ISSN 2095-8137
CN 53-1229/Q
Volume 46 Issue 1
18 January 2025

ZR

Zoological Research



CODEN: DOYADI
www.zoores.ac.cn

KUNMING INSTITUTE OF ZOOLOGY, CAS
CHINA ZOOLOGICAL SOCIETY
Science Press

ZOOLOGICAL RESEARCH

Volume 46, Issue 1 18 January 2025

CONTENTS

The golden era of scientific publishing in China ····· *Yong-Gang Yao*, *Yong-Tang Zheng* (1)

Palmitoylation-mediated NLRP3 inflammasome activation in teleosts highlights evolutionary divergence in immune regulation ····· *Li Nie*, *Xiang-Yu Wu*, *Zi-Yue Zhao*, *Chen-Jie Fei*, *Ting-Fang Zhu*, *Jian-Zhong Shao*, *Jiong Chen* (3)

Enhanced risk assessment framework integrating distribution dynamics, genetically inferred populations, and morphological traits of *Diploderma* lizards ····· *Qi Xiao*, *Xiu-Dong Shi*, *Lin Shi*, *Zhong-Yi Yao*, *You-Hua Chen*, *Wei-Zhao Yang*, *Zi-Yan Liao*, *Yin Qi* (15)

Both 20S and 19S proteasome components are essential for meiosis in male mice ····· *Ting-Ting Han*, *Li-Ying Wang*, *Qiu-Xing Zhou*, *Wei Wei*, *Yan-Jie Ma*, *Ying-Hong Chen*, *Wei Li*, *Zhen-Yu Ju*, *Chao Liu* (27)

Novel mouse model of Alzheimer's disease exhibits pathology through synergistic interactions among amyloid-β, tau, and reactive astrogliosis ····· *Young-Eun Han*, *Sunhwa Lim*, *Seung Eun Lee*, *Min-Ho Nam*, *Soo-Jin Oh* (41)

Ribosome profiling and single-cell RNA sequencing identify the unfolded protein response as a key regulator of pigeon lactation ····· *Jing Liu*, *San-Feng Liu*, *Hui-Rong Mao*, *Hong-Xia Jiang*, *Shui-Bing Liu*, *Xiao-Fei Xu*, *Jin-Tao Wu*, *Xun Liu*, *Wen-Tao Zhang*, *Xiao-Long Hu*, *Biao Chen* (54)

Transposable elements shape the landscape of heterozygous structural variation in a bird genome ··· *Bo-Ping Li*, *Na Kang*, *Zao-Xu Xu*, *Hao-Ran Luo*, *Shi-Yu Fan*, *Xiao-Han Ao*, *Xing Li*, *Ya-Peng Han*, *Xiao-Bin Ou*, *Luo-Hao Xu* (75)

DNA2 knockout aggravates cerebral ischemia/reperfusion injury by reducing postsynaptic Homer1a ····· *Ting Ma*, *Yu-Meng Li*, *Peng-Yu Ren*, *Shi-Quan Wang*, *Xiang-Long Liu*, *Wen-Bo Lv*, *Wu-Gang Hou*, *Wen-Qiang Zuo*, *Wei-Qiang Lin*, *Jian Sima*, *An-Qi Geng* (87)

Identification of novel rodent and shrew orthohepeviruses sheds light on hepatitis E virus evolution ····· *Fu-Li Li*, *Bo Wang*, *Pei-Yu Han*, *Bei Li*, *Hao-Rui Si*, *Yan Zhu*, *Hong-Min Yin*, *Li-Dong Zong*, *Yi Tang*, *Zheng-Li Shi*, *Ben Hu*, *Xing-Lou Yang*, *Yun-Zhi Zhang* (103)

Transcription coactivator YAP1 promotes CCND1/CDK6 expression, stimulating cell proliferation in cloned cattle placentas ····· *Shan-Shan Wu*, *Xiao-Yu Zhao*, *Lei Yang*, *Chao Hai*, *Di Wu*, *Xue-Fei Liu*, *Li-Shuang Song*, *Chun-Ling Bai*, *Guang-Hua Su*, *Guang-Peng Li* (122)

Reciprocal translocation experiments reveal gut microbiome plasticity and host specificity in a Qinghai-Xizang Plateau lizard ····· *Wei Yu*, *Jing Yang*, *Li-Wei Teng*, *Xiao-Long Zhao*, *Ze-Yu Zhu*, *Shuang Cui*, *Wei-Guo Du*, *Zhen-Sheng Liu*, *Zhi-Gao Zeng* (139)

sRNA113 regulates *Pseudomonas plecoglossicida* motility to affect immune response against infection in pearl gentian grouper ····· *Li He*, *Mei-Qin Mao*, *Ling-Min Zhao*, *Qi Li*, *Hui Ge*, *Jiao-Nan Zhang*, *Jiao-Lin Zhang*, *Qing-Pi Yan* (152)

Deciphering the toxic effects of polystyrene nanoparticles on erythropoiesis at single-cell resolution ····· *Eun Jung Kwon*, *Hyeon Mi Sung*, *Hansong Lee*, *Soyul Ahn*, *Yejin Kim*, *Chae Rin Lee*, *Kihun Kim*, *Kyungjae Myung*, *Won Kyu Kim*, *Dokyoung Kim*, *Sanghwa Jeong*, *Chang-Kyu Oh*, *Yun Hak Kim* (165)

Multi-omics analysis and experimental verification reveal testicular fatty acid metabolism disorder in non-obstructive azoospermia ····· *Zhou Li*, *Yi-Jian Xiang*, *Zhi-Chuan Zou*, *Yu-Ming Feng*, *Hui Wang*, *Wei-Qing Chen*, *Xie Ge*, *Jin-Zhao Ma*, *Jun Jing*, *Bing Yao* (177)

Cross-species single-cell transcriptomics reveals neuronal similarities and heterogeneity in amniote pallium ····· *Fu-Bao-Qian Huang*, *Kuo Liao*, *Yu-Nong Sun*, *Zi-Hao Li*, *Yan-Ru Zhang*, *Ping-Fang Liao*, *Si-Yuan Jiang*, *Zhi-Yong Zhu*, *Duo-Yuan Chen*, *Ying Lei*, *Shi-Ping Liu*, *You-Ning Lin*, *Zhen-Kun Zhuang* (193)

Intranasal iron administration induces iron deposition, immunoactivation, and cell-specific vulnerability in the olfactory bulb of C57BL/6 mice ····· *Xiao-Qing Mi*, *Bao-Chen Liu*, *Le Qu*, *Yu Yuan*, *Hui Li*, *Ao-Yang Xu*, *Yu-Lin Zhang*, *Jun-Xia Xie*, *Ning Song* (209)

Elevated CXCL1 triggers dopaminergic neuronal loss in the substantia nigra of C57BL/6J mice: Evaluation of a novel Parkinsonian mouse model ····· *Xi-Zhen Ma*, *Guo-Rui Jia*, *Meng-Yu Li*, *Sheng-Han Zhang*, *Zhao-Xin Wang*, *Ning Song*, *Ying-Juan Liu*, *Jun-Xia Xie* (225)

Functional evolution of thyrotropin-releasing hormone neuropeptides: Insights from an echinoderm ····· *Yingqiu Zheng*, *Huachen Liu*, *Xin Dang*, *Juan Diego Gaitán-Espitia*, *Muyan Chen* (236)

Advances and applications of genome-edited animal models for severe combined immunodeficiency ····· *Xiao Zheng*, *Chun-Hui Huang*, *Sen Yan*, *Ming-Deng Rong* (249)

Cover image: The golden pheasant (*Chrysolophus pictus*). Photo by Hua Ding. **Design**: Li-Bin Wu

Inside back cover design: Li-Bin Wu

Niu et al. *Zool. Res.* 2024, 45(1): 1−12
https://doi.org/10.24272/j.issn.2095-8137.2023.171

ZR | Zoological Research

Article Open Access

Surviving winter on the Qinghai-Xizang Plateau: Extensive reversible protein phosphorylation plays a dominant role in regulating hypometabolism in hibernating *Nanorana parkeri*

Yong-Gang Niu[1,*], Deng-Bang Wei[2], Xue-Jing Zhang[3], Ti-Sen Xu[1], Xiang-Yong Li[1], Hai-Ying Zhang[1], Zhi-Fang An[2], Kenneth B. Storey[4], Qiang Chen[3]

[1] *Department of Life Sciences*, *Dezhou University*, *Dezhou*, *Shandong* 253023, *China*
[2] *State Key Laboratory of Plateau Ecology and Agriculture*, *Qinghai University*, *Xining*, *Qinghai* 810016, *China*
[3] *School of Life Sciences*, *Lanzhou University*, *Lanzhou*, *Gansu* 730000, *China*
[4] *Department of Biology*, *Carleton University*, *Ottawa*, *ON K1S 5B6*, *Canada*

ABSTRACT

Changes in protein abundance and reversible protein phosphorylation (RPP) play important roles in regulating hypometabolism but have never been documented in overwintering frogs at high altitudes. To test the hypothesis that protein abundance and phosphorylation change in response to winter hibernation, we conducted a comprehensive and quantitative proteomic and phosphoproteomic analysis of the liver of the Xizang plateau frog, *Nanorana parkeri*, living on the Qinghai-Xizang Plateau. In total, 5170 proteins and 5695 phosphorylation sites in 1938 proteins were quantified. Based on proteomic analysis, 674 differentially expressed proteins (438 up-regulated, 236 down-regulated) were screened in hibernating *N. parkeri* versus summer individuals. Functional enrichment analysis revealed that higher expressed proteins in winter were significantly enriched in immune-related signaling pathways, whereas lower expressed proteins were mainly involved in metabolic processes. A total of 4251 modified sites (4147 up-regulated, 104 down-regulated) belonging to 1638 phosphoproteins (1555 up-regulated, 83 down-regulated) were significantly changed in the liver. During hibernation, RPP regulated a diverse array of proteins involved in multiple functions, including metabolic enzymatic activity, ion transport, protein turnover, signal transduction, and alternative splicing. These changes contribute to enhancing protection, suppressing energy-consuming processes, and inducing metabolic depression. Moreover, the activities of phosphofructokinase, glutamate dehydrogenase, and ATPase were all significantly lower in winter compared to summer. In conclusion, our results support the hypothesis and demonstrate the importance of RPP as a regulatory mechanism when animals transition into a hypometabolic state.

Keywords: *Nanorana parkeri*; Proteomic; Phosphoproteomic; Hibernation; Reversible protein phosphorylation; Metabolism

INTRODUCTION

Winter dormancy and hibernation are common survival strategies used by many species, including both endo- and ectotherms, to cope with prolonged cold exposure and poor food availability (Pinder et al., 1992; Tattersall & Ultsch, 2008). By entering dormancy or hibernation states, animals can reduce their metabolic rates, extend the duration for which stored food or endogenous energy reserves sustain survival, and retreat to safe locations conducive to protection. Depression of the metabolic rate involves the suppression of various biochemical (transcription, translation, cell cycle) and physiological processes (heart rate, respiratory rate) (Boutilier et al., 1997; Donohoe & Boutilier, 1998; Storey & Storey, 2004, 2007). Enhancement of defense mechanisms, such as the up-regulation of antifreeze proteins, antioxidant enzymes, and chaperone proteins, is also essential for long-term survival under hypometabolic conditions (Storey & Storey, 2007). This state allows for substantial energy conservation through targeted inhibition of various cellular and metabolic processes and decreased substrate utilization (Storey & Storey, 1990, 2004). Despite this strong suppression, essential metabolic functions must be retained to enable rapid

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 05 July 2023; Accepted: 11 September 2023; Online: 12 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (32001110), Training Program for Cultivating High-level Talents by the China Scholarship Council (2021lxjjw01), and Open Project of State Key Laboratory of Plateau Ecology and Agriculture, Qinghai University (2021-KF-004)

*Corresponding author, E-mail: yonggangniu@126.com

restoration upon arousal (Storey & Storey, 1990, 2004). Thus, hypometabolism is generally characterized not by a significant reduction in metabolic capacity but by reversible control mechanisms (Storey & Storey, 2004).

The molecular mechanisms that underlie the hibernation phenotype have been elucidated in a variety of ectothermic vertebrates, including the Chinese alligator (*Alligator sinensis*) (Lin et al., 2020; Sun et al., 2018; Zhang et al., 2021), Australian central bearded dragon (*Pogona vitticeps*) (Capraro et al., 2019), Asiatic toad (*Bufo gargarizans*) (Jin et al., 2018), and wood frog (*Rana sylvatica*) (Kiss et al., 2011). In the case of *A. sinensis*, integrative multi-omics analysis revealed a down-regulation of pathways related to nutrient absorption and metabolism, muscle contraction, urinary excretion, and immunity function during hibernation (Lin et al., 2020). In addition, transcriptomic and proteomic studies of *P. vitticeps* revealed enrichment of protective mechanisms across various tissues during hibernation, including skeletal muscle atrophy resistance, neuroprotection in the brain, and cardiac hypertrophic processes (Capraro et al., 2019). Transcriptomic analysis of overwintering female toads found significant down-regulation of most genes in the liver during hibernation, involving metabolic depression and shifts in energy utilization (Jin et al., 2018). For *R. sylvatica*, hepatoproteomic analysis revealed significant up-regulation of proteins involved in cytoprotection (heat shock proteins and antioxidants) in the liver during hibernation, with significant down-regulation of proteins involved in cell proliferation, protein synthesis, and mitochondrial function (Kiss et al., 2011). Although the reorganization of specific proteomic components is crucial for survival under a hypometabolic state, the coordinated inhibition of non-essential or energy-intensive cellular functions is also crucial for successful overwintering (Storey & Storey, 2013). The extensive degradation of proteins and mRNA transcripts during dormant states would be counterproductive for organisms, as these components are immediately required upon arousal from torpor or rewarming. Therefore, mechanisms that can reversibly and comprehensively inhibit metabolic functions while differentially regulating various cellular processes are essential. One such efficient control mechanism is reversible protein phosphorylation (RPP). This fast-acting and easily reversible control mechanism necessitates minimal expenditure of adenosine triphosphate (ATP) and serves as a regulatory mechanism for fuel metabolism, ion channel inactivation, and protein synthesis inhibition during hypometabolic states (Storey & Storey, 2004, 2007). For example, various enzymes, such as glucose-6-phosphate dehydrogenase, catalase, creatine kinase, and hexokinase, are regulated by reversible phosphorylation in overwintering and/or frozen wood frogs (Dawson & Storey, 2016; Dieni & Storey, 2009, 2010, 2011). Although many studies have addressed the molecular mechanisms of overwintering in frog species inhabiting North America or high-latitude regions, the mechanisms governing metabolic depression in overwintering species residing in high-altitude regions remain unclear.

The Xizang plateau frog (*Nanorana parkeri*) (Anura: Dicroglossidae) inhabits the Qinghai-Xizang Plateau at elevations ranging from 2 850 to 5 100 m above sea level (a.s.l.), representing the highest known altitude of any amphibian species (Ma et al., 2009). As such, this species has been extensively used to explore adaptations to extreme high-altitude environments (Sun et al., 2015). Our previous studies on *N. parkeri* primarily focused on its physiological and biochemical mechanisms of adaptation to high altitude, as well as the physiological ecology of its winter hibernation (Niu et al., 2022a, 2022b). For example, *N. parkeri* individuals living at high elevations exhibit modified hematological parameters, such as higher blood hemoglobin content, increased hematocrit values, and higher red cell counts, in addition to stronger antioxidant defenses and reduced oxidative damage compared to low-elevation residing frogs (Niu et al., 2022b). Furthermore, during hibernation, *N. parkeri* demonstrates greater body mass, body mass index, liver index, liver glycogen concentration, and plasma bacteria-killing ability, as well as higher levels of oxidative stress but lower antioxidant defenses compared to summer-active individuals (Niu et al., 2018, 2022a). These frogs hibernate in shallow ponds within damp caves and display reduced freeze tolerance compared to *R. sylvatica* (Larson et al., 2014), although freeze exposure triggers substantial changes in liver metabolic profiles and tissue-specific activation of antioxidant systems (Niu et al., 2021a, 2021b). Moreover, acute warming leads to oxidative stress and a significant increase in total antioxidant capacity in most tissues in hibernating *N. parkeri* (Zhang et al., 2022a). Overwintering *N. parkeri* frogs also exhibit seasonal and temperature-independent metabolic depression, accompanied by marked changes in plasma metabolomic profiles (Niu et al., 2020, 2021b). Furthermore, integrated transcriptomic and metabolomic analyses have shown that adaptive regulation occurs in gene expression and metabolic processes of the liver in overwintering *N. parkeri* (Niu et al., 2023). However, information on the regulatory roles of protein abundance and RPP in hypometabolism in overwintering *N. parkeri* remains limited.

In the present study, we hypothesized that the adaptation of *N. parkeri* frogs to winter stress (such as low temperature, low oxygen, lack of food) also necessitates adjustments in protein abundance for maintenance of homeostasis, as well as reversible mechanisms (e.g., protein phosphorylation) for the differential regulation of metabolic processes in winter conditions. To test this hypothesis, we used quantitative proteomic and phosphoproteomic analyses to assess the seasonal variations in the abundance and phosphorylation status of proteins in the livers of *N. parkeri*. The liver was chosen as it is a key organ for the synthesis, metabolism, storage, and redistribution of carbohydrates, proteins, and lipids. The findings of this study provide insights into the macromolecular responses and regulatory roles of reversible phosphorylation that contribute to the winter survival of this species.

MATERIALS AND METHODS

Animals and sample collection

Adult male *N. parkeri* frogs (n=9 in each season) were captured and their liver tissues were collected, as reported previously (Niu et al., 2023). In brief, liver tissue samples were dissected and snap-frozen in liquid nitrogen, followed by storage at −80 °C in the laboratory. The mean snout-vent length (SVL) of frogs was 4.08±0.06 cm and mean body mass (BM) was 4.99±0.14 g in summer (n=9); comparable data were 4.06±0.06 cm and 5.14±0.16 g in winter. All animal procedures were approved by the Ethics Committee of Animal Experiments at Dezhou University (Approval No. DZXY2021003) and met the guidelines of the China Council on Animal Care.

Protein extraction and trypsin digestion

Protein extraction and trypsin digestion were performed as described by Lu et al. (2022) with minor modifications. Briefly, frozen samples (n=3 for each season) were fully powdered under liquid nitrogen and then sonicated in lysis buffer (8 mol/L urea (Sigma-Aldrich, USA), 1% protease inhibitor (Merck Millipore, USA), and 1% phosphatase inhibitor (Millipore, USA); w:v=1:4), followed by centrifugation at 12 000 ×g for 10 min at 4 °C. Supernatants were collected to determine protein concentration using a BCA kit (Beyotime Biotechnology, China). Supernatants were added to 20% (w:v) trichloroacetic acid (TCA) (Sigma-Aldrich, USA), followed by incubation at 4 °C for 2 h. After centrifugation at 4 500 ×g for 5 min at 4 °C, the precipitates were collected and washed three times with cooled acetone, then dissolved in 200 mmol/L tetraethylammonium bromide (TEAB) (Sigma-Aldrich, USA). For digestion, trypsin (Promega, USA) (w:w=1:50) was added and left overnight. Samples were then reduced with 5 mmol/L dithiothreitol (Sigma-Aldrich, USA) for 60 min at 37 °C and alkylated with 11 mmol/L iodoacetamide (Sigma-Aldrich, USA) for 45 min at room temperature in darkness. Finally, the peptides were desalted using a Strata X SPE column (Dr. Maisch GmbH, GER) and vacuum-dried.

Tandem mass tag (TMT) labeling, high-performance liquid chromatography (HPLC) fractionation, and affinity enrichment of phosphopeptides

Peptides from each sample were labeled with TMT reagent (Thermo Fisher Scientific, USA) and fractionated by high-pH reverse-phase HPLC (Thermo Fisher Scientific, USA), as reported in Lu et al. (2022) with minor modification. Briefly, the peptides were separated with a gradient of 8%–32% acetonitrile (Thermo Fisher Scientific, USA) in 10 mmol/L ammonium bicarbonate (Sigma-Aldrich, USA) (pH=9) over 60 min into 60 fractions, then combined into nine fractions and dried by vacuum centrifugation for liquid chromatography-tandem mass spectrometry (LC-MS-MS). Enrichment of phosphopeptides was carried out according to Lu et al. (2022).

LC-MS/MS analysis

LC-MS/MS analyses were performed as reported by Lu et al. (2022) with minor modification. Briefly, after dissolution in mobile phase A (0.1% formic acid, 2% acetonitrile in water, Thermo Fisher Scientific, USA), the peptides were separated using the EASY-nLC 1200 ultra-high-performance liquid system (Thermo Fisher Scientific, USA), followed by ionization in an NSI source and analysis using the Q Exactive™ HF-X system (Thermo Fisher Scientific, USA). The mass spectrometry scan range was 350–1 600 m/z with a resolution of 120 000 and a second-order mass scan resolution of 30 000. The top 20 peptide parent ions with the highest signal intensity were used for higher-energy collisional dissociation (HCD) fragmentation at a normalized collision energy (NCE) of 28%. Mobile phase B contained 0.1% formic acid and 90% acetonitrile. For proteomic analysis, the elution settings were: 40 min, 6%–22% B; 14 min, 22%–32% B; 3 min, 32%–80% B; 3 min, 80% B, with a constant flow rate of 500 nL/min. For phosphoproteomic analysis, the gradient was as follows: 38 min, 5%–22% B; 15 min 22%–35% B; 4 min 35%–80% B; 3 min, 80% B, with a flow rate of 350 nL/min at 0–53 min and 600 nL/min at 53–60 min.

Proteomic and phosphoproteomic database search

The raw MS/MS data were processed using the MaxQuant search engine (v.1.5.2.8), and tandem mass spectra were searched against the *Nanorana_parkeri* database (37 248 entries) concatenated with the reverse decoy database. The database search parameters were set as described in Zhang et al. (2022b). Briefly, trypsin/P was designated as a cleavage enzyme and missing cleavages of less than 2 were permitted. The mass error for precursor ions was set at 5 ppm in the main search and 20 ppm in the first search, and a value of 0.02 Da was used for mass error of fragment ions. Carbamidomethyl on Cys was specified as a fixed modification, and variable modifications included acetylation on protein N-terminal, oxidation on Met, deamidation (NQ), and phosphorylation on Ser, Thr, and Tyr. TMT-6plex was performed for quantification. The false discovery rate (FDR) was adjusted to <1% and the minimum score for peptides was set to >40.

Bioinformatics analysis for proteomics and phosphoproteomics

Gene Ontology (GO) annotation of proteomic and phosphoproteomic data was conducted using the UniProt-GOA database (http://www.ebi.ac.uk/GOA/) and included three categories: biological processes, cellular components, and molecular functions. Protein domains and pathways were annotated using the InterPro (http://www.ebi.ac.uk/interpro/) database and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (http://www.kegg.jp/), respectively. Subcellular localization was predicted using WoLF PSORT software (v.2.0), and sequence patterns around the phosphorylation modification site (spanning six amino acids both upstream and downstream) were identified using Soft motif-x (v.5.0.2). Two-tailed Fisher's exact test was used to test functional enrichment of differentially expressed proteins with a critical corrected P-value of 0.05. Functional interaction network analysis of differentially expressed proteins and phosphorylated proteins was performed using STRING (v.11.0) and visualized using Cytoscape (v.3.9.1).

Enzymatic activity assays

All enzymatic activities were determined using an assay kit (Solarbio Science & Technology, China). Frozen liver tissues (n=6 for each season) were weighed and immediately homogenized on ice according to the manufacturer's instructions. The resulting supernatant was immediately used to determine the activities of Na^+/K^+-ATPase, Ca^{2+}-ATPase, phosphofructokinase (PFK; an important rate-controlling enzyme in the glycolytic pathway), and glutamate dehydrogenase (GDH). All enzymatic assays were conducted at 25 °C using an automatic microplate reader (BioTek Instruments, USA) and were run in triplicate.

Statistical analysis

All data were analyzed using SPSS v.20.0 (SPSS, USA) and expressed as mean±standard error of the mean (SEM). After verification of normality and homogeneity of variances to satisfy the assumptions for parametric tests, differences in enzyme activity between the two groups were evaluated using Student's t-tests for independent samples. A threshold level of P<0.05 was used to establish statistical significance.

RESULTS

Summary of quantitative proteome and protein phosphoproteome analysis

Principal component analysis (PCA) of the proteomic and

phosphoproteomic datasets showed good clustering between biological replicates, as well as a clear summer and winter separation (Figure 1A, B). In total, 5 411 protein groups were identified, from which 5 170 proteins were quantified (Supplementary Table S1). Altogether, 12 299 phosphorylation sites in 4 119 protein groups were identified, among which 5 695 sites in 1 938 proteins were quantified (Supplementary Table S2).

Analysis of differentially expressed proteins and phosphorylated proteins

Based on established cut-off criteria (fold-change (FC)>1.5 (up-regulated) or <0.667 (down-regulated) and P<0.05), a total of 674 proteins were differentially expressed, among which 438 proteins (65%) showed higher abundance and 236 proteins (35%) showed lower abundance in winter compared to summer (Figure 1C; Supplementary Table S3). Phosphorylation levels changed significantly in 4 251 modified sites belonging to 1 638 phosphoproteins (FC>1.5 (up-regulated) or <0.667 (down-regulated) and P<0.05). Among them, 4 147 sites in 1 555 proteins were up-regulated and 104 sites in 83 proteins were down-regulated in the liver of hibernating *N. parkeri* compared to summer-active individuals (Figure 1D; Supplementary Table S4). The phosphorylation levels of ATPases showed a significant increase in overwintering *N. parkeri*, including plasma membrane calcium-transporting ATPase 1, v-type proton ATPase subunit d 1, v-type proton ATPase catalytic subunit A, cation-transporting ATPase, sarcoplasmic/endoplasmic reticulum calcium ATPase 1, sodium/potassium-transporting ATPase subunit alpha-1, and transitional endoplasmic reticulum ATPase. Based on motif-x analysis, 100 different phosphorylation motifs were identified from hibernation-regulated phosphopeptides (Supplementary Table S5). Among the top 10 phosphorylation motifs ranked by fold increase (Table 1), aspartic acid was conserved at the −1 position, while those motifs containing aspartic and glutamic acids at the +1, +2, or +3 positions of the central serine residue were highly enriched. Motifs containing arginine at the −3 position of the central serine and tyrosine residues were also strongly conserved.

Functional analysis of differentially quantified proteins

GO enrichment analysis indicated that higher expressed proteins in hibernating *N. parkeri* were enriched in cell-cell

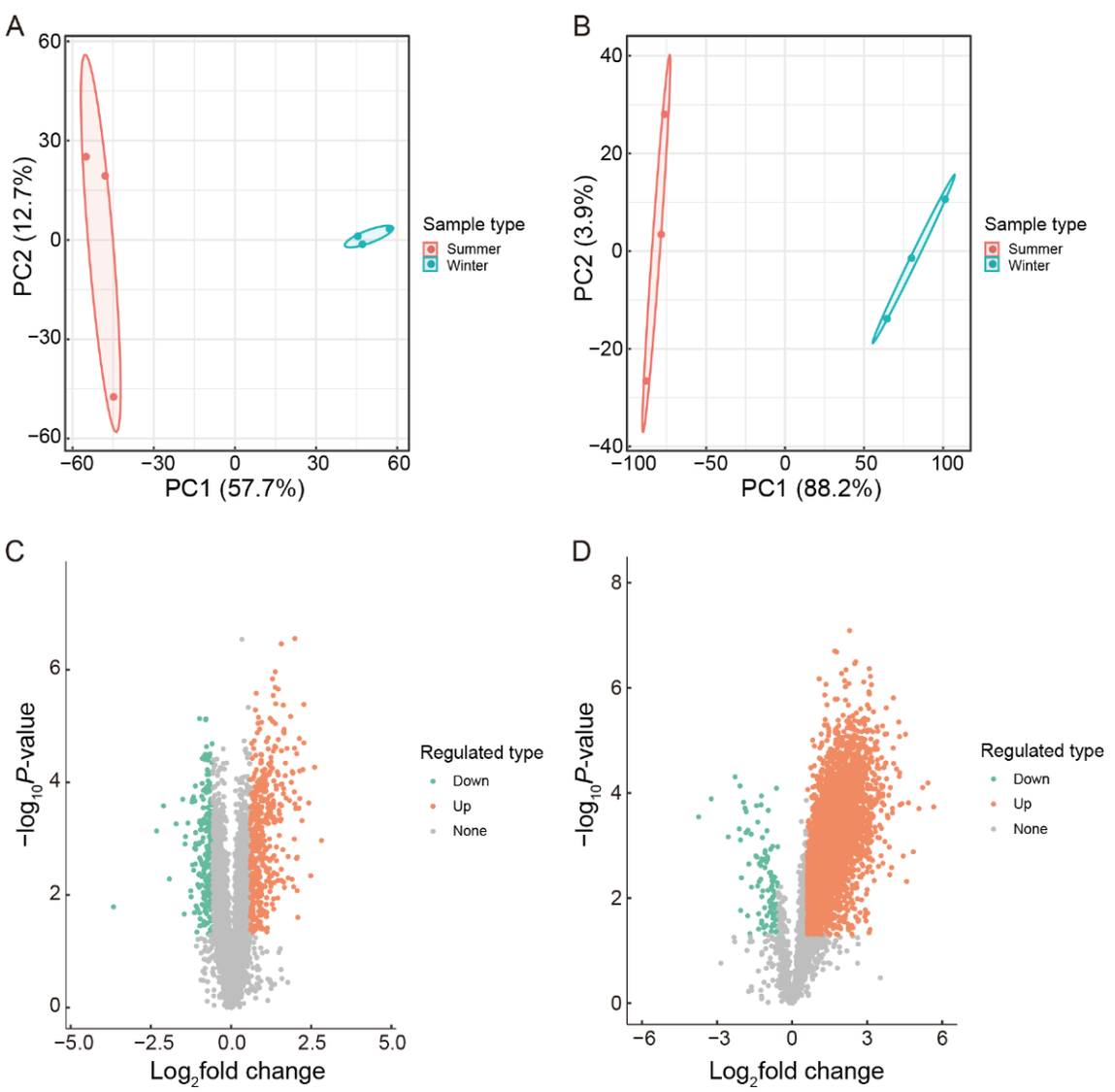


Figure 1 PCA and volcano plots showing differentially expressed proteins and phosphorylated proteins

A, C: Proteomic analysis. B, D: Phosphoproteomic analysis. Screening criteria were FC>1.5 (up-regulated, red) or <0.667 (down-regulated, green) and P<0.05.

Table 1 Motif-x analysis of significantly different phosphopeptides in *Nanorana parkeri* liver between summer and winter

Motif logo	Motif	Motif score	Foreground		Background		Fold increase
			Matches	Size	Matches	Size	
	KxxxxD_S_ExExxx	54.17	25	7 919	51	1 005 390	62.2
	xxxxxx_S_DDDxxx	47.05	59	6 676	322	980 303	26.9
	xxxxxD_S_DxExxx	48.00	117	9 266	493	1 024 378	26.2
	xxxxxx_T_PPxxxx	32.00	167	1 375	3349	680 998	24.7
	xxxxxx_S_DDExxx	44.42	66	8 344	352	1 012 577	22.8
	xxxxxD_S_DxDxxx	40.76	35	5 657	282	952 528	20.9
	xRxRxx_S_xDxxxx	39.02	40	8 971	223	1 020 862	20.4
	xxxRxx_S_PxPxxx	48.00	76	10 662	366	1 041 899	20.3
	xxxRRx_T_xxxxxx	32.00	68	1 030	2196	669 591	20.1
	xxxxxD_S_xEDxxx	43.65	45	7 146	312	990 174	20.0

junction organization, regulation of cell morphogenesis, and cell junction assembly in the biological process category (Supplementary Figure S1A), double-stranded DNA binding, cytoskeletal protein binding, and DNA binding in the molecular function category (Supplementary Figure S1B), and in cell junction, anchoring junction, and adherens junction in the cellular component category (Supplementary Figure S1C). KEGG pathway enrichment analysis showed that higher expressed proteins in overwintering *N. parkeri* were mainly enriched in tight junction, vasopressin-regulated water reabsorption (Supplementary Figure S1D), and signaling pathways, including the Toll and Imd, T cell receptor, Toll-like receptor, RIG-I-like receptor, IL-17, and Ras signaling pathways (Figure 2A).

Proteins with lower expression in the liver of hibernating *N. parkeri* were mainly enriched in oxidative phosphorylation, respiratory electron transport chain, cellular respiration, and ribose phosphate metabolic process in the biological process category (Supplementary Figure S2A), NADH dehydrogenase activity, oxidoreductase activity, and electron transfer activity in the molecular function category (Supplementary Figure S2B), and respirasome, respiratory chain complex, mitochondrial respirasome, and mitochondrial membrane in the cellular component category (Supplementary Figure S2C). KEGG pathway enrichment analysis showed that lower expressed proteins in overwintering *N. parkeri* were primarily enriched in metabolism-related pathways, including thermogenesis, oxidative phosphorylation, steroid hormone biosynthesis, retinol metabolism, primary bile acid biosynthesis, linoleic acid metabolism, cholesterol metabolism, and peroxisome proliferator-activated receptor (PPAR) signaling (Supplementary Figure S2D; Figure 2B).

Functional analysis of differentially phosphorylated proteins

GO enrichment analysis showed that the higher abundances of phosphorylated proteins in hibernating *N. parkeri* were enriched in carboxylic acid metabolic process, carboxylic acid biosynthetic process, cellular amino acid metabolic process, Golgi vesicle transport, and lipid oxidation in the biological process category (Supplementary Figure S3A), oxidoreductase activity, mRNA binding, and cell adhesion molecule binding in the molecular function category (Supplementary Figure S3B), and mitochondrial matrix, intercalated disc, and endocytic vesicle in the cellular component category (Supplementary Figure S3C). For KEGG pathways analysis, a higher abundance of phosphorylated proteins in hibernating *N. parkeri* were enriched in metabolic pathways (such as valine, leucine and isoleucine degradation, fatty acid degradation, glycolysis/gluconeogenesis, and tryptophan metabolism), hypoxia-inducible factor-1 (HIF-1) signaling pathway, and protein processing in endoplasmic reticulum and protein export (Figure 3; Supplementary Figure S3D).

Proteins expressed at lower levels were mainly enriched in ribosome biogenesis, glucan metabolic processes, and energy reserve metabolic processes in the biological process category (Supplementary Figure S4A), rRNA methyltransferase activity, catalytic activity, acting on an rRNA, and poly(G) binding in the molecular function category (Supplementary Figure S4B), and mitotic spindle midzone and nephrocyte diaphragm in the cellular component category (Supplementary Figure S4C). KEGG pathway enrichment analysis showed that lower expressed proteins were enriched in starch and sucrose metabolism, metabolism of xenobiotics by cytochrome P450, and pathogenic *Escherichia coli* infection (Supplementary Figure S4D).

Protein-protein interaction (PPI) networks of differentially expressed proteins and phosphorylated proteins

PPI network analysis revealed significant enrichment in differentially expressed proteins and phosphorylated proteins in five major connected subnetworks, including oxidative phosphorylation, ubiquitin-mediated proteolysis, spliceosomes, fatty acid elongation, and RNA transport (Figure 4). In oxidative phosphorylation, the expression of most proteins was significantly decreased, but the phosphorylation levels of proteins did not change significantly.

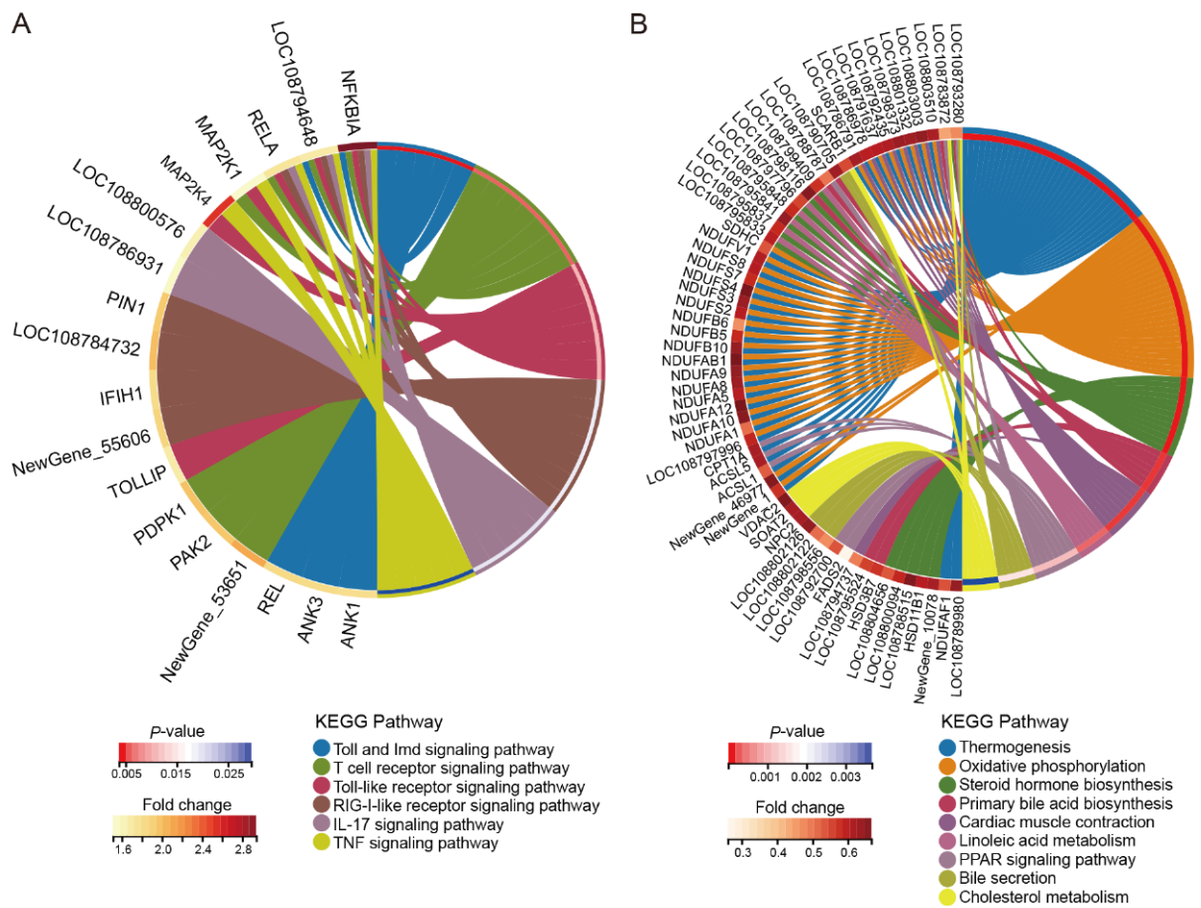


Figure 2 Circle diagram showing KEGG enrichment analysis of differentially expressed proteins

A: Up-regulated proteins related to signaling pathways. B: Down-regulated proteins related to nutrient absorption and energy metabolism pathways.

In contrast, for ubiquitin-mediated proteolysis, spliceosomes, fatty acid elongation, and RNA transport, phosphorylation levels increased substantially without apparent changes in protein expression, suggesting that protein phosphorylation plays an important regulatory role in these processes.

Enzymatic activities

Overall, the activity of the four enzymes was significantly lower in winter compared to summer. Na^+/K^+ ATPase and Ca^{2+} ATPase activities were 34% and 75% lower, respectively, in the liver of hibernating *N. parkeri* frogs than those in summer-active individuals. Moreover, PFK and GDH activities decreased by 58% and 73%, respectively, in winter compared to summer (Figure 5).

DISCUSSION

To the best of our knowledge, this study provides the first evidence of changes in the liver proteome and phosphorylated proteome during hibernation in a high-altitude frog (*N. parkeri*). Remarkably, in hibernating frogs, regulated changes were observed in both nutrient absorption and energy metabolism processes. These changes were affected though the down-regulation of protein levels as well as the utilization of RPP to modulate energy metabolism, ion motive ATPases, protein turnover, signal transduction cascades, and alternative splicing. In addition, the expression of proteins associated with protective mechanisms (a conserved response often seen in other hypometabolic states), such as chaperone proteins and antioxidant defense enzymes, was not extensively increased in hibernating *N. parkeri* (Hermes-Lima et al., 1998; Wu et al., 2018). These results highlight the commonality and specificity of regulatory mechanisms between hibernating *N. parkeri* and other frog species.

Nutrient absorption and energy metabolism

Nanorana parkeri frogs overwinter in shallow ponds within damp cold caves for up to 6 months without feeding (Niu et al., 2021a). Thus, not only is digestion in the stomach and intestines halted, but pathways associated with nutrient absorption and processing by the liver are also down-regulated. Indeed, our results showed that the expression levels of proteins involved in pathways related to primary bile acid biosynthesis and bile secretion were significantly down-regulated in hibernating *N. parkeri*. This observation aligns with our earlier transcriptomic and metabolomic data, showing a significant reduction in pathway-related genes, such as *ACOT8*, *HSD3B7*, and *AMACR*, and cholesterol (a precursor of bile acid synthesis), in winter (Niu et al., 2023). Similar results have also been found for the Chinese alligator (*Alligator sinensis*) (Lin et al., 2020) and Chinese soft-shelled turtle (*Pelodiscus sinensis*) (Huang et al., 2021). Cellular functions involving both ATP utilization and production are inhibited in all forms of natural hypometabolism, especially "optional" cellular functions (e.g., gene expression, protein turnover, cell proliferation, growth) (Storey & Storey, 1990, 2004, 2007). In the current study, proteins with reduced expression levels in winter were predominantly associated with the oxidative phosphorylation pathway. This is in line with our prior transcriptomic study showing a down-regulation of the ATP supply pathway (oxidative phosphorylation) in the

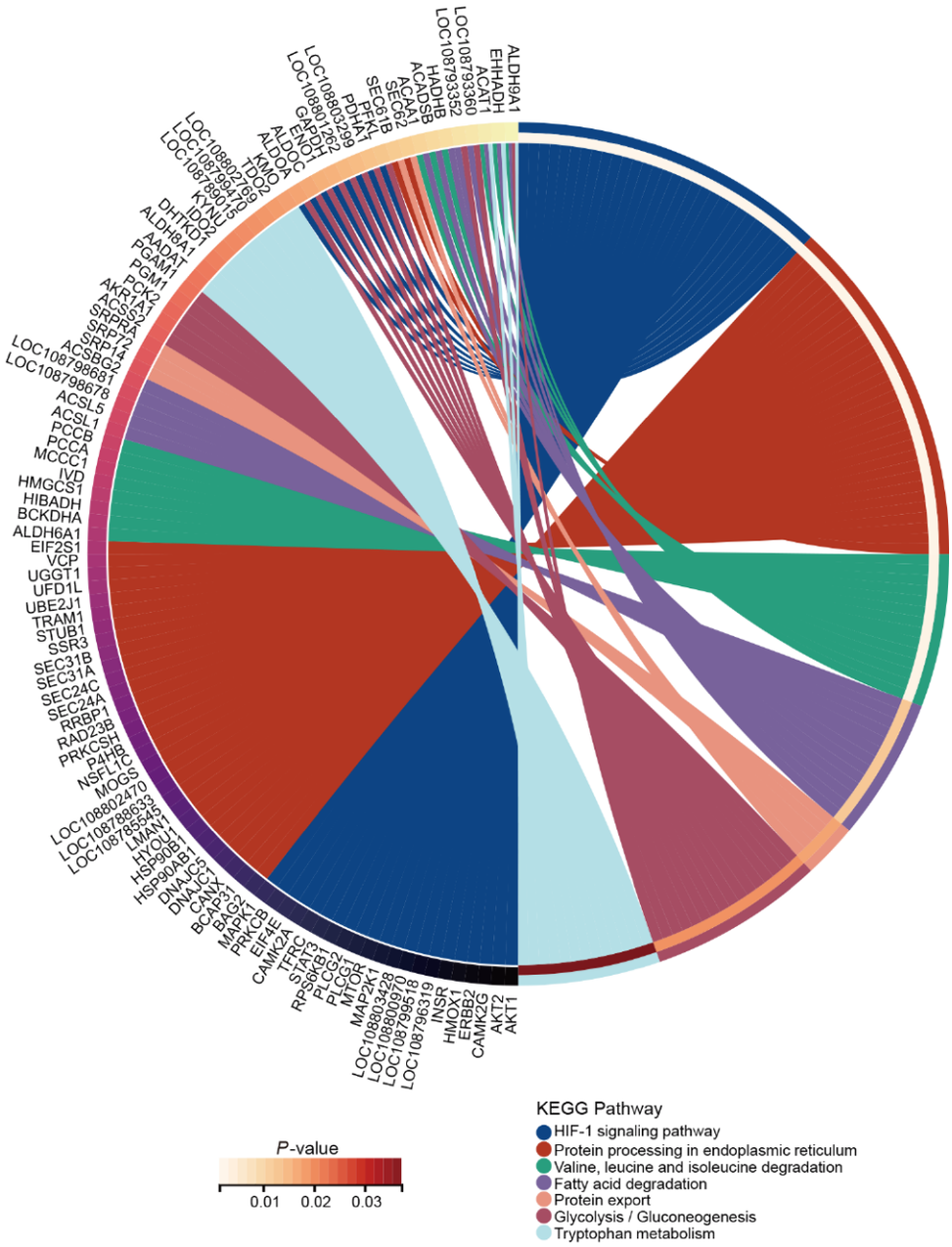


Figure 3 Circle diagram showing KEGG enrichment analysis of up-regulated phosphorylated proteins related to signaling pathways, protein synthesis, and energy metabolism pathways in hibernating *Nanorana parkeri*

liver of winter-collected frogs (Niu et al., 2023). Similar results have also been observed in overwintering *A. sinensis*, in which the expression levels of most genes involved in the oxidative phosphorylation pathway are significantly reduced in winter (Lin et al., 2020).

We also found that proteins involved in several metabolic pathways showed higher phosphorylation levels during hibernation, including the oxidative phosphorylation pathway, fatty acid degradation, glycolysis/gluconeogenesis, and valine, leucine, and isoleucine degradation. Increased phosphorylation and lower activity of glycolytic enzymes have also been reported in freeze-tolerant *R. sylvatica* after 24 h of freezing exposure (Hawkins et al., 2019). Our results also showed that the phosphorylation of PFK and GDH enzymes was up-regulated during winter, accompanied by a simultaneous decrease in their activities. These results suggest that RPP plays an important regulatory role in suppressing metabolic enzymatic activities by altering their conformation and charge state in winter. Our previous studies also showed a lower level of urea in winter (Niu et al., 2022a, 2021c), mainly attributed to suppressed GDH activity. Fatty acid oxidation is regulated by carnitine O-acetyltransferase (CPT1A, a rate-limiting enzyme) and PPARs (nuclear transcription factors) (Barger & Kelly, 2000; Britton et al., 1997). In this study, a lower level of CPT1A and elevated phosphorylation levels of the PPAR signaling pathway suggest suppressed fatty acid oxidation during winter. Moreover, the phosphorylation levels of acetyl-CoA acetyltransferase, which plays a crucial role in the degradation of fats and proteins, and acetyl-CoA carboxylase (ACC), which is the rate-limiting enzyme for fatty acid biosynthesis, were up-regulated in winter. These findings suggest that these ATP-consuming pathways were inhibited by RPP in hibernating *N. parkeri*, contributing to long-term survival during food-depleted winters.

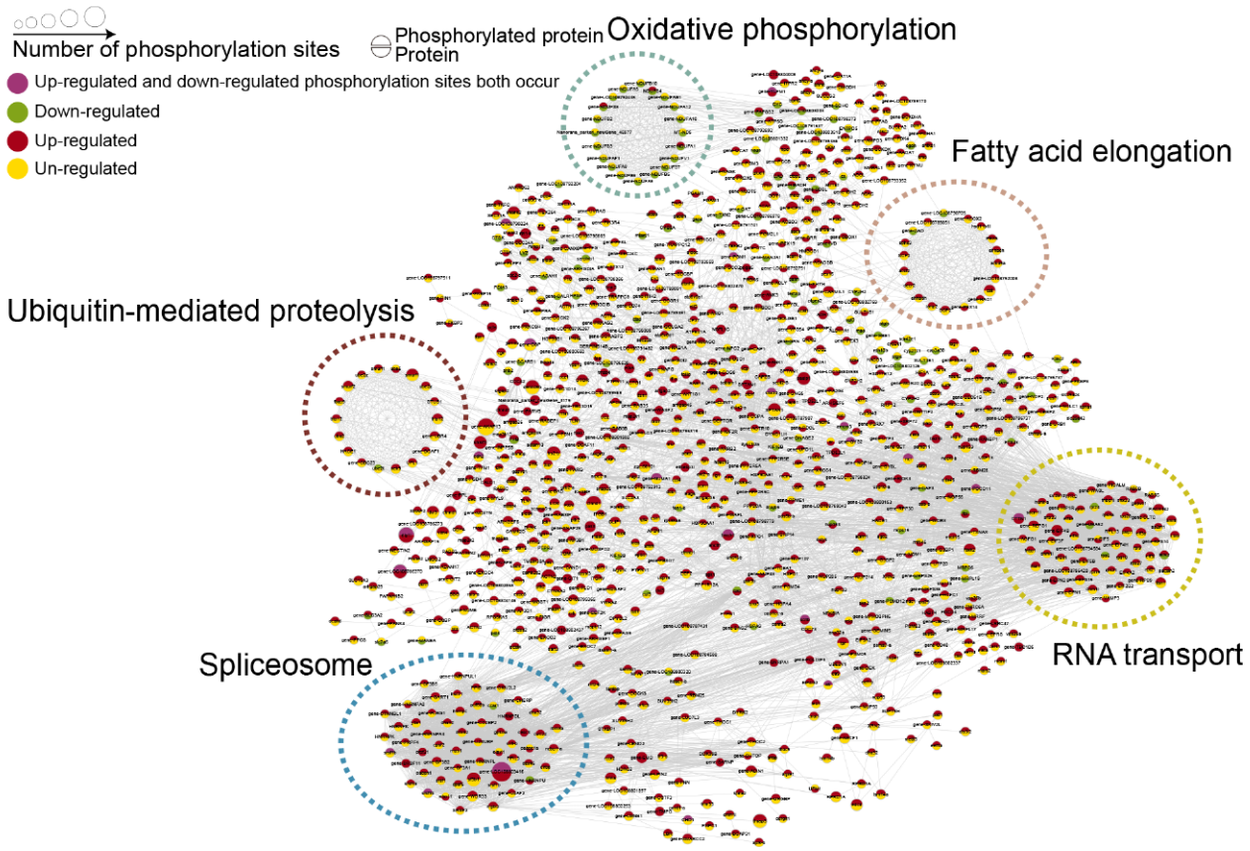


Figure 4 PPI network of differentially regulated phosphopeptides and expressed proteins

Enriched pathways are surrounded by dotted circles. Green, red, and yellow in circles indicate down-, up-, or un-regulated phosphopeptides (upper half) and proteins (lower half). Purple in circle indicates that up-regulated and down-regulated phosphorylation sites occurred simultaneously.

Ion motive ATPases and protein turnover

Energy-consumptive processes are strongly inhibited in a hypometabolic state, among which ion-driven ATPases are particularly metabolically expensive (Clausen, 1986). In this study, increased phosphorylation and decreased activities of Na^+/K^+-ATPase were observed in overwintering *N. parkeri* frogs, suggesting that RPP plays a crucial role in depressing ion pump activity during hibernation. Inhibition of ion motive pumps is essential in the overall reduction of ATP turnover during hibernation. Similar results have also been reported in freeze-tolerant gall fly larvae (*Eurosta solidaginis*), with the activity of both Na^+/K^+-ATPase and Ca^{2+}-ATPase strongly reduced by reversible phosphorylation in midwinter compared to early autumn (McMullen et al., 2010; McMullen & Storey, 2008).

As another major energy consumer in cells, protein turnover (both synthesis and degradation) is often suppressed when animals are subjected to stress or nutrient/energy limitations in hypometabolic systems (Degracia et al., 2002; Storey & Storey, 2004). In the present study, the phosphorylation levels of proteins associated with RNA transport, endoplasmic reticulum protein processing, and protein export were significantly up-regulated during winter, indicating that protein synthesis is inhibited by RPP. Previous research has indicated that RPP plays a pivotal role in regulating protein synthesis, particularly through ribosome initiation and elongation factor activities (Storey & Storey, 1990, 2004). Key sites for regulation include ribosome initiation factor 2 (eIF2), eIF4, and elongation factor 2 (eEF2). The relative content of phosphorylated eIF-2α is considered a sensitive indicator of protein synthesis activity in all eukaryotes (Storey & Storey, 2010). Here, the phosphorylation levels of many translation initiation factors and elongation factors, including eIF-3α, eIF-3f, eIF-2α, eIF-2β, eIF-2C, elongation factor 1 α/β/δ chain, and eEF2, were significantly increased in hibernating *N. parkeri*. Similarly, previous research has shown that the relative amount of phosphorylated eIF-2α is increased in the liver of aestivating desert frogs (*Neobatrachus sutor*) (Pakay et al., 2003). Protein degradation is also likely to be suppressed due to the strong inhibition of protein synthesis during dormancy, as reported in several hypometabolic systems (Storey & Storey, 2004). In this study, ubiquitin-mediated proteolysis exhibited increased phosphorylation during winter, suggesting that protein degradation is also inhibited by RPP in hibernating *N. parkeri*.

Cytoprotective mechanisms

Chaperone proteins, also known as heat shock proteins (HSPs), play important roles in the protection, renewal, and modification of the proteome to sustain cellular functions under hypometabolic states (Storey & Storey, 2004). Higher levels of HSPs have been observed in wood frogs during hibernation, facilitating the refolding and preventing the aggregation of denatured proteins (Kiss et al., 2011). In contrast, our results showed that molecular chaperones (stress-70 protein, HSP 90-β, DnaJ homolog subfamily B member 6) were significantly down-regulated in winter. Concordantly, our previous transcriptomic study showed a significant down-regulation of genes encoding HSPs, such as *HSPD1*, *HSPE1*, *TRAP1*, and *HSPH1* (Niu et al., 2023). This reduced expression of HSPs during hibernation may be a

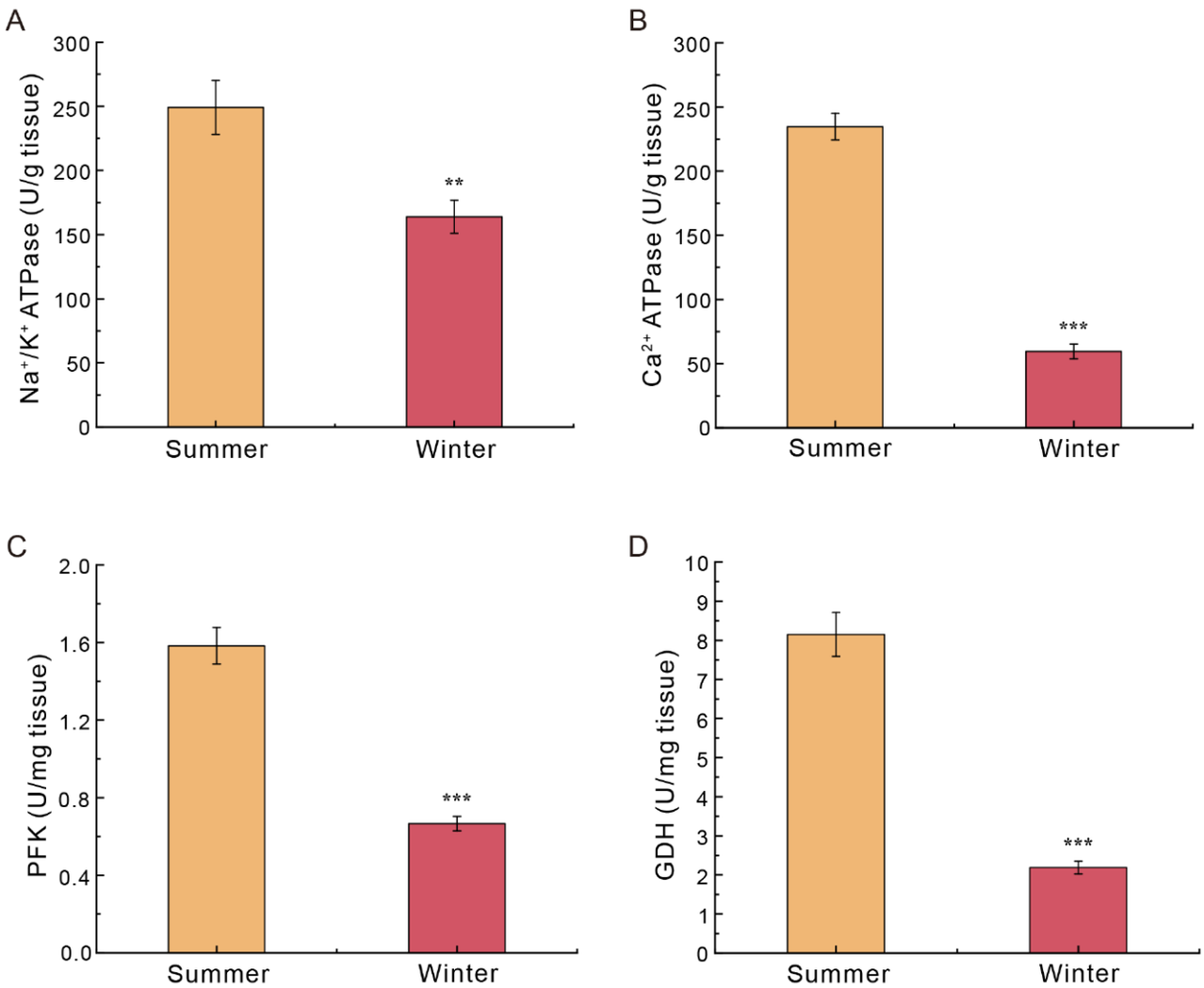


Figure 5 Enzymatic activities in liver of summer- and winter-collected *Nanorana parkeri*

A: Na^+/K^+ ATPase. B: Ca^{2+} ATPase. C: Phosphofructokinase (PFK). D: Glutamate dehydrogenase (GDH). Data are presented as mean±SEM (n=6 for each season). Asterisks (**: P<0.01; ***: P<0.001) indicate a significant difference between two seasons.

consequence of metabolic inhibition. Similar findings have been reported in water frogs (*Pelophylax ridibundus*) (Feidantsis et al., 2012), with the lowest levels of HSP expression observed during hibernation. Enhancing antioxidant defenses is another prevalent feature of hypometabolism across phylogeny (Hermes-Lima et al., 1998; Storey & Storey, 2007), serving to mitigate oxidative damage to macromolecules during hibernation and recovery. We found that the phosphorylation levels of superoxide dismutase (SOD1), catalase, and glutathione S-transferase were significantly increased in winter. Although no significant changes in protein abundance levels were detected, enzymatic activities were markedly reduced, as also shown in our previous study (Niu et al., 2018). These results suggest that RPP plays a critical role in down-regulating the activity of antioxidant enzymes. Moreover, glutaredoxin 2, thioredoxin (TXN2), and glucose-6-phosphate dehydrogenase, which are implicated in antioxidant defense (Lillig et al., 2008; Lu & Holmgren, 2014), were significantly down-regulated in hibernating *N. parkeri*, providing evidence of reduced antioxidant defense. However, the genes encoding these antioxidants did not change significantly at the mRNA level. The inconsistency between the results obtained at the mRNA and protein level may be attributed to post-transcriptional regulation, which needs to be verified in future studies.

Evidence indicates that frogs exhibit reduced immune capabilities during winter, as mounting an effective immune response is energetically demanding (Zapata et al., 1992). In contrast, our previous study suggested an enhanced level of innate immunity in winter may be needed to protect frogs when sequestered in cold water hibernacula (Niu et al., 2022a). Notably, the up-regulated proteins identified in this study and up-regulated genes identified in our prior study (Niu et al., 2023) were significantly enriched in immune response-related pathways. Similar findings have been reported in juvenile Chinese soft-shelled turtles (*Pelodiscus sinensis*), in which cold-induced stress stimulates immune responses in the spleen and intestine, enhancing the complement pathway via up-regulation of the *C1S*, *C3*, and *C6* genes (Zhang et al., 2018). Hence, our results are tentatively consistent with the winter immunity enhancement hypothesis (Nelson & Demas, 1996), although further studies of innate immune parameters, such as complement levels, phagocytes, pattern recognition receptors, natural killer cells, skin antimicrobial peptides, cytokines, chemokines, macrophages, and dendritic cells, are warranted.

Signal transduction pathways

Cell signaling is crucial for reorganizing multiple metabolic pathways and achieving a hypometabolic state (MacDonald, 2004; Storey & Storey, 2012). HIF-1 is the main transcriptional mediator of the hypoxic response in eukaryotic cells, activating expression of target genes involved in angiogenesis, oxygen transport, iron metabolism, glycolysis, glucose uptake, and apoptosis (Bárdos & Ashcroft, 2005; Semenza, 2001). Increasing evidence suggests that HIF-1α regulation occurs at the protein level and accumulates under hypoxic conditions (Kallio et al., 1997). In the present study, we observed higher levels of HIF-1α in overwintering frogs, suggesting potential hypoxia, consistent with our previous findings on the physiological ecology of hibernation in this species (Niu et al., 2022a). Moreover, higher phosphorylation levels of proteins were significantly enriched in the HIF-1 signaling pathway in the winter. HIF-1 plays an important role in inducing the expression of genes encoding glycolytic enzymes, vascular endothelial growth factor, and erythropoietin (Semenza, 2001). However, in our study, the expression of these genes showed no significant changes in hibernating *N. parkeri*. Therefore, we hypothesized that the regulation of HIF-1 by phosphorylation modification does not necessarily activate the expression of downstream targets. HIF-1 is also regulated by other factors, such as stabilization of the HIF-1α subunit, redox regulation, and interaction with HSP90 or selected coactivators (Semenza, 2001). The AMP-activated protein kinase (AMPK), a highly conserved serine/threonine protein kinase, is thought to be a sensor of cellular energy status and is activated by phosphorylation in response to low levels of intracellular ATP and/or high levels of AMP (Hardie et al., 2012). In the present study, AMPK was hyperphosphorylated in winter. The hyperphosphorylation of AMPK in winter has also been reported in hibernating leeches (*Whitmania pigra*) (Shi et al., 2020). To maintain energy homeostasis, AMPK phosphorylates downstream targets to turn on the ATP-producing pathway and turn off ATP-consuming processes such as lipid biosynthesis by phosphorylating ACC or protein synthesis by inactivating eEF2 (Hardie, 2007; Horman et al., 2002). Indeed, the phosphorylation levels of both ACC and eEF2 were up-regulated in hibernating *N. parkeri*, contributing to the down-regulation of lipid biosynthesis and inhibition of translational processes. Numerous studies have shown that AMPK is activated in a hypometabolic state (Bartrons et al., 2004; Rider, 2016). Our results also validate the role of AMPK in mediating metabolic depression through the phosphorylation of downstream targets.

Alternative splicing

Alternative splicing can generate multiple mRNA transcripts from a single mRNA precursor, greatly diversifying the transcriptome and proteome, and thereby altering protein activity, subcellular localization, and protein-protein interactions (Matlin et al., 2005). Alternative splicing also plays an important role in phenotypic plasticity in response to temperature at the cellular level. For instance, GO enrichment analyses have revealed that genes showing differential expression in fish under chronic low temperatures are enriched in spliceosomes (Healy & Schulte, 2019). In line with this, our results suggest that higher phosphorylation of spliceosomal proteins may facilitate an increase in alternative splicing events during cold shock in hibernating *N. parkeri*. Indeed, our previous transcriptomic study showed a significantly higher occurrence of alternative splicing events in winter compared to summer, including alternative 5' first exon (transcription start site) the first exon splicing, alternative 3' last exon (transcription terminal site) the last exon splicing, skipped exon single exon skipping, approximate SKIP single exon skipping (fuzzy boundary), approximate MSKIP multi-exon skipping (fuzzy boundary), and alternative exon ends (5', 3', or both) (Niu et al., 2023).

CONCLUSIONS

In summary, we performed comprehensive proteomic and phosphoproteomic profiling of the liver from summer- and winter-collected *N. parkeri*. Our study demonstrated significant down-regulation in various proteins that participate in active metabolic processes associated with warm-weather life (e.g., growth, reproduction, food acquisition, chaperones, and antioxidant defenses) as well as an up-regulation in proteins associated with the immune signaling pathways in hibernating *N. parkeri*. Moreover, we identified RPP as a crucial regulatory mechanism, which controlled the activities of metabolic enzymes and cellular signaling cascades, regulated transcription via mediation of alternative splicing, and inhibited energy-consuming processes, such as ion-motive ATPases and protein turnover, in hibernating *N. parkeri*. Collectively, our results provide insights into the genetic and epigenetic mechanisms underlying hibernation in a high-altitude frog species.

DATA AVAILABILITY

The mass spectrometry proteomic and phosphoproteomic data have been deposited in the ProteomeXchange Consortium via the PRIDE partner repository (dataset identifier PXD042165). The dataset is deposited in the Science Data Bank (https://www.scidb.cn/) under DOI: 10.57760/sciencedb.j00139.00079.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y.G.N. and D.B.W. conceived and designed the project. Y.G.N., X.J.Z., T.S.X., X.Y.L., H.Y.Z., and Z.F.A. performed the experiments. Y.G.N. and X.J.Z. analyzed the data. T.S.X. and X.Y.L. collected the samples. Y.G.N. and X.J.Z. prepared the first draft of the manuscript. D.B.W., K.B.S., and Q.C. reviewed and edited the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We are grateful to Ruo-Yu Yang and Shi-Han Guo for their help in this study.

REFERENCES

Bárdos JI, Ashcroft M. 2005. Negative and positive regulation of HIF-1: a complex network. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, **1755**(2): 107−120.

Barger PM, Kelly DP. 2000. PPAR signaling in the control of cardiac energy metabolism. *Trends in Cardiovascular Medicine*, **10**(6): 238−245.

Bartrons M, Ortega E, Obach M, et al. 2004. Activation of AMP-dependent protein kinase by hypoxia and hypothermia in the liver of frog *Rana perezi*. *Cryobiology*, **49**(2): 190–194.

Boutilier RG, Donohoe PH, Tattersall GJ, et al. 1997. Hypometabolic

homeostasis in overwintering aquatic amphibians. *Journal of Experimental Biology*, **200**(2): 387−400.

Britton CH, Mackey DW, Esser V, et al. 1997. Fine chromosome mapping of the genes for human liver and muscle carnitine palmitoyltransferase I (CPT1A and CPT1B). *Genomics*, **40**(1): 209−211.

Capraro A, O'meally D, Waters SA, et al. 2019. Waking the sleeping dragon: gene expression profiling reveals adaptive strategies of the hibernating reptile Pogona vitticeps. *BMC Genomics*, **20**(1): 460.

Clausen T. 1986. Regulation of active Na^+-K^+ transport in skeletal muscle. *Physiological Reviews*, **66**(3): 542−580.

Dawson NJ, Storey KB. 2016. A hydrogen peroxide safety valve: the reversible phosphorylation of catalase from the freeze-tolerant North American wood frog. *Rana sylvatica. Biochimica et Biophysica Acta (BBA)-General Subjects*, **1860**(3): 476−485.

Degracia DJ, Kumar R, Owen CR, et al. 2002. Molecular pathways of protein synthesis inhibition during brain reperfusion: implications for neuronal survival or death. *Journal of Cerebral Blood Flow & Metabolism*, **22**(2): 127−141.

Dieni CA, Storey KB. 2009. Creatine kinase regulation by reversible phosphorylation in frog muscle. *Comparative Biochemistry and Physiology Part B:Biochemistry and Molecular Biology*, **152**(4): 405−412.

Dieni CA, Storey KB. 2010. Regulation of glucose-6-phosphate dehydrogenase by reversible phosphorylation in liver of a freeze tolerant frog. *Journal of Comparative Physiology B*, **180**(8): 1133−1142.

Dieni CA, Storey KB. 2011. Regulation of hexokinase by reversible phosphorylation in skeletal muscle of a freeze-tolerant frog. *Comparative Biochemistry and Physiology Part B:Biochemistry and Molecular Biology*, **159**(4): 236−243.

Donohoe PH, Boutilier RG. 1998. The protective effects of metabolic rate depression in hypoxic cold submerged frogs. *Respiration Physiology*, **111**(3): 325−336.

Feidantsis K, Anestis A, Vasara E, et al. 2012. Seasonal variations of cellular stress response in the heart and gastrocnemius muscle of the water frog (*Pelophylax ridibundus*). *Comparative Biochemistry and Physiology Part A:Molecular & Integrative Physiology*, **162**(4): 331−339.

Hardie DG. 2007. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nature Reviews Molecular Cell Biology*, **8**(10): 774−785.

Hardie DG, Ross FA, Hawley SA. 2012. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nature Reviews Molecular Cell Biology*, **13**(4): 251−262.

Hawkins LJ, Wang MJ, Zhang BW, et al. 2019. Glucose and urea metabolic enzymes are differentially phosphorylated during freezing, anoxia, and dehydration exposures in a freeze tolerant frog. *Comparative Biochemistry and Physiology Part D:Genomics and Proteomics*, **30**: 1−13.

Healy TM, Schulte PM. 2019. Patterns of alternative splicing in response to cold acclimation in fish. *Journal of Experimental Biology*, **222**(5): jeb193516.

Hermes-Lima M, Storey JM, Storey KB. 1998. Antioxidant defenses and metabolic depression. The hypothesis of preparation for oxidative stress in land snails. *Comparative Biochemistry and Physiology Part B:Biochemistry and Molecular Biology*, **120**(3): 437−448.

Horman S, Browne GJ, Krause U, et al. 2002. Activation of AMP-activated protein kinase leads to the phosphorylation of elongation factor 2 and an inhibition of protein synthesis. *Current Biology*, **12**(16): 1419−1423.

Huang YF, Yang S, Bai XB, et al. 2021. Molecular and cellular mechanisms of lipid droplet breakdown in the liver of Chinese soft-shelled turtle (*Pelodiscus sinensis*). *Frontiers in Marine Science*, **8**: 633425.

Jin L, Yu JP, Yang ZJ, et al. 2018. Modulation of gene expression in liver of hibernating Asiatic toads (*Bufo gargarizans*). *International Journal of Molecular Sciences*, **19**(8): 2363.

Kallio PJ, Pongratz I, Gradin K, et al. 1997. Activation of hypoxia-inducible factor 1α: posttranscriptional regulation and conformational change by recruitment of the Arnt transcription factor. *Proceedings of the National Academy of Sciences of the United States of America*, **94**(11): 5667−5672.

Kiss AJ, Muir TJ, Lee Jr RE, et al. 2011. Seasonal variation in the hepatoproteome of the dehydration- and freeze-tolerant wood frog. *Rana sylvatica. International Journal of Molecular Sciences*, **12**(12): 8406−8414.

Larson DJ, Middle L, Vu H, et al. 2014. Wood frog adaptations to overwintering in Alaska: new limits to freezing tolerance. *Journal of Experimental Biology*, **217**(12): 2193−2200.

Lillig CH, Berndt C, Holmgren A. 2008. Glutaredoxin systems. *Biochimica et Biophysica Acta (BBA)-General Subjects*, **1780**(11): 1304−1317.

Lin JQ, Huang YY, Bian MY, et al. 2020. A unique energy-saving strategy during hibernation revealed by multi-omics analysis in the Chinese alligator. *iScience*, **23**(6): 101202.

Lu J, Holmgren A. 2014. The thioredoxin antioxidant system. *Free Radical Biology and Medicine*, **66**: 75−87.

Lu Y, Li XY, Zhao K, et al. 2022. Proteomic and phosphoproteomic profiling reveals the oncogenic role of protein kinase D family kinases in cholangiocarcinoma. *Cells*, **11**(19): 3088.

Ma X, Lu X, Merilä J. 2009. Altitudinal decline of body size in a Tibetan frog. *Journal of Zoology*, **279**(4): 364−371.

MacDonald JA. 2004. Signal transduction pathways and the control of cellular responses to external stimuli. *In*: Storey KB. Functional Metabolism: Regulation and Adaptation. New York: Wiley-Liss, 87–123.

Matlin AJ, Clark F, Smith CWJ. 2005. Understanding alternative splicing: towards a cellular code. *Nature Reviews Molecular Cell Biology*, **6**(5): 386−398.

McMullen DC, Ramnanan CJ, Storey KB. 2010. In cold-hardy insects, seasonal, temperature, and reversible phosphorylation controls regulate sarco/endoplasmic reticulum Ca^{2+}-ATPase (SERCA). *Physiological and Biochemical Zoology*, **83**(4): 677−686.

McMullen DC, Storey KB. 2008. Suppression of Na^+K^+-ATPase activity by reversible phosphorylation over the winter in a freeze-tolerant insect. *Journal of Insect Physiology*, **54**(6): 1023−1027.

Nelson RJ, Demas GE. 1996. Seasonal changes in immune function. *The Quarterly Review of Biology*, **71**(4): 511−548.

Niu YG, Cao WJ, Storey KB, et al. 2020. Metabolic characteristics of overwintering by the high-altitude dwelling Xizang plateau frog. *Nanorana parkeri. Journal of Comparative Physiology B*, **190**(4): 433−444.

Niu YG, Cao WJ, Wang JZ, et al. 2021a. Freeze tolerance and the underlying metabolite responses in the Xizang plateau frog. *Nanorana parkeri. Journal of Comparative Physiology B*, **191**(1): 173−184.

Niu YG, Cao WJ, Zhao YF, et al. 2018. The levels of oxidative stress and antioxidant capacity in hibernating *Nanorana parkeri*. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **219**–**220**: 19–27.

Niu YG, Chen Q, Storey KB, et al. 2022a. Physiological ecology of winter hibernation by the high-altitude frog *Nanorana parkeri*. *Physiological and Biochemical Zoology*, **95**(3): 201–211.

Niu YG, Zhang XJ, Men SK, et al. 2023. Integrated analysis of transcriptome and metabolome data reveals insights for molecular mechanisms in overwintering Tibetan frogs. *Nanorana parkeri. Frontiers in Physiology*, **13**: 1104476.

Niu YG, Zhang XJ, Xu TS, et al. 2022b. Physiological and biochemical adaptations to high altitude in Tibetan frogs. *Nanorana parkeri. Frontiers in Physiology*, **13**: 942037.

Niu YG, Zhang XJ, Zhang HY, et al. 2021b. Antioxidant and non-specific immune defenses in partially freeze-tolerant Xizang plateau frogs. *Nanorana parkeri. Journal of Thermal Biology*, **102**: 103132.

Niu YG, Zhang XJ, Zhang HY, et al. 2021c. Metabolic responses of plasma to extreme environments in overwintering Tibetan frogs *Nanorana parkeri*: a

metabolome integrated analysis. *Frontiers in Zoology*, **18**(1): 41.

Pakay JL, Hobbs AA, Kimball SR, et al. 2003. The role of eukaryotic initiation factor 2α during the metabolic depression associated with estivation. *Journal of Experimental Biology*, **206**(14): 2363−2371.

Pinder AW, Storey KB, Ultsch GR. 1992. Estivation and hibernation. *In*: Feder ME, Burggren WW. Environmental Physiology of the Amphibians. Chicago: University of Chicago Press, 250–274.

Rider MH. 2016. Role of AMP-activated protein kinase in metabolic depression in animals. *Journal of Comparative Physiology B*, **186**(1): 1−16.

Semenza GL. 2001. HIF-1 and mechanisms of hypoxia sensing. *Current Opinion in Cell Biology*, **13**(2): 167−171.

Shi HZ, Wang J, Liu F, et al. 2020. Proteome and phosphoproteome profiling reveals the regulation mechanism of hibernation in a freshwater leech (*Whitmania pigra*). *Journal of Proteomics*, **229**: 103866.

Storey KB, Storey JM. 1990. Metabolic rate depression and biochemical adaptation in anaerobiosis, hibernation and estivation. *The Quarterly Review of Biology*, **65**(2): 145−174.

Storey KB, Storey JM. 2004. Metabolic rate depression in animals: transcriptional and translational controls. *Biological Reviews*, **79**(1): 207−233.

Storey KB, Storey JM. 2007. Tribute to P. L. Lutz: putting life on `pause'–molecular regulation of hypometabolism. *Journal of Experimental Biology*, **210**(10): 1700−1714.

Storey KB, Storey JM. 2010. Metabolic regulation and gene expression during aestivation. *In*: Navas CA, Carvalho JE. Aestivation. Berlin: Springer, 25–45.

Storey KB, Storey JM. 2012. Aestivation: signaling and hypometabolism. *Journal of Experimental Biology*, **215**(9): 1425−1433.

Storey KB, Storey JM. 2013. Molecular biology of freezing tolerance. *Comprehensive Physiology*, **3**(3): 1283−1308.

Sun HJ, Zuo XB, Sun L, et al. 2018. Insights into the seasonal adaptive mechanisms of Chinese alligators (*Alligator sinensis*) from transcriptomic analyses. *Australian Journal of Zoology*, **66**(2): 93−102.

Sun YB, Xiong ZJ, Xiang XY, et al. 2015. Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proceedings of the National Academy of Sciences of the United States of America*, **112**(11): E1257−E1262.

Tattersall GJ, Ultsch GR. 2008. Physiological ecology of aquatic overwintering in ranid frogs. *Biological Reviews*, **83**(2): 119−140.

Wu CW, Tessier SN, Storey KB. 2018. Stress-induced antioxidant defense and protein chaperone response in the freeze-tolerant wood frog *Rana sylvatica*. *Cell Stress and Chaperones*, **23**(6): 1205–1217.

Zapata AG, Varas A, Torroba M. 1992. Seasonal variations in the immune system of lower vertebrates. *Immunology Today*, **13**(4): 142−147.

Zhang HY, Zhang XJ, Xu TS, et al. 2022a. Effects of acute heat exposure on oxidative stress and antioxidant defenses in overwintering frogs. *Nanorana parkeri. Journal of Thermal Biology*, **110**: 103355.

Zhang JH, Cai RQ, Liang JJ, et al. 2021. Molecular mechanism of Chinese alligator (*Alligator sinensis*) adapting to hibernation. *Journal of Experimental Zoology Part B:Molecular and Developmental Evolution*, **336**(1): 32−49.

Zhang WY, Niu CJ, Chen BJ, et al. 2018. Digital gene expression profiling reveals transcriptional responses to acute cold stress in Chinese soft-shelled turtle *Pelodiscus sinensis* juveniles. *Cryobiology*, **81**: 43−56.

Zhang Y, Wang Q, Liang JQ et al. 2022b. Serum proteomic analysis of differentially expressed proteins and pathways involved in the mechanism of endemic osteoarthritis. *Molecular Omics*, **18**(8): 745−753.

Liang et al. *Zool. Res.* 2024, 45(1): 13−24
https://doi.org/10.24272/j.issn.2095-8137.2023.053

ZR | Zoological Research

Article Open Access

Full T-cell activation and function in teleosts require collaboration of first and co-stimulatory signals

Wei Liang[1,#], Kang Li[1,#], Haiyou Gao[1], Kunming Li[1], Jiansong Zhang[1], Qian Zhang[1], Xinying Jiao[1], Jialong Yang[1,2,*], Xiumei Wei[1,*]

[1] *State Key Laboratory of Estuarine and Coastal Research*, *School of Life Sciences*, *East China Normal University*, *Shanghai* 200241, *China*

[2] *Laboratory for Marine Biology and Biotechnology*, *Qingdao National Laboratory for Marine Science and Technology*, *Qingdao*, *Shandong* 266237, *China*

ABSTRACT

Mammalian T-cell responses require synergism between the first signal and co-stimulatory signal. However, whether and how dual signaling regulates the T-cell response in early vertebrates remains unknown. In the present study, we discovered that the Nile tilapia (*Oreochromis niloticus*) encodes key components of the LAT signalosome, namely, LAT, ITK, GRB2, VAV1, SLP-76, GADS, and PLC-γ1. These components are evolutionarily conserved, and CD3ε mAb-induced T-cell activation markedly increased their expression. Additionally, at least ITK, GRB2, and VAV1 were found to interact with LAT for signalosome formation. Downstream of the first signal, the NF-κB, MAPK/ERK, and PI3K-AKT pathways were activated upon CD3ε mAb stimulation. Furthermore, treatment of lymphocytes with CD28 mAbs triggered the AKT-mTORC1 pathway downstream of the co-stimulatory signal. Combined CD3ε and CD28 mAb stimulation enhanced ERK1/2 and S6 phosphorylation and elevated NFAT1, c-Fos, IL-2, CD122, and CD44 expression, thereby signifying T-cell activation. Moreover, rather than relying on the first or co-stimulatory signal alone, both signals were required for T-cell proliferation. Full T-cell activation was accompanied by marked apoptosis and cytotoxic responses. These findings suggest that tilapia relies on dual signaling to maintain an optimal T-cell response, providing a novel perspective for understanding the evolution of the adaptive immune system.

Keywords: *Oreochromis niloticus*; CD3; CD28; T cells; Adaptive immunity; Evolution

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

INTRODUCTION

The emergence of the T-cell-mediated adaptive immune system in fish played a critical role in facilitating the rapid occupation of a favorable ecological niche by jawed vertebrates (Anderson et al., 2004). Recently, many studies have highlighted the importance of T-cell-mediated adaptive immunity in teleosts (Ai et al., 2022; Jung et al., 2021; Mu et al., 2022; Tafalla et al., 2016). Once T cells recognize the major histocompatibility complex (MHC)-antigenic peptide complex on the surface of antigen-presenting cells (APCs), they rapidly activate, proliferate, differentiate, and release cytokines to resist pathogenic bacterial invasion (Ashfaq et al., 2019). In recent years, our understanding of the immune system in certain fish species has seen significant improvements. However, the scarcity of adequate tools to study the fish immune response continues to impede research on the mechanisms and evolution of the fish immune system.

Appropriate T-cell activation is essential to initiate an appropriate immune response. Mammalian T-cell activation, proliferation, and differentiation require two critical stimulatory signals: the interaction of the T-cell receptor (TCR)/CD3 complex with MHC on APCs and the binding of the co-stimulatory molecule CD28 to the B7 family ligands on APCs (Cheng et al., 2011; Liu et al., 2020b; Mariuzza et al., 2020; Sanchez-Lockhart et al., 2011). Dependent on conformational changes, the immunoreceptor tyrosine-based activation motif (ITAM) in the ζ chain of the TCR/CD3 complex is phosphorylated by the leukocyte C-terminal Src kinase (LCK), which further initiates the recruitment and activation of zeta chain of T-cell receptor-associated protein kinase 70 kDa (ZAP-70) (Au-Yeung et al., 2018). Subsequent phosphorylation of the SH2-domain-containing leukocyte protein 76 kDa (SLP-76) by ZAP-70 is indispensable for

Received: 06 May 2023; Accepted: 08 September 2023; Online: 09 September 2023

Foundation items: This study was supported by the National Key Research and Development Program (2022YFD2400804), National Natural Science Foundation of China (32022086, 31972822), and Natural Science Foundation of Shanghai (20ZR1417500)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: jlyang@bio.ecnu.edu.cn; xmwei@bio.ecnu.edu.cn

interleukin 2 (IL-2) inducible T-cell kinase (ITK) activation, which is then recruited to the plasma membrane (Sela et al., 2011). Changes in the conformation of ITK facilitate the activation of downstream effectors via formation of the linker for activation of T cells (LAT) signalosome (Schwartzberg et al., 2005). Growth factor receptor-bound protein 2 (GRB2), ITK, SLP-76, GRB2-related adaptor protein 2 (GADS), vav guanine nucleotide exchange factor 1 (VAV1), and phospholipase C gamma 1 (PLC-γ1) are known to interact with LAT in mammals, which is for downstream signaling cascades (Bunnell et al., 2000). PLC-γ1, once activated, hydrolyzes the membrane lipid phosphatidylinositol 4,5-bisphosphate into two second messengers: inositol polyphosphates and diacylglycerols (Smith-Garvin et al., 2009). Subsequently, with the help of second messengers, downstream signaling cascades are activated, including Ca^{2+}-calcineurin-NFAT, NF-κB, MAPK/ERK, and mTOR signaling (Gorentla & Zhong, 2012). These signaling cascades ultimately facilitate T-cell activation, proliferation, and effector function (Gorentla & Zhong, 2012). The first signal facilitates antigen binding to specific co-receptors on the T-cell surface, prompting initial T-cell activation. As a critical co-receptor, CD3 is involved in the assembly, stabilization, and signal transduction of the TCR/CD3 complex. This complex transmits the activation signal, which is generated when the TCR binds to the antigen, into the intracellular compartment, ultimately leading to T-cell activation (Kuhns et al., 2006). Monoclonal antibodies targeting the CD3 ε chain (CD3ε mAbs) have been developed to induce mammalian T-cell activation *in vitro* (Chitnis et al., 2022). The binding of CD3 mAbs strongly initiates T-cell activation and IL-2 receptor expression. Moreover, CD3 mAb stimulation can enhance the production of IL-2, IL-3, and interferon γ (IFN-γ) (Sauerwein et al., 1988).

However, without the second signal, the interaction of the TCR with the MHC-antigenic peptide complex is not sufficient to trigger full T-cell activation, leading to a hypo-responsive state (Caporali et al., 2014). The second signal is a non-specific, co-stimulatory signal generated by the interaction of multiple pairs of co-stimulatory molecules on APCs and corresponding TCRs, e.g., CD28 and CD80/CD86, 4-1BB and 4-1BBL, and CD40 and CD40L (Chen & Flies, 2013). CD28, as the most important co-stimulatory molecule, is characterized by an extracellular variable immunoglobulin-like structural domain and is constitutively expressed in T cells (Beyersdorf et al., 2015; Esensten et al., 2016). The involvement of CD28 enhances full T-cell activation, cell survival, and cytokine secretion (Marinari et al., 2004). Moreover, cross-linking CD28 with CD28 mAbs in the presence of alloantigen, T-cell mitogen (PHA), or CD3 mAbs greatly enhances the IL-2 production, activation, proliferation, and cytotoxicity of T cells (June et al., 1990; Jung et al., 1987; Lesslauer et al., 1986). However, CD28 signaling alone does not induce T-cell proliferation (Hara et al., 1985). Although numerous mammalian studies have elucidated that dual signaling effectively initiates T-cell function (Nguyen et al., 2016; Sloan-Lancaster et al., 1993), its potential regulatory role in T-cell response remains unknown in early vertebrates.

An increasing number of studies have demonstrated that signal transduction downstream of the TCR in teleosts is similar to that in mammals (Rodríguez-Caparrós et al., 2020), but many obstacles remain in identifying the different components and mechanisms that trigger TCR signaling, including the lack of mAbs (Randelli et al., 2008). Recently, various mAbs have been developed against T-cell markers in teleosts, including DLT15, which specifically targets peripheral and thymocyte T cells in European sea bass, and 6D1 and 6C10, which specifically target CD4 and CD8, respectively, in common carp (Abelli et al., 1999; Laing & Hansen, 2011). Moreover, several mAbs have been developed against the first signal (CD3) in teleosts, including rainbow trout, olive flounder, Japanese flounder, and grass carp (Boardman et al., 2012; Jung et al., 2017; Qin et al., 2021; Tang et al., 2017). Furthermore, mAbs have been developed to target the co-stimulatory molecule CD28 in olive flounder (Xing et al., 2021). However, how the first signal and co-stimulatory signal together regulate the T-cell response in teleosts remains poorly understood. In the present study, we revealed conserved components of the LAT signalosome in Nile tilapia (*Oreochromis niloticus*) that may be involved in TCR downstream signal transduction upon CD3ε mAb-induced T-cell activation. Furthermore, we demonstrated the regulatory role of the dual signal in full T-cell activation, proliferation, apoptosis, and cytokine release. To the best of our knowledge, this is the first study to describe the regulatory mechanism of dual signaling-mediated T-cell response in a non-mammalian vertebrate. Collectively, these findings provide a valuable reference for understanding the role of dual signaling-induced TCR signal transduction, as well as the evolution of T-cell-mediated adaptive immunity.

MATERIALS AND METHODS

Ethics statement and experimental animals

Nile tilapias (2 g in weight) were purchased from a fish farm in Guangzhou, Guangdong Province, China, and cultured in 120 L tanks at 28 °C with continuous aeration and daily feeding at the biological station of East China Normal University. Healthy tilapias were used for the experiments once reaching 15 g in weight. All experimental procedures were conducted in accordance with the Guide for the Care and Use of Laboratory Animals of the Ministry of Science and Technology of China and were approved by the East China Normal University Experimental Animal Ethics Committee (permit No. AR2021-245).

Sequence, structure, and phylogenetic analysis

The cDNA and amino acid sequences of related genes were obtained from the NCBI GenBank database (https://www.ncbi.nlm.nih.gov/) and were analyzed using the BLAST algorithm (https://blast.ncbi.nlm.nih.gov/Blast.cgi). The protein domains were predicted by SMART (http://smart.embl.de/) and domain organization was displayed using DOG v.2.0 software. Multiple sequence alignment of the amino acids was performed using Clustalx v.1.83. The protein tertiary structures were predicted using SWISS-MODEL (https://swissmodel.expasy.org/) and displayed with PyMOL software v.2.3.2.0. The phylogenetic tree was constructed using MEGA v.7.0 with the neighbor-joining algorithm. The accession numbers of selected genes are listed in Supplementary Table S1.

Leukocyte isolation

Spleen and head kidney leukocytes were isolated according to our previous study (Wei et al., 2019). Briefly, Percoll (GE Healthcare, USA) was mixed with 10×phosphate-buffered saline (PBS) at a ratio of 9:1, then diluted to 52% Percoll and 34% Percoll with L-15 medium (Gibco, USA), respectively.

The spleen was harvested, and a single-cell suspension was prepared with L-15 medium. We added 4 mL of 52% Percoll, 34% Percoll, and cell suspension, respectively, into a 15 mL centrifuge tube, which was then centrifuged at 500 ×*g* for 35 min at room temperature with the lowest acceleration and deceleration. Leukocytes between the 52% and 34% Percoll layer were collected, washed with L-15 medium, and resuspended with Dulbecco's modified Eagle medium (DMEM, Sangon Biotech, China) with 10% fetal bovine serum (FBS, Gibco, USA) for future assay. According to our previous study, almost all isolated spleen leukocytes are lymphocytes (Ai et al., 2022; Wei et al., 2019).

Leukocyte stimulation

The spleen leukocytes (3×10^6) were cultured in a 24-well plate with DMEM containing 10% FBS and 1% penicillin/streptomycin (Sangon Biotech, China) at 28 °C. The cells were then stimulated with previously generated CD3ε and CD28 mAbs (2 μg/mL) (Li et al., 2023), and collected at indicated times for assay. Unstimulated leukocytes were used for the control. For the phosphorylation assays, the spleen leukocytes were resuspended in Dulbecco's PBS at 28 °C for 30 min to rest the phosphorylated proteins. After that, the cells were transferred to DMEM (10% FBS) and stimulated with CD3ε or CD28 mAbs at 28 °C. The leukocytes were collected at indicated times for western blotting and immunofluorescence assays.

Reverse transcription-quantitative real-time polymerase chain reaction (RT-qPCR)

Total RNA was extracted from the spleen leukocytes using TRIzol reagent (Invitrogen, USA) according to the provided instructions. After treatment with gDNA Purge for 5 min at 42 °C, the RNA was reverse-transcribed using the First-Strand cDNA Synthesis SuperMix (Novoprotein, China). The resulting cDNA was diluted 1:10 to serve as a template, and RT-qPCR was performed using the NovoStart SYBR qPCR SuperMix Plus (Novoprotein, China) on a CFX Connect Real-Time System (BioRad, USA). Relative mRNA expression was calculated using the $2^{-\Delta\Delta CT}$ method, with β-actin used as the reference gene. All primer information is listed in Supplementary Table S2.

Western blot assay

The spleen leukocytes were lysed in NP40 lysis buffer containing 1 mmol/L phenylmethylsulfonyl fluoride (PMSF), protease inhibitor cocktail, and phosphatase inhibitor cocktail III (MedChemExpress, USA) on ice for 30 min. The supernatant was harvested after centrifugation at 10 000 ×*g* and 4 °C for 10 min, then mixed with sodium dodecyl sulfate (SDS)-loading buffer and boiled at 100 °C for 10 min. The sample was separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and subjected to western blot analysis using primary antibodies (1:1 000), including anti-p-IKKα/β (Ser176/180, #2697), anti-IKKα/β (#2694), anti-p-IκBα (Ser32, #2859), anti-IκBα (#4812), anti-p-NF-κB p65 (Ser468, #3039), anti-NF-κB p65 (#9936), anti-p-c-Raf Ser-338 (#9427), anti-c-Raf (#9422), anti-p-MEK1/2 (Ser217/221,#9936), anti-Mek1/2 (#8727), anti-p-p44/42 MAPK (ERK1/2) (Thr202/Tyr204, #4370), anti-p44/42 MAPK (ERK1/2,#9102), anti-p-MEK1/2 (Ser217/221, #8221), anti-MEK1/2 (#8727), anti-p-c-Fos (ser32, #5348), anti-c-Fos (#2250), anti-PI3 Kinase p110α (#4249), anti-p-AKT (Thr308,#13038), anti-p-AKT (Thr473, #4060), anti-AKT (#4697), anti-p-S6 (Ser240/244, #5364), anti-S6 (#2217), anti-p-4EBP1 (Thr37/46, #5123), anti-4EBP1 (#9959), and anti-β-actin (#3700) purchased from Cell Signaling Technology (USA). Primary antibody binding was then detected using secondary antibodies, 1:30 000 diluted goat anti-rabbit IgG H&L Alexa Fluor 800 (#5151, Cell Signaling Technology, USA) or 1:10 000 diluted goat anti-mouse IgG H&L Alexa Fluor 680 (#ab175775, Abcam, UK), at room temperature for 1 h. After each incubation, the membranes were washed three times using PBS with Tween 20 (PBST). Finally, the blots were scanned using Odyssey CLx Image Studio (USA).

Immunofluorescence assay

The spleen leukocytes were centrifuged onto slides using Cytospin 4 (Thermo Scientific, USA) at 1 000 ×*g* for 3 min, then fixed with methyl alcohol for 5 min. After the slides were blocked with 1% bovine serum albumin (BSA) at 37 °C for 1 h, the samples were washed twice with PBST and once with PBS. The samples were subsequently incubated with 1:200 diluted mouse-anti CD3ε or CD28 mAbs, rabbit anti-p-S6 (Ser240/244), rabbit anti-p-ERK1/2 (Thr202/Tyr 204), or mouse anti-β-actin primary antibodies at 37 °C for 1 h. After washing, the cells were incubated with 1:800 diluted Alexa Fluor 488-conjugated goat anti-mouse IgG H+L (Abcam, UK), Alexa Fluor 594-conjugated goat anti-mouse IgG H+L (Abcam), or Alexa Fluor 594-conjugated goat anti-rabbit IgG H+L (Abcam, UK) secondary antibodies at 37 °C for 1 h. For the CD3 and CD28 co-localization assay, the cells were first stained with mouse anti-CD28 mAbs and Alexa Fluor 594-conjugated goat anti-mouse IgG H+L, followed by incubation with FITC-conjugated mouse anti-CD3ε mAbs at 37 °C for 1 h. After washing three times, 4',6-diamidino-2-phenylindole (DAPI) staining solution (Beyotime, China) was added, with the slides then sealed with a coverslip and observed using a Zeiss ApoTome microscope (Germany).

Co-immunoprecipitation (Co-IP) assay

The coding sequence (CDS) regions with a Flag or HA tag, including LAT-Flag, GRB2-HA, ITK-HA, VAV1-HA, PLC-γ1-HA, GADS-HA, and SLP-76-HA, were amplified from the tilapia cDNA with corresponding primers (Supplementary Table S2), then ligated into pEGFP-C1 plasmid to construct transfection vectors. The HEK293T cells (2×10^6) were seeded into 6 cm^2 dishes overnight, which were co-transfected with 5 μg of the LAT-Flag-pEGFP-C1 plasmid and 5 μg of the partner-HA-pEGFP-C1 plasmid. The cells were collected at 48 h post-transfection, then lysed with NP40 as above. The supernatants were collected and incubated with 10 μL of anti-Flag Ab-conjugated agarose beads (Sigma-Aldrich, USA) at 4 °C overnight with shaking. Subsequently, the beads were washed five times with 1 mL of NP40 lysis buffer, mixed with SDS-loading buffer and heated at 100 °C for 10 min, followed by western blot analysis.

Flow cytometry

The spleen and head kidney leukocytes were stained with mouse anti-tilapia CD28 mAbs (1 : 400) in FACS buffer (2% FBS in PBS) on ice for 30 min, then washed twice with FACS buffer. The cells were subsequently incubated with Alexa Fluor 647-conjugated goat anti-mouse IgG H+L (1 : 2 000; Abcam, UK) on ice for 30 min, then washed twice. After that, the leukocytes were stained with FITC-conjugated CD3ε mAbs (1 : 400) on ice for another 30 min and washed twice. The cells were finally resuspended in FACS buffer and collected

on a BD Canto-II flow cytometer. The data were analyzed using FlowJo software v.10.0.7.

Cell proliferation assay

Spleen leukocytes (1×10^6) were labeled with 10 μmol/L carboxyfluorescein succinimidyl ester (CFSE, Invitrogen, USA) according to the manufacturer's instructions. After washing twice with L15 medium, the cells were resuspended in DMEM (10% FBS) and cultured in a 24-well plate with or without 2 μg/mL CD3ε or CD28 mAbs. The cells were harvested at 48 h and stained with CD3ε mAbs as above. Subsequently, 7AAD (1:400; BioLegend, USA) was added before the samples were subjected to flow cytometry assay.

Apoptosis assay

The spleen leukocytes were stimulated with CD3ε and CD28 mAbs for specified durations, as described above. The cells were first stained with CD3ε mAbs, then twice washed with FACS buffer. Subsequently, the cells were stained with APC-conjugated Annexin V antibodies (1 : 400; BioLegend, USA) in Annexin V binding buffer (0.14 mol/L NaCl, 0.01 mol/L HEPES/NaOH, 2.5 mmol/L $CaCl_2$, pH 7.4,) at room temperature for 15 min, with 7AAD (1:400) then added before flow cytometry assay.

Statistical analysis

All results are presented as mean±standard error of the mean (SEM). Significant differences were determined using a two-tailed Student *t*-test. *P*-values were indicated as *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$.

RESULTS

Tilapia possesses intact and evolutionarily conserved LAT signalosome components

Downstream of TCR signaling cascades, the LAT signalosome plays an essential role in signal transduction to ensure a proper T-cell response. Here, the central component of the LAT signalosome was found to be located on chromosome LG4 in the Nile tilapia (Figure 1A). Based on synteny analysis, both the location and orientation of LAT and its adjacent genes, including *SH2B1, ATXN2L, SPNS1, NFATC2IP, and SGF29*, were evolutionarily conserved among the teleost species (Figure 1A). In addition to LAT, other components that form the signalosome in mammals, including ITK, GRB2, SLP-76, GADS, VAV1, and PLC-γ1, were identified in tilapia in both our previous (Liang et al., 2022) and present study (Figure 1B), suggesting the potential presence of the LAT signalosome in this early vertebrate species. Furthermore, the tilapia components exhibited conserved functional domain organization analogous to mouse homologs (Figure 1B). Notably, the key components of the tilapia LAT signalosome showed high tertiary structural congruence with their mouse counterparts (Figure 1C). Phylogenetic trees were constructed to further investigate the evolutionary relationships of the LAT signalosome components. Similar to ITK in our previous study (Liang et al., 2022), all LAT signalosome components from tilapia formed distinct clusters with their counterparts from diverse teleost species (Figure 1D, E; Supplementary Figure S1), suggesting close evolutionary distance. Collectively, these observations indicate that Nile tilapias harbor intact and evolutionarily conserved components of the LAT signalosome.

First signal induces activation and formation of LAT signalosome

As the first signal triggered by the TCR/CD3 complex is essential for T-cell activation, we first determined the exact signal during CD3ε mAb-induced T-cell activation in tilapia. Downstream of the TCR, the mRNA levels of early T-cell activation molecules, including LCK and ZAP-70, were significantly increased in the spleen lymphocytes upon CD3ε mAb activation (Figure 2A). Furthermore, phosphorylation of these molecules was concordantly enhanced, as observed in our previous study (Li et al., 2021). First signal-induced T-cell activation lead to an increase in the transcriptional levels of LAT (Figure 2B), as well as ITK, GRB2, SLP-76, GADS, VAV1, and PLC-γ1 (Figure 2C), suggesting activation of the LAT signalosome components downstream of TCR signaling. In mammalian systems, engagement of the TCR typically results in the interaction of LAT with other signaling components to initiate a downstream cascade. Thus, we next investigated the constitutive associations of the LAT signalosome in tilapia. As revealed by the Co-IP assay, tilapia LAT directly interacted with ITK (Figure 2D), GRB2 (Figure 2E), and VAV1 (Figure 2F), but showed no direct association with SLP-76, GADS, or PLC-γ1 (Figure 2G–I). These findings suggest that while the roles of SLP-76, GADS, and PLC-γ1 in the LAT signalosome remain uncertain, LAT, ITK, GRB2, and VAV1 are integral components for intracellular signaling transduction in tilapia T cells. Overall, these results underscore the role of the first signal in modulating LAT signalosome activation and formation in tilapia T cells.

First signal activates T cell canonical pathways in tilapia

Next, we investigated whether the first signal regulates T-cell activation in tilapia. Downstream of TCR signaling, several canonical pathways, such as NF-κB and MAPK/ERK, are essential prerequisites for orchestrating T-cell-mediated adaptive immunity (Gorentla & Zhong, 2012). Here, upon CD3ε mAb-induced T-cell activation, there was a significant increase in the transcription levels of IKKα/β, IκBα, and NF-κB p65 in the spleen lymphocytes (Figure 3A). Furthermore, concurrent enhancements were observed in the phosphorylation of IKKα/β, IκBα, and NF-κB p65, and degradation of IκBα (Figure 3B), suggesting robust activation of the NF-κB pathway. In line with these findings, the MAPK/ERK pathway was also activated downstream of the first signal, as evidenced by the elevated transcription and phosphorylation of MAPK/ERK components (i.e., c-Raf, MEK1/2 and ERK1/2) and transcription factor c-Fos upon CD3ε mAb stimulation (Figure 3C, D). Previous studies have indicated that the first signal activates the AKT-mTOR pathway through the NF-κB and MAPK/ERK axis in mammals (Gorentla et al., 2011). Our results revealed that this regulation is not exclusive to mammals. Notably, we observed a marked increase in the mRNA and phosphorylation levels of mTOR components, including AKT, S6, and 4E-BP1, in tilapia during CD3ε mAb-induced T-cell activation (Figure 3E, F). Thus, these results suggest that the first signal activates multiple canonical pathways crucial for the activation of T cells in tilapia.

CD28-mediated co-stimulatory signal initiates AKT-mTORC1 pathway in tilapia

In addition to the first signal, CD28-mediated co-stimulatory signaling is also essential for T-cell activation, proliferation,

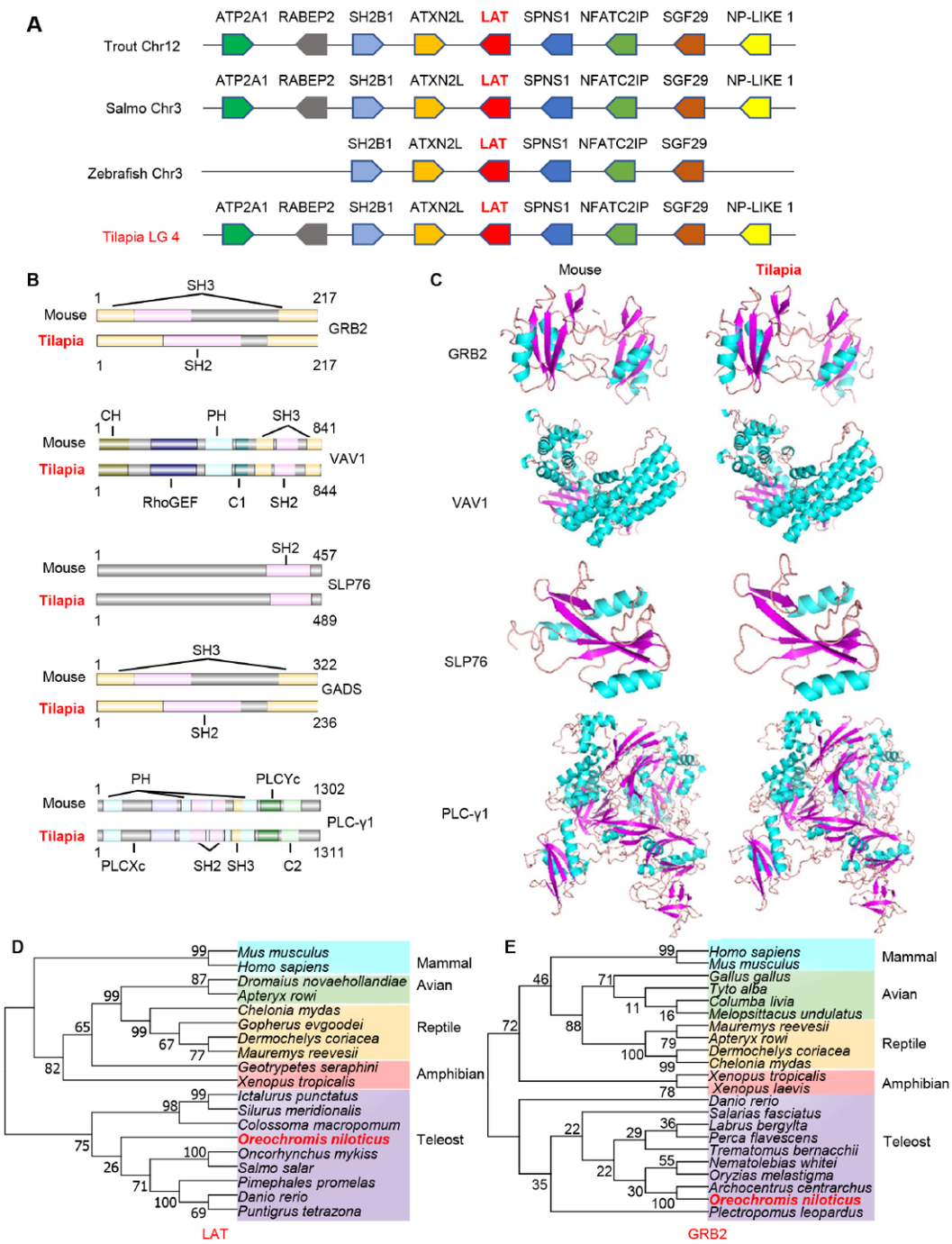


Figure 1 Evolutionary conservation of LAT signalosome molecules in tilapia

A: Synteny and chromosomal location of LAT gene in teleosts. B: Comparative analysis of domain organization of LAT signalosome components in Nile tilapia and mouse. C: Tertiary structure of LAT signalosome molecules in Nile tilapia and mouse were predicted by SWISS-MODEL. D, E: Phylogenetic trees of LAT and GRB2 were constructed using MEGA v.7.0 with the neighbor-joining method. Bootstrap values for 1 000 replicates (%) are indicated on each branch.

differentiation, and function (Marinari et al., 2004). Through immunofluorescence and flow cytometry assays, we detected CD28 on the surface of the tilapia lymphocytes (Figure 4A), with tilapia CD28$^+$ cells accounting for approximately 30% and 20% of the lymphocytes in the spleen and head kidney, respectively (Figure 4B). To confirm the expression of CD28 in tilapia T cells, the spleen and head kidney lymphocytes were stained with CD28 and CD3ε mAbs for flow cytometry. Results showed that CD28 was predominantly expressed in the CD3$^+$ T-cell population in both the spleen and head kidney (Figure 4C), as supported by the co-localization of CD28 and CD3 based on immunofluorescence assay (Figure 4D). To further investigate whether the CD28-mediated co-stimulatory signal regulates T-cell activation, the spleen lymphocytes were stimulated with CD28 mAbs. Compared to the unstimulated control, CD28 and CD28 mAb ligation enhanced both the transcription and phosphorylation of AKT (Figure 4E, F), subsequently activating downstream mTORC1 signaling, as evidenced by the increased transcription and phosphorylation of S6 and 4E-BP1 (Figure 4E, F). These results suggest that the CD28-mediated co-stimulatory signal may regulate T-cell responses in tilapia via AKT-mTORC1 pathway initiation.

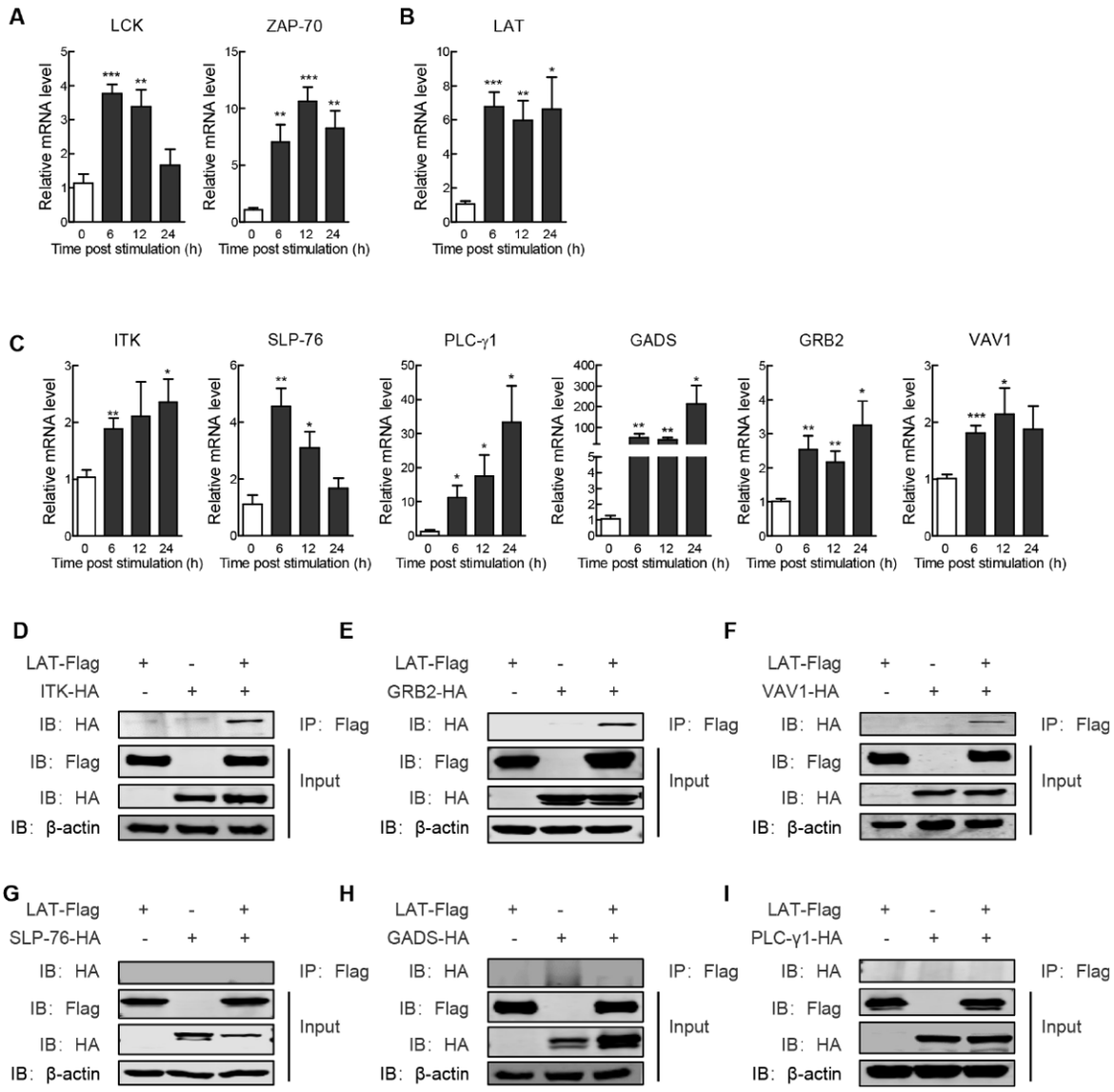


Figure 2 Expression patterns and interactions of tilapia LAT signalosome molecules

A–C: Relative mRNA levels of indicated molecules in spleen lymphocytes upon CD3ε mAb stimulation, as determined by RT-qPCR, $n \geq 4$. D–I: Interactions of Nile tilapia LAT with ITK (D), GRB2 (E), VAV1 (F), SLP-76 (G), GADS (H), and PLC-γ (I), as determined by Co-IP and western blot assay. Experiments were conducted at least twice independently. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$.

Dual signals are critical for full activation of T cells in tilapia

We next examined the essential role of dual signals in T-cell activation of tilapia. Stimulation of spleen lymphocytes with CD3ε and CD28 mAbs resulted in a pronounced increase in the phosphorylation of ERK1/2 and its nuclear translocation (Figure 5A). A similar trend was observed for the phosphorylation and nuclear translocation of S6 (Figure 5B). These events, related to MAPK/ERK and mTORC1 activation, suggest the potential involvement of dual signals in promoting T-cell activation of tilapia. Although stimulation with either CD3ε or CD28 mAbs alone can induce mRNA expression of T-cell activation transcription factors NFAT1 and c-Fos, combined stimulation with CD3ε and CD28 mAbs led to much higher expression (Figure 5C). Stimulation with the first signal, but not the second signal, also increased the transcription of T-cell activation markers CD122 and CD44 in the spleen lymphocytes (Figure 5D), whereas activation of both signals resulted in more pronounced up-regulation (Figure 5D). Thus, these observations suggest that full T-cell activation requires the synergistic cooperation of dual signaling. To assess the effects of the first signal, co-stimulatory signal, and dual signals on tilapia T cell proliferation, CFSE-labeled spleen lymphocytes were stimulated with CD3ε, CD28, or CD3ε plus CD28 mAbs *in vitro*. At 48 h after stimulation, synergistic CD3ε and CD28 mAbs treatment induced significant $CD3^+$ T cell proliferation, while neither the CD3ε nor αCD28 mAbs alone triggered such proliferation (Figure 5E), highlighting the indispensable role of both the first and co-stimulatory signals in T-cell proliferation in tilapia. Taken together, these findings indicate that the dual signals are essential for complete activation and proliferation of tilapia T cells.

Dual signals promote T-cell function and triggers T-cell apoptosis

As the dual signals triggered T-cell activation and proliferation, we further explored their potential roles in regulating tilapia T-cell function. Dual signal stimulation induced high mRNA expression of IL-2 and tumor necrosis factor alpha (TNF-α) (Figure 6A), critical factors in T-cell proliferation, survival, and

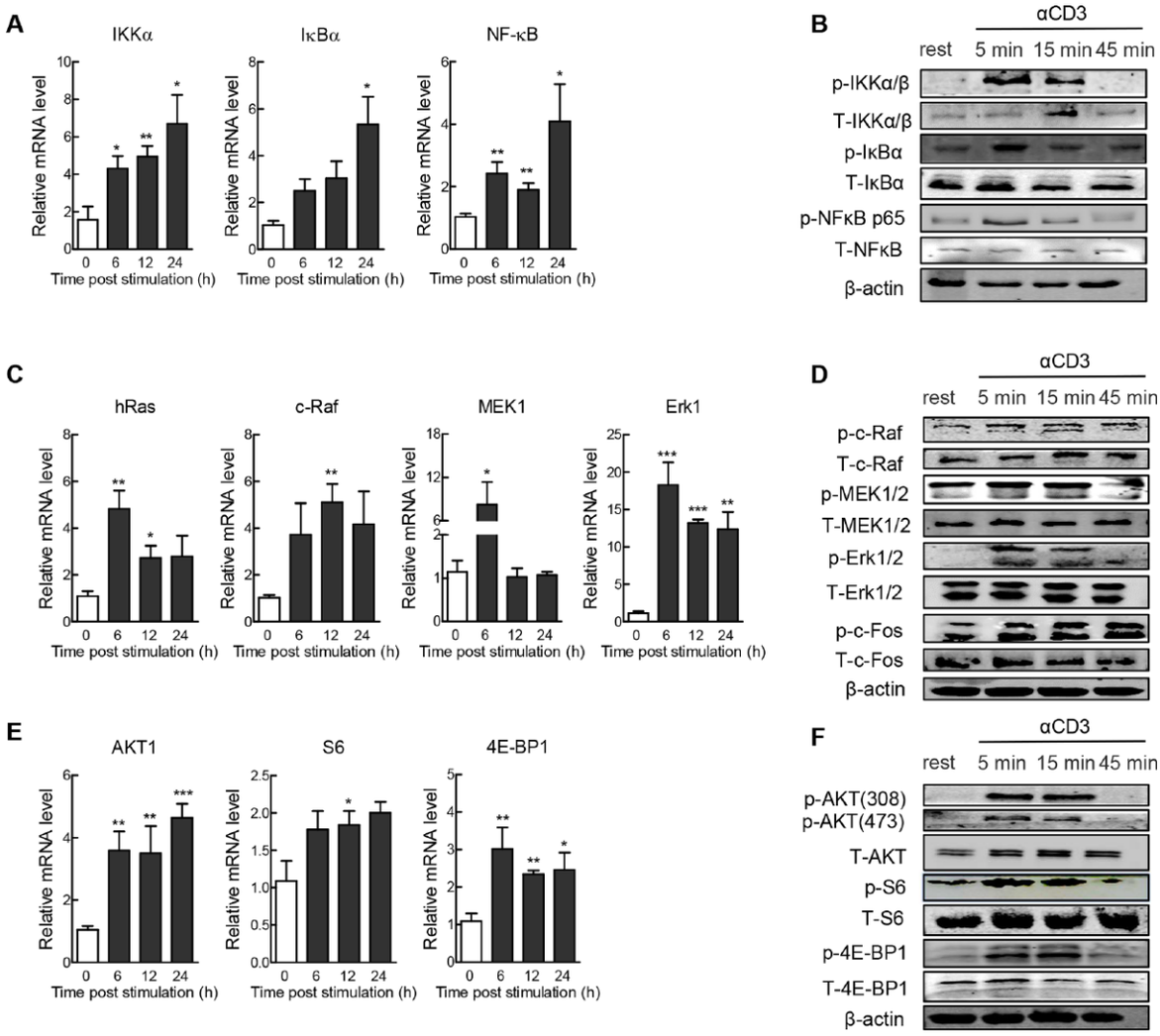


Figure 3 Activation of canonical T-cell pathways by first signal in tilapia

A, B: mRNA and phosphorylation levels of NF-κB signaling components in spleen (A) or head-kidney (B) leukocytes following CD3ε mAb stimulation at indicated times. C, D: mRNA and phosphorylation levels of MAPK/ERK signaling components in spleen (C) or head-kidney (D) leukocytes following CD3ε mAb stimulation at indicated times. E, F: mRNA and phosphorylation levels of PI3K-AKT signaling components in spleen (E) or head-kidney (F) leukocytes following CD3ε mAb stimulation at indicated times. Experiments were conducted at least twice independently. *: P<0.05; **: P<0.01; ***: P<0.001.

function (Smith-Garvin et al., 2009). Given the crucial roles of both helper and cytotoxic T cells in T-cell function and infection elimination, we evaluated the immune responses related to helper and cytotoxic T cells in lymphocytes upon CD3ε plus CD28 mAb stimulation. Compared to the unstimulated control, dual signal-induced T-cell activation resulted in a marked up-regulation in the mRNA levels of T-bet and GATA3, key factors driving the differentiation of $CD4^+$ T cells towards Th1 and Th2 phenotypes (Figure 6B). The transcript levels of cytotoxic genes perforin A and granzyme B, were also significantly induced (Figure 6C). Additionally, synergistic CD3ε and CD28 mAb stimulation increased the expression of apoptosis-related caspase genes, including caspase-3, caspase-6, caspase-8, and caspase-9 (Figure 6D). This resulted in an increase in $CD3^+$ T cell apoptosis (Figure 6E), suggesting that the dual signals triggered activation-induced apoptosis in tilapia T cells. Overall, our findings support the notion that dual signal cooperation is essential for T-cell function and plays a potential role in T-cell apoptosis to maintain immune homeostasis in tilapia.

DISCUSSION

In response to the variable and complex microbial environment, teleosts have evolved a complete T-cell-mediated adaptive immune system based on their innate immune system (Dickerson & Findly, 2017; Lu & Chen, 2019; Ren et al., 2019; Wilson, 2017). Recent studies on the identification of multiple T-cell subpopulations and response patterns after infection strongly underpin the indispensable role of T-cell immunity in fish (Ai et al., 2022; Tafalla et al., 2016; Yamaguchi et al., 2019). To date, however, it remains unclear whether the “two-signal hypothesis” described in mammals also governs the T-cell response in fish. Using the tilapia CD3ε and CD28 mAbs previously developed in our laboratory, we examined the impact of the first and second signals on T-cell activation in teleosts.

T-cell activation is initiated by complex intercellular interactions. Assembly of the CD3-TCR complex is essential for recognition of the MHC-antigenic peptide complex and transmission of TCR signals (Gorentla & Zhong, 2012; Hara et al., 1985; Smith-Garvin et al., 2009). CD3 aids in stabilizing

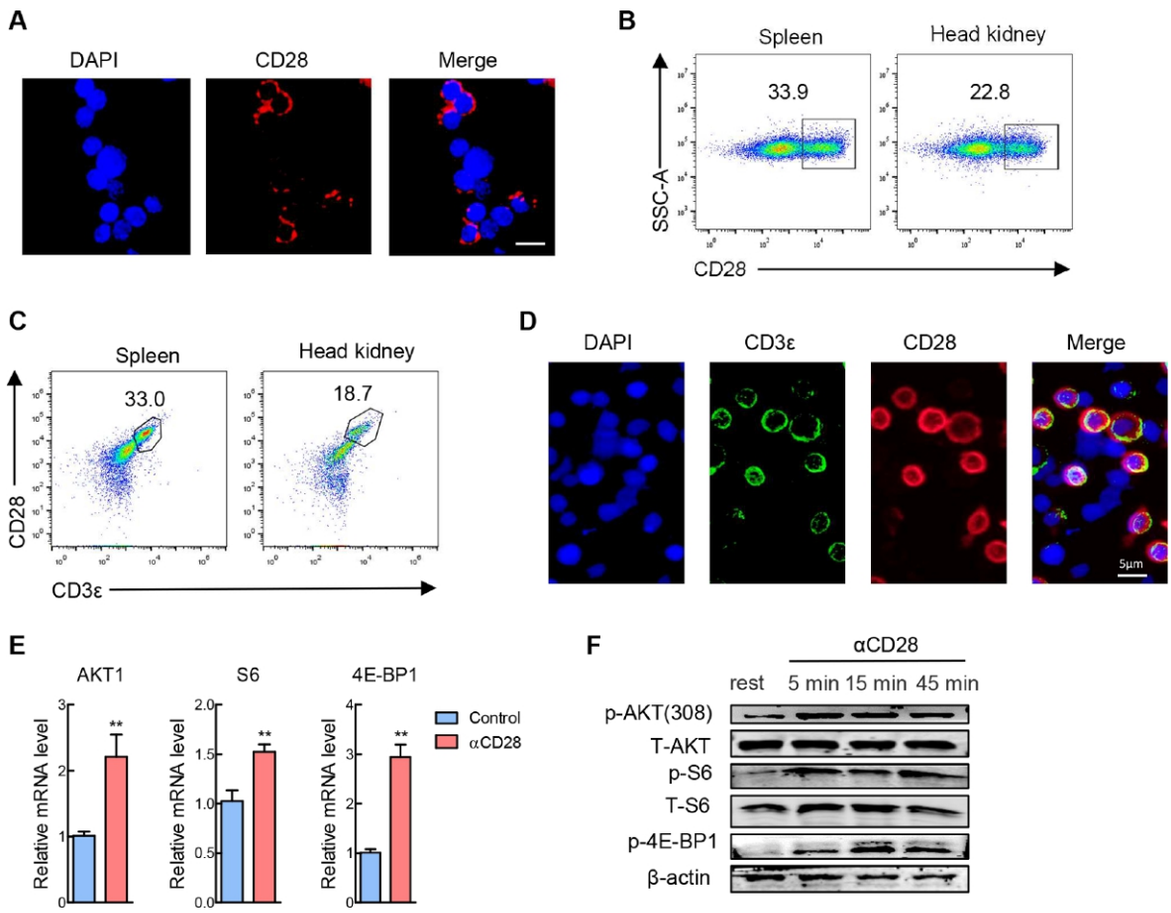


Figure 4 CD28 mAb stimulation induces AKT-mTORC1 signaling in tilapia

A: Localization of CD28 in spleen lymphocytes detected by confocal microscopy. B: $CD28^+$ lymphocytes in spleen and head kidney detected by flow cytometry. C: $CD3\varepsilon^+CD28^+$ lymphocytes in spleen and head kidney detected by flow cytometry. D: Localization of CD3εCD28 in spleen lymphocytes detected by confocal microscopy. E: mRNA levels of mTORC1 signaling components determined 12 h after spleen lymphocytes were stimulated with CD28 mAb, $n \geq 4$. F: Phosphorylation levels of mTORC1 signaling components determined after head-kidney leukocytes were stimulated with CD28 mAbs. Experiments were conducted at least twice independently. **: $P<0.01$.

the immunoreceptor tyrosine activation motif in its cytoplasmic region and transmitting TCR activation signals, triggering a kinase activation cascade to promote T-cell activation (Alcover & Alarcón, 2000; Kuhns et al., 2006). In mammals, functional mAbs against CD3ε chains have been developed and used to induce T-cell activation *in vitro*, providing a valuable tool for investigating T-cell-mediated immune mechanisms (Woodle et al., 1991). While CD3 antibodies have been established in various teleosts, no antibodies have been generated for stimulating T-cell activation, with previous teleost studies relying on PHA or PMA plus ionomycin stimulation to induce T-cell activation *in vitro* (Miyazawa et al., 2018; Mu et al., 2022; Tang et al., 2017). In the present study, Nile tilapia T-cell activation was successfully induced *in vitro* using CD3ε mAbs. The first signal mimicked by CD3ε mAbs induced LCK and ZAP-70 activation, implying the formation of TCR signaling equivalent to that observed in mammals (Gorentla & Zhong, 2012). Downstream of ZAP-70, LAT signalosome formation through the cooperative interactions of LAT with other molecules is crucial for T-cell responses (Liu et al., 2020a). Upon TCR engagement, the interactions of LAT with GADS, GRB2, and PLC-γ1 are indispensable for T-cell development (Zhu et al., 2003). Furthermore, the multivalent adaptor protein SLP-76 interacts with LAT and GADS to stabilize their persistent signaling clusters for T-cell activation (Bunnell et al., 2006). To date, however, no previous studies have identified the LAT signalosome in early vertebrates, particularly teleosts. In our research, we not only confirmed the presence of potential LAT signalosome components in tilapia, including LAT, ITK, GRB2, VAV1, SLP-76, GADS, and PLC-γ1, but also demonstrated that their expression was induced by activation of the first signal. Of these, LAT cooperatively interacted with ITK, GRB2, and VAV1 to transmit TCR signals. Intriguingly, the binding of tilapia LAT to SLP-76, GADS, or PLC-γ1 was not detected by Co-IP assay, implying the possibility of functional differences between teleosts and mammals. However, considering that Co-IP assays based on overexpression systems cannot exclude transient or indirect binding, the detailed contribution of these molecules in forming LAT signalosomes remains unclear. Thus, further study is required to elucidate whether SLP-76, GADS, and PLCγ1 are involved in the formation of LAT signalosomes to regulate T-cell activation in tilapia.

Activation of PLC-γ1 may facilitate the response of components within the NF-κB and MAPK/ERK pathways (Nishibe et al., 1990). In mammals, accumulating evidence suggests that TCR engagement may drive NF-κB signaling, which not only governs T-cell activation, proliferation, and cytokine release, but also T-cell development, survival, and function (Hayden & Ghosh, 2011; Li & Verma, 2002). Our previous studies on tilapia revealed that PHA or combined PMA plus ionomycin stimulation activated the NF-κB and

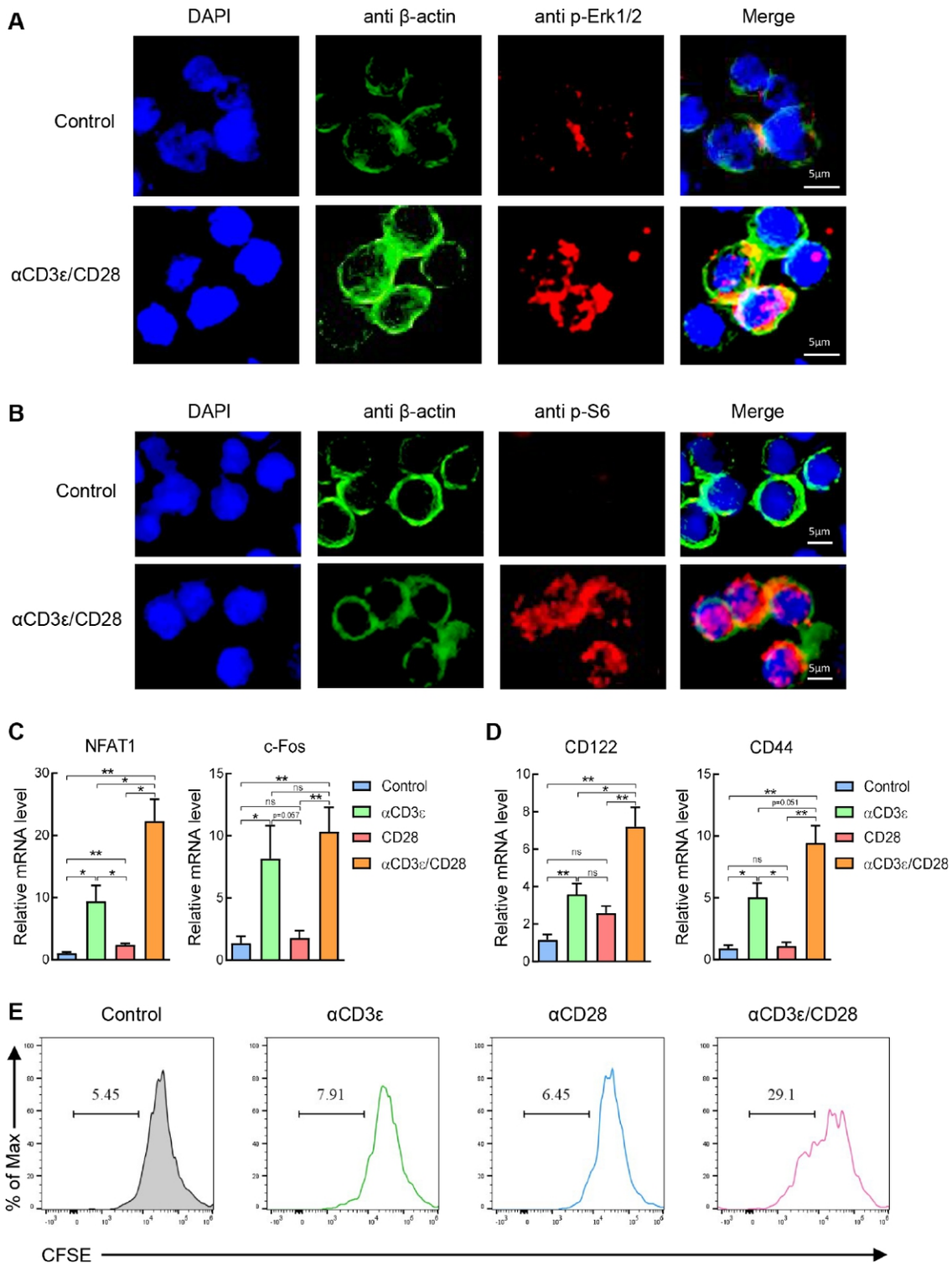


Figure 5 Dual signals trigger full activation of T cells in tilapia

A, B: Phosphorylation of ERK1/2 or S6 examined by confocal microscopy after spleen lymphocytes were stimulated with CD3ε plus CD28 mAbs for 12 h. C, D: mRNA levels of indicated molecules determined 12 h after spleen lymphocytes were stimulated with CD3ε mAbs, CD28 mAbs, or CD3ε plus CD28 mAbs, n≥4. E: Proliferation of spleen lymphocytes examined by flow cytometry 48 h after cells were stimulated with CD3ε mAbs, CD28 mAbs, or CD3ε plus CD28 mAbs, with unstimulated lymphocytes used as a control. Experiments were conducted at least twice independently. ns: No significance; *: P<0.05; **: P<0.01; ***: P<0.001.

MAPK/ERK pathways, thereby orchestrating the activation, proliferation, differentiation, and function of T cells (Wei et al., 2020, 2021). The present study supports our previous findings, asserting the involvement of the NF-κB and MAPK/ERK signaling pathways in T-cell activation after CD3ε mAb stimulation. Downstream of MAPK/ERK signaling, the transcription factor c-Fos, an AP-1 component member, plays an important role in the regulation of numerous processes, including cell migration, proliferation, differentiation, apoptosis, and angiogenesis (You et al., 2016). In fish species such as

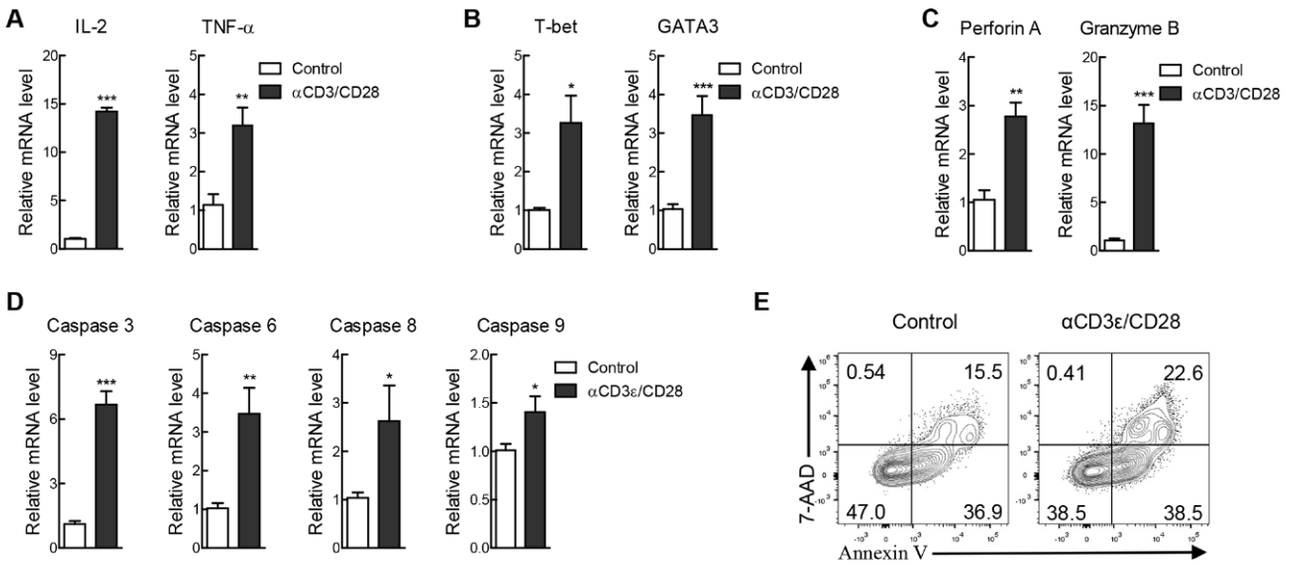


Figure 6 Dual signal-induced T-cell activation is accompanied by cytotoxic responses and apoptosis

A–D: mRNA levels of indicated molecules in spleen lymphocytes examined 12 h after cell stimulation with CD3ε plus CD28 mAbs, $n \geq 4$. E: Representative contour plot of Annexin V and 7AAD staining of CD3ε$^+$ T cells 12 h after spleen lymphocytes were stimulated with CD3ε plus CD28 mAbs. Experiments were conducted at least twice independently. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$.

Liza haematocheila, c-Fos expression can be strongly induced by poly I:C or *Lactococcus garvieae* stimulation, suggesting the possible involvement of c-Fos in the fish immune response (Janson et al., 2019). Consistent with this finding, we demonstrated the response of c-Fos to TCR signaling in tilapia T cells, highlighting its potential role in regulating T-cell biological processes. Notably, the PI3K/AKT signaling pathway was similarly induced upon CD3ε mAb stimulation. Drawing parallels from mammalian systems, we theorize that heightened Ras activity may modulate PI3K/AKT signaling (Chappell et al., 2011), although this hypothesis requires further investigation in future studies. Nevertheless, our results support the notion that the first signal plays a vital role in initiating T-cell activation in teleosts.

Over the last fifty years, extensive research has explored many co-stimulatory receptors (CD28–CD80/CD86) on the surface of T cells, which spatiotemporally regulate the biology of these cells (Esensten et al., 2016; June et al., 1990). The binding of the key co-stimulatory receptor CD28 to the B7 family molecules CD80/CD86 generates co-stimulatory signals that promote T-cell activation, proliferation, survival, and cytokine secretion (Esensten et al., 2016; June et al., 1990; Tavano et al., 2006). Positive stimulatory effects of CD28 molecules have been reported in several teleost species, including rainbow trout, grass carp, and half-smooth tongue soles (Hu et al., 2012; Lu et al., 2022; Zhang et al., 2009). For instance, the cytoplasmic tail of trout CD28 mediates ERK phosphorylation, and the chimeric receptor formed by the fusion of the extracellular structure of human CD28 with the cytoplasmic tail of rainbow trout CD28 promotes TCR-induced IL-2 production in human T-cell lines, indicating that trout CD28 is indeed a positive co-stimulatory factor (Bernard et al., 2006). In the present study, we found that tilapia CD28 was highly expressed on the surface of T cells isolated from the spleen and head kidney. Furthermore, the mTORC1 pathway was strongly activated by ligation of CD28 mAbs, suggesting that tilapia CD28 is involved in T-cell activation as a positive co-stimulator.

Recent mammalian studies have shown that full T-cell activation, proliferation, survival, and cytokine release require synergistic stimulation by the first signal and CD28-mediated co-stimulatory signals (Esensten et al., 2016; June et al., 1990). Our results also showed that dual signaling by CD3ε plus CD28 mAb stimulation strongly induced mTORC1 and ERK pathway activation in T cells, demonstrating the importance of dual signaling for T-cell activation. In the absence of CD28-mediated co-stimulatory signals, T cells fail to achieve full activation, instead entering an anergic state categorized by an inability to proliferate and become hypo-responsive (Esensten et al., 2016; Gimmi et al., 1993). Similarly, our results showed that synergistic stimulation by CD3ε plus CD28 mAbs induced tilapia T-cell proliferation, while stimulation by CD3ε or CD28 mAbs alone did not induce the same proliferation, implying that tilapia T-cell proliferation requires synergistic stimulation by dual signaling. This finding may be related to the expression of IL-2, a key T cell growth factor, as the absence of CD28 signaling in mammals impairs the expression of IL-2 and further affects the expression of various effector molecules (Coppola et al., 2020). Similarly, our study showed that tilapia T cells synergistically stimulated by dual signaling robustly induced IL-2 expression, further promoting T-cell differentiation. Notably, previous research has shown that combined treatment with CD28 antibodies and PHA can stimulate the proliferation of flounder peripheral blood leukocytes and increase the transcription of cytokines IL-2, TNF-α, IFN-γ, and IL-6 (Xing et al., 2021). Remarkably, in half-smooth tongue soles, the binding of CD28 antibodies alone significantly induces kidney lymphocyte proliferation, suggesting considerable differences among teleost CD28 molecules (Hu et al., 2012). In addition, our results emphasized that dual signaling synergism induced T-cell apoptosis in tilapia, a phenomenon likely attributed to activation-induced apoptosis as opposed to apoptosis resulting from CD28 deficiency (Green et al., 2003).

In summary, our study demonstrated that T-cell-mediated adaptive immune responses in tilapia are regulated by synergy between the CD3/TCR-mediated first signal and CD28-mediated co-stimulatory signal. Notably, dual signaling

is required for complete T-cell activation, subsequently inducing T cells to enter a proliferative and responsive state. Moreover, co-stimulation by dual signaling plays a critical role in activation-induced T-cell apoptosis as a mechanism for immune homeostasis maintenance. Thus, our results suggest that dual signaling is essential for T-cell functions in response to variable microbial environments. To the best of our knowledge, this study is the first to describe dual signaling-mediated T-cell responses in teleosts. Overall, these findings provide a robust basis for identifying T-cell immune mechanisms in teleosts and offer a novel perspective for understanding the evolution of the adaptive immune system.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

W.L. designed and performed the experiments, analyzed the data, and drafted the manuscript. Kang L., H.G., Kunming L., J.Z., Q.Z., and X.J. performed the experiments. J.Y. and X.W. acquired funding, conceived the project, designed the experiments, analyzed the data, and drafted the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the Instrument-Sharing Platform of the School of Life Sciences of East China Normal University (ECNU) for instrument sharing and the Flow Cytometry Core Facility of the School of Life Sciences of ECNU for FACS analysis.

REFERENCES

Abelli L, Baldassini MR, Mastrolia L, et al. 1999. Immunodetection of lymphocyte subpopulations involved in allograft rejection in a teleost, *Dicentrarchus labrax* (L.). *Cellular Immunology*, **191**(2): 152−160.

Ai K, Li K, Jiao XY, et al. 2022. IL-2-mTORC1 signaling coordinates the STAT1/T-bet axis to ensure Th1 cell differentiation and anti-bacterial immune response in fish. *PLoS Pathogens*, **18**(10): e1010913.

Alcover A, Alarcón B. 2000. Internalization and intracellular fate of TCR-CD3 complexes. *Critical Reviews in Immunology*, **20**(4): 325−346.

Anderson MK, Pant R, Miracle AL, et al. 2004. Evolutionary origins of lymphocytes: ensembles of T cell and B cell transcriptional regulators in a cartilaginous fish. *The Journal of Immunology*, **172**(10): 5851−5860.

Ashfaq H, Soliman H, Saleh M, et al. 2019. CD4: a vital player in the teleost fish immune system. *Veterinary Research*, **50**(1): 1.

Au-Yeung BB, Shah NH, Shen L, et al. 2018. ZAP-70 in signaling, biology, and disease. *Annual Review of Immunology*, **36**: 127−156.

Bernard D, Riteau B, Hansen JD, et al. 2006. Costimulatory receptors in a teleost fish: typical CD28, elusive CTLA4. *The Journal of Immunology*, **176**(7): 4191−4200.

Beyersdorf N, Kerkau T, Hünig T. 2015. CD28 co-stimulation in T-cell homeostasis: a recent perspective. *ImmunoTargets and Therapy*, **4**: 111−122.

Boardman T, Warner C, Ramirez-Gomez F, et al. 2012. Characterization of an anti-rainbow trout (*Oncorhynchus mykiss*) CD3ε monoclonal antibody. *Veterinary Immunology and Immunopathology*, **145**(1-2): 511−515.

Bunnell SC, Diehn M, Yaffe MB, et al. 2000. Biochemical interactions integrating Itk with the T cell receptor-initiated signaling cascade. *Journal of Biological Chemistry*, **275**(3): 2219−2230.

Bunnell SC, Singer AL, Hong DI, et al. 2006. Persistence of cooperatively stabilized signaling clusters drives T-cell activation. *Molecular and Cellular Biology*, **26**(19): 7155−7166.

Caporali R, Bugatti S, Cavagna L, et al. 2014. Modulating the co-stimulatory signal for T cell activation in rheumatoid arthritis: could it be the first step of the treatment?. *Autoimmunity Reviews*, **13**(1): 49−53.

Chappell WH, Steelman LS, Long JM, et al. 2011. Ras/Raf/MEK/ERK and PI3K/PTEN/Akt/mTOR inhibitors: rationale and importance to inhibiting these pathways in human health. *Oncotarget*, **2**(3): 135−164.

Chen LP, Flies DB. 2013. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nature Reviews Immunology*, **13**(4): 227−242.

Cheng J, Montecalvo A, Kane LP. 2011. Regulation of NF-κB induction by TCR/CD28. *Immunologic Research*, **50**(2-3): 113−117.

Chitnis T, Kaskow BJ, Case J, et al. 2022. Nasal administration of anti-CD3 monoclonal antibody modulates effector CD8+ T cell function and induces a regulatory response in T cells in human subjects. *Frontiers In Immunology*, **13**: 956907.

Coppola C, Hopkins B, Huhn S, et al. 2020. Investigation of the impact from IL-2, IL-7, and IL-15 on the growth and signaling of activated $CD4^+$ T cells. *International Journal of Molecular Sciences*, **21**(21): 7814.

Dickerson HW, Findly RC. 2017. Vertebrate adaptive immunity-comparative insights from a teleost model. *Frontiers in Immunology*, **8**: 1379.

Esensten JH, Helou YA, Chopra G, et al. 2016. CD28 costimulation: from mechanism to therapy. *Immunity*, **44**(5): 973−988.

Gimmi CD, Freeman GJ, Gribben JG, et al. 1993. Human T-cell clonal anergy is induced by antigen presentation in the absence of B7 costimulation. *Proceedings of the National Academy of Sciences of the United States of America*, **90**(14): 6586−6590.

Gorentla BK, Wan CK, Zhong XP. 2011. Negative regulation of mTOR activation by diacylglycerol kinases. *Blood*, **117**(15): 4022−4031.

Gorentla BK, Zhong XP. 2012. T cell receptor signal transduction in T lymphocytes. *Journal of Clinical Immunology*, **2012**(S12): 5.

Green DR, Droin N, Pinkoski M. 2003. Activation-induced cell death in T cells. *Immunological Reviews*, **193**(1): 70−81.

Hara T, Fu SM, Hansen JA. 1985. Human T cell activation. II. A new activation pathway used by a major T cell population via a disulfide-bonded dimer of a 44 kilodalton polypeptide (9.3 antigen). *Journal of Experimental Medicine*, **161**(6): 1513−1524.

Hayden MS, Ghosh S. 2011. NF-κB in immunobiology. *Cell Research*, **21**(2): 223−244.

Hu YH, Sun BG, Deng T, et al. 2012. Molecular characterization of *Cynoglossus semilaevis* CD28. *Fish & Shellfish Immunology*, **32**(5): 934−938.

Janson ND, Jehanathan N, Jung S, et al. 2019. Insight into the molecular function and transcriptional regulation of activator protein 1 (AP-1) components c-Jun/c-Fos ortholog in red lip mullet (*Liza haematocheila*). *Fish & Shellfish Immunology*, **93**: 597−611.

June CH, Ledbetter JA, Linsley PS, et al. 1990. Role of the CD28 receptor in T-cell activation. *Immunology Today*, **11**: 211−216.

Jung G, Ledbetter JA, Müller-Eberhard HJ. 1987. Induction of cytotoxicity in resting human T lymphocytes bound to tumor cells by antibody heteroconjugates. *Proceedings of the National Academy of Sciences of the United States of America*, **84**(13): 4611−4615.

Jung JW, Lee AR, Kim J, et al. 2021. Elucidating the functional roles of helper and cytotoxic T cells in the cell-mediated immune responses of olive flounder (*Paralichthys olivaceus*). *International Journal of Molecular Sciences*, **22**(2): 847.

Jung JW, Lee JS, Kim YR, et al. 2017. Development of a monoclonal antibody against the CD3ε of olive flounder (*Paralichthys olivaceus*) and its application in evaluating immune response related to CD3ε. *Fish & Shellfish Immunology*, **65**: 179−185.

Kuhns MS, Davis MM, Garcia KC. 2006. Deconstructing the form and function of the TCR/CD3 complex. *Immunity*, **24**(2): 133−139.

Laing KJ, Hansen JD. 2011. Fish T cells: recent advances through genomics. *Developmental & Comparative Immunology*, **35**(12): 1282−1295.

Lesslauer W, Koning F, Ottenhoff T, et al. 1986. T90/44 (9.3 antigen). A cell surface molecule with a function in human T cell activation. *European Journal of Immunology*, **16**(10): 1289−1296.

Li JQ, Liang W, Li K, et al. 2021. ZAP70 activation is an early event of T cell immunity that involved in the anti-bacterial adaptive immune response of Nile tilapia. *Developmental & Comparative Immunology*, **124**: 104177.

Li K, Wei XM, Jiao XY, et al. 2023. Glutamine metabolism underlies the functional similarity of T cells between nile tilapia and tetrapod. *Advanced Science*, **10**(12): e2201164.

Li QT, Verma IM. 2002. NF-κB regulation in the immune system. *Nature Reviews Immunology*, **2**(10): 725−734.

Liang W, Li KM, Zhang Q, et al. 2022. Interleukin-2 inducible T cell kinase (ITK) may participate in the anti-bacterial immune response of Nile tilapia via regulating T-cell activation. *Fish & Shellfish Immunology*, **127**: 419−426.

Liu C, Xu XY, Han L, et al. 2020a. LRCH1 deficiency enhances LAT signalosome formation and $CD8^+$ T cell responses against tumors and pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(32): 19388−19398.

Liu Y, Liu YE, Tong CC, et al. 2020b. CD28 deficiency attenuates primary blast-induced renal injury in mice via the PI3K/Akt signalling pathway. *BMJ Military Health*, **166**(E): e66−e69.

Lu TZ, Liu X, Wu CS, et al. 2022. Molecular and functional analyses of the primordial costimulatory molecule CD80/86 and its receptors CD28 and CD152 (CTLA-4) in a teleost fish. *Frontiers in Immunology*, **13**: 885005.

Lu XJ, Chen J. 2019. Specific function and modulation of teleost monocytes/macrophages: polarization and phagocytosis. *Zoological Research*, **40**(3): 146−150.

Marinari B, Costanzo A, Marzano V, et al. 2004. CD28 delivers a unique signal leading to the selective recruitment of RelA and p52 NF-κB subunits on *IL-8* and *Bcl-xL* gene promoters. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(16): 6098−6103.

Mariuzza RA, Agnihotri P, Orban J. 2020. The structural basis of T-cell receptor (TCR) activation: an enduring enigma. *Journal of Biological Chemistry*, **295**(4): 914−925.

Miyazawa R, Murata N, Matsuura Y, et al. 2018. Peculiar expression of CD3-epsilon in kidney of ginbuna crucian carp. *Frontiers In Immunology*, **9**: 1321.

Mu PF, Huo JY, Li XF, et al. 2022. IL-2 signaling couples the MAPK and mTORC1 axes to promote T cell proliferation and differentiation in teleosts. *The Journal of Immunology*, **208**(7): 1616−1631.

Nguyen TD, Carrascal M, Vidal-Cortes O, et al. 2016. The phosphoproteome of human Jurkat T cell clones upon costimulation with anti-CD3/anti-CD28 antibodies. *Journal of Proteomics*, **131**: 190−198.

Nishibe S, Wahl MI, Hernandez-Sotomayor SM, et al. 1990. Increase of the catalytic activity of phospholipase C-γ1 by tyrosine phosphorylation. *Science*, **250**(4985): 1253−1256.

Qin YT, Sun ZS, Wang W, et al. 2021. Characterization of $CD3\gamma/\delta^+$ cells in grass carp (*Ctenopharyngodon idella*). *Developmental & Comparative Immunology*, **114**: 103791.

Randelli E, Buonocore F, Scapigliati G. 2008. Cell markers and determinants in fish immunology. *Fish & Shellfish Immunology*, **25**(4): 326−340.

Ren Y, Liu SF, Nie L, et al. 2019. Involvement of ayu NOD2 in NF-κB and MAPK signaling pathways: insights into functional conservation of NOD2 in antibacterial innate immunity. *Zoological Research*, **40**(2): 77−88.

Rodríguez-Caparrós A, Álvarez-Santiago J, DEL Valle-Pastor MJ, et al. 2020. Regulation of T-cell receptor gene expression by three-dimensional locus conformation and enhancer function. *International Journal of Molecular Sciences*, **21**(22): 8478.

Sanchez-Lockhart M, Kim M, Miller J. 2011. Cutting edge: a role for inside-out signaling in TCR regulation of CD28 ligand binding. *The Journal of Immunology*, **187**(11): 5515−5519.

Sauerwein RW, Der Van Meer WG, Van Oostveen JW, et al. 1988. Anti-CD3 antibodies induce T helper function for human B cell differentiation in vitro by an interleukin 2-independent pathway. *European Journal of Immunology*, **18**(1): 133−137.

Schwartzberg PL, Finkelstein LD, Readinger JA. 2005. TEC-family kinases: regulators of T-helper-cell differentiation. *Nature Reviews Immunology*, **5**(4): 284−295.

Sela M, Bogin Y, Beach D, et al. 2011. Sequential phosphorylation of SLP-76 at tyrosine 173 is required for activation of T and mast cells. *The EMBO Journal*, **30**(15): 3160−3172.

Sloan-Lancaster J, Evavold BD, Allen PM. 1993. Induction of T-cell anergy by altered T-cell-receptor ligand on live antigen-presenting cells. *Nature*, **363**(6425): 156−159.

Smith-Garvin JE, Koretzky GA, Jordan MS. 2009. T cell activation. *Annual Review of Immunology*, **27**: 591−619.

Tafalla C, Leal E, Yamaguchi T, et al. 2016. T cell immunity in the teleost digestive tract. *Developmental and Comparative Immunology*, **64**: 167−177.

Tang XQ, Qin YH, Sheng XZ, et al. 2017. Characterization of $CD3^+$ T lymphocytes of Japanese flounder (*Paralichthys olivaceus*) and its response after immunization with formalin-inactivated *Edwardsiella tarda*. *Fish & Shellfish Immunology*, **63**: 220–227.

Tavano R, Contento RL, Baranda SJ, et al. 2006. CD28 interaction with filamin-A controls lipid raft accumulation at the T-cell immunological synapse. *Nature Cell Biology*, **8**(11): 1270−1276.

Wei XM, Ai K, Li HY, et al. 2019. Ancestral T cells in fish require mTORC1-coupled immune signals and metabolic programming for proper activation and function. *The Journal of Immunology*, **203**(5): 1172−1188.

Wei XM, Li C, Zhang Y, et al. 2021. Fish NF-κB couples TCR and IL-17 signals to regulate ancestral T-cell immune response against bacterial infection. *FASEB Journal*, **35**(4): e21457.

Wei XM, Zhang Y, Li C, et al. 2020. The evolutionarily conserved MAPK/ERK signaling promotes ancestral T-cell immunity in fish via c-Myc-mediated glycolysis. *Journal of Biological Chemistry*, **295**(10): 3000−3016.

Wilson AB. 2017. MHC and adaptive immunity in teleost fishes. *Immunogenetics*, **69**(8-9): 521−528.

Woodle ES, Thistlethwaite JR, Jolliffe LK, et al. 1991. Anti-CD3 monoclonal antibody therapy. An approach toward optimization by in vitro analysis of new anti-CD3 antibodies. *Transplantation*, **52**(2): 361−368.

Xing J, Liu WJ, Tang XQ, et al. 2021. The expression of CD28 and its synergism on the immune response of flounder (*Paralichthys olivaceus*) to thymus-dependent antigen. *Frontiers in Immunology*, **12**: 765036.

Yamaguchi T, Takizawa F, Furihata M, et al. 2019. Teleost cytotoxic T cells. *Fish & Shellfish Immunology*, **95**: 422−439.

You L, Ren XX, Du YX, et al. 2016. c-Fos/ERK promotes the progression from pancreatic intraepithelial neoplasia to pancreatic ductal adenocarcinoma. *Oncology Reports*, **36**(6): 3413−3420.

Zhang YA, Hikima JI, Li J, et al. 2009. Conservation of structural and functional features in a primordial CD80/86 molecule from rainbow trout (*Oncorhynchus mykiss*), a primitive teleost fish. *The Journal of Immunology*, **183**(1): 83−96.

Zhu MH, Janssen E, Zhang WG. 2003. Minimal requirement of tyrosine residues of linker for activation of T cells in TCR signaling and thymocyte development. *The Journal of Immunology*, **170**(1): 325−333.

Ning et al. *Zool. Res.* 2024, 45(1): 25−35
https://doi.org/10.24272/j.issn.2095-8137.2022.520

ZR | Zoological Research

Article Open Access

LncRNA pol-lnc78 as a ceRNA regulates antibacterial responses via suppression of pol-miR-n199-3p-mediated SARM down-regulation in *Paralichthys olivaceus*

Xian-Hui Ning[1,2,3], Bing Han[1], Ye Peng[1], Shao-Wu Yin[1,3,*]

[1] *College of Marine Science and Engineering*, *Nanjing Normal University*, *Nanjing*, *Jiangsu* 210023, *China*

[2] *CAS Key Laboratory of Experimental Marine Biology*, *Center for Ocean Mega-Science*, *Institute of Oceanology*, *Chinese Academy of Sciences*, *Qingdao*, *Shandong* 266071, *China*

[3] *Co-Innovation Center for Marine Bio-Industry Technology of Jiangsu Province*, *Lianyungang*, *Jiangsu* 222005, *China*

ABSTRACT

Long non-coding RNAs (lncRNAs) function as key modulators in mammalian immunity, particularly due to their involvement in lncRNA-mediated competitive endogenous RNA (ceRNA) crosstalk. Despite their recognized significance in mammals, research on lncRNAs in lower vertebrates remains limited. In the present study, we characterized the first immune-related lncRNA (pol-lnc78) in the teleost Japanese flounder (*Paralichthys olivaceus*). Results indicated that pol-lnc78 acted as a ceRNA for pol-miR-n199-3p to target the sterile alpha and armadillo motif-containing protein (SARM), the fifth discovered member of the Toll/interleukin 1 (IL-1) receptor (TIR) adaptor family. This ceRNA network regulated the antibacterial responses of flounder via the Toll-like receptor (TLR) signaling pathway. Specifically, SARM acted as a negative regulator and exacerbated bacterial infection by inhibiting the expression of inflammatory cytokines IL-1β and tumor necrosis factor-α (TNF-α). Pol-miR-n199-3p reduced SARM expression by specifically interacting with the 3' untranslated region (UTR), thereby promoting SARM-dependent inflammatory cytokine expression and protecting the host against bacterial dissemination. Furthermore, pol-lnc78 sponged pol-miR-n199-3p to ameliorate the inhibition of SARM expression. During infection, the negative regulators pol-lnc78 and SARM were significantly down-regulated, while pol-miR-n199-3p was significantly up-regulated, thus favoring host antibacterial defense. These findings provide novel insights into the mechanisms underlying fish immunity and open new horizons to better understand ceRNA crosstalk in lower vertebrates.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Keywords: LncRNA; SARM; miRNA; ceRNA; Antibacterial response

INTRODUCTION

Sterile alpha and armadillo motif-containing protein (SARM), the last discovered Toll/interleukin 1 (IL-1) receptor (TIR) adaptor, exhibits a highly conserved sequence across evolution, from nematodes to mammals (Jault et al., 2004; Lindsay & Wasserman, 2014; Vériepe et al., 2015; Zhou et al., 2013). Indeed, phylogenetic analysis suggests a potential bacterial origin for animal SARM, likely arising from horizontal gene transfer (Malapati et al., 2017; Zhang et al., 2011). Like the other four TIR adaptors, myeloid differentiation primary response 88 (MyD88), MyD88 adaptor-like (Mal), TIR domain-containing adapter-inducing interferon-β (TRIF), and TRIF-related adaptor molecule (TRAM), SARM features a TIR domain, which facilitates its role in Toll-like receptor (TLR) signaling and innate immunity (Panneerselvam & Ding, 2015). Functional analyses demonstrated that, across all examined species, from worms and flies to mammals, SARM significantly influences immune responses to pathogenic infections, with varied outcomes despite sequence conservation. In *Caenorhabditis elegans*, the SARM ortholog TIR-1 regulates host defense against fungal and bacterial infections, enhancing survival without relying on the sole Toll receptor (Couillault et al., 2004). In *Drosophila melanogaster*, the SARM ortholog Ect4 is vital for survival post-viral infection, possibly reducing Toll signaling to restore homeostasis (Monsanto-Hearne et al., 2017). In mammals, SARM acts as a negative regulator in immune defense against live bacterial infection by suppressing TLR signaling (Carty & Bowie, 2019). Although limited in scope, investigations conducted on teleost fish, which serve as representatives of lower vertebrates, have

Received: 27 April 2023; Accepted: 07 September 2023; Online: 08 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (42006082), Natural Science Foundation of Jiangsu Province of China (BK20221323), "JBGS" Project of Seed Industry Revitalization in Jiangsu Province (JBGS [2021] 034), and State Key Laboratory of Developmental Biology of Freshwater Fish (2021KF009)

*Corresponding author, E-mail: yinshaowu@163.com

indicated the involvement of SARM in the immune responses of grass carp to grass carp reovirus (GCRV) infection and mandarin fish to stimulation with poly (I:C) and lipopolysaccharide (LPS) (Wang et al., 2021a; Yan et al., 2015). Nonetheless, the precise role of SARM in bacterial infections in fish remains undetermined, and its immunological mechanism remains elusive. Recent research on worms and flies has revealed that SARM expression is mediated by non-coding RNAs (ncRNAs), specifically miR-956 and miR-71, respectively (Finger et al., 2019; Monsanto-Hearne et al., 2017), providing a crucial clue for elucidating the regulatory mechanisms of SARM.

NcRNAs, including microRNAs (miRNAs), lncRNAs, and circular RNAs (circRNAs), have attracted considerable attention in gene regulation. As highly conserved small ncRNAs (~22 nucleotides long) (Cheng et al., 2005), miRNAs play vital roles in biological processes, such as growth, development, reproduction, and immunity (Andreassen & Høyheim, 2017; Hatfield et al., 2005; Sarma et al., 2014; Vasadia et al., 2019; Wienholds & Plasterk, 2005). The regulatory mechanisms governing miRNAs are well characterized, involving the repression of gene expression by promoting mRNA degradation or suppressing mRNA translation (Bagga et al., 2005). LncRNAs, a class of ncRNAs longer than 200 nucleotides, can be distinguished by their transcriptional but non-translational nature (Kapranov et al., 2007). LncRNAs exert diverse effects on development, differentiation, metabolism, and immunity (Cesana et al., 2011; Du et al., 2017; Sarangdhar et al., 2018; Wang et al., 2017b). LncRNAs can regulate gene expression directly through *cis*- or *trans*-acting mechanisms (Ponting et al., 2009). Recently, a novel mechanism of lncRNA-mediated regulation has emerged, whereby lncRNAs function as miRNA sponges to modulate target genes, known as competitive endogenous RNA (ceRNA) activity (Wang et al., 2013). LncRNA-mediated ceRNA crosstalk has been extensively detected in mammalian immunity (Cong et al., 2019; Song et al., 2017; Wu et al., 2017). In the context of fish, it is anticipated that ceRNA networks participate in immune responses (Wu et al., 2021, 2022; Ye et al., 2021). However, given that the intricacies of the lncRNA-miRNA-mRNA regulatory mechanism remain poorly understood, with only recent studies clarifying this mechanism in miiuy croaker antipathogen infections (Chu et al., 2020, 2021b; Zheng et al., 2021a, 2021b), assessing the extent of the ceRNA mechanism across different fish species remains crucial.

Teleosts are representative of early vertebrate evolution and can serve as key models in lower vertebrate studies. However, these economically valuable aquaculture species face significant threats from bacterial diseases, amplifying their importance in immunology research. Notably, vibriosis, a prevalent aquatic disease caused primarily by *Vibrio anguillarum* (Egidius, 1987), has prompted great research efforts into fish defense mechanisms (Ning & Sun, 2020a, 2020b, 2021a, 2021b; Qi et al., 2021; Wang et al., 2021b; Xu et al., 2019). Nevertheless, in-depth investigations on fish defense responses remain severely impeded due to limitations in experimental techniques, including the lack of suitable cell lines. Furthermore, fish immune processes and regulatory cascades exhibit notable differences when contrasted with the well-characterized mammalian immunity.

In this study, we identified a lncRNA-miRNA-mRNA ceRNA network, consisting of pol-miR-n199-3p, pol-lnc78, and SARM, which exhibited the capacity to regulate the host immune response in the teleost Japanese flounder (*Paralichthys olivaceus*). SARM acted as a negative regulator, inhibiting the expression of inflammatory cytokines, and was declined upon *V. anguillarum* infection. Pol-miR-n199-3p suppressed SARM expression by targeting the 3' untranslated region (UTR), thereby contributing to host defense against pathogens. Moreover, pol-lnc78 functioned as a ceRNA, counteracting the inhibitory effects of pol-miR-n199-3p on SARM, and thus regulating SARM-mediated immune responses. Our results represent the first ceRNA regulatory mechanism identified in flounder. These findings should help advance our understanding of ceRNA crosstalk in lower vertebrates.

MATERIALS AND METHODS

Fish and bacterial infection

Clinically healthy Japanese flounder (~50 g) were purchased from a commercial fish farm in Shandong Province, China, and maintained at 20±1 °C in aquariums equipped with aerated seawater and digital temperature sensing controllers as reported previously (Ning & Sun, 2020a). Fish were acclimated in the laboratory for at least one week before any experimental manipulation. Bacterial challenge was performed as reported previously (Ning & Sun, 2020a). Briefly, flounder individuals were injected intramuscularly with 100 μL of *V. anguillarum* C312, a pathogenic strain isolated from diseased flounder (Zheng et al., 2010), at a concentration of 5×10^6 colony forming units (CFU)/mL. The same volume of phosphate-buffered saline (PBS) was injected into flounder as a control. At 12 h post-infection (hpi), fish were euthanized, and spleen tissue samples were collected for RNA extraction. All experimental procedures were conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals in China and were approved by the Nanjing Normal University Animal Ethics Committee (Permit No. SYXK2015-0028).

Bio-samples from the spleen, previously utilized in a transcriptome study examining the responses of flounder to *V. anguillarum* over various time points (Ning & Sun, 2020a), were used to evaluate SARM and pol-lnc78 expression at 6, 12, and 24 hpi and for subsequent correlation analyses.

RNA extraction and reverse transcription-quantitative real-time polymerase chain reaction (RT-qPCR)

Total RNA from spleen samples was extracted using TRIzol reagent (Invitrogen, USA) following the manufacturer's protocols. The cDNA and miRNA cDNA templates were synthesized using ReverTra Ace qPCR RT Master Mix (TOYOBO, Japan) and miRNA First-Strand cDNA Synthesis Kit (Vazyme, China), respectively, according to the manufacturer's instructions. The RT-qPCR (mRNA and lncRNA) and stem-loop RT-qPCR (miRNA) procedures were carried out with SYBR Premix Ex TaqII (TaKaRa, China) using the QuantStudio 3 Real-Time PCR System (Thermo Fisher Scientific, USA) according to the manufacturer's protocols. The expression level was determined using the $2^{-\Delta\Delta Ct}$ comparative Ct method (Ning & Sun, 2020b). β-actin and α-tubulin were used as the internal controls for samples of uninfected and *V. anguillarum*-infected flounder, respectively. 5S rRNA was used as the internal control for pol-miR-n199-3p. The primers used for RT-qPCR are listed in Supplementary Table S1. Correlation analysis was performed using the cor.test in R (v.3.5.2).

Plasmid construction

To construct the SARM expression plasmid (pCN3-SARM), the *SARM* gene sequence in flounder (Gene ID 109630267) was amplified using PCR with specific primers, with the PCR product then cloned into the expression vector pCN3 at the EcoR V site as reported previously (Jiao et al., 2009). Similarly, the pol-lnc78 expression plasmid (pCN3-lnc78) was constructed by cloning the pol-lnc78 sequence (accession number PRJNA554220) into pCN3 at the EcoR V site. To construct the SARM 3'UTR reporter plasmid (pmir-SARM-wt), the 3'UTR region of SARM was amplified by PCR, with the PCR product then inserted into the pmirGLO vector at the Sac I/Xba I sites. The mutant form of the SARM 3'UTR reporter plasmid (pmir-SARM-mut) was identical to the SARM 3'UTR reporter plasmid, except that the sequence (5'-CAGTGTT-3') complementary to the seed sequence of pol-miR-n199-3p was mutated to 5'-GTCACAA-3'. Similarly, the pol-lnc78 reporter plasmid (pmir-lnc78-wt) and mutated-pol-lnc78 reporter plasmid (pmir-lnc78-mut) were constructed by inserting the pol-lnc78 sequence and pol-lnc78 with mutated sequence (5'-CAGTGTT-3' mutated to 5'-GTCACAA-3') into the pmirGLO vector, respectively. The primers used to construct plasmids are listed in Supplementary Table S1.

MiRNA mimic and agomir

Pol-miR-n199-3p mimic and agomir, as well as respective negative control miR-NC and agomir-NC, were synthesized by GenePharma (China).

Dual-luciferase reporter assay

Human embryonic kidney epithelial (HEK293T) cells were cultured in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen, USA) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin and streptomycin at 37 °C with 5% CO_2 in 24-cell plates. To investigate the interaction between pol-miR-n199-3p and SARM 3'UTR, the HEK293T cells were transfected with pmir-SARM-wt alone or co-transfected with pmir-SARM-wt or pmir-SARM-mut, along with pol-miR-n199-3p mimic or miR-NC using Lipofectamine™ 3000 (Invitrogen, USA) according to the manufacturer's instructions. After transfection for 8 h, the culture medium was renewed. At 24 h post-transfection, the cells were lysed, and luciferase activity was measured using a Dual Luciferase Reporter Assay Kit (Vazyme, China). Similarly, the interaction between pol-miR-n199-3p and pol-lnc78 was determined with HEK293T cells transfected with pmir-lnc78-wt alone or co-transfected with pmir-lnc78-wt or pmir-lnc78-mut, along with pol-miR-n199-3p mimic or miR-NC. To detect the competition between pol-lnc78 and SARM in binding to pol-miR-n199-3p, the HEK293T cells were co-transfected with pol-miR-n199-3p mimic or miR-NC, together with the pCN3 vector or pol-lnc78 expression plasmid, along with SARM 3'UTR, or transfected with SARM 3'UTR alone. All luciferase activity values were normalized against the *Renilla* luciferase control.

***In vivo* overexpression and effects on bacterial infection**

Flounder with SARM or pol-lnc78 overexpression were constructed, with four groups of flounder used. Fish in the former two groups were injected with expression plasmid pCN3-SARM or pCN3-lnc78 with 2 μg plasmid/g fish, respectively. Fish in the latter two groups were injected with or without the same volume of empty vector pCN3 as a control. Spleen tissues were collected at 3 d post-plasmid administration, and the expression levels of SARM, pol-lnc78, and pol-miR-n199-3p were examined by RT-qPCR.

To investigate the effects of SARM or pol-lnc78 overexpression on *V. anguillarum* infection, flounder were injected with the plasmid pCN3-SARM or pCN3-lnc78, as described above. At 2.5 d post-plasmid administration, flounder were injected with *V. anguillarum* (5×10^5 CFU/fish). At 12 h post *V. anguillarum* injection, number of bacteria was determined using the plate count method, as reported previously (Zhang et al., 2020).

Flounder with pol-miR-n199-3p overexpression or pol-miR-n199-3p and pol-lnc78 co-overexpression were constructed, with a total of five groups used. For the former three groups, flounder were injected with or without pol-miR-n199-3p agomir or agomir-NC (2 μg/g fish) for 12 h. For the latter two groups, fish were injected with pCN3-lnc78 or pCN3 for 2.5 d, then injected with pol-miR-n199-3p agomir (2 μg/g fish) for another 12 h. The expression levels of pol-miR-n199-3p, SARM, and pol-lnc78 were determined by RT-qPCR.

To investigate the effects of pol-miR-n199-3p overexpression on *V. anguillarum* infection, fish were injected with pol-miR-n199-3p agomir (2 μg/g fish) together with *V. anguillarum* (5×10^5 CFU/fish). At 12 hpi, the number of *V. anguillarum* in the spleen and liver was determined by plate count.

Statistical analysis

Data are presented as means±standard deviation (SD) from three independent replicates. Student's *t*-test was used to compare the values between different groups. Statistical analyses were performed using SPSS software (v.23.0). Statistical significance was considered at $P<0.05$.

RESULTS

Flounder SARM negatively regulates bacterial infection

Symptoms in fish challenged with *V. anguillarum* were first confirmed. Notably, after infection, common pathological findings of vibriosis were observed in flounder, including ulcerative skin, fin hemorrhage, and swollen spleen (Figure 1A). To evaluate the expression profiles of flounder SARM in response to *V. anguillarum* challenge, RT-qPCR was conducted. Results showed that SARM expression was significantly reduced during *V. anguillarum* infection at 6, 12, and 24 hpi (Figure 1B). In mammals, SARM is an inhibitor of TLR signaling via the suppression of IL-1β and TNF-α expression after LPS stimulation. Accordingly, we examined whether flounder SARM could inhibit these inflammatory cytokines upon bacterial infection. Fish exhibiting SARM overexpression were established using plasmid pCN3-SARM injection. Enhanced SARM expression was validated by RT-qPCR at 3 d post-administration (Supplementary Figure S1). As depicted in Figure 1C, D, *in vivo* overexpression of SARM significantly down-regulated IL-1β and TNF-α expression in *V. anguillarum*-infected flounder. To investigate the effect of SARM on pathogen infection, bacterial counts were determined in *V. anguillarum*-infected flounder with or without SARM overexpression. Results showed that bacterial loads in fish with SARM overexpression were significantly higher than that in fish without SARM overexpression in both the spleen and liver (Figure 1E, F). These findings suggest that antibacterial immune defense is triggered via suppression of SARM expression, which activates the TLR signaling pathway and inflammatory response, resulting in the clearance of bacterial pathogens.

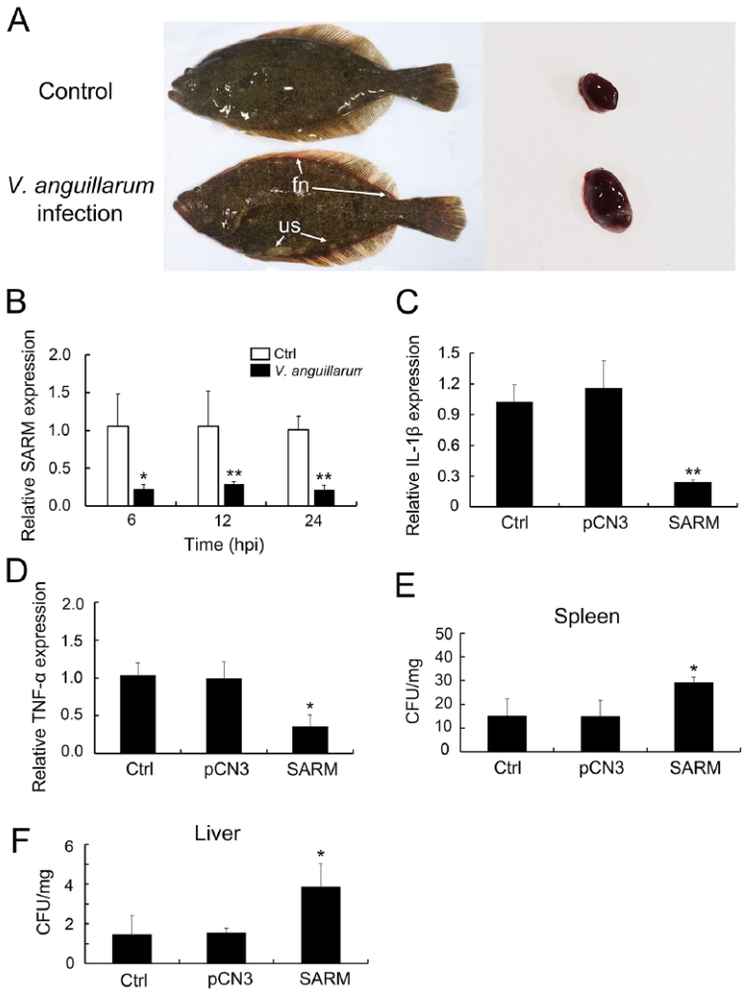


Figure 1 SARM acts as a negative regulator in host defense against *Vibrio anguillarum* infection

A: Symptoms of flounder infected with *V. anguillarum*. Compared to control fish, *V. anguillarum*-infected flounder exhibited ulcerative skin (us) and fin hemorrhage (fh). Swollen spleen was observed in *V. anguillarum*-infected flounder compared to control fish. B: SARM expression decreased in response to *V. anguillarum* infection. SARM expression in flounder infected with or without (control) *V. anguillarum* for 6, 12, and 24 h was examined. C, D: SARM inhibited expression of inflammatory cytokines. Flounder were administered with SARM expression plasmid or pCN3 vector for 2.5 d, then injected with *V. anguillarum* for another 12 h. Expression levels of IL-1β (C) and TNF-α (D) were examined. E, F: SARM exacerbated *V. anguillarum* infection. Flounder were administered with or without (control) SARM expression plasmid for 2.5 d, then injected with *V. anguillarum* for another 12 h. Bacterial loads in spleen (E) and liver (F) were determined through plate count. Ctrl: Control; hpi, hours post infection. *P*-values were calculated with Student's *t*-test. *: $P<0.05$; **: $P<0.01$. Error bars indicate SD.

SARM is a target gene of pol-miR-n199-3p

In our previous investigation on the micro-transcriptome of *V. anguillarum*-infected flounder, SARM was predicted to be a potential target of pol-miR-n199-3p (Ning & Sun, 2020b). To elucidate the mechanism underlying SARM modulating *V. anguillarum* infection, we examined the regulatory role of candidate miRNAs on SARM. The sequence features were first analyzed. As shown in Figure 2A, the sequence of SARM 3'UTR was completely complementary to the seed sequence of pol-miR-n199-3p. To confirm the specific binding site between pol-miR-n199-3p and SARM 3'UTR, a mutated form of SARM 3'UTR (SARM-mut) was constructed (Figure 2A). The dual-luciferase reporter assay was then conducted using HEK293T cells to verify the interaction between pol-miR-n199-3p and SARM 3'UTR. Results showed that pol-miR-n199-3p significantly reduced luciferase activity when co-transfected with the wild-type SARM 3'UTR but not with the mutated type (Figure 2B). Moreover, no significant decrease in luciferase activity was observed when pol-miR-n199-3p was replaced by miR-NC or wild-type SARM 3'UTR alone (Figure 2B). Finally, to determine the regulatory effect of pol-miR-n199-3p on SARM expression, a flounder model with pol-miR-n199-3p overexpression was established. Elevated pol-miR-n199-3p expression was confirmed by RT-qPCR (Supplementary Figure S2). SARM expression was significantly suppressed in fish with pol-miR-n199-3p overexpression, but not in the control fish (Figure 2C). Thus, these results indicate that pol-miR-n199-3p represses SARM expression via specific interactions with its 3'UTR.

Pol-miR-n199-3p enhances host antibacterial immunity by suppressing SARM

In our previous study on miRNA expression profiles in *V. anguillarum*-infected flounder, we found that pol-miR-n199-3p is significantly up-regulated in response to *V. anguillarum* infection at 6, 12, and 24 hpi (Ning & Sun, 2020b). In the present study, the expression levels of pol-miR-n199-3p and SARM were shown to be significantly negatively correlated

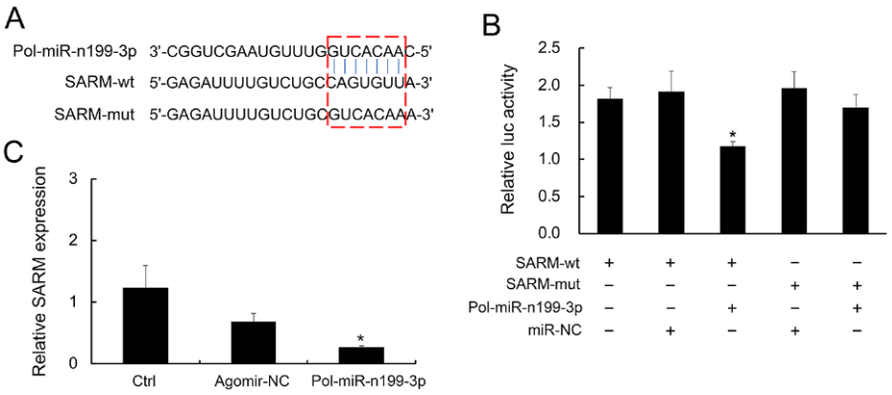


Figure 2 *SARM* is a target gene of pol-miR-n199-3p

A: SARM contained binding site for pol-miR-n199-3p. Sequence alignments between pol-miR-n199-3p seed sequence and its binding site in wild-type SARM 3'UTR (SARM-wt), as well as mutant-type with binding site mutation (SARM-mut), are shown. Red box indicates seed sequence in pol-miR-n199-3p, binding site in SARM-wt, and mutated sequence in SARM-mut. B: Pol-miR-n199-3p targeted 3'UTR of SARM. HEK293T cells were co-transfected with pol-miR-n199-3p mimic or miR-NC, along with SARM-wt or SARM-mut, or transfected with SARM-wt alone, with Luc activity then examined and normalized. NC, negative control. miR-NC, negative control of pol-miR-n199-3p mimic. Luc, luciferase. C: Pol-miR-n199-3p inhibited expression of SARM. Flounder were administered with or without (control, Ctrl) pol-miR-n199-3p agomir or administered with negative control of pol-miR-n199-3p agomir (agomir-NC) for 12 h. Expression levels of SARM were examined by RT-qPCR. *P*-values were calculated with Student's *t*-test. *: $P<0.05$. Error bars indicate SD.

($P<0.05$) (Figure 3A). Thus, we subsequently explored whether pol-miR-n199-3p had any effect on SARM expression. Results showed that pol-miR-n199-3p significantly suppressed SARM expression in flounder (Figure 3B). Given that SARM exacerbated bacterial infection by suppressing inflammatory cytokines, we wondered whether pol-miR-n199-3p could affect SARM-mediated IL-1β and TNF-α. As shown in Figure 3C, D, overexpression of pol-miR-n199-3p markedly increased the expression of IL-1β and TNF-α. To examine the direct effects of pol-miR-n199-3p on bacterial infection, bacterial counts were determined in fish with or without pol-miR-n199-3p overexpression upon *V. anguillarum* challenge. Our findings indicated that the up-regulation of pol-miR-n199-3p contributed to the inhibition of bacterial disseminations in both the spleen and liver (Figure 3E, F). These observations suggest that pol-miR-n199-3p promotes host defense against bacterial infection via increasing the expression of SARM-mediated inflammatory cytokines.

Pol-lnc78 regulates pol-miR-n199-3p expression via interaction

In our previous study on lncRNAs in flounder, we predicted that pol-lnc78 may target pol-miR-n199-3p (Ning & Sun, 2021a). To determine whether pol-lnc78 interacts with pol-miR-n199-3p, a dual-luciferase reporter assay was performed in HEK293T cells. The sequence of pol-lnc78 was first analyzed to identify the binding site for pol-miR-n199-3p (Figure 4A). Subsequently, both a wild-type pol-lnc78 reporter and mutated-type reporter with a mutated binding site were constructed. The dual-luciferase assay revealed that pol-miR-n199-3p strongly inhibited luciferase activity of the wild-type pol-lnc78 reporter but elicited no response in the mutated-type reporter (Figure 4B). To investigate the regulatory roles of pol-lnc78 on the target, a flounder model with pol-lnc78 overexpression was established using plasmid pCN3-lnc78 injection. After verification of increased pol-lnc78 expression at 3 d post-administration (Supplementary Figure S3), pol-miR-n199-3p expression was determined. As shown in Figure 4C, pol-lnc78 significantly reduced the expression of pol-miR-n199-3p *in vivo*. Together, these observations indicate that pol-lnc78 negatively regulates pol-miR-n199-3p expression through direct interaction.

Pol-lnc78 aggravates bacterial infection via suppression of pol-miR-n199-3p

In our previous research, we identified pol-lnc78 as a potential immune-related lncRNA engaged in *V. anguillarum* infection (Ning & Sun, 2021a). Here, validation of expression patterns indicated that pol-lnc78 was significantly down-regulated in the *V. anguillarum*-challenged flounder at 6, 12, and 24 h (Figure 5A). Correlation analysis showed that pol-lnc78 was significantly and negatively correlated ($P<0.05$) with pol-miR-n199-3p expression upon bacterial stimulation (Figure 5B). Moreover, *in vivo* overexpression of pol-lnc78 significantly decreased the expression of pol-miR-n199-3p in *V. anguillarum*-infected flounder (Figure 5C). Exploration of the potential regulatory effects of pol-lnc78 on inflammatory cytokines mediated by pol-miR-n199-3p during infection showed that overexpression of pol-lnc78 significantly repressed IL-1β and TNF-α expression in *V. anguillarum*-infected flounder (Figure 5D, E). An *in vivo* infection study was conducted using flounder with or without pol-lnc78 overexpression. Results indicated significantly increased bacterial loads in both the spleen and liver of fish overexpressing pol-lnc78 (Figure 5F, G). Collectively, these findings suggest that pol-lnc78 promotes bacterial dissemination by inhibiting pol-miR-n199-3p.

Pol-lnc78, as a ceRNA, sponges pol-miR-n199-3p to increase SARM expression

In our previous study on lncRNA-miRNA-mRNA networks in *V. anguillarum*-infected flounder, we predicted that the pol-lnc78-pol-miR-n199-3p-SARM axis may be involved in immune-related ceRNA networks (Ning & Sun, 2021a). In the current study, we validated the interaction between pol-lnc78 and pol-miR-n199-3p and the targeting of SARM by pol-miR-n199-3p. Subsequently, we explored whether pol-lnc78 could regulate SARM expression. *In vivo* analysis showed that overexpression of pol-lnc78 significantly up-regulated SARM expression (Figure 6A). Moreover, correlation analysis showed that the expression levels of pol-lnc78 and SARM

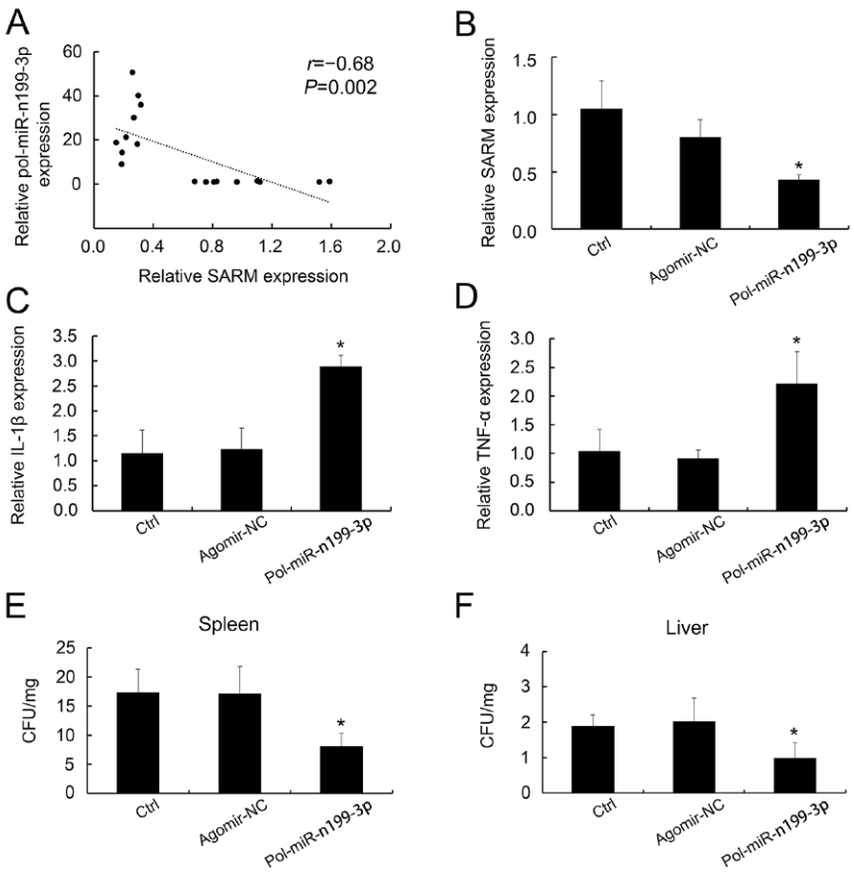


Figure 3 Pol-miR-n199-3p enhances host antibacterial defense by suppressing SARM expression

A: SARM and pol-miR-n199-3p were significantly and negatively correlated. Expression levels of SARM and pol-miR-n199-3p in flounder challenged with or without *Vibrio anguillarum* for 6, 12, and 24 h were tested, then subjected to correlation analysis. *r*, correlation coefficient. B: Pol-miR-n199-3p suppressed SARM expression upon *V. anguillarum* stimulation. Expression levels of SARM in flounder administered with pol-miR-n199-3p agomir or negative control (agomir-NC), along with *V. anguillarum* infection for 12 h, were detected. C, D: Pol-miR-n199-3p promoted expression of inflammatory cytokines that were reduced by SARM. Flounder were administered with pol-miR-n199-3p agomir or agomir-NC, along with *V. anguillarum* infection for 12 h. Expression levels of IL-1β (C) and TNF-α (D) were then examined. E, F: Pol-miR-n199-3p inhibited *V. anguillarum* infection. Flounder were administered with pol-miR-n199-3p agomir or agomir-NC, along with *V. anguillarum* infection for 12 h. Bacterial loads in spleen (E) and liver (F) were then determined by plate count. Ctrl: Control. *P*-values were calculated with Student's *t*-test. *: $P<0.05$. Error bars indicate SD.

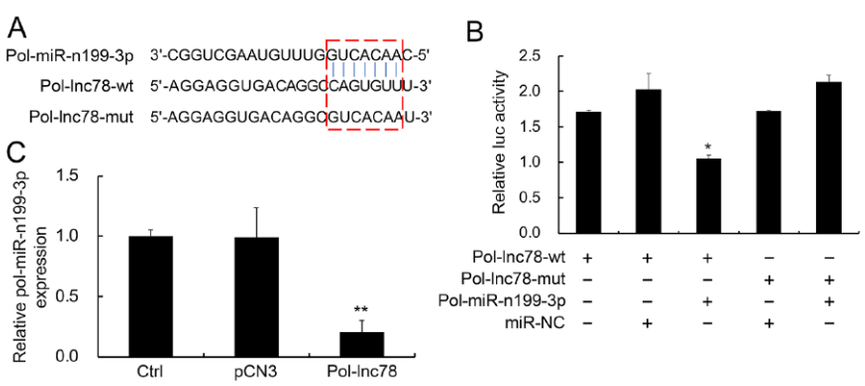


Figure 4 Pol-lnc78 regulates pol-miR-n199-3p expression by interactions

A: Pol-lnc78 contained the binding site for pol-miR-n199-3p. Sequence alignments between pol-miR-n199-3p seed sequence and its binding site in wild-type of pol-lnc78 (pol-lnc78-wt), as well as mutant-type with binding site mutated (pol-lnc78-mut), are shown. Red box indicates seed sequence in pol-miR-n199-3p, binding site in pol-lnc78-wt, and mutated sequence in pol-lnc78-mut. B: Pol-lnc78 interacted with pol-miR-n199-3p. HEK293T cells were co-transfected with pol-miR-n199-3p mimic or miR-NC, along with pol-lnc78-wt or pol-lnc78-mut, or transfected with pol-lnc78-wt alone, with Luc activity then examined and normalized. miR-NC, negative control of pol-miR-n199-3p mimic. Luc, luciferase. C: Pol-lnc78 inhibited pol-miR-n199-3p expression. Flounder were administered with or without pol-lnc78 expression plasmid or administered with pCN3 vector for 3 d. Expression of pol-miR-n199-3p was determined by RT-qPCR. Ctrl: Control. *P*-values were calculated with Student's *t*-test. *: $P<0.05$; **: $P<0.01$. Error bars indicate SD.

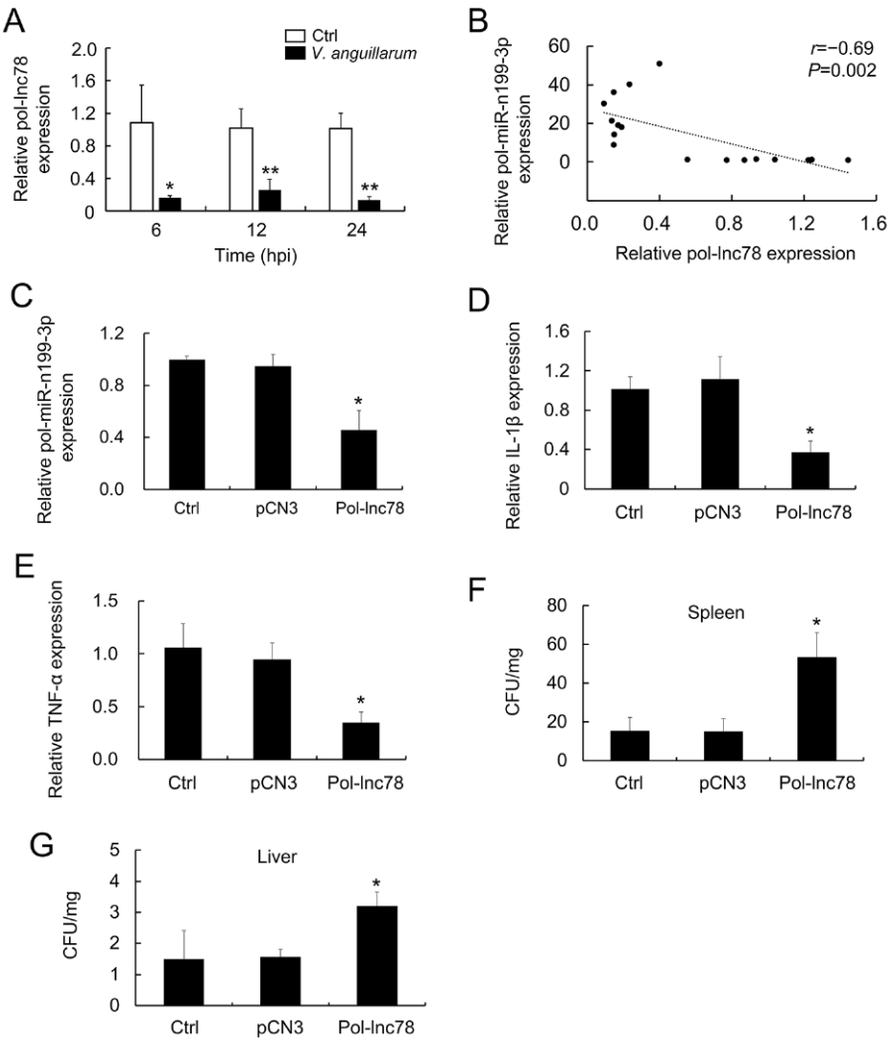


Figure 5 Pol-lnc78 suppresses anti-*Vibrio anguillarum* response by inhibiting pol-miR-n199-3p

A: Pol-lnc78 expression declined in response to *V. anguillarum* infection. Expression levels of pol-lnc78 in flounder infected with or without (control) *V. anguillarum* for 6, 12, and 24 h were examined. hpi: Hours post infection. B: Pol-lnc78 and pol-miR-n199-3p were significantly and negatively correlated. Expression levels of pol-lnc78 and pol-miR-n199-3p in flounder stimulated with or without *V. anguillarum* for 6, 12, and 24 h were tested, then subjected to correlation analysis. *r*, correlation coefficient. C–E: Pol-lnc78 suppressed expression of pol-miR-n199-3p and its mediated inflammatory cytokines upon *V. anguillarum* stimulation. Flounder were administered with or without pol-lnc78 expression plasmid or administered with pCN3 vector for 2.5 d, then injected with *V. anguillarum* for another 12 h. Expression levels of pol-miR-n199-3p (C), IL-1β (D), and TNF-α (E) were examined. F, G: Pol-lnc78 promoted *V. anguillarum* dissemination. Flounder were administered with or without pol-lnc78 expression plasmid or administered with pCN3 vector for 2.5 d, then injected with *V. anguillarum* for another 12 h. Bacterial loads in spleen (F) and liver (G) were determined by plate count. Ctrl: Control. *P*-values were calculated with Student's *t*-test. *: $P<0.05$; **: $P<0.01$. Error bars indicate SD.

were significantly and positively correlated during bacterial infection (Figure 6B), suggesting that pol-lnc78 regulates SARM expression. Thus, we next tested whether the regulatory effects of pol-lnc78 on SARM could be mediated by pol-miR-n199-3p. A dual-luciferase reporter assay was performed using HEK293T cells co-transfected with SARM 3' UTR reporter, pol-miR-n199-3p mimic, and/or pol-lnc78 expression plasmid. Results showed that pol-lnc78 counteracted the inhibitory effects of pol-miR-n199-3p on SARM 3'UTR (Figure 6C). Furthermore, comparing the expression of SARM in flounder with pol-miR-n199-3p overexpression, and with pol-miR-n199-3p and pol-lnc78 co-overexpression, we found that pol-lnc78 relieved the inhibitory effects of pol-miR-n199-3p on SARM expression (Figure 6D). Together, these observations suggest that pol-lnc78 can act as a ceRNA and sponge pol-miR-n199-3p to regulate the expression of SARM as well as SARM-mediated inflammatory cytokines, thereby modulating host antibacterial immunity (Figure 7).

DISCUSSION

Bacterial diseases greatly impact the fish farming industry, with vibriosis, one of the most prevalent aquaculture diseases globally, causing substantial economic losses (Egidius, 1987). Among the causative agents of vibriosis, *V. anguillarum* is noteworthy due to its pathogenicity and ubiquitous existence in the water environment, leading to widespread infections in various aquatic animals, including Atlantic cod (Lokesh et al., 2012), orange-spotted grouper (Huang et al., 2014), tongue sole (Zhang et al., 2015), turbot (Gao et al., 2016), miiuy croaker (Chu et al., 2017), and flounder (Gao et al., 2013). The innate immune system acts as the first and central line of defense in teleost fish, with the TLR signaling pathway playing

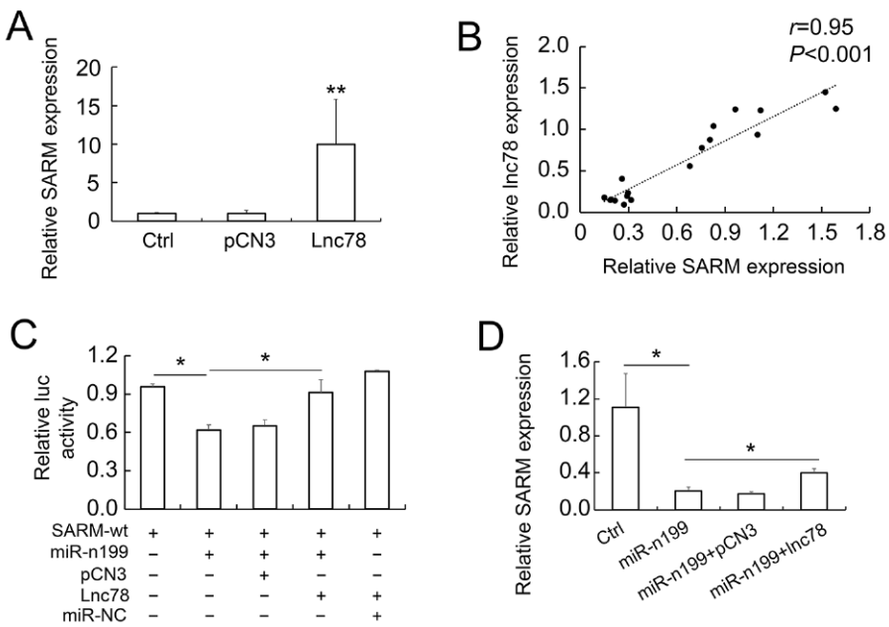


Figure 6 Pol-lnc78 as a ceRNA sponges pol-miR-n199-3p to enhance SARM expression

A: Pol-lnc78 up-regulates SARM expression. Flounder were administered with or without pol-lnc78 expression plasmid or administered with pCN3 vector for 3 d. Expression levels of SARM were determined. B: SARM and pol-lnc78 were significantly and positively correlated. Expression levels of SARM and pol-lnc78 in flounder infected with or without *Vibrio anguillarum* for 6, 12, and 24 h were examined, then subjected to correlation analysis. *r*, correlation coefficient. C: Pol-lnc78 counteracted inhibition of pol-miR-n199-3p on SARM 3'UTR. HEK293T cells were co-transfected with pol-miR-n199-3p mimic or miR-NC, together with pCN3 vector or pol-lnc78 expression plasmid, along with SARM 3'UTR, or transfected with SARM 3'UTR alone. Luc activity was examined and normalized. miR-NC, negative control of pol-miR-n199-3p mimic. Luc, luciferase. D: Pol-lnc78 counteracted inhibitory effect of pol-miR-n199-3p on SARM expression. Flounder were administered with or without pol-lnc78 expression plasmid for 2.5 d, then administered with pol-miR-n199-3p agomir for another 12 h. Expression levels of SARM were determined. Ctrl: Control. *P*-values were calculated with Student's *t*-test. *: $P<0.05$; **: $P<0.01$. Error bars indicate SD.

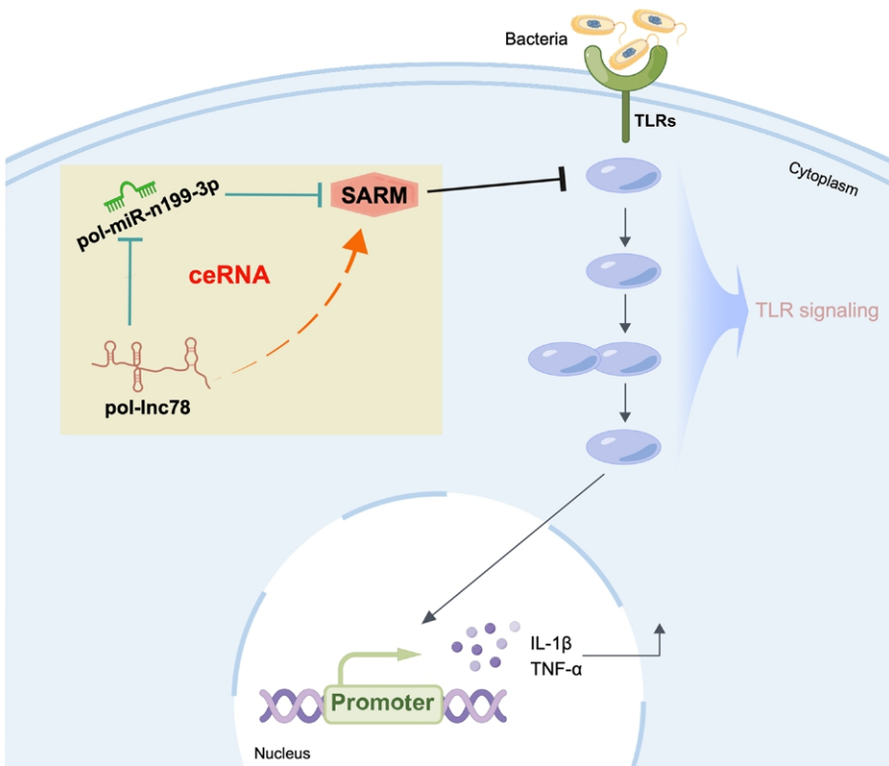


Figure 7 Mechanistic diagram of ceRNA network and antibacterial regulation of pol-lnc78, pol-miR-n199-3p, and SARM

Fish SARM negatively regulates immune response by inhibiting inflammatory cytokines IL-1β and TNF-α. Pol-miR-n199-3p targets SARM and promotes antibacterial defense by up-regulating SARM-dependent inflammatory cytokine expression. Pol-lnc78 acts as a ceRNA sponging pol-miR-n199-3p to enhance SARM, thus ensuring appropriate inflammatory and antibacterial responses for host survival.

a crucial role in antibacterial responses. Among the five TIR adaptors, SARM is uniquely inhibitory and highly conserved (O'Neill & Bowie, 2007). In the present study, we observed a significant decline in the expression of flounder SARM during *V. anguillarum* infection. The inhibitory role of SARM on the TLR pathway also appears to be conserved across mammals. For example, in human HEK293T cells, overexpression of SARM significantly represses LPS-mediated up-regulation of inflammatory cytokines, such as IL-1β and TNF-α (Carlsson et al., 2016). *Burkholderia pseudomallei* infection can increase SARM expression in mice, resulting in the suppression of TRIF and interferon β (IFNβ) production (Pudla et al., 2011). Similarly, we found that flounder SARM significantly reduced the expression of inflammatory cytokines IL-1β and TNF-α upon *V. anguillarum* stimulation. Previous studies have shown divergence in the function of SARM between invertebrates and mammals. In worms and flies, SARM orthologs contribute to host resistance against microbial pathogens (Liberati et al., 2004; Monsanto-Hearne et al., 2017), while $SARM^{-/-}$ mice exhibit higher survival from LPS-stimulated sepsis due to increased IL-1β release (Carty et al., 2019). Here, we found that overexpression of SARM in lower vertebrates such as flounder significantly exacerbated bacterial dissemination, indicating the functional resemblance of fish SARM to that of mammals.

The molecular mechanism by which SARM regulates immune function remains enigmatic, even in mammals. Studies on invertebrate model animals have suggested a potential miRNA-mediated modulation of SARM (Finger et al., 2019; Monsanto-Hearne et al., 2017). In teleost fish, miRNAs are recognized as ubiquitous and versatile regulators, critical in immune responses. For instance, in the orange-spotted grouper, miR-122 modulates the immune response to *Aeromonas hydrophila* infection by targeting IL-15 (Liu et al., 2020); in miiuy croaker, miR-144 and miR-217 act as negative modulators of the inflammatory response against *Vibrio harveyi* infection (Chu et al., 2021a). Recent studies have also suggested that miRNAs are involved in the flounder response to *V. anguillarum* challenge (Ning & Sun, 2020b). In the present study, pol-miR-n199-3p was shown to target the 3' UTR of SARM, leading to a reduction in expression. These findings are consistent with the established mechanism of miRNA modulation, whereby miRNAs interact with the 3'UTR of the target gene, thus facilitating its degradation or suppressing its translation at the post-transcriptional level (Cannell et al., 2008). Accumulating evidence suggests that fish miRNAs can affect pathogens. For example, flounder pol-miR-novel_171 promotes gram-negative bacterial dissemination (Li et al., 2020), pol-miR-150 negatively regulates gram-positive bacterial infection (Sun et al., 2021), and miR-206 increases *Mycobacterium marinum* burden in zebrafish (Wright et al., 2021). Here, we found that pol-miR-n199-3p played a protective role against pathogen infection in flounder. Of note, the inhibition of *V. anguillarum* dissemination by pol-miR-n199-3p was attributable to the increase in SARM-mediated inflammatory cytokines. Moreover, we found that SARM was linked to the regulation of lncRNA.

Despite continued research on lncRNAs since the first lncRNA (Xist) was discovered in humans (Borsani et al., 1991), identification of lncRNAs in non-model species, especially fish, has only gained momentum in recent years. In our previous study, we systematically identified flounder lncRNAs through bioinformatics analysis, highlighting pol-lnc78 as an immune-related lncRNA (Ning & Sun, 2021a). In the current study, functional experiments confirmed the importance of pol-lnc78 in the host immune response to bacterial challenge, with pol-lnc78 shown to serve as a negative regulator. Recently, lncRNA-mediated ceRNA crosstalk has become a hotspot of research. In mammals, ceRNAs have been strongly implicated in the pathogenesis of cancers, providing crucial diagnostic biomarkers and therapeutic targets (Abdollahzadeh et al., 2019; Chen et al., 2017; Wang et al., 2017a). In lower vertebrates, limited research has been conducted on lncRNA-mediated ceRNA mechanisms, except for a few studies associated with miiuy croaker immunity. Specifically, in miiuy croaker, lncRNA IRL promotes the immune response against bacterial infection by counteracting the suppressive effects of miR-27c-3p on IRAK4 (Zheng et al., 2021a), while lncRNA AANCR increases MITA expression through competitively binding to miR-210 and regulates host antiviral responses (Chu et al., 2021b). Given this context, determining whether ceRNA crosstalk occurs widely among teleost fish is important. Here, we revealed that flounder pol-lnc78 acts as a ceRNA to increase SARM expression through sponging pol-miR-n199-3p, with this ceRNA network regulating host defense responses to bacterial infection. These results highlight the essential effects of lncRNAs in host immunity and lend further support to the ubiquity of ceRNA regulation in fish.

CONCLUSIONS

Our study identified SARM, pol-miR-n199-3p, and pol-lnc78 as key regulators in the response of flounder to *V. anguillarum* infection through modulation of inflammatory processes. Pol-miR-n199-3p repressed SARM expression by targeting its 3' UTR. Concurrently, pol-lnc78 acted as a ceRNA, counteracting the inhibitory effects of pol-miR-n199-3p on SARM. These observations provide new insights into the mechanisms of SARM, as well as the roles of ceRNA in lower vertebrates.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Conceptualization, X.H.N. and S.W.Y.; Funding acquisition, X.H.N. and S.W.Y.; Methodology, X.H.N.; Investigation, B.H., Y.P., and X.H.N.; Resources, X.H.N.; Original draft, X.H.N.; Review & editing, X.H.N. and S.W.Y. All authors read and approved the final version of the manuscript.

REFERENCES

Abdollahzadeh R, Daraei A, Mansoori Y, et al. 2019. Competing endogenous RNA (ceRNA) cross talk and language in ceRNA regulatory networks: a new look at hallmarks of breast cancer. *Journal of Cellular Physiology*, **234**(7): 10080−10100.

Andreassen R, Høyheim B. 2017. miRNAs associated with immune response in teleost fish. *Developmental & Comparative Immunology*, **75**: 77−85.

Bagga S, Bracht J, Hunter S, et al. 2005. Regulation by *let*-7 and *lin*-4 miRNAs results in target mRNA degradation. *Cell*, **122**(4): 553−563.

Borsani G, Tonlorenzi R, Simmler MC, et al. 1991. Characterization of a murine gene expressed from the inactive X chromosome. *Nature*,

351(6324): 325−329.
Cannell IG, Kong YW, Bushell M. 2008. How do microRNAs regulate gene expression. *Biochemical Society Transactions*, **36**(6): 1224−1231.
Carlsson E, Ding JL, Byrne B. 2016. SARM modulates MyD88-mediated TLR activation through BB-loop dependent TIR-TIR interactions. *Biochimica et Biophysica Acta* (*BBA*) - *Molecular Cell Research*, **1863**(2): 244–253.
Carty M, Bowie AG. 2019. SARM: from immune regulator to cell executioner. *Biochemical Pharmacology*, **161**: 52−62.
Carty M, Kearney J, Shanahan KA, et al. 2019. Cell survival and cytokine release after inflammasome activation is regulated by the Toll-IL-1R protein SARM. *Immunity*, **50**(6): 1412−1424.e6.
Cesana M, Cacchiarelli D, Legnini I, et al. 2011. A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell*, **147**(2): 358−369.
Chen DL, Lu YX, Zhang JX, et al. 2017. Long non-coding RNA UICLM promotes colorectal cancer liver metastasis by acting as a ceRNA for microRNA-215 to regulate ZEB2 expression. *Theranostics*, **7**(19): 4836−4849.
Cheng AM, Byrom MW, Shelton J, et al. 2005. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Research*, **33**(4): 1290−1297.
Chu Q, Bi DK, Zheng WW, et al. 2021a. MicroRNA negatively regulates NF-κB-mediated immune responses by targeting NOD1 in the teleost fish *Miichthys miiuy*. *Science China Life Sciences*, **64**(5): 803–815.
Chu Q, Song WH, Cui JX, et al. 2017. Genome-guided transcriptome analysis of miiuy croaker provides insights into pattern recognition receptors and cytokines in response to *Vibrio anguillarum*. *Developmental & Comparative Immunology*, **73**: 72–78.
Chu Q, Xu TJ, Zheng WW, et al. 2020. Long noncoding RNA MARL regulates antiviral responses through suppression miR-122-dependent MAVS downregulation in lower vertebrates. *PLoS Pathogens*, **16**(7): e1008670.
Chu Q, Xu TJ, Zheng WW, et al. 2021b. Long noncoding RNA AANCR modulates innate antiviral responses by blocking miR-210-dependent MITA downregulation in teleost fish. *Miichthys miiuy*. *Science China Life Sciences*, **64**(7): 1131−1148.
Cong ZZ, Diao YF, Xu Y, et al. 2019. Long non-coding RNA linc00665 promotes lung adenocarcinoma progression and functions as ceRNA to regulate AKR1B10-ERK signaling by sponging miR-98. *Cell Death & Disease*, **10**(2): 84.
Couillault C, Pujol N, Reboul J, et al. 2004. TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. *Nature Immunology*, **5**(5): 488−494.
Du M, Yuan L, Tan X, et al. 2017. The LPS-inducible lncRNA Mirt2 is a negative regulator of inflammation. *Nature Communications*, **8**(1): 2049.
Egidius E. 1987. Vibriosis: pathogenicity and pathology. A review. *Aquaculture*, **67**(1-2): 15−28.
Finger F, Ottens F, Springhorn A, et al. 2019. Olfaction regulates organismal proteostasis and longevity via microRNA-dependent signalling. *Nature Metabolism*, **1**(3): 350−359.
Gao CB, Fu Q, Su BF, et al. 2016. Transcriptomic profiling revealed the signatures of intestinal barrier alteration and pathogen entry in turbot (*Scophthalmus maximus*) following *Vibrio anguillarum* challenge. *Developmental & Comparative Immunology*, **65**: 159−168.
Gao H, Wu L, Sun JS, et al. 2013. Molecular characterization and expression analysis of Toll-like receptor 21 cDNA from *Paralichthys olivaceus*. *Fish & Shellfish Immunology*, **35**(4): 1138–1145.
Hatfield SD, Shcherbata HR, Fischer KA, et al. 2005. Stem cell division is regulated by the microRNA pathway. *Nature*, **435**(7044): 974−978.
Huang JB, Wu YC, Chi SC. 2014. Dietary supplementation of *Pediococcus pentosaceus* enhances innate immunity, physiological health and resistance to *Vibrio anguillarum* in orange-spotted grouper (*Epinephelus coioides*). *Fish & Shellfish Immunology*, **39**(2): 196−205.
Jault C, Pichon L, Chluba J. 2004. Toll-like receptor gene family and TIR-domain adapters in *Danio rerio*. *Molecular Immunology*, **40**(11): 759–771.
Jiao XD, Zhang M, Hu YH, et al. 2009. Construction and evaluation of DNA vaccines encoding *Edwardsiella tarda* antigens. *Vaccine*, **27**(38): 5195−5202.
Kapranov P, Cheng J, Dike S, et al. 2007. RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science*, **316**(5830): 1484−1488.
Li WR, Guan XL, Jiang S, et al. 2020. The novel fish miRNA pol-miR-novel_171 and its target gene FAM49B play a critical role in apoptosis and bacterial infection. *Developmental & Comparative Immunology*, **106**: 103616.
Liberati NT, Fitzgerald KA, Kim DH, et al. 2004. Requirement for a conserved Toll/interleukin-1 resistance domain protein in the *Caenorhabditis elegans* immune response. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(17): 6593−6598.
Lindsay SA, Wasserman SA. 2014. Conventional and non-conventional *Drosophila* Toll signaling. *Developmental & Comparative Immunology*, **42**(1): 16−24.
Liu X, Hao Y, Peng LP, et al. 2020. MiR-122 is involved in immune response by regulating Interleukin-15 in the orange-spotted grouper (*Epinephelus coioides*). *Fish & Shellfish Immunology*, **106**: 404−409.
Lokesh J, Fernandes JMO, Korsnes K, et al. 2012. Transcriptional regulation of cytokines in the intestine of Atlantic cod fed yeast derived mannan oligosaccharide or β-Glucan and challenged with *Vibrio anguillarum*. *Fish & Shellfish Immunology*, **33**(3): 626–631.
Malapati H, Millen SM, W JB. 2017. The axon degeneration gene SARM1 is evolutionarily distinct from other TIR domain-containing proteins. *Molecular Genetics and Genomics*, **292**(4): 909−922.
Monsanto-Hearne V, Tham ALY, Wong ZS, et al. 2017. *Drosophila* miR-956 suppression modulates Ectoderm-expressed 4 and inhibits viral replication. *Virology*, **502**: 20−27.
Ning XH, Sun L. 2020a. Gene network analysis reveals a core set of genes involved in the immune response of Japanese flounder (*Paralichthys olivaceus*) against *Vibrio anguillarum* infection. *Fish & Shellfish Immunoloy*, **98**: 800−809.
Ning XH, Sun L. 2020b. Micro-transcriptome analysis reveals immune-related MicroRNA regulatory networks of *Paralichthys olivaceus* induced by *Vibrio anguillarum* infection. *International Journal of Molecular Sciences*, **21**(12): 4252.
Ning XH, Sun L. 2021a. Identification and characterization of immune-related lncRNAs and lncRNA-miRNA-mRNA networks of *Paralichthys olivaceus* involved in *Vibrio anguillarum* infection. *BMC Genomics*, **22**(1): 447.
Ning XH, Sun L. 2021b. Systematic identification and analysis of circular RNAs of Japanese flounder (*Paralichthys olivaceus*) in response to *Vibrio anguillarum* infection. *Genes* (*Basel*), **12**(1): 100.
O'Neill LAJ, Bowie AG. 2007. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nature Reviews Immunology*, **7**(5): 353−364.
Panneerselvam P, Ding JL. 2015. Beyond TLR signaling-the role of SARM in antiviral immune defense, apoptosis & development. *International Reviews of Immunology*, **34**(5): 432−444.
Ponting CP, Oliver PL, Reik W. 2009. Evolution and functions of long noncoding RNAs. *Cell*, **136**(4): 629−641.
Pudla M, Limposuwan K, Utaisincharoen P. 2011. *Burkholderia pseudomallei*-induced expression of a negative regulator, sterile-α and Armadillo motif-containing protein, in mouse macrophages: a possible

mechanism for suppression of the MyD88-independent pathway. *Infection and Immunity*, **79**(7): 2921−2927.

Qi LJ, Chen YD, Shi KP, et al. 2021. Combining of transcriptomic and proteomic data to mine immune-related genes and proteins in the liver of *Cynoglossus semilaevis* challenged with *Vibrio anguillarum*. *Comparative Biochemistry and Physiology Part D*: *Genomics and Proteomics*, **39**: 100864.

Sarangdhar MA, Chaubey D, Srikakulam N, et al. 2018. Parentally inherited long non-coding RNA Cyrano is involved in zebrafish neurodevelopment. *Nucleic Acids Research*, **46**(18): 9726−9735.

Sarma NJ, Tiriveedhi V, Crippin JS, et al. 2014. Hepatitis C virus-induced changes in microRNA 107 (miRNA-107) and miRNA-449a modulate CCL2 by targeting the interleukin-6 receptor complex in hepatitis. *Journal of Virology*, **88**(7): 3733−3743.

Song YX, Sun JX, Zhao JH, et al. 2017. Non-coding RNAs participate in the regulatory network of CLDN4 via ceRNA mediated miRNA evasion. *Nature Communications*, **8**(1): 289.

Sun YL, Li XP, Sun L. 2021. Pol-miR-150 regulates anti-bacterial and viral infection in Japanese flounder (*Paralichthys olivaceus*) *via* the lysosomal protein LMP2L. *Comparative Biochemistry and Physiology Part B:Biochemistry and Molecular Biology*, **254**: 110578.

Vasadia DJ, Zippay ML, Place SP. 2019. Characterization of thermally sensitive miRNAs reveals a central role of the FoxO signaling pathway in regulating the cellular stress response of an extreme stenotherm. *Trematomus bernacchii. Marine Genomics*, **48**: 100698.

Vériène J, Fossouo L, Parker JA. 2015. Neurodegeneration in *C. elegans* models of ALS requires TIR-1/Sarm1 immune pathway activation in neurons. *Nature Communications*, **6**: 7319.

Wang H, Huo XS, Yang XR, et al. 2017a. STAT3-mediated upregulation of lncRNA HOXD-AS1 as a ceRNA facilitates liver cancer metastasis by regulating SOX4. *Molecular Cancer*, **16**(1): 136.

Wang KL, Chen SN, Li L, et al. 2021a. Functional characterization of four TIR domain-containing adaptors, MyD88, TRIF, MAL, and SARM in mandarin fish *Siniperca chuatsi*. *Developmental & Comparative Immunology*, **122**: 104110.

Wang P, Xu JF, Wang YJ, et al. 2017b. An interferon-independent lncRNA promotes viral replication by modulating cellular metabolism. *Science*, **358**(6366): 1051−1055.

Wang XA, Ma AJ, Sun ZB. 2021b. Genetic parameters of seven immune factors in turbot (*Scophthalmus maximus*) infected with *Vibrio anguillarum*. *Journal of Fish Diseases*, **44**(3): 263–271.

Wang Y, Xu ZY, Jiang JF, et al. 2013. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Developmental Cell*, **25**(1): 69−80.

Wienholds E, Plasterk RHA. 2005. MicroRNA function in animal development. *FEBS Letters*, **579**(26): 5911−5922.

Wright K, de Silva K, Plain KM, et al. 2021. Mycobacterial infection-induced miR-206 inhibits protective neutrophil recruitment via the CXCL12/CXCR4 signalling axis. *PLoS Pathogens*, **17**(4): e1009186.

Wu Q, Ning XH, Sun L. 2021. Megalocytivirus induces complicated fish immune response at multiple rna levels involving mRNA, miRNA, and circRNA. *International Journal of Molecular Sciences*, **22**(6): 3156.

Wu SJ, Huang JQ, Li YJ, et al. 2022. Integrated analysis of lncRNA and circRNA mediated ceRNA regulatory networks in skin reveals innate immunity differences between wild-type and yellow mutant rainbow trout (*Oncorhynchus mykiss*). *Frontiers in Immunology*, **13**: 802731.

Wu XS, Wang F, Li HF, et al. 2017. LncRNA-PAGBC acts as a microRNA sponge and promotes gallbladder tumorigenesis. *EMBO Reports*, **18**(10): 1837−1853.

Xu HS, Xing J, Tang XQ, et al. 2019. Immune response and protective effect against *Vibrio anguillarum* induced by DNA vaccine encoding Hsp33 protein. *Microbial Pathogenesis*, **137**: 103729.

Yan NN, Su JG, Yang CR, et al. 2015. Grass carp SARM1 and its two splice variants negatively regulate IFN-I response and promote cell death upon GCRV infection at different subcellular locations. *Developmental & Comparative Immunology*, **48**(1): 102−115.

Ye WD, Duan Y, Zhang WT, et al. 2021. Comprehensive analysis of hub mRNA, lncRNA and miRNA, and associated ceRNA networks implicated in grass carp (*Ctenopharyngodon idella*) growth traits. *Genomics*, **113**(6): 4004−4014.

Zhang P, Yu C, Sun L. 2020. Japanese flounder (*Paralichthys olivaceus*) Bmal1 is involved in the regulation of inflammatory response and bacterial infection. *Aquaculture*, **525**: 735330.

Zhang Q, Zmasek CM, Cai XH, et al. 2011. TIR domain-containing adaptor SARM is a late addition to the ongoing microbe-host dialog. *Developmental & Comparative Immunology*, **35**(4): 461−468.

Zhang X, Wang SL, Chen SL, et al. 2015. Transcriptome analysis revealed changes of multiple genes involved in immunity in *Cynoglossus semilaevis* during *Vibrio anguillarum* infection. *Fish & Shellfish Immunology*, **43**(1): 209−218.

Zheng WJ, Hu YH, Xiao ZZ, et al. 2010. Cloning and analysis of a ferritin subunit from turbot (*Scophthalmus maximus*). *Fish & Shellfish Immunology*, **28**(5-6): 829−836.

Zheng WW, Chu Q, Xu TJ. 2021a. Long noncoding RNA IRL regulates NF-κB-mediated immune responses through suppression of miR-27c-3p-dependent IRAK4 downregulation in teleost fish. *Journal of Biological Chemistry*, **296**: 100304.

Zheng WW, Chu Q, Xu TJ. 2021b. The long noncoding RNA NARL regulates immune responses via microRNA-mediated NOD1 downregulation in teleost fish. *Journal of Biological Chemistry*, **296**: 100414.

Zhou X, Jiang TF, Du XC, et al. 2013. Molecular characterization of porcine *SARM1* and its role in regulating TLRs signaling during highly pathogenic porcine reproductive and respiratory syndrome virus infection *in vivo*. *Developmental & Comparative Immunology*, **39**(1–2): 117–126.

Fan et al. *Zool. Res.* 2024, 45(1): 36−38
https://doi.org/10.24272/j.issn.2095-8137.2023.038

ZR | Zoological Research

Letter to the editor Open Access

Population size estimates based on GPS telemetry

DEAR EDITOR,

Accurate estimation of population size is a crucial issue in wildlife population ecology and conservation. While Global Positioning System (GPS) collars are well recognized as an effective method for monitoring wildlife behavior, their application in direct wildlife population size estimation remains underutilized. In this study, GPS telemetry was employed to survey a rhesus macaque (*Macaca mulatta*) population on Neilingding Island in Guangdong, China. From May 2021 to April 2022, 32 macaques across nine groups were fitted with GPS collars, producing a dataset of 109 739 location fixes. Analyzing data from all tracked individuals revealed the annual home ranges of the nine groups, which varied from 13.9 ha to 55.4 ha. Substantial home range overlap among groups was observed, with the highest level of exclusivity occurring in March. Based on the data, the study area was projected to support approximately 38.6 macaque groups within their respective home ranges, with a total population size of 957±193 macaques, derived from a mean group size of 24.8±5.0. Further analysis highlighted a consistent spatial utilization pattern between individual macaques and their corresponding groups, suggesting that data from a single individual may represent the home range and utilization patterns of the entire group. The presented methodology is time-effective, operates independently of surveyor expertise, demonstrates high repeatability, and yields extensive data suitable for further ecological studies.

Population monitoring is a core aspect of wildlife population ecology and conservation. Such monitoring provides crucial information regarding population status and its response to threats and conservation interventions, particularly in the context of rapid biodiversity decline driven by unsustainable anthropogenic activities. Traditionally, population monitoring has relied on the assessment of direct counts or indirect traces, such as distance and mark-recapture sampling approaches (Li, 2021). These conventional methods, while reliable, are labor- and time-intensive and necessitate substantial surveyor expertise. Technological advances, combined with field studies, have provided alternative tools for surveying wildlife populations, including infrared cameras and passive acoustic sensors (Piel et al., 2022). Notably, GPS tracking is a promising technology for population monitoring as it overcomes many sampling limitations and can automatically generate long-term and scientifically robust data (Nathan et al., 2022). While such technology has been used to estimate the population size of intra-sexually territorial species such as fishers (*Martes pennanti*) (Koen et al., 2007), its high cost and installation difficulties have limited its broader application in population censuses.

In the current study, we explored the feasibility of using GPS tracking technology to estimate the size of a rhesus macaque (*Macaca mulatta*) population located on Neilingding Island (N22°23′49″−22°25′35″, E113°46′18″−113°49′49″) in Guangdong, China (Figure 1A), a small enclosed island (total land area of 4.98 km^2) with minimal variation in elevation (highest elevation of 340.9 m a.s.l.). Historical surveys have regularly monitored this macaque population, with the most recent 2016 survey estimating a population size of approximately 1 000 individuals (Chu et al., 2019). These historical data provide a robust foundation to evaluate the viability of GPS tracking. During the study period, macaques were equipped with GPS collars to track their trajectories, ascertain their home ranges, and estimate population size by determining the number of groups that the study area could accommodate based on the specified home ranges. From May 2021 to April 2022, 32 individuals from nine groups were tracked, yielding a dataset of 109 739 location fixes. To ensure the accuracy of locations, we adopt a data screening method to filter fixes with locating deviation. After data screening, 26 667 fixes (24.3%) were excluded. Detailed information regarding GPS collar implementation, data validation and screening processes, group size assessment, and tracked individuals can be found in the Supplementary Materials and Methods.

Annual and monthly home range sizes were derived using kernel density estimates. Results indicated that macaques on Neilingding Island maintained constrained home ranges (mean: 31.2±11.8 ha, range: 13.9–55.4 ha) with a high degree of overlap (Figure 1B). The extent of monthly home ranges fluctuated (Figure 1C), with a maximum in August (coefficient of variation (CV) 0.40, mean: 35.26 ha, range: 14.2–57.3 ha) and a minimum in April (CV=0.30, mean: 18.27 ha, range: 11.5–26.1 ha). For each group, the exclusive home range (home range not overlapping with any other tracked group) and exclusive ratio (proportion of the exclusive area to total home range size) were calculated. As the exclusive home range may be overestimated due to potential undetected overlaps with neighboring groups not subject to tracking, and a lower exclusive ratio indicates a higher extent of this overlap, monthly home range with the highest exclusive ratio

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 10 May 2023; Accepted: 08 September 2023; Online: 09 September 2023

Foundation items: This study was supported by the Forestry Administration of Guangdong Province (33000-42090019), Shenzhen Science and Technology Innovation Committee (Basic Research 20180285), Nature Science Foundation of Guangdong, China (2021A 1515010641), and Neilingding-Futian National Nature Reserve

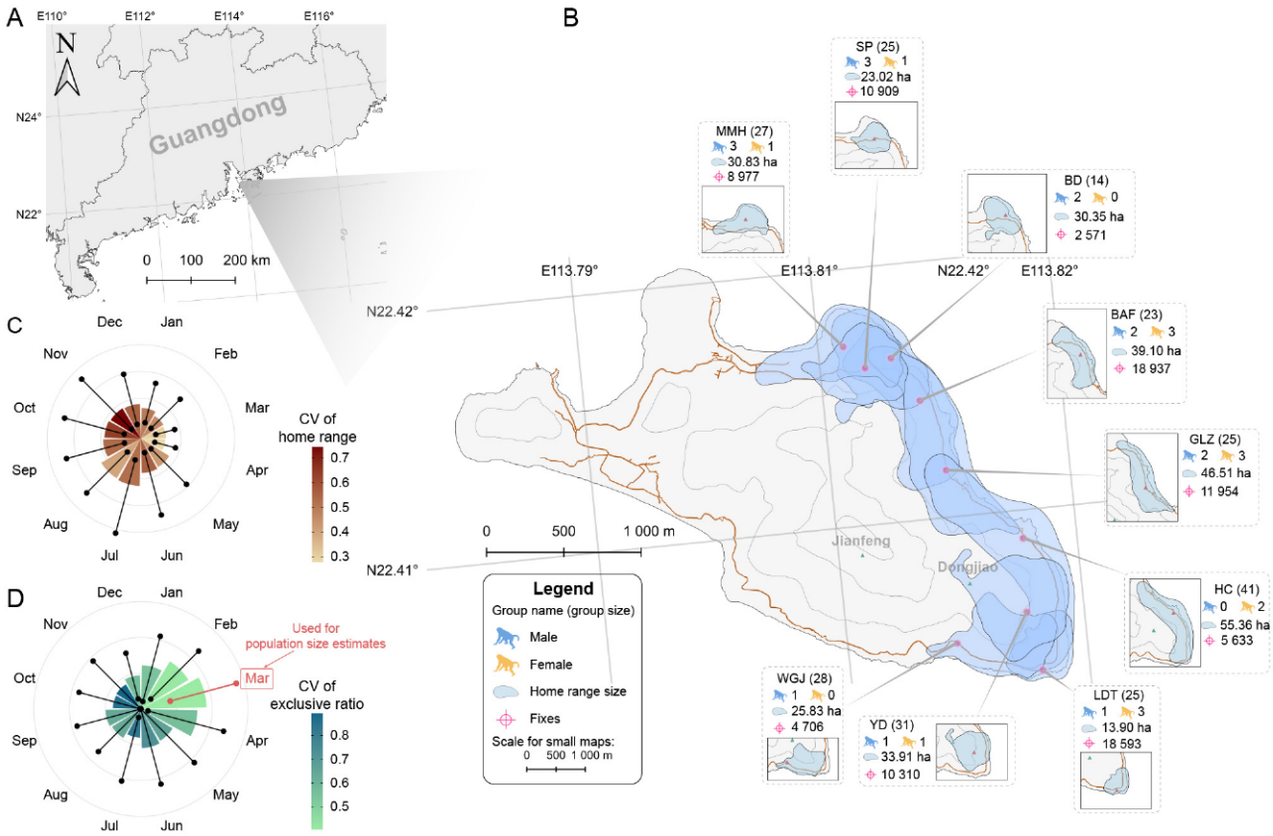


Figure 1 Locations and home range features of nine macaque groups with tracked individuals on Neilingding Island, Guangdong, China

A: Geographical location of Neilingding Island. B: Basic information of each tracked group. Group name, group size, number of collared males and females in each group, annual home range size, and number of fixes acquired are listed from top to bottom. C: Monthly variation in home range of nine macaque groups. Mean (brown bar) and range (black solid line) of monthly home range size are shown for each month. D: Monthly variation in exclusive ratio of nine macaque groups. Mean (green bar) and range (black solid line) of exclusive ratio of monthly home range are shown. March shows the highest exclusive ratio (red), which was used to estimate population size. For abbreviations see text.

was used to minimize the impact of undetected overlap. The highest exclusive home range ratio (55.2%) occurred in March (Figure 1D). To determine the potential capacity of the island, the aggregate exclusive home range ratio (derived from the proportion between the exclusive ranges of the nine groups and total home range) was multiplied by island area. The number of groups was estimated by dividing the island capacity by the average exclusive home range size. Thus, we estimated that the island could support 38.6 groups based on the mean exclusive area value (12.79 ha) in March.

To estimate the population size of rhesus macaques on Neilingding Island, field surveys were conducted to evaluate the size of the nine tracked groups. Group size was variable, ranging from 14 individuals in group BD (including two adult males and five adult females) to 41 individuals in group HC (including four adult males and thirteen adult females), with a mean of 26.4 (SE=7.1; Figure 1B). Considering that the largest group HC is a semi-provisioned group, which has been feeding regularly by reserve managers on the island, thus this group was omitted from population size calculation. And the adjusted mean group size decreased to 24.8 (SE=5.0).

By multiplying mean group size by the number of groups, the macaque population size was estimated at 957 (SE=193.0) individuals. This is comparable to the most recent estimation of 989 individuals (24 groups) using direct counting (Chu et al., 2019), although group number and mean group size differed. Due to significant home range overlap among adjacent groups, differentiating between groups and achieving accurate individual counts was challenging. This may have led to the previous study overestimating group size and underestimating group number. Alternatively, certain larger groups may have split into smaller groups between 2016 and 2021. Notably, after the conclusion of this study, several macaques from the HC group separated and established a new smaller group comprising 21 members.

To compare the utilization distribution (UD) similarity between each individual and their group, the Bhattacharyya's affinity (BA) overlap index was applied. The mean BA value among all groups was 0.98 (SE=0.06, range: 0.40–1.00). Of note, 93% of the BA values exceeded 0.95, with only 20 values falling below this threshold (Supplementary Figure S1), 19 of which were calculated from a small sample size of less than 100 fixes, indicating that increasing sample size may improve similarity between individuals and groups. These results suggest that for species in cohesive groups, like the macaques in this study, a single GPS collar may adequately represent both the home range size and utilization patterns of the entire group, thereby offering a cost-effective approach for such research.

For animal species residing in dense forests that vocalize infrequently, options for population assessment remain limited. This research presents an approach that uses GPS collars to estimate animal population sizes. In contrast to conventional survey methods, the presented approach confers multiple

distinct benefits. First, utilizing GPS technology to track wild animals overcomes challenges associated with long-term sampling in inaccessible or remote terrain, thereby reducing the need for human resource requirements. Large field teams often face significant logistic challenges. In our study, however, only three team members were required for all fieldwork, including macaque anesthetization, GPS collar attachment, and group size determination. Second, this approach is independent of the experience and observational ability of surveyors, thus ensuring high repeatability and reliability. Sustained and scientifically rigorous monitoring is important for delineating population dynamics (Chapman et al., 2010). Unlike the application of GPS devices, it is difficult for a single researcher to engage in continuous population monitoring to ensure comparability between surveys. Third, GPS collars can generate extensive datasets capturing the movements of multiple individuals simultaneously. Such data can play an important role in addressing ecological questions, such as those related to habitat selection, inter- and intra-group spatial interactions, and ecological constraints.

Our study was conducted on a small island with limited heterogeneity in terrain, elevation, and forest type, with the macaque groups exhibiting a relatively uniform distribution (Chu et al., 2019). Consequently, we did not consider habitat heterogeneity in our study. However, in environments with pronounced heterogeneity, including groups from different habitats in the sampling process is necessary to ensure comprehensive representation of the area (Kühl et al., 2008). The macaque population density is notably high on the island, and neighboring groups exhibit considerable overlap in home range. Consequently, we introduced an exclusive utilization ratio into the estimation function for quantifying groups. It should be noted that this function may not be applicable for areas with low population density, where the home ranges of neighboring individuals or groups do not overlap. In such instances, alternative methods exist to estimate population density, such as the Thiessen polygon approach (Bermejo et al., 2006), which employs home range centroids to construct polygons regardless of home range area or overlap. Lastly, when confronted with obvious variation in monthly home range, populations should be monitored for at least one year.

In conclusion, our study demonstrates the viability of integrating GPS telemetry and group censuses as a promising approach for wildlife population surveys. This method is particularly effective for assessing medium- to large-sized animals that occupy stable home ranges and live in dense forests with low visibility. The advantages of low labor costs, high repeatability, and inherent accuracy make it a promising approach for broad utilization across different taxa. We anticipate that the future incorporation of novel technologies will further facilitate wildlife monitoring in the fight against biodiversity loss.

SCIENTIFIC FIELD SURVEY PERMISSION INFORMATION

Permission for field surveys in Neilingding Island was granted by the Neilingding-Futian National Reserve. Permission for GPS collar implementation was granted by the Forestry Administration of Guangdong Province.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

P.F.F. and Y.X.F. contributed to the conception and design of the study. P.F.F., C.M., Y.X.F., P.Z.X., T.C., X.Y., H.L.X., and Q.Y. conducted the field study. Y.X.F. and P.F.F. wrote the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Dr. Li Yang for helping with statistical analyses and Shi-Ming Peng and Wei-Jun Xie from Guangzhou Zoo for assistance in anesthetizing macaques. We are grateful to colleagues from the Neilingding-Futian National Nature Reserve for assisting with field work.

Yu-Xuan Fan[1], Pu-Zheng Xie[1], Chi Ma[2], Ting Chen[3], Xiao Ye[3], Hua-Lin Xu[3], Qiong Yang[3], Peng-Fei Fan[1,*]

[1] *School of Life Sciences*, *Sun Yat-sen University*, *Guangzhou*, *Guangdong* 510275, *China*

[2] *College of Agriculture and Biological Sciences*, *Dali*, *Yunnan* 671003, *China*

[3] *Neilingding-Futian National Nature Reserve of Guangdong*, *Shenzhen*, *Guangdong* 518040, *China*

*Corresponding author, E-mail: fanpf@mail.sysu.edu.cn

REFERENCES

Bermejo M, Rodríguez-Teijeiro JD, Illera G, et al. 2006. Ebola outbreak killed 5000 gorillas. *Science*, **314**(5805): 1564.

Chapman CA, Struhsaker TT, Skorupa JP, et al. 2010. Understanding long-term primate community dynamics: implications of forest change. *Ecological Applications*, **20**(1): 179−191.

Chu YMR, Zan QJ, Yang Q, et al. 2019. Population dynamic and viability analysis of rhesus macaque (*Macaca mulatta*) in Neilingding Nature Reserve, Guangdong Province. *Chinese Journal of Wildlife*, **40**(2): 259−266. (in Chinese)

Koen EL, Bowman J, Findlay CS, et al. 2007. Home range and population density of fishers in eastern ontario. *Journal of Wildlife Management*, **71**(5): 1484−1493.

Kühl H, Maisels F, Ancrenaz M, et al. 2008. Best Practice Guidelines for Surveys and Monitoring of Great Ape Populations. Gland, Switzerland: IUCN SSC Primate Specialist Group (PSG).

Li YH. 2021. A review on estimating population size of large and medium-sized mammals. *Biodiversity Science*, **29**(12): 1700−1717. (in Chinese)

Nathan R, Monk CT, Arlinghaus R, et al. 2022. Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science*, **375**(6582): eabg1780.

Piel AK, Crunchant A, Knot IE, et al. 2022. Noninvasive technologies for primate conservation in the 21st century. *International Journal of Primatology*, **43**(1): 133−167.

Zhao et al. *Zool. Res.* 2024, 45(1): 39−54
https://doi.org/10.24272/j.issn.2095-8137.2023.047

ZR | Zoological Research

Article Open Access

Dynamic foraging strategy adaptation to heterogeneous environments contributes to social aggregation in snub-nosed monkeys

Lan Zhao[1,2,#], Sheng-Nan Ji[3,#], Xiao-Bing Du[1], Jia-Hui Liu[1], Bo-Lun Zhang[1], Pei-Hua Li[1], Yi-Jun Yang[1], Bao-Guo Li[1], Yan-Qing Guo[1], Xiao-Guang Qi[1,*]

[1] *College of Life Sciences*, *Northwest University*, *Xi'an*, *Shaanxi* 710069, *China*

[2] *Chengdu Institute of Biology*, *Chinese Academy of Sciences*, *Chengdu*, *Sichuan* 610041, *China*

[3] *State Key Laboratory of Environmental Criteria and Risk Assessment*, *Chinese Research Academy of Environmental Sciences*, *Beijing* 100012, *China*

ABSTRACT

The dynamics of animal social structures are heavily influenced by environmental patterns of competition and cooperation. In folivorous colobine primates, prevailing theories suggest that larger group sizes should be favored in rainforests with a year-round abundance of food, thereby reducing feeding competition. Yet, paradoxically, larger groups are frequently found in high-altitude or high-latitude montane ecosystems characterized by a seasonal scarcity of leaves. This contradiction is posited to arise from cooperative benefits in heterogeneous environments. To investigate this hypothesis, we carried out a six-year field study on two neighboring groups of golden snub-nosed monkey (*Rhinopithecus roxellana*), a species representing the northernmost distribution of colobine primates. Results showed that the groups adjusted their movement and habitat selection in response to fluctuating climates and spatiotemporal variability of resources, indicative of a dynamic foraging strategy. Notably, during the cold, resource-scarce conditions in winter, the large group occupied food-rich habitats but did not exhibit significantly longer daily travel distances than the smaller neighboring group. Subsequently, we compiled an eco-behavioral dataset of 52 colobine species to explore their evolutionary trajectories. Analysis of this dataset suggested that the increase in group size may have evolved via home range expansion in response to the cold and heterogeneous climates found at higher altitudes or latitudes. Hence, we developed a multi-benefits framework to interpret the formation of larger groups by integrating environmental heterogeneity. In cold and diverse environments, even smaller groups require larger home ranges to meet their dynamic survival needs. The spatiotemporal distribution of high-quality resources within these expanded home ranges facilitates more frequent interactions between groups, thereby encouraging social aggregation into larger groups. This process enhances the benefits of collaborative actions and reproductive opportunities, while simultaneously optimizing travel costs through a dynamic foraging strategy.

Keywords: Social evolution; Folivore paradox; MLS; *Rhinopithecus roxellana*; Multi-benefits framework

INTRODUCTION

Given the diversity of mammalian social systems, many important questions regarding their origins and impacts remain unresolved (Clutton-Brock, 2021). Although variation in feeding competition associated with differences in food distribution plays an important role in restricting group size, our understanding of this variation remains incomplete (Clutton-Brock & Janson, 2012). Several socioecological models predict that within-group feeding competition is weak among folivorous primates because food resources are presumed to be superabundant and evenly distributed, theoretically allowing individuals the freedom to form larger groups in areas with a high and consistent food supply (Isbell, 1991; Isbell & Young, 2002; Janson & Goldsmith, 1995). In contradiction to these models, however, many folivores live in relatively small groups, a phenomenon referred to as the

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 04 May 2023; Accepted: 08 September 2023; Online: 09 September 2023

Foundation items: This research was supported by the Biodiversity Survey and Assessment Project of the Ministry of Ecology and Environment, China (2019HJ2096001006), National Natural Science Foundation of China (32001099, 32170512, 32370524), and China Postdoctoral Science Foundation (2020M683539)

[#]Authors contributed equally to this work

*Corresponding author, E-mail: qixg@nwu.edu.cn

folivore paradox (Steenbeek & Van Schaik, 2001; Sterck & Roth, 2022).

The ecological constraints model posits that folivorous primates, which typically forage for food located in discrete patches, face increasing challenges with increasing group size. Specifically, the depletion of resources within these patches and escalating travel costs associated with searching for new patches intensify with group expansion (Chapman & Chapman, 2000; Snaith & Chapman, 2005; Teichroeb & Sicotte, 2009). The increase in travel costs predicts the occurrence of within-group scramble competition, effectively constraining folivore groups to smaller sizes (Chapman & Teichroeb, 2012). This hypothesis provides an ecological rationale for why most colobine primates inhabiting leaf-rich tropical areas tend to form small groups, characterized by one-male multi-female families, often consisting of fewer than 20 individuals (Figure 1A).

However, certain species exhibit considerable deviations from the predictions of the ecological constraint model. Notably, snub-nosed monkeys (*Rhinopithecus* species), Rwenzori Angolan colobus monkeys (*Colobus angolensis ruwenzorii*), Himalayan langurs (*Semnopithecus schistaceus*), and Indochinese gray langurs (*Trachypithecus crepusculus*), which inhabit high-altitude or high-latitude montane forests with relatively scarce and patchy vegetation, typically form large, cohesive groups consisting of hundreds of individuals (Figure 1A) (Adams et al., 2021; Grueter et al., 2020; Qi et al., 2014, 2023). Among these species, the endangered golden snub-nosed monkey (*Rhinopithecus roxellana*), recognized as the most northerly distributed colobine primate (Yu et al., 2022), resides in a complex multi-level society (MLS) comprised of hundreds of individuals organized into extensive social matrices (Figure 1A). The foundational one-male multi-female families within their social system do not exist autonomously, but rather integrate into larger groups where multiple such families coalesce, forming a cohesive breeding band. Concurrently, non-breeding males congregate in all-male bachelor bands, which are composed of several all-male units (Grueter et al., 2017; Qi et al., 2014). Bachelor males within these units typically travel in pursuit of different breeding bands in order to secure potential mating opportunities, leading to an unstable unit composition (Fang et al., 2018a; Li et al., 2020; Qi et al., 2017). A cohesive breeding band forms associations with one or more dynamic all-male bands, together constituting a herd. These herds aggregate seasonally to form a troop, during which members are exchanged to facilitate gene flow and prevent inbreeding (Qi et al., 2014). Ancestral state reconstruction studies have revealed that this complex MLS evolved from the aggregation of independent one-male, multi-female groups, a social structure characteristic of ancestral Asian colobines and similar to that found in most tropical colobines (Grueter & Van Schaik, 2010; Qi et al., 2023). Among the four hierarchical levels of the MLS—unit, band, herd, and troop—the breeding band is regarded as a relatively stable core structure. Members of breeding bands engage in various collective activities, such as traveling, resting, and foraging, necessitating access to ample food resources (Qi et al., 2014, 2017). Nonetheless, colobine species residing in high-altitude regions encounter substantial energetic challenges due to cold temperatures and seasonal food scarcity (Grow et al., 2014; Hemingway & Bynum, 2005). Understanding how these species resolve such energetic constraints to form large, cohesive bands is a critical question in socioecological research (Grueter et al., 2020).

Social complexity in animal groups is shaped by both the size of the social unit and the interactions among its members (Kappeler, 2019). The heterogeneity hypothesis posits that in montane ecosystems, where climate and resources vary both spatially and temporally, there is an increased likelihood of encounters between one-male multi-female groups in colobines (Grueter, 2022), particularly when food resources are low in quality and spatiotemporally and heterogeneously distributed. In such cases, a small home range may not be sufficient for group survival across all seasons, necessitating an expansion of the home range to encompass diverse habitat types year-round. The larger the home range, the higher the probability of encountering other groups or individuals with similar spatial and dietary requirements. Considering the trade-offs in group living (Garcia et al., 2021; Markham & Gesquiere, 2017), larger group aggregation may be favored over smaller independent groups if it offers greater benefits, such as foraging and reproduction, assuming travel costs are not a limiting factor. Therefore, we speculate that high-altitude Asian colobines may adapt to their heterogeneous environments by extending their home ranges, and therein gain higher foraging or other benefits from living in larger groups, leading to the formation of a breeding band.

The above hypothesis has not yet been tested and requires empirical evidence for validation. First, it is essential to establish the relationship between environmental heterogeneity and group size, specifically whether colobines residing in large bands encounter highly heterogeneous resources. Second, it is necessary to clarify whether bands modify their ranging behaviors to access heterogeneous resources. Third, it is also important to determine whether larger groups gain enhanced benefits, such as access to food and reproductive opportunities, without a significant increase in travel costs. Additionally, when exploring evolutionary trajectories, two further hypotheses should be examined based on comparative eco-behavioral analysis across colobine species, notably, whether there is an expansion of home range in heterogeneous environments and whether an increase in home range size is correlated with larger group sizes.

To test the socioecological hypothesis, extensive, long-term monitoring of intraspecific variation within a model species is necessary, alongside comparative studies that integrate ecological and phylogenetic data (Clutton-Brock & Janson, 2012; Kappeler et al., 2019). The golden snub-nosed monkey, which inhabits temperate montane forests with seasonally variable food availability and is characterized by an MLS (Tan et al., 2016), serves as an excellent model species for such research. While golden snub-nosed monkey bands typically select habitats according to food availability (Fan et al., 2019; Huang et al., 2021; Li et al., 1999, 2000; Tan et al., 2007), it remains unclear whether they adjust their ranging behaviors to exploit heterogeneous resources, as predicted by the heterogeneity hypothesis. Tracking these primates poses a significant challenge. Direct observations and long-term monitoring of wild populations in montane forests are exceptionally demanding tasks, with the reliability of radio telemetry often hindered by fluctuating weather conditions and difficult terrain (Beyer & Haufler, 1994). The adoption of global positioning system (GPS) collars has helped address these tracking challenges (Tomkiewicz et al., 2010), allowing for the

collection of accurate activity data (Fang et al., 2018b; Qi et al., 2014). In 2012, we initiated a continuous monitoring program using GPS collars to determine the home range and core areas of the golden snub-nosed monkey, which enabled vegetation sampling across diverse habitats to evaluate habitat quality. Given the close association between travel behaviors and habitat quality, the precise movement data obtained from GPS tracking also provided the opportunity to track the dynamic patterns of habitat utilization (Van Moorter et al., 2016).

Over six years, we collected meteorological, home range, diet, and habitat data of two golden snub-nosed monkey bands. Based on these data, we explored three predictions of the heterogeneity hypothesis. (1) We analyzed spatial and temporal variations in the home range of a single band to clarify habitat utilization, then compared habitat parameters, including food richness and food diversity across different levels of habitat use, providing direct evidence of resource heterogeneity. (2) We tracked spatiotemporal movement, diet, and habitat characteristics of a single band to determine how it adjusts its ranging behaviors to access heterogeneous food resources. (3) We compared the social organization, home range, daily travel distance (DTD), and habitat quality between two neighboring bands to determine whether the larger band achieves superior feeding and reproductive opportunities and manages travel costs more efficiently compared to the smaller band. Subsequently, we compiled an eco-behavioral dataset covering bioclimate, elevation, body mass, group size, and home range across colobines to evaluate the last two aspects of our hypothesis. (4) We conducted a comparative phylogenetic analysis to examine the correlations between ecological variables and group size to determine whether heterogeneous climates affect group size. (5) We employed causal models with phylogenetic path analysis to explore the evolutionary relationship between home range and group size and ascertain whether (and how) an expanded home range leads to a larger group size. Based on the results, we assessed the validity of the heterogeneity hypothesis in the social evolution of colobines, which, if valid, can be incorporated into existing theoretical frameworks for a better understanding of the social evolution of these primates and potentially of mammals.

MATERIALS AND METHODS

Study sites and subjects

This field study focused on wild groups of the Qinling golden snub-nosed monkey (*R. roxellana qinlingensis*) inhabiting the Yuhuangmiao area within Zhouzhi National Nature Reserve on the northern slopes of the Qinling Mountains, Shaanxi Province, China (Figure 1B). The area experiences fluctuating temperatures typical of its temperate climate, leading to seasonal withering and prolonged periods of leaflessness in winter. The mountainous terrain includes three forest types distributed along an elevational gradient: (1) deciduous broadleaf forest (1 400–2 200 m above sea level (a.s.l.)), (2) mixed coniferous and broadleaf forest (2 200–2 600 m a.s.l.), and (3) coniferous forest (above 2 600 m a.s.l.) (Li et al., 2000).

The study area is home to the wild West Ridge troop (WRT), which includes the Gongnigou herd (GNG) and Dujiafen herd (DJF) (Figure 1C, D). In December and June each year, we recorded the number of adults (males >7 years old, females >5 years old), subadults (males 3–5 years old, females 4–7 years old), juveniles (males 1–4 years old, females 1–3 years old), and infants (<1 year old) in both herds. From 2013 to 2017, the GNG-herd consisted of 160–190 individuals, including a breeding band with 11–15 one-male units (OMUs) and an all-male band of 24–40 individuals, while the DJF-herd consisted of 60–80 individuals, including a breeding band with 6–7 OMUs and an all-male band of 11–12 individuals. The focus of this study was on the cohesive breeding bands due to the frequent migration of members of the all-male bands (Huang, 2015). The GNG breeding band (GNG-BB) has been habituated to receiving food provisions in spring and autumn since 2001 (Qi et al., 2009). Thus, the GPS data obtained in summer (June to August) and winter (December to February) without provisioning are reflective of its natural living conditions. The DJF breeding band (DJF-BB) is a free-ranging group. Additional details are provided in the Supplementary Materials (Supplementary Method S1).

Ethics approval

All animal procedures performed in this research were conducted following the ethical guidelines on the care and use of animals for scientific purposes, as outlined by the Animal Care Committee of the Wildlife Protection Society of China (SL-2012-42) and the Specialist Committee of the National Forestry Administration of China (SFA: LHXZ-2012-2788) and following the regulatory requirements of Zhouzhi National Nature Reserve, China.

Meteorological data collection

A meteorological station was established in the study area (E108°14′, N33°49′) in 2003 (Qi et al., 2008), which monitored air temperature, relative humidity, solar radiation, and evaporation hourly. Daily average, maximum, and minimum air temperatures, relative humidity, and solar radiation were determined using the meteorological data. Climate data were collected from October 2013 to April 2017 (1 278 days).

Estimation of GPS location and ranging behaviors

From December 2012 to January 2018, seven GPS collars were attached to adult members of GNG-BB and eight to those of DJF-BB, enabling the tracking of both groups (Supplementary Table S1). The GPS collars were programmed to record locations every 2 h from 0500h to 1900h each day (Supplementary Method S2). To eliminate significant positional errors, we applied data screening criteria, including three-dimensional fixes and a dilution of precision (*DOP*) no greater than 10, as a higher *DOP* indicates poor satellite geometry and less accurate measurements (Bjørneraas et al., 2010). A total of 38 374 GPS location records (33 059 valid data points after filtering) were used to analyze the ranging behaviors of GNG-BB and DJF-BB. To mitigate the impact of food provisioning on GNG-BB, we excluded its data from spring and autumn. In addition, we removed duplicate locations recorded for individuals at the same time to prevent pseudoreplication. In total, 6 158 location data points (longitude, latitude, elevation) were used for calculating the ranging behavior (home range, core area, and DTD) of GNG-BB in summer and winter and 8 192 location data points were used to calculate the ranging behavior of DJF-BB across all seasons. The kernel density estimation method in ArcGIS v.10.6 (ESRI., USA) was used to calculate the home range and core area each year

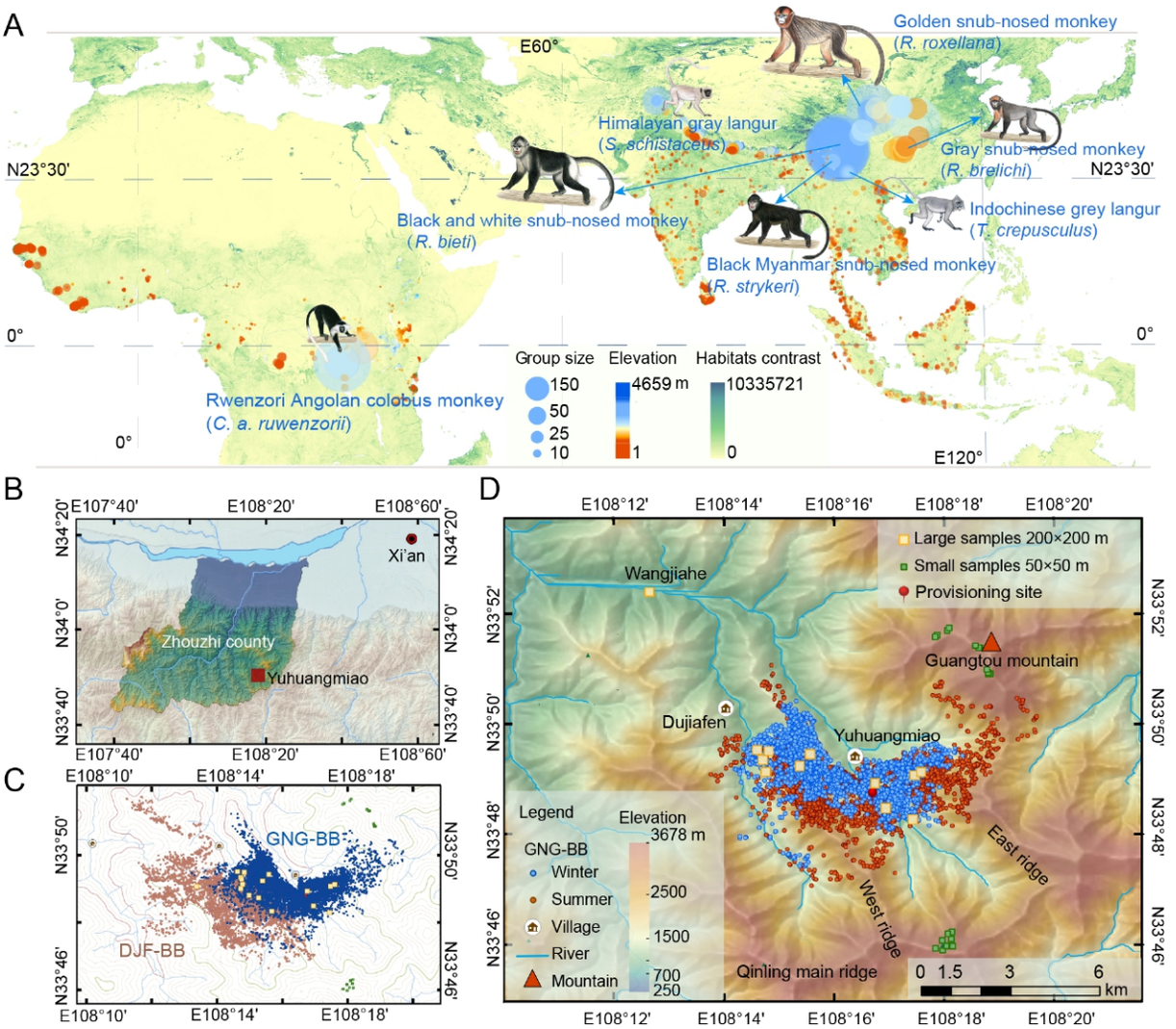


Figure 1 Distribution of colobines and geographical location of study area

A: Distribution of colobines. Map displays exponentially weighted difference in Enhanced Vegetation Index (EVI) between adjacent pixels, indicating spatial heterogeneity of habitats, i.e., habitat contrast. Distribution coordinates are scaled by average group size or population size and colored by elevation. B: Geographical location of study sites. C: Mountainous terrain around study area. D: GPS coordinates and sample sites in this study. Illustrations of monkeys: copyright 2013, Stephen D. Nash/IUCN SSC Primate Specialist Group. All illustrations are used with permission.

(Supplementary Method S3). Habitats were then classified into four categories: core area (50%), moderately utilized area (non-core area within the home range, 50%–95%), seldom utilized area (area outside the home range but within the range scope of activity, 95%–100%), and non-utilized area (area outside the range scope of activity) (Supplementary Figure S1; Supplementary Tables S2–S3). The Tracking Analyst module in ArcGIS v.10.6 was employed to track monkey movements and compute distances between adjacent anchor points based on time series coordinates (Supplementary Method S3).

A multi-year home range calculation was conducted to compare home range sizes between the two bands using data from periods when location information was simultaneously available for both bands.

Vegetation sampling

Quadrats for vegetation sampling were established from July 2014 to October 2016. Seventeen large quadrats (200 m×200 m) below 2 200 m a.s.l. and 19 small quadrats (50 m×50 m) above 2 200 m a.s.l. were sampled (Figure 1D; Supplementary Table S2), covering the altitudinal gradient. For better comparability and higher precision when measuring vegetation characteristics under complex terrains, we subdivided each of the original large samples into 16 small quadrats (50 m×50 m). Based on average seasonal habitat utilization over five years, we classified samples into four types of areas (core, moderate, seldom-utilized, and non-utilized) in two separate seasons (Supplementary Tables S2–S3). For example, the S-H-1 quadrat was classified as a highly utilized (core) area during the summer season, while the S-M-3 quadrat was classified as a moderately utilized area in summer but a core area in winter and was thus resampled as W-H-2 (Supplementary Figure S1). As a result, a total of 358 small quadrats were analyzed in the two seasons for GNG-BB. In addition, to minimize conflicts with humans, the study group avoided certain areas (Huang et al., 2021). Consequently, we selected plots within regions frequented by the monkeys that experienced minimal human disturbance to determine the dynamic patterns between habitat characteristics and movements.

A plant dataset encompassing 85 387 live trees within the study area and a diet dataset comprising 1 117.5 h of

observational records of the study group were used to determine habitat characteristics and dietary changes (Huang et al., 2021; Huang, 2015). The main food species were defined as edible foods that ranked within the top 75% of all food plants consumed by individuals in the BB (Huang, 2015). We calculated six basic parameters: the number of species and quantity of plants (PS, PQ), food (FS, FQ), and main food (MFS, MFQ). Additionally, nine vegetation parameters were calculated to evaluate plant and food resources, including the diversity, richness, and evenness of plants (PR, PD, PE), food (FR, FD, FE), and main food (MFR, MFD, MFE) within each quadrat. In addition, average diameter at breast height (a-DBH), total tree basal area (t-TBA), average tree height (a-TH), as well as average and total canopy density (a-CD, t-CD) were determined to characterize the trees in the quadrat. Detailed calculations are presented in the Supplementary Materials (Supplementary Method S4).

Construction of eco-behavioral dataset

An ecological dataset including 19 bioclimatic variables and elevation data for 70 colobine species (including subspecies) was established based on occurrence data from the Global Biodiversity Information Facility (GBIF) and bioclimatic data from WorldClim (Supplementary Method S5 and Dataset S1). A dataset detailing social and ranging behaviors was also complied for 72 colobine species (including subspecies) (Supplementary Method S5 and Dataset S2, S3). A colobine phylogenetic tree was then constructed by integrating existing phylogenetic data (Supplementary Method S5). Finally, data for 50 species and two subspecies were obtained for comparative phylogenetic analysis (Supplementary Dataset S4).

Data analysis

The differences in ranging parameters of GNG-BB between summer and winter were assessed using independent samples *t*-tests (parametric data), Mann-Whitney *U*-test (nonparametric data), and paired *t*-test (paired data). Similarly, seasonal differences in DJF-BB were determined using analysis of variance (ANOVA) and *post hoc* least significant difference tests. The Kaiser-Meyer-Olkin (KMO) and Bartlett's tests were used to measure partial correlations between variables. Due to the significant correlations among multiple parameters, principal component analysis (PCA) was conducted to extract the primary characteristics of the nine meteorological and 20 vegetation variables. The new principal components (PCs) served as proxies for the original variables. We examined the correlations between DTD and meteorological variables (original variables and PCs) using Pearson or Spearman correlation analysis to determine whether ranging behaviors were influenced by climatic changes. Correlations with original variables helped validate the PCA results, while the PCs themselves reflected the principal characteristics of the original variables. Furthermore, spatial and seasonal vegetation differences in the core area were determined using the Kruskal-Wallis test and *post-hoc* Dunn's all-pairs test with false discovery rate (FDR) correction. To understand the spatiotemporal dynamics of habitat utilization, we constructed a time-series framework for foraging behaviors and habitat selection for GNG-BB, integrating real-time movements, meteorological data, dietary records, and habitat parameters. K-means clustering was used to classify DTD into three categories: long, medium, and short. According to Huang (2015), the distances between two OMUs within GNG-BB ranged from 14 to 385 m. To ascertain whether the band stayed in the same habitat patch or migrated to another, we calculated the multiplier of the clustering center of each category to the maximum distance (385 m). A multiplier less than 2 indicated that the band either remained at the same location or moved to a neighboring area within the same habitat patch, while a multiplier greater than 4 or 5 indicated that the band migrated a long distance to another habitat patch in that category.

Next, we compared the social organization, home range, DTD, and habitat quality differences between GNG-BB and DJF-BB using the methods mentioned above. The results were used to evaluate the reproductive benefits of mate access, travel costs, and foraging benefits between the two groups. Analyses were performed using R v.4.1.0 (R Core Team, 2022), SPSS v.26 (SPSS, USA), and Origin Pro 2018 (Origin Lab, USA). Specific analyses and detailed processes are illustrated in the Supplementary Materials (Supplementary Method S6).

For the interspecific eco-behavioral data, we performed correlations and extracted PCs. To accurately identify the effect of environmental variables in shaping home range and group size during evolutionary history, we used phylogenetic generalized least-squares (PGLS) analysis to model the phylogenetic regression relationship among the candidate variables. PGLS can control for the non-independence of data that arise from shared ancestry during analysis by incorporating a phylogenetic tree (Symonds & Blomberg, 2014). After identifying the most significant variables impacting home range and group size in the PGLS analysis, we performed phylogenetic path analysis (PPA) to estimate statistical support for the hypothesized causal links among variables using the R package "phylopath" (Van Der Bijl, 2018). As several variables in the PGLS analysis approached the significance threshold, their impact on the PPA results was uncertain. Consequently, we established three different variable sets (eight, nine, and six variables, respectively) by incorporating or omitting the variables with uncertain significance.

For each set of variables, we designed 16 candidate models to examine both direct or indirect associations among variables and to determine the direction of causality. Using the eight-variable dataset as an example, two models were categorized as direct models to assess the direct effect of elevation on group size, while the remaining models were categorized as indirect models, positing that elevation influences group size indirectly by altering bioclimate variables. To examine the direction of causality, we used odd ordinal models where home range size affects group size, and even ordinal models where group size affects home range size. This allowed us to investigate whether an increase in home range size across species led to an increase in group size. Similarly, the relationships between other variables were also tested by establishing control models for comparison. We assessed the goodness of fit of the candidate models to elucidate potential causal links, identifying the most likely pathways in the best-supported model. The three best-supported models for the three sets of variables were combined to confirm the fifth prediction of the heterogeneity hypothesis, i.e., whether an increase in home range would promote group size.

For all statistical analyses, significant differences were identified at the 95% confidence level (P<0.05).

RESULTS

Range pattern and habitat quality

From 2013 to 2016, the GNG-BB size was approximately 131±13, with an average of 18 more individuals in summer (141±6) compared to winter (123±11) (paired *t*-test: t=3.83, df=3, P=0.031). After accounting for age structure, significant seasonal differences were observed in juveniles (paired *t*-test: t=3.37, df=3, P=0.043) and infants (paired *t*-test: t=3.40, df=3, P=0.043), but not in adults (paired *t*-test: t=1.44, df=3, P=0.25) or subadults (paired *t*-test: t=0, df=3, P=1). The average home range size in summer (13.84±1.59 km^2) was more extensive than that in winter (7.03±2.87 km^2) (paired *t*-test: t=5.00, df=3, P=0.015), including shaded areas with low sunlight radiation that were seldom used in winter. The core areas, comprising only a small portion of the home range (summer: 4.73%, winter: 9.60%), did not show significant size differences between summer and winter (paired *t*-test: t=-0.06, df=3, P=0.873). However, the core area varied based on geographical location, with minimal seasonal overlap (8.5%±6.2%) (Figure 2A; Supplementary Figure S1 and Table S4). Several sites, especially where quadrat S-H-1 was located, were consistently selected by GNG-BB as core areas almost every summer but were seldom utilized in winter (Supplementary Figure S1).

From 2013 to 2016, the DTD for GNG-BB also varied between seasons and was significantly longer in summer (1 529.45±128.31 m) than in winter (939.18±48.18 m) (paired *t*-test: t=7.68, df=3, P=0.005; Figure 2B; Supplementary Table S5). To clarify the factors affecting DTD, we found no statistically significant correlation between band size and DTD (Pearson correlation: r=0.61, n=8, P=0.105). For meteorological variables, PCA was conducted, informed by Bartlett's test (χ^2=3 114.02, n=36, P<0.001) and the KMO test (KMO value=0.72), which indicated that the first two PCs accounted for 87.94% (PC1 49.00%, PC2 38.94%) of the variance (Supplementary Figure S2A). Daily evaporation (20.53%) and solar radiation (18.42%) contributed the most to PC1 (Supplementary Figure S2B); thus, PC1 was considered the main environmental parameter determining how much thermal energy the golden snub-nosed monkey can obtain from the environment. Correlation analysis showed that the lower the thermal energy available, the shorter the DTD of the breeding bands (Pearson correlation: r=0.44, n=332, P<0.01, Figure 2C). PC2 was primarily related to three humidity parameters (Supplementary Figure S2B) and was thus indicative of overall ambient humidity. PC2 was positively correlated with elevation and DTD (Figure 2C). Different patterns were observed when the data for each season were analyzed separately. DTD was significantly shorter on colder and drier days in winter, but it did not respond to PC1 and PC2 in summer (Figure 2D).

Habitats were reclassified into four levels based on the ranging patterns of the breeding bands. For each quadrat within these habitats, 20 variables were separately calculated in summer and winter. The correlation matrix revealed significant correlations among numerous habitat variables, such as food diversity, food richness, and food evenness (Supplementary Figure S3A, B). PCA was performed based on Bartlett's test (summer: χ^2=3050.98, df=190, P<0.01; winter: χ^2=3221.37, df=190, P<0.01) and the KMO test (summer: KMO value=0.76; winter: 0.8).

The initial three PCs from the summer PCA were not readily interpretable or visualizable in their original form. Thus, we rotated these PCs to obtain a new set of components with a more intuitive interpretation. These three rotated PCs (RCs) explained 73% of the variance in the original 20 variables (Figure 2E). RC1 was positively correlated with PS, PR, and PD (Figure 2E), indicating that a high RC1 score denoted a wide variety of plants. Given the herbivorous behavior of the golden snub-nosed monkey, plant parameters showed significant positive correlations with food-related parameters (Supplementary Figure S3A), implying that RC1 also represented the diversity and richness of food resources (i.e., FR, FS, and FD) (Figure 2E). Analysis using the RC plot and multiple comparisons showed that the RC1 scores of the utilized areas were significantly higher than those of the non-utilized areas; however, there was no significant difference between highly and moderately utilized areas (Figure 2F, G).

RC2 predominantly reflected the evenness, quantity, and diversity of the main foods (MFE, MFQ, and MFD) (Figure 2E). Main foods consisted of the 19 types of foods frequently consumed from the 86 available foods consumed in summer. Higher RC2 scores signified both a large quantity and high diversity of main foods in the plot, as well as their balanced proportions, serving as a comprehensive indicator of the quality and quantity of preferred foods. The RC2 scores decreased significantly in the order: core area, moderately utilized area, seldom utilized area, and non-utilized area (Figure 2F, G), suggesting that RC2 was a critical factor in differentiating levels of habitat utilization. RC3 showed positive correlations between tree characteristics and a-DBH, PQ, and a-CD (Figure 2E). Higher RC3 values suggested that trees in these quadrat areas were taller and thicker than those in the other areas. The value of RC3 in the core area was moderate but lower than that in the moderately utilized area and higher than that in the non-utilized area (Figure 2G).

In winter, PCA revealed three RCs accounting for 79% of the total variance (Figure 2H). RC1 was positively correlated with PD, PR, and PS, as well as FR, indicating that RC1 may be a representative variable for plant quality and food richness (Figure 2H). Both RC plot and multiple comparison analyses showed that RC1 scores were higher in utilized areas than in non-utilized areas but the differences between the three types of utilized areas were not significant (Figure 2I, J). Similarly, RC3 served as an indicator of tall and large trees within the quadrats, with findings suggesting that trees in the core area were of moderate size (Figure 2H–J). In contrast to summer, RC2 in winter was positively correlated with the quantity of food, main foods, and plants (FQ, MFQ, PQ), as well as total canopy density (Figure 2H), representing overall food abundance. The RC2 scores decreased significantly in the order: core area, moderately utilized area, and seldom utilized area (Figure 2J).

Spatiotemporal dynamics of habitat utilization

To understand the spatiotemporal dynamics of habitat utilization, we combined the dietary records of GNG-BB with their movement patterns and habitat characteristics. Using clustering analysis, DTD was divided into three categories: long, medium, and short, with respective clustering centers at 2.18 km, 1.37 km, and 0.74 km. The short-distance category indicated that the breeding band remained within the same habitat patch, with its cluster center to maximum diameter ratio being less than 2. Conversely, the long-distance category was indicative of migration over a long distance to gain access

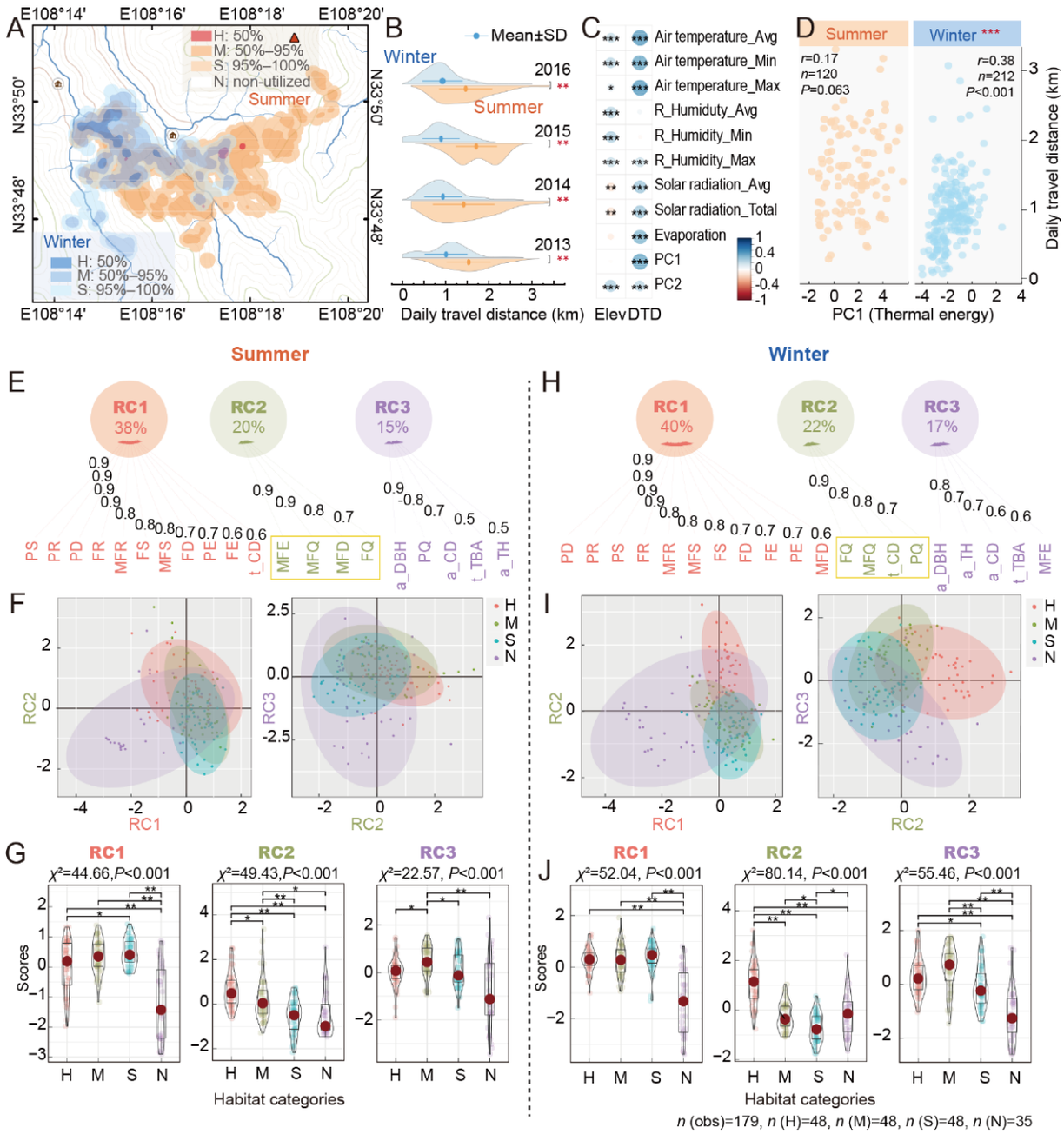


Figure 2 Seasonal ranging patterns and habitat characteristics of GNG-BB

A: Home range and core area in summer and winter for GNG-BB in 2014–2015. B: Differences in DTD between seasons in 2013–2016. C: Correlation matrix between elevation, daily movement parameters, and meteorological variables; Elev: daily average elevation, DTD: daily travel distance. D: Scatterplots showing correlation between DTD and PC scores. E–J: Habitat characteristics of GNG-BB in summer (E–G) and winter (H–J). E, H: Correlation between original variables and RCs in summer (E) and winter (H). Values above lines show correlation coefficient of variables to each RC. F, I: Scatterplot of three RCs in summer (F) and winter (I). G, J: Multiple comparisons of RC scores between different habitats in summer (G) and winter (J). Violin plots show distribution features of data. Kruskal-Wallis test detected differences between three categories and Dunn's test with FDR correction was used for *post-hoc* multiple comparisons. *: P<0.05; **: P<0.01; ***: P<0.001. H: highly utilized area; M: moderately utilized area; S: seldom utilized area; N: non-utilized area; PS: number of plant species; FS: number of food species; MFD: number of main food species; PQ/FQ/MFQ: quantity of all plants/food species/main food species; PD/FD/MFD: diversity of plants/food/main food species; PR/FR/MFR: richness of plants/food/main food species; PE/FE/MFE: evenness of plants/food/main food species; t-TBA: total tree basal area; a-DBH: average diameter at breast height; a-TH: average tree height; t-CD: total canopy density; a-CD: average canopy density.

to another patch, with its multiplier being greater than 5. The time series curve showed that DTD fluctuated dynamically throughout the year, with alternating peaks and valleys. A recurring pattern emerged where periods of long-distance migration were typically followed by phases of moderate and short-distance travel (Figure 3A).

In early summer 2014, GNG-BB often occupied the area encompassing the large S-H-1 quadrat, a site containing many diurnal and nocturnal locations (Figure 3B). This quadrat was predominantly populated with Liaodong oak trees (*Quercus liaotungensis* Koidz., 29.3% quantity proportion, same below), vines (*Celastrus gemmatus*, 14.9%), and all other main food

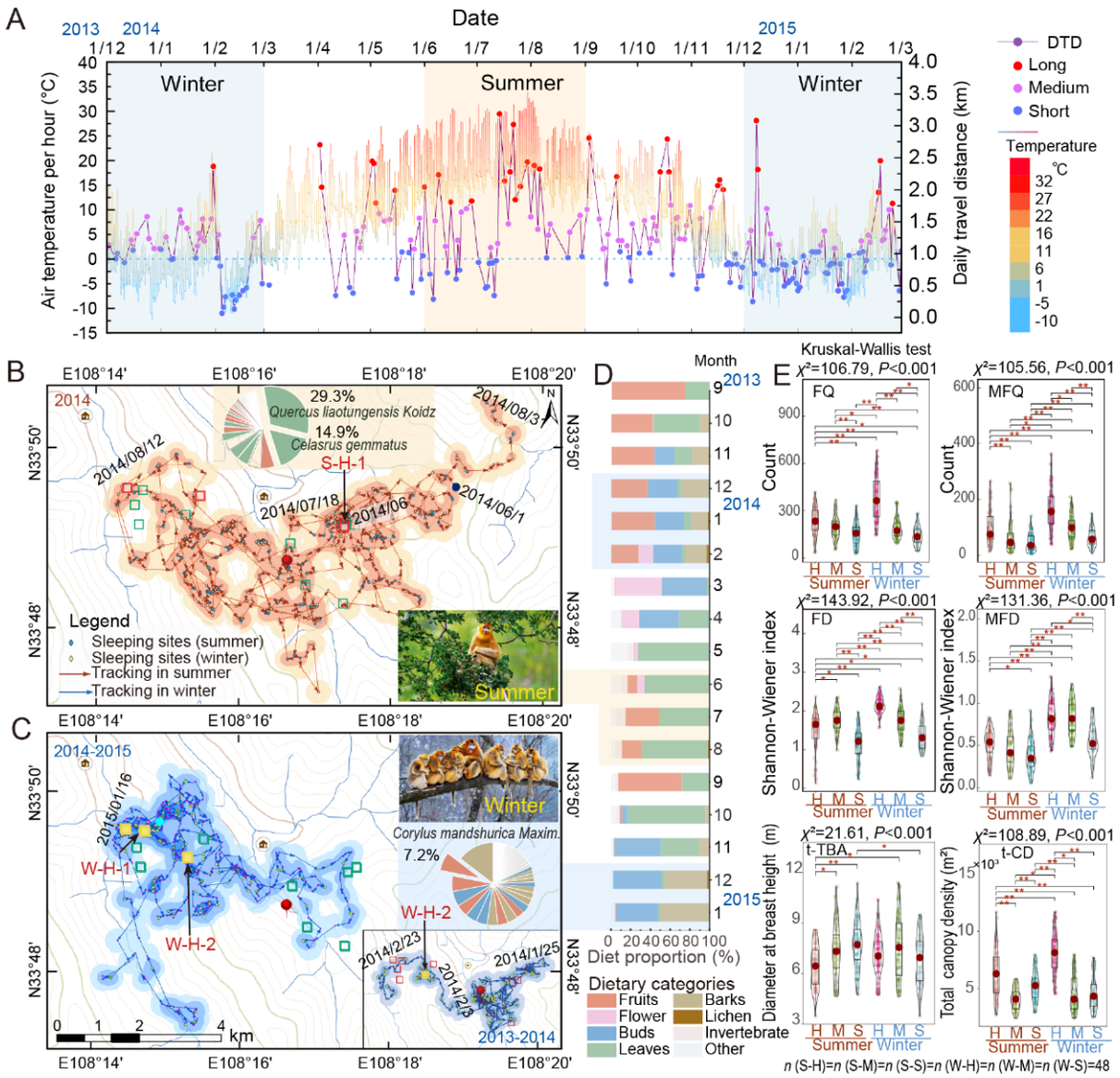


Figure 3 Seasonal differences in daily travel patterns and habitat utilization of GNG-BB

A: Dynamic changes in DTD and temperature from December 2013 to February 2015. Color of dots represents categories of distances. Red: long; purple: moderate; blue: short. B, C: Travel routes in summer (B) and winter (C). Arrows indicate direction of travel, and travel data are labeled above arrows. Pie chart highlights top 20 plant species in quadrats, with colors representing dietary categories. Diamond-shaped points are locations at night (1900h and 0500h), i.e., sleeping sites. D: Diet composition from September 2013 to January 2015. E: Differences in habitat characteristics between summer and winter. Abbreviations are as in Figure 2. Photographs of habitat are copyrighted by Kuan-Liang Ding and used with permission.

species typical of June (Figure 3B). Liaodong oak leaves are not only edible, but the tree branches are also suitable for roosting (Li, 2006). Despite the abundance of Liaodong oak in this area, monkeys preferred the tender leaves of *C. gemmatus*, which constituted the highest proportion of their diet in June (22.8%) (Figure 3B, D). Correspondingly, a large and dense number of feeding and sleeping locations were found in this small core area (<0.2 km^2) (Figure 3A). At the end of June, as the tender leaves of *C. gemmatus* matured, the group migrated to other areas, with a shift in their dominant food to *Actinidia kolomikta* (July 19.1%) and *Ilex polyneura* (August 26.8%).

With the onset of winter and declining temperatures, the monkeys gradually shortened their travel distances (Figure 3A), migrated to sunnier mountain slopes (Figure 3C), and shifted their diet to fruits (including seeds), bark, and buds (Figure 3D). In the cold winter of 2013 (December 2013 to February 2014), when oak and birch fruits were abundant, the band traveled long distances to use the habitat covering quadrat W-H-2, which was rich in oak and hazel trees, for over half a month (Figure 3A, C, D). However, in the winter of 2014, characterized by an unusual climate leading to a scarcity of these seeds and fruits, GNG-BB reduced the utilization of W-H-2, instead expanding its home range to cover other habitat types such as W-H-1 and W-H-3 (Figure 3C). Overall, during the coldest periods of both years, the band traveled the shortest daily distances within the core areas, where food availability and canopy density were highest (Figure 3A, C, E).

Comparison of eco-behaviors between breeding bands

From 2013 to 2016, GNG-BB comprised approximately 12±1 adult males and 51±6 adult females with reproductive ability

(Figure 4A). The all-male band following GNG-BB, seeking access to adult females, expanded from 24 in June 2013 to 40 in June 2016, with the number of adult males increasing from 8 to 14. In contrast, DJF-BB contained 57±1 individuals, including 6–7 adult males and 14–16 adult females. The all-male band following DJF-BB consisted of 11–12 males, with the number of adults decreasing from 10 in 2013 to 8 in 2016. Consequently, the overall size of GNG-BB, along with its count of adult males and females, was approximately two to three times larger than that of DJF-BB during the same period (paired *t*-test: t=8.56, df=3, P=0.003; Figure 4A).

Similar to GNG-BB, DJF-BB also exhibited seasonal variations in home range size and DTD (Figure 4C; Supplementary Figure S4A and Tables S6, S7). Throughout the year, DTD displayed dynamic fluctuations with alternating peaks and valleys (Supplementary Figure S4B). These variations were significantly positively correlated with environmental thermal energy in spring and winter but were less influenced by it in summer and autumn (Supplementary Figure S4C, D). However, notable differences were found between the two breeding bands. The multi-year home range of GNG-BB was smaller than that of DJF-BB in both summer (GNG: 16.83 km^2; DJF: 21.93 km^2) and winter (GNG: 11.27 km^2; DJF: 14.82 km^2), a pattern that persisted when combining the home ranges across both seasons (Figure 4B; Supplementary Figure S4A and Table S8). In addition, the multi-year DTD of GNG-BB (median: 1 495.15 m) was significantly greater than that of DJF-BB (median: 1 150.18 m) in summer (Mann-Whitney *U*-test, U=1 524, Z=−4.53, P<0.001), which also persisted when comparing the two bands separately in 2015 and 2016 (Figure 4C). However, there was no significant difference in DTD between the two bands in winter (Mann-Whitney *U*-test, U=4 786, Z=−1.50, P=0.135; Figure 4C; Supplementary Table S8).

When assessing the quality of habitats used by the bands, multiple variables, including food quantity and quality, were markedly higher in the core area of GNG-BB than in the core area of DJF-BB in both summer and winter (Figure 4D, E;

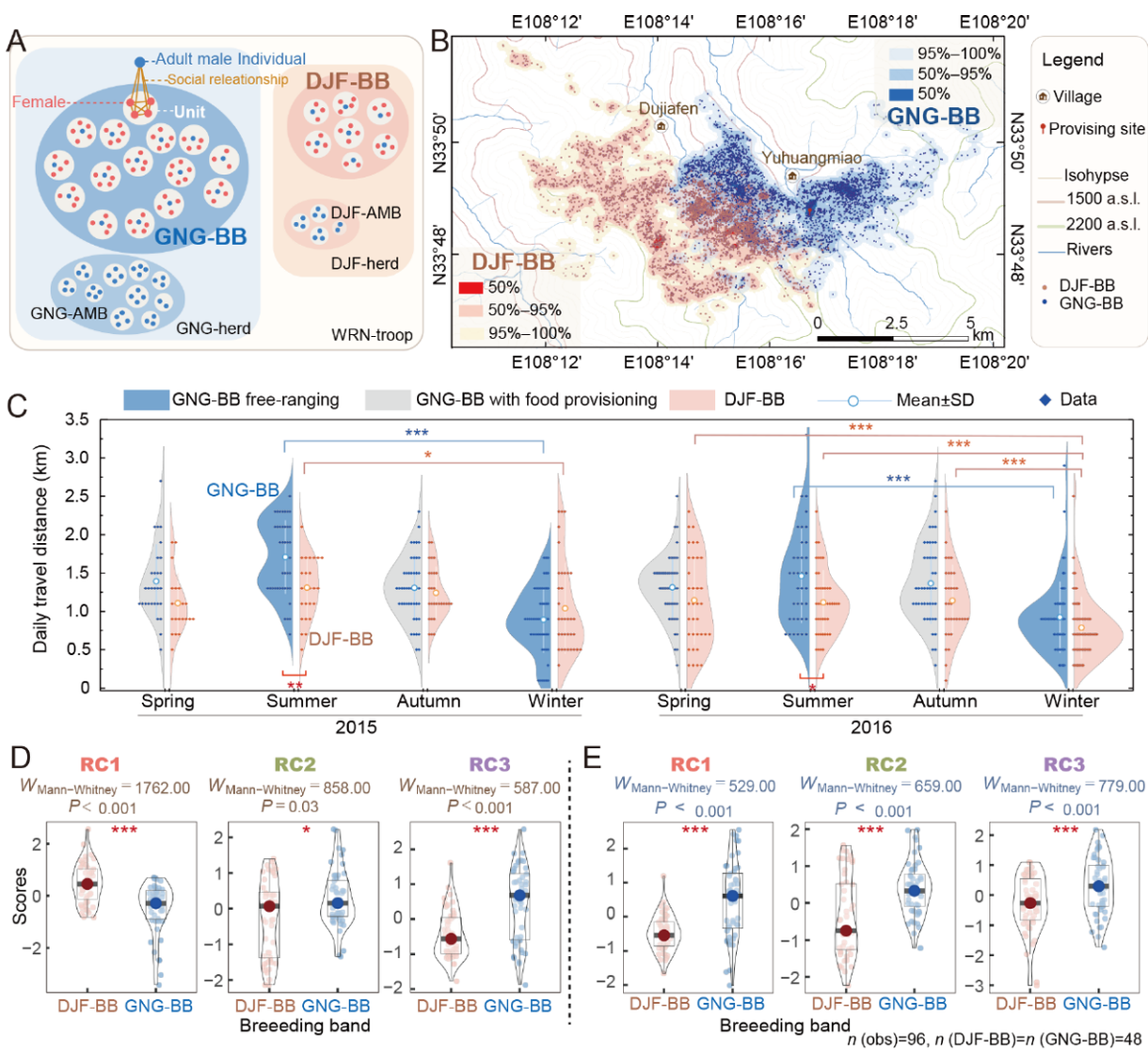


Figure 4 Differences in social organization, ranging behaviors, and habitat quality between bands

A: Differences in social organization between GNG-BB and DJF-BB. GNG-BB: GNG breeding band; GNG-AMB: GNG all-male band; DJF-BB: DJF breeding band; DJF-AMB: DJF all-male band. B: Multi-year (2014–2017) merged home range (summer and winter) of GNG-BB and DJF-BB. C: Differences in DTD in four seasons between GNG-BB and DJF-BB in 2015 and 2016. D, E: Comparison of RC scores between GNG-BB and DJF-BB in summer (D) and winter (E). Mann-Whitney *U*-test compared differences between GNG and DJF.

Supplementary Figure S5). In winter, almost all variables were greater for GNG-BB habitat (Supplementary Figure S5B), indicating that the larger group resided in higher quality habitats than the smaller group (DJF-BB).

Evolutionary pathways of colobines

Among the colobine species, those with larger groups inhabited larger home ranges, particularly in colder regions (Figure 5A). Notably, supergroups emerged in the cold high-altitude mountainous forests and occupied large home ranges, while smaller groups formed in the warmer environments and resided in smaller home ranges (Figure 5A). PGLS analysis revealed a significant positive relationship between home range and group size (λ=0.57, *n*=52, estimate=1.23, *t*=5.96, *P*<0.001). Further analysis detected significant correlations with multiple bioclimatic variables (log-transformed for non-normal data) (Supplementary Figure S6). Four PCs explaining 92.56% of the total variance in 19 bioclimatic variables. These PCs were characterized by isothermality, mean temperature of the warmest quarter, precipitation of the wettest quarter, and diurnal range of temperature (Supplementary Tables S9–S11). Total scores of the four PCs (PCT) were also calculated by weighting the variance percentages. Higher PCT scores indicated a stable, warm, and humid climate, i.e., a benign environment. Conversely, lower score indicated a cold, dry, and fluctuating climate, i.e., a harsh environment. The phylogenetic correlations among home range, group size, and all bioclimatic variables were further examined. Single-factor PGLS revealed that home range was negatively correlated with mean temperature of the coldest quarter/month (bio11 and bio6). After log-transforming, group size had a positive effect on home range and elevation and was negatively correlated with PCT scores (Figure 5B; Supplementary Table S12). The effects of bio11, bio6, and temperature seasonality (bio4) on home range were statistically significant when group size and body mass were included as covariates (Supplementary Table S12).

Based on the PGLS results, we conducted phylogenetic path analysis to determine the inter-relationships among the factors affecting home range and group size (Supplementary Method S7). Of the 16 candidate models for the eight-variable set, the best-supported model suggested that cold winters and seasonal climatic fluctuations, often associated with high elevation or latitude in montane ecosystems, were predictive of group size via three pathways (Supplementary Table S13; Supplementary Figure S7). The first path predicted that an increase in seasonality would decrease the group size; the second path predicted that low PCT scores would influence large group size directly or increase body mass, leading to an expanded home range, ultimately favoring larger groups; and the third path suggested that cold winters predicted an increase in annual home range and enlargement of group size (Figure 5C).

The three best models from the three sets of variables supported the home range to group size direction (Figure 5C; Supplementary Figures S8, S9 and Table S14), whereas the

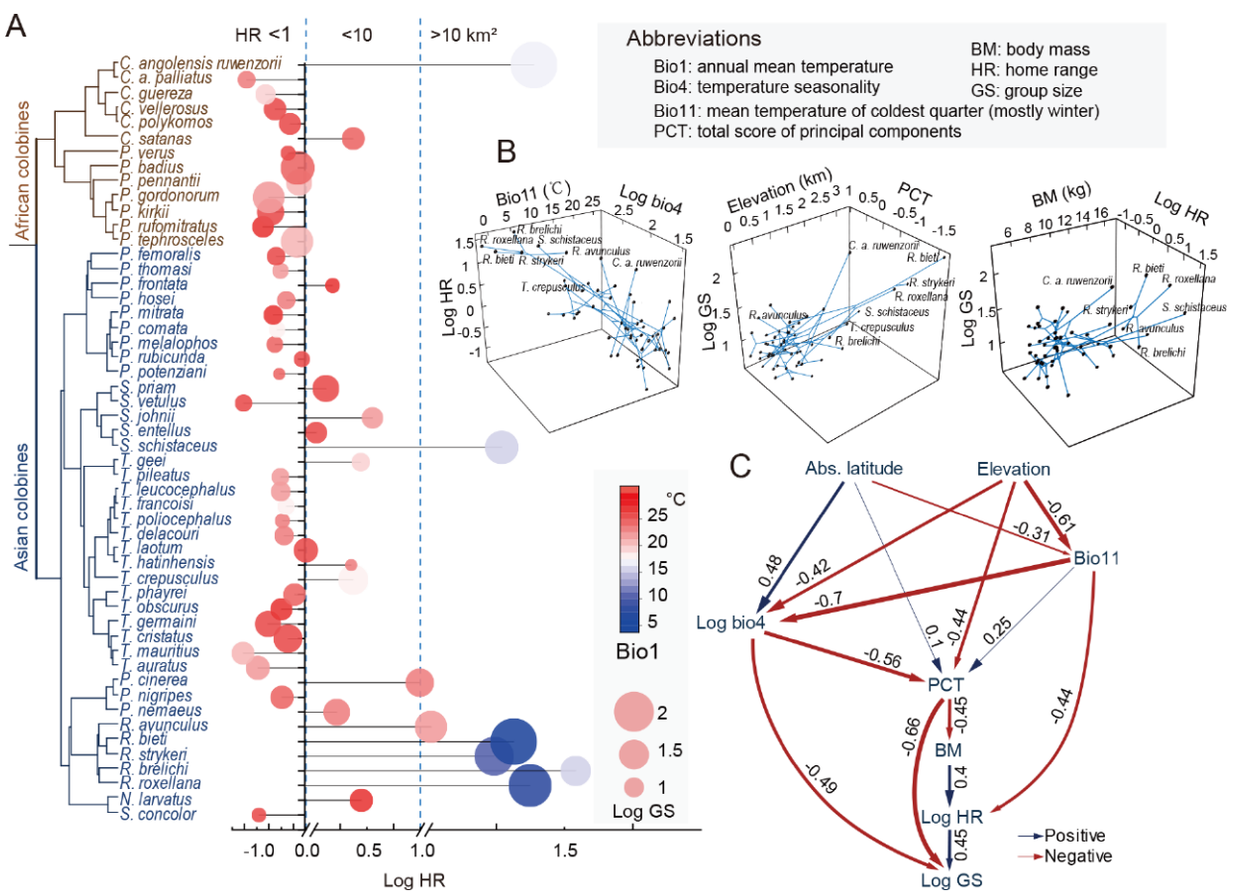


Figure 5 Ecology-associated evolution of home range and group size in colobine primates

A: Pattern of home range (HR) and group size (GS) along colobine phylogeny. B: Three-dimensional phylomorphospace plots of correlations of HR and GS with climatic and topographical factors, as well as with body size. C: Best-supported model with standardized path coefficients obtained from phylogenetic path analysis.

opposite direction was detected in the average model of the six variables (Supplementary Figure S9C). The evolutionary direction of home range to group size was more strongly supported because the standardized regression coefficient of home range to group size was greater than that of group size to home range (Supplementary Figure S9D).

DISCUSSION

Adaptative energetic strategy of golden snub-nosed monkeys to heterogeneous habitat

As elevation and latitude increases, the richness, diversity, and productivity of plant species generally decline due to reductions in temperature (energy) and humidity (water) (Kreft & Jetz, 2007). As such, montane ecosystems at high latitudes likely provide lower food densities and longer periods of resource scarcity for golden snub-nosed monkeys compared to their tropical leaf-eating colobine counterparts. To investigate this, we analyzed the ranging behaviors and food resources in the habitats of golden snub-nosed monkey breeding bands. Our goal was to determine whether this species faces highly heterogeneous resources as predicted by the heterogeneity hypothesis and to understand how they adapt to such environmental conditions.

Although the monkey breeding bands used a total home range of approximately 20 km^2, half of their locations were concentrated within the core area each season, each less than 1 km² in size (Figure 2). Moreover, these core areas were discretely distributed and exhibited seasonal variation within the home range, highlighting spatiotemporal habitat utilization. Quadrat sampling in summer identified higher diversity, evenness, and abundance of main food species in these core areas compared to other areas. In winter, these core areas showed the highest food resource abundance and canopy density (Figures 2, 3), providing protection from predators and cold winds (Davies et al., 2017; Li et al., 2002). These findings confirmed that habitat quality for golden snub-nosed monkeys varied spatially and temporarily in the mountainous forest, supporting the existence of heterogeneous resources.

This study involved tracking the spatiotemporal movements, diets, and habitat characteristics of monkey breeding bands to elucidate their utilization of heterogeneous food resources. Typically, folivores are thought to require more rest for digestion and low-quality fiber fermentation but spend less time traveling because leaves tend to be uniformly distributed and widely abundant. This foraging strategy, feeding on low-quality food and traveling shorter distances to minimize energy expenditure, is referred to as an energy minimization strategy (Pyke et al., 1977) and has been reported in various colobine species such as white-headed langurs (*Trachypithecus leucocephalus*) (Zhang et al., 2021). In contrast, frugivores are predicted to spend more time traveling between food patches in pursuit of nutrient-dense fruits, which tend to occur in discrete clusters. This prediction has led to an energy maximization strategy, in which groups attempt to maximize the amount of energy obtained per unit of feeding time. Our results demonstrated that the golden snub-nosed monkey bands frequently traveled long distances to reach high-quality habitats in summer (Figure 3). This indicates that, although golden snub-nosed monkeys are largely folivorous, they employ the energy maximization strategy due to the spatiotemporal distribution of high-quality food resources in summer. In contrast, home ranges and DTD decreased in winter compared to summer, with DTD showing a positive correlation with thermal energy only in winter and spring, suggesting that the band adjusted its movement in response to climatic variations. During winter, when temperatures dropped and food became scarce, long-distance travel was more energetically costly than in summer. Spatiotemporal eco-behavioral tracking revealed that the band reduced its DTD to minimize energy expenditure, shifted their diet to lower-quality food, and remained in areas with abundant food and high canopy density (Figure 3), indicating a typical energy minimization strategy. Furthermore, during brief temperature rises during winter, the time-series temperature-movement curve demonstrated that the band traveled long distances to access high-energy diets, such as nutrient- and fat-rich oak and hazel seeds, reflecting an energy maximization strategy. Collectively, these results indicate a dynamic foraging strategy, whereby the band alternates between two strategies depending on climatic shifts and varying availability and distribution of different food items.

Previous studies have documented different foraging strategies among mountain-living colobines. Within the same genus as golden snub-nosed monkeys, most species exhibit shorter DTD in winter than in summer, except for the gray snub-nosed monkey (*R. brelichi*) (Supplementary Table S15). Black and white snub-nosed monkeys (*R. bieti*), which are distributed at the highest altitude (2 500–5 000 m a.s.l.), show reduced traveling time and DTD with decreasing temperatures, indicating an energy minimization strategy in winter (Grueter et al., 2013; Ren et al., 2009a, 2009b). However, these monkeys also display patterns of long-distance movement and concentrated use of small areas, indicative of an energy maximization strategy (Grueter et al., 2008). Recent comparative research on activity budgets suggests that while geographic or climatic gradients do not significantly impact colobine activities, *Rhinopithecus* species display a tendency towards energy maximization rather than energy minimization (Kraus & Strier, 2022). In African colobines, the Rwenzori Angolan colobus monkey, which inhabits high-altitude mountainous forests and lives in large multilevel societies (Adams et al., 2021), employs an energy maximization strategy likely shaped clustered resources (Arseneau-Robar et al., 2021; Miller et al., 2020). Thus, mountain-living colobines appear to utilize an energy maximization strategy in addition to the general energy minimization strategy of other colobines, suggesting that the dynamic foraging strategy may be a general pattern among these species. This behavior may offer further insights into how folivorous primates adapt to mountainous climates, whereby the fluctuating climate in montane ecosystems shapes the distribution of resources over time and space, prompting these species to adapt their foraging strategies dynamically, similar to the approach observed in golden snub-nosed monkeys.

By comparing the eco-behaviors of two neighboring bands, we further explored how heterogeneous resources affect the costs and benefits associated with group living. The larger breeding band contained twice as many adult males and three times as many adult females as the smaller band. The greater number of females in the larger band offered more potential reproduction opportunities for adult males, especially the bachelor males. This may have led to the increased size of the GNG all-male band, thereby providing females with additional mating partners outside their primary unit, potentially

improving their fitness through extra-unit mating (Qi et al., 2020). The ecological constraints model proposes that movement patterns and foraging efficiency are key determinants of the food resources available to the group; thus, ecological factors that affect them can also constrain group size (Chapman & Chapman, 2000; Chapman & Teichroeb, 2012). An essential component of the ecological constraints model is that an increase in group size must lead to an increase in within-group feeding competition (Chapman & Teichroeb, 2012), thereby predicting an increase in DTD and home range size as group size increases in order to meet the dietary needs of additional group members (Ganas & Robbins, 2005). Contrary to expectations from the ecological constraints model, our study revealed that the larger band did not occupy a home range significantly larger than that of the smaller band, but rather occupied a smaller home range characterized by higher richness, diversity, and evenness of food resources, especially in winter (Figure 4). As a result, during winter, the large band did not travel significantly longer distances to meet dietary needs compared to the small band, thereby mitigating travel costs. Similar patterns in home range sizes have been reported in other geographically distributed golden snub-nosed monkeys, with group sizes ranging from 60 to 236 but annual home ranges only ranging from 18.3 to 22.5 km^2 (Supplementary Table S16). The deviation of the golden snub-nosed monkey from the ecological constraints model could be attributed to the spatial and temporal heterogeneity of resources in high-latitude mountain ecosystems. The seasonal and uneven distribution of food resources allows larger groups to occupy higher-quality habitats. At the same time, their dynamic foraging strategies enable efficient resource utilization, thus avoiding a sharp increase in DTD. These results imply that, compared to the smaller group, the larger band secured greater reproductive benefits without experiencing a corresponding increase in travel costs to acquire feeding resources in the high-latitude montane ecosystem, supporting the third prediction of the heterogeneity hypothesis.

Social evolution of colobines is associated with environmental heterogeneity and cold temperature

Research has indicated that the complex MLS of snub-nosed monkeys evolved from the social aggregation of independent one-male, multi-female groups (Grueter & Van Schaik, 2010; Qi et al., 2023). In the current study, we compared eco-behaviors among species to explore the impact of environmental heterogeneity on social aggregation. PGLS analysis revealed a positive correlation between group size and home range size in colobines, along with significant correlations with climate variables (Figure 5). Specifically, species residing at higher altitudes with colder temperatures and stronger climatic fluctuations tended to have larger home ranges and group sizes. These findings support the fourth prediction of the heterogeneity hypothesis, suggesting an expansion of home range in heterogeneous environments. Furthermore, PPA confirmed that high-elevation and high-latitude mountain ecosystems resulted in strong temperature seasonality and cold winters, driving group expansion through three distinct pathways.

The first pathway, indicating that a stronger seasonal climate leads to smaller group sizes, aligns with the ecological constraints model. This pathway functions independently of cold temperatures and may explain the social organization of colobines in warmer areas such as tropical rainforests, where plants are abundant and foliage remains available throughout the year, even in seasonal climates. In periods when food availability is reduced due to climate change, groups may mitigate within-group competition by reducing their size, avoiding the need to expand their home range or travel long distances for new food sources. This adaptation may explain the reduction in group size under this pathway, independent of changes in home range size.

The second pathway suggests that the combination of cold and seasonal climates leads to an increase in group size, distinct from the first pathway due to the significant role of cold. Studies have shown that glacial periods in the past six million years imposed intense survival pressures related to cold, influencing the evolution of genes linked to cold-adaptive energy metabolism and neurohormonal regulation in odd-nosed monkeys (Qi et al., 2023). Key among these adaptations are more efficient dopamine and oxytocin pathways, which have been implicated in extending maternal care, enhancing infant survival in colder environments. These genetic changes are thought to strengthen social bonds, increase male-male tolerance, and foster social affiliations, thereby encouraging social aggregation (Qi et al., 2023). This pathway aligns with the concept of cold-driven social aggregation.

The third pathway presents an indirect process wherein cold, dry, and fluctuating climates contribute to an increase in body size and home range, ultimately leading to the growth of group size (Figure 5C). Prior studies have typically considered home range as a consequence of group size, arguing that a greater number of individuals necessitates more resources, leading to a larger home range (Ganas & Robbins, 2005; Grove, 2012). However, this pathway supports the heterogeneity hypothesis, suggesting that expansive home ranges have historically led to the formation of large groups. In challenging environments, where resources are limited and unpredictable, and plants die off completely in winter, maintaining a sufficiently extensive home range to satisfy seasonal food requirements becomes essential for survival. This larger home range, in turn, facilitates more frequent encounters among groups, thereby creating conditions conducive to social aggregation.

Multi-benefits framework of social evolution in colobine primates

We demonstrated that group size in colobine primates is influenced through multiple pathways (Figure 5C; Supplementary Figures S7–S9). Our results indicated that the ecological constraints model, cold-related neurohormonal regulation model, and heterogeneity hypothesis each contribute to understanding the social evolution of colobine primates. Consequently, we integrated these concepts into a comprehensive framework to explain the evolution of social organization in colobine clades. A dual-benefit framework has previously been used to resolve the evolutionary transition from solitary living to cooperative breeding in groups (Shen et al., 2017), identifying two main categories of group benefits: i.e., resource defense (RD) benefits, derived from group-defense critical resources, and collective action (CA) benefits, derived from social cooperation such as foraging, breeding, and predator detection. For colobines, the evolutionary transition from small one-male units to large social bands is more complex. Notably, in addition to the above benefits, the

reproductive benefit (RP) of increased mating opportunities also plays a crucial role, as evidenced by comparisons of reproductive opportunities between the different bands. Furthermore, feeding competition (FC) and travel costs (TC) are critical factors constraining group size (GS). Thus, the net benefits of individuals in a group (G_f) can be simplified as:

$$G_f = \mathrm{RD} + \mathrm{CA} + \mathrm{RP} - \mathrm{FC} - \mathrm{TC} + \varepsilon \quad (1)$$

where the ε is the error term representing the effects that are either not included or are unknown. For a group to be stable, it is essential that the benefits outweigh or balance the costs (Figure 6A). Furthermore, for evolution towards a larger group, the net benefits for individuals in a larger group must be greater than those in a smaller group ($G_{fL}>G_{fS}$).

In diverse environments, the benefits and costs associated with group living vary. In particularly, harsh, cold, and heterogeneous environments are characterized by limited resources per unit area and substantial spatiotemporal variations, making a small home range insufficient for meeting the survival needs of the ancestral OMUs in different seasons. Energy demands during cold winters necessitate access to a variety of high-quality complementary food resources, leading to the requirement of a larger home range during spring (Supplementary Figure S2). In addition, high-quality habitats are not only sparsely distributed but also exhibit variation both within and between seasons. Notably, an ideal summer habitat may be less so in winter, requiring a large home range covering high-quality habitats in different seasons. Within such an extensive home range, low-quality foods, which serve as fallback options, are widely available, and certain high-nutrient foods, like fruits, are abundant in specific seasons or brief periods. Consequently, habitat within this large home range would be underutilized by an ancestral OMU, with its size being smaller than the potential carrying capacity of the environment. In this context, additional individuals joining the group may gain collective action and reproductive benefits without significantly increasing within-group food competition, until reaching the maximum environmental capacity of the home range. In this study, we observed that larger bands of golden snub-nosed monkeys accessed higher-quality habitats and more mating opportunities compared to smaller bands, illustrating the collective action and reproductive benefits of larger groups. Previous studies have also reported the reproductive benefits of OMU aggregation in MLS, whereby breeding males in MLS can more effectively ward off bachelor males through collective defense, creating a safety-in-numbers effect (Grueter & Van Schaik, 2010). This dynamic may result in bachelor males uniting into larger bands to obtain reproductive opportunities, with such an "arms race" between breeding and bachelor males potentially driving evolution towards larger band size for both groups (Grueter & Van Schaik, 2010; Qi et al., 2017). Furthermore, the increased number of breeding and bachelor males provides females with more mating choices, potentially enhancing female fitness through extra-unit mating or selection of alternative partners (Qi et al., 2020). Regarding travel costs, both our intraspecific data, as well as previously reported interspecific data, indicate that DTD does not increase proportionally with group size in cold and heterogeneous environments. For example, the largest band of *R. bieti*, encompassing around 400 individuals, has a reported average DTD of 1 514 m, only slightly more than the 1 068 m traveled by a significantly smaller seven-member group of Thomas's langurs (*Presbytis thomasi*) in a tropical primary rainforest (Grueter and Van Schaik, 2010). Hence, in cold and heterogeneous environments, the increased benefits of collective action and reproduction, coupled with constrained travel costs, likely contribute to the formation of larger groups. This, in turn, raises demand for resources, further facilitating home range enlargement and explaining the correlated evolution of social organization and home range (Supplementary Figure S9).

In tropical rainforests, however, the pattern may differ. In such warm, humid, and stable environments, abundant, evenly distributed, and regular food resources are available for colobines, and thus a small home range is sufficient to meet the year-round resource needs of an independent OMU. Typically, colobine species in tropical rainforests inhabit small home ranges of less than 1 km^2 (Figure 5A). We hypothesize that the evolutionary persistence of the one-male, multi-female social system may be attributed to the complete utilization of the habitat within these small home ranges by ancestral

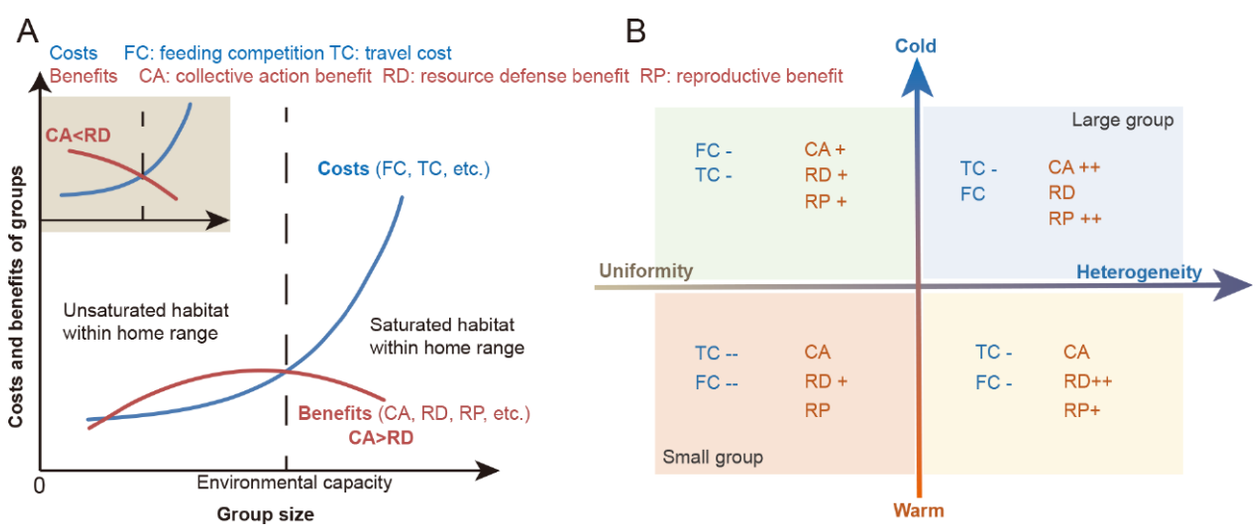


Figure 6 Multi-benefits framework for social evolution in colobines

A: Expected diagram of trade-offs between the costs and benefits of introducing environmental capacity. Expected relationship between group size and resource benefits (RD and CA) was obtained from Shen et al. (2017). B: Concept sketch of costs and benefits per additional capita to initial group in different environment types. Increasing effect is labeled "+" for benefits and "–" for costs; " ++" and "--"indicate stronger effects.

OMUs. In these saturated habitats, the addition of more individuals to an OMU may lead to increased costs rather than benefits, thus favoring autonomous living over integration into larger bands in tropical rainforests (Figure 6A). This hypothesis is supported by evidence from previous studies on population density and travel costs in these regions (Figure 6A). For example, the population density of pig-tailed langurs (*Simias concolor*) distributed in tropical rainforests reaches up to 220 animals/km^2, indicating extremely high habitat saturation (Watanabe, 1981) compared with the 7.2 animals/km^2 for golden snub-nosed monkeys (Tan et al., 2007). Furthermore, a striking contrast in travel costs is observed among different colobine species. For example, *R. bieti*, residing in cold and heterogeneous mountainous ecosystems, shows an increase of only 2 m in DTD for each additional group member, while tropical colobine *P. thomasi*, living in a more benign environment, requires an additional 60 m of travel per added individual (Grueter & Van Schaik, 2010), indicating considerably higher travel costs. This scenario, where high population density results in high feeding competition, may necessitate a strategy focused on resource defense within the group against outsiders (Figure 6B). According to the "insider-outsider conflict theory", when resource defense benefits are the primary driver of grouping, insiders gain higher fitness by cooperating with closely related individuals than by accommodating unrelated outsiders (Shen et al., 2017). This may contribute to the prevalence of colobine groups in tropical rainforests consisting of a male with mates and their offspring, often exhibiting territorial defense between groups.

In environments that are neither exceptionally harsh nor entirely benign, colobine species experience a nuanced balance of benefits and costs. This balance encompasses collective action benefits, resource defense benefits, and reproductive advantages. Such ecological conditions permit the inclusion of both distantly related and unrelated individuals, thereby fostering the development of diverse social group structures among these species (Figure 6B).

Apart from colobines, two other primate species known to form MLS have also adapted to harsh environments (Schreier & Swedell, 2012; Snyder-Mackler et al., 2012). Hamadryas baboons (*Papio hamadryas*) are distributed in sub-desert alpine grasslands (~3 300 m a.s.l.), while geladas (*Theropithecus gelada*) inhabit cold and windy highlands (1 800–4 400 m a.s.l.) (Rowe & Myers, 2017). The occurrence of large social groups in such challenging environments suggests the potential for larger-scale interspecies comparisons in future research to assess the applicability of the multi-benefits framework across different primate species.

CONCLUSIONS

Based on long-term fieldwork with the golden snub-nosed monkey and comparative analysis across colobine species, we proposed a multi-benefits framework integrating the ecological constraints model and heterogeneity hypothesis to explain the social evolution of colobines. The framework highlights that various benefits, including collective action, resource defense, reproduction, and the costs of food competition and travel associated with group living, differ across environments. In warm, tropical rainforests, the high abundance of plant resources available throughout the year enables colobines to satisfy their dietary needs within a relatively small home range. These resource-rich habitats have prompted the adoption of a resource defense strategy, often involving close relatives, resulting in both small home ranges and small family groups. In contrast, during historical glacial periods, ancestral odd-nosed and snub-nosed monkeys, which dispersed to high latitude/altitude montane ecosystems, experienced survival stress due to the cold and dry environments and heterogeneous food resources (Qi et al., 2023). To adapt, these monkeys expanded their home range to accommodate year-round fluctuating resources and developed a dynamic foraging strategy integrating energy maximization to exploit resources and energy minimization to reduce energy expenditure in cold climates. This adaptation allowed large groups to occupy higher-quality habitats and access more mates without a proportional increase in travel costs for foraging, thus providing an ecological basis for social aggregation into larger groups in snub-nosed monkeys from the perspective of resource utilization. Additionally, during periods of intense survival stress from cold conditions, the ancestral odd-nosed and snub-nosed monkeys underwent selection for more efficient dopamine and oxytocin pathways. These pathways are known to facilitate social affiliation behaviors and provide a genetic and physiological basis for social aggregation (Qi et al., 2023). These cold-driven genetic changes and habitat-associated ecological adaptations jointly facilitated social aggregation from independent OMUs to larger breeding bands in high-altitude montane ecosystems. Our study combining fieldwork and comparative analysis distinguished the mixed effects of eco-dependent factors on group-level costs and benefits. Our findings shed light on how various factors have shaped social evolution via multiple pathways in a colobine clade of primates, offering insights into the complex and diverse social structures observed across a broad range of animal taxa.

SCIENTIFIC FIELD SURVEY PERMISSION INFORMATION

Permission for field survey in Mts. Qinling was granted by the Shaanxi Forestry Bureau.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

X.G.Q. conceived and designed the study. L.Z. and S.N.J. compiled the dataset. L.Z. analyzed the data. X.B.D, J.H.L, B.L.Z., and P.H.L. performed experiments. L.Z., S.N.J., J.H.L., Y.J.Y., B.G.L., Y.Q.G., and X.G.Q. wrote and revised the manuscript. All authors added materials, helped to revise the manuscript, and approved the final manuscript.

ACKNOWLEDGMENTS

We greatly thank the Zhouzhi National Nature Reserve for permission to carry out this study. We appreciate Qi Li, Kun Bian, and Zhi-Pang Huang for their assistance with data collection. We thank Qi Li and Dr. Dunn Derek for their contributions to the early draft of the manuscript. We thank photographer Kuan-Liang Ding for permission to use photographs, as well as Stephen D. Nash for permission to use illustrations. We would like to thank Editage (www.editage.cn) for English language editing. We thank the associate editors and the reviewers for their helpful feedback that improved this paper.

REFERENCES

Adams FV, Arseneau-Robar TJM, Bonnell TR, et al. 2021. Temporal

patterns in the social network of core units in Rwenzori Angolan colobus monkeys: Effects of food availability and interunit dispersal. *Ecology and Evolution*, **11**(7): 3251−3263.

Arseneau-Robar TJM, Changasi AH, Turner E, et al. 2021. Diet and activity budget in *Colobus angolensis ruwenzorii* at nabugabo, uganda: are they energy maximizers?. *Folia Primatologica*, **92**(1): 35−48.

Beyer DE JR, Haufler JB. 1994. Diurnal versus 24-hour sampling of habitat use. *The Journal of Wildlife Management*, **58**(1): 178−180.

Bjørneraas K, Van Moorter B, Rolandsen CM, et al. 2010. Screening global positioning system location data for errors using animal movement characteristics. *The Journal of Wildlife Management*, **74**(6): 1361−1366.

Chapman CA, Chapman LJ. 2000. Determinants of group size in primates: the importance of travel costs. *In*: Boinski S, Garber PA. On the Move: How and Why Animals Travel in Groups. Chicago: University of Chicago Press, 24–41.

Chapman CA, Teichroeb JA. 2012. What influences the size of groups in which primates choose to live?. *Nature Education Knowledge*, **3**(10): 9.

Clutton-Brock T. 2021. Social evolution in mammals. *Science*, **373**(6561): eabc9699.

Clutton-Brock T, Janson C. 2012. Primate socioecology at the crossroads: Past, present, and future. *Evolutionary Anthropology:Issues, News, and Reviews*, **21**(4): 136−150.

Davies AB, Ancrenaz M, Oram F, et al. 2017. Canopy structure drives orangutan habitat selection in disturbed Bornean forests. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(31): 8307−8312.

Fan PL, Li YM, Stanford CB, et al. 2019. Home range variation of two different-sized groups of golden snub-nosed monkeys (*Rhinopithecus roxellana*) in Shennongjia, China: implications for feeding competition. *Zoological Research*, **40**(2): 121−128.

Fang G, Chen J, Pan RL, et al. 2018a. Female choice impacts residential male takeover in golden snub-nosed monkeys (*Rhinopithecus roxellana*). *Zoological Research*, **39**(4): 266−271.

Fang G, Li M, Liu XJ, et al. 2018b. Preliminary report on Sichuan golden snub-nosed monkeys (*Rhinopithecus roxellana roxellana*) at Laohegou Nature Reserve, Sichuan, China. *Scientific Reports*, **8**(1): 16183.

Ganas J, Robbins MM. 2005. Ranging behavior of the mountain gorillas (*Gorilla beringei beringei*) in Bwindi Impenetrable National Park, Uganda: a test of the ecological constraints model. *Behavioral Ecology and Sociobiology*, **58**(3): 277−288.

Garcia C, Bouret S, Druelle F, et al. 2021. Balancing costs and benefits in primates: ecological and palaeoanthropological views. *Philosophical Transactions of the Royal Society B Biological Sciences*, **376**(1819): 20190667.

Grove M. 2012. Space, time, and group size: a model of constraints on primate social foraging. *Animal Behaviour*, **83**(2): 411−419.

Grow NB, Gursky-Doyen S, Krzton A. 2014. High Altitude Primates. New York: Springer.

Grueter CC. 2022. Causes and consequences of the formation of Multilevel Societies in Colobines. *In*: Matsuda I, Grueter CC, Teichroeb JA. The Colobines: Natural History, Behaviour and Ecological Diversity. Cambridge: Cambridge University Press, 293–311.

Grueter CC, Li DY, Ren BP, et al. 2013. Overwintering strategy of Yunnan snub-nosed monkeys: adjustments in activity scheduling and foraging patterns. *Primates*, **54**(2): 125−135.

Grueter CC, Li DY, Van Schaik CP, et al. 2008. Ranging of *Rhinopithecus bieti* in the samage forest, China. I. Characteristics of range use. *International Journal of Primatology*, **29**(5): 1121−1145.

Grueter CC, Qi XG, Li BG, et al. 2017. Multilevel societies. *Current Biology*, **27**(18): R984−R986.

Grueter CC, Qi XG, Zinner D, et al. 2020. Multilevel organisation of animal sociality. *Trends in Ecology & Evolution*, **35**(9): 834−847.

Grueter CC, Van Schaik CP. 2010. Evolutionary determinants of modular societies in colobines. *Behavioral Ecology*, **21**(1): 63−71.

Hemingway CA, Bynum N. 2005. The influence of seasonality on primate diet and ranging. *In*: Brockman DK, Van Schaik CP. Seasonality in Primates: Studies of Living and Extinct Human and Non-Human Primates. Cambridge: Cambridge University Press, 57–104.

Huang PZ, Bian K, Huang ZP, et al. 2021. Human activities and elevational constraints restrict ranging patterns of snub-nosed monkeys in a mountainous refuge. *Integrative Zoology*, **16**(2): 202−213.

Huang ZP. 2015. Ecological Mechanism of Fission-Fusion of Golden Snub-Nosed Monkey (*Rhinopithecus roxellana*) in Qingling Mountains. Ph. D. dissertation, Northwest University, Xi'an.

Isbell LA. 1991. Contest and scramble competition: patterns of female aggression and ranging behavior among primates. *Behavioral Ecology*, **2**(2): 143−155.

Isbell LA, Young TP. 2002. Ecological models of female social relationships in primates: similarities, disparities, and some directions for future clarity. *Behaviour*, **139**(2-3): 177−202.

Janson CH, Goldsmith ML. 1995. Predicting group size in primates: foraging costs and predation risks. *Behavioral Ecology*, **6**(3): 326−336.

Kappeler PM. 2019. A framework for studying social complexity. *Behavioral Ecology and Sociobiology*, **73**(1): 13.

Kappeler PM, Clutton-Brock T, Shultz S, et al. 2019. Social complexity: patterns, processes, and evolution. *Behavioral Ecology and Sociobiology*, **73**(1): 5.

Kraus JB, Strier KB. 2022. Geographic, climatic, and phylogenetic drivers of variation in colobine activity budgets. *Primates*, **63**(6): 647−658.

Kreft H, Jetz W. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(14): 5925−5930.

Li BG, Chen C, Ji WH, et al. 2000. Seasonal home range changes of the Sichuan snub-nosed monkey (*Rhinopithecus roxellana*) in the Qinling Mountains of China. *Folia Primatologica*, **71**(6): 375−386.

Li BG, Pan RL, Oxnard CE. 2002. Extinction of Snub-Nosed Monkeys in China During the Past 400 Years. *International Journal of Primatology*, **23**(6): 1227−1244.

Li BG, Ren BP, Gao YF. 1999. A change in the summer home range of Sichuan snub-nosed monkeys in Yuhuangmiao, Qinling mountains. *Folia Primatologica*, **70**(5): 269−273.

Li YL, Wang L, Wu JW, et al. 2020. Bachelor groups in primate multilevel society facilitate gene flow across fragmented habitats. *Current Zoology*, **66**(2): 113−122.

Li YM. 2006. Seasonal variation of diet and food availability in a group of Sichuan snub-nosed monkeys in Shennongjia Nature Reserve, China. *American Journal of Primatology*, **68**(3): 217−233.

Markham AC, Gesquiere LR. 2017. Costs and benefits of group living in primates: an energetic perspective. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **372**(1727): 20160239.

Miller A, Judge D, Uwingeneye G, et al. 2020. Diet and use of fallback foods by rwenzori black-and-white colobus (*Colobus angolensis ruwenzorii*) in rwanda: implications for supergroup formation. *International Journal of Primatology*, **41**(3): 434−457.

Pyke GH, Pulliam HR, Charnov EL. 1977. Optimal foraging: A selective review of theory and tests. *The Quarterly Review of Biology*, **52**(2): 137−154.

Qi XG, Garber PA, Ji WH, et al. 2014. Satellite telemetry and social modeling offer new insights into the origin of primate multilevel societies. *Nature Communications*, **5**(1): 5296.

Qi XG, Grueter CC, Fang G, et al. 2020. Multilevel societies facilitate infanticide avoidance through increased extrapair matings. *Animal*

Behaviour, **161**: 127−137.
Qi XG, Huang K, Fang G, et al. 2017. Male cooperation for breeding opportunities contributes to the evolution of multilevel societies. *Proceedings of the Royal Society B:Biological Sciences*, **284**(1863): 20171480.
Qi XG, Li BG, Garber PA, et al. 2009. Social dynamics of the golden snub-nosed monkey (*Rhinopithecus roxellana*): female transfer and one-male unit succession. *American Journal of Primatology*, **71**(8): 670−679.
Qi XG, Li BG, Ji WH. 2008. Reproductive parameters of wild female *Rhinopithecus roxellana*. *American Journal of Primatology*, **70**(4): 311–319.
Qi XG, Wu JW, Zhao L, et al. 2023. Adaptations to a cold climate promoted social evolution in Asian colobine primates. *Science*, **380**(6648): eabl8621.
R Core Team. 2022. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Ren BP, Li M, Long YC, et al. 2009a. Influence of day length, ambient temperature, and seasonality on daily travel distance in the Yunnan snub-nosed monkey at Jinsichang, Yunnan, China. *American Journal of Primatology*, **71**(3): 233−241.
Ren BP, Li M, Long YC, et al. 2009b. Home range and seasonality of Yunnan snub-nosed monkeys. *Integrative Zoology*, **4**(2): 162−171.
Rowe N, Myers M. 2017. All the World's Primates. Charlestown: Primate Conservation.
Schreier AL, Swedell L. 2012. The socioecology of network scaling ratios in the multilevel society of hamadryas baboons (*Papio hamadryas hamadryas*). *International Journal of Primatology*, **33**(5): 1069−1080.
Shen SF, Emlen ST, Koenig WD, et al. 2017. The ecology of cooperative breeding behaviour. *Ecology Letters*, **20**(6): 708−720.
Snaith TV, Chapman CA. 2005. Towards an ecological solution to the folivore paradox: patch depletion as an indicator of within-group scramble competition in red colobus monkeys (*Piliocolobus tephrosceles*). *Behavioral Ecology and Sociobiology*, **59**(2): 185−190.
Snyder-Mackler N, Beehner JC, Bergman TJ. 2012. Defining higher levels in the multilevel societies of geladas (*Theropithecus gelada*). *International Journal of Primatology*, **33**(5): 1054−1068.
Steenbeek R, Van Schaik CP. 2001. Competition and group size in Thomas's langurs (*Presbytis thomasi*): the folivore paradox revisited. *Behavioral Ecology and Sociobiology*, **49**(2-3): 100−110.
Sterck EHM, Roth TS. 2022. Socioecology of asian colobines. *In*: Matsuda I, Grueter CC, Teichroeb JA. The Colobines: Natural History, Behaviour and Ecological Diversity. Cambridge: Cambridge University Press, 250–270.
Symonds MRE, Blomberg SP. 2014. A primer on phylogenetic generalised least squares. *In*: Garamszegi LZ. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Berlin: Springer, 105–130.
Tan CL, Guo ST, Li BG. 2007. Population structure and ranging patterns of *Rhinopithecus roxellana* in Zhouzhi national nature reserve, Shaanxi, China. *International Journal of Primatology*, **28**(3): 577−591.
Tan CL, Li YM, Li BG. 2016. *Rhinopithecus roxellana*. *In*: Rowe N and Myers M. All the World's Primates. Charlestown, Rhode Island: Pogonias Press, 625–626.
Teichroeb JA, Sicotte P. 2009. Test of the ecological-constraints model on ursine colobus monkeys (*Colobus vellerosus*) in Ghana. *American Journal of Primatology*, **71**(1): 49−59.
Tomkiewicz SM, Fuller MR, Kie JG, et al. 2010. Global positioning system and associated technologies in animal behaviour and ecological research. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **365**(1550): 2163−2176.
Van Der Bijl W. 2018. *phylopath*: Easy phylogenetic path analysis in R. *PeerJ*, **6**: e4718.
Van Moorter B, Rolandsen CM, Basille M, et al. 2016. Movement is the glue connecting home ranges and habitat selection. *Journal of Animal Ecology*, **85**(1): 21−31.
Watanabe K. 1981. Variations in group composition and population density of the two sympatric Mentawaian leaf-monkeys. *Primates*, **22**(2): 145−160.
Yu Y, He G, Li DY, et al. 2022. Climate change challenge, extinction risk, and successful conservation experiences for a threatened primate species in China: golden snub-nosed monkey (*Rhinopithecus roxellana*). *Zoological Research*, **43**(6): 940−944.
Zhang KC, Zhou QH, Xu HL, et al. 2021. Diet, food availability, and climatic factors drive ranging behavior in white-headed langurs in the limestone forests of Guangxi, southwest China. *Zoological Research*, **42**(4): 406−411.

Xiao et al. *Zool. Res.* 2024, 45(1): 55−68
https://doi.org/10.24272/j.issn.2095-8137.2023.007

ZR | Zoological Research

Article Open Access

Single-cell profiling of the pig cecum at various developmental stages

Yan-Yuan Xiao[1], Qing Zhang[1], Fei Huang[1], Lin Rao[1], Tian-Xiong Yao[1], Si-Yu Yang[1], Lei Xie[1], Xiao-Xiao Zou[1], Li-Ping Cai[1], Jia-Wen Yang[1], Bin Yang[1,*], Lu-Sheng Huang[1,*]

[1] *National Key Laboratory for Swine Genetic Improvement and Germplasm Innovation*, *Ministry of Science and Technology of China*, *Jiangxi Agricultural University*, *Nanchang*, *Jiangxi* 330045, *China*

ABSTRACT

The gastrointestinal tract is essential for food digestion, nutrient absorption, waste elimination, and microbial defense. Single-cell transcriptome profiling of the intestinal tract has greatly enriched our understanding of cellular diversity, functional heterogeneity, and their importance in intestinal tract development and disease. Although such profiling has been extensively conducted in humans and mice, the single-cell gene expression landscape of the pig cecum remains unexplored. Here, single-cell RNA sequencing was performed on 45 572 cells obtained from seven cecal samples in pigs at four different developmental stages (days (D) 30, 42, 150, and 730). Analysis revealed 12 major cell types and 38 subtypes, as well as their distinctive genes, transcription factors, and regulons, many of which were conserved in humans. An increase in the relative proportions of $CD8^+$ T and Granzyme A (low expression) natural killer T cells ($GZMA^{low}$ NKT) cells and a decrease in the relative proportions of epithelial stem cells, Tregs, $RHEX^+$ T cells, and plasmacytoid dendritic cells (pDCs) were noted across the developmental stages. Moreover, the post-weaning period exhibited an up-regulation in mitochondrial genes, *COX2* and *ND2*, as well as genes involved in immune activation in multiple cell types. Cell-cell crosstalk analysis indicated that $IBP6^+$ fibroblasts were the main signal senders at D30, whereas $IBP6^-$ fibroblasts assumed this role at the other stages. NKT cells established interactions with epithelial cells and $IBP6^+$ fibroblasts in the D730 cecum through mediation of *GZMA-F2RL1/F2RL2* pairs. This study provides valuable insights into cellular heterogeneity and function in the pig cecum at different development stages.

Keywords: Single-cell RNA-seq; Cecum; Bama Xiang pigs; Various developmental stages; Cellular heterogeneity

INTRODUCTION

Pigs are one of the most important farm animals, providing approximately 40% of global red meat consumption (Gerrits et al., 2005). Additionally, they are valuable biomedical models due to their anatomical, structural, immunological, physiological, and genomic similarities to humans (Lunney et al., 2021). The gastrointestinal tract, essential for food digestion and waste expulsion (Liao et al., 2009; Zorn & Wells, 2009), is also the largest immune organ in the body, accommodating trillions of microorganisms (Chassaing et al., 2014; James, 1993). The cecum, a pouch or large tubelike structure situated immediately after the ileocecal junction, demarcates the transition from the small to large intestine. Lying between the non-fermentative and fermentative parts of the gut, the cecum serves as a fermentation chamber, facilitating the conversion of indigestible fibers into energy-rich compounds (Stumpff et al., 2019) and plays an important role in electrolyte and water absorption (Dabareiner & White II, 1997). Cecum removal has been shown to severely affect short-chain fatty acid production and reduce resistance to *Salmonella enterica* serovar Enteritidis infection in mice (Brown et al., 2018). Proper gut functionality requires the cooperation of multiple cell lineages. Single-cell RNA sequencing (scRNA-seq) has enabled detailed gene expression profiling of tissues and organs at the single-cell level (Han et al., 2018, 2020, 2022; Li et al., 2022). Notably, scRNA-seq has been extensively applied in human and mouse studies to profile the intestinal tract (Elmentaite et al., 2021; Fawkner-Corbett et al., 2021; Wang et al., 2020), shedding light on cell types and genes associated with intestinal tract development and diseases. However, a comprehensive examination of the single-cell gene expression profiles of the pig cecum is yet to be conducted.

The weaning process represents an important stage in pig farming. During weaning, piglets transition from a diet of highly

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 16 March 2023; Accepted: 12 September 2023; Online: 13 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (31790410, 32160781)

*Corresponding authors, E-mail: binyang@live.cn; lushenghuang@hotmail.com

digestible and palatable liquid milk to less digestible and palatable solid dry feed (Campbell et al., 2013). This abrupt dietary shift can induce weaning stress, potentially leading to intestinal dysfunction, reduced growth, compromised health, and reduced feed intake, particularly during the first week of weaning (Campbell et al., 2013; Tang et al., 2022). However, what effects the weaning process has on the gut at the single-cell level remain poorly understood.

In this study, scRNA-seq was conducted on seven cecal samples obtained from pigs aged 30, 42, 150, and 730 days, yielding 45 572 cells after quality control. We identified a total of 12 major cell types and 38 subtypes, including major intestinal epithelial cells, such as goblet cells, enterocytes, and stem cells, and various lymphocyte and myeloid immune cell lineages, and characterized cell-type-specific marker genes, transcription factors (TFs), and regulons. We observed substantial cellular heterogeneity in the cecum across the four developmental stages and identified dynamic changes in gene expression in different cell types during weaning. Furthermore, we also revealed cell-cell interaction heterogeneity across the different developmental stages, as well as cell-type conservation in cecal tissue between pigs and humans. These findings enhance our understandings of cellular diversity across different developmental stages in the pig cecum.

MATERIALS AND METHODS

Pig cecal tissue collection and ethics statement

Cecal samples were obtained from seven healthy Bama Xiang pigs at four different developmental stages: D30 (30 days after birth, five days before weaning), D42 (42 days after birth, seven days after weaning), D150 (150 days after birth), and D730 (730 days after birth). The cecal samples were collected from one boar at D150 and from three pairs of full siblings (both sexes) at the other three developmental stages, with a single sample per individual (Supplementary Table S1). All samples were consistently harvested from the bottom region

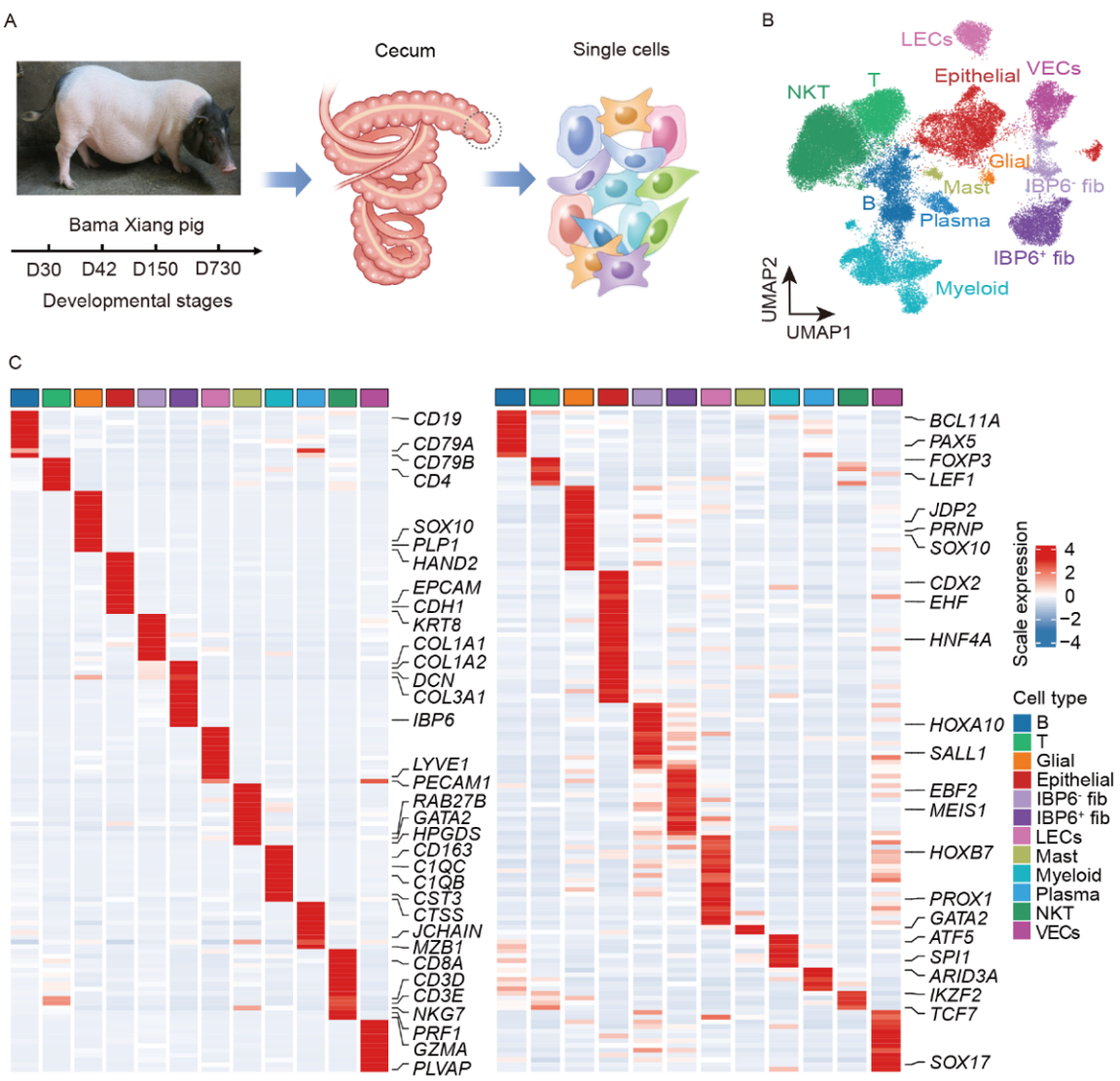


Figure 1 ScRNA-seq profiling of pig cecum at four developmental stages

A: Experimental schematic. Circled section in middle figure is the specific sampling site. D30: 30 days after birth (five days before weaning); D42: 42 days after birth (seven days after weaning); D150: 150 days after birth; D730: 730 days after birth. B: UMAP of 45 572 cells colored by major cell types. B: B cells, T: T cells, Glial: enteric glial cells, Epithelial: epithelial cells, IBP6$^-$ fib: IBP6$^-$ fibroblasts; IBP6$^+$ fib: IBP6$^+$ fibroblasts; LECs: lymphatic endothelial cells, Mast: mast cells, Myeloid: myeloid cells, Plasma: plasma cells, NKT: natural killer T cells, VECs: vascular endothelial cells. C: Heatmap showing highly expressed genes in cell types (left) and cell-type-specific TFs (right).

of the cecum (Figure 1A). Pigs were fed their mother's milk before weaning (D30), then a solid dry feed diet (corn-soyabean-based diet composed of 15%–18% crude protein, ≥ 1.05% lysine, 0.40%–1.20% calcium, and ≥0.40% phosphorus) after weaning. All animal experiments were approved by the Animal Ethics Committee of Jiangxi Agricultural University.

Cell suspension for scRNA-seq

Cecal tissue was harvested and washed in ice-cold RPMI 1640. Tissue slices (0.4 mm^2) were then dissociated using a Multi-Tissue Dissociation Kit 2 (Miltenyi Biotec, Germany) as per the manufacturer's instructions. After DNase treatment, fresh cells were washed twice in RPMI 1640 and resuspended at 1×10^6 cells/mL in 1×phosphate buffered saline (PBS) and 0.04% bovine serum albumin.

Library construction and sequencing

ScRNA-seq libraries were prepared using a SeekOne® MM Single-Cell 3' Library Preparation Kit. The PE150 libraries were sequenced using the Illumina NovaSeq 6000 platform (USA).

ScRNA-seq data analysis

The raw scRNA-seq reads were first trimmed using Fastp (v.0.20.1) (Chen et al., 2018) to obtain clean reads. The specific parameters were: --cut_front --cut_front_Window_size 4 --cut_front_mean_quality 10, --cut_tail --cut_tail_window_size 1 --cut_tail_mean_quality 3, detect_adapter_for_pe, --length_required 60. The obtained clean reads were processed using the Seekone Tools pipeline (https://www.seekgene.com/fxrj) to generate a transcript expression matrix. A counting matrix was then imported into the Python package Scanpy (v.1.7.1). Quality control procedures were performed on each sample: (1) removal of genes expressed in fewer than three cells and (2) exclusion of cells containing less than 300–450 genes, more than 2 500–4 000 genes, more than 13 000–20 000 unique molecular identifier (UMI) counts, or mitochondrial read fraction of more than 40%–45%. Furthermore, doublet exclusion was performed using Scrublet (v.0.2.3) with a cut-off score of 0.25 and expected doublet rate of 0.0225–0.0445 depending on the number of captured cells.

Library size was normalized and the effects of UMI counts and percentage of mitochondrial reads on gene expression were regressed out. Batch corrections of the seven cecal samples were performed using Harmony (Korsunsky et al., 2019). Dimensionality reduction and Louvain clustering using the top 2 000 highly variable genes were carried out with default parameters. Differential expression analysis was performed using the *sc.tl.rank_genes_groups* (method= "Wilcoxon") function in Scanpy.

Cell-type-specific TFs

The list of TFs was downloaded from the SCENIC GitHub repository (Van de Sande et al., 2020). The Tau value, frequently used to measure tissue or cell-type-specific gene expression (Li et al., 2022; Zhu et al., 2022), was calculated using the following formula (Kryuchkova-Mostacci & Robinson-Rechavi, 2017):

$$Tau = \frac{\sum_{i=1}^{n}(1-\hat{x}_i)}{n-1}; \hat{x}_i = \frac{x_i}{\max\limits_{1\le i\le n}(x_i)} \quad (1)$$

where x_i represents the average expression value of a gene across all cells in the ith cell type out of the n cell types. Cell-type-specific TFs were identified based on a tau value of >0.85.

Gene regulatory network inference

The gene regulatory network was inferred using pySCENIC (Python-based version of SCENIC, v.0.11.2). The list of TFs was retrieved from the SCENIC GitHub repository (Van de Sande et al., 2020). Gene regulatory network inference using pySCENIC consisted of three steps. First, co-expression analysis between a predefined list of TFs and putative target genes was carried out to construct candidate regulatory modules. Next, motif enrichment analysis was performed to refine co-expression modules using human TF motif information. Finally, the activities of the discovered regulons in each individual cell were measured. To further reveal cell-type-specific regulons, differential expression analysis was performed on the activity matrix of regulons using the *FindAllMarkers* function in the Seurat package. Only those with Benjamini-Hochberg adjusted $P<0.01$ and avg_$\log_2$fold-change (FC)>0.12 were defined as cell-type-specific regulons.

Gene Ontology (GO) enrichment analysis

The clusterProfiler package (v.3.18.1) in R and ClueGO plugin (v.2.5.7) in Cytoscape were used to perform GO enrichment analysis. Significantly enriched terms were determined based on a Benjamini-Hochberg corrected *P*-value cutoff of <0.05.

Inference of differentiation trajectory

Trajectory inference was performed using PAGA (v.1.2) with a k-nearest neighbor graph of 10 neighbors based on the top 40 principal components. The connectivity threshold between cell types was set to 0.08 to filter spurious connections.

Transcriptomic comparative analysis before and after weaning

Differential expression analysis was performed in each cell type before and after weaning. Differentially expressed genes (DEGs) were calculated using the *FindMarkers* function in Seurat and defined by Benjamini-Hochberg adjusted $P<0.05$ and $|\text{avg_log}_2\text{FC}|>0.25$.

Cell-cell communication analysis

Cell-cell communication analysis was performed using CellChat (v.1.1.3) to infer the interactions between different cell types. Ligand-receptor pairs were obtained from the receptor-ligands library of humans. Normalized counts were used as input data to CellChat and followed an analysis workflow with default parameters.

Cross-species analysis

Cross-species comparative analysis was performed to explore conservation between human and pig ceca. First, single-cell transcriptome data of the human cecum were downloaded from the Gut Cell Atlas (https://www.gutcellatlas.org/), with single-cell data of postnatal and healthy people extracted for downstream analysis. A total of 45 572 cells from the pig cecum and 20 693 cells from the human cecum were integrated using various functions in the Seurat package based on homologous genes in the two species. The top 2 000 highly variable genes in the two species were identified using the *FindVariableFeatures* function. Their intersection was used to find anchors using the *FindIntegrationAnchors* function. The two datasets were then integrated using the *IntegrateData* function. To further reveal conserved genes between the pig and human ceca, signature genes were calculated for each cell type in pigs and humans separately

using the *FindAllMarker* function in Seurat with default parameters. In both species, signature genes were defined with Benjamini-Hochberg adjusted $P<0.05$ and avg_$\log_2$FC>0.25. Finally, the signature genes of the two species were intersected to obtain cell type-conserved genes.

RESULTS

Integrated single-cell map of developing pig cecum

We profiled single-cell transcriptomes of the pig cecum from seven individuals at four different developmental stages (Figure 1A; Supplementary Figure S1A). After quality control, we obtained 45 572 cells, which were integrated using Harmony (Korsunsky et al., 2019) and clustered into 12 major cell types (Figure 1B, C), including epithelial cells (*EPCAM*, *CDH11*, and *KRT8*), B cells (*CD19*, *CD79A*, and *CD79B*), plasma cells (*JCHAIN*, *XBP1*, and *MZB1*), myeloid cells (*CD163*, *C1QC*, and *CST3*), mast cells (*RAB27B*, *GATA2*, and *HPGDS*), T cells (*CD4*, *CD3D*, and *CD3E*), NKT cells (*CD3D*, *CD3E*, *NKG7*, and *PRF1*), enteric glial cells (*SOX10*, *PLP1*, and *HAND2*), IBP6$^-$ fibroblasts (*COL1A1*, *COL1A2*, and *DCN*), IBP6$^+$ fibroblasts (*COL1A1*, *COL1A2*, *DCN*, and *IBP6*), lymphatic endothelial cells (*PECAM1* and *LYVE1*), and vascular endothelial cells (*PECAM1* and *PLVAP*), comparable to the cell types reported in humans and mice (Elmentaite et al., 2021; Han et al., 2018, 2020; James et al., 2020; Smillie et al., 2019). Most cell types were robustly identified in all samples (Supplementary Figure S1B, C), indicating minimal impact of samples on cell-type discovery. Nevertheless, we observed substantial heterogeneity in cell composition in the cecum at different developmental stages, including an expansion of immune cells from pre-weaning (D30) to post-weaning (D42) (Supplementary Figure S1C), which may be associated with the dramatic changes in diet and ambient environment between the two stages.

At a tau value threshold of >0.85, we identified 140 TFs displaying cell-type-specific expression (Figure 1C), including *BCL11A* and *PAX5* in B cells, *FOXP3* and *LEF1* in T cells, *JDP2*, *PRNP*, and *SOX10* in enteric glial cells, *CDX2*, *EHF*, and *HNF4A* in epithelial cells, and *SOX17* in VECs. Among the identified TFs, many have been implicated in corresponding cell types in other species, e.g., *HNF4A* (Boyd et al., 2009; Tunçer et al., 2020), *SOX17* (Yao et al., 2019), and *PAX5* (Cobaleda et al., 2007; Medvedovic et al., 2011) are essential for intestinal epithelial cell, VECs, and B cell differentiation, respectively. Furthermore, several TFs displayed age-dependent expression patterns (Supplementary Figure S2A). For example, all 10 B cell-specific TFs (*BACH2*, *BCL11A*, *E2F2*, *E2F8*, *MEF2B*, *MYBL1*, *MYBL2*, *PAX5*, *PAU2F3*, and *ZNF70*) showed higher expression in the adult stages (D150 and D730), reflecting greater activity of B cells at adults than during the weaning stages; the myeloid cell-specific expression of *ARG1*, a well-characterized marker of immunosuppressive myeloid cells (Rodriguez et al., 2017; Su et al., 2021), was higher at the D30 stage and showed reduced expression with development, suggesting immunity of myeloid cells at the pre-weaning stage was more suppressed.

To further explore the key driver of cell fate and regulator of cellular functions, SCENIC (Aibar et al., 2017; Van de Sande et al., 2020) was used to infer the TF regulatory networks determined by co-expression and motif enrichment analysis. In total, 305 regulons were identified, 22 of which were cell-type-specific (Supplementary Figure S1D), including FOXP3 and GFI1 in T cells, SOX6 and SOX10 in enteric glial cells, ISX and CDX2 in epithelial cells, and SOX17 in VECs. Among these, *GFI1* is reported to regulate intestinal secretory cell differentiation and allocation, e.g., *GFI1*$^{-/-}$ mice intestines show excess enteroendocrine cells, fewer goblet cells, and no Paneth cells (Shroyer et al., 2005). *ISX* is associated with the maintenance of vitamin A metabolism (Seino et al., 2008) and controls intestinal β,β-carotene absorption (Lobo et al., 2010) in mice. These results provide a basis for future studies on the mechanisms of cell differentiation and function in the pig cecum.

Epithelial cell heterogeneity in pig cecum

Epithelial cells play important roles in digestion, absorption, and microbial interactions (Ali et al., 2020; Kong et al., 2018). Here, based on sub-clustering analysis, seven cecal epithelial subtypes were obtained, including stem cells (*OLFM4* and *SMOC2*), goblet cells (*MUC2*, *SPINK4*, and *FCGBP*), immature enterocytes (*APOBEC1* and *CA1*), mature enterocytes (*APOBEC1*, *CA1*, *SLC26A3*, and *AQP8*), BEST4$^+$ epithelial cells (*CA7*, *BEST4*, and *OTOP2*), enterocyte progenitor cells (*OLFM4*, *APOBEC1*, and *CA1*), and COL3A1$^+$ epithelial cells (*EPCAM1* and *COL3A1*) (Figure 2A, B). UMAP analysis indicated that the goblet cells were separated from other cells (Figure 2A). Partition-based graph abstraction (PAGA) analysis (Wolf et al., 2019) revealed two differentiation trajectories from stem cells via enterocyte progenitor cells to mature enterocytes and BEST4$^+$ epithelial cells, respectively (Figure 2C).

Intestinal epithelial stem cells located in crypts are responsible for the continuous renewal of the intestinal epithelial cells (Yang et al., 2021). Here, stem cells were characterized by high expression of *OLFM4* (Figure 2B), consistent with previous human colonic study (van der Flier et al., 2009). The top 200 DEGs included 50 ribosomal genes (Supplementary Table S2), consistent with previous findings in human intestinal stem cells (Burclaff et al., 2022), reflecting a state of active gene translation in intestinal epithelial stem cells.

Enterocytes perform passive barrier function and participate in digestive processes (Snoeck et al., 2005). In the current study, three enterocyte subtypes were identified, i.e., enterocyte progenitor cells, immature enterocytes, and mature enterocytes (Figure 2A, B). The enterocyte progenitor cells expressed both enterocyte marker genes *APOBEC1* and *CA1* and intestinal stem cell marker gene *OLFM4*. Immature enterocytes expressed *APOBEC1* and *CA1* but lacked *AQP8* (a water channel protein). Mature enterocytes were identified based on the expression of *APOBEC1*, *CA1*, and *AQP8*. Solute carrier (*SLC*) family genes are critical for the transport of nutrients and metabolites, such as amino acids, organic solutes, and glucose (Lin et al., 2015; Wang et al., 2020). Here, the expression of *SLC26A3* (chloride ions transporter) and *SLC16A1* (organic solute transporter) gradually increased with differentiation trajectory toward mature enterocytes (Figure 2B), reflecting gradual enhancement of transporter function in the enterocytes. The mature enterocytes exhibited a uniquely restricted expression of *AQP8* (Figure 2B), consistent with observations in the human colon (Burclaff et al., 2022).

We identified BEST4$^+$ epithelial cells in the cecum at all developmental stages (Figure 2C). This subtype is also observed in human intestines but is absent in mice (Burclaff et al., 2022; Elmentaite et al., 2021; Ito et al., 2013; Parikh

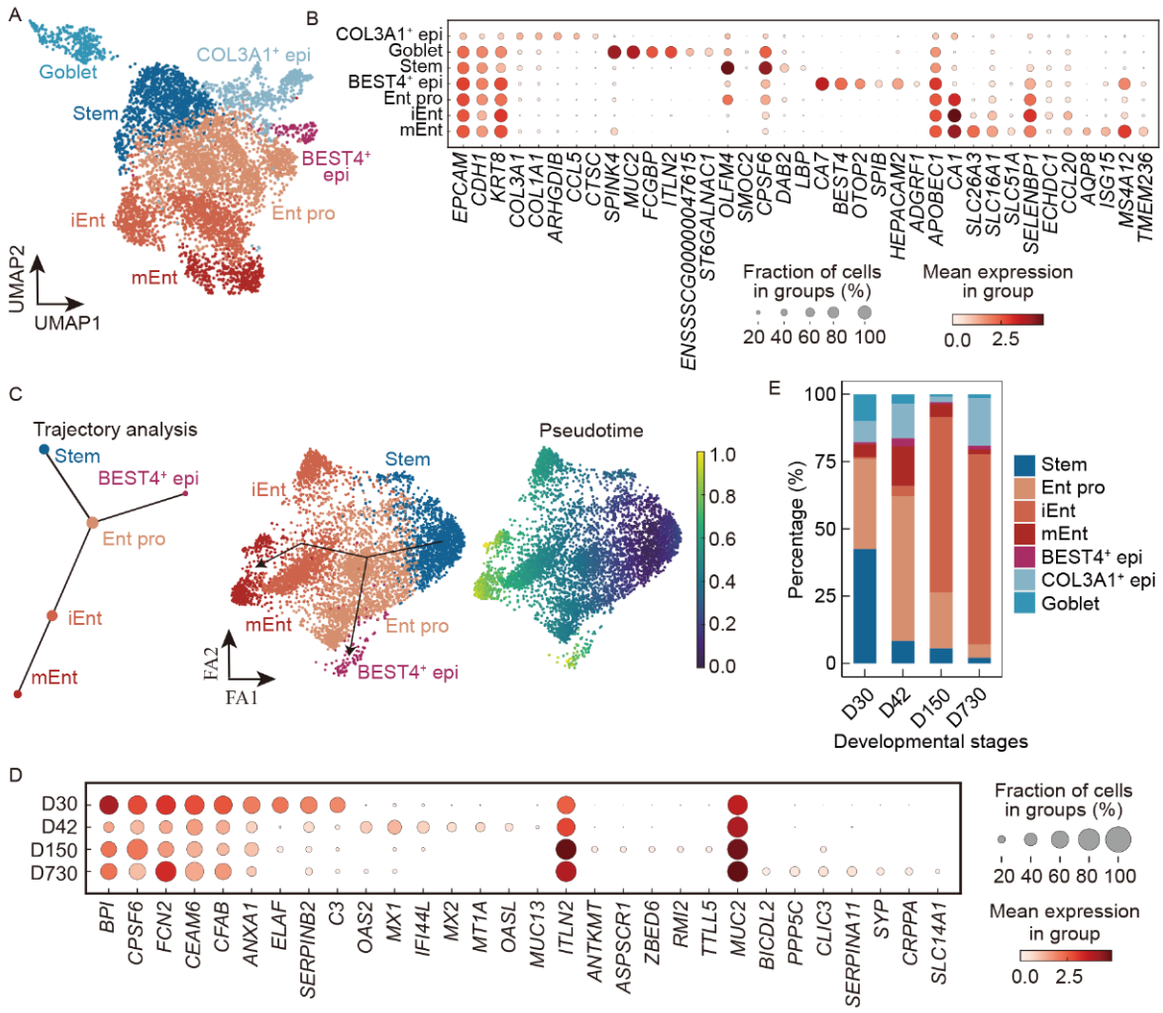


Figure 2 Heterogeneity of epithelial cells

A: UMAP of epithelial cell subtypes. Ent pro: enterocytes progenitor cells; iEnt: immature enterocytes; mEnt: mature enterocytes; epi: epithelial cells. B: Dot plot showing expression of highly expressed genes in epithelial cell subtypes. C: Trajectory analysis of enterocytes using PAGA, showing connectivity between stem cells and subpopulations of enterocytes (left), colored by cell types (middle) and pseudotime (right). D: Dot plot showing expression of highly expressed genes in goblet cells at different developmental stages. E: Relative proportions of epithelial cell subtypes at each developmental stage.

et al., 2019; Smillie et al., 2019). Notably, several marker genes, i.e., *BEST4*, *OTOP2*, and *CA7*, were conserved in both pigs and humans (Figure 2B). Based on the differentiation trajectory of BEST4$^+$ epithelial cells (Figure 2C), 67 TFs showed increased expression with differentiation from stem cells to BEST4$^+$ epithelial cells (Supplementary Table S3).

Goblet cells are responsible for producing and secreting mucin, which forms a mucus layer to support commensal bacteria (Sicard et al., 2017), protect the host epithelium from invading pathogens (Grondin et al., 2020; Specian & Oliver, 1991), and participate in the immune response (Zhang & Wu, 2020). Enrichment analysis showed that goblet cells genes were significantly enriched in GO terms "endoplasmic reticulum to Golgi vesicle-mediated transport" and "response to unfolded protein" (Supplementary Figure S3A), similar to the characteristics of goblet cells in the human gut (Burclaff et al., 2022). Based on these findings, we identified developmental stage-dependent expression of genes in goblet cells. For example, mucin-encoding *MUC2* and *ITLN2*, which participate in defense against pathogens, were expressed across all developmental stages; *C3*, an important member of complement system involved in innate immunity, showed higher expression at the D30 stage; several antiviral-related genes, such as *OAS2*, *MX1*, and *MX2*, were specifically up-regulated at the D42 stage (Figure 2D).

We observed substantial heterogeneity in the proportions of the seven epithelial cell subtypes across the four developmental stages (Figure 2E). For example, stem cells and enterocyte progenitor cells accounted for the majority of the cecal epithelium at the weaning period (D30 and D42), while enterocytes were the main cell type in the D150 and D730 stages. This reflected a physiological transition in the cecum from proliferation and development at the weaning period to food absorption and digestion at the adult stages. Notably, a marked decline in stem cells and increase in enterocyte progenitor cells and enterocytes fractions were observed from the pre-weaning (D30) to post-weaning stages (D42), which may be induced by the abrupt change in diet during the weaning period, when stem cells differentiate into enterocyte progenitor cells to prepare for absorbing nutrition from more solid and coarse feed rather than milk. Diet-induced changes in the function and fate of intestinal stem cells have

also been observed in mice (Aliluev et al., 2021).

Immune cell heterogeneity in pig cecum

Based on sub-clustering analysis, a total of 11 T lymphocyte subtypes were identified, including CD8$^+$ T (*CD8A* and *CD3E*), CD4$^+$CD8$^+$ double positive T cells (DPT) (*CD3E*, *CD8A*, and *CD4*), Tregs (*FOXP3*), γδ T (*CTSW* and *TRGC2*), RHEX$^+$ T, innate lymphoid cells (ILCs) (*KIT* and *IL23R*), cycling T (*CD3D*, *HMGB2*, *TOP2A*, and *UBE2C*), Granzyme A (high expression) natural killer T cells (GZMA^{high} NKT) (*CD3D*, *NKG7*, and *PRF1*), GZMA^{low} NKT, GZMA$^-$ NKT, and cycling NKT (*NKG7*, *HMGB2*, *TOP2A*, and *UBE2C*) (Figure 3A, B). CD8$^+$ T cells were not detected in the pre-weaning stage (D30) (Figure 3C), consistent with previous research reporting that CD8$^+$ T cells began to appear at five weeks of age in piglets (Stokes et al., 2004).

The identified RHEX$^+$ T cell subtype (Figure 3A) primarily occurred in the pre-weaning stage (D30), with its distribution decreasing over subsequent developmental stages (Figure 3C). This suggests that RHEX$^+$ T cells may play important roles in the pig cecum at the early developmental

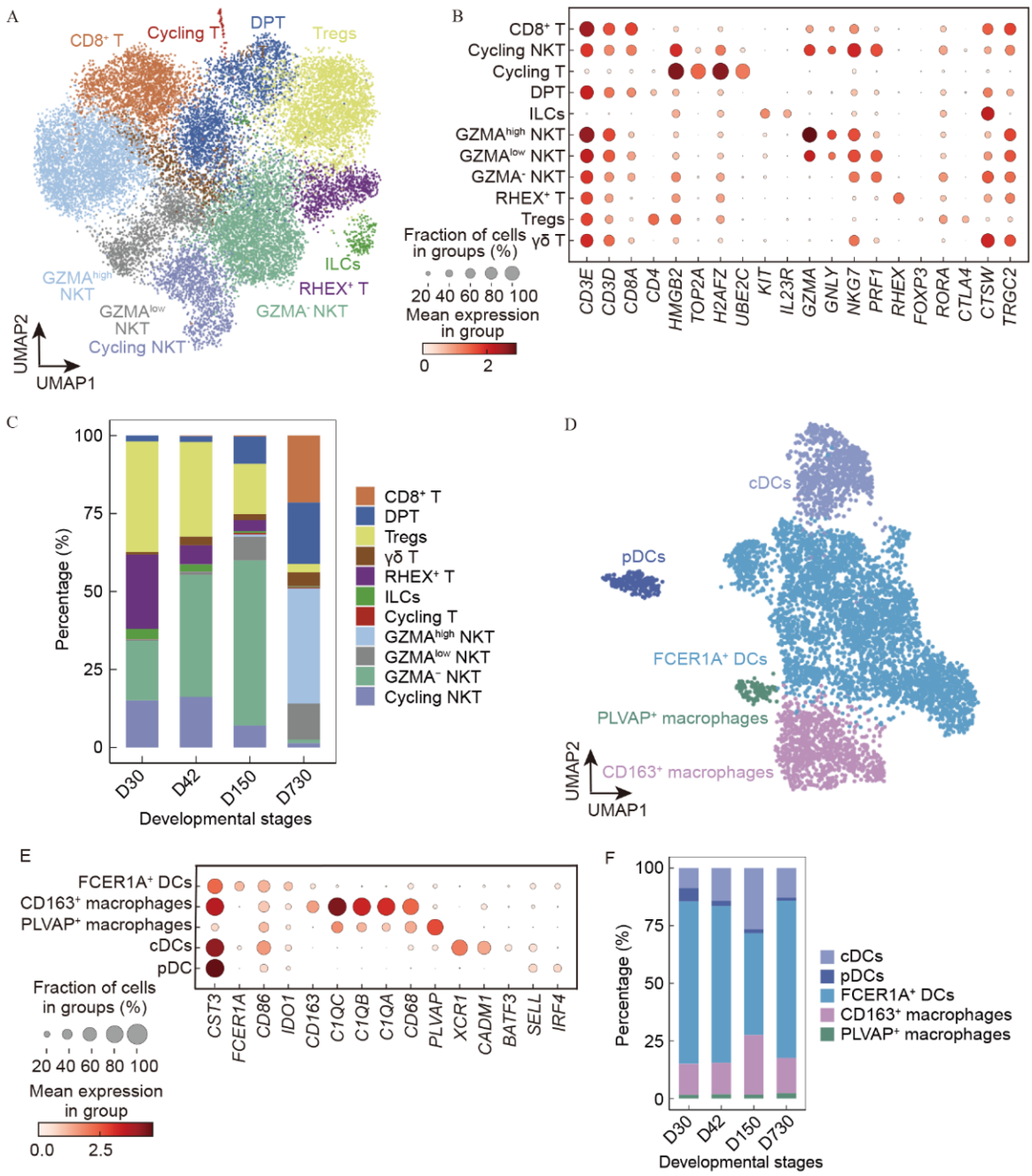


Figure 3 Heterogeneity of immune cells

A: UMAP of T cells (NKT cells plus T cells) subtypes. DPT: CD8$^+$ CD4$^+$ double positive cells; Tregs: regulatory T cells; ILCs: innate lymphocyte cells; NKT: natural killer T cells. B: Dot plot showing expression of highly expressed genes in T cells subtypes. C: Relative proportions of T cell subtypes at each developmental stage. D: UMAP of myeloid cells subtypes. cDCs: classical dendritic cells; pDCs: plasmacytoid dendritic cells. E: Dot plot showing expression of highly expressed genes in myeloid cells subtypes. F: Relative proportions of myeloid cell subtypes at each developmental stage.

stage. The relative proportion of CD8$^+$ T cells and GZMA^{low} NKT cells tended to increase, while Tregs appeared to decrease with development (Figure 3C; Supplementary Figure S3B), suggesting an increased immune response by these cells in the cecum during development. Cycling NKT cells were also enriched in the D30 and D42 stages, while DPT cells were enriched in the D150 and D730 stages (Figure 3C).

The myeloid cells were clustered into five sub-populations, including CD163$^+$ macrophages (*CD163*, *C1QC*, and *C1QB*), PLVAP$^+$ macrophages, FCER1A$^+$ DCs (*FCER1A* and *IDO1*), cDCs (*XCR1*, *CADM1*, and *BATF3*) and plasmacytoid dendritic cells (pDCs, *SELL* and *IRF4*) (Figure 3D, E; Supplementary Figure S3C). Among these, PLVAP$^+$ macrophages expressed both *C1QC* and *C1QB* genes, as well as the VECs marker *PLVAP*, but lacked the *CD163* gene (Figure 3E). The pDCs, a relatively rare cell type linking innate and adaptive immunity (McKenna et al., 2005), showed a gradient decrease from the D30 to D730 stages (Figure 3F).

Fibroblasts and endothelial cell heterogeneity in pig cecum

The fibroblasts were clustered into five sub-populations, including two IBP6$^-$ fibroblast subtypes (IBP6$^-$CCL5$^-$ fib and IBP6$^-$CCL5$^+$ fib) and three IBP6$^+$ fibroblast subtypes (IBP6$^+$IL6$^+$ fib, IBP6$^+$SEMA3C$^+$ fib, and IBP6$^+$CPXM2$^+$ fib), which all expressed fibroblast marker genes *COL1A1*, *COL1A2*, and *DCN* (Figure 4A, B). The IBP6$^-$CCL5$^+$ fib subtype was distinguished by *CCL5* expression but a lack of *IBP6* (Figure 4B), while IBP6$^-$CCL5$^+$ fib also showed *IL2RB* and *CCL4* expression (Figure 4B). Enrichment analysis revealed that the IBP6$^-$CCL5$^+$ fib subtype genes were significantly enriched in the GO term “positive regulation of developmental growth” (Supplementary Figure S4A). The IBP6$^-$CCL5$^+$ fib subtype appeared at the D42 stage and showed a gradient increase with development (Figure 4C). These findings suggested that the IBP6$^-$CCL5$^+$ fib subtype plays an important role in promoting growth and development, especially in late-stage cecal tissue. Enrichment analysis revealed that the three IBP6$^+$ fibroblasts subtypes were significantly enriched in genes involved in “positive regulation of angiogenesis”, implying a key role in angiogenesis (Supplementary Figure S4A). The relative proportion of the IBP6$^+$CPXM2$^+$ fib subtype tended to increase with development (Figure 4C), while the IBP6$^+$IL6$^+$ fib subtype was mainly enriched in the D30 and D42 stages, showing a gradient decrease with development (Figure 4C).

Endothelial cells (ECs) were clustered into six sub-populations, including LYVEL$^+$ LECs (*PROX1*, *RELN*, and *LYVE1*), LYVE1$^-$ LECs (*PROX1* and *RELN*), venous ECs (*PLVAP*, *ADGRG6*, and *ACKR1*), venous capillary ECs (*PLVAP, RGCC*, and *SGK1*), capillary ECs (*EDNRB*, *RGCC*, and *CD36*), and arterial ECs (*TMEM100*, *SEMA3G*, and *GJA5*) (Figure 4A, B; Supplementary Figure S4B). Among these, LYVE1$^-$ LECs expressed LECs markers *PROX1* and *RELN* but did not express *PLVAP*. Furthermore, LYVE1$^-$ LECs also differed based on the expression of *SCG3*, *ECRG4*, *SLC41A1*, and *S100A4* (Figure 4B). Venous capillary ECs expressed both venous marker *PLVAP* and capillary markers *RGCC* and *SGK1* (Figure 4B), similar to observations

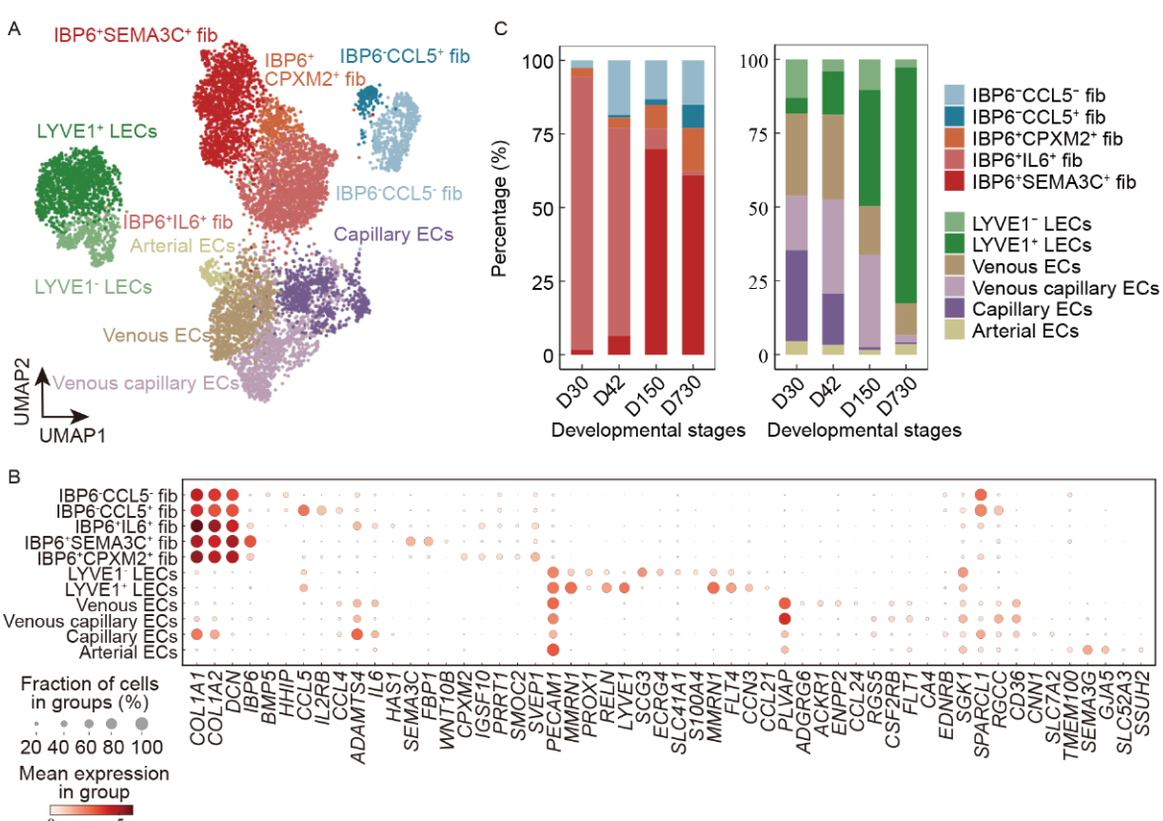


Figure 4 Heterogeneity of fibroblasts and endothelial cells

A: UMAP of fibroblasts and endothelial cell subtypes. fib: fibroblasts; LECs: lymphatic endothelial cells; ECs: endothelial cells. B: Dot plot showing expression of highly expressed genes in fibroblasts and endothelial cell subtypes. C: Relative proportions of fibroblast subtypes (left) and endothelial cell subtypes (right) at each developmental stage.

in humans and mice (Fawkner-Corbett et al., 2021; Kalucka et al., 2020). Additionally, the relative proportion of LYVE1$^+$ LECs tended to increase with developmental stages, particularly from D150 to D730 (Figure 4C).

Age-associated gene expression changes in different cell types in pig cecum

The weaning process, an important event in pig production, induces dramatic functional and morphological changes in the intestine (Campbell et al., 2013; Tang et al., 2022; Zheng et al., 2021). To explore the changes in molecular characteristics during the weaning process at the single-cell level, we identified weaning-associated gene transcriptions in different cecal cell types. In total, we identified 1 869 DEGs, which were differentially expressed in at least one cell type of the D42 cecum compared to the D30 cecum (Figure 5A; Supplementary Table S4). The most affected cell types during weaning included epithelial cells, myeloid cells, IBP6$^+$ fibroblasts, IBP6$^-$ fibroblasts, and NKT cells, with 648, 300, 180, 174, and 151 DEGs in the D42 compared to the D30 cecal cell types, respectively (Figure 5A). Mitochondrial genes were up-regulated in multiple cell types in the post-weaning cecum compared to the pre-weaning cecum (Figure 5B). *COX2*, the inducible form of cyclo-oxygenase, which catalyzes arachidonic acid to produce prostaglandins (Choi et al., 2009), was significantly up-regulated in the NKT, T, myeloid, and epithelial cells. *ND2*, which encodes mitochondrially NADH: ubiquinone oxidoreductase core subunit 2, was significantly up-regulated in NKT, T, VECs, and epithelial cells. Enrichment analysis of DEGs revealed the activation of immune-related pathways in several immune cells, including T, NKT, and B cells (Figure 5C). In addition, genes involved in the transportation of nutrients were down-regulated in epithelial cells (Figure 5C). These results imply that dramatic changes occur in gut gene transcription due to weaning stress.

Receptor-ligand analysis revealed heterogeneity in cell-cell interactions across developmental stages

Tissues achieve functionality through interactions among different cell types. Based on cellular communication analysis using CellChat (Jin et al., 2021), we explored potential cell-type interactions and heterogeneities across different developmental stages. On average, 100 interactions at each developmental stage were detected based on a *P*-value threshold significance of 0.05 (Figure 6A; Supplementary Table S5). Total cellular interaction strength gradually increased with development (Figure 6A).

We next compared outgoing and incoming interaction strengths to identify cell types showing significant changes in signaling activities across different developmental stages. Results revealed that IBP6$^+$ fibroblasts were the main signal senders at the D30 stage, while IBP6$^-$ fibroblasts assumed this role at other stages (Figure 6B). IBP6$^+$ but not IBP6$^-$ fibroblasts used the vascular endothelial growth factor (*VEGFD* and *VEGFC*) signaling pathway for endothelial cell communication (Figure 6C, Supplementary Figure S5A). Given the pivotal role of VEGFs and their endothelial tyrosine kinase receptors in angiogenesis and lymph-angiogenesis (Lohela et al., 2009), this suggests an important regulatory function of IBP6$^+$ fibroblasts in angiogenesis.

Extensive interactions between immune and non-immune cells were also identified. The incoming interaction strength for T cells gradually increased with development (Figure 6B), suggesting increasing involvement of different cell types in mediating the T cell immune response. Additionally, *GZMA* was identified as an outgoing signal from NKT cells, predicted to interact with *F2RL1/F2RL2* in epithelial cells and IBP6$^+$ fibroblasts at the D730 stage (Figure 6D; Supplementary Figure S5A, B). The *GZMA*-encoded serine protease promotes aberrant cell death in tumorigenic and virus-infected cells (Catalan et al., 2012; Chowdhury & Lieberman, 2008; Topham & Hewitt, 2009), suggesting a potential host strategy to lyse infected cells, and thereby provide protection and promote cell turnover. Furthermore, *CCL5-ACKR1* interactions between NKT cells and VECs were also observed at D730 (Figure 6D; Supplementary Figure S5A, B). *CCL5* plays an important role in immune responses and inflammation (Marques et al., 2013), while *ACKR1* is thought to clear inflammatory CC and CXC chemokines to regulate acute inflammatory responses (Davis et al., 2015; Wan et al., 2015). Overall, these results reveal the important role of immune and non-immune cell interactions in intestinal homeostasis.

Conservativeness of cell-type gene expression in pig and human ceca

To better understand the conservation of genes between human and pig intestines, we compared our data with published transcriptome data of the human cecum (Elmentaite et al., 2021). In total, 45 572 cells from the pig cecum and 20 693 cells from the human cecum were integrated for cross-species comparison (Haghverdi et al., 2018). Unsupervised graph-based clustering showed that the same cell types in humans and pigs were grouped together (Figure 7A). We next examined gene expression conservation between the pig and human ceca. Results showed that marker genes defining cell types were conserved in both pigs and humans (Figure 7B), including *MUC2* and *FCGBP* in goblet cells, *PECAM1* and *PLVAP* in endothelial cells, *SLC26A3* and *AQP8* in enterocytes, and *CD19* and *CD79A* in B cells. *MUC2* is secreted by goblet cells to form an insoluble mucous barrier to protect the gut lumen (Grondin et al., 2020), while *AQP8* plays an important role in water absorption (Wang et al., 2020). In addition, results also revealed several conserved TFs in both the pig and human single-cell datasets (Figure 7C), including *SOX18* in endothelial cells, *CREB3L1* and *SPDEF* in goblet cells, *CDX2* in enterocytes, and *PAX5* in B cells. Of note, *SOX18* is essential for endothelial cell differentiation (Yao et al., 2019), while *SPDEF* is required for goblet cell mucus production and differentiation (Song et al., 2017). In summary, these observations confirmed the conservation of gene expression and functions in the pig and human ceca.

DISCUSSION

The intestinal tract, vital for digestion, absorption, and immunological functions, has been extensively characterized through single-cell transcriptomics in humans and mice (Burclaff et al., 2022; Elmentaite et al., 2021; Haber et al., 2017). However, in pigs, investigations into cellular compositions and gene expression profiles in the intestinal tract at different developmental stages remain unexplored. In the present study, we conducted a single-cell transcriptomic analysis of the pig cecum at ages 30, 42, 150 and 730 days, and further explored cellular heterogeneity across different developmental stages, particularly before and after weaning, which is characterized by marked shifts in diet and environmental conditions (Blavi et al., 2021). Our results indicated an expansion in immune cells from the pre-weaning

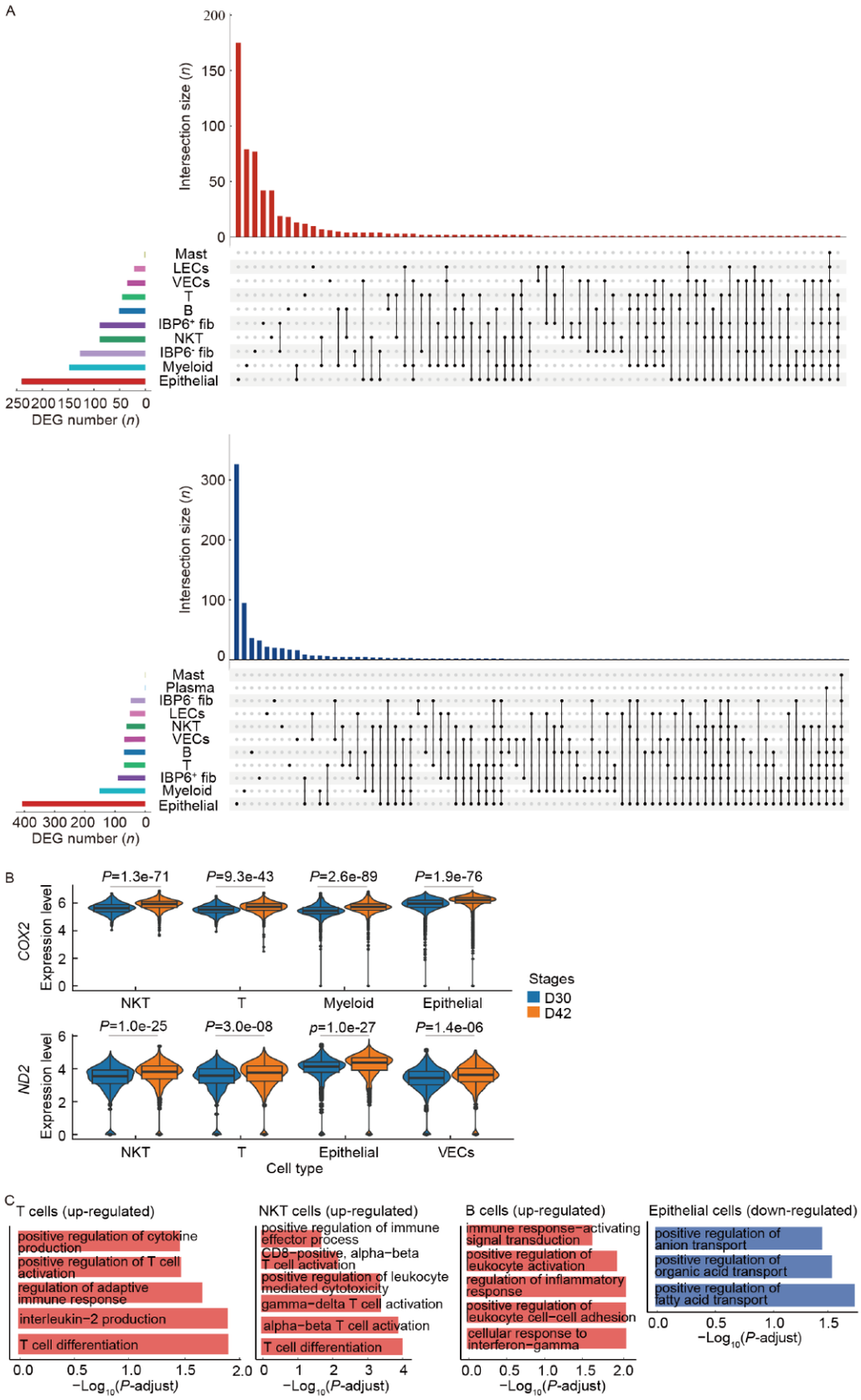


Figure 5 Weaning-associated changes in gene expression levels in different cell types

A: Intersection plot showing number of DEGs (top: up-regulated genes; bottom: down-regulated genes) in different cell types. Left horizontal bar plot shows total number of DEGs in each cell type. Top bar plot shows number of DEGs for each intersection. Abbreviations of cell types are the same as in Figure 1. B: Violin plot combined with box plot showing transcription levels of *COX2* and *ND2* in D30 and D42 stages, respectively. *P*-value refers to Benjamini-Hochberg adjusted *P*-value. C: GO pathways significantly enriched in up-regulated or down-regulated genes in T, NKT, B, and epithelial cells at D42 stage compared to D30 stage.

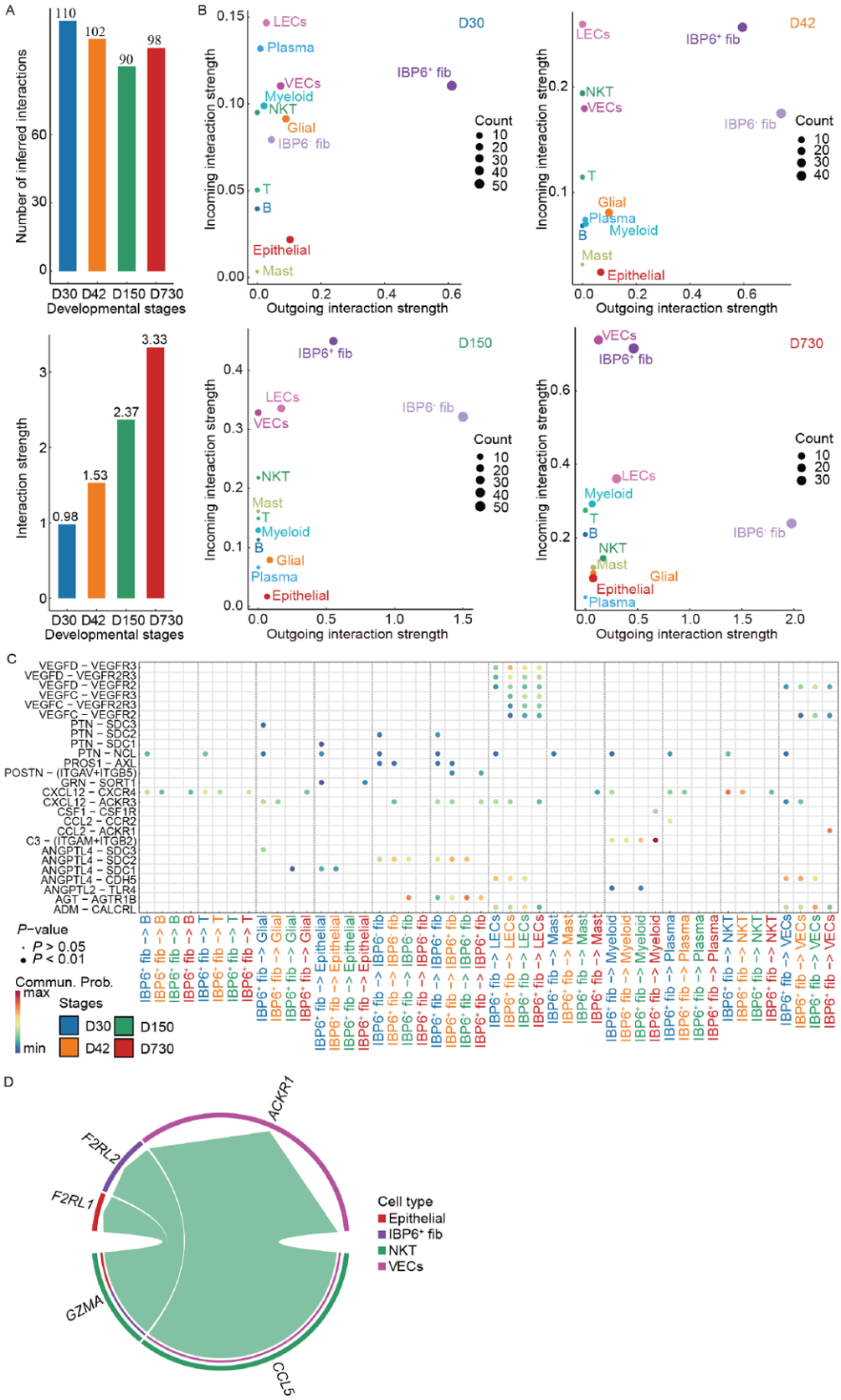


Figure 6 Cell-cell communication in developing pig cecum

A: Total number and strength of inferred cell-cell interactions at each developmental stage. B: Scatter plot showing strength of signals sent and received by each cell type at each developmental stage. Abbreviations of cell types are the same as in Figure 1. C: Bubble plot showing significant receptor-ligand pairs of IBP6$^+$ fibroblasts for communication with other cell types at different developmental stages. D: Chord diagram showing all interactions sent from NKT cells at D730 stage.



Figure 7 Comparison of human and pig cecum single-cell data

A: UMAP plot showing integration of human and pig cecum single-cell data, colored by cell type (left) and species (right). DCs: dendritic cells. B: Dot plot showing expression of marker genes between human and pig cecum single-cell data. ENS*35-MUC2: ENSSSCG00000040035-MUC2, ENS*15-FCGBP: ENSSSCG00000027515-FCGBP. C: Dot plot showing expression of TFs between human and pig cecum single-cell data.

(D30) to post-weaning stages (D42), as well as a decline in epithelial stem cells along developmental stages, particularly from D30 to D42. We identified 848 up-regulated and 1 021 down-regulated genes after weaning. Among these genes, certain mitochondrial genes were up-regulated in different cell types in the D42-stage cecum, including *COX2* in T, NKT, epithelial, and myeloid cells and *COX1* in epithelial cells. *COX1* and *COX2* are key enzymes involved in the catalysis of arachidonic acid to produce prostaglandins (Choi et al., 2009), which are implicated in inflammatory processes (Ricciotti & Fitzgerald, 2011). Concurrently, we noted that up-regulated genes were predominantly enriched in immune-activation related pathways, including "CD8-positive, alpha-beta T cell activation" and "positive regulation of immune effector process" in NKT cells, and "immune response-activation signal transduction" and "positive regulation of leukocyte activation" in B cells. These findings provide insight into physiological changes that occur during weaning and lay the groundwork for understanding the underlying mechanisms of weaning stress in the pig cecum.

We identified two major fibroblast subtypes, which were distinguished by *IBP6* expression and displayed distinct cell-cell crosstalk patterns across the different developmental stages. Notably, IBP6$^+$ fibroblasts acted as the dominant signal senders at the D30 stage, while IBP6$^-$ fibroblasts assumed this role at the other stages. Results indicated that IBP6$^+$ fibroblasts used *PTN/NCL* to communicate with all other cell types, but only at the D30 stage. *PTN*, a secreted growth factor, is associated with cell differentiation, growth, and proliferation (González-Castillo et al., 2015). IBP6$^+$ but not IBP6$^-$ fibroblasts used *C3-(ITGAM_ITGB2)* to communicate with myeloid cells at the D30 and D42 stages. *C3*, also known as complement C3, plays a critical role in the innate immune response (Bohlson et al., 2014). Additionally, IBP6$^+$ but not IBP6$^-$ fibroblasts communicated with endothelial cells (LECs and VECs) through *VEGFD/VEGFC-VEGFR3/VEGFR2* pairs. VEGFs and their endothelial tyrosine kinase receptors play key roles in angiogenesis (Lohela et al., 2009). These observations suggest important roles of IBP6$^+$ fibroblasts in the regulation of innate immunity and angiogenesis, as well as the promotion of cell growth and differentiation, especially in early-stage cecal tissue.

Gut immune cells account for 70%–80% of all immune cells in the body (Agace & McCoy, 2017; Mowat & Agace, 2014; Wiertsema et al., 2021). These immune cells are essential for normal physiology and pathogen defense (James et al., 2022). Our data revealed extensive interactions between immune and non-immune cells at the different developmental stages, including interactions between NKT and epithelial cells, plasma and enteric glial cells, and myeloid cells and LECs. At the D730 stage, NKT cells sent GZMA signals to interact with epithelial cells and IBP6$^+$ fibroblasts. *GZMA* encodes serine protease, which induces aberrant cell death of tumorigenic and virus-infected cells (Catalan et al., 2012; Chowdhury & Lieberman, 2008; Topham & Hewitt, 2009). Plasma cells used *MIF-ACKR3* to communicate with non-immune cells at the D30 and D42 stages, including enteric glial cells, VECs, LECs, and IBP6$^+$ fibroblasts. The *MIF* ligand is an immunoregulatory cytokine and displays direct chemokine-like functions (Grieb et al., 2010), while receptor *ACKR3* is considered a scavenger receptor gene (Duval et al., 2022). Based on these results, we hypothesize that intestinal immune and non-immune cells may communicate with each other using ligand and receptor pairs to maintain intestinal homeostasis. However, the interactions between these receptor ligands and cells need to be further verified with spatial transcriptomic analysis.

This study has several limitations worth mentioning. Due to the limited number of cells, not all epithelial subtypes were captured, including tuft and enteroendocrine cells, possibly due to insufficient or prolonged tissue digestion. Consequently, observations regarding cell-type proportions and their temporal shifts during development, particularly the hypothesis suggesting diet-induced changes in stem cell composition from the pre-weaning to post-weaning stages, require further experimental verification based on other approaches, such as immunofluorescence quantification and flow sorting. Future research should expand single-cell datasets and incorporate multi-omics analyses to elucidate the cellular characteristics of the pig cecum more comprehensively.

DATA AVAILABILITY

Datasets generated in this study were deposited in the Genome Sequence Archive (GSA) database (https://ngdc.cncb.ac.cn/gsa/) under accession number CRA010843, Science Data Bank (doi:10.57760/sciencedb.j00139.00082) and NCBI under BioProjectID PRJNA1036035.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Experimental design and conception: L.S.H.; investigation: Y.Y.X., B.Y., Q.Z., F.H., L.R., T.X.Y., S.Y.Y., L.P.C., L.X., X.X.Z., and J.W.Y.; formal analysis: Y.Y.X.; writing-original draft: Y.Y.X.; writing-review & editing: L.S.H. and B.Y.; experimental organization and analysis supervision: L.S.H. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We would like to thank colleagues in the National Key Laboratory for Swine Genetic Improvement and Germplasm Innovation, Jiangxi Agricultural University for their help in collecting samples and providing suggestions during analysis.

REFERENCES

Agace WW, McCoy KD. 2017. Regionalized development and maintenance of the intestinal adaptive immune landscape. *Immunity*, **46**(4): 532−548.

Aibar S, González-Blas CB, Moerman T, et al. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nature Methods*, **14**(11): 1083−1086.

Ali A, Tan HY, Kaiko GE. 2020. Role of the intestinal epithelium and its interaction with the microbiota in food allergy. *Frontiers in Immunology*, **11**: 604054.

Aliluev A, Tritschler S, Sterr M, et al. 2021. Diet-induced alteration of intestinal stem cell function underlies obesity and prediabetes in mice. *Nature Metabolism*, **3**(9): 1202−1216.

Blavi L, Solà-Oriol D, Llonch P, et al. 2021. Management and feeding strategies in early life to increase piglet performance and welfare around weaning: a review. *Animals*, **11**(2): 302.

Bohlson SS, O'Conner SD, Hulsebus HJ, et al. 2014. Complement, C1q, and C1q-related molecules regulate macrophage polarization. *Frontiers in Immunology*, **5**: 402.

Boyd M, Bressendorff S, Møller J, et al. 2009. Mapping of HNF4α target genes in intestinal epithelial cells. *BMC Gastroenterology*, **9**: 68.

Brown K, Abbott DW, Uwiera RRE, et al. 2018. Removal of the cecum affects intestinal fermentation, enteric bacterial community structure, and acute colitis in mice. *Gut Microbes*, **9**(3): 218−235.

Burclaff J, Bliton RJ, Breau KA, et al. 2022. A proximal-to-distal survey of healthy adult human small intestine and colon epithelium by single-cell transcriptomics. *Cellular and Molecular Gastroenterology and Hepatology*, **13**(5): 1554−1589.

Campbell JM, Crenshaw JD, Polo J. 2013. The biological stress of early weaned piglets. *Journal of Animal Science and Biotechnology*, **4**(1): 19.

Catalan E, Sanchez-Martinez D, Pardo J. 2012. GZMA (granzyme A (granzyme 1, cytotoxic T-lymphocyte-associated serine esterase 3)). *Atlas of Genetics and Cytogenetics in Oncology and Haematology*, **16**(2): 123−126.

Chassaing B, Kumar M, Baker MT, et al. 2014. Mammalian gut immunity. *Biomedical Journal*, **37**(5): 246−258.

Chen SF, Zhou YQ, Chen YR, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884−i890.

Choi SH, Aid S, Bosetti F. 2009. The distinct roles of cyclooxygenase-1 and -2 in neuroinflammation: implications for translational research. *Trends in Pharmacological Sciences*, **30**(4): 174−181.

Chowdhury D, Lieberman J. 2008. Death by a thousand cuts: granzyme pathways of programmed cell death. *Annual Review of Immunology*, **26**: 389−420.

Cobaleda C, Schebesta A, Delogu A, et al. 2007. Pax5: the guardian of B cell identity and function. *Nature Immunology*, **8**(5): 463−470.

Dabareiner RM, White II NA. 1997. Diseases and surgery of the cecum. *Veterinary Clinics of North America:Equine Practice*, **13**(2): 303−315.

Davis MB, Walens A, Hire R, et al. 2015. Distinct transcript isoforms of the atypical chemokine receptor 1 (ACKR1)/duffy antigen receptor for chemokines (DARC) gene are expressed in lymphoblasts and altered isoform levels are associated with genetic ancestry and the duffy-null allele. *PLoS One*, **10**(10): e0140098.

Duval V, Alayrac P, Silvestre JS, et al. 2022. Emerging roles of the atypical chemokine receptor 3 (ACKR3) in cardiovascular diseases. *Frontiers in Endocrinology*, **13**: 906586.

Elmentaite R, Kumasaka N, Roberts K, et al. 2021. Cells of the human intestinal tract mapped across space and time. *Nature*, **597**(7875): 250−255.

Fawkner-Corbett D, Antanaviciute A, Parikh K, et al. 2021. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell*, **184**(3): 810−826.e23.

Gerrits RJ, Lunney JK, Johnson LA, et al. 2005. Perspectives for artificial insemination and genomics to improve global swine populations. *Theriogenology*, **63**(2): 283−299.

González-Castillo C, Ortuño-Sahagún D, Guzmán-Brambila C, et al. 2015. Pleiotrophin as a central nervous system neuromodulator, evidences from the hippocampus. *Frontiers in Cellular Neuroscience*, **8**: 443.

Grieb G, Merk M, Bernhagen J, et al. 2010. Macrophage migration inhibitory factor (MIF): a promising biomarker. *Drug News & Perspectives*, **23**(4): 257−64.

Grondin JA, Kwon YH, Far PM, et al. 2020. Mucins in intestinal mucosal defense and inflammation: learning from clinical and experimental studies. *Frontiers in Immunology*, **11**: 2054.

Haber AL, Biton M, Rogel N, et al. 2017. A single-cell survey of the small intestinal epithelium. *Nature*, **551**(7680): 333−339.

Haghverdi L, Lun ATL, Morgan MD, et al. 2018. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology*, **36**(5): 421−427.

Han L, Wei XY, Liu CY, et al. 2022. Cell transcriptomic atlas of the non-human primate *Macaca fascicularis*. *Nature*, **604**(7907): 723–731.

Han XP, Wang RY, Zhou YC, et al. 2018. Mapping the mouse cell atlas by microwell-seq. *Cell*, **172**(5): 1091−1107.e17.

Han XP, Zhou ZM, Fei LJ, et al. 2020. Construction of a human cell landscape at single-cell level. *Nature*, **581**(7808): 303−309.

Ito G, Okamoto R, Murano T, et al. 2013. Lineage-specific expression of bestrophin-2 and bestrophin-4 in human intestinal epithelial cells. *PLoS One*, **8**(11): e79693.

James KR, Elmentaite R, Teichmann SA, et al. 2022. Redefining intestinal immunity with single-cell transcriptomics. *Mucosal Immunology*, **15**(4): 531−541.

James KR, Gomes T, Elmentaite R, et al. 2020. Distinct microbial and immune niches of the human colon. *Nature Immunology*, **21**(3): 343−353.

James SP. 1993. The gastrointestinal mucosal immune system. *Digestive Diseases*, **11**(3): 146−156.

Jin SQ, Guerrero-Juarez CF, Zhang LH, et al. 2021. Inference and analysis of cell-cell communication using CellChat. *Nature Communications*, **12**(1): 1088.

Kalucka J, de Rooij LPMH, Goveia J, et al. 2020. Single-cell transcriptome atlas of murine endothelial cells. *Cell*, **180**(4): 764−779.e20.

Kong SS, Zhang YH, Zhang WQ. 2018. Regulation of intestinal epithelial cells properties and functions by amino acids. *BioMed Research International*, **2018**: 2819154.

Korsunsky I, Millard N, Fan J, et al. 2019. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods*, **16**(12): 1289−1296.

Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics*, **18**(2): 205−214.

Li HJ, Janssens J, De Waegeneer M, et al. 2022. Fly cell atlas: a single-nucleus transcriptomic atlas of the adult fruit fly. *Science*, **375**(6584): eabk2432.

Liao DH, Zhao JB, Gregersen H. 2009. Gastrointestinal tract modelling in health and disease. *World Journal of Gastroenterology*, **15**(2): 169−176.

Lin L, Yee SW, Kim RB, et al. 2015. SLC transporters as therapeutic targets: emerging opportunities. *Nature Reviews Drug Discovery*, **14**(8): 543−560.

Lobo GP, Hessel S, Eichinger A, et al. 2010. ISX is a retinoic acid-sensitive gatekeeper that controls intestinal β, β-carotene absorption and vitamin A production. *The FASEB Journal*, **24**(6): 1656−1666.

Lohela M, Bry M, Tammela T, et al. 2009. VEGFs and receptors involved in angiogenesis versus lymphangiogenesis. *Current Opinion in Cell Biology*, **21**(2): 154−165.

Lunney JK, Van Goor A, Walker KE, et al. 2021. Importance of the pig as a human biomedical model. *Science Translational Medicine*, **13**(621): eabd5758.

Marques RE, Guabiraba R, Russo RC, et al. 2013. Targeting CCL5 in inflammation. *Expert Opinion on Therapeutic Targets*, **17**(12): 1439−1460.

McKenna K, Beignon AS, Bhardwaj N. 2005. Plasmacytoid dendritic cells: linking innate and adaptive immunity. *Journal of Virology*, **79**(1): 17−27.

Medvedovic J, Ebert A, Tagoh H, et al. 2011. Pax5: a master regulator of B cell development and leukemogenesis. *Advances in Immunology*, **111**: 179−206.

Mowat AM, Agace WW. 2014. Regional specialization within the intestinal immune system. *Nature Reviews Immunology*, **14**(10): 667−685.

Parikh K, Antanaviciute A, Fawkner-Corbett D, et al. 2019. Colonic epithelial cell diversity in health and inflammatory bowel disease. *Nature*, **567**(7746): 49−55.

Ricciotti E, Fitzgerald GA. 2011. Prostaglandins and inflammation. *Arteriosclerosis, Thrombosis, and Vascular Biology*, **31**(5): 986−1000.

Rodriguez PC, Ochoa AC, Al-Khami AA. 2017. Arginine metabolism in myeloid cells shapes innate and adaptive immunity. *Frontiers in Immunology*, **8**: 93.

Seino Y, Miki T, Kiyonari H, et al. 2008. Isx participates in the maintenance of vitamin A metabolism by regulation of β-carotene 15, 15'-monooxygenase (Bcmo1) expression. *Journal of Biological Chemistry*, **283**(8): 4905−4911.

Shroyer NF, Wallis D, Venken KJT, et al. 2005. *Gfi1* functions downstream of *Math1* to control intestinal secretory cell subtype allocation and differentiation. *Genes & Development*, **19**(20): 2412−2417.

Sicard JF, Le Bihan G, Vogeleer P, et al. 2017. Interactions of intestinal bacteria with components of the intestinal mucus. *Frontiers in Cellular and Infection Microbiology*, **7**: 387.

Smillie CS, Biton M, Ordovas-Montanes J, et al. 2019. Intra- and inter-cellular rewiring of the human colon during ulcerative colitis. *Cell*, **178**(3): 714−730.e22.

Snoeck V, Goddeeris B, Cox E. 2005. The role of enterocytes in the intestinal barrier function and antigen uptake. *Microbes and Infection*, **7**(7-8): 997−1004.

Song J, Heijink IH, Kistemaker LEM, et al. 2017. Aberrant DNA methylation and expression of SPDEF and FOXA2 in airway epithelium of patients with COPD. *Clinical Epigenetics*, **9**: 42.

Specian RD, Oliver MG. 1991. Functional biology of intestinal goblet cells. *American Journal of Physiology-Cell Physiology*, **260**(2 Pt 1): C183–C193.

Stokes CR, Bailey M, Haverson K, et al. 2004. Postnatal development of intestinal immune system in piglets: implications for the process of weaning. *Animal Research*, **53**(4): 325−334.

Stumpff F, Manneck D, Martens H. 2019. Unravelling the secrets of the caecum. *Pflügers Archiv - European Journal of Physiology*, **471**(7): 925−926.

Su XM, Xu YL, Fox GC, et al. 2021. Breast cancer-derived GM-CSF regulates arginase 1 in myeloid cells to promote an immunosuppressive microenvironment. *Journal of Clinical Investigation*, **131**(20): e145296.

Tang XP, Xiong KN, Fang RJ, et al. 2022. Weaning stress and intestinal health of piglets: a review. *Frontiers in Immunology*, **13**: 1042778.

Topham NJ, Hewitt EW. 2009. Natural killer cell cytotoxicity: how do they pull the trigger?. *Immunology*, **128**(1): 7−15.

Tunçer S, Sade-Memişoğlu A, Keşküş AG, et al. 2020. Enhanced expression of HNF4αduring intestinal epithelial differentiation is involved in the activation of ER stress. *The FEBS Journal*, **287**(12): 2504−2523.

Van de Sande B, Flerin C, Davie K, et al. 2020. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nature Protocols*, **15**(7): 2247−2276.

van der Flier LG, Haegebarth A, Stange DE, et al. 2009. OLFM4 is a robust marker for stem cells in human intestine and marks a subset of colorectal cancer cells. *Gastroenterology*, **137**(1): 15−17.

Wan WZ, Liu Q, Lionakis MS, et al. 2015. Atypical chemokine receptor 1 deficiency reduces atherogenesis in *ApoE*-knockout mice. *Cardiovascular Research*, **106**(3): 478−487.

Wang YL, Song WL, Wang JL, et al. 2020. Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *Journal of Experimental Medicine*, **217**(2): e20191130.

Wiertsema SP, van Bergenhenegouwen J, Garssen J, et al. 2021. The Interplay between the gut microbiome and the immune system in the context of infectious diseases throughout life and the role of nutrition in optimizing treatment strategies. *Nutrients*, **13**(3): 886.

Wolf FA, Hamey FK, Plass M, et al. 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology*, **20**(1): 59.

Yang L, Yang HY, Chu YX, et al. 2021. CREPT is required for murine stem cell maintenance during intestinal regeneration. *Nature Communications*, **12**(1): 270.

Yao YC, Yao JY, Boström KI. 2019. SOX transcription factors in endothelial differentiation and endothelial-mesenchymal transitions. *Frontiers in Cardiovascular Medicine*, **6**: 30.

Zhang MM, Wu CC. 2020. The relationship between intestinal goblet cells and the immune response. *Bioscience Reports*, **40**(10): BSR20201471.

Zheng L, Duarte ME, Sevarolli Loftus A, et al. 2021. Intestinal health of pigs upon weaning: challenges and nutritional intervention. *Frontiers in Veterinary Science*, **8**: 628258.

Zhu YL, Zhou ZM, Huang T, et al. 2022. Mapping and analysis of a spatiotemporal H3K27ac and gene expression spectrum in pigs. *Science China Life Sciences*, **65**(8): 1517−1534.

Zorn AM, Wells JM. 2009. Vertebrate endoderm development and organ formation. *Annual Review of Cell and Developmental Biology*, **25**: 221−251.

Wang et al. *Zool. Res.* 2024, 45(1): 69−78
https://doi.org/10.24272/j.issn.2095-8137.2023.138

ZR | Zoological Research

Article Open Access

Unexpected divergence in magnetoreceptor MagR from robin and pigeon linked to two sequence variations

Shun Wang[1,2], Peng Zhang[2,3], Fan Fei[2,3], Tianyang Tong[2,4], Xiujuan Zhou[2,3], Yajie Zhou[1,2], Jing Zhang[2,3], Mengke Wei[1,2], Yanqi Zhang[2,3], Lei Zhang[2,3], Yulong Huang[5], Lin Zhang[2], Xin Zhang[1,2,3,6], Tiantian Cai[2,3,*], Can Xie[2,3,6,*]

[1] *Institutes of Physical Science and Information Technology*, *Anhui University*, *Hefei*, *Anhui* 230039, *China*
[2] *High Magnetic Field Laboratory*, *Hefei Institutes of Physical Science*, *Chinese Academy of Sciences*, *Science Island*, *Hefei*, *Anhui* 230031, *China*
[3] *Science Island Branch of Graduate School*, *University of Science and Technology of China*, *Hefei*, *Anhui* 230036, *China*
[4] *Department of Anatomy*, *Anhui Medical University*, *Hefei*, *Anhui* 230032, *China*
[5] *Department of Mechanical and Aerospace Engineering*, *University at Buffalo*, *State University of New York*, *Buffalo*, *NY* 14260 *USA*
[6] *International Magnetobiology Frontier Research Center*, *Science Island*, *Hefei*, *Anhui* 230031, *China*

ABSTRACT

Birds exhibit extraordinary mobility and remarkable navigational skills, obtaining guidance cues from the Earth's magnetic field for orientation and long-distance movement. Bird species also show tremendous diversity in navigation strategies, with considerable differences even within the same taxa and among individuals from the same population. The highly conserved iron and iron-sulfur cluster binding magnetoreceptor (MagR) protein is suggested to enable animals, including birds, to detect the geomagnetic field and navigate accordingly. Notably, MagR is also implicated in other functions, such as electron transfer and biogenesis of iron-sulfur clusters, raising the question of whether variability exists in its biochemical and biophysical features among species, particularly birds. In the current study, we conducted a comparative analysis of MagR from two different bird species, including the migratory European robin (*Erithacus rubecula*) and the homing pigeon (*Columba livia*) . Sequence alignment revealed an extremely high degree of similarity between the MagRs of these species, with only three sequence variations. Nevertheless, two of these variations underpinned significant differences in metal binding capacity, oligomeric state, and magnetic properties. These findings offer compelling evidence for the marked differences in MagR between the two avian species, potentially explaining how a highly conserved protein can mediate such diverse functions.

Keywords: Homing and migration; Animal navigation; Magnetoreceptor (MagR); Diverse navigation pattern; Conserved protein

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

INTRODUCTION

Birds are among the most agile and mobile organisms in the animal kingdom, possessing an innate ability to navigate vast distances and reliably return to their points of origin. The Arctic tern (*Sterna paradisaea*) holds the record for the longest migration of any animal in the world, making a return trip of approximately 70 900 km from pole to pole each year (Egevang et al., 2010). Birds, as well as many other animals, can detect the Earth's magnetic field, which they utilize as a navigational tool to maintain spatial orientation and direction (Johnsen & Lohmann, 2008; Mouritsen, 2018; Wiltschko & Wiltschko, 2019). Homing and migration represent two distinct navigational strategies used by avian species, which are crucial for locating food sources and nesting sites (Dingle & Drake, 2007). The European robin (*Erithacus rubecula*), a widely distributed species across Europe, is noted for its semi-migratory behavior, with some populations moving between wintering grounds and summer breeding areas, and others remaining permanent residents throughout the year. In contrast, the pigeon (*Columba livia*) is one of the best-studied subjects for understanding animal homing behavior (Meade et al., 2005). The navigation systems of both the robin and pigeon have been extensively studied as model systems in the field of animal navigation.

For navigation, animals utilize two types of geomagnetic field cues: the vector of magnetic field lines, providing

Received: 11 October 2023; Accepted: 28 November 2023; Online: 30 November 2023

Foundation items: This study was supported by the National Natural Science Foundation of China (31640001 and T2350005 to C.X., U21A20148 to X.Z. and C.X.), Ministry of Science and Technology of China (2021ZD0140300 to C.X.), and Presidential Foundation of Hefei Institutes of Physical Science, Chinese Academy of Sciences (Y96XC11131, E26CCG27, and E26CCD15 to C.X., E36CWGBR24B and E36CZG14132 to T.C.)

*Corresponding authors, E-mail: ttcai@hmfl.ac.cn; canxie@hmfl.ac.cn

directional information via polarity, and the gradient declines in magnetic intensity and inclination from the magnetic poles to the equator, which are integral components of the navigational "map" (Wiltschko & Wiltschko, 2019). In recent decades, several primary models have been proposed to explain how animals sense magnetic fields from different perspectives, including the magnetite-based model (Kirschvink et al., 2001), radical pair model (Hore & Mouritsen, 2016; Maeda et al., 2008; Rodgers & Hore, 2009; Xu et al., 2021), and biocompass model (Lohmann, 2016; Qin et al., 2016; Xie, 2022).

Magnetoreceptor (MagR), a highly conserved A-type iron and iron-sulfur cluster binding protein (Guo et al., 2021; Zhou et al., 2023) involved in the biogenesis of iron-sulfur clusters and electron transfer reactions (Bilder et al., 2004; Cózar-Castellano et al., 2004), has been suggested as a potential "universal mechanism for animal magnetoreception" in the biocompass model (Lohmann, 2016; Qin et al., 2016). However, the precise mechanism by which MagR may serve as a biocompass in response to the geomagnetic field is the subject of ongoing debate. Meister (2016) argued that the iron content in MagR is insufficient by several orders of magnitude to produce thermally induced fluctuations in the magnetic moment, while Nimpf & Keays (2022) suggested that the magnetic properties of MagR may result from laboratory contamination. Interestingly, several theories based on the MagR/Cry biocompass model, including long-range intermolecular electron transfer (Xie, 2022) and quantum mechanics (Cao & Yan, 2018; Xiao et al., 2020; Zhao et al., 2022), have provided alternative theoretical explanations of how MagR responds to the geomagnetic field. For example, Xiao et al. (2020) proposed a microscopic quantum dynamical model, which posits that radical pairs may facilitate geomagnetic field detection through induction of magnetic fluctuation in the MagR structure rather than permanent magnetism. Additionally, Arai et al. (2022) suggested that an external magnetic field may induce assembly of the MagR protein.

An increasing number of applications leveraging the magnetic properties of MagR have been reported, utilizing MagR or the MagR/Cry complex as magnetic tags (Jiang et al., 2017; Kang et al., 2021; Wang et al., 2019), bionic magnetic sensors (Cheng et al., 2023; Xue et al., 2020), and MRI contrast enhancers (Li et al., 2023). The functional significance of MagR has been extensively explored, with studies revealing daily oscillations in the expression patterns of MagR in certain species and higher expression levels in migratory populations, hinting at the physiological relevance of MagR in magnetoreception (Chang et al., 2017, 2018; Jin et al., 2023; Mannino et al., 2023). Notably, a recent *in vivo* study demonstrated that knockdown of MagR in oriental armyworms (*Mythimna separata*) abolished their magnetic orientation, suggesting a vital role of MagR in magnetoreception (Tong et al., 2022).

Considering the extensive variety of navigational strategies, such as homing and migration, across species, populations, and individuals in very different environments (Kentie et al., 2023), as well as the diverse functions attributed to MagR, the question arises: how does such a highly conserved protein function as a magnetoreceptor and mediate distinct biological processes? In our previous work, we determined that pigeon MagR (clMagR) binds to two types of iron-sulfur clusters, [2Fe-2S] and [3Fe-4S], each conferring distinct magnetic properties, suggesting a possible mechanism for magnetic modulation of MagR via iron-sulfur cluster binding (Guo et al., 2021). Additionally, we discovered that mononuclear iron and iron-sulfur clusters can both bind to pigeon MagR, with each element playing a synergistic and indispensable role in establishing the magnetic properties of clMagR (Zhou et al., 2023). Species-specific iron and iron-sulfur cluster binding has been confirmed in multiple species, from bacteria and yeast to pigeons and humans (Ding & Clark, 2004; Landry et al., 2013; Lill et al., 2012; Lu et al., 2010; Mühlenhoff et al., 2011; Ollagnier-De-Choudens et al., 2001; Yang et al., 2015; Zhou et al., 2023). Birds are the most species-rich class of tetrapod vertebrates and exhibit remarkable long-distance navigation abilities (Feng et al., 2020; Zhang et al., 2014). Therefore, it is important to investigate the diversity and conservation of magnetoreception in birds at the molecular level.

In the current study, MagRs from the migratory European robin and homing pigeon were analyzed and compared to investigate potential differences. Despite the high sequence identity (95.5%) between the robin MagR and pigeon MagR, unexpected biochemical and biophysical differences were identified. Notably, the robin MagR demonstrated much higher iron and iron-sulfur binding capacity and existed mainly as a dimer, where the pigeon MagR primarily presented as a tetramer. Only three variations in the sequences of robin MagR and pigeon MagR were found, with two contributing to the observed difference in metal binding and oligomerization. Our study clearly demonstrated the occurrence of marked and unexpected differences in two highly conserved avian MagR proteins, which may help explain the diverse functions and different but closely related navigational strategies (migration and homing) associated with such a conserved protein.

MATERIALS AND METHODS

Protein expression and purification

The expression vectors containing the pigeon and robin MagR genes were constructed with a Strep-II tag on the N-terminal of MagR, as described previously (Qin et al., 2016; Zhou et al., 2023). The expression vectors of PtoR-K20R, PtoR-E47D, PtoR-delSAS, RtoP-R17K, RtoP-D44E, and RtoP-insSAS were obtained by site-directed mutagenesis (Tiangen, China) using robin MagR and pigeon MagR plasmid as templates in accordance with the manufacturer's instructions.

All proteins were expressed in the *Escherichia coli* strain BL21 (DE3). Bacterial cells were harvested after induction with 20 μmol/L isopropyl-D-1-thiogalactopyranoside (IPTG) at 288 K (15 °C) for 20–22 h. The cell culture was then harvested by centrifugation at 7 000 r/min for 8 min at 4 °C in 500 mL centrifuge bottles with a JA-10 rotor (Beckman Coulter, USA). To extract proteins, the cells were lysed by sonication on ice and resuspended in lysis buffer (20 mmol/L Tris, 500 mmol/L NaCl, pH 8.0). After centrifugation at 4 °C and 17 500 r/min for 45 min using a JA-20 rotor (Beckman Coulter, USA), the supernatant was collected and loaded onto a Strep-Tactin affinity column (IBA GmbH, Germany). The column was washed with approximately 20 column volumes (CV) of buffer W (20 mmol/L Tris, 500 mmol/L NaCl, pH 8.0) to remove unbound proteins. After washing, robin MagR, pigeon MagR protein, and their respective mutants were eluted from the Strep-Tactin affinity column with buffer E (20 mmol/L Tris, 150 mmol/L NaCl, 5 mmol/L d-desthiobiotin, pH 8.0). For sodium dodecyl-sulfate polyacrylamide gel electrophoresis

(SDS-PAGE), a protein ladder (PageRuler Prestained Protein Ladder, Thermo Scientific, Product# 26616, USA) was used as the molecular weight standard.

Ultraviolet-visible (UV-Vis) absorption analysis

Wild-type MagR and mutant proteins (100 μmol/L) were prepared in Buffer E (20 mmol/L Tris, 150 mmol/L NaCl, 5 mmol/L d-desthiobiotin, pH 8.0). UV-Vis spectra were recorded using a NanoDrop spectrophotometer (NanoDrop OneC, Thermo Scientific, USA) at room temperature in the near UV-Vis range (300–600 nm). Buffer E was used as the blank control.

Size‐exclusion chromatography (SEC)

Analytical SEC was used to monitor and distinguish various oligomeric states of MagR and its mutants. Purified proteins were injected into a Superdex 200 10/300 column (GE Healthcare, USA) connected to a GE AKTA FPLC purifier system at a flow rate of 0.50 mL/min. Columns were pre-equilibrated with TBS buffer (20 mmol/L Tris, 150 mmol/L NaCl, pH 8.0).

Cross-linking assay

Purified robin MagR and pigeon MagR were prepared in 10 mmol/L PBS at 100 μmol/L, then incubated with 1 mmol/L cross-linking reagent BS3 (sulfo-DSS, bis (sulfosuccinimidyl) suberate, Thermo Fisher Scientific, USA) at room temperature for 10 min. The reaction was then quenched with 100 mmol/L Tris-HCl (pH 8.0) for 15 min.

Ferrozine assay

Ferrous iron reacts with ferrozine to form an intense, purple-colored complex that can be quantified spectrophotometrically at 562 nm using a microplate reader. A series of ferric chloride (0–800 μmol/L) in 1 mol/L HCl was used to generate a standard curve. Total iron content in the MagRs and their mutant proteins was quantified by reducing iron with hydroxylamine hydrochloride (HAHCl, 10% (w/v) HAHCl in 1 mol/L HCl) and performing analysis using the ferrozine assay based on the stand curve, as described previously (Zhou et al., 2023). In brief, the protein and HAHCl mixtures (80 μL of HAHCl and 20 μL of proteins at 100 μmol/L, total 100 μL) were incubated at 37 °C for 30 min in the dark in a 96-well plate, followed by the addition of 100 μL of ferrozine reagent (0.1% (w/v) ferrozine in 50% (w/v) ammonium acetate) in each well and incubation at 37 °C for an additional 15 min in the dark. The iron-ferrozine complex was measured at 562 nm on a microplate reader (Tecan Spark, Switzerland). Total iron content in the MagRs and mutants was calculated based on the standard curve by linear regression analysis. There were three replicates for each protein sample. Histograms and statistical analyses were performed using GraphPad Prism v8.0. Student's *t*-test was used to test for differences in total iron between protein samples, with significance determined at $P<0.05$.

Circular dichroism (CD) spectroscopy

CD spectroscopy is a robust method for evaluating secondary structures in proteins in the far UV range (190–260 nm) or for monitoring protein-bound co-factors such as metals or iron-sulfur clusters in the near UV-vis range (300–600 nm) (Kelly & Price, 2000). To characterize protein-bound metals and iron-sulfur clusters, purified robin MagR and pigeon MagR (100 μmol/L) were measured using a CD MOS-500 spectrometer (Biologic, France) at room temperature with 1 cm path quartz cells. Phosphate-buffered saline (PBS) buffer (0.01 mol/L, NaCl, $Na_2HPO_4{\cdot}12H_2O$, $NaH_2PO_4{\cdot}2H_2O$, pH 7.4) was used as a blank control.

Electron paramagnetic resonance (EPR) measurement

X-band (9.6 GHz) EPR spectra of the MagR proteins were recorded using an EMX plus 10/12 spectrometer (Bruker, USA), equipped with an Oxford ESR-910 liquid helium cryostat. Briefly, 1 mmol/L oxidized proteins (as-purified) in buffer E (20 mmol/L Tris, 150 mmol/L NaCl, 5 mmol/L d-desthiobiotin, pH8.0) were mixed in a total volume of 200 μL with 50 μL of glycerol, respectively. The 1 mmol/L reduced proteins were obtained by adding 2 μL of $Na_2S_2O_4$ (1 mol/L) to the protein samples. The protein samples were then transferred into a 4 mm diameter quartz EPR tube (Wilmad 707-SQ-250 M, USA) and frozen in liquid nitrogen. The EPR signals of the oxidized and reduced proteins were recorded at different temperatures (10, 25, 45, and 60 K). The EPR conditions were: microwave frequency, 9.4 GHz; microwave power, 2 mW; modulation frequency, 100 kHz; modulation amplitude, 2 G; receive gain, 1.0×10^4.

Magnetic measurement

Protein samples, including wild-type robin MagR and pigeon MagR (in buffer E containing 20 mmol/L Tris, 150 mmol/L NaCl, 5 mmol/L D-desthiobiotin, pH 8.0) and blank control (buffer E), were lyophilized using a freezer dryer (Heto PowerDry LL3000, Thermo Scientific, USA). Magnetic measurements were performed on lyophilized samples (mass of approximately 5 mg) using a MPMS-3 magnetometer (Quantum Design, USA) equipped with a superconducting quantum interference device (SQUID) sensor at different temperatures (300 and 5 K). The fields applied were between −2 and 2 T. MH curves (magnetization (M) curves measured versus applied fields (H)) of proteins were obtained after subtracting the background from the buffer control. The volume magnetic susceptibility versus magnetic field of robin and pigeon MagRs were calculated based on the collected data, with a protein density of about 0.15 g/cm^3 (Beech et al., 2015).

RESULTS

Divergent features of robin and pigeon MagRs with highly conserved sequences

Although various navigational behaviors have been observed in animals, the underlying mechanisms remain unclear. MagR has been suggested to play an essential role in animal magnetoreception. The MagR sequence is highly conserved across species (Supplementary Figure S1), especially in birds (Figure 1A). However, the degree to which MagRs from different avian species with closely related sequences share functional characteristics is not yet known. In this study, we examined MagRs from the migratory European robin and non-migratory homing pigeon to explore their species-specific traits. The MagR sequences from the two species displayed a high degree of conservation, sharing 95.5% identity, with variations in only three regions (Figure 1A).

To determine whether the three variable regions or less conserved sites influence MagR functionality, we conducted comparative analysis of co-factor binding and oligomeric states of MagR proteins from different avian species, using the prokaryotic homologous protein IscA from *E. coli* as the control in some experiments. Strep-tagged MagRs from robin

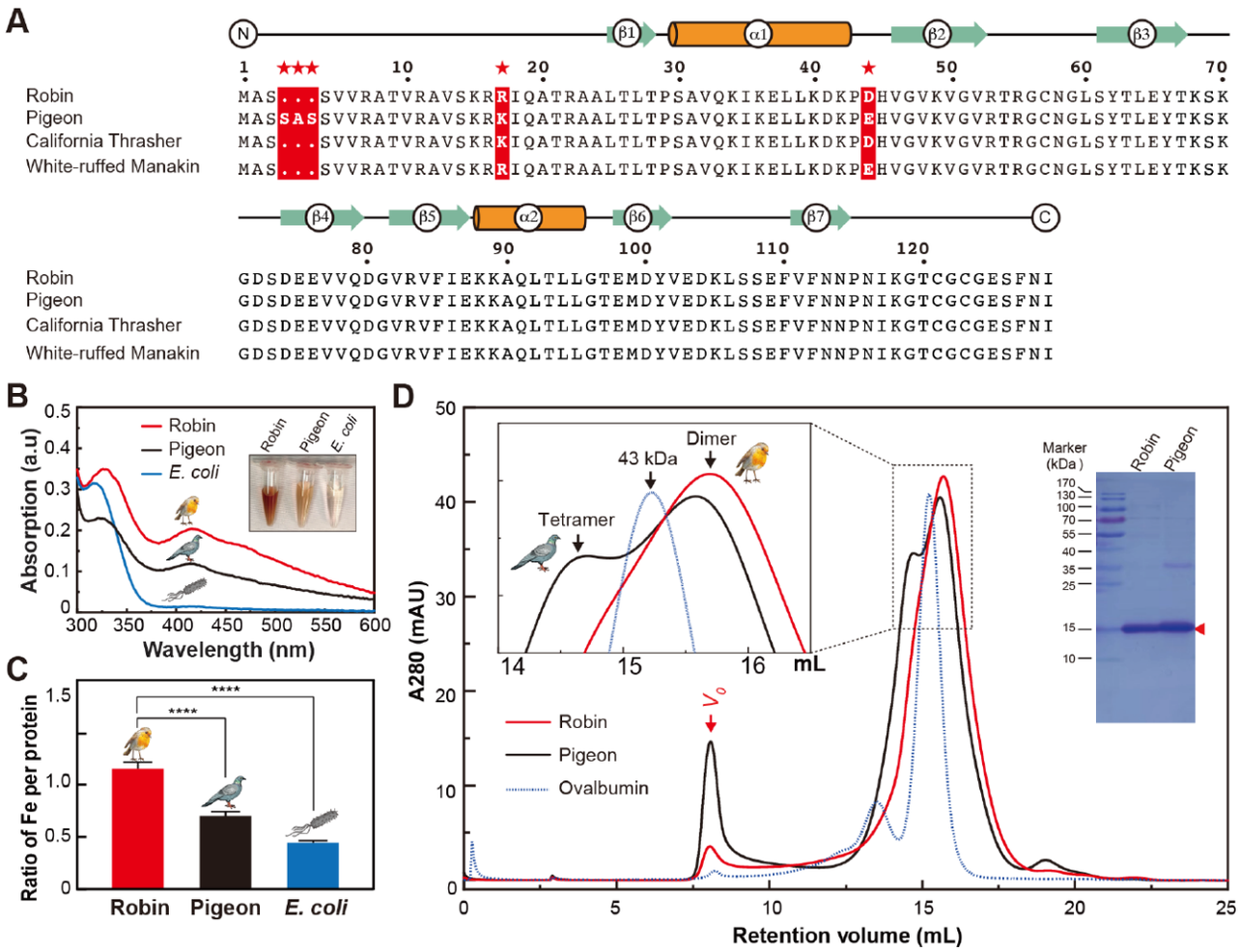


Figure 1 Differences in highly conserved avian MagRs

A: Sequence alignment of MagR from four different species of birds. Predicted secondary structures are shown in upper lines. Sequence variation sites are shown in white with a red background and marked by red stars (see Methods and Supplementary Materials for details about sequence ID). B: UV-Vis absorption spectra of robin MagR (red line), pigeon MagR (black line), and *E. coli* IscA (blue line). Insert pictures show color variation of MagR proteins in solution at a concentration of 800 μmol/L. C: Total iron content in purified robin MagR, pigeon MagR, and *E. coli* IscA measured by ferrozine assay as the ratio of iron atoms per protein monomer. Student's *t*-test, ****: $P<0.0001$, error bars represent mean±standard error of the mean (*SEM*). D: SEC of robin MagR (red line) and pigeon MagR (black line) analyzed using a Superdex 200 10/300 GL column. Ovalbumin (blue, dashed line) was used as a control to reveal elution volume of globular protein with relative molecular mass around 43 kDa. Arrows indicate voids (red) and peaks corresponding to MagR dimer and tetramer (black). SDS-PAGE of protein preparation under reducing conditions is shown as insert.

and pigeon and *E. coli* IscA were expressed and purified under aerobic conditions. UV-Vis absorbance spectra from 300 nm to 600 nm were measured to characterize iron and iron-sulfur cluster binding in MagR proteins. Significant differences in the UV-Vis profile were observed (Figure 1B). The robin and pigeon MagRs showed typical absorption peaks at 320 and 420 nm and a shoulder at 456 nm, indicating the presence of iron-sulfur cluster binding in the protein, consistent with previous reports (Guo et al., 2021; Zhou et al., 2023). In contrast, *E. coli* IscA only showed an absorption peak at 320 nm, suggesting iron binding but the absence of iron-sulfur clusters (Ding & Clark, 2004). Interestingly, robin MagR showed markedly higher UV-Vis absorption than pigeon MagR and *E. coli* IscA, in accordance with the coloration of protein solutions (as shown in Figure 1B). Freshly purified robin MagR was deep brown, pigeon MagR was light brown, and *E. coli* IscA was colorless at the same concentration. Total iron content was highest in robin MagR, followed by pigeon MagR and *E. coli* IscA, as quantified by ferrozine assay (Figure 1C).

There were notable differences in the oligomerization of MagR between the two avian species. Analytical SEC revealed that robin MagR exhibited a major peak with a retention volume of 16.0 mL, while pigeon MagR exhibited two major peaks with retention volumes of 15.02 and 15.89 mL, respectively. Comparing with ovalbumin, a reference protein with a 43 kDa mass that elutes at 15.54 mL, robin MagR was predominantly found in dimer form, whereas pigeon MagR exhibited a combination of dimers and tetramers in solution (Figure 1D), in alignment with the molecular weights of the robin and pigeon MagR monomers (16.18 and 16.41 kDa, respectively) determined by reducing SDS-PAGE. The formation of MagR dimers and tetramers was further confirmed by chemical crosslinking assay (Supplementary Figure S2).

Collectively, despite minimal sequence variations, robin MagR demonstrated a significantly greater capacity for binding iron and iron-sulfur compared to pigeon MagR and was consistently found as a dimer in solution, unlike the mixture of dimers and tetramers observed with pigeon MagR.

Pigeon MagR showed more stable iron-sulfur cluster binding compared to robin MagR

To elucidate the types and stabilities of iron-sulfur clusters in MagR across different species, the robin and pigeon MagR proteins were assessed using EPR, CD (300–600 nm), and UV-Vis absorption spectroscopy. The EPR spectra revealed identical g values for both the oxidized and reduced states of robin MagR and pigeon MagR (Supplementary Figure S3A–D), consistent with previous research (Guo et al., 2021). Correspondingly, the CD spectra of robin and pigeon MagRs showed similar signal profiles, including two positive peaks at 371 and 426 nm and three negative peaks at 324, 396, and 463 nm (Supplementary Figure S3E), indicating the presence of [2Fe-2S] and [3Fe-4S] clusters.

However, unexpected differences in the stability of iron-sulfur clusters between robin and pigeon MagRs were also observed. Freshly purified MagR proteins (200 μmol/L) were incubated at room temperature for 168 h and monitored by UV-Vis scanning every couple of hours. A reduction in the iron-sulfur cluster content in the MagR proteins was revealed by the decrease in absorption in UV-Vis spectra. As depicted in Figure 2A, B, robin MagR exhibited much higher UV-Vis absorption compared to pigeon MagR initially, indicating greater iron-sulfur cluster binding in robin MagR compared to pigeon MagR (as also seen in Figure 1A, C). During incubation, the UV-Vis absorption of both MagR proteins decreased with time, eventually reaching similar absorption levels after 168 h, reflecting the dissociation or decay of the iron-sulfur clusters. The decrease in UV-Vis absorption at 320 and 420 nm was much faster for robin MagR than for pigeon MagR, suggesting a lower stability of iron-sulfur clusters in the former, despite its higher initial cluster binding (Figure 2C, D).

Differences in robin and pigeon MagRs can be traced to two sequence variations

Marked divergence in the properties of an identical protein from two closely related species with high sequence identity is atypical. To determine the key sequence variations responsible for the differences in MagR proteins, we generated a series of substitution mutants based on robin and pigeon MagR sequence alignment to convert robin MagR to pigeon MagR (abbreviated as "RtoP", Figure 3A) and *vice versa* (abbreviated as "PtoR", Figure 4A).

The wild-type and mutant MagRs (three RtoP mutants and three PtoR mutants) were then expressed and purified to homogeneity under aerobic conditions. UV-Vis spectral analysis showed that all mutants exhibited typical absorption at 320 and 420 nm, and a shoulder at 456 nm, indicating that the mutations did not change the iron-sulfur cluster binding type in the MagR proteins (Figures 3B, 4B), as confirmed by CD spectral analysis (Figures 3E, 4E).

Two of the three "RtoP" single-site substitutions or insertions, specifically RtoP-R17K (R17 to lysine substitution) and RtoP-insSAS (SAS insertion), showed a significant decrease in UV-Vis absorption, resembling that of pigeon

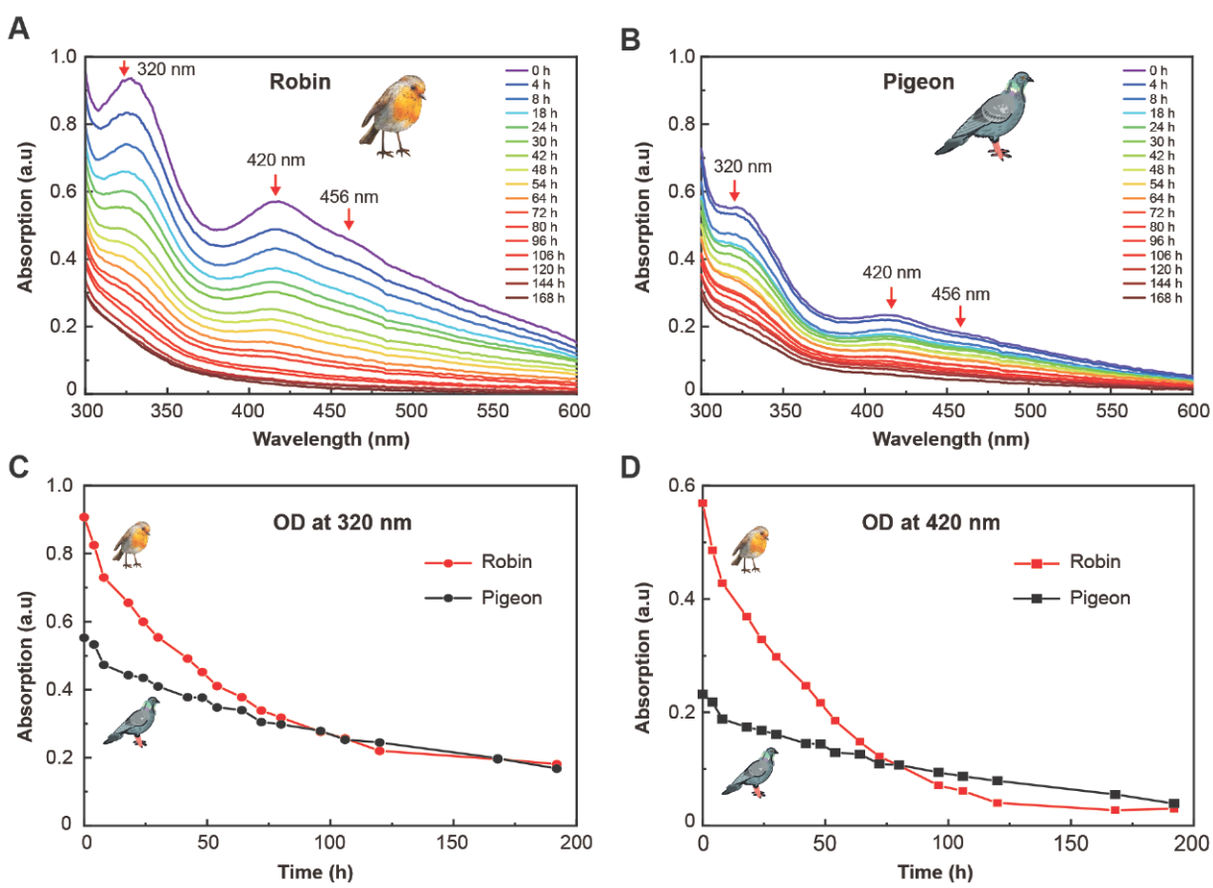


Figure 2 Stability of iron-sulfur cluster binding in robin MagR and pigeon MagR

A, B: UV-Vis absorption of purified robin MagR (A) and pigeon MagR (B) recorded at room temperature for 168 h. C, D: Decay curve of UV-Vis absorption of purified robin MagR and pigeon MagR at 320 nm (C, filled circles) and 420 nm (D, filled squares) during room temperature incubation. Data were derived from A and B.

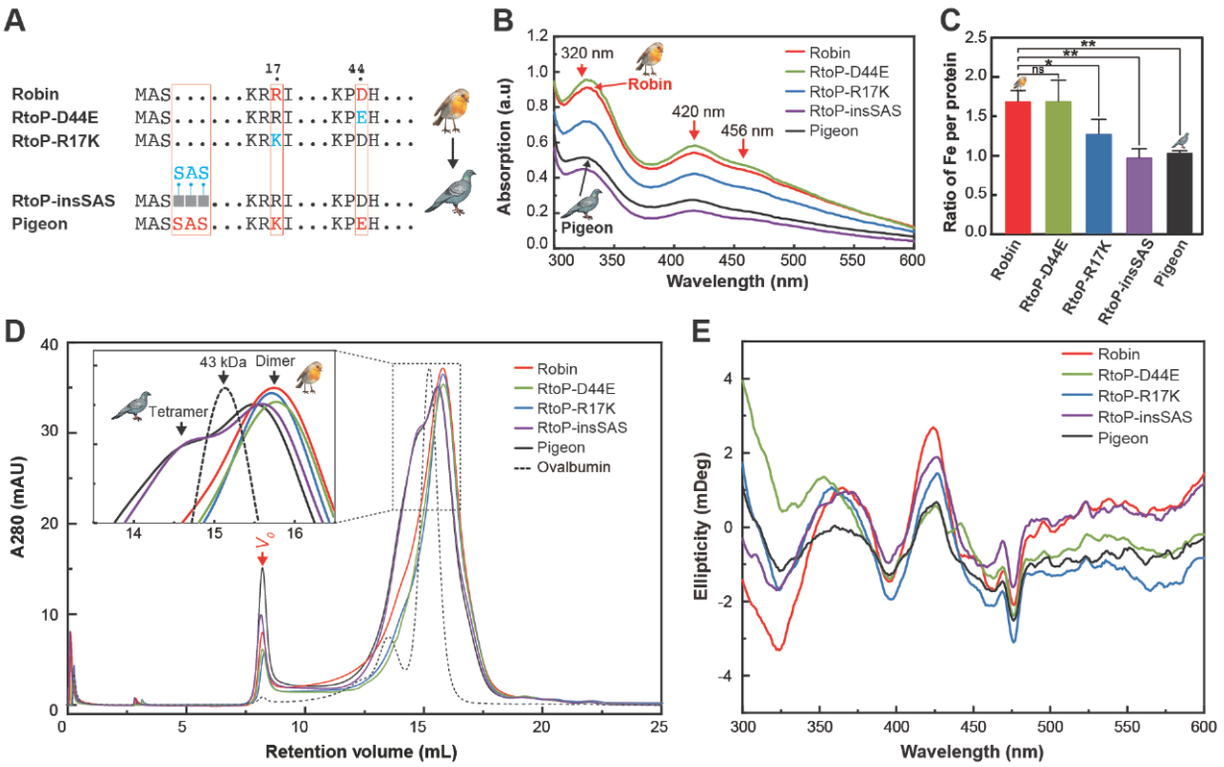


Figure 3 Differences in robin MagR and pigeon MagR are determined by two sequence variations, validated by "robin to pigeon" (RtoP) mutations

A: Robin to pigeon (RtoP) mutation design based on sequence alignment of robin MagR and pigeon MagR. Mutated amino acids are shown in blue. B: UV-Vis absorption of purified robin MagR (red line), its mutants (RtoP-D44E (green line), RtoP-R17K (blue line), and RtoP-insSAS (purple line)), and pigeon MagR (black line). C: Total iron content in purified robin MagR, its mutants (RtoP-D44E, RtoP-R17K, RtoP-insSAS), pigeon MagR, as measured by ferrozine assay. Student's *t*-test, *: $P<0.05$; **: $P<0.005$; ns: $P>0.05$, no significance. Error bars represent mean±*SEM*. D: SEC of wild-type robin MagR (red line), RtoP-D44E (green line), RtoP-R17K (blue line), RtoP-insSAS (purple line), pigeon MagR (black line), and ovalbumin (black, dashed line) analyzed on Superdex 200 10/300 GL column. Arrows indicate void (red) and two major peaks representing dimeric MagR and tetrameric MagR (black). E: CD spectra of as-purified robin MagR (red line), its mutants (RtoP-D44E (green line), RtoP-R17K (blue line), RtoP-insSAS (purple line)), and pigeon MagR (black line) measured at a concentration of 100 μmol/L.

MagR, while RtoP-D44E (D44 to glutamic acid) remained unchanged, displaying almost the same absorption as robin MagR (Figure 3B). Consistently, total iron content in RtoP-R17K and RtoP-insSAS decreased significantly compared to robin MagR but was approximately equivalent to pigeon MagR, whereas RtoP-D44E remained unchanged, akin to robin MagR (Figure 3C). Similar trends were observed in the oligomerization of the two wild-type MagRs and three RtoP mutants (Figure 3D), especially the "RtoP" mutant RtoP-insSAS (purple line in Figure 3D), which showed an almost identical oligomeric state as pigeon MagR. Thus, these findings suggested the N-terminal sequence insertion SAS and R17 in robin MagR (corresponding to K20 in pigeon) may play an active role in iron and iron-sulfur binding, and the SAS site may also contribute to tetramer formation and the modulation of higher order assembly in MagR.

Similarly, mutations were engineered to convert pigeon MagR to robin MagR (herein named "PtoR") to verify which sequence variations account for the observed differences (Figure 4A). Mutants PtoR-delSAS (deletion of N-terminal sequence SAS from pigeon sequence) and PtoR-K20R (substitution of K20 in pigeon sequence to R as in robin sequence) both exhibited significantly increased UV-Vis absorption (Figure 4B) and higher iron content (Figure 4C) compared to the wild-type pigeon MagR, closely resembling the pattern observed in robin MagR. Deletion of the N-terminal SAS sequence from pigeon MagR successfully converted the oligomeric state to a dimeric state, almost identical to that of robin MagR (Figure 4D). These results suggest that the N-terminal sequence SAS and K20 in pigeon (corresponding to R17 in robin) may play active roles in iron and iron-sulfur binding, and the SAS site may also be involved in tetramer formation and the modulation of higher order assembly in MagR.

Robin MagR and pigeon MagR show different magnetic properties

As described above, robin MagR contained higher iron and iron-sulfur content than pigeon MagR. To explore whether differences in iron and iron-sulfur content influence magnetic properties of MagR, the magnetic moments of robin MagR and pigeon MagR were measured using SQUID (MPMS3) magnetometry at varying temperatures (300 and 5 K). MH curves were then generated to compare the magnetic susceptibility of both proteins (Figure 5A, B). The MH curves of both MagRs exhibited similar diamagnetic properties at 300 K and paramagnetic properties at 5 K, although robin MagR demonstrated slightly stronger paramagnetic effects than pigeon MagR. Overall, the magnetic susceptibility of robin MagR was approximately 2.6 times that of pigeon MagR.

To compare the magnetic susceptibilities of MagR to the known volume magnetic susceptibilities of typical tissues

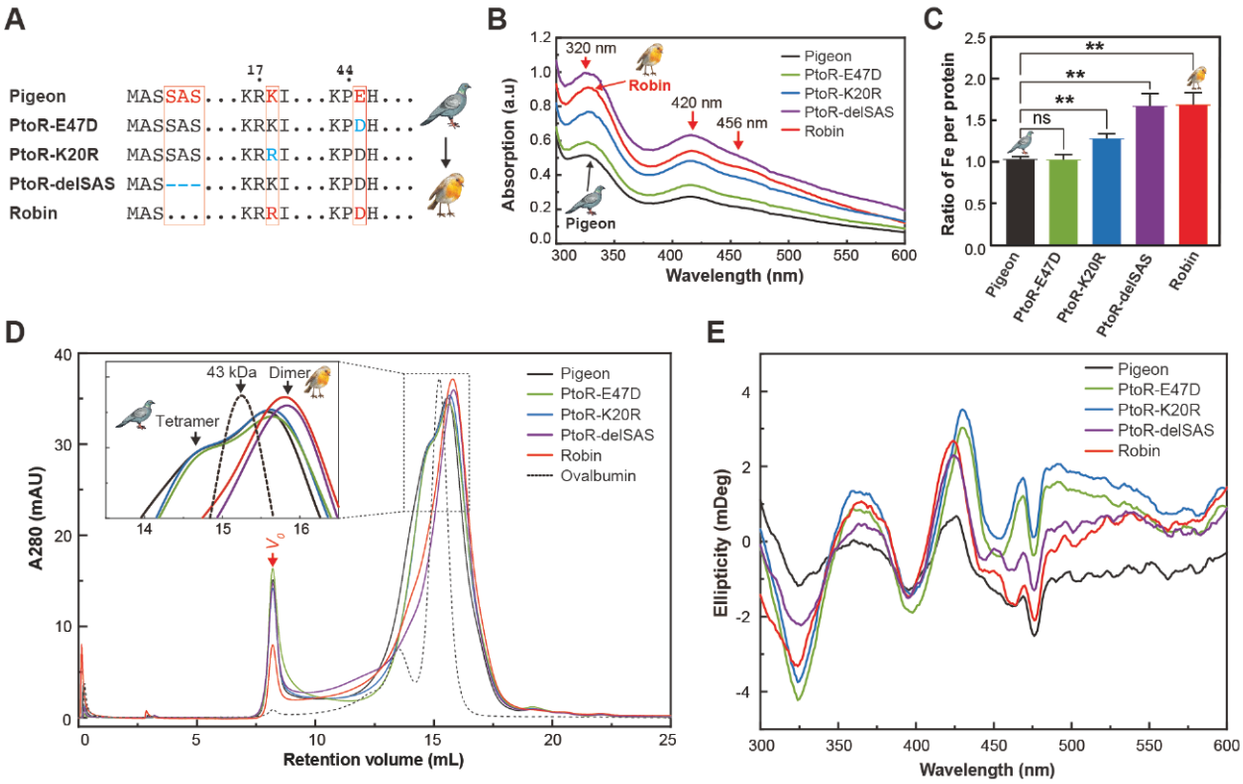


Figure 4 Differences in robin MagR and pigeon MagR are determined by two sequence variations, as validated by "pigeon to robin" (PtoR) mutations

A: Pigeon to robin (PtoR) mutation design based on sequence alignment of pigeon and robin MagR. Mutated amino acids are shown in blue. B: UV-Vis absorption of purified pigeon MagR (black line), its mutants (PtoR-E47D (green line), PtoR-K20R (blue line), and PtoR-delSAS (purple line)), and robin MagR (red line). C: Total iron content in purified pigeon MagR, its mutants (PtoR-E47D, PtoR-K20R, and PtoR-delSAS), and robin MagR, as measured by ferrozine assay. Student's *t*-test, **: $P<0.005$; ns: $P>0.05$, no significance. Error bars represent mean±*SEM*. D: SEC of pigeon MagR (black line), PtoR-E47D (green line), PtoR-K20R (blue line), PtoR-delSAS (purple line), robin MagR (red line), and ovalbumin (black, dashed line) analyzed on Superdex 200 10/300 GL column. Arrows indicate void (red) and two major peaks corresponding to tetrameric and dimeric MagR (black). E: CD spectra of as-purified pigeon MagR (black line), its mutants (PtoR-E47D (green line), PtoR-K20R (blue line), and PtoR-delSAS (purple line)), and robin MagR (red line) measured at a concentration of 100 μmol/L.

(Klohs & Hirt, 2021), we converted mass magnetic susceptibilities in CGS units of MagR to volume magnetic susceptibilities (Figure 5C, D). The volume magnetic susceptibilities of robin MagR and pigeon MagR were -1.4204×10^{-6} and -1.2808×10^{-6}, respectively, at 300 K, which are more paramagnetic than those of heart and brain tissues (-1.932×10^{-5} and -6.928×10^{-6}, respectively).

DISCUSSION

Navigation in animals, the innate ability of diverse species to find their way accurately without the aid of maps or tools, is an evolutionary adaptation to spatiotemporal fluctuations in resources (Dingle & Drake, 2007). However, when examining various species, there is a notable variation in navigational behaviors, even among members of the same taxonomic category. Various insect species may engage in annual migrations or multiple migrations within a breeding season across generations, while other species, such as honeybees and pigeons, adopt very different homing strategies. Additionally, some animals undertake round-trip migrations, whereas others embark on "one-way" migrations, often triggered by habitat deterioration or food scarcity.

Animal navigation usually manifests as either migration or homing, with both processes involving the movement of organisms within their environment. Migration typically refers to the regular, annual, or seasonal mass movements of animals that cycle between breeding grounds and wintering areas (Greenberg & Marra, 2005). In contrast, homing is the inherent ability of an animal (e.g., honeybees and pigeons) to return to a specific starting point from various locations, often navigating through unfamiliar territories over considerable distances. This considerable diversity in navigation patterns has led to speculation about whether a universal mechanism or magnetoreceptor is present in animals across different phyla.

MagR, along with its prokaryotic homologous protein IscA, is one of the most highly conserved and abundant proteins across all organisms, ranging from bacteria and yeasts to plants and animals. Upon its initial discovery and potential involvement in magnetoreception, one question was naturally raised: How might such a highly conserved protein contribute to the vast array of animal behaviors related to magnetoreception (including both migratory and residential species) and different navigation patterns? Therefore, in this study, we examined MagRs from two closely related avian species, the migratory robin and the homing pigeon, to elucidate the potential differences between their respective proteins.

The sequence identity between robin MagR and pigeon MagR reached 95.5% (with a similarity of 97%), with only

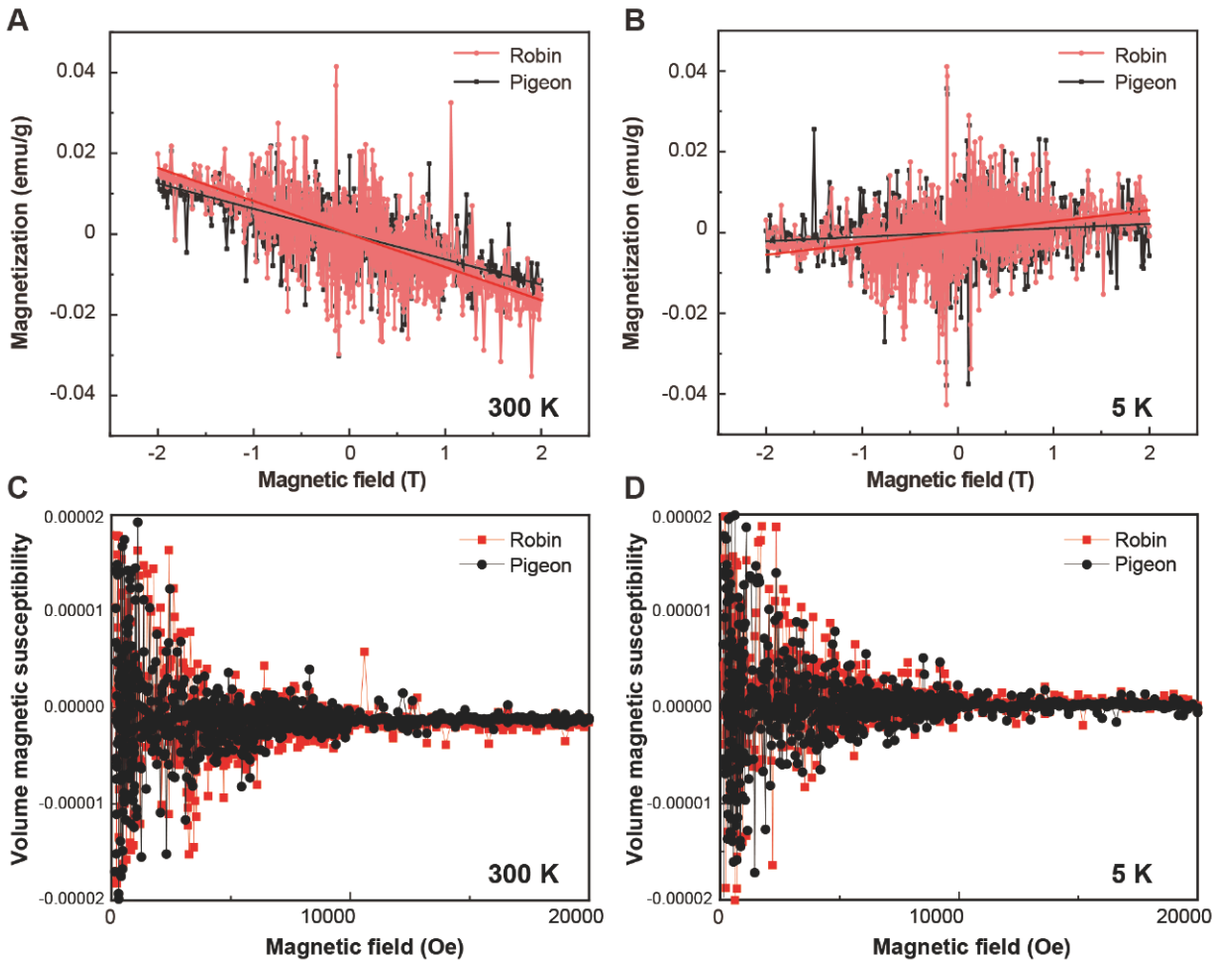


Figure 5 Robin MagR and pigeon MagR showed different magnetic properties

A: Field-dependent magnetization curves (MH) at 300 K for pigeon MagR (black) and robin MagR (red). Magnetic susceptibilities of pigeon MagR and robin MagR were –6.26395e-7 and –8.17783e-7, respectively. B: Field-dependent magnetization curves (MH) at 5 K for pigeon MagR (black) and robin MagR (red). Magnetic susceptibilities of pigeon MagR and robin MagR were 1.06894e-7 and 2.75509e-7, respectively. C, D: Volume magnetic susceptibility versus magnetic field of robin MagR (red) and pigeon MagR (black) at 300 K (C) and 5 K (D).

three sequence variations identified through alignment. Notable differences in the biochemical and biophysical properties of the two proteins were observed, including variations in iron and iron-sulfur binding capacities, oligomeric states, and magnetic features. These differences were attributed to the small sequence variations through site-directed mutagenesis, in which the robin-type was converted to the pigeon-type, and *vice versa*. The pronounced disparities in the MagR proteins between these two species prompted a reconsideration of the initial question: Can a protein with such conserved sequences be integral in both migratory and residential species, and potentially also be involved in diverse navigational behaviors? Based on the findings of this study, we propose that this is indeed plausible.

Proteins adapt through evolution to perform specific biological functions, with these functions shaping the divergence in their sequences. MagR follows this evolutionary pattern. Initially identified as a scaffolding protein for iron-sulfur cluster assembly and implicated in mitochondrial dysfunction syndrome in humans (Bilder et al., 2004; Cupp-Vickery et al., 2004; Jacobson et al., 1989; Ollagnier-De-Choudens et al., 2001), MagR was subsequently proposed as a putative magnetoreceptor upon the discovery of its interaction with cryptochrome (Cry). Both the redox potential involved in electron transportation and magnetic properties relevant to animal magnetoreception may have influenced the evolutionary trajectory of MagR. Our findings indicated that the variable sequences or less conserved regions in the proteins evolved slowly, which may not directly impact primary protein function, but may exert significant influence on protein structure and stability. Consequently, these regions may contribute to the diversity of physiological functions across organisms, including animal magnetoreception and navigation patterns.

Several significant directions for future research are apparent. First, it is worth investigating whether the observed differences between MagRs from migratory species such as robins and homing species such as pigeons are reflective of their distinct migratory and homing behaviors. It is also possible that robin MagR and pigeon MagR may function similarly in magnetoreception and navigation, with the differing properties observed in this study being related to other MagR functions, such as iron-sulfur cluster biogenesis and electron transfer reactions. Therefore, the link between the biochemical properties of MagR and the diversity of navigational patterns remains an open question. Second, a comprehensive and comparative study of MagRs from a wide range of species across different phyla could be crucial in understanding the

biochemical and biophysical diversification of this ancient and highly conserved protein. Such research could illuminate the ways in which speciation and functional selection pressures for magnetoreception/navigation have driven sequence divergence and evolution of MagR, greatly enhancing our comprehension of the mechanisms underlying animal magnetoreception and navigation.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

C.X. and T.C. conceived the idea and designed the study. S.W. carried out protein purification, site-directed mutagenesis, UV-vis measurements, ferrozine assay, CD spectroscopy, and EPR experiments. S.W. and C.X. performed data analysis. P.Z., F.F., Y.H., and L.Z. helped with SQUID measurement and data analysis. F.F. helped with size-exclusion chromatography data analysis. P.Z. and T.T. helped with EPR experiments. T.T., Y.Z., and X.Z. contributed to the ferrozine assay data analysis. Y.Z. and T.T. helped with CD measurements. J.Z., X.Z., M.W., and Y.Z. provided valuable suggestions on data analysis. S.W and C.X. wrote the paper. T.C. and C.X. revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank the Steady High Magnetic Field Facilities (High Magnetic Field Laboratory, CAS) for assistance with EPR measurements, and thank Dr. Wei Tong and Jinxing Li for their technical support. We are grateful to the core facilities of the College of Life Sciences of the University of Science and Technology of China and College of Life Sciences of Anhui University for their help in CD spectral measurements. We also thank Huangtao Xu for discussions.

REFERENCES

Arai S, Shimizu R, Adachi M, et al. 2022. Magnetic field effects on the structure and molecular behavior of pigeon iron–sulfur protein. *Protein Science*, **31**(6): e4313.

Beech KE, Biddlecombe JG, Van Der Walle CF, et al. 2015. Insights into the influence of the cooling profile on the reconstitution times of amorphous lyophilized protein formulations. *European Journal of Pharmaceutics and Biopharmaceutics*, **96**: 247−254.

Bilder PW, Ding HG, Newcomer ME. 2004. Crystal structure of the ancient, Fe-S scaffold IscA reveals a novel protein fold. *Biochemistry*, **43**(1): 133−139.

Cao YS, Yan P. 2018. Role of atomic spin-mechanical coupling in the problem of a magnetic biocompass. *Physical Review E*, **97**(4): 042409.

Chang H, Fu XW, Zhao SY, et al. 2017. Molecular characterization, tissue, and developmental expression profiles of *MagR* and *Cryptochrome* genes in *Agrotis ipsilon* (Lepidoptera: Noctuidae). *Annals of the Entomological Society of America*, **110**(4): 422−432.

Chang H, Guo JL, Fu XW, et al. 2018. Molecular characterization and expression profiles of *IscA1* gene in a long-distance migrant. *Agrotis segetum*. *Journal of Asia-Pacific Entomology*, **21**(4): 1299−1306.

Cheng Q, Sun J, Ge Y, et al. 2023. Bionic magnetic sensor based on the MagR/Cry4 complex-configured graphene transistor with an integrated on-chip gate. *ACS Sens*, **8**(2): 793−802.

Cózar-Castellano I, Del Valle Machargo M, Trujillo E, et al. 2004. hIscA: a protein implicated in the biogenesis of iron-sulfur clusters. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, **1700**(2): 179−188.

Cupp-Vickery JR, Silberg JJ, Ta DT, et al. 2004. Crystal structure of IscA, an iron-sulfur cluster assembly protein from *Escherichia coli*. *Journal of Molecular Biology*, **338**(1): 127–137.

Ding HE, Clark RJ. 2004. Characterization of iron binding in IscA, an ancient iron-sulphur cluster assembly protein. *Biochemical Journal*, **379**(2): 433−440.

Dingle H, Drake VA. 2007. What is migration?. *BioScience*, **57**(2): 113−121.

Egevang C, Stenhouse IJ, Phillips RA, et al. 2010. Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(5): 2078−2081.

Feng SH, Stiller J, Deng Y, et al. 2020. Dense sampling of bird diversity increases power of comparative genomics. *Nature*, **587**(7833): 252−257.

Greenberg R, Marra PP. 2005. Birds of Two Worlds: The Ecology and Evolution of Migration. Baltimore: Johns Hopkins University Press.

Guo Z, Xu S, Chen X, et al. 2021. Modulation of MagR magnetic properties via iron–sulfur cluster binding. *Scientific Reports*, **11**(1): 23941.

Hore PJ, Mouritsen H. 2016. The radical-pair mechanism of magnetoreception. *Annual Review of Biophysics*, **45**: 299−344.

Jacobson MR, Brigle KE, Bennett LT, et al. 1989. Physical and genetic map of the major *nif* gene cluster from *Azotobacter vinelandii*. *Journal of Bacteriology*, **171**(2): 1017–1027.

Jiang M, Zhang LJ, Wang FQ, et al. 2017. Novel application of magnetic protein: convenient one-step purification and immobilization of proteins. *Scientific Reports*, **7**(1): 13329.

Jin MH, Liu B, Zheng WG, et al. 2023. Chromosome-level genome of black cutworm provides novel insights into polyphagy and seasonal migration in insects. *BMC Biology*, **21**(1): 2.

Johnsen S, Lohmann KJ. 2008. Magnetoreception in animals. *Physics Today*, **61**(3): 29−35.

Kang J, Kang D, Yeom G, et al. 2021. Molecular diagnostic system using engineered fusion protein-conjugated magnetic nanoparticles. *Analytical Chemistry*, **93**(50): 16804−16812.

Kelly SM, Price NC. 2000. The use of circular dichroism in the investigation of protein structure and function. *Current Protein & Peptide Science*, **1**(4): 349−384.

Kentie R, Morgan Brown J, Camphuysen KCJ, et al. 2023. Distance doesn't matter: migration strategy in a seabird has no effect on survival or reproduction. *Proceedings of the Royal Society B:Biological Sciences*, **290**(1997): 20222408.

Kirschvink JL, Walker MM, Diebel CE. 2001. Magnetite-based magnetoreception. *Current Opinion in Neurobiology*, **11**(4): 462−467.

Klohs J, Hirt AM. 2021. Investigation of the magnetic susceptibility properties of fresh and fixed mouse heart, liver, skeletal muscle and brain tissue. *Physica Medica*, **88**: 37−44.

Landry AP, Cheng ZS, Ding HE. 2013. Iron binding activity is essential for the function of IscA in iron-sulphur cluster biogenesis. *Dalton Transactions*, **42**(9): 3100−3106.

Li N, Wang P, Xie YY, et al. 2023. Expression of clMagR/clCry4 protein in mBMSCs provides T_2-contrast enhancement of MRI. *Acta Biomaterialia*, doi:https://doi.org/10.1016/j.actbio.2023.09.039.

Lill R, Hoffmann B, Molik S, et al. 2012. The role of mitochondria in cellular iron-sulfur protein biogenesis and iron metabolism. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, **1823**(9): 1491−1508.

Lohmann KJ. 2016. A candidate magnetoreceptor. *Nature Materials*, **15**(2): 136−138.

Lu JX, Bitoun JP, Tan GQ, et al. 2010. Iron-binding activity of human iron-sulfur cluster assembly protein hIscA1. *Biochemical Journal*, **428**(1): 125−131.

Maeda K, Henbest KB, Cintolesi F, et al. 2008. Chemical compass model of avian magnetoreception. *Nature*, **453**(7193): 387−390.

Mannino G, Casacci LP, Bianco Dolino G, et al. 2023. The geomagnetic

field (GMF) is necessary for black garden ant (*Lasius niger* L.) foraging and modulates orientation potentially through aminergic regulation and *MagR* expression. *International Journal of Molecular Sciences*, **24**(5): 4387.

Meade J, Biro D, Guilford T. 2005. Homing pigeons develop local route stereotypy. *Proceedings of the Royal Society B:Biological Sciences*, **272**(1558): 17−23.

Meister M. 2016. Physical limits to magnetogenetics. *eLife*, **5**: e17210.

Mouritsen H. 2018. Long-distance navigation and magnetoreception in migratory animals. *Nature*, **558**(7708): 50−59.

Mühlenhoff U, Richter N, Pines O, et al. 2011. Specialized function of yeast Isa1 and Isa2 proteins in the maturation of mitochondrial [4Fe-4S] proteins. *Journal of Biological Chemistry*, **286**(48): 41205−41216.

Nimpf S, Keays DA. 2022. Myths in magnetosensation. *iScience*, **25**(6): 104454.

Ollagnier-De-Choudens S, Mattioli T, Takahashi Y, et al. 2001. Iron-sulfur cluster assembly: characterization of IscA and evidence for a specific and functional complex with ferredoxin. *Journal of Biological Chemistry*, **276**(25): 22604−22607.

Qin SY, Yin H, Yang CL, et al. 2016. A magnetic protein biocompass. *Nature Materials*, **15**(2): 217−226.

Rodgers CT, Hore PJ. 2009. Chemical magnetoreception in birds: the radical pair mechanism. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(2): 353−360.

Tong DD, Zhang L, Wu NN, et al. 2022. The oriental armyworm genome yields insights into the long-distance migration of noctuid moths. *Cell Reports*, **41**(12): 111843.

Wang L, Xu HX, Liu ZW, et al. 2019. Magnetic immobilization of a quorum sensing signal hydrolase, AiiA. *Microbiologyopen*, **8**(8): e00797.

Wiltschko R, Wiltschko W. 2019. Magnetoreception in birds. *Journal of The Royal Society Interface*, **16**(158): 20190295.

Xiao DW, Hu WH, Cai YF, et al. 2020. Magnetic noise enabled biocompass. *Physical Review Letters*, **124**(12): 128101.

Xie C. 2022. Searching for unity in diversity of animal magnetoreception: from biology to quantum mechanics and back. *The Innovation*, **3**(3): 100229.

Xu JJ, Jarocha LE, Zollitsch T, et al. 2021. Magnetic sensitivity of cryptochrome 4 from a migratory songbird. *Nature*, **594**(7864): 535−540.

Xue L, Hu T, Guo Z, et al. 2020. A novel biomimetic magnetosensor based on magneto‐optically involved conformational variation of MagR/Cry4 complex. *Advanced Electronic Materials*, **6**(4): 1901168.

Yang J, Tan GQ, Zhang T, et al. 2015. Deletion of the proposed iron chaperones IscA/SufA results in accumulation of a red intermediate cysteine desulfurase IscS in *Escherichia coli*. *Journal of Biological Chemistry*, **290**(22): 14226–14234.

Zhang GJ, Li C, Li QY, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, **346**(6215): 1311−1320.

Zhao X, Chen HB, Lu LH, et al. 2022. A mechanism of compass-free migratory navigation. *Journal of Physics D:Applied Physics*, **55**(24): 245004.

Zhou YJ, Tong TY, Wei MK, et al. 2023. Towards magnetism in pigeon MagR: iron- and iron-sulfur binding work indispensably and synergistically. *Zoological Research*, **44**(1): 142−152.

Zhu et al. *Zool. Res.* 2024, 45(1): 79−94
https://doi.org/10.24272/j.issn.2095-8137.2023.022

ZR | Zoological Research

Article Open Access

Loss of *LBP* triggers lipid metabolic disorder through H3K27 acetylation-mediated C/EBPβ-*SCD* activation in non-alcoholic fatty liver disease

Ya-Ling Zhu[1,2,#], Lei-Lei Meng[1,#], Jin-Hu Ma[1,#], Xin Yuan[1], Shu-Wen Chen[1], Xin-Rui Yi[1], Xin-Yu Li[1], Yi Wang[1], Yun-Shu Tang[1,2], Min Xue[1], Mei-Zi Zhu[1], Jin Peng[1], Xue-Jin Lu[1], Jian-Zhen Huang[4], Zi-Chen Song[1], Chong Wu[1], Ke-Zhong Zheng[3], Qing-Qing Dai[3], Fan Huang[3,*], Hao-Shu Fang[1,2,*]

[1] *Department of Pathophysiology*, *Anhui Medical University*, *Hefei*, *Anhui* 230032, *China*
[2] *Laboratory Animal Research Center*, *School of Basic Medical Sciences*, *Anhui Medical University*, *Hefei*, *Anhui* 230032, *China*
[3] *Department of Hepatobiliary Surgery*, *First Affiliated Hospital of Anhui Medical University*, *Hefei*, *Anhui* 230022, *China*
[4] *College of Animal Science and Technology*, *Jiangxi Agricultural University*, *Nanchang*, *Jiangxi* 330045, *China*

ABSTRACT

Non-alcoholic fatty liver disease (NAFLD) is associated with mutations in lipopolysaccharide-binding protein (*LBP*), but the underlying epigenetic mechanisms remain understudied. Herein, *LBP*$^{-/-}$ rats with NAFLD were established and used to conduct integrative targeting-active enhancer histone H3 lysine 27 acetylation (H3K27ac) chromatin immunoprecipitation coupled with high-throughput and transcriptomic sequencing analysis to explore the potential epigenetic pathomechanisms of active enhancers of NAFLD exacerbation upon *LBP* deficiency. Notably, *LBP*$^{-/-}$ reduced the inflammatory response but markedly aggravated high-fat diet (HFD)-induced NAFLD in rats, with pronounced alterations in the histone acetylome and regulatory transcriptome. In total, 1 128 differential enhancer-target genes significantly enriched in cholesterol and fatty acid metabolism were identified between wild-type (WT) and *LBP*$^{-/-}$ NAFLD rats. Based on integrative analysis, CCAAT/enhancer-binding protein β (C/EBPβ) was identified as a pivotal transcription factor (TF) and contributor to dysregulated histone acetylome H3K27ac, and the lipid metabolism gene *SCD* was identified as a downstream effector exacerbating NAFLD. This study not only broadens our understanding of the essential role of *LBP* in the pathogenesis of NAFLD from an epigenetics perspective but also identifies key TF C/EBPβ and functional gene *SCD* as potential regulators and therapeutic targets.

Keywords: Non-alcoholic fatty liver disease; C/EBPβ; Lipopolysaccharide-binding protein; H3K27ac; Integrative analysis; Enhancer

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

INTRODUCTION

Non-alcoholic fatty liver disease (NAFLD), characterized by ectopic lipid accumulation in the liver, is the most prevalent chronic liver disease worldwide, affecting almost 25% of the global population (Friedman et al., 2018; Samuel & Shulman, 2018; Younossi et al., 2019). Etiologically, NAFLD demonstrates a close correlation with systemic metabolic disorders, such as obesity, insulin resistance, dyslipidemia, and hyperglycemia (Yu et al., 2019), in addition to certain genetic polymorphisms. Notably, several gene single-nucleotide polymorphisms, including Semaphorin 7A (*SEMA7A*) (Zhao et al., 2022a) and Abhydrolase domain containing 5 (*ABHD5*) (Youssefian et al., 2019), have been reported to increase susceptibility to lipid metabolic disorders in human NAFLD.

Lipopolysaccharide-binding protein (*LBP*), which is primarily produced in the hepatocytes (Zhou et al., 2016), plays a pivotal role in the inflammatory response (Pfalzgraff & Weindl, 2019; Song et al., 2021). Emerging evidence has also shown that *LBP* regulates adipogenesis, affecting hepatic lipid metabolism, and is highly associated with various metabolic

Received: 21 May 2023; Accepted: 24 August 2023; Online: 25 August 2023

Foundation items: This study was supported by the National Natural Science Foundation of China (81971875, 82300661), Natural Science Foundation of Anhui province (2308085QH246), Natural Science Foundation of the Anhui Higher Education Institutions (KJ2021A0205), Basic and Clinical Cooperative Research Program of Anhui Medical University (2019xkjT002, 2019xkjT022, 2022xkjT013), Talent Training Program, School of Basic Medical Sciences, Anhui Medical University (2022YPJH102), and National College Students Innovation and Entrepreneurship Training Program of China (202210366024)

#Authors contributed equally to this work

*Corresponding authors, E-mail: huang_f@vip.126.com; fanghaoshu@ahmu.edu.cn

disorders, including obesity (Moreno-Navarrete et al., 2015; Sun et al., 2010). Of note, patients harboring *LBP* mutations also exhibit high susceptibility to metabolic disorders (Meng et al., 2021; Takeuchi et al., 2007), raising the question of the mechanisms by which *LBP* influences the metabolic dysregulation observed in NAFLD, especially in the context of epigenetic processes.

Various studies have explored the epigenetic mechanisms related to the etiology and pathogenesis of disease (Feinberg, 2007; Ospelt, 2022). As a primary type of epigenetic modification and post-translational modification, histone H3 lysine 27 acetylation (H3K27ac) is widely recognized as a marker of active enhancers and a contributor to NAFLD pathogenesis (Rada-Iglesias et al., 2011; Smith & Shilatifard, 2014; Zhao et al., 2018; Zhu et al., 2021). For instance, Becares et al. (2019) demonstrated that liver X receptor alpha can alter H3K27ac and affect NAFLD development. Liu et al. (2018) revealed that Snail1 can mediate the deacetylation of H3K27 and prevent NAFLD. Hu et al. (2022) found that H3K27ac can activate long non-coding RNA (lncRNA) *NEAT1* and promote hepatic lipid accumulation in NAFLD. However, despite considerable efforts, the epigenetic mechanisms involved in the pathogenesis of NAFLD under *LBP* deficiency are poorly defined.

In this study, we explored the epigenomic H3K27ac and transcriptomic alterations induced by *LBP* deficiency in an experimental NAFLD model, resulting in outcomes analogous to those observed in clinical NAFLD cases with *LBP* mutation. We employed integrative epigenome H3K27ac chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq) and transcriptome high-throughput RNA sequencing (RNA-Seq). Results revealed widespread NAFLD-associated histone acetylome and transcriptome changes and further identified a core transcription factor (TF) involved in the regulation and exacerbation of lipid metabolic disorder progression in NAFLD rats with *LBP* deletion. Overall, our study provided genome-wide profiling of histone acetylome and transcriptome signals underlying NAFLD pathogenesis and elucidated the role of *LBP* deficiency in accelerating NAFLD from an epigenetic perspective. These findings provide novel insights and strategies for the potential prevention and treatment of NAFLD in patients harboring *LBP* loss-of-function.

MATERIALS AND METHODS

Ethics statement

All applicable international, national, and/or institutional guidelines for the care and use of animals were strictly followed. All animal collection protocols complied with the current laws of China. All animal protocols were approved by the Ethics Committee of Anhui Medical University (Approval No. SYXK 2017-006).

Animals and treatment

$LBP^{+/+}$ (wild-type (WT)) and $LBP^{-/-}$ rats and $LBP^{+/+}$ (WT) and $LBP^{-/-}$ mice were generated at the Nanjing Biomedical Research Institute of Nanjing University, as described previously (Song et al., 2021). Male WT (n=12) and $LBP^{-/-}$ (n=12) rats aged 6 weeks and male WT (n=12) and $LBP^{-/-}$ (n=12) mice aged 8 weeks were used in this study. Animals were housed under specific pathogen-free (SPF) conditions under a 12-h light/12-h dark cycle at a temperature of 24±2 °C and humidity of 50%±10%. The NAFLD models were established by inducing rats and mice with a high-fat diet (HFD; 20.2% protein, 45.4% fat, and 34.5% carbohydrates) (XTHF45-1, Xietong Pharma, China) for 8 and 24 weeks, respectively. Rats and mice fed a normal diet (ND; 20.6% protein, 12% fat, and 67.4% carbohydrates) (SWS9102, Xietong Pharma, China) were used as controls. The animals were randomly divided into four groups: WT_ND, $LBP^{-/-}$_ND, WT_HFD, and $LBP^{-/-}$_HFD groups. At the end of the experiments, body and liver weights were determined and serum levels of lipids, including triglycerides (TG) and total cholesterol (TC), and alanine amino transferase (ALT) and aspartate amino transferase (AST) were measured using a Hitachi 3100 Automatic Analyzer (Japan) according to the manufacturer's protocols.

Histological analysis and staining

All histological analyses were performed by Servicebio Technology (China). Briefly, fresh liver tissues were dissected and fixed in 10% neutral buffered formation, with paraffin sections then stained with a Hematoxylin and Eosin (H&E) Staining Kit (Beyotime, China). The histological samples were observed and imaged using a light microscope (ECLIPSE 80i, Nikon, Japan). To calculate the NAFLD activity score (NAS), H&E images were analyzed according to previously described criteria (Brunt et al., 2011). NAS is defined as the sum of the scores of three components: steatosis (0–3), lobular inflammation (0–3), and hepatocyte ballooning (0–2).

Reverse transcription-quantitative real-time polymerase chain reaction (RT-qPCR)

To validate mRNA expression levels, total RNA was extracted using Trizol reagent (Invitrogen, USA) and reverse-transcribed into cDNA using a Reverse Transcription Kit (Vazyme Biotech, China). A SYBR Green Premix Kit (Vazyme Biotech, China) was then used for RT-qPCR amplification, pre-denaturing cDNA at 95 °C for 10 min, followed by 40 cycles at 95 °C and 60 °C for 1 min. The relative mRNA expression levels of the target genes were normalized to *β-actin* expression using the $2^{-\Delta\Delta Ct}$ method. The primers used for RT-qPCR analysis are listed in Supplementary Table S1.

Cell culture and cell model of lipid accumulation

The hepatocyte cell line BRL-3A was purchased from the Type Culture Collection of the Chinese Academy of Sciences (Beijing, China). Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% fetal bovine serum (FBS) and placed in a humidified atmosphere with 5% CO_2 at 37 °C. After corresponding pretreatment and transfection with indicated plasmids, 0.5 mmol/L palmitic acid (PA) or 0.2 mmol/L oleic acid (OA) was added to the medium to establish an *in vitro* model of lipid accumulation in BRL-3A hepatocytes. For intracellular lipid accumulation examination, cells were fixed with 4% paraformaldehyde for 20 min and stained with 60% Oil Red O (ORO) (O1391; Sigma-Aldrich, USA) working solution.

Exploration of tissue and cellular localization of *LBP*

To explore the tissue localization of *LBP* in humans, mice, and rats, *LBP* expression in different tissues of these species was analyzed using the PRJEB4337, PRJNA66167, and PRJNA238328 datasets, respectively, from the NCBI public database (http://www.ncbi.nlm.nih.gov). The localization of *LBP* in different cell types was obtained from the Single Cell

Type database in the Human Protein Atlas (https://www.proteinatlas.org/).

High-throughput RNA sequencing (RNA-Seq) and differential gene expression analysis

RNA was isolated from the liver tissues of WT and *LBP*$^{-/-}$ rats fed with ND or HFD separately using a Trizol Kit (Invitrogen, USA) according to the manufacturer's protocols. The extracted samples were then used for library construction, and 150 bp paired-end sequencing was performed using the HiSeq 4000 platform (Illumina, USA).

Sequenced reads were mapped to the rat reference genome *Rattus_norvegicus*_6.0 by STAR-2.5.3a (Dobin et al., 2013). Read quality was assessed using the FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Low-quality and adapter sequences were trimmed using Trimmomatic. Afterwards, the expression of transcripts was quantified using the Salmon tool (Sewe et al., 2022), and estimates of transcript abundance for gene-level analysis were imported and summarized using the tximport (Soneson et al., 2015) function in the R/Bioconductor software suite. Finally, DESeq2 (Love et al., 2014) was employed to identify differentially expressed genes (DEGs) between WT and *LBP*$^{-/-}$ rats based on the criteria $|\log_2$ (fold-change)$|\geq 1.5$ and $P<0.05$.

Chromatin immunoprecipitation with high-throughput sequencing (ChIP-Seq) and differential analysis of histone acetylome signals

Chromatin immunoprecipitation assay was performed using a SimpleChIP Enzymatic Chromatin IP Kit (9005, Cell Signaling Technology, USA). Firstly, 150 mg of liver tissue was crosslinked with 1% formaldehyde, with the process then terminated with ice-cold phosphate-buffered saline (PBS) containing 0.125 mol/L glycine and protease inhibitor. The samples were lysed by the addition of lysis buffer and disrupted with a homogenizer, followed by the sheared of chromatin to an average size of 100–500 bp using sonication and micrococcal nuclease. The sheared chromatin was then immunoprecipitated with 500 μg of chromatin and 5 μg of anti-H3K27ac antibody (39133, Active Motif, China). Histone H3 (4620, Cell Signaling Technology, USA) and immunoglobulin G (IgG) (2729, Cell Signaling Technology, USA) were used as positive and negative controls, respectively. DNA fragments were collected and purified using protease and RNAase, and qPCR was also conducted using 1% input DNA to confirm ChIP success. Input library construction and sequencing procedures were carried out according to the Illumina protocols with minor modifications (Illumina, USA).

Sequencing reads were aligned to the rat reference genome *Rattus_norvegicus*_6.0 using a Burrows-Wheeler Aligner (BWA) (Abuín et al., 2015). The peak caller of Model-based Analysis for ChIP-Seq v.2.1.0 (MACS 2.1.0) was then used to identify the region enriched by H3K27ac, setting the *Q*-value threshold to 1e-5 (Zhang et al., 2008). The output peak files then were sorted and merged to skip short segments with small gaps using bedtools v.2.27.0 (Quinlan & Hall, 2010), merging those with more than 1 kb distance (merged regions were considered putative active enhancers if the region boundaries were ≥1 kb from the transcript start site (TSS)), following previous study (Villar et al., 2015). Subsequently, the SAMtools (v.1.2) "bedcov" function (Li et al., 2009) was adopted to quantify acetylated peak coverage for peak read depth according to the read counts of BAM files. ChIP and input read counts were normalized by peak lengths, subtracting input coverage from ChIP coverage to reduce the impact of fragmentation bias.

Differentially acetylated peaks between WT and *LBP*$^{-/-}$ individuals were identified using the DESeq2 R package (Love et al., 2014) in R v.3.5.1, with threshold criteria of $P<0.05$ and $|\log_2$(fold-change)$|\geq 1$. Super-enhancers were identified using the ROSE algorithm following previous reported methods (Whyte et al., 2013), defined as groups of putative enhancers in close genomic proximity (12.5 kb) with signal densities higher than the threshold (plot slope>1).

Gene Ontology (GO) and pathway enrichment analysis

DAVID (http://david-d.ncifcrf.gov/) and Metascape (https://metascape.org/gp/) were used to interpret the over-represented terms and biological processes of DEGs and differentially expressed peak-genes. The significance threshold criteria were set to $P<0.05$ and false discovery rate (FDR)≤0.05.

Integrative analysis of H3K27ac ChIP-Seq and RNA-Seq

To identify putative H3K27ac peak target genes, acetylome H3K27ac ChIP-Seq and transcriptome RNA-Seq integrative analyses were conducted by calculating the correlation between each H3K27ac peak and gene transcriptional level (Takahashi et al., 2015). Selection criteria for peak-gene correlation were defined by Pearson correlation coefficient≥0.8 and $P<0.01$. Genome-wide "four-way" analysis was performed based on the regulatory direction of differential peak-genes influenced by H3K27ac, with threshold criteria set to |Cor (peak-gene correlation)|≥0.5 and $|\log_2$(fold-change)$|\geq 1.5$, to screen PP peak-genes (up-regulated peak-genes positively regulated by H3K27ac ($\log_2$(fold-change) (ChIP≥0 & RNA≥0))) for further investigation.

Determination of core TFs

To screen core TFs in the *LBP* deficiency model, the PP peak-genes were first compared with the public Human ENCODE database (http://genome.ucsc.edu/ENCODE), then classified into 15 overlapping PP peak-genes in the livers of *LBP*$^{-/-}$ rats fed with ND and HFD. The sequences of the overlapping PP peak-genes were separately submitted to PROMO (http://alggen.lsi.upc.es/) for the identification of putative TFs. Subsequently, to investigate the interaction of putative TFs and identify core TFs, the Search Tool for the Retrieval of Interacting Genes (STRING) v,11.5 database (http://string-db.org/) was used to construct a protein-protein interaction (PPI) (Szklarczyk et al., 2017) network of predicted TFs. Combined scores of interactions higher than 0.4 were considered as significant. Metascape (https://metascape.org/gp/) was employed to examine the biological processes potentially regulated by these PP peak-genes.

To further confirm core TFs, gene homology analysis was conducted using UniProt (http://www.uniprot.org/) to determine TF conservation among humans, rats, and mice. The *LBP* promoter sequences of humans, rats, and mice were obtained from the UCSC Genome Browser (https://genome-store.ucsc.edu/) and *LBP* enhancer sequences were detected using our H3K27ac ChIP-Seq data. The sequences were then separately submitted to PROMO (http://alggen.lsi.upc.es/) to verify the core TF CCAAT/enhancer-binding protein β (C/EBPβ).

To determine the binding sites of C/EBPβ with the target gene stearoyl-CoA desaturase (*SCD*), the promoter sequence of *SCD* was obtained from the UCSC Genome Browser

(https://genome-store.ucsc.edu/) and binding site prediction was conducted using the JASPAR database (http://jaspar.genereg.net/).

Western blot analysis

Fresh liver tissues were washed with PBS and lysed in Radio Immunoprecipitation Assay (RIPA) protein lysis buffer. Insoluble material was then removed by centrifugation at 12 000 ×*g* for 20 min at 4 °C, with the resulting supernatant collected and stored at –80 °C until further use. Total protein was extracted and quantified using a BCA Protein Assay Kit (Beyotime, China). The proteins were then separated by 10%, 12%, or 15% sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and transferred to polyvinylidene fluoride (PVDF) membranes (ISEQ00010; Millipore, USA). The blotted membranes were blocked with 5% skim milk in Tris-buffered saline with Tween® 20 (TBST), then incubated with primary antibodies against H3K27ac (39133, Active motif, China, 1:1 000 dilution), C/EBPβ (23431-1-AP, Proteintech, China, 1:800 dilution), SCD (28678-1-AP, Proteintech, China, 1:10 000 dilution), LBP (23559-1-AP, Proteintech, 1:1 000 dilution), β-actin (66009-1-Ig, Proteintech, China, 1:50 000 dilution) at 4 °C overnight, then finally coated with appropriate horseradish peroxidase (HRP)-conjugated secondary antibodies (Proteintech, China, 1:10 000 dilution). Protein expression was detected using the Tanon 5200 Series Fully Automated Chemiluminescence/Fluorescence Image Analysis System (Tanon, China).

Plasmid constructs and SCD inhibition

For *LBP* and *SCD* knockdown in BRL-3A hepatocytes, small interfering RNA (siRNA) targeting *LBP* or *SCD* was designed by Shanghai GenePharma (China). For SCD inhibition, the SCD inhibitor A939572 was purchased from MedChemExpress (China). For *C/EBPβ* overexpression in the BRL-3A cells, the *C/EBPβ* overexpression plasmid (pcDNA3.1-*C/EBPβ*) was purchased from Shanghai GenePharma (China). Cells were transfected using Lipofectamine® 2000 transfection reagent (Invitrogen, USA) according to the manufacturer's protocols. Cells were harvested 48 h after transfection for further experiments.

Luciferase reporter assay

To confirm the regulatory role of C/EBPβ in *SCD* expression, sequences of three predicted C/EBPβ-binding sites in the *SCD* promoter region were cloned into the pGL3-promoter vector (containing an SV-40 promoter upstream of the luciferase gene) (Promega, USA). *C/EBPβ* was also cloned into the pCDNA 3.1 vector. For the luciferase assay, BRL-3A cells were cultured in six-well tissue culture dishes to 60% confluency by the time of transfection. Plasmids containing C/EBPβ-binding sites in the *SCD* promoter region were co-transfected with a plasmid carrying C/EBPβ using Lipofectamine® 2000 reagent (Invitrogen, USA). Six hours after transfection, the medium was changed to HDMEM supplemented with 10% FBS before overnight culture at 37 °C. To test luciferase activity, cells were collected and assessed using a Dual-Luciferase Reporter Assay Kit (Promega, USA). All primer information for plasmids is shown in Supplementary Table S2.

DNA pull-down assay

Biotin- and non-biotin-labeled DNA fragments of the *SCD* promoter were amplified by PCR. BRL-3A cells were first lysed, after which nucleoprotein was extracted and quantified using standard procedures. The desired biotin- or non-biotin-labeled DNA was added to nucleoprotein supernatant and placed in tubes on a rocker and incubated at 4 °C for 1 h. Then, 40 μL of streptavidin-conjugated agarose was added to each tube, which were again placed on the rocker for 30 min to allow affinity binding of streptavidin with biotin. The collected streptavidin-conjugated agarose beads were added with 500 μL of lysis buffer five times, followed by the addition of 40 μL of 2×SDS sample buffer to each sample for western blot analysis.

Statistical analysis

All data are presented as mean±standard deviation (SD) and were analyzed using GraphPad Prism v.9.0 (GraphPad Software, USA). Differences between groups were analyzed using Student's *t*-test or one-way analysis of variance (ANOVA) as appropriate. Here, $P<0.05$ was considered statistically significant.

RESULTS

LBP knockout attenuates inflammatory response but aggravates HFD-induced NAFLD in rats

To elucidate the role of *LBP* in NAFLD pathogenesis, rat and mouse strains with *LBP* deletions (*LBP*$^{-/-}$), mimicking clinical cases of *LBP* mutations, were constructed. Deletion efficiency was validated by agarose gel electrophoresis and western blot (Supplementary Figure S1A, B). Both WT and *LBP*$^{-/-}$ rats were fed ND and HFD for 8 weeks (Figure 1A). Notably, minimal differences were observed between the *LBP*$^{-/-}$_ND rats and WT_ND controls (Figure 1B–K). Examination of the morphology, H&E staining, and immunostaining of livers from *LBP*$^{-/-}$_HFD rats revealed reduced inflammation but increased lipid deposition relative to the WT_HFD rats (Figure 1B, C; Supplementary Figure S2A, B). Consistently, decreased mRNA expression of proinflammatory cytokines (Tumor necrosis factor-α (TNF-α) and Interleukin-6 (IL6)) was also observed, with more severe hypertriglyceridemia and hypercholesterolemia (TG and TC) in *LBP*$^{-/-}$_HFD rats compared to WT_HFD rats (Figure 1D–G). In addition, although *LBP* knockout reduced the body weight of HFD-fed rats, the *LBP*$^{-/-}$_HFD rats showed increased liver damage and NAFLD progression, as evidenced by higher serum transaminase (e.g., ALT and AST) levels, liver indices, and NAS values compared to the HFD-fed controls (Supplementary Figure S2C, D; Figure 1H, K). Together, these results suggest that *LBP* knockout may reduce the inflammatory response but aggravate HFD-induced NAFLD in rats.

Hepatocyte-specific _LBP_ ablation deteriorates lipid accumulation _in vitro_

To further confirm the biological consequences of *LBP* depletion, the tissue-specific expression and cellular localization of *LBP* were first determined. Results showed that *LBP* expression in the liver far exceeded that in other tissues across humans, rats, and mice (Figure 2A–C). Among the different liver cells, hepatocytes exhibited predominant *LBP* cellular localization, as evidenced by RNA single-cell type specificity analysis (Figure 2D, E). *LBP*-knockdown BRL-3A hepatocytes were then constructed through transduction using siRNA targeting the *LBP* gene *in vitro*, which was confirmed by western blot and RT-qPCR (Supplementary Figure 1C, D). Consistent with the *in vivo* NAFLD model, *LBP*-ablated

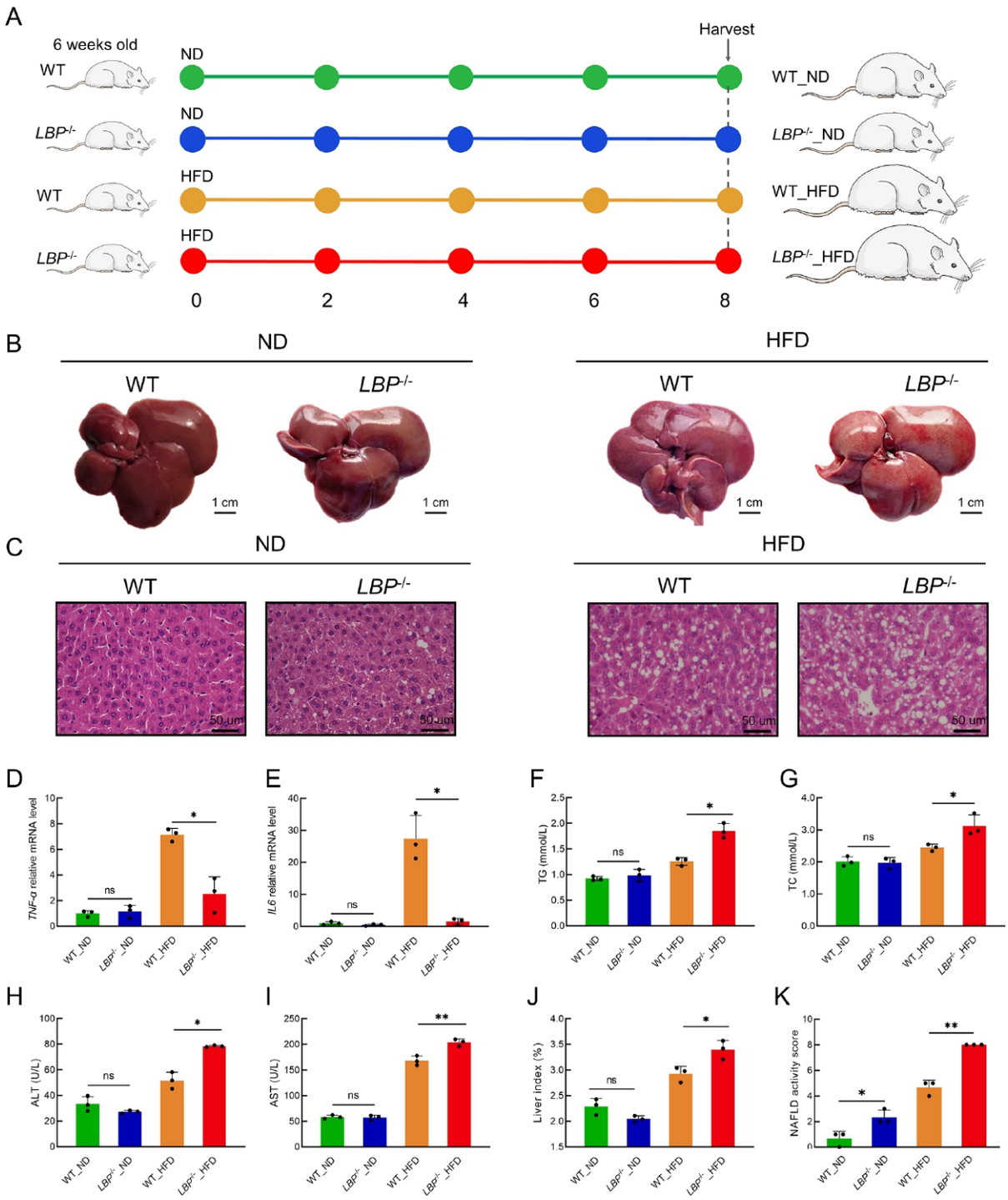


Figure 1 *LBP* knockout reduces inflammatory response but exacerbates HFD-induced NAFLD process in rats

A: Schematic showing groupings and steps of animal experiments. WT and *LBP*$^{-/-}$ rats were fed with ND or HFD for 8 weeks (*n*=6/group). B, C: Representative morphology (B) (scale bar: 1 cm) and H&E staining (C) (scale bar: 50 μm) of liver sections from WT_ND, *LBP*$^{-/-}$_ND, WT_HFD, and *LBP*$^{-/-}$_HFD rats (*n*=6/group). D, E: Relative mRNA expression levels of *TNF-α* (D) and *IL6* (E) in liver of WT_ND, *LBP*$^{-/-}$_ND, WT_HFD, and *LBP*$^{-/-}$_HFD rats (*n*=3/group). F–I: Serum content of TG (F), TC (G), ALT (H), and AST (I) in indicated groups (*n*=3/group). J: Liver index in indicated groups (*n*=3/group). K: Quantification of NAS in indicated groups *(n*=3/group). NAS is defined as the sum of scores of three components: steatosis (0–3), lobular inflammation (0–3), and hepatocyte ballooning (0–2). One-way ANOVA of data was performed. ns: Not significant; *: *P<0.05*; **: *P*<0.01. All data are shown as mean±SD. ND: Normal diet, HFD: High-fat diet, *TNF-α*: Tumor necrosis factor-α, *IL6*: Interleukin 6, TG: triglyceride, TC: total cholesterol, ALT: alanine transaminase, AST: aspartate transaminase, Liver index: liver weight to body weight ratio, NAFLD: non-alcoholic fatty liver disease, NAS: NAFLD activity score.

hepatocytes exhibited no significant differences compared to their controls under bovine serum albumin (BSA) treatment (Figure 2F, G). However, hepatocyte-specific ablation of *LBP* markedly decreased the expression of inflammation-related genes (*TNF-α*, *IL6*) (Figure 2F) and markedly increased the expression of lipid storage-related genes (*FASN*, *SREBP1*c) under PA treatment (Figure 2G). Together, these results indicate that hepatocyte-specific ablation of *LBP* may

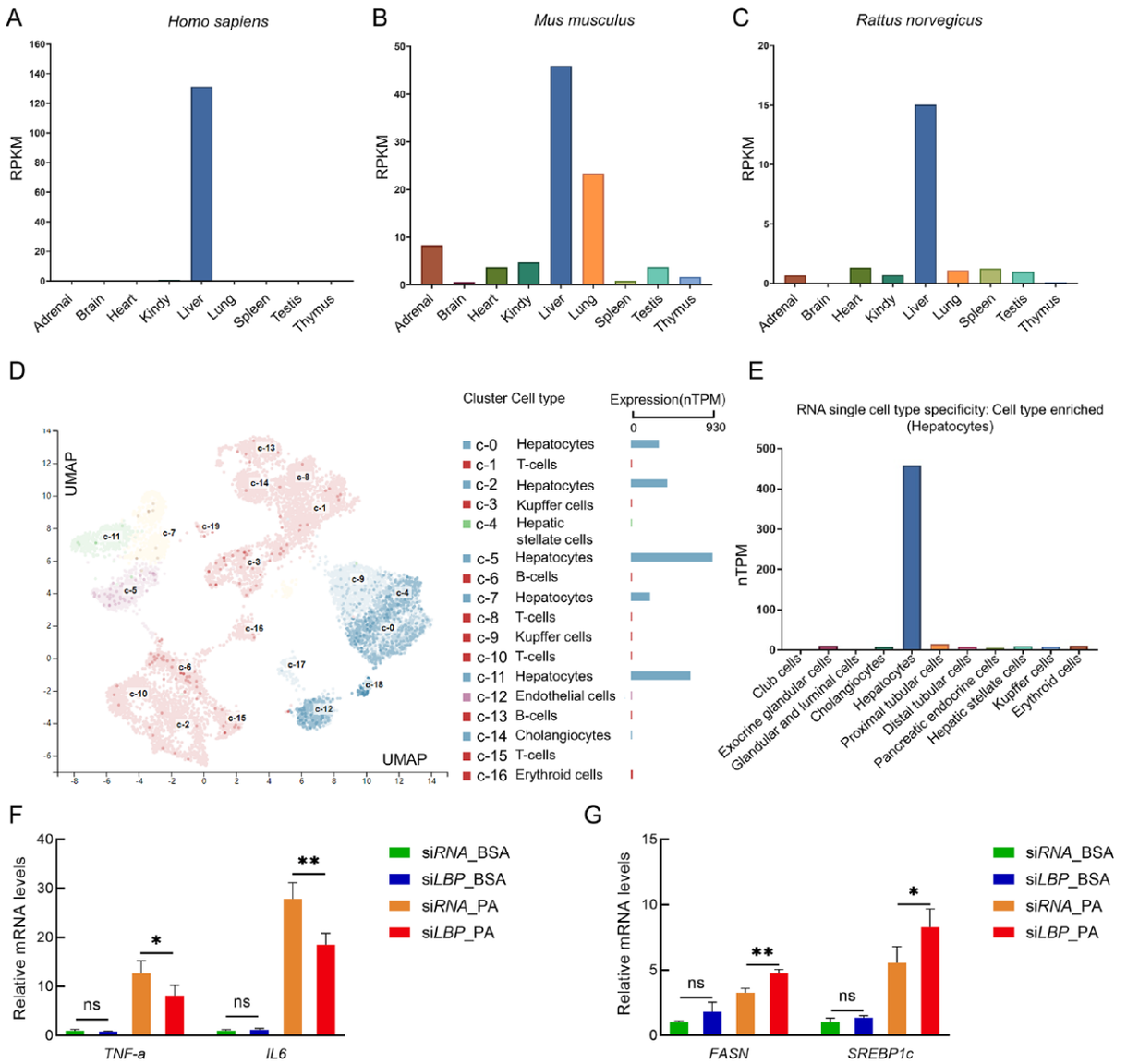


Figure 2 Hepatocyte-specific knockdown of *LBP* diminishes lipid accumulation

A–C: Tissue localization analysis of *LBP* in different tissues of *Homo sapiens* (A), *Mus musculus* (B), and *Rattus norvegicus* (C) using PRJEB443, PRJNA66167, and PRJNA238328 datasets, respectively, from NCBI public database (http://www.ncbi.nlm.nih.gov). D, E: Cellular localization analysis of *LBP* in different cell types in liver tissue (D) and all tissues (E) of humans using HPA public database (https://www.proteinatlas.org/). F, G: RT-qPCR analysis of mRNA levels of inflammation-related genes (F) and lipid deposition-related genes (G) of BRL-3A hepatocytes infected with indicated vectors, followed by stimulation with PA or BSA (n=3 independent experiments). One-way ANOVA of data was performed. ns: Not significant; *: P<0.05; **: P<0.01. All data are shown as mean±SD. PA: Palmitic acid, BSA: Bovine serum albumin, *TNF-α*: Tumor necrosis factor-α, *IL6*: Interleukin 6, *FASN*: Fatty acid synthase, *SREBP1c*: Sterol regulatory element binding protein 1c.

deteriorate lipid accumulation at the cellular level, corroborating observations noted in the animal models.

Transcriptome analysis reveals *LBP* deletion triggered lipid metabolic disorder in NAFLD rats

To further understand lipid deposition induced by *LBP* deletion, RNA-Seq of liver tissues from *LBP*$^{-/-}$ and WT rats fed a ND or HFD was conducted. In total, 168 and 3 871 significant DEGs (|$\log_2$(fold-change)|≥1.5 and P<0.05) were identified in the ND-fed *LBP*$^{-/-}$ rats and HFD-fed *LBP*$^{-/-}$ rats, respectively (Supplementary Figure S3A, B), consistent with our previous results showing more pronounced biological alterations in HFD-fed *LBP*-deficient rats compared to ND-fed *LBP*-deficient rats. The top 10 up-regulated and down-regulated genes in *LBP*$^{-/-}$_ND and *LBP*$^{-/-}$_HFD rats are shown in Supplementary Table S3 and S4, respectively. Functional analysis of these DEGs was conducted using DAVID (https://david.ncifcrf.gov/summary.jsp#) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses (https://www.genome.jp/kegg/). In the ND-fed rats, the DEGs induced by *LBP* deletion, including 101 up-regulated and 67 down-regulated genes, were markedly enriched in several biological processes, including “Endoplasmic reticulum membrane” and “Oxidation-reduction process” (Supplementary Figure S3A, C). For the HFD-fed rats, the DEGs induced by *LBP* deletion, including 2 402 up-regulated and 1 469 down-regulated genes, were enriched in various lipid metabolic processes, including “Oxidative acid metabolism”, “Steroid metabolic process”, and “Non-alcoholic fatty liver disease (NAFLD)” (Supplementary Figure S3B, D). Collectively, these results showed that *LBP* deletion

significantly interfered with the transcriptome profiles of HFD-fed rats, triggering abundant variation in gene expression and dysfunction in lipid metabolism, which may account for the excessive lipid deposition observed in the livers of HFD-induced NAFLD rats subjected to *LBP* knockout.

LBP deficiency contributes to H3K27 hyper-acetylation through the C/EBPβ TF

To examine the genome-wide cis-regulatory profiles of histone acetylome and transcriptome, ChIP-Seq analysis of H3K27ac, an important epigenetic marker of active enhancers, was applied to capture acetylated DNA regions and to determine alterations in acetylation levels induced by *LBP* deletion. Intriguingly, *LBP* deletion induced robust histone acetylome alteration in the HFD-fed rats compared to ND-fed rats. In total, 815 differential H3K27ac peaks ($|\log_2(\text{fold-change})|\geq 1$ and $P<0.05$) were obtained in the livers of $LBP^{-/-}$_ND rats compared to WT_ND rats, including 279 hyper-acetylated and 536 hypo-acetylated peaks (Figure 3A, B). Furthermore, 2 116 differential H3K27ac peaks ($|\log_2(\text{fold-change})|\geq 1$ and $P<0.05$) were obtained in the livers of $LBP^{-/-}$_HFD rats compared to $LBP^{-/-}$_ND rats, including 891 hyper-acetylated and 1 225 hypo-acetylated peaks (Figure 3E, F). The top 10 ranked hyper- and hypo-acetylated peaks identified in the ND- and HFD-fed rats are shown in Supplementary Tables S5 and S6, respectively. Combined histone acetylome and transcriptome analysis revealed a consistent trend in gene expression and H3K27ac activity, suggesting a potential role of H3K27ac in regulating transcriptome gene expression.

To better understand the relationship between the acetylome and transcriptome, integrative acetylome H3K27ac ChIP-Seq and transcriptome RNA-Seq analysis was performed. Genome-wide "four-way" analysis identified significantly up-regulated peak-genes positively regulated by H3K27ac (PP peak-genes with up-regulated mRNA expression and positive H3K27ac density correlation) based on specific criteria ($|\log_2(\text{fold-change})|\geq 1.5$ and a Cor≥0.5). These findings were in accordance with expectations, with fewer PP peak-genes in $LBP^{-/-}$_ND rats (PP=48) than in $LBP^{-/-}$_HFD rats (PP=1 128) (Figure 3C, G). Regarding the PP peak-genes, loss of *LBP* in ND-fed rats primarily impacted fundamental biological processes, including the "Cytosol" (Figure 3D). In contrast, HFD-fed rats exhibited a pronounced over-representation of lipid deposition-related biological processes, including "Cholesterol metabolic process", "Fatty acid metabolic process", and "Lipid homeostasis" (Figure 3H), in accordance with the RNA-Seq results (Supplementary Figure S3C, D). These findings suggest that changes in H3K27ac activity may contribute to excessive lipid deposition in the liver of $LBP^{-/-}$ rats through changes in gene expression, prompting further investigation into the underlying mechanisms linking *LBP* knockout and the regulation of H3K27ac activity and lipid metabolism.

To identify core TFs, 15 PP peak-genes up-regulated in both $LBP^{-/-}$_ND and $LBP^{-/-}$_HFD rats were analyzed. Based on the novel method for core TF identification using integrative histone acetylome and transcriptome analysis, as detailed in Supplementary Figure S4, C/EBPβ was identified as a core TF. Notably, C/EBPβ was extensively modulated by the 15 intersecting PP peak-genes (Figure 3I; Supplementary Figure S5A) and emerged as the most relevant TF among 20 overlapping ones. Interestingly, previous studies have highlighted the role of C/EBPβ in epigenetics (Tamura et al., 2021) and lipid metabolism (Dörr et al., 2022). Both PPI network and GO enrichment analyses of the top 20 predicted TFs also emphasized the vital regulatory role of C/EBPβ (Figure 3J, Supplementary Figure S5B), especially regarding H3K27ac activity and lipid metabolism. To further validate the relationship between C/EBPβ and H3K27ac, H3K27ac peak Chr1:264178426-264189376 was identified as a core peak, indicative of genome-wide acetylation, with a shared differential peak induced by *LBP* deletion in both ND-fed and HFD-fed groups and markedly increased density in the C/EBPβ gene region (Supplementary Figure S5C, D). Further investigation confirmed the increase in C/EBPβ expression after *LBP* deletion and its positive correlation with H3K27ac in HFD-fed rats (Figure 3K–N).

The consistency of TF C/EBPβ and *LBP* gene interactions across species was analyzed following the workflow outlined in Supplementary Figure S4. Results indicated that *C/EBPβ* and *LBP* showed high homology among humans, rats, and mice, indicating high consistency in C/EBPβ-triggered transcriptional activation of *LBP* among these species (Figure 4A, B). The relationship between C/EBPβ and *LBP* was further confirmed by TF prediction based on *LBP* promoter sequences of the three species and *LBP* enhancer sequences of rats from the ChIP-Seq data. In total, 43 overlapping TFs were identified as targeting both *LBP* promoters and enhancers (Figure 4C). Among them, C/EBPβ emerged as the most relevant TF due to its extensive interactions with other TFs and related functions in epigeneticsand lipid metabolism (Figure 4D), consistent with previous results (Dörr et al., 2022; Tamura et al., 2021). Moreover, the Cistrome database (Zheng et al., 2019) also lists *C/EBPβ* as a potential transcriptional regulator of *LBP* in liver tissue (Figure 4E). Further RT-qPCR analysis validated the positive correlation between *C/EBPβ* and *LBP* (Figure 4F). Cumulatively, these findings indicate that the regulatory role C/EBPβ on *LBP* is highly consistent across humans, rats, and mice.

Based on the above results, *LBP* deletion appears to contribute to H3K27 hyper-acetylation in NAFLD rats through stimulation of C/EBPβ, providing further evidence for potential application in NAFLD cases in other species.

C/EBPβ significantly up-regulates _SCD_ to promote liver lipid deposition

To investigate the potential mechanisms underlying altered H3K27ac-associated lipid accumulation, H3K27ac ChIP-Seq data analysis revealed increased genome-wide H3K27 acetylation within 3 kb of the TSS and transcript termination site (TES) in $LBP^{-/-}$_HFD rats, whereas a decline was observed in $LBP^{-/-}$_ND rats compared to their WT controls (Figure 5A, B). Hub targets of H3K27ac were subsequently searched using differential expression analysis of lipid metabolism-related genes based on the RNA-Seq data. Of note, *SCD* was identified as the most differentially expressed lipid metabolism-related gene (Figure 5C), in line with the ChIP-Seq results (Figure 5D), emphasizing the essential role of H3K27ac in the transcriptional activation of *SCD*. Further correlation and western blot analyses confirmed the positive correlation between *SCD* and H3K27ac, with enhanced protein levels after *LBP* knockout under HFD conditions (Figure 5E–G). Super-enhancer profiles of each group, derived from the H3K27ac ChIP-Seq data, revealed 456, 672, 837, and 892 super-enhancers in the livers of WT_ND, $LBP^{-/-}$

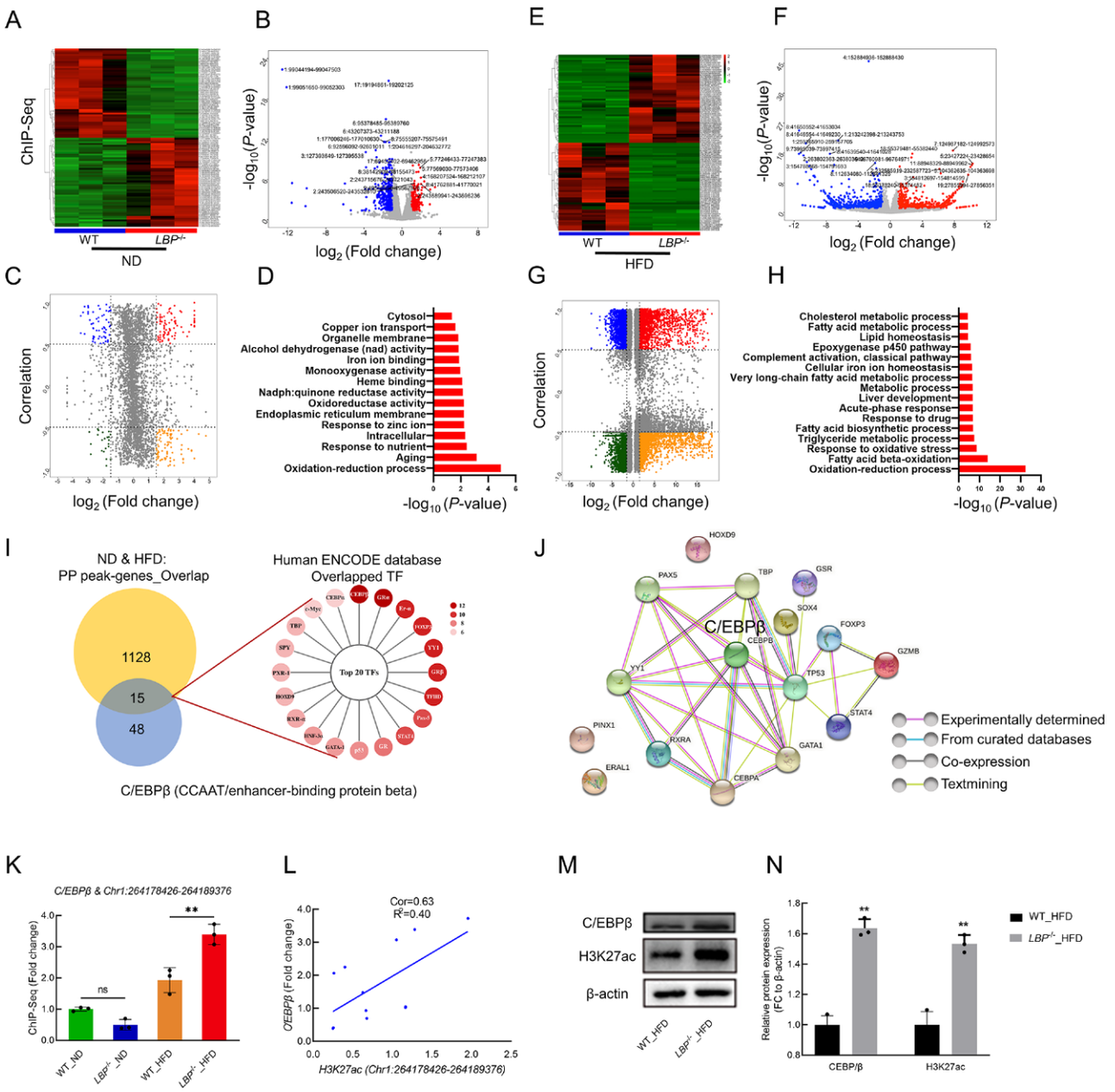


Figure 3 Integrative analysis of acetylome and transcriptome reveals upstream transcription factor C/EBPβ

A, B: Heatmap (A) and volcano plot (B) of differential H3K27ac regions between WT_ND and *LBP*$^{-/-}$_ND livers. Red and blue indicate up-regulation and down-regulation, respectively. C: Genome-wide "four-way" plot showing genes with |Cor (peak-gene correlation)|>0.5 and |$\log_2$(fold-change)|>1.5 generated by integrative H3K27ac ChIP-Seq and RNA-Seq analysis between WT_ND and *LBP*$^{-/-}$_ND rats. PP peak-genes with up-regulation and positive correlation are in red. D: Regulatory terms of PP peak-genes from ND_*LBP*$^{-/-}$ rats by DAVID (https://david-d.ncifcrf.gov/). E, F: Heatmap (E) and volcano plot (F) of differential H3K27ac regions between WT_ HFD and *LBP*$^{-/-}$_HFD livers. Red and blue indicate up-regulation and down-regulation, respectively. G: Genome-wide "four-way" plot showing genes with |Cor (peak-gene correlation)|>0.5 and |$\log_2$(fold-change)|>1.5 between WT and *LBP*$^{-/-}$ rats fed HFD, referred to in the legend of (C). H: Regulatory terms of PP peak-genes from *LBP*$^{-/-}$_ND rats by DAVID (https://david-d.ncifcrf.gov/). I: Venn diagram indicating overlapping PP peak-genes from livers of rats fed ND or HFD (left panel). Orange and blue circles indicate PP peak-genes from ND group (PP=48) and HFD group (PP=1 128), respectively. Right panel shows top 20 overlapping TFs identified in both ND and HFD groups using the Human ENCODE database (http://genome.ucsc.edu/ENCODE), identifying the most related TF (C/EBPβ). J: PPI diagram of most relevant TFs, using STRING v11.5 (https://cn.string-db.org/). K: Relative expression of *C/EBPβ&Chr1:264178426-264189376* from H3K27ac CHIP-Seq data in each group (*n*=3/group). L: Correlation between *C/EBPβ* expression level and H3K27ac peak (*Chr1:264178426-264189376*) density (Cor=0.63, R^2=0.4). M, N: Representative western blots (M) and quantification (N) of expression of C/EBPβ and H3K27ac protein levels in liver tissues from indicated rats (*n*=3/group). Student's *t*-test or one-way ANOVA was performed for data analysis as appropriate. ns: Not significant; **: *P*<0.01. Data are shown as mean±SD. ChIP-Seq: Chromatin immunoprecipitation coupled with high-throughput sequencing, RNA-Seq: Transcriptome high-throughput RNA sequencing, PP: up-regulated peak-genes positively regulated by H3K27ac ($\log_2$fold-change (ChIP≥0 & RNA≥0)), TF: Transcription factor, C/EBPβ: CCAAT/enhancer binding protein beta, H3K27ac: Histone H3 lysine 27 acetylation.

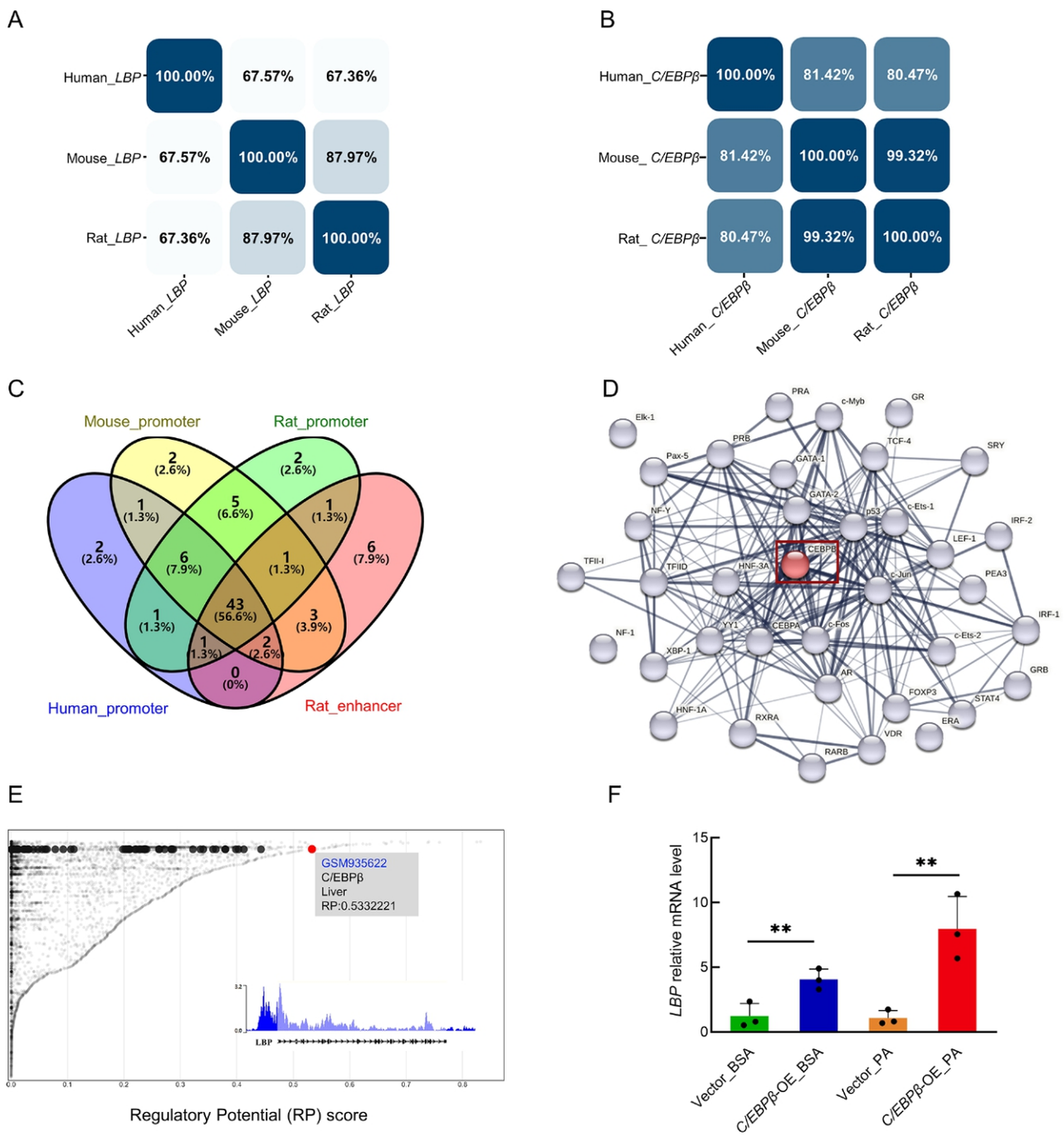


Figure 4 Regulatory process of C/EBPβ on *LBP* is highly consistent across humans, rats, and mice

A, B: Gene homology analysis of *LBP* (A) and *C/EBPβ* (B) among humans, rats, and mice using UniProt (https://www.uniprot.org/uniprot/). C: Venn diagram showing 43 overlapping TFs predicted by *LBP* promoter sequences of humans, mice, and rats from UCSC (http://www.genome.ucsc.edu/) and *LBP* enhancer sequences detected by H3K27ac ChIP-seq data. D: PPI network of top overlapping TFs in (C). E: Identification of C/EBPβ as a potential regulator of *LBP* gene in liver according to Cistrome database (accession number: GSM935622). All Cistrome data were carefully curated and processed with a streamlined analysis pipeline and evaluated with comprehensive quality control metrics. F: Relative mRNA levels of *LBP* in *C/EBPβ*-overexpressing BRL-3A cells and controls stimulated with PA or BSA (n=3/group). One-way ANOVA was performed for data analysis. **: P<0.01. Data are shown as mean±SD.

_ND, WT_HFD and *LBP*$^{-/-}$_HFD rats, respectively (Figure 6A, B, D, E). Interestingly, the *Chr1:264151393-264184386* super-enhancer, located proximal to the *SCD* gene region, exhibited pronounced activity in the *LBP*$^{-/-}$_HFD rats compared to the WT_HFD rats (Figure 6C, G). These findings suggest that H3K27 acetylation may facilitate the transcription and sustained expression of *SCD*, thus playing a vital role in the progression of HFD-induced NAFLD harboring *LBP* deletion.

Both RT-qPCR and Cistrome database analyses (Zheng et al., 2019) were performed, revealing a positive correlation between C/EBPβ and *SCD* (Figure 6F, H). C/EBPβ motifs and the top 10 potential C/EBPβ-binding sites in the SCD promoter region were identified using JASPAR (https://jaspar.genereg.net/) for subsequent regulatory functional analysis (Supplementary Figure S6). Of these, the top three binding sites were validated via a luciferase reporter assay, which

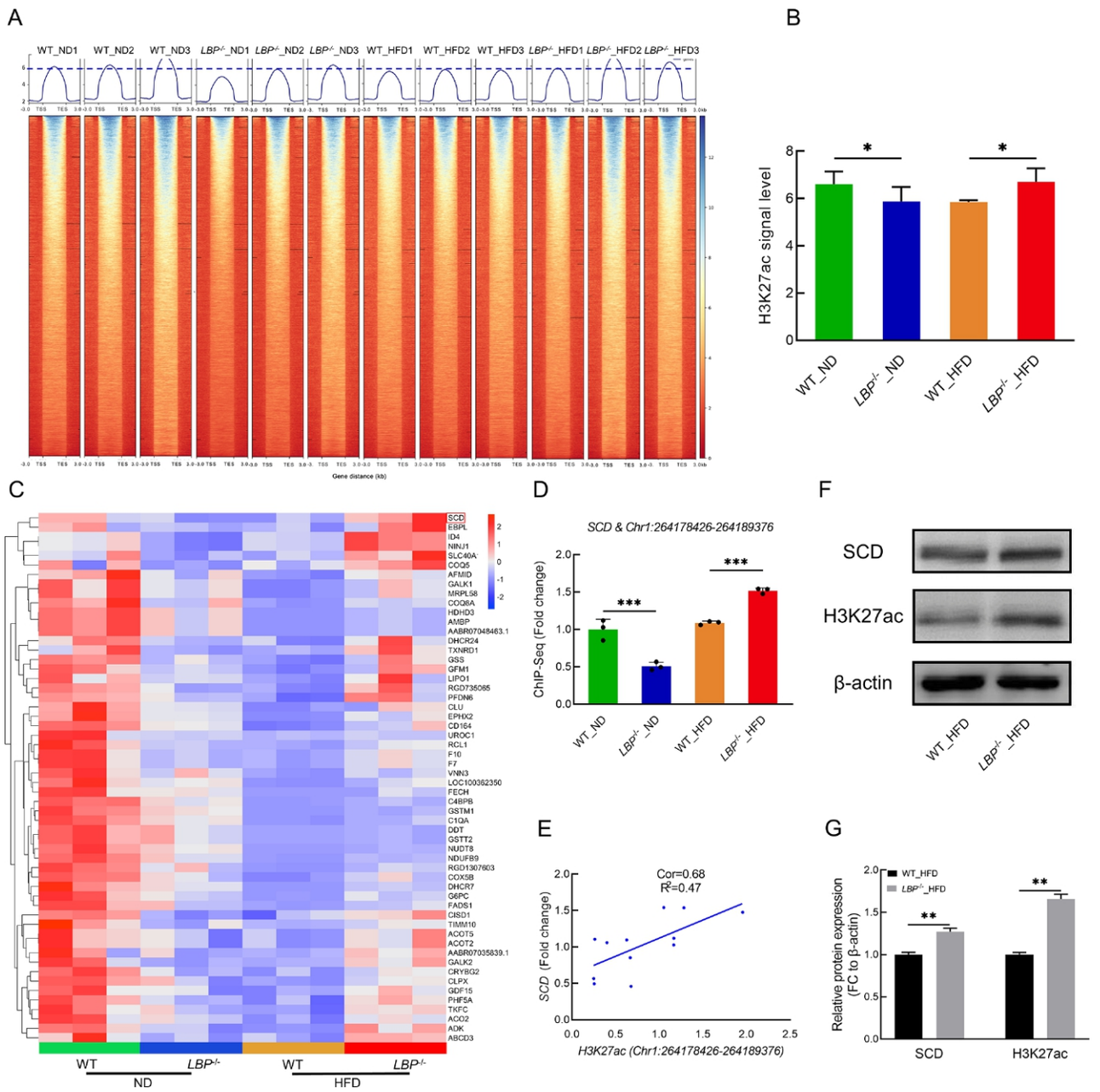


Figure 5 ***LBP*** **deficiency regulates expression of downstream lipid metabolism gene** ***SCD*** **through H3K27ac**

A: Heatmap showing H3K27ac signals at 3 kb near TSS and TES in the livers of WT_ND, *LBP*$^{-/-}$_ND, WT_HFD, and *LBP*$^{-/-}$_HFD rats (*n*=3/group). B: Quantification of H3K27ac signal level in each group, derived from ChIP-seq data shown in A (*n*=3/group). C: Heatmap of changes in mRNA levels of lipid metabolism-related genes in RNA-Seq data of liver from indicated groups (*n*=3/group). D: Relative expression of *SCD*&*Chr1:264178426-264189376* obtained by H3K27ac CHIP-Seq data in each group (*n*=3/group). E: Correlation between *SCD* expression level and H3K27ac peak (*Chr1:264178426-264189376*) density (Cor=0.68, R^2=0.47); F, G: Representative western blots (F) and quantification (G) of SCD and H3K27ac protein levels in liver tissues of WT_HFD and *LBP*$^{-/-}$_HFD rats (*n*=3/group). Student's *t*-test or one-way ANOVA was performed for data analysis. *: *P*<0.05; **: *P*<0.01; ***: *P*<0.001. Data are shown as mean±SD. TSS: Transcript start site, TES: Transcript termination site, *SCD*: Stearoyl-CoA desaturase.

signified that C/EBPβ substantially increased *SCD* activity, thereby highlighting the potential role of C/EBPβ in the transcriptional activation of *SCD* (Figure 6I, J). The DNA pull-down assay provided evidence of an interaction between *SCD* and C/EBPβ (Figure 6K), with the increase in the *SCD* mRNA level due to *C/EBPβ* overexpression further substantiated by RT-qPCR (Figure 6L). Together, these data suggest that C/EBPβ may promote the transcriptional activation of the lipid metabolism-related gene *SCD* during NAFLD in the context of *LBP* deficiency.

SCD accelerates excessive lipid deposition by inhibition of fatty acid oxidation (FAO)

To fully understand the functional metabolism of *SCD* in excessive lipid deposition, *SCD*-knockdown hepatocytes were established through siRNA transduction targeting the *SCD* gene, as confirmed by western blot and RT-qPCR (Supplementary Figure S7A, B). Results showed that *SCD* knockdown prominently decreased PA-induced lipid deposition but had little effect on BSA-treated hepatocytes (Supplementary Figure S7C). Consistently, hepatic TG was

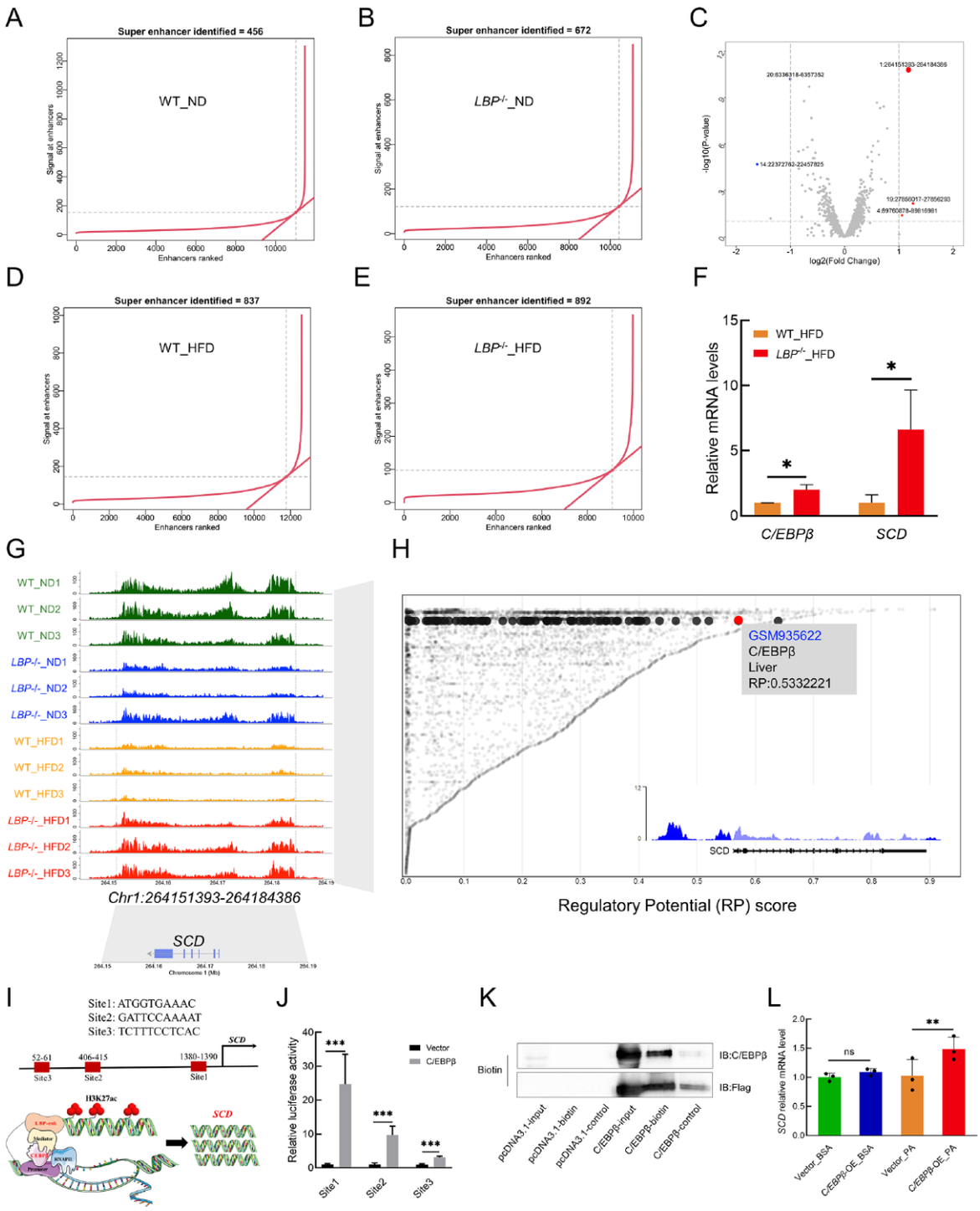


Figure 6 C/EBPβ facilitates expression of downstream lipid metabolism gene *SCD*

A, B: Enhancer regions of WT_ND (A) and *LBP*$^{-/-}$_ND (B) were plotted in increasing order based on their H3K27ac ChIP-seq signal. Enhancers above the curve inflection point were defined as super-enhancers. C: Volcano plots indicating significantly differential super-enhancers between WT_HFD and *LBP*$^{-/-}$_HFD samples. Red, blue, and gray dots indicate up-regulated, down-regulated, and unchanged H3K27ac peaks, respectively. D, E: Enhancer regions of WT_HFD (D) and *LBP*$^{-/-}$_HFD (E) were plotted in increasing order based on their H3K27ac ChIP-seq signal. Enhancers above the curve inflection point were defined as super-enhancers. F: Relative mRNA levels of *C/EBPβ* and *SCD* in HFD-fed *LBP*$^{-/-}$ and WT rats (*n*=3/group). G: Genome browser representation showing activity of *Chr1:264151393-264184386* super-enhancer, which tracks at the *SCD* locus in indicated individuals (*n*=3/group). H: Identification of C/EBPβ as a potential regulator of *SCD* gene in liver according to the Cistrome database (accession number: GSM935622). I: Schematic of three potential binding sites for C/EBPβ in *SCD* promoter region (upper panel) and model of enhanced *SCD* expression induced by C/EBPβ (lower panel). J: Relative activities of luciferase vectors in BRL-3A hepatocytes transfected with C/EBPβ and vector at predicted C/EBPβ-binding sites (site 1, site 2, site 3) in *SCD* promoter region (*n*=3/group). K: DNA pull-down assay showing interaction between C/EBPβ-binding site 1 in *SCD* promoter region and exogenous C/EBPβ. L: Relative mRNA levels of *SCD* in *C/EBPβ*-overexpressing BRL-3A cells and controls stimulated with PA or BSA (*n*=3/group). Student's *t*-test or one-way ANOVA was performed for data analysis. ns: Not significant; *: P<0.05; **: P<0.01; ***: P<0.001. Data are shown as mean±SD.

also significantly decreased following *SCD* knockdown and PA exposure compared to the controls, while the BSA group demonstrated no significant changes (Supplementary Figure S7D). These data suggest that *SCD* is essential for the progression of excessive lipid deposition in PA-induced hepatocytes. We further investigated the critical role of *SCD* in the progression of NAFLD deterioration induced by *LBP* deficiency *in vitro* and *in vivo*. Metabolomic analysis revealed elevated fatty acid content in *LBP*$^{-/-}$_HFD livers compared to WT_HFD livers (Supplementary Figure S8), as well as positive correlations between *SCD* expression and NAFLD parameters, including serum TG, TC, ALT, and AST levels, in *LBP*$^{-/-}$ rats (Supplementary Figure S9). In line with the *in vivo* observations, SCD inhibition in hepatocytes also significantly reversed the accelerated lipid accumulation induced by si*LBP* (Figure 7A; Supplementary Figure S10), further confirming the critical role of *SCD* in *LBP* deficiency-induced accelerated hepatocyte lipid deposition. Given the potential effects of *SCD* on accelerating lipid deposition through regulation of FAO (Kim et al., 2011), the mRNA expression levels of key genes positively associated with FAO activity (Supplementary Table S7) were measured in the livers of *LBP*-deficient HFD-induced NAFLD mice, confirming the role of *SCD* in excessive lipid deposition through dysfunctional FAO. Interestingly, *LBP* depletion-triggered overexpression of *SCD* robustly impaired the mRNA expression levels of FAO-related genes compared to the controls (Figure 7B–I). Together, these results suggest that *SCD* may contribute to excessive lipid deposition in the liver by inhibiting FAO in NAFLD harboring *LBP* deficiency.

DISCUSSION

In the present study, the potential epigenetic

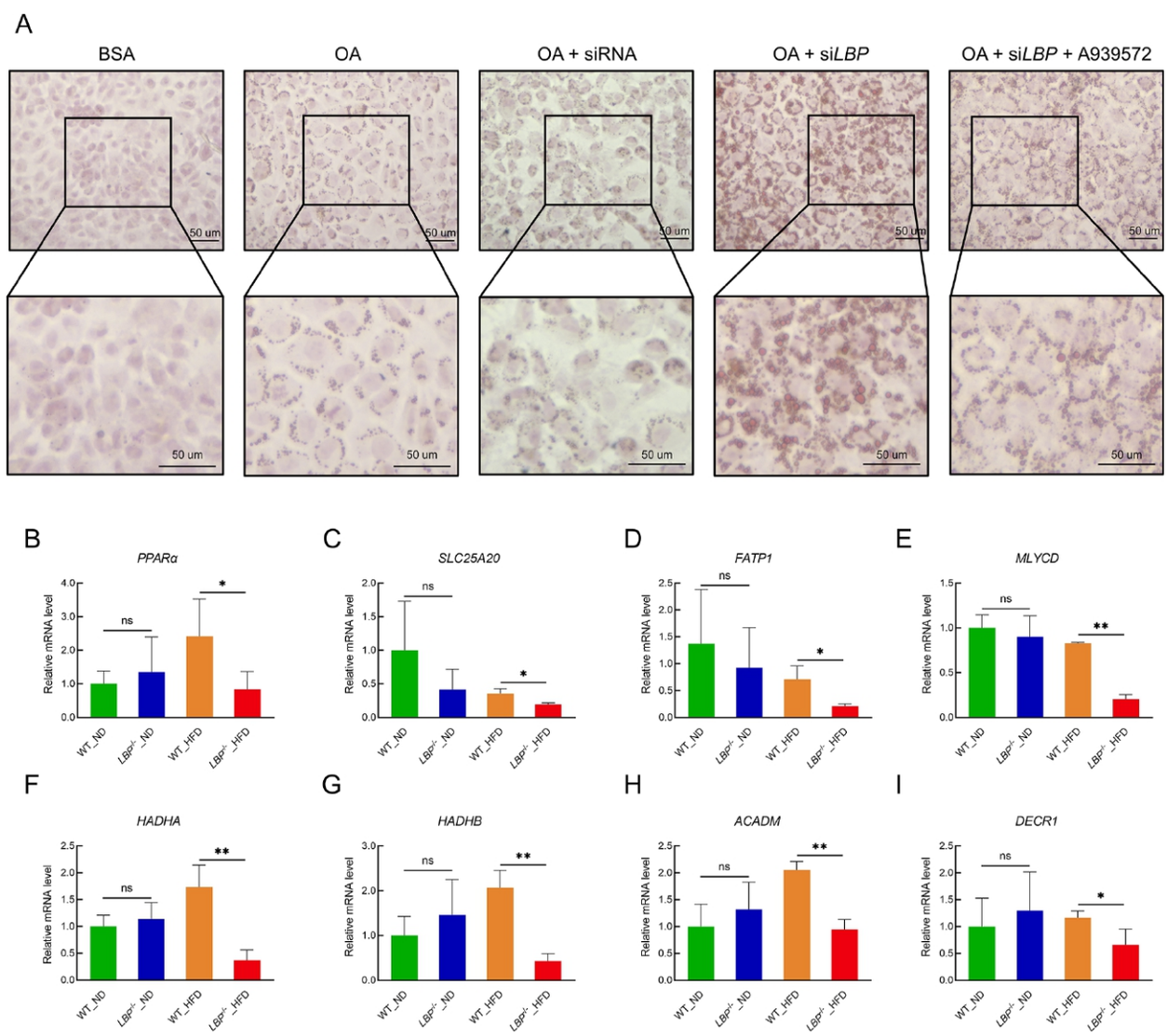


Figure 7 Regulatory mechanism of *SCD* aggravating lipid metabolic disorder of HFD-induced NAFLD with *LBP* deficiency

A: Images of Oil-red O staining of BRL-3A hepatocytes in indicated groups (n=3/group). Scale bar: 50 μm. B–I: Relative mRNA expression levels of FAO-related genes *PPARα* (B), *SLC25A20* (C), *FATP1* (D), *MLYCD* (E), *HADHA* (F), *HADHB* (G), *ACADM* (H), and *DECR1* (I) in liver tissues from WT_ND, *LBP*$^{-/-}$_ND, WT_HFD, and *LBP*$^{-/-}$_HFD rats (n=3/group). One-way ANOVA was performed for data analysis. ns: Not significant; *: P<0.05; **: P<0.01. Data are shown as mean±*SD*. *PPARα*: Peroxisome proliferator-activated receptor α, *SLC25A20*: Solute carrier family 25 member 20, *FATP1*: Fatty acid transport protein 1, *MLYCD*: Malonyl-CoA decarboxylase, *HADHA*: Hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit alpha, *HADHB*: Hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta, *ACADM*: Acyl-CoA dehydrogenase medium chain, *DECR1*: Mitochondrial 2,4-dienoyl-CoA reductase 1.

pathomechanisms of *LBP* deficiency as a potent contributor to NAFLD exacerbation were investigated. Results indicated that *LBP* knockout reduced inflammation but exacerbated HFD-induced NAFLD development through excessive lipid deposition. Through integrative histone acetylome H3K27ac ChIP-Seq and transcriptome RNA-Seq analysis, several related TFs and genes were identified in NAFLD livers harboring *LBP* deficiency. Among these, C/EBPβ was identified as an essential TF for dysregulated H3K27ac and *SCD*, transcriptionally activated by C/EBPβ, was identified as a hub gene and key mediator of NAFLD progression in *LBP*$^{-/-}$ rats. Taken together, this study provides insights into the epigenetic mechanisms underpinning HFD-induced NAFLD pathogenesis in the context of *LBP* deficiency, identifying potential TFs and key genes as potential candidate therapeutic targets.

LBP, recognized for its crucial role in inflammatory response regulation (Meng et al., 2021), has also been implicated in nutritional metabolic diseases (Nien et al., 2018; Sun et al., 2010). However, the precise mechanism by which *LBP* functions in NAFLD remains unclear, particularly in terms of epigenetic processes. In the current study, an *LBP*-knockout NAFLD model was utilized, replicating outcomes observed in clinical cases of *LBP* mutation. Based upon this model, *LBP* deficiency was identified as a potent contributor to NAFLD exacerbation from an epigenetics perspective. Results also revealed prominent liver specificity of *LBP* and distinct dysfunction in lipid metabolism-related processes in NAFLD with *LBP* knockout, consistent with the potential induction role of *LBP*$^{-/-}$ in liver lipid metabolism disorders. To further explore the epigenetic mechanisms and pathogenesis underlying the effects of *LBP* on NAFLD, we constructed H3K27ac alteration profiles and transcriptome maps, revealing widespread NAFLD-associated histone acetylome and transcriptome variation induced by *LBP* depletion. These observations align with the notion that dysregulation of H3K27ac is instrumental in modulating the development of fatty liver diseases by facilitating gene transcriptional activation (Li et al., 2021; Sun et al., 2015; Zhu et al., 2021).

Moreover, mediation of TF binding motifs is viewed as a critical mechanism of histone acetylome-induced gene transcriptional activation (Hsieh et al., 2022; Lambert et al., 2018; Marzi et al., 2018), highlighting the importance of TFs as potential therapeutic targets for NAFLD (Liang et al., 2019; Zhao et al., 2022b). Nevertheless, methods for the systematic prediction of core TFs are still lacking. Here, a model for the systematic screening of core TFs was developed using integrative histone acetylome and transcriptome analysis. Results revealed the potential significance of C/EBPβ, recognized as a key early regulator of adipogenesis (Feng et al., 2016; Moseti et al., 2016; Xi et al., 2016), in the progression of H3K27 hyper-acetylation induced by *LBP* deficiency in HFD-fed rats. Previous studies have identified C/EBPβ as a critical factor in the promoter and enhancer regions of genes (Takagi et al., 2022) and in the genome-wide increase of H3K27ac and various genes (Tamura et al., 2021), consistent with our findings of C/EBPβ-induced H3K27 hyper-acetylation and enhanced *SCD* gene expression. These results broaden our understanding of the biological functions of C/EBPβ in NAFLD and support C/EBPβ as a key player in the pathogenesis of NAFLD under *LBP* deficiency.

Our results also indicated that *SCD* may serve as a downstream effector of H3K27ac and C/EBPβ. Deficiency of *LBP* resulted in pronounced H3K27 hyper-acetylation, which facilitated the binding of C/EBPβ to the promoter region, leading to the activation of *SCD*. This gene, a vital marker of lipid metabolism, is frequently associated with the acceleration of fatty liver disorder progression due to its role in lipid deposition (Klepp et al., 2022; Zhang et al., 2020). Emerging evidence has suggested that *SCD* plays a potential role in triggering lipid accumulation through FAO suppression (Dobrzyn et al., 2004; Kim et al., 2011; Ntambi et al., 2002), as confirmed by the decreased expression of various FAO-related genes in our study. Differential super-enhancers with increased intensity in the *SCD* gene region were observed, consistent with H3K27ac ChIP-Seq data showing marked H3K27ac occupancy post-*LBP* deletion and HFD feeding, correlated with elevated mRNA expression. Collectively, these findings emphasize the key roles of C/EBPβ in *SCD* expression elevation and FAO inhibition in the current NAFLD model with *LBP* depletion. However, comprehensive research is needed to fully delineate the downstream network contributing to the hepatic roles of C/EBPβ.

Based on dysregulated acetylation and peak-genes, as well as the enriched TF binding motif of C/EBPβ and *SCD* gene, we proposed a potential epigenetic regulation mechanism in HFD-induced NAFLD livers harboring *LBP* deletion (Figure 8). In this model, *LBP* deficiency may trigger C/EBPβ, driving genome-wide histone hyper-acetylation and transcriptional activation of *SCD* through the binding motif of C/EBPβ, resulting in FAO inhibition in hepatocytes and exacerbation of HFD-induced NAFLD pathogenesis.

To the best of our knowledge, this is the first study to present genome-wide profiling of acetylation and transcription signals underlying the pathogenesis of NAFLD related to *LBP* deficiency. However, several study limitations should be mentioned. Firstly, the applicability of the results to other diet-induced NAFLD models, such as Western-style diet (WSD)-induced or methionine- and choline-deficient diet (MCD)-induced NAFLD models, is unknown. Notably, Jin et al. (2017) reported that loss of *LBP* can attenuate the development of WSD-induced NAFLD, which is inconsistent with our results. It is well established that the clinical pathogenesis and characteristics of NAFLD are highly heterogeneous (Yang et al., 2021). Thus, different diets may trigger different hepatic pathological features, mimicking different clinical states of NAFLD, which may explain the different influences of *LBP* deficiency on NAFLD. Additionally, although our sample size met the requirements for statistical analysis and was sufficient for determining the findings described above, future studies with larger sample sizes are essential for mitigating potential confounding factors and reinforcing these findings. Further in-depth experimental validation and molecular mechanism elucidation are required, which will be the focus of our future work.

In conclusion, while certain questions remain to be elucidated, these findings provide valuable insights into the epigenetic pathomechanisms underlying HFD-induced NAFLD with *LBP* depletion. Our results revealed robust alterations in histone acetylome H3K27ac and transcriptome as well as marked dysfunction in lipid metabolism-related biological processes induced by *LBP* knockout. Furthermore, based on integrative H3K27ac ChIP-seq and RNA-seq data analysis, C/EBPβ was identified as a core TF and driver of pathogenesis. Additionally, downstream effector *SCD*, a lipid

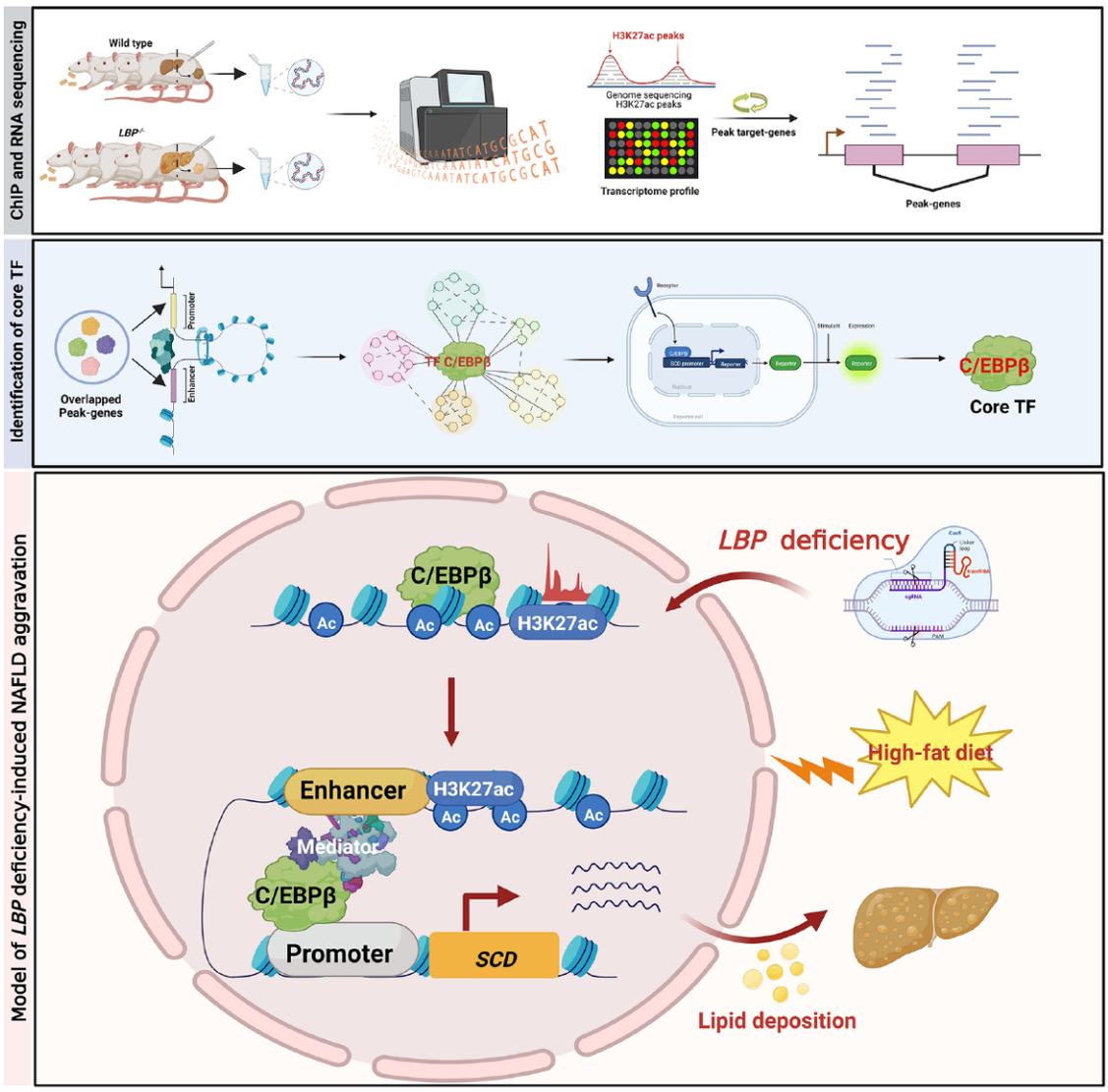


Figure 8 Proposed model of *LBP* deficiency exacerbating HFD-induced NAFLD pathogenesis

In this epigenetic mechanism-based model, C/EBPβ was identified as a pivotal TF and contributor to the progression of HFD-induced NAFLD with *LBP* deletion based on integrative H3K27ac ChIP-Seq and RNA-Seq analysis. Thus, *LBP* deficiency may trigger C/EBPβ to drive genome-wide histone H3 hyper- acetylation and activate transcriptional expression of downstream effector *SCD* to exacerbate NAFLD pathogenesis. Graph was created with BioRender software (https://biorender.com/)

metabolism gene with elevated super-enhancer activity in the gene region, was confirmed to accelerate excessive liver lipid deposition by inhibiting FAO. C/EBPβ and *SCD* were screened as core drivers of the pathogenesis of HFD-induced NAFLD with *LBP* depletion, and thus may serve as effective therapeutic targets.

DATA AVAILABILITY

The raw data reported in this paper were deposited in the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA: PRJNA1044849), National Genomics Data Center (GSA: PRJCA017968 and PRJCA002667), and Science Data Bank (DOI:10.57760/sciencedb.09322). All data that support the findings of this study are available from the corresponding authors upon request.

COMPETING INTERESTS

The authors declare they have no competing interests.

AUTHOR'S CONTRIBUTATIONS

H.S.F. and F.H. designed the experiments. Y.L.Z., L.L.M., and J.H.M. performed the experiments and completed the manuscript. X.Y., Y.S.T., M.X., and S.W.C. performed the cell experiments. X.R.Y. and X.Y.L. preformed animal experiments. Y.W. and M.Z.Z. analyzed the H3K27ac ChIP-Seq and RNA-Seq data. J.P. and X.J.L. conducted the core transcription factor prediction. J.Z.H. conducted the integrative ChIP-Seq and RNA-Seq analyses. Z.C.S., C.W., K.Z.Z., and Q.Q.D. analyzed most of the experimental data. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We are grateful to colleagues from the Laboratory Animal Research Center, Anhui Medical University, for animal husbandry and sample collection, and the College of Animal Science and Technology, Jiangxi Agricultural University, for the ChIP experiments.

REFERENCES

Abuín JM, Pichel JC, Pena TF, et al. 2015. BigBWA: approaching the burrows-wheeler aligner to big data technologies. *Bioinformatics*, **31**(24): 4003−4005.

Becares N, Gage MC, Voisin M, et al. 2019. Impaired LXRα phosphorylation attenuates progression of fatty liver disease. *Cell Reports*, **26**(4): 984−995.e6.

Brunt EM, Kleiner DE, Wilson LA, et al. 2011. Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology*, **53**(3): 810−820.

Dobin A, Davis CA, Schlesinger F, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**(1): 15−21.

Dobrzyn P, Dobrzyn A, Miyazaki M, et al. 2004. Stearoyl-CoA desaturase 1 deficiency increases fatty acid oxidation by activating AMP-activated protein kinase in liver. *Proceedings of the National Academy of Sciences of the United States of America*, **101**(17): 6409−6414.

Dörr D, Obermayer B, Weiner JM, et al. 2022. C/EBPβ regulates lipid metabolism and *Pparg* isoform 2 expression in alveolar macrophages. *Science Immunology*, **7**(75): eabj0140.

Feinberg AP. 2007. Phenotypic plasticity and the epigenetics of human disease. *Nature*, **447**(7143): 433−440.

Feng S, Reuss L, Wang Y. 2016. Potential of natural products in the inhibition of adipogenesis through regulation of PPARγ expression and/or its transcriptional activity. *Molecules*, **21**(10): 1278.

Friedman SL, Neuschwander-Tetri BA, Rinella M, et al. 2018. Mechanisms of NAFLD development and therapeutic strategies. *Nature Medicine*, **24**(7): 908−922.

Hsieh WC, Sutter BM, Ruess H, et al. 2022. Glucose starvation induces a switch in the histone acetylome for activation of gluconeogenic and fat metabolism genes. *Molecular Cell*, **82**(1): 60−74.e5.

Hu MJ, Long M, Dai RJ. 2022. Acetylation of H3K27 activated lncRNA NEAT1 and promoted hepatic lipid accumulation in non-alcoholic fatty liver disease via regulating miR-212-5p/GRIA3. *Molecular and Cellular Biochemistry*, **477**(1): 191−203.

Jin CJ, Engstler AJ, Ziegenhardt D, et al. 2017. Loss of lipopolysaccharide-binding protein attenuates the development of diet-induced non-alcoholic fatty liver disease in mice. *Journal of Gastroenterology and Hepatology*, **32**(3): 708−715.

Kim E, Lee JH, Ntambi JM, et al. 2011. Inhibition of stearoyl-CoA desaturase1 activates AMPK and exhibits beneficial lipid metabolic effects *in vitro*. *European Journal of Pharmacology*, **672**(1–3): 38–44.

Klepp TD, Sloan ME, Soundararajan S, et al. 2022. Elevated stearoyl-CoA desaturase 1 activity is associated with alcohol-associated liver disease. *Alcohol*, **102**: 51−57.

Lambert SA, Jolma A, Campitelli LF, et al. 2018. The human transcription factors. *Cell*, **172**(4): 650−665.

Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078−2079.

Li XZ, Yuan BC, Lu M, et al. 2021. The methyltransferase METTL3 negatively regulates nonalcoholic steatohepatitis (NASH) progression. *Nature Communications*, **12**(1): 7213.

Liang N, Damdimopoulos A, Goñi S, et al. 2019. Hepatocyte-specific loss of GPS2 in mice reduces non-alcoholic steatohepatitis via activation of PPARα. *Nature Communications*, **10**(1): 1684.

Liu Y, Jiang L, Sun CX, et al. 2018. Insulin/snail1 axis ameliorates fatty liver disease by epigenetically suppressing lipogenesis. *Nature Communications*, **9**(1): 2751.

Love MI, Huber W, Anders S. 2014. Moderated estimation of Fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12): 550.

Marzi SJ, Leung SK, Ribarska T, et al. 2018. A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nature Neuroscience*, **21**(11): 1618−1627.

Meng LL, Song ZC, Liu AD, et al. 2021. Effects of lipopolysaccharide-binding protein (LBP) single nucleotide polymorphism (SNP) in infections, inflammatory diseases, metabolic disorders and cancers. *Frontiers in Immunology*, **12**: 681810.

Moreno-Navarrete JM, Escoté X, Ortega F, et al. 2015. Lipopolysaccharide binding protein is an adipokine involved in the resilience of the mouse adipocyte to inflammation. *Diabetologia*, **58**(10): 2424−2434.

Moseti D, Regassa A, Kim WK. 2016. Molecular regulation of adipogenesis and potential anti-adipogenic bioactive molecules. *International Journal of Molecular Sciences*, **17**(1): 124.

Nien HC, Sheu JC, Chi YC, et al. 2018. One-year weight management lowers lipopolysaccharide-binding protein and its implication in metainflammation and liver fibrosis. *Plos One*, **13**(11): e0207882.

Ntambi JM, Miyazaki M, Stoehr JP, et al. 2002. Loss of stearoyl-CoA desaturase-1 function protects mice against adiposity. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(17): 11482−11486.

Ospelt C. 2022. A brief history of epigenetics. *Immunology Letters*, **249**: 1−4.

Pfalzgraff A, Weindl G. 2019. Intracellular lipopolysaccharide sensing as a potential therapeutic target for sepsis. *Trends in Pharmacological Sciences*, **40**(3): 187−197.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6): 841−842.

Rada-Iglesias A, Bajpai R, Swigut T, et al. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*, **470**(7333): 279−283.

Samuel VT, Shulman GI. 2018. Nonalcoholic fatty liver disease as a nexus of metabolic and hepatic diseases. *Cell Metabolism*, **27**(1): 22−41.

Sewe SO, Silva G, Sicat P, et al. 2022. Trimming and validation of illumina short reads using trimmomatic, trinity assembly, and assessment of RNA-seq data. *In*: Edwards D. Plant Bioinformatics: Methods and Protocols. New York: Springer, 211–232.

Smith E, Shilatifard A. 2014. Enhancer biology and enhanceropathies. *Nature Structural & Molecular Biology*, **21**(3): 210−219.

Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, **4**: 1521.

Song ZC, Meng LL, He ZX, et al. 2021. LBP protects hepatocyte mitochondrial function via the PPAR-CYP4A2 signaling pathway in a rat sepsis model. *Shock*, **56**(6): 1066−1079.

Sun C, Fan JG, Qiao L. 2015. Potential epigenetic mechanism in non-alcoholic fatty liver disease. *International Journal of Molecular Sciences*, **16**(3): 5161−5179.

Sun L, Yu ZJ, Ye XW, et al. 2010. A marker of endotoxemia is associated with obesity and related metabolic disorders in apparently healthy Chinese. *Diabetes Care*, **33**(9): 1925−1932.

Szklarczyk D, Morris JH, Cook H, et al. 2017. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Research*, **45**(D1): D362−D368.

Takagi H, Tamura I, Fujimura T, et al. 2022. Transcriptional coactivator PGC-1α contributes to decidualization by forming a histone-modifying complex with C/EBPβ and p300. *Journal of Biological Chemistry*, **298**(5): 101874.

Takahashi JS, Kumar V, Nakashe P, et al. 2015. ChIP-seq and RNA-seq methods to study circadian control of transcription in mammals. *Methods in Enzymology*, **551**: 285−321.

Takeuchi F, Yanai K, Inomata H, et al. 2007. Search of type 2 diabetes susceptibility gene on chromosome 20q. *Biochemical and Biophysical Research Communications*, **357**(4): 1100−1106.

Tamura I, Maekawa R, Jozaki K, et al. 2021. Transcription factor C/EBPβ induces genome-wide H3K27ac and upregulates gene expression during

decidualization of human endometrial stromal cells. *Molecular and Cellular Endocrinology*, **520**: 111085.

Villar D, Berthelot C, Aldridge S, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell*, **160**(3): 554−566.

Whyte WA, Orlando DA, Hnisz D, et al. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**(2): 307−319.

Xi Y, Shen WJ, Ma LL, et al. 2016. HMGA2 promotes adipogenesis by activating C/EBPβ-mediated expression of PPARγ. *Biochemical and Biophysical Research Communications*, **472**(4): 617−623.

Yang X, Sun DT, Xiang H, et al. 2021. Hepatocyte SH3RF2 deficiency is a key aggravator for NAFLD. *Hepatology*, **74**(3): 1319−1338.

Younossi Z, Tacke F, Arrese M, et al. 2019. Global perspectives on nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Hepatology*, **69**(6): 2672−2682.

Youssefian L, Vahidnezhad H, Saeidian AH, et al. 2019. Inherited non-alcoholic fatty liver disease and dyslipidemia due to monoallelic *ABHD5* mutations. *Journal of Hepatology*, **71**(2): 366−370.

Yu Y, Cai JJ, She ZG, et al. 2019. Insights into the epidemiology, pathogenesis, and therapeutics of nonalcoholic fatty liver diseases. *Advanced Science*, **6**(4): 1801585.

Zhang N, Wang YL, Zhang JL, et al. 2020. N-glycosylation of CREBH improves lipid metabolism and attenuates lipotoxicity in NAFLD by modulating PPARα and SCD-1. *The FASEB Journal*, **34**(11): 15338−15363.

Zhang Y, Liu T, Meyer CA, et al. 2008. Model-based analysis of ChIP-seq (MACS). *Genome Biology*, **9**(9): R137.

Zhao N, Zhang XX, Ding JJ, et al. 2022a. *SEMA7A*^{R148W} mutation promotes lipid accumulation and NAFLD progression via increased localization on the hepatocyte surface. *JCI insight*, **7**(15): e154113.

Zhao S, Zhang XR, Li HT. 2018. Beyond histone acetylation—writing and erasing histone acylations. *Current Opinion in Structural Biology*, **53**: 169−177.

Zhao YC, Gao LC, Jiang CL, et al. 2022b. The transcription factor zinc fingers and homeoboxes 2 alleviates NASH by transcriptional activation of phosphatase and tensin homolog. *Hepatology*, **75**(4): 939−954.

Zheng RB, Wan CX, Mei SL, et al. 2019. Cistrome data browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Research*, **47**(D1): D729−D735.

Zhou Z, Xu MJ, Gao B. 2016. Hepatocytes: a key cell type for innate immunity. *Cellular & Molecular Immunology*, **13**(3): 301−315.

Zhu YL, Zeng QJ, Li F, et al. 2021. Dysregulated H3K27 acetylation is implicated in fatty liver hemorrhagic syndrome in chickens. *Frontiers in Genetics*, **11**: 574167.

Wang et al. *Zool. Res.* 2024, 45(1): 95−107
https://doi.org/10.24272/j.issn.2095-8137.2023.008

ZR | Zoological Research

Article Open Access

Comprehensive analysis of the gut microbiome and post-translational modifications elucidates the route involved in microbiota-host interactions

Hai-Yang Wang[1,2,3,#], Lan-Xiang Liu[4,#], Xue-Yi Chen[5,#], Yang-Dong Zhang[6,#], Wen-Xia Li[1], Wen-Wen Li[5], Lian Wang[1], Xiao-Long Mo[1,6], Hong Wei[7,*], Ping Ji[2,3,*], Peng Xie[1,2,6,*]

[1] *NHC Key Laboratory of Diagnosis and Treatment on Brain Functional Diseases*, *First Affiliated Hospital of Chongqing Medical University*, *Chongqing* 400016, *China*
[2] *College of Stomatology and Affiliated Stomatological Hospital of Chongqing Medical University*, *Chongqing* 401147, *China*
[3] *Chongqing Key Laboratory for Oral Diseases and Biomedical Sciences*, *Chongqing* 401147, *China*
[4] *Department of Neurology*, *Yongchuan Hospital of Chongqing Medical University*, *Chongqing* 402160, *China*
[5] *Department of Pathology*, *Faculty of Basic Medicine*, *Chongqing Medical University*, *Chongqing* 400016, *China*
[6] *Department of Neurology*, *First Affiliated Hospital of Chongqing Medical University*, *Chongqing* 400016, *China*
[7] *Yu-Yue Pathology Scientific Research Center*, *Chongqing* 401329, *China*

ABSTRACT

The gut microbiome interacts with the host to maintain body homeostasis, with gut microbial dysbiosis implicated in many diseases. However, the underlying mechanisms of gut microbe regulation of host behavior and brain functions remain unclear. This study aimed to elucidate the influence of gut microbiota on brain functions via post-translational modification mechanisms in the presence or absence of bacteria without any stimulation. We conducted succinylome analysis of hippocampal proteins in germ-free (GF) and specific pathogen-free (SPF) mice and metagenomic analysis of feces from SPF mice. These results were integrated with previously reported hippocampal acetylome and phosphorylome data from the same batch of mice. Subsequent bioinformatics analyses revealed 584 succinylation sites on 455 proteins, including 54 up-regulated succinylation sites on 91 proteins and 99 down-regulated sites on 51 proteins in the GF mice compared to the SPF mice. We constructed a panoramic map of gut microbiota-regulated succinylation, acetylation, and phosphorylation, and identified cross-talk and relative independence between the different types of post-translational modifications in modulating complicated intracellular pathways. Pearson correlation analysis indicated that 13 taxa, predominantly belonging to the Bacteroidetes phylum, were correlated with the biological functions of post-translational modifications. Positive correlations between these taxa and succinylation and negative correlations between these taxa and acetylation were identified in the modulation of intracellular pathways. This study highlights the hippocampal physiological changes induced by the absence of gut microbiota, and proteomic quantification of succinylation, phosphorylation, and acetylation, contributing to our understanding of the role of the gut microbiome in brain function and behavioral phenotypes.

Keywords: Gut microbiota; Hippocampal protein; Post-translational modifications; Succinylation; Acetylation; Phosphorylation

INTRODUCTION

Microbial communities in natural and host-associated environments are large and delicate ecosystems comprised of trillions of bacteria, archaea, fungi, microbial eukaryotes, and viruses (Marcelino et al., 2020). These communities play significant roles in human health and disease (Fan and & Pedersen, 2021). As the largest microbial community in the human body, the gut microbiota not only regulates host metabolism, neurological signaling, immune responses, gut hormone secretion, and gastrointestinal function, but also

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 06 April 2023; Accepted: 08 September 2023; Online: 09 September 2023

Foundation items: This work was supported by the Natural Science Foundation Project of China (81820108015, 82201683), China Postdoctoral Science Foundation (2021M693926, 2020TQ0393, 2020M683634XB), Chongqing Science & Technology Commission (cstc2021jcyj-bshX0150, cstc2021jcyj-bshX0201), and Special Funding for Chongqing Postdoctoral Research Projects (2021XMT001)

#Authors contributed equally to this work

*Corresponding authors, E-mail: Weihong63528@163.com; jiping@hospital.cqmu.edu.cn; xiepeng@cqmu.edu.cn

protects against infection (Fan & Pedersen, 2021; Wei et al., 2019).

The conventional brain-centric perspective proposed that the peripheral system was unidirectionally driven by the central nervous system (CNS), with the influence of peripheral metabolic homeostasis and immune balance on the CNS largely neglected or simply overlooked (Morais et al., 2021). However, in more recent years, research has begun to recognize that microorganisms can influence the brain through their ability to produce and modify many metabolic, immunological, and neurochemical factors in the gut, which ultimately impact the nervous system (Lukić et al., 2019). Although the exact mechanisms involved remain unclear, accumulating evidence has demonstrated that gut microbiota dysbiosis can adversely affect behavior and neural biochemistry through the microbiota-gut-brain axis (Cryan & Dinan, 2012; Fung et al., 2017; Long-Smith et al., 2020). In previous studies, we showed that the gut microbiota modulates gene transcription and protein expression (Chen et al., 2019; Liu et al., 2020; Rao et al., 2021; Zeng et al., 2016; Zhou et al., 2020) and drives metabolic alterations in the brains of rodents and nonhuman primates (Li et al., 2018a; Wang et al., 2020a; Zheng et al., 2021). The gut microbiota varies with lifecycle and exerts profound influences on the host (Yang et al., 2022), with bacterial infection found to precipitate stress-induced memory dysfunction (Gareau et al., 2011). Research has also shown that the maternal gut microbiome can modulate fetal neurodevelopment and thalamocortical axon genesis in mice (Vuong et al., 2020) and maternal sleep deprivation can induce gut microbial dysbiosis (Yao et al., 2022). However, two crucial questions remain to be addressed. First, how does the gut microbiota influence host behavior and brain function via the gut-brain axis? Second, which key taxa are responsible for mediating the interactions between the microbiota and host.

Germ-free (GF) animals, which lack microbiota and are free from bacterial contamination, serve as suitable models for exploring host responses to disordered, normal, or absent flora. GF mice have been an invaluable tool for investigating microbiota-host interactions (Cryan & Dinan, 2012; Cryan et al., 2019; Foster & McVey Neufeld, 2013; Long-Smith et al., 2020; Zheng et al., 2019) and exhibit cognitive deficits, altered sociability, and modifications in neurogenesis, neurotransmitter, synapse, and neuronal activity-related genes (Cryan et al., 2019; Luczynski et al., 2016; Schretter, 2020; Wang et al., 2020a). We have also previously identified decreased depressive and anxiety-like behavior in GF mice accompanied by molecular and neurochemical changes (Chen et al., 2017b, 2019; Zeng et al., 2016; Zheng et al., 2016). However, although earlier studies have confirmed the role of gut microbiota in regulating host behavior and brain function, the precise mechanisms remain largely unknown.

Proteins are pivotal biomolecules necessary for mediating diverse biological functions. Most cellular proteins require further modification to perform their specific functions (Fang et al., 2018). Post-translational modifications (PTMs), with the addition of covalent bonds or enzymatic modification on amino acid side chains or at the C- or N-termini of proteins (Gao et al., 2019), play crucial roles in modulating protein function by regulating their activity, localization, and interaction. Protein PTMs are critical regulators of cellular function and activity, including responses to environmental stimuli (Zhu et al., 2022). In our previous work using GF mice, we found that an absence of microbiota induces global dysregulation in protein phosphorylation and acetylation (Wang et al., 2020b; Yu et al., 2021), while gut microbiota dysbiosis induces changes in phosphorylation, acetylation, and succinylation in hippocampal proteins (Liu et al., 2021; Wang et al., 2020b). These findings suggest that PTMs may help reveal the mechanisms underlying gut microbiota regulation of brain function and behavior. However, little is known regarding the extent to which PTMs are influenced by gut microbiota and which type of PTM is more susceptible to gut microbiota.

In the current study, proteomic quantification of lysine succinylation in the hippocampus of GF and specific pathogen-free (SPF) mice was performed using tandem-mass-tag (TMT) labeling followed by high-resolution liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. The resulting succinylome was integrated with our previously published acetylome and phosphorylome data from the same batch of GF and SPF mice (Wang et al., 2020b; Yu et al., 2021) and a panoramic map of succinylation, phosphorylation, and acetylation of hippocampal proteins in GF mice was constructed. To further explore how the gut microbiota regulates this panoramic map of PTMs, metagenomics sequencing was used to analyze gut microbiota composition in SPF mice, with results then integrated with the PTM data to reveal microbial/PTM interactions.

MATERIALS AND METHODS

Animals

Eight-week-old GF and SPF male Kunming mice (*n*=8 per group, 30–40 g) were provided by the Laboratory Animal Center at the Third Military Medical University (Chongqing, China), as previously described (Wang et al., 2020b). The GF mice were housed in sterile conditions using flexifilm gnotobiotic isolators and fed with irradiated food and sterile water. Stool samples from the GF mice were collected weekly and examined by polymerase chain reaction (PCR) and culture-based methods to verify GF status. The SPF mice were housed in a standard animal facility. All mice were maintained under consistent environmental conditions: noise less than 60 dB, 12 h light/dark cycle (lights on at 0800–2000h), temperature of 22–23 °C, and relative humidity of 50%±5%. All regulations regarding the use of research animals were observed and all protocols were approved by the Ethics Committee of Chongqing Medical University (Chongqing, China; Approval No. 2017013).

Sample preparation, TMT labeling, and high-performance liquid chromatography (HPLC) fractionation

Sample preparation, TMT labeling, and HPLC fractionation procedure were performed as described previously (Liu et al., 2021; Wang et al., 2020b). Briefly, GF and SPF mice were anesthetized and perfused before sacrifice by decapitation. The hippocampus was immediately dissected on ice, frozen in liquid nitrogen, and stored at −80 °C. Proteins were extracted and their concentrations were determined using a BCA kit, followed by trypsin digestion. Each peptide sample was labeled with the respective TMT reagent (based on the manufacturer's protocols, Thermo Scientific, USA) and pooled samples were desalted with Strata X C18 SPE column (Phenomenex, USA), dried by vacuum centrifugation, and fractionated by high pH reverse-phase HPLC (Agilent

300Extend C18 column, 5 μm particles, 4.6 mm ID, 250 mm length, see Supplemental Materials for further details).

Affinity enrichment of Ksucc-modified peptides and LC-MS/MS analysis

Ksucc-modified peptides were enriched by dissolving tryptic peptides in NETN buffer (100 mmol/L NaCl, 1 mmol/L EDTA, 50 mmol/L Tris-HCl, 0.5% NP-40, pH 8.0), followed by incubation with pre-washed succinylated antibody beads at 4 °C overnight with gentle shaking. The beads were washed four times with NETN buffer and twice with ddH_2O, with the bound peptides then eluted with 0.1% trifluoroacetic acid (TFA). Eluted fractions were combined and vacuum-dried. The peptides were desalted with C18 ZipTips (Millipore, USA), according to the manufacturer's instructions, and LC-MS/MS was subsequently performed (see Supplemental Materials for further details).

Conserved sequence analysis

Conserved sequences comprising amino acids at specific locations within modify-21-mers (encompassing 10 amino acids both upstream and downstream of the modification site) in all protein sequences were analyzed using the motif-x algorithm. The minimal number of peptides occurring in one motif "occurrence" was 20, and the significance threshold value was 0.0000001. Motif-based clustering was visualized as a heatmap. Icelogo (https://iomics.ugent.be/icelogoserver/) was used to examine the properties of amino acids surrounding the modification sites using *t*-test (P<0.05), and the "choosing scoring system" was set to "percentage difference".

Secondary structure prediction and cellular localization of Ksucc proteomes

The sequence profiles of target proteins were used to predict secondary structures of succinylation, acetylation, and phosphorylation using NetSurfP-2.0 (Klausen et al., 2019) (http://www.cbs.dtu.dk/services/NetSurfP-2.0/). Predictions with a minimum probability of 0.5 were then selected. The Wilcoxon's test was used to calculate *P*-values and Wolfpsort software (https://www.wolfpsort.hgc.jp/) was used to predict subcellular localization of succinylated, acetylated, and phosphorylated proteins in the mouse hippocampus.

Functional annotation and pathway analysis

Gene Ontology (GO) annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using Omicsbean (http://www.omicsbean.cn/). *P*-values were calculated using Fisher's exact test (hypergeometric test), with P<0.05 indicating significance.

Prediction of phosphorylation kinase-substrate and kinase activities

Site-specific kinase-substrate relationships (ssKSRs) were predicted using iGPS1.0 software and the GPS2.0 algorithm5. Protein-protein interaction (PPI) information served as the contextual factor to filter out potential false-positives. The parameter "Interaction" was set to "Exp./String" and a "medium" threshold was selected. Kinase activity was predicted by gene set enrichment analysis (GSEA) and normalized enrichment scores (NES) were considered to represent kinase activity scores. Kinases predicted to have positive or negative activity and significantly differentially expressed phosphorylation sites were used to construct a kinase-substrate regulatory network.

PPI network analysis and construction of multiple PTM regulatory networks

PPI networks were constructed to identify succinylated, acetylated, and phosphorylated proteins using the STRING database (http://string-db.org/), with visualization performed using Cytoscape (v.3.9.1).

Metagenomic analysis of fecal samples

Total genomic DNA was extracted from fecal samples of SPF mice (8 weeks old, n=6), according to the manufacturer's instructions. The DNA was fragmented to an average size of approximately 400 bp for paired-end library construction using NEXTFLEX® Rapid DNA-Seq (Bioo Scientific, USA). Paired-end sequencing was performed on the Illumina NovaSeq 6000 platform (Illumina, USA) at Majorbio Bio-Pharm Technology (China). Data were analyzed using the Majorbio Cloud Platform (www.majorbio.com). Open reading frames (ORFs) from each assembled contig were predicted by Prodigal/MetaGene (http://metagene.cb.k.u-tokyo.ac.jp/, see Supplementary Materials for further details).

Correlation between gut microbiome and PTMs

Gene taxonomy was identified from the NR (RefSeq non-redundant proteins) database in the bacterial domain. The operational taxonomic unit (OTU) table was collapsed at the species level and filtered for rare taxa by only including taxa with an average relative abundance of at least 0.1% in all samples (Dayama et al., 2020). In total, 134 taxa (species) were identified. Overlapping KEGG pathways shared between modified proteins and genes corresponding to microbiota were screened. Pearson correlation analysis of overlapping pathways and the 134 taxa was performed.

Statistical analysis

All statistical analyses (*t*-test, Fisher's exact test, Pearson correlation analysis) were conducted using IBM SPSS software (v.21.0; SPSS, USA), with P<0.05 indicating statistical significance.

RESULTS

Global landscape of microbial regulation of succinylation, acetylation, and phosphorylation proteomes in mouse hippocampus

Succinylome and metagenomic analyses were integrated with previously reported lysine acetylome and protein phosphorylome data from the hippocampus of GF mice (Wang et al., 2020b; Yu et al., 2021). We aimed to reveal the role of the gut microbiota in PTM regulation from both physiological and biological perspectives. The study design and workflow are shown in Figure 1. In total, 584 succinylation sites were found on 455 proteins with a distribution of mass error near zero (<0.02 Da), indicating mass accuracy appropriate for MS data (Figure 2A). Most peptides ranged between seven and 20 amino acids in length (Figure 2B), suggesting satisfactory sample preparation. Between 1–9 Ksuc sites were found per peptide, with most containing only one (Figure 2C). A fold-change cutoff of >1.2 or <0.83 and a *P*-value threshold of <0.05 were used to select differentially modified sites based on volcano plots, including 54 up-regulated succinylation sites on 91 proteins and 99 down-regulated succinylation sites on 51 proteins in GF mice compared to SPF mice (Figure 2D; Supplementary Table S1). Most succinylated protein sites enriched in oxidative phosphorylation were down-regulated,

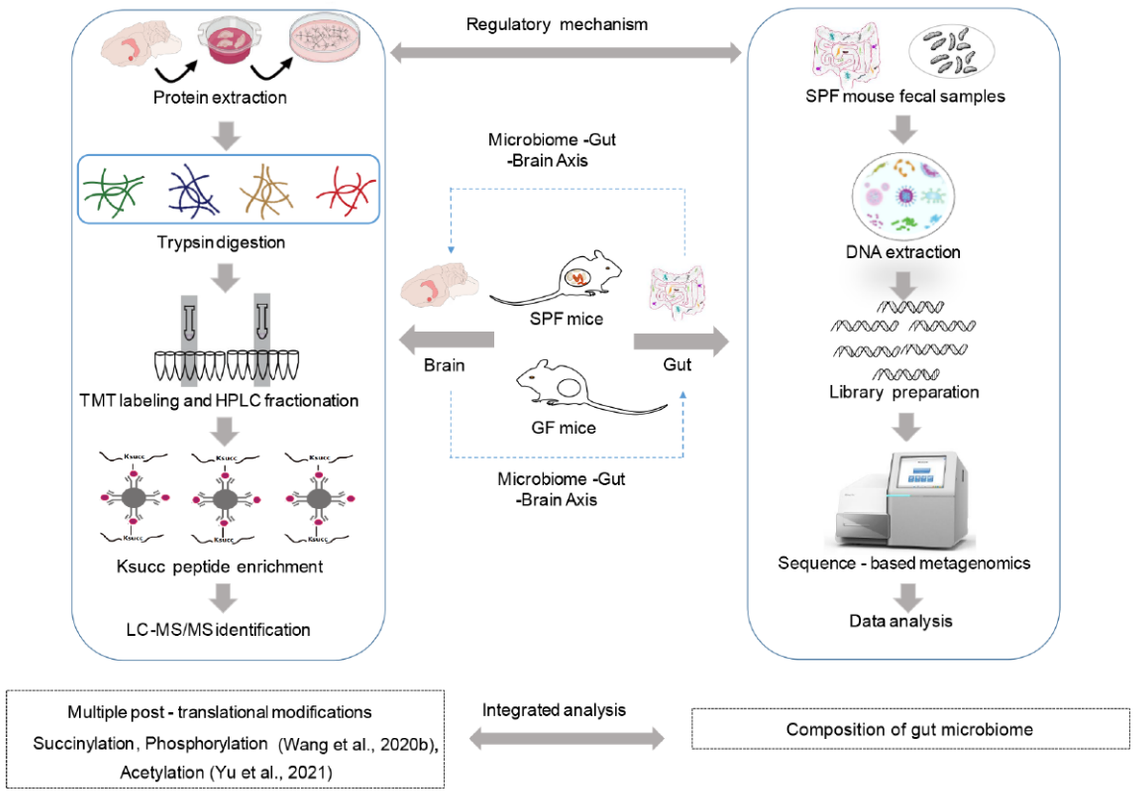


Figure 1 Experimental workflow

GF, germ-free; SPF, specific pathogen-free; LC-MS/MS, liquid chromatography-tandem mass spectrometry.

including Atp5a1, Atp6v1e1, Atp5h, and Ndufv3.

Acetylation and phosphorylation were more susceptible to regulation by the gut microbiome than succinylation in GF mice. In total, 986 acetylation sites on 543 acetylated proteins and 6 945 sites on 2 370 phosphorylated proteins were identified in GF mice (Wang et al., 2020b) (Figure 2F), including 106 up-regulated phosphorylation sites on 88 proteins, 503 down-regulated phosphorylation sites on 349 proteins (Supplementary Table S2), 138 up-regulated acetylation sites on 104 proteins, and 43 down-regulated acetylation sites on 32 proteins (Supplementary Table S3). Overlap analysis identified 84 proteins co-modified by phosphorylation, succinylation, and acetylation (Figure 2E), with four proteins differentially expressed between GF and SPF mice (Figure 2F), including two up-regulated (Figure 2G) and one down-regulated (Figure 2H). Interactions between succinylation and acetylation were analyzed, revealing competitive binding at lysine residues. Of these, 63 sites were modified by both succinylation and acetylation (Figure 2I), eight of which were differentially modified between GF and SPF mice (Figure 2J). These findings provide a comprehensive overview of multiple PTMs in hippocampal proteins regulated by gut microbiota.

Motif analysis of microbial modified sites

Using the motif-x algorithm, conserved sequences surrounding the modified sites were analyzed. A preference for Pro and Lys residues at the +1 position, Glu residues between the −3 and −1 position and at the +2 position, and Asp residues at the +2 position relative to the Ksucc sites was observed (Figure 3A). Thus, *EE*Ksucc* motifs were the most conserved sequences for Ksucc sites (Figure 3B; Supplementary Table S4). The Ksucc*E motif has been identified as the most conserved sequence for Ksucc sites regulated by gut microbiota dysbiosis in a mouse model of depression (Liu et al., 2021). This suggests different structural preferences for different subtypes of gut microbiota dysbiosis.

Prediction of secondary structures for different PTMs

To further ascertain whether the function and structure of proteins were influenced by gut microbiota under physiological conditions, analyses were conducted on proteins showing differential PTMs. In total, 40.36% of Ksucc sites were found in ordered regions (34.12% in α-helices and 6.24% in β-strands), while 59.63% of sites were located in disordered regions (Figure 3C). Previous studies identified Kac sites primarily in disordered regions (62.1% in coils), with a smaller proportion in ordered regions (31.63% in α-helices and 6.26% in β-strands) (Wang et al., 2020b; Yu et al., 2021) (Supplementary Figure S1). Our results also showed 83.32% of serine phosphorylation sites were located in disordered regions, while only 16.66% of sites were located in ordered regions (11.68% in α-helices and 4.98% in β-strands); 81.96% of threonine phosphorylation sites were located in disordered regions, while 18.02% of sites were located in ordered regions (11.91% in α-helices and 6.11% in β-strands) (Supplementary Figure S2). Gut microbiota-modified succinylation sites were more prevalent in α-helices, while acetylation or phosphorylation sites were more prevalent in disordered coils. Surface accessibility analysis identified fewer phosphorylation, acetylation, or succinylation sites compared with corresponding non-modified residues. Thus, phosphorylation, acetylation, and succinylation produced by gut microbiota under physiological conditions may influence the functions and surface properties of proteins.

Subcellular localization of different PTMs

Subcellular localization predictions for differentially expressed

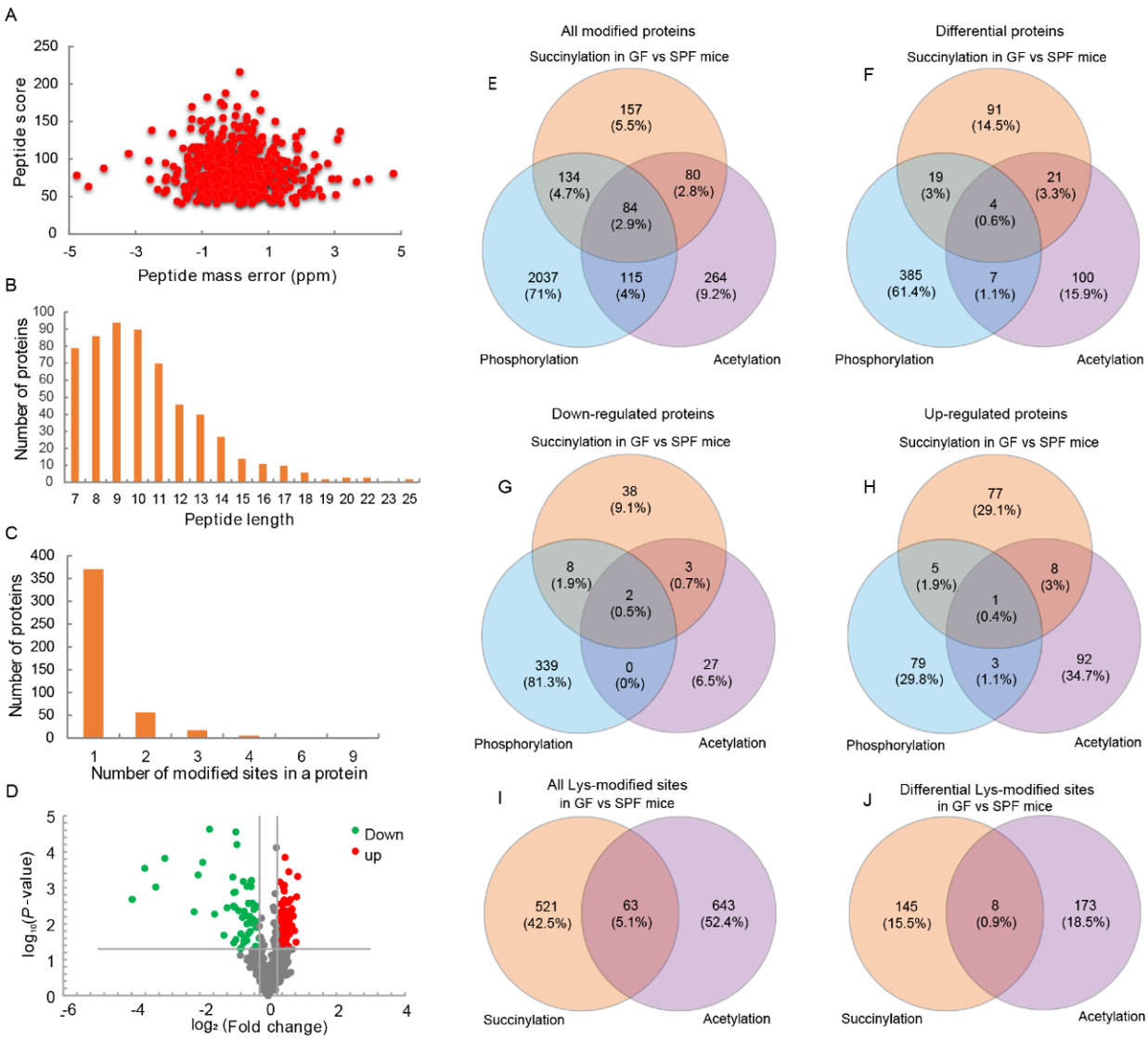


Figure 2 Phosphorylation, acetylation, and succinylation proteome profiles in hippocampal tissue from GF mice

A: Distribution of mass error of all succinylated peptides. B: Distribution of peptide lengths of all succinylated peptides. C: Number of succinylation sites within each modified protein. D: Volcano plot of differentially expressed succinylated proteins in GF vs. SPF mice. E: Venn diagram showing total number of common and distinct hippocampal proteins modified by phosphorylation, acetylation, and succinylation in GF vs. SPF mice. F: Venn diagram showing number of common and distinct differentially expressed hippocampal proteins modified by phosphorylation, acetylation, and succinylation in GF vs. SPF mice. G: Venn diagram showing number of common and distinct down-regulated hippocampal proteins modified by phosphorylation, acetylation, and succinylation in GF vs. SPF mice. H: Venn diagram showing number of common and distinct up-regulated hippocampal proteins modified by phosphorylation, acetylation, and succinylation in GF vs. SPF mice. I: Venn diagram showing overlap between Lys acetylation and succinylation sites in GF vs. SPF mice. J: Venn diagram showing overlap between differentially expressed Lys acetylation and succinylation sites in GF vs. SPF mice.

succinylated, acetylated, and phosphorylated proteins showed that Ksucc modifications were abundant in the cytoplasm (84 proteins, 63%), plasma membrane (47 proteins, 35%), nucleus (31 proteins, 23%) and mitochondria (24 proteins, 18%) (Figure 3D). Earlier studies identified gut microbiota-regulated acetylated proteins in the mitochondria (35%), cytoplasm (32%), and nucleus (Wang et al., 2020b; Yu et al., 2021). Most microbiota-regulated phosphorylated proteins were localized in the nucleus (181 proteins, 55%), plasma membrane (56 proteins, 17%), cytoplasm (49 proteins, 14%), and mitochondria (17 proteins, 5%) (Wang et al., 2020b). The patterns suggest a prominence of Ksucc modifications in cytoplasmic proteins, Kac sites in mitochondrial proteins, and phosphorylation in nuclear proteins. These findings suggest that the gut microbiota may exert regulatory effects on various PTMs across distinct cellular compartments.

Functional annotation and enrichment analysis of different PTMs

GO functional annotations were conducted for all modified proteins. The analysis produced classifications in the cell component (CC) category, biological process (BP) category (including single-organism cellular processes, cellular component organization, and biogenesis), and molecular function (MF) category (Supplementary Figure S4).

Previous studies have reported GO annotations for acetylated and phosphorylated proteins (Wang et al., 2020b; Yu et al., 2021). In our study, succinylated proteins were enriched in cytoplasmic part (80%), single-organism cellular

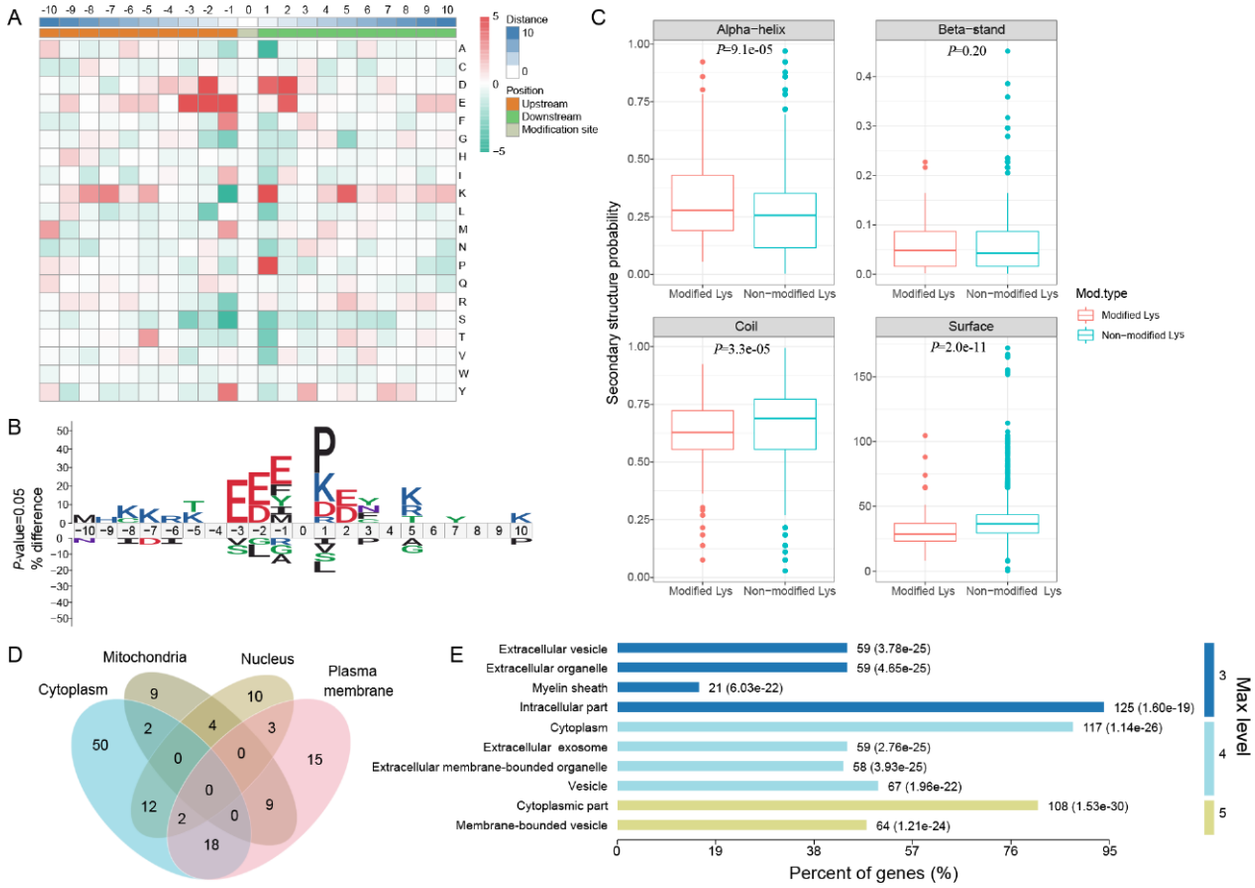


Figure 3 Properties of gut microbiota-regulated succinylation proteome

A: Heatmaps showing features of flanking sequences for all Lys succinylation sites. B: IceLogo diagram showing over-represented amino acids around succinylation site. C: Secondary structure distribution and surface accessibility prediction of significant succinylation sites. D: Venn diagram showing cellular localization of differentially succinylated proteins. E: Bar graph showing cellular component enrichment based on GO annotation analysis of differential succinylated proteins. Max level means maximal annotated level of this term in GO graph (tree) and number indicates depth of GO term level.

process (29%), and protein binding within the CC, BP, and MF categories, respectively (Supplementary Figure S5).

KEGG pathway enrichment analysis showed that succinylated proteins were primarily enriched in metabolic pathways, carbon metabolism, and oxidative phosphorylation (Supplementary Table S5), while acetylated proteins were mainly enriched in carbon metabolism, tricarboxylic acid (TCA) cycle, and oxidative phosphorylation (Supplementary Table S6), and phosphorylated proteins were mainly enriched in glutamatergic synapse, synaptic vesicle cycle, cAMP signaling pathway, and oxytocin signaling pathways (Supplementary Table S7). These observations suggest that different PTMs play distinct roles in different biological processes, with acetylation predominantly affecting metabolic pathways and phosphorylation primarily impacting intracellular signaling pathways.

Overlap analysis revealed the shared pathways among the three PTMs, including oxytocin signaling, glucagon signaling, long-term potentiation, endocrine, and other factor-regulated calcium reabsorption, synaptic vesicle cycle, and dopaminergic synapse pathways (Figure 4). Notably, both phosphorylation and succinylation influenced the GnRH signaling, cholinergic synapse, and axon guidance pathways. Both phosphorylation and acetylation regulated the GABAergic synapse and glutamatergic synapse pathways, while succinylation and acetylation regulated the carbon metabolism and oxidative phosphorylation pathways (Supplementary Table S8).

Construction of PPI network for modified proteins

To investigate the complicated interactions between differentially modified proteins and their synergistic effects on functional pathways, a PPI network of differentially expressed succinylated, acetylated, and phosphorylated proteins was generated using high-confidence interactions (≥0.7) with STRING. Highly connected node clusters, indicative of interactions with other modified proteins, were defined using the MCODE plug-in. Four such clusters, which spanned various pathways, including oxidative phosphorylation, ribosome, PI3K-Akt signaling pathway, and TCA cycle, were identified (Figure 5). Among these pathways, oxidative phosphorylation and TCA cycle were modified by succinylation and acetylation, while ribosome and PI3K-Akt signaling pathway were modified by phosphorylation.

The energy production pathways, oxidative phosphorylation, and TCA cycle were further explored. Most oxidative phosphorylation proteins were modified by acetylation and/or succinylation, with only NADH dehydrogenase (ubiquinone) Fe-S protein 8 (Ndufs8), a core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I), modified by phosphorylation (Figure 6). TCA cycle proteins were mainly modified by acetylation, except for the pyruvate dehydrogenase complex component dihydrolipoamide

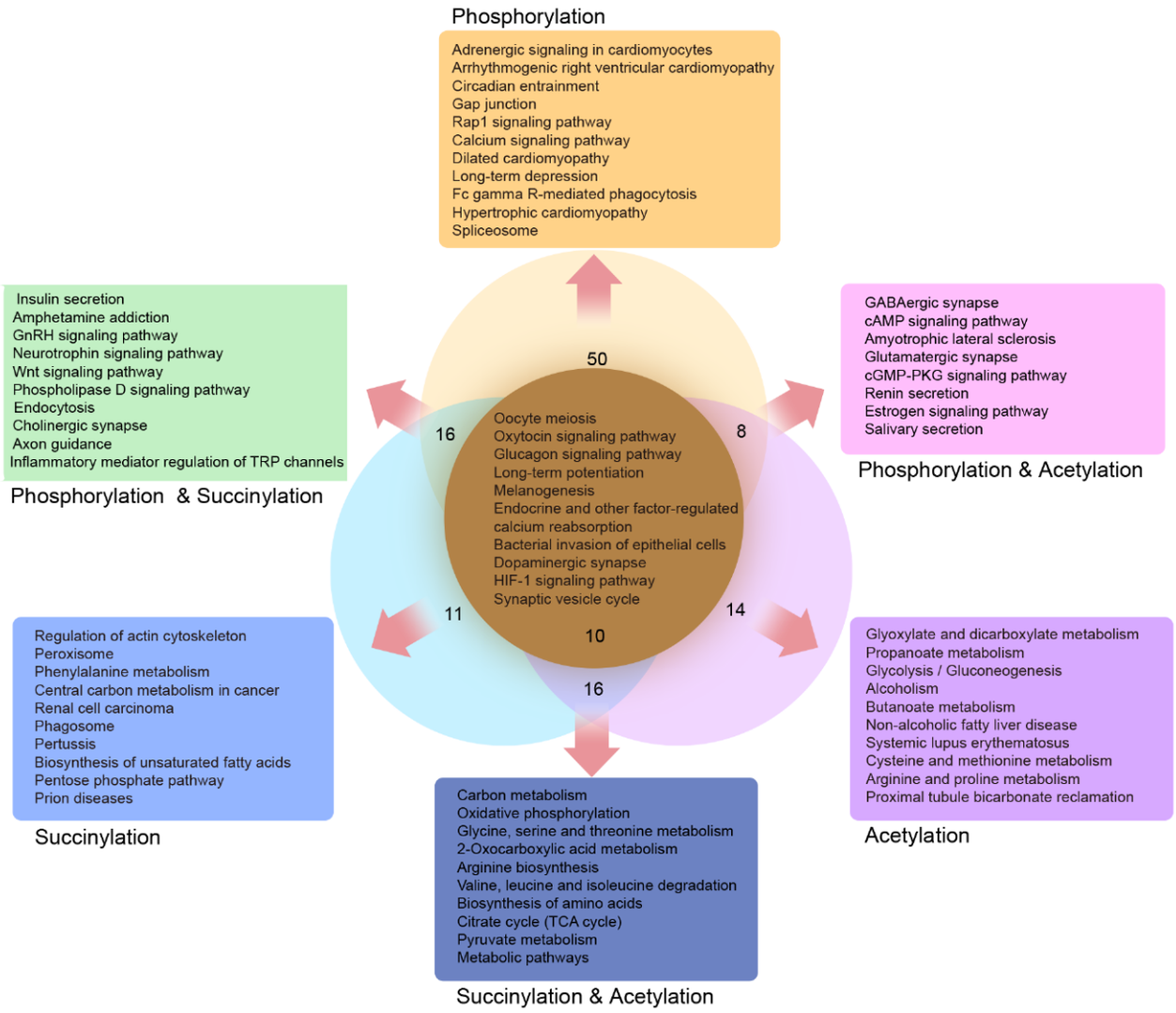


Figure 4 KEGG pathway overlap among differentially expressed phosphorylated, acetylated, and succinylated proteins

Venn diagram showing number of common and distinct proteins modified by phosphorylation, acetylation, and succinylation. Brown area indicates common KEGG pathways enriched in phosphorylated, succinylated, and acetylated proteins; blue area indicates unique KEGG pathways enriched in succinylated proteins; purple area indicates unique KEGG pathways enriched in acetylated proteins; and orange area indicates unique KEGG pathways enriched in phosphorylated proteins.

dehydrogenase (Dld), which was modified by both succinylation and acetylation. These findings highlight both cross-talk as well as specific roles for different PTMs in modulating intracellular pathways.

Comprehensive analysis of gut microbiome and PTMs

KEGG pathway analysis was conducted on gut microbial genes to reveal specifically enriched pathways (Supplementary Table S9). Further overlap analysis identified eight shared pathways between microbial genes and modified proteins, including biosynthesis of amino acids, glycolysis/gluconeogenesis, cysteine and methionine metabolism, alanine, aspartate and glutamate metabolism, pyruvate metabolism, ribosome, oxidative phosphorylation, and 2-oxocarboxylic acid metabolism. Pearson correlation analysis indicated that 13 of the 134 taxa were significantly correlated with these common pathways, which were enriched in 37 differentially expressed succinylated and acetylated proteins. These common pathways were used as a “bridge” to illustrate correlations between gut microbiota and modified proteins (Figure 7). Most taxa belonged to Bacteroidetes, Muribaculaceae_bacterium, *Bacteroides_acidifaciens*, Porphyromonadaceae_bacterium, and *Alistipes*_sp._Z76. A positive correlation between succinylation and taxa was found in these pathways, while a negative association was detected between acetylation and taxa. In the absence of gut microbiota, the GF mice showed reduced pathway activity, decreased succinylation, and increased acetylation. These findings suggest that specific taxa may play roles in modulating brain functions via PTM regulation.

DISCUSSION

The gut microbiota is associated with various biological processes essential for maintaining CNS and peripheral nervous system homeostasis. Various molecular and signaling pathways enable bidirectional communication between the gut microbiota and brain, thus establishing the microbiota-gut-brain axis (Osadchiy et al., 2019; Schoch et al., 2022). Advancements in sequencing technology have provided a powerful approach for exploring the associations between the gut microbiome and brain diseases and facilitating the

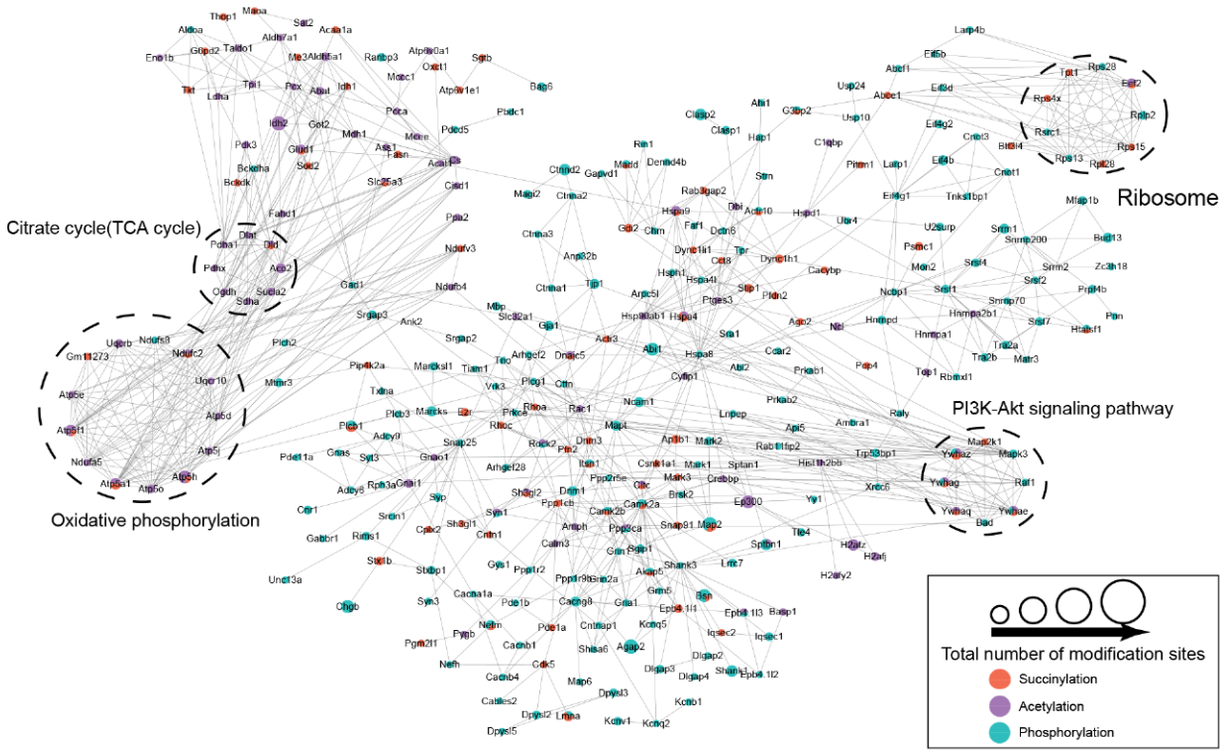


Figure 5 PPI network analysis of differentially expressed phosphorylated, acetylated, and succinylated proteins

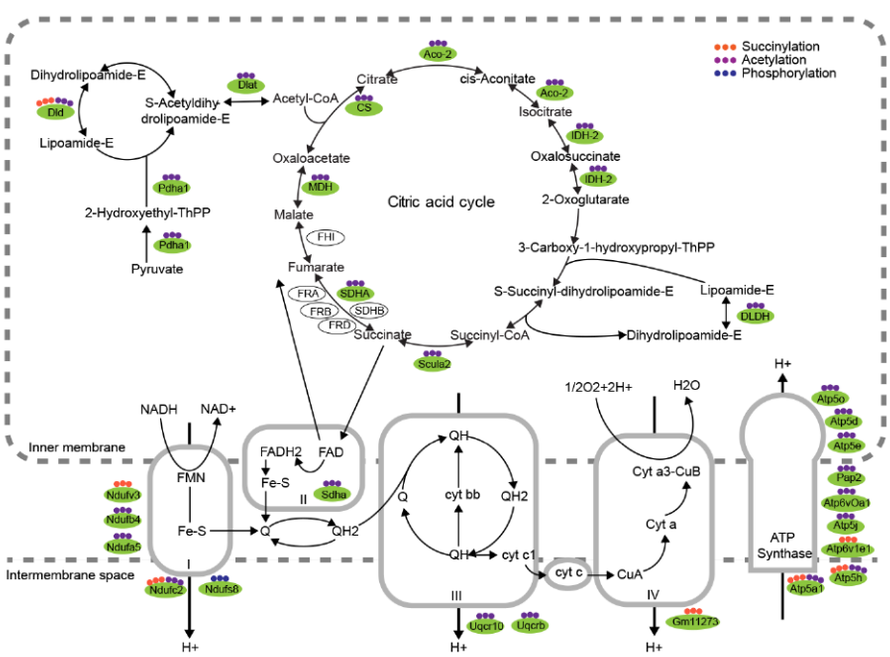


Figure 6 Details of phosphorylation, acetylation, and succinylation alterations in relative proteins involved in oxidative phosphorylation pathway

development of novel interventions targeting the gut to regulate brain function and create "gut treatments for brain diseases".

In practice, there is a pressing need to elucidate both the identity and the mechanisms through which the gut microbiota governs brain function. While previous studies have focused on changes in gut microbiota and regulatory roles during specific diseases in patients or animal models (Parker et al., 2022), structural variants in the gut microbiome during a state of disease cannot be ruled out (Zeevi et al., 2019). Microbial strains may exhibit different genomic structures under different pathological and physiological states (Durrant & Bhatt, 2019; Wang et al., 2021); this complexity adds to the challenge of developing gut microbiota-based interventions for brain diseases using gut microbiota obtained from healthy donors. Thus, exploring the composition of gut microbiota and their interactions with the brain under physiological conditions is vital, but remains relatively understudied.

PTMs are sensitive to environmental stimuli (Zhu et al., 2022). By analyzing global PTM profiles, one can potentially

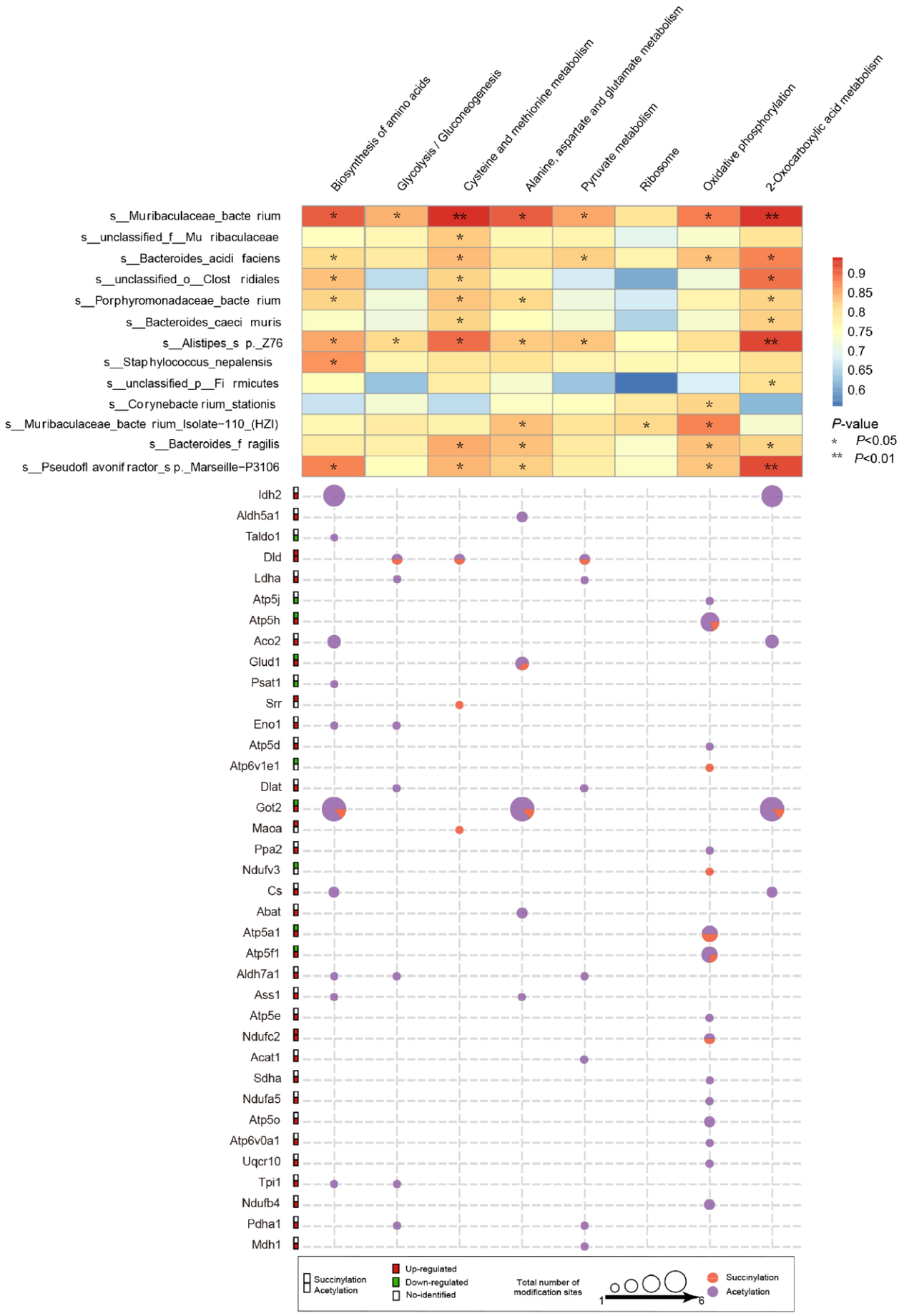


Figure 7 Integration analysis of gut microbiota and protein post-translational modifications

elucidate the mechanisms by which the gut microbiota regulates brain function and behavioral phenotypes. Here, we generated a panoramic map of the microbiota-regulated succinylome, acetylome, and phosphorylome in the hippocampus of GF and SPF mice to explore the regulatory role of the gut microbiota in modulating PTMs and better understand the gut-brain axis from a PTM perspective.

Our results showed that acetylation and phosphorylation were more susceptible to gut microbiome effects than succinylation and together constituted the vast majority of PTMs, as shown in previous research (Ebert et al., 2022). Phosphorylation, catalyzed by more than 800 protein kinases, is one of the most pervasive PTM in eukaryotic cells, affecting up to one third of the proteome (Humphrey et al., 2015; Pieroni et al., 2020) and showing involvement in almost every cellular process (Humphrey et al., 2015). Acetylation is involved in physiology and disease, including DNA damage repair, gene transcription, autophagy, and metabolism (Narita et al., 2019). In earlier work, we noted that the deacetylases, NAD-dependent deacetylase sirtuin 1 (SIRT1) and SIRT2, were down-regulated in the hippocampus of GF mice compared to SPF mice (Liu et al., 2021), while SIRT5, which regulates de-succinylation, remained unaffected. This may account for the observed dominance of acetylation over succinylation in GF mice.

In addition, previous research has shown that succinylation regulated by the gut microbiota tended to occur near acidic amino acids in cytoplasmic proteins, while acetylation was prominent in mitochondrial proteins and phosphorylation was dominant in nuclear proteins, underlining the specific biological processes tied to each PTM type (Liu et al., 2021). We also identified distinct pathways enriched by different PTMs. Notably, peroxisome and biosynthesis of unsaturated fatty acids were associated with succinylated proteins. In contrast, glycolysis/gluconeogenesis and cysteine and methionine metabolism with acetylated proteins and calcium signaling pathway, Fc gamma R-mediated phagocytosis, and spliceosome were associated with phosphorylated proteins. Cross-talk between different PTM types was also evident in the nervous system. For example, neurotransmitter release, including cholinergic, GABAergic, glutamatergic, and dopaminergic systems, were co-modified by at least two types of PTMs. Neurotransmitters play many roles in the CNS (Wu et al., 2022), and disruptions in their function are hallmark features of neuropsychiatric diseases (Snyder & Ferris, 2000) such as depression (Duman et al., 2019). Neurotransmitters also mediate bidirectional gut-brain communication (Lai et al., 2021; Mittal et al., 2017). Thus, the gut microbiota appears to regulate brain function via cross-talk of PTMs affecting proteins in various biological processes.

In contrast to phosphorylation, both acetylation and succinylation target lysine, suggesting an increased potential for cross-talk between them. Our results showed that many proteins involved in metabolic pathways, such as oxidative phosphorylation and the TCA cycle, are co-modified by acetylation and succinylation. Oxidative phosphorylation and the TCA cycle are pivotal processes in living organisms, representing the main mechanisms initially providing energy for neuronal activity (Hall et al., 2012; Wang et al., 2019). Oxidative phosphorylation comprises five multi-subunit complexes, designated CI–CV. Previous studies have identified the occurrence of phosphorylation on CI–CV, acetylation on CI, CII, and CV, and glycosylation on CI, CII, and CV (Chen et al., 2008; Guo et al., 2017; Koopman et al., 2013). In the present study, the absence of gut microbiota precipitated increased acetylation levels of complexes, particularly adenosine triphosphate (ATP) synthase subunits (CV), such as Atp5d, Atp5h, Atp5o, Atp5a1, ATP5f1, Atp5e, as well as decreased activity and mitochondrial respiration (Rahman et al., 2014; Vassilopoulos et al., 2014). Impairment of ATP synthase increases the NAD+/NADH ratio, subsequently reducing protein succinylation (Chen et al., 2017a) and regulating Ndufv3, Ndufb4, Ndufa5, Ndufc2, and Ndufs8. This may explain the divergent patterns of acetylation and succinylation observed for certain related proteins, like Atp5h, Atp5a1, and Atp5f1, in our results. Key rate-limiting enzymes of the TCA cycle, including citrate synthase (Krug et al., 2019), isocitrate dehydrogenase (IDH), and α-ketoglutarate dehydrogenase (α-KGDH), were enriched with acetylation and succinylation sites, underscoring the regulatory role of gut microbiota in energy metabolism.

These energy-producing pathways are also co-modified by acetylation and succinylation in mice introduced with gut microbiota from patients diagnosed with major depressive disorder (Liu et al., 2021), accompanied by a down-regulation in the expression of pathway-relevant genes (Qi et al., 2020). However, even for the same pathway, the source of gut microbiota, whether from pathological and physiological conditions, can differentially regulate host health. Our previous work revealed differences in the modification of complex subunits in the oxidative respiratory chain and enzymes in the TCA cycle when modified by gut microbiota from patients versus healthy individuals (Liu et al., 2021; Wang et al., 2020b; Yu et al., 2021). Furthermore, gut microbiota-produced butyrate can induce the phosphorylation of Ser37 and Tyr105, acetylation of Lys305, and succinylation of Lys311 on pyruvate kinase isoform 2 (PKM2), a key enzyme in pyruvate metabolism, to suppress the proliferation of colorectal cancer cells (Bettaieb et al., 2013; Li et al., 2018b; Lv et al., 2011; Wang et al., 2017; Yang et al., 2012). Depression-associated taxa can induce hyper-succinylation of Lys224 and Lys247 of PKM, without affecting acetylation and phosphorylation (Liu et al., 2021). Notably, these changes are absent under normal physiological states. Even when focusing on the same modification site of a specific protein, discrepancies emerge between pathological and physiological states. For example, depression-related gut microbiota dysbiosis can reduce acetylation of Lys321 on the Cs protein (Liu et al., 2021), whereas an absence of gut microbiota can enhance its acetylation. These findings suggest that different taxa may target different aspects for brain function regulation, and the mechanisms underlying microbiota-host interactions should be explored in further studies.

Bacteroidetes and Firmicutes are the dominant phyla in the gut (The Human Microbiome Project Consortium, 2012), with Bacteroidetes comprising over half of the gut microbiome (Qin et al., 2010; The Human Microbiome Project Consortium, 2012) and playing a key role in providing their host with energy (Johnson et al., 2017). In this study, we found a significant correlation between Bacteroidetes bacteria and metabolic pathways related to energy metabolism, including Muribaculaceae_bacterium, *Bacteroides_acidifaciens*, and *Alistipes*_sp._Z76, which participate in glycolysis/gluconeogenesis and pyruvate metabolism processes. Many energy-related metabolic pathways are influenced by taxa from various phyla. For example, Bacteroidetes and

Firmicutes exhibit synergistic effects on oxidative phosphorylation, while *Bifidobacterium* modulates metabolic processes and enhances mitochondrial activity in gut regulatory T cells (Tregs) (Sun et al., 2020), suggesting that a group of gut microorganisms, rather than just a single microorganism, modulates physiological processes essential for health (Miyauchi et al., 2020). Negative correlations between acetylation and taxa suggest that such taxa may reduce the activity of enzymes involved in energy metabolism by promoting the acetylation of relevant proteins (Rahman et al., 2014; Vassilopoulos et al., 2014). Our study highlights several potential targets associated with these taxa of interest, indicating possible gut-brain axis mechanisms through which the gut microbiota affects host health by modifying protein function.

This study has several limitations to mention. First, to avoid potential confounders arising from the influence of sex hormones, only male mice were used. Second, while the SPF mice used for PTM analysis and for metagenomes were the same age and housed in the same environmental conditions, they were not from the same batch. Third, although our results are promising and were obtained using highly reliable methods, experimental validation has not yet been performed. Further studies involving different brain regions and including female mice are necessary to understand how the gut microbiome influences host health via PTMs.

CONCLUSIONS

In conclusion, we used GF and SPF mice to determine proteomic changes in the hippocampus, particularly quantification of succinylation, phosphorylation, and acetylation, arising from the absence of gut microbiota. Our objective was to clarify the role of gut microbiota on brain functionality via PTM dynamics under physiological conditions. We further constructed a panoramic map of gut microbiota-regulated succinylation, acetylation, and phosphorylation and identified cross-talk/relative independence between different types of PTMs in modulating intracellular pathways. Notably, we identified positive correlations between succinylation and taxa and negative correlations between acetylation and taxa, with implications for the modulation of intracellular pathways. These findings should help advance our understanding of how the gut microbiome regulates brain function and behavioral phenotypes by modulating PTMs under physiological conditions.

DATA AVAILABILITY

The raw metagenomics sequencing reads can be downloaded from the NCBI (PRJNA1003305), China National Center for Bioinformation (PRJCA018916), and Science Data Bank databases (DOI: 10.57760/sciencedb.j00139.00058).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

H.W., P.J., and P.X. conceived and designed the study, H.Y.W. and L.X.L. wrote the manuscript, X.Y.C. and Y.D.Z. analyzed the data, W.W.L., W.X.L., L.W., and X.L.M. revised the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We greatly appreciate the technical support and helpful suggestions of other Xie laboratory members.

REFERENCES

Bettaieb A, Bakke J, Nagata N, et al. 2013. Protein tyrosine phosphatase 1B regulates pyruvate kinase M2 tyrosine phosphorylation. *Journal of Biological Chemistry*, **288**(24): 17360−17371.

Chen CL, Chen JF, Rawale S, et al. 2008. Protein tyrosine nitration of the flavin subunit is associated with oxidative modification of mitochondrial complex II in the post-ischemic myocardium. *Journal of Biological Chemistry*, **283**(41): 27991−28003.

Chen HL, Xu H, Potash S, et al. 2017a. Mild metabolic perturbations alter succinylation of mitochondrial proteins. *Journal of Neuroscience Research*, **95**(11): 2244−2252.

Chen JJ, Xie J, Zeng BH, et al. 2019. Absence of gut microbiota affects lipid metabolism in the prefrontal cortex of mice. *Neurological Research*, **41**(12): 1104−1112.

Chen JJ, Zeng BH, Li WW, et al. 2017b. Effects of gut microbiota on the microRNA and mRNA expression in the hippocampus of mice. *Behavioural Brain Research*, **322**: 34−41.

Cryan JF, Dinan TG. 2012. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nature Reviews Neuroscience*, **13**(10): 701−712.

Cryan JF, O'Riordan KJ, Cowan CSM, et al. 2019. The microbiota-gut-brain axis. *Physiological Reviews*, **99**(4): 1877−2013.

Dayama G, Priya S, Niccum DE, et al. 2020. Interactions between the gut microbiome and host gene regulation in cystic fibrosis. *Genome Medicine*, **12**(1): 12.

Duman RS, Sanacora G, Krystal JH. 2019. Altered connectivity in depression: GABA and glutamate neurotransmitter deficits and reversal by novel treatments. *Neuron*, **102**(1): 75−90.

Durrant MG, Bhatt AS. 2019. Microbiome genome structure drives function. *Nature Microbiology*, **4**(6): 912−913.

Ebert T, Tran N, Schurgers L, et al. 2022. Ageing - Oxidative stress, PTMs and disease. *Molecular Aspects of Medicine*, **86**: 101099.

Fan Y, Pedersen O. 2021. Gut microbiota in human metabolic health and disease. *Nature Reviews Microbiology*, **19**(1): 55−71.

Fang XP, Xin Y, Sheng ZL, et al. 2018. Systematic identification and analysis of lysine succinylation in strawberry stigmata. *Journal of Agricultural and Food Chemistry*, **66**(50): 13310−13320.

Foster JA, McVey Neufeld KA. 2013. Gut-brain axis: how the microbiome influences anxiety and depression. *Trends in Neurosciences*, **36**(5): 305−312.

Fung TC, Olson CA, Hsiao EY. 2017. Interactions between the microbiota, immune and nervous systems in health and disease. *Nature Neuroscience*, **20**(2): 145−155.

Gao Y, Lee H, Kwon OK, et al. 2019. Global proteomic analysis of lysine succinylation in zebrafish (*Danio rerio*). *Journal of Proteome Research*, **18**(10): 3762−3769.

Gareau MG, Wine E, Rodrigues DM, et al. 2011. Bacterial infection causes stress-induced memory dysfunction in mice. *Gut*, **60**(3): 307−317.

Guo RY, Zong S, Wu M, et al. 2017. Architecture of human mitochondrial respiratory megacomplex $I_2III_2IV_2$. *Cell*, **170**(6): 1247−1257.e12.

Hall CN, Klein-Flugge MC, Howarth C, et al. 2012. Oxidative phosphorylation, not glycolysis, powers presynaptic and postsynaptic mechanisms underlying brain information processing. *Journal of Neuroscience*, **32**(26): 8940−8951.

Humphrey SJ, James DE, Mann M. 2015. Protein phosphorylation: a major switch mechanism for metabolic regulation. *Trends in Endocrinology & Metabolism*, **26**(12): 676−687.

Johnson EL, Heaver SL, Walters WA, et al. 2017. Microbiome and metabolic disease: revisiting the bacterial phylum Bacteroidetes. *Journal of Molecular Medicine*, **95**(1): 1−8.

Klausen MS, Jespersen MC, Nielsen H, et al. 2019. NetSurfP-2.0: improved prediction of protein structural features by integrated deep learning. *Proteins*, **87**(6): 520−527.

Koopman WJH, Distelmaier F, Smeitink JAM, et al. 2013. OXPHOS mutations and neurodegeneration. *The EMBO Journal*, **32**(1): 9−29.

Krug K, Mertins P, Zhang B, et al. 2019. A curated resource for phosphosite-specific signature analysis. *Molecular & Cellular Proteomics*, **18**(3): 576−593.

Lai YJ, Liu CW, Yang YF, et al. 2021. High-coverage metabolomics uncovers microbiota-driven biochemical landscape of interorgan transport and gut-brain communication in mice. *Nature Communications*, **12**(1): 6000.

Li B, Guo KN, Zeng L, et al. 2018a. Metabolite identification in fecal microbiota transplantation mouse livers and combined proteomics with chronic unpredictive mild stress mouse livers. *Translational Psychiatry*, **8**(1): 34.

Li QR, Cao LJ, Tian Y, et al. 2018b. Butyrate suppresses the proliferation of colorectal cancer cells via targeting pyruvate kinase M2 and metabolic reprogramming. *Molecular & Cellular Proteomics*, **17**(8): 1531−1545.

Liu LX, Wang HY, Rao XC, et al. 2021. Comprehensive analysis of the lysine acetylome and succinylome in the hippocampus of gut microbiota-dysbiosis mice. *Journal of Advanced Research*, **30**: 27−38.

Liu LX, Wang HY, Yu Y, et al. 2020. Microbial regulation of a lincRNA-miRNA-mRNA network in the mouse hippocampus. *Epigenomics*, **12**(16): 1377−1387.

Long-Smith C, O'Riordan KJ, Clarke G, et al. 2020. Microbiota-gut-brain axis: new therapeutic opportunities. *Annual Review of Pharmacology and Toxicology*, **60**: 477−502.

Luczynski P, McVey Neufeld KA, Oriach CS, et al. 2016. Growing up in a bubble: using germ-free animals to assess the influence of the gut microbiota on brain and behavior. *International Journal of Neuropsychopharmacology*, **19**(8): pyw020.

Lukić I, Getselter D, Ziv O, et al. 2019. Antidepressants affect gut microbiota and *Ruminococcus flavefaciens* is able to abolish their effects on depressive-like behavior. *Translational Psychiatry*, **9**(1): 133.

Lv L, Li D, Zhao D, et al. 2011. Acetylation targets the M2 isoform of pyruvate kinase for degradation through chaperone-mediated autophagy and promotes tumor growth. *Molecular Cell*, **42**(6): 719−730.

Marcelino VR, Clausen PTLC, Buchmann JP, et al. 2020. CCMetagen: comprehensive and accurate identification of eukaryotes and prokaryotes in metagenomic data. *Genome Biology*, **21**(1): 103.

Mittal R, Debs LH, Patel AP, et al. 2017. Neurotransmitters: the critical modulators regulating gut-brain axis. *Journal of Cellular Physiology*, **232**(9): 2359−2372.

Miyauchi E, Kim SW, Suda W, et al. 2020. Gut microorganisms act together to exacerbate inflammation in spinal cords. *Nature*, **585**(7823): 102−106.

Morais LH, Schreiber IV HL, Mazmanian SK. 2021. The gut microbiota-brain axis in behaviour and brain disorders. *Nature Reviews Microbiology*, **19**(4): 241−255.

Narita T, Weinert BT, Choudhary C. 2019. Functions and mechanisms of non-histone protein acetylation. *Nature Reviews Molecular Cell Biology*, **20**(3): 156−174.

Osadchiy V, Martin CR, Mayer EA. 2019. The gut-brain axis and the microbiome: mechanisms and clinical implications. *Clinical Gastroenterology and Hepatology*, **17**(2): 322−332.

Parker A, Romano S, Ansorge R, et al. 2022. Fecal microbiota transfer between young and aged mice reverses hallmarks of the aging gut, eye, and brain. *Microbiome*, **10**(1): 68.

Pieroni L, Iavarone F, Olianas A, et al. 2020. Enrichments of post-translational modifications in proteomic studies. *Journal of Separation Science*, **43**(1): 313−336.

Qi XZ, Zhong XG, Xu SH, et al. 2020. Extracellular matrix and oxidative phosphorylation: important role in the regulation of hypothalamic function by gut microbiota. *Frontiers in genetics*, **11**: 520.

Qin JJ, Li RQ, Raes J, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, **464**(7285): 59−65.

Rahman M, Nirala NK, Singh A, et al. 2014. *Drosophila* Sirt2/mammalian SIRT3 deacetylates ATP synthase β and regulates complex V activity. *Journal of Cell Biology*, **206**(2): 289−305.

Rao XC, Liu LX, Wang HY, et al. 2021. Regulation of gut microbiota disrupts the glucocorticoid receptor pathway and inflammation-related pathways in the mouse hippocampus. *Experimental Neurobiology*, **30**(1): 59−72.

Schoch SF, Castro-Mejía JL, Krych L, et al. 2022. From Alpha Diversity to Zzz: interactions among sleep, the brain, and gut microbiota in the first year of life. *Progress in Neurobiology*, **209**: 102208.

Schretter CE. 2020. Links between the gut microbiota, metabolism, and host behavior. *Gut Microbes*, **11**(2): 245−248.

Snyder SH, Ferris CD. 2000. Novel neurotransmitters and their neuropsychiatric relevance. *American Journal of Psychiatry*, **157**(11): 1738−1751.

Sun S, Luo LJ, Liang WH, et al. 2020. *Bifidobacterium* alters the gut microbiota and modulates the functional metabolism of T regulatory cells in the context of immune checkpoint blockade. *Proceedings of the National Academy of Sciences of the United States of America*, **117**(44): 27509−27515.

The Human Microbiome Project Consortium. 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, **486**(7402): 207−214.

Vassilopoulos A, Pennington JD, Andresson T, et al. 2014. SIRT3 deacetylates ATP synthase F_1 complex proteins in response to nutrient- and exercise-induced stress. *Antioxidants & Redox Signaling*, **21**(4): 551−564.

Vuong HE, Pronovost GN, Williams DW, et al. 2020. The maternal microbiome modulates fetal neurodevelopment in mice. *Nature*, **586**(7828): 281−286.

Wang DM, Doestzada M, Chen LM, et al. 2021. Characterization of gut microbial structural variations as determinants of human bile acid metabolism. *Cell Host & Microbe*, **29**(12): 1802−1814.e5.

Wang F, Wang K, Xu W, et al. 2017. SIRT5 desuccinylates and activates pyruvate kinase M2 to block macrophage IL-1β production and to prevent DSS-induced colitis in mice. *Cell Reports*, **19**(11): 2331−2344.

Wang HY, Liu LX, Rao XC, et al. 2020a. Commensal microbiota regulation of metabolic networks during olfactory dysfunction in mice. *Neuropsychiatric Disease and Treatment*, **16**: 761−769.

Wang HY, Liu LX, Rao XC, et al. 2020b. Integrated phosphoproteomic and metabolomic profiling reveals perturbed pathways in the hippocampus of gut microbiota dysbiosis mice. *Translational Psychiatry*, **10**(1): 346.

Wang XB, Chen XZ, Li JR, et al. 2019. Global analysis of lysine succinylation in patchouli plant leaves. *Horticulture Research*, **6**: 133.

Wei MY, Shi S, Liang C, et al. 2019. The microbiota and microbiome in pancreatic cancer: more influential than expected. *Molecular Cancer*, **18**(1): 97.

Wu ZF, Lin DY, Li YL. 2022. Pushing the frontiers: tools for monitoring neurotransmitters and neuromodulators. *Nature Reviews Neuroscience*, **23**(5): 257−274.

Yang WW, Zheng YH, Xia Y, et al. 2012. ERK1/2-dependent phosphorylation and nuclear translocation of PKM2 promotes the Warburg effect. *Nature Cell Biology*, **14**(12): 1295−1304.

Yang YP, Lu Y, Yu PJ, et al. 2022. Characterization of gut microbial alterations in cynomolgus macaques during growth and maturation.

Zoological Research, **43**(2): 176−179.

Yao ZY, Li XH, Zuo L, et al. 2022. Maternal sleep deprivation induces gut microbial dysbiosis and neuroinflammation in offspring rats. *Zoological Research*, **43**(3): 380−390.

Yu Y, Wang HY, Rao XC, et al. 2021. Proteomic profiling of lysine acetylation indicates mitochondrial dysfunction in the hippocampus of gut microbiota-absent mice. *Frontiers in Molecular Neuroscience*, **14**: 594332.

Zeevi D, Korem T, Godneva A, et al. 2019. Structural variation in the gut microbiome associates with host health. *Nature*, **568**(7750): 43−48.

Zeng L, Zeng BH, Wang HY, et al. 2016. Microbiota modulates behavior and protein kinase C mediated cAMP response element-binding protein Signaling. *Scientific Reports*, **6**: 29998.

Zheng P, Wu J, Zhang HP, et al. 2021. The gut microbiome modulates gut-brain axis glycerophospholipid metabolism in a region-specific manner in a nonhuman primate model of depression. *Molecular Psychiatry*, **26**(6): 2380−2392.

Zheng P, Zeng B, Zhou C, et al. 2016. Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host's metabolism. *Molecular Psychiatry*, **21**(6): 786−796.

Zheng P, Zeng BH, Liu ML, et al. 2019. The gut microbiome from patients with schizophrenia modulates the glutamate-glutamine-GABA cycle and schizophrenia-relevant behaviors in mice. *Science Advances*, **5**(2): eaau8317.

Zhou CJ, Rao XC, Wang HY, et al. 2020. Hippocampus-specific regulation of long non-coding RNA and mRNA expression in germ-free mice. *Functional & Integrative Genomics*, **20**(3): 355−365.

Zhu GJ, Jin LF, Sun WC, et al. 2022. Proteomics of post-translational modifications in colorectal cancer: discovery of new biomarkers. *Biochimica et Biophysica Acta* (*BBA*) - *Reviews on Cancer*, **1877**(4): 188735.

Lu et al. *Zool. Res.* 2024, 45(1): 108−124
https://doi.org/10.24272/j.issn.2095-8137.2023.101

ZR | Zoological Research

Article Open Access

Cath-KP, a novel peptide derived from frog skin, prevents oxidative stress damage in a Parkinson's disease model

Huanpeng Lu[1,3,4,#], Jinwei Chai[2,#], Zijian Xu[1,3,4], Jiena Wu[2], Songzhe He[1,3,4], Hang Liao[2], Peng Huang[1,3,4], Xiaowen Huang[5], Xi Chen[1,3,4], Haishan Jiang[1], Shaogang Qu[6,1,3,4,*], Xueqing Xu[2,*]

[1] *Department of Neurology*, *Nanfang Hospital*, *Southern Medical University*, *Guangzhou*, *Guangdong* 510515, *China*

[2] *Guangdong Provincial Key Laboratory of New Drug Screening*, *School of Pharmaceutical Sciences*, *Southern Medical University*, *Guangzhou*, *Guangdong* 510515, *China*

[3] *Guangdong-Hong Kong-Macao Greater Bay Area Center for Brain Science and Brain-Inspired Intelligence*, *Guangzhou*, *Guangdong* 510515, *China*

[4] *Key Laboratory of Mental Health of the Ministry of Education*, *Southern Medical University*, *Guangzhou*, *Guangdong* 510515, *China*

[5] *Department of Dermatology*, *Nanfang Hospital*, *Southern Medical University*, *Guangzhou*, *Guangdong* 510515, *China*

[6] *Department of Neurology*, *Ganzhou Hospital-Nanfang Hospital*, *Southern Medical University*, *Ganzhou*, *Jiangxi* 341001, *China*

ABSTRACT

Parkinson's disease (PD) is a neurodegenerative condition that results in dyskinesia, with oxidative stress playing a pivotal role in its progression. Antioxidant peptides may thus present therapeutic potential for PD. In this study, a novel cathelicidin peptide (Cath-KP; GCSGRFCNLF NNRRPGRLTLIHRPGGDKRTSTGLIYV) was identified from the skin of the Asiatic painted frog (*Kaloula pulchra*). Structural analysis using circular dichroism and homology modeling revealed a unique *αββ* conformation for Cath-KP. *In vitro* experiments, including free radical scavenging and ferric-reducing antioxidant analyses, confirmed its antioxidant properties. Using the 1-methyl-4-phenylpyridinium ion (MPP^+)-induced dopamine cell line and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-induced PD mice, Cath-KP was found to penetrate cells and reach deep brain tissues, resulting in improved MPP^+-induced cell viability and reduced oxidative stress-induced damage by promoting antioxidant enzyme expression and alleviating mitochondrial and intracellular reactive oxygen species accumulation through Sirtuin-1 (Sirt1)/Nuclear factor erythroid 2-related factor 2 (Nrf2) pathway activation. Both focal adhesion kinase (FAK) and p38 were also identified as regulatory elements. In the MPTP-induced PD mice, Cath-KP administration increased the number of tyrosine hydroxylase (TH)-positive neurons, restored TH content, and ameliorated dyskinesia. To the best of our knowledge, this study is the first to report on a cathelicidin peptide demonstrating potent antioxidant and neuroprotective properties in a PD model by targeting oxidative stress. These findings expand the known functions of cathelicidins, and hold promise for the development of therapeutic agents for PD.

Keywords: Cath-KP; Peptide; Parkinson's disease; Oxidative stress; Neuroprotection

INTRODUCTION

Parkinson's disease (PD), related to excessive levels of reactive oxygen species (ROS) and oxidative injury, is the most prevalent neurological condition worldwide after Alzheimer's disease (Bloem et al., 2021). In 2005, the number of PD patients over 50 years of age in the 10 most populous countries was 4.6 million, which is expected to increase to 9.3 million by 2030 (Dorsey et al., 2007). Current medications are unable to halt degenerative progression in PD patients, emphasizing the necessity for the identification and development of novel, efficacious, and safe therapeutic compounds. Neuropathologically, PD is characterized by the formation of intracytoplasmic Lewy bodies and the selective and progressive loss of dopaminergic (DA) neurons in the substantia nigra pars compacta (SNpc) (Bloem et al., 2021). While the precise etiology of PD remains unclear, imbalanced redox homeostasis derived from mitochondrial dysfunction and/or dopamine metabolism underpins both sporadic and

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 29 May 2023; Accepted: 11 September 2023; Online: 12 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (31772476 and 31911530077 to X.X., 81870991 and U1603281 to S.Q.), Guangdong Basic and Applied Basic Research Foundation (2023A1515010914 to X.X.), and Natural Science Foundation of Guangdong Province (2022A1515010352 to S.Q.)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: sgq9528@smu.edu.cn; Xu2003@smu.edu.cn

idiopathic manifestations (Bloem et al., 2021). Furthermore, various environmental toxins, such as 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP), 6-hydroxydopamine, and rotenone, can compromise mitochondrial functionality, induce DA neuron damage, and produce PD-like motor neurological symptoms (Dauer & Przedborski, 2003). For example, following systemic exposure, highly lipophilic MPTP can rapidly cross the blood-brain barrier (BBB), convert into 1-methyl-4-phenylpyridinium ions (MPP^+) in astrocytes, and suppress mitochondrial respiratory complex I (first enzyme of the respiratory chain) within the mitochondria (Zeng et al., 2020b). This can lead to diminished adenosine triphosphate (ATP) generation and mitochondrial membrane potential ($\Delta\psi m$), as well as increased ROS formation, especially superoxide (O_2^-), which can induce mitochondrial malfunction and fragmentation, oxidative stress, and even death (Bloem et al., 2021; Dauer & Przedborski, 2003). Many cell death-related pathways, such as apoptosis, cell necrosis, and ferroptosis, are associated with oxidative stress (Callizot et al., 2019; Li et al., 2023; Qu et al., 2022). Among them, apoptosis is considered the predominant programmed cell death mechanism associated with PD (Tatton et al., 2003). Similarly, dopamine metabolism is implicated in the increased production of ROS in PD patients, thereby enhancing the susceptibility of DA neurons to oxidative injury (Vidoni et al., 2018). Therefore, antioxidants that are effective at scavenging excessive free radicals or limiting their production have received increasing attention in PD treatment due to the significant role of ROS in this disease.

Amphibians inhabit harsh environments characterized by a multitude of potential threats, including pathogenic microbes, parasites, and physical and chemical stressors. Therefore, their skin has evolved numerous glands that secrete diverse defensive molecules, including peptides, a noted resource for novel drug development (Mwangi et al., 2019; Qi et al., 2019; Xu & Lai, 2015) or as important channels (Zhang et al., 2021). Cathelicidins, unique to vertebrates, are multifaceted defense peptides essential for immune regulation, inflammatory response, wound healing, and blood vessel formation. Since the discovery of cathelicidin-AL in *Amolops loloensis* in 2012 (Hao et al., 2012), more than 20 cathelicidins from 13 amphibian species have been reported. These peptides exhibit a range of biological functions, including antimicrobial, antioxidant, lipopolysaccharide (LPS) neutralization, chemotaxis, antitumor, and wound healing activities (Cao et al., 2018; Chai et al., 2021; Wei et al., 2013). While certain cathelicidins, e.g., cathelicidin-PN (*Pelophylax nigromaculata*) (Wang et al., 2021), cathelicidin-NV (*Nanorana ventripunctata*) (Feng et al., 2022), cathelicidin-OA1 (*Odorrana andersonii*) (Cao et al., 2018), and cathelicidin-HR (*Hoplobatrachus rugulosus*) (Wiriyaampaiwong et al., 2022), are known to exhibit *in vitro* free radical scavenging and antioxidant activity (Xu & Lai, 2015), their therapeutic effects and mechanisms regarding neurodegenerative diseases, especially PD, remain unexplored.

The Asiatic painted frog (*Kaloula pulchra*: Microhylidae), characterized by a triangular-shaped body (55–77 mm long in males and 56–76 mm long in females), is predominantly found in low-altitude areas of South China and Southeast Asia (Fei et al., 2012). Upon disturbance or capture, these frogs inflate their bodies to a near-spherical shape and secrete a protective white substance from their skin (Zhang et al., 2010). Microhylid species, including *K. pulchra*, have been used in traditional Chinese medicine for several centuries to alleviate symptoms described as "dispelling wind", a condition associated with PD (Fauna CGOCMA, 1983). While previous research has identified an antimicrobial peptide (brevinin-2KP) and two trypsin inhibitors from the skin of microhylids (Conlon & Kim, 2000; Gao et al., 2022; Zhang et al., 2010), no studies have explored the presence of cathelicidins or antioxidants. In this research, a novel cathelicidin (Cath-KP) with potent antioxidant effects was identified from *K. pulchra* frog skin secretions. Based on circular dichroism (CD) and homology modeling, the secondary structure of the identified peptide was determined. *In vitro* antioxidant properties were assessed using 2,2'-azinobis-(3-ethylbenzthiazoline-6-sulphonate) (ABTS) and 1,1-diphenyl-2-picrylhydrazyl (DPPH) radical scavenging assays and ferric-reducing antioxidant power (FRAP) analysis. The possible role of Cath-KP in neurological diseases was explored by examining its underlying mechanisms and effects on viability, delivery, mitochondrial $\Delta\psi m$, and ROS accumulation in MPP^+-induced MN9D cells. The neuroprotective effects of Cath-KP in MPTP-induced PD mice were also evaluated. To the best of our knowledge, this is the first cathelicidin demonstrated to show potent antioxidant and neuroprotective properties in PD mice by targeting oxidative stress. These findings may help guide the development of innovative PD therapeutic agents and expand our understanding of the biological functions of cathelicidins.

MATERIALS AND METHODS

Reagents and cells

Anti-TH (F-11, sc-25269), SOD-1 (24, sc-101523), SOD-2 (E-10, sc-137254), and HO-1 (F-4, sc-390991) antibodies were obtained from Santa Cruz Biotechnology (USA). Nrf2 (ab137550), Keap1 (ab227828), Bcl-2 (ab182858), and anti-histone H3 (ab1791) antibodies were obtained from Abcam (UK). Sirt1 (#9475), phospho-FAK (#3283), FAK (#13009), phospho-p38 (#4511), and p38 (#14451) antibodies were purchased from Cell Signaling Technology (USA). Bax (50599-2-Ig) was purchased from Proteintech Group (USA). Human integrin α1 (FAB56761G) and β1 (FAB17781G) 488-conjugated antibodies were purchased from R&D Systems (USA). The GAPDH antibody (AB0037) was purchased from Abways Technology (China). Goat anti-mouse IgG (H+L) (DyLight 488 conjugate) (BA1126), goat anti-mouse IgG (H+L) (DyLight 550 conjugate) (BA1133), and goat anti-rabbit IgG (H+L) (DyLight 550 conjugate) (BA1135) secondary antibodies were purchased from Boster Biological Technology (USA). Tubulin antibody (AT819), actin antibody (AA128), horseradish peroxidase (HRP)-labeled goat anti-rabbit IgG (H+L) secondary antibody, and HRP-labeled goat anti-mouse IgG (H+L) secondary antibody were purchased from Beyotime Biotechnology (China). MPTP (M0896) and MPP^+ (D048) were obtained from Sigma-Aldrich (USA). Cath-KP and fluorescein isothiocyanate (FITC)-Cath-KP (>95% purity) were synthesized, purified, and confirmed by GL Biochemistry (China). The mouse DA neuronal cell line (MN9D) was obtained from the American Type Culture Collection (ATCC, USA) and maintained in Dulbecco's Modified Eagle Medium (DMEM, Gibco, USA) with 10% fetal bovine serum (FBS, Gibco, USA) and 1% penicillin-streptomycin-amphotericin B solution (P7630, Solarbio, China) in an incubator at 37 °C and 5% CO_2. Human umbilical vein endothelial cells (HUVECs) obtained from ATCC were maintained in DMEM-F12 (Gibco,

USA) with 10% FBS, growth factors, and 1% penicillin-streptomycin-amphotericin B solution in an incubator at 37 °C and 5% CO_2. The Sirtuin-1 (Sirt1)-specific inhibitor EX527 (SI541, Selleck, USA) was dissolved in dimethyl sulfoxide (DMSO).

Animals and ethical statement

Adult *K. pulchra* specimens (weighing 40–60 g) were collected from Baiyun District, Guangzhou City, Guangdong Province, China, and humanely sacrificed. Their skin tissue was immediately removed and placed in liquid nitrogen until use. Male C57BL/6 mice (8–10 weeks old, 25–30 g) were obtained from BesTest Bio-Tech Co., Ltd. (Zhuhai, China) and housed at the experimental animal center at Southern Medical University under a 12:12 h light/dark cycle at 25±2 °C and 55%±10% humidity, with free access to food and water. Prior to the studies, the animals were given at least one week to acclimatize to their new environment. All animal procedures were authorized by the Animal Care and Use Ethics Committee of Nanfang Hospital (NFYY-2021-1074).

Molecular cloning and sequence analyses

Molecular cloning of Cath-KP-encoding cDNA followed our previous studies, with minor modification (Chai et al., 2021; Zeng et al., 2018). In brief, after obtaining the 5' fragments of cDNA encoding the Cath-KP precursor through polymerase chain reaction (PCR) amplification using the primer encoding the cathelin domain and the 5' PCR primer from the SMART cDNA Library Construction Kit, a sense primer (5'-AAATG CAGAGCTGCTGGTGGGGGGC-3') was synthesized and coupled to the 3' PCR primer CDS III from the same kit to screen the full cDNA encoding Cath-KP. SignalP v.6.0 (http://www.cbs.dtu.dk/services/SignalP) was used to predict the signal peptide. The Bioinformatics Resources Portal (http://www.expasy.org/tools/) was used to determine the physicochemical characteristics of Cath-KP. The Basic Local Alignment Search Tool (BLAST) was used to identify cathelicidin sequences with high homology to Cath-KP, which were then aligned using ClustalW (http://embnet.vital-it.ch/software/ClustalW.html). PEP-FOLD v.3.5 (http://bioserv.rpbs.univ-paris-diderot.fr/services/PEP-FOLD3) was used to predict the structure of Cath-KP.

CD analysis

The secondary structure and stability of Cath-KP under different solvent conditions were determined using a Chirascan-plus CD spectrophotometer (Applied Photophysics, UK). Spectra (180–260 nm) were determined in a cuvette with a 0.1 cm path length, bandwidth of 1 nm, response time of 1 s, scan rate of 100 nm/min, and temperature of 25 °C. Cath-KP was dissolved in 0, 30, 60, and 90 mmol/L sodium dodecyl sulfate (SDS) solution or 60 mmol/L SDS solution with 0, 100, 200, and 400 mmol/L NaCl at a final concentration of 50 μmol/L. To further determine the effect of temperature on secondary structure stability, Cath-KP (final concentration of 50 μmol/L) was dissolved in 60 mmol/L SDS and incubated at 25, 37, 50, 70, and 90 °C for 1 h before measuring the CD spectra. The CD data (deg·cm^2/dmol) were reported as the mean residual ellipticity (θ) of two successive scans.

Measurement of antioxidant activities *in vitro*

ABTS (S0119, Beyotime Biotechnology, China) and DPPH (Sigma-Aldrich, USA) radical scavenging activity assays were performed as reported previously, with slight modifications (Zeng et al., 2020a). In brief, 10 μL of phosphate-buffered saline (PBS) or Cath-KP (final concentrations ranging from 2.5–40 μmol/L) was added to a 96-well plate containing 200 μL of $ABTS^+$ or DPPH working solution and left at room temperature for 16 min and 9 min, respectively. Absorbance was detected at 734 nm and 517 nm using a microplate reader (Infinite M1000 Pro, Tecan, Switzerland). Total reducing power of Cath-KP was evaluated using a FRAP Kit (S0116, Beyotime Biotechnology, China) according to the methods described in our prior research, with minor modification (Zeng et al., 2020a). Briefly, 5 μL of PBS or Cath-KP (final concentrations ranging from 2.5–40 μmol/L) was added to a 96-well plate with 180 μL of FRAP working solution and left at room temperature for 5 min. Absorbance was detected at 593 nm using a microplate reader. Total antioxidant capacity was determined using a standard curve and expressed as the equivalent concentration of $FeSO_4$ solution. All assays were performed in triplicate.

PD mouse model induction

A MPTP-induced PD mouse model was established to examine the potential neuroprotective effects of Cath-KP based on previous reports (Zhang et al., 2017). In brief, mice were administered an intraperitoneal injection of MPTP at a dose of 30 mg/kg at around 0800h for 5 consecutive days to establish a subacute PD mouse model. Control mice were administered an intraperitoneal injection of sterilized 0.9% saline. Simultaneously, mice were continuously administered an intraperitoneal injection of 1 mg/kg of Cath-KP and 0.9% saline at approximately 1600h for 6 consecutive days. On day 9 after the first dose, behavioral tests, including open field, grasping, pole, and rotarod tests, were carried out to assess mobility and anxiety, upper limb locomotion, anti-fatigue ability, and motor coordination of mice. On day 12 after the first dose, mice were anesthetized with 1% sodium pentobarbital (50 mg/kg; Ceva Santé Animale, France). Blood was removed with sterilized 0.9% saline and their brain tissues were quickly removed for immunohistochemical, immunofluorescence, and western blot analyses.

Behavioral study

All tests were performed between 0900h and 1400h and were assessed by the same rater in an observation room under low-intensity light. The mice were acclimated for at least 1 h prior to the start of the tests. To prevent olfactory cues, all equipment was cleaned with 10% ethanol between each animal test.

Open field test: An open field test was conducted to assess both movement and anxiety-related behaviors in mice. Each mouse was individually placed at the center of a box (50 cm×50 cm×40 cm) and videotaped for 10 min using EthoVision XT software (China). Parameters such as total distance traveled, total duration of activity, and distance traveled in the center were noted.

Grasping test: A grasping test was used to assess the motor capacity of the forelimbs of mice. In brief, mice were suspended by their two front paws on a horizontal cable (diameter 1 mm, 30 cm from the ground). The duration for which the mice remained suspended was recorded.

Rotarod test: An accelerated rotarod test was used to determine endurance capacity of mice. In brief, mice were trained twice daily for three days on a TSE Rotarod (TSE Systems, Bad Homburg, Germany) at a speed of 5 r/min. In the formal experiments, the time to fall was recorded as the speed of the spinning rod gradually accelerated from 0 to 30

r/min over 5 min.

Pole-climbing test: A pole-climbing test was implemented to assess motor coordination of the mice. In brief, mice were positioned at the top of a vertical pole (55-cm length, 1-cm diameter), facing downwards. Prior to the formal experiment, mice received three days of training. The total time required to descend to the base was recorded.

Cell viability assay

MN9D cells (3×10^3 cells/well) were grown in 96-well plates (Corning, Germany) for 24 h at 37 °C, then incubated with Cath-KP (0–5 μmol/L), MPP^+ (0–1 mmol/L), and MPP^+ (1 mmol/L) plus Cath-KP (0–5 μmol/L) for another 24 h at 37 °C. After that, cell viability was determined using CCK-8 reagent (C0037, Beyotime Biotechnology, China) following the manufacturer's manual. A multifunctional microwell plate tester (Infinite 200 Pro, Tecan Group AG, Switzerland) was used to detect absorbance at 450 nm. All experiments were repeated at least three times.

Intracellular ROS detection of live cells

The conversion of non-fluorescent 2,7-dichlorodihydrofluorescein (DCFH) to fluorescein by intracellular ROS can serve as the basis for measuring intracellular ROS, with increased levels reflected by fluorescence intensity. Here, MN9D cells seeded in black 96-well plates (3×10^3 cells/well) were washed with PBS three times and stained with 20 μmol/L DCFH-DA (S0033M, Beyotime Biotechnology, China) for 2 h at 37 °C after the incubation experiments (Chai et al., 2022). Intracellular fluorescence intensities were measured using an IX73 Inverted Microscope (Olympus, Japan) and FACScan flow cytometry (Becton Dickinson Company, USA).

Mitochondrial ROS detection of live cells

MN9D cells (3×10^3 cells/well) were treated with 1 mmol/L MPP^+ and/or 5 μmol/L Cath-KP for an additional 24 h after being grown in 96-well plates for 24 h at 37 °C. MitoSOX (M36008, Thermo Fisher Scientific, USA) and Hoechst 33258 staining (C1017, Beyotime Biotechnology, China) solutions were added. After incubation for 15 min at 37 °C, the cells were washed three times, and images were scanned using an IX73 Inverted Microscope (Olympus, Japan).

Mitochondrial *Δψm* detection of live cells

MN9D cells (3×10^3 cells/well) were treated with 1 mmol/L MPP^+ and/or 5 μmol/L Cath-KP for an additional 24 h after being grown in 96-well plates for 24 h at 37 °C. The cells were then stained with JC-1 (C2003S, Beyotime Biotechnology, China) after washing three times with PBS (pH 7.4). The *Δψm* of labeled cells was determined using flow cytometry or fluorescence microscopy, with green fluorescence signals (J-Monomers) indicating low *Δψm* and red fluorescence signals (J-Aggregates) indicating high *Δψm*.

Intracellular superoxide dismutase (SOD), catalase (CAT), and nitric oxide (NO) detection assays

MN9D cells (3×10^5 cells/well) were treated with 1 mmol/L MPP^+ and/or 5 μmol/L Cath-KP for an additional 24 h after being grown in 6-well plates for 24 h at 37 °C. The cells were then collected and split with RIPA lysis buffer (P0013, Beyotime Biotechnology, China). According to the manufacturer's instructions, levels of SOD, CAT, and NO were detected using assay kits (Cu/Zn-SOD and Mn-SOD Assay Kit with WST-8 (S0103), Catalase Assay Kit (S0051), and Nitrate/Nitrite Assay Kit (S0023), respectively) from Beyotime Biotechnology (China).

Immunohistochemistry and immunofluorescence assays

A freezing microtome (CM1950, Leica, Germany) was used to slice embedded mouse brains into 7 μm slices, which were then incubated with appropriate primary antibodies overnight at 4 °C. After that, brain slices were treated with fluorescent-labeled secondary antibodies and 4',6-diamidino-2-phenylindole (DAPI) for the immunofluorescence test. Images were scanned using a confocal laser scanning microscope (LSM 980, Zeiss, Germany). Quantitative analyses were carried out using ImageJ software. For immunohistochemical analysis, the brain slices were exposed to biotin-tagged secondary antibodies and stained with diaminobenzidine (DAB). Stereomicroscopy (N23977, Zeiss, Germany) was used to scan the images.

Western blot assay

MN9D cells (3×10^5 cells/well) were treated with 1 mmol/L MPP^+ and/or 5 μmol/L Cath-KP for an additional 24 h after being seeded into 6-well plates for 24 h. Total proteins were extracted from the cells and substantia nigra tissue using RIPA lysis buffer (P0013, Beyotime Biotechnology, China), then quantified using a BCA Protein Assay Kit (P0010, Beyotime Biotechnology, China) before western blotting was carried out according to our previous study (Chai et al., 2021). Primary antibodies against TH, B-cell lymphoma 2 (Bcl-2), Bcl-2 associated X (Bax), Sirt1, Nuclear factor erythroid 2-related factor 2 (Nrf2), Heme oxygenase 1 (HO-1), Kelch-like ECH-associated protein 1 (Keap1), SOD-1, SOD-2, phospho-focal adhesion kinase (FAK), FAK, phospho-p38, p38, actin, Glyceraldehyde 3-phosphate dehydrogenase (GAPDH), histone H3, tubulin, and HRP-conjugated secondary antibodies were applied for western blot analysis. Immobilon western chemiluminescent HRP substrate (Millipore, USA) was used to incubate the attached antibodies, and Tanon 5200 (China) was used to detect chemiluminescence. ImageJ was used to determine optical density of the bands. All experiments were performed in triplicate.

Peptide internalization analysis

For fluorescence intensity detection, MN9D cells (1×10^5 cells/well) were cultured overnight in 24-well plates, then subsequently treated with FITC-Cath-KP (0, 2.5, and 5 μmol/L) for 3 h at 4 °C or 37 °C and assayed using confocal laser scanning microscopy (LSM 980, Zeiss, Germany) and FACScan flow cytometry (Becton Dickinson Company, USA). To evaluate its effect of heparin sulphate on FITC-Cath-KP internalization, heparin sulphate (50 μg/mL) was preincubated with 5 μmol/L FITC-Cath-KP for 30 min, followed by co-incubation with the cells for another 3 h and flow cytometry analysis. All experiments were performed in triplicate.

Assessment of brain targeting delivery

Male C57BL/6 mice received an intraperitoneal injection of FITC-Cath-KP (1 mg/kg) and were then anesthetized at 0, 2, and 6 h post-injection with 1% sodium pentobarbital. Blood was removed with sterilized 0.9% saline and 4% paraformaldehyde via cardiac perfusion. Brains were surgically removed, embedded in paraffin, cut into 7 μm sections using a Leica CM1950 freezing microtome, and stained with tyrosine hydroxylase (TH) antibodies and DAPI. The SNpc slices were immediately scanned using an IX73 Inverted Microscope (Olympus, Japan). Control mice received an intraperitoneal injection of sterilized 0.9% saline.

Statistical analysis

Statistical analyses were performed using GraphPad Prism v.9.0 (GraphPad Software, USA). All data were evaluated using unpaired Student's *t*-test and ordinary one-way analysis of variance (ANOVA) with Bonferroni's multiple comparison test. Ten individuals from each group were used for the behavioral tests. Three to five samples from each group were used for *in vivo* immunofluorescence, immunohistochemical, and western blot analyses. Three individuals per group were employed for *in vitro* western blot and immunofluorescence assays. Values are given as mean±standard error of the mean (SEM). Significant *P*-values were determined at *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$; ****: $P<0.0001$.

RESULTS

Identification and structural analysis of Cath-KP

The cDNA sequence of Cath-KP was derived from *K. pulchra* skin using a PCR-based technique. As shown in Figure 1A, the Cath-KP precursor, encoded by 540 base pairs, consisted of 169 amino acid residues and shared high sequence similarity to antimicrobial peptides (AMPs) in the cathelicidin family based on a NCBI BLAST search and sequence alignment, showing 58.49%, 57.89%, and 57.89% homologies with cathelicidin-HR (*H. rugulosus*), cathelicidin-PY1 (*Paa yunnanensis*), and cathelicidin-NV (*N. ventripunctata*), respectively (Figure 1B). However, the mature peptide, named Cath-KP (GCSGRFCNLFNNRRPGRLTLIHRPGGDKRTS TGLIYV) after its classical RR protease cleavage site, showed no similarity with any reported AMP (Figure 1B). The theoretical isoelectric point (PI) of Cath-KP was 11.35, indicating a positive charge under physiological conditions. The relative mass of Cath-KP was 4 131.70 Da, while the relative mass of FITC-Cath-KP was 4 636.35 Da (Supplementary Figure S1).

The structure of Cath-KP was predicted using the PEP-FOLD3 method. The structure exhibited an *αββ* conformation, with an *α*-helix (Arg-5-Asn-11) in the N-terminal and a *β*-hairpin (Leu-18-Asn-Val-37) at the C-terminal (Figure 1C). Notably, an RTS sequence, typically recognized as a disintegrin motif, was located in the loop between two *β*-strands. Consistent with the predicted results, Cath-KP in aqueous solution showed a broad negative peak near 200 nm, indicative of an unfolded peptide in equilibrium with a *β*-sheet structure (Eiríksdóttir et al., 2010). Upon dissolution in SDS solution, the negative minimum of Cath-KP shifted from 200 nm to 206 nm, indicative of enhanced secondary structure formation. Results from CD spectral analysis showed that Cath-KP contained 5.40% *α*-helices, 38.30% *β*-sheets, 21.90% turns, and 38.30% random coils in aqueous solution, with the *α*-helix ratio increasing gradually with SDS concentration from 5.40% to 7.80% (Figure 1D; Supplementary Table S1). As shown in Figure 1E, F and Supplementary Table S1, increasing NaCl concentrations disrupted the secondary structure components of Cath-KP, as evidenced by the increase in *α*-helix ratio and decrease in *β*-sheet ratio. In contrast, increasing temperature resulted in minimal alterations in CD spectra, indicating that Cath-KP was stable at high temperature but sensitive to high salt concentrations.

Cath-KP improved motor performance in MPTP-treated mice and reversed DA neuron loss

The prominent clinical manifestation of PD is the impairment of motor function. Prior to the formal experiments, preliminary

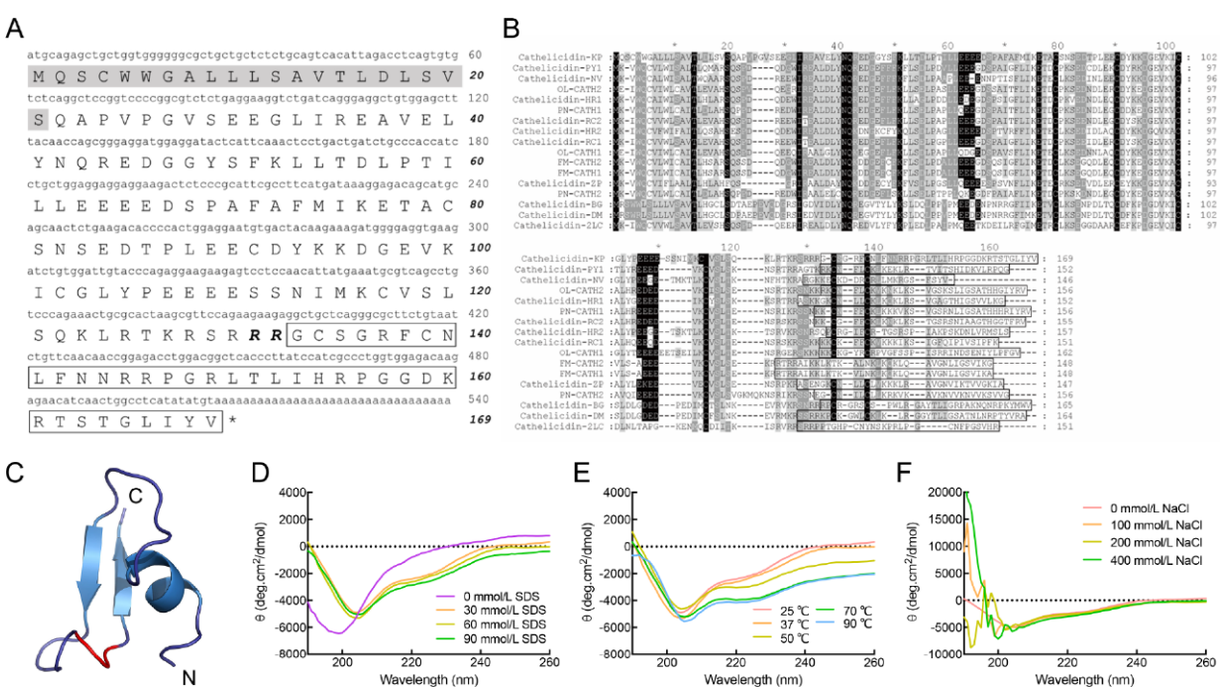


Figure 1 Sequence characterization and secondary structure determination of Cath-KP

A: cDNA encoding Cath-KP and deduced amino acid sequences. Signal peptide is marked in gray, RR residues are shown in bold and italic. Mature peptide sequence is boxed, and stop codon is denoted by an asterisk (*). Nucleotide and amino acid numbers are shown after sequence. B: Alignment of cathelicidin peptides from amphibians. To better present sequence identity, hyphens (-) have been added to some sequences to indicate residue deletion. Same and highly conserved residues are represented in black and shadows, respectively. All mature peptide domains are boxed. C: Cartoon image of Cath-KP conformation with labeled N and C termini and RTS motif in red. CD spectra of Cath-KP dissolved in: D: SDS solutions; E: temperatures; or F: NaCl solutions.

animal studies were conducted using varying doses of Cath-KP (0.5, 1, and 2 mg/kg) to determine the most appropriate dosage for use *in vivo*. As shown in Supplementary Figure S2, 1 mg/kg of Cath-KP yielded the best improvement in dyskinesia in PD mice. Furthermore, no significant hepatic or renal toxicity was observed in the mice at this dose (Supplementary Figure S3). Thus, 1 mg/kg of Cath-KP was selected as the optimal concentration for subsequent animal experiments. The open field, grasping, pole, and rotarod tests were used to assess the effects of Cath-KP on MPTP-induced motor impairment in mice (Figure 2A). As shown in Figure 2B, MPTP administration markedly decreased locomotor activity in the open field test. Furthermore, total distance of spontaneous motion, number of spontaneous activities, and central distance of spontaneous motion in MPTP-treated mice decreased by 50.00%, 23.52%, and 69.95%, respectively, compared to the control group in the open field test(Figure 2C–E), suggesting that MPTP exposure impaired motor coordination. Additionally, compared to the corresponding control groups, MPTP injection led to reductions in the average holding time and time spent on the rotarod by 76.78% and 69.03%, respectively (Figure 2F, G). The pole-climbing test was utilized to evaluate mouse coordination and balance. Results revealed that the MPTP administration extended the pole-climbing time by 153.25% relative to the control group (Figure 2H). In comparison to the MPTP-only group, Cath-KP treatment at 1 mg/kg ameliorated motor dysfunction in PD mice. Notably, total distance of spontaneous motion increased from 16.17 m to 26.24 m, number of spontaneous activities increased from 865.80 to 1 009.10, and central distance of spontaneous motion increased from 1.78 m to 4.32 m in the open field test (Figure 2C–E), average holding time and time spent increased from 6.20 s to 16.10 s and 32.80 s to 75.80 s in the grasping and rotarod tests (Figure 2F, G), while climbing time

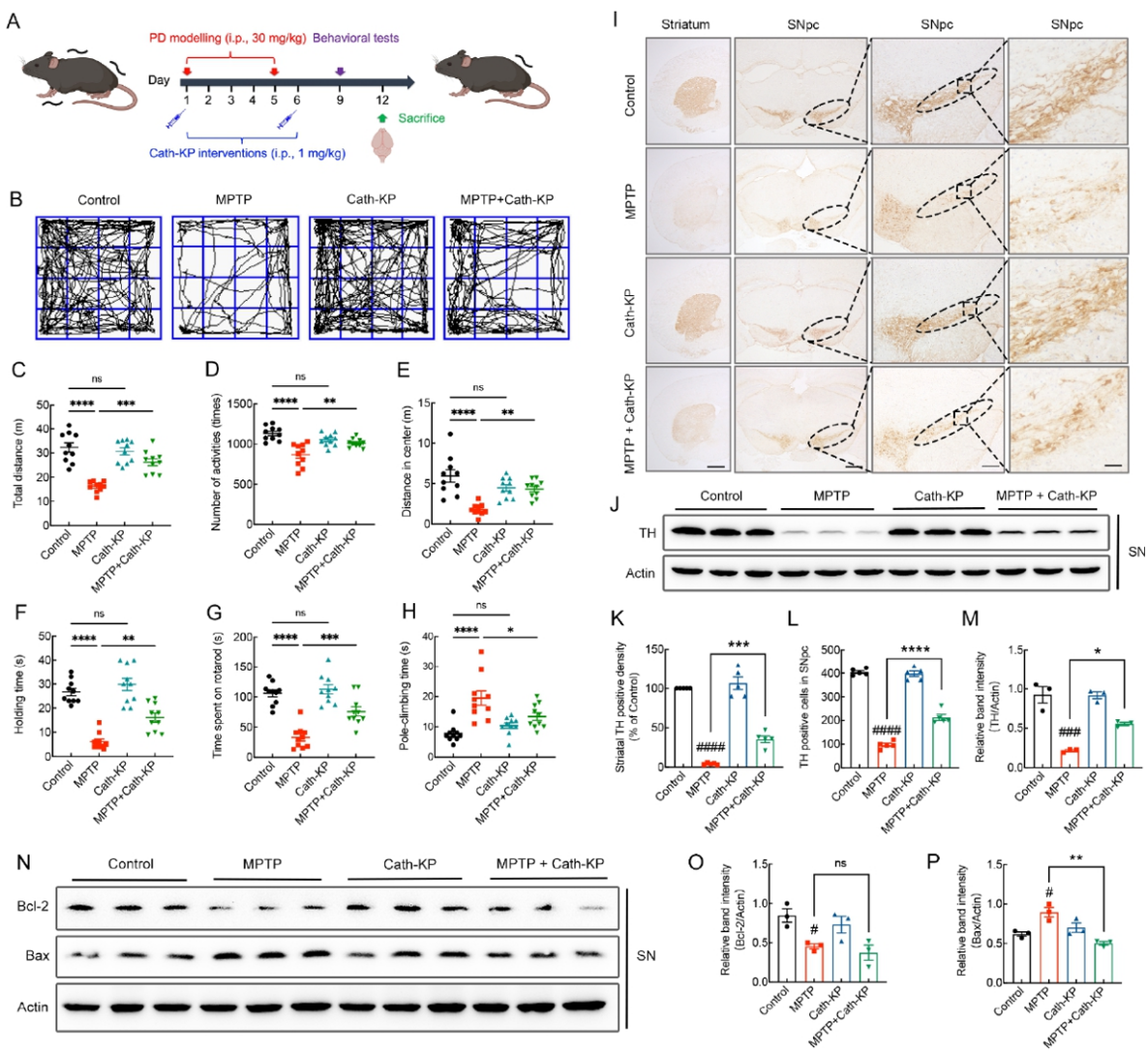


Figure 2 Cath-KP decreased motor dysfunction and dopaminergic (DA) neuron degeneration in Parkinson's disease (PD) mice

A: Experimental design involving Cath-KP treatment (intraperitoneal (i.p.) 1 mg/kg/day) in MPTP-induced PD mice (i.p. 30 mg/kg/day). B: Travelled trace in open area. C–E: Analysis of total distance, number of activities, and central distance in open field test (n=10). F–H: Analysis of grasping, rotarod, and pole-climbing tests (n=10). I, K, L: Immunohistochemical staining and analysis of TH-positive neurons in the striatum and SNpc (n=5). Borders of the SNpc are shown by ellipses and the box represents further enlarged area. Scale bar: 1 mm for images in the striatum; 800, 100, and 50 μm for images in the SNpc (from left to right). J, M–P: Representative western blot images and analysis of TH, Bax, and Bcl-2 expression levels in the SNpc (n=3). Data are represented as means±SEM with ordinary one-way ANOVA. ns: Not significant. $^{\#}$: P<0.05, $^{\#\#\#}$: P<0.001, and $^{\#\#\#\#}$: P<0.0001 vs. Control. *: P<0.05, **: P<0.01, ***: P<0.001, and ****: P<0.0001 vs. PD models.

decreased from 19.50 s to 13.40 s in the pole-climbing test (Figure 2H). Thus, these findings demonstrate the potential benefits of Cath-KP in treating motor dysfunction.

The primary pathological feature of PD is the loss of DA neurons in the SNpc and a consequent reduction in dopamine levels in the striatum. The potential protective effects of Cath-KP on DA neurons in the SNpc following MPTP administration were assessed. As shown by TH immunohistochemical analysis (Figure 2I), SNpc DA neuronal degeneration was clearly observed in the brains of the MPTP group. Post-MPTP exposure, a marked decrease in DA neurons was detected, as evidenced by immunohistochemical analysis of TH-positive cell terminals in the striatum (Figure 2K) and SNpc (Figure 2L). These findings were confirmed by western blot analysis (Figure 2J), revealing a 76.62% reduction in TH protein expression in the model group compared to the control group (Figure 2M). Notably, Cath-KP treatment reversed these changes both histologically (725.87% increase in the striatum and 120.95% increase in the SNpc) and at the protein expression level (2.57-fold increase in the SNpc).

To explore whether the protective effects of Cath-KP on DA neuron loss induced by MPTP in the SNpc were associated with its anti-apoptotic effects, the concentrations of Bax and Bcl-2 in the SNpc were determined by western blot analysis. As illustrated in Figure 2N–P, MPTP exposure significantly reduced Bcl-2 expression but increased Bax expression, and thus induced apoptosis in the SNpc. Cath-KP treatment reversed the MPTP-induced changes in Bax protein expression in the SNpc, but not that of Bcl-2 (Figure 2N–P). These findings suggest that the protective effects of Cath-KP against PD are, at least partially, mediated by inhibition of SNpc apoptosis.

Cath-KP enhanced TH expression in MPP^+-induced MN9D cells

To further explore the protective mechanism of Cath-KP against MPTP-induced DA neuron loss in mice, we assessed its effects on TH expression, a cytoplasmic marker of DA neurons (Daubner et al., 2011), in MN9D cells. As depicted in Figure 3A, B, the cultured MN9D cells exhibited TH-positive staining and showed a 76.99% reduction in fluorescence intensity after MPP^+ treatment relative to the control group. Notably, this reduction in MPP^+-treated cells was ameliorated (2.04-fold increase) following Cath-KP treatment, which did not affect the MN9D cells without MPP^+ treatment when comparing the model and control groups. Furthermore, western blot analysis demonstrated a 45.89% decrease in TH protein content in the MN9D cells exposed to MPP^+ compared to the control group; however, Cath-KP treatment significantly attenuated this reduction by 1.59-fold (Figure 3C, D). These results not only validate the utility of MN9D cells for subsequent mechanistic studies but also suggest that Cath-KP offers neuronal protection against MPP+ damage.

Decreased TH levels are closely associated with the progression of DA neuron death (Daubner et al., 2011). To further explore whether Cath-KP can mitigate the death of MN9D cells induced by MPP^+, cell viability was measured. As shown in Figure 3E, Cath-KP did not exhibit cytotoxicity to MN9D cells, even at doses of 5 μmol/L after 24 h. Furthermore, MPP^+ inhibited MN9D cell proliferation concentration dependently, with an EC_{50} value of approximately 1 mmol/L (Figure 3F). In contrast, Cath-KP attenuated the cytotoxic effects of MPP^+ on MN9D cells in a concentration-dependent manner, with cell viabilities enhanced by 2.37%, 32.33%, and 60.85% when co-incubated with 1.25, 2.5, and 5 μmol/L of Cath-KP and 1 mmol/L of MPP^+ for 24 h (Figure 3G). To explore the mechanism by which Cath-KP suppressed MPP^+-induced DA neuron death, MN9D cell apoptosis was analyzed. As illustrated in Figure 3H–J, treatment with MPP^+ markedly reduced Bcl-2 expression and increased Bax expression, thus inducing MN9D cell apoptosis. However, Cath-KP administration reversed the MPP^+-induced changes in Bcl-2 and Bax protein expression in the MN9D cells (Figure 3H–J). Together, these findings indicate that the neuroprotective effects of Cath-KP against PD are, at least partially, achieved by preventing mitochondrial apoptosis.

Cath-KP attenuated oxidative stress in MN9D cells based on antioxidant activity

MPP^+ can induce mitochondrial ROS generation, leading to the accumulation of intracellular ROS and ultimately to neurotoxicity in DA neurons (Subramaniam & Chesselet, 2013). Therefore, the effects of Cath-KP on intracellular and mitochondrial ROS levels in MN9D cells were measured. As displayed in Figure 4A, D, green fluorescence intensity, reflecting intracellular ROS levels, was significantly increased after MPP^+ treatment. This elevation was significantly suppressed by the addition of Cath-KP. These findings were confirmed by representative flow cytometry analysis of ROS generation (Supplementary Figure S4A). In parallel, Cath-KP alleviated MPP^+-induced accumulation of mitochondrial ROS in MN9D cells (Figure 4B, E).

ROS generation is closely related to mitochondrial $\Delta\psi m$. A reduction in $\Delta\psi m$ is indicative of mitochondrial malfunction, under oxidative stress, and apoptosis. As shown in Figure 4C and F, prominent red fluorescence was observed in all cells, while green fluorescence was absent in the control cells and those treated with Cath-KP only. MPP^+ treatment led to a reduction in the red/green fluorescence ratio, indicating a decline in $\Delta\psi m$. Nevertheless, co-treatment with Cath-KP (5 μmol/L) effectively reversed this ratio. Similar findings were obtained by representative flow cytometry analysis of mitochondrial $\Delta\psi m$ (Supplementary Figure S4B). In accordance with its modulation of ROS levels, Cath-KP enhanced the activities of SOD and CAT, which function as cellular antioxidants, while diminishing the activity of NO, which inhibits complex I of the mitochondrial electron transport chain to facilitate ROS generation (Figure 4G–I) (Aquilano et al., 2008; Bloem et al., 2021).

In investigating the underlying mechanism for the antioxidant effects of Cath-KP in MPP^+-induced MN9D cells, the *in vitro* antioxidant activity of Cath-KP was further assessed. As shown in Figure 4J, K, Cath-KP was scavenged ABTS and DPPH radicals in a dose-dependent manner. Specifically, 40 μmol/L Cath-KP eliminated approximately 64.92% of ABTS radicals at 16 min and 26.47% of DPPH radicals at 9 min. Consistently, as shown in Figure 4L, the capacity of Cath-KP to reduce Fe^{3+} was improved with increasing concentration. Together, these findings indicate that Cath-KP exhibits antioxidant activity, diminishes ROS accumulation and its mitochondrial damage, and suppresses MPP^+-induced neuron death via inhibition of the mitochondrial apoptotic pathway.

Cath-KP modified the Sirt1/Nrf2 and FAK/p38-mediated pathways

The effects of Cath-KP on the Sirt1/Nrf2 pathway, which is

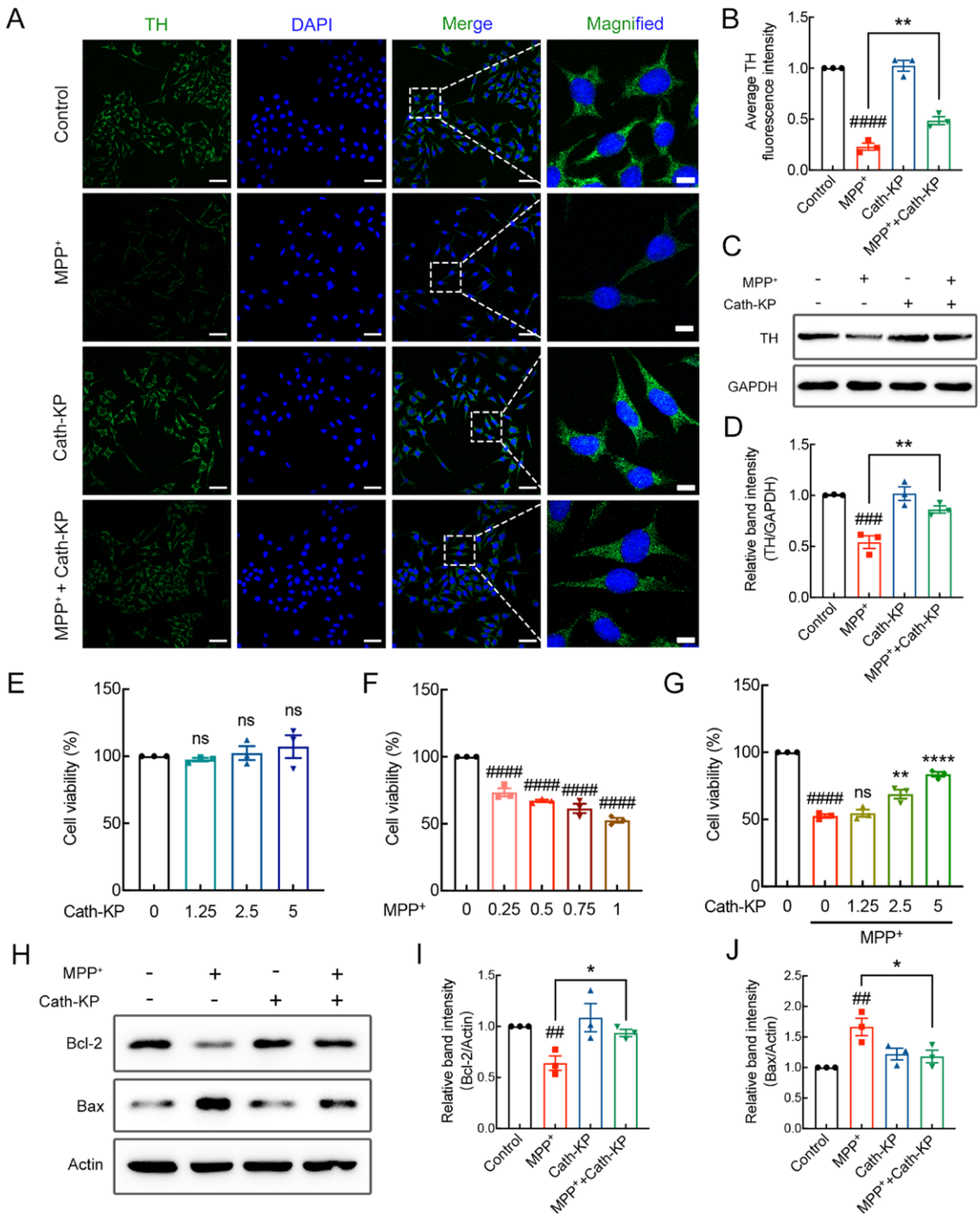


Figure 3 Cath-KP attenuated the decrease in tyrosine hydroxylase (TH) protein levels and reduced apoptosis of MPP$^+$-treated MN9D cells

A, B: Example images from confocal laser scanning microscopy (A) and quantification of TH expression (B) in MN9D cells. MN9D cells were incubated with Cath-KP (5 μmol/L) for 24 h with MPP$^+$ (1 mmol/L) present or absent before representative confocal laser scanning microscopy. TH, green fluorescence; DAPI, blue fluorescence. Scale bar: 50 μm for original and 10 μm for magnified images. C, D: Representative blots (C) and quantification analysis (D) of TH expression levels in MPP$^+$ (1 mmol/L) and Cath-KP (5 μmol/L)-treated MN9D cells. E–G: Viability of MN9D cells treated with Cath-KP (0–5 μmol/L), MPP$^+$ (0–1 mmol/L), J"and MPP$^+$ (1 mmol/L) plus Cath-KP (0–5 μmol/L) for 24 h. H–J: Representative western blot images (H) and statistical analysis of Bcl-2 (I) and Bax (J) expression levels in MN9D cells. Data are represented as means±SEM with Student' s *t*-test and ordinary one-way ANOVA (n=3). ns: Not significant. $^{\#\#}$: P<0.01, $^{\#\#\#}$: P<0.001, and $^{\#\#\#\#}$: P<0.0001 vs. Control. *: P<0.05, **: P<0.01, and ****: P<0.0001 vs. PD models.

reported to control inflammatory reactions and redox homeostasis (Fão et al., 2019), were evaluated. As shown in Figure 5A, C, the expression levels of Sirt1, Keap1, Nrf2, SOD-1, SOD-2, and HO-1 in MN9D cells were markedly down-regulated by MPP$^+$ treatment compared to the control group. In contrast, their expression levels were substantially elevated by Cath-KP treatment compared to MPP$^+$ treatment (Figure 5A, C). Notably, preincubation with the Sirt1-specific inhibitor EX527 (1 μmol/L) abolished the Cath-KP-enhanced expression levels of Sirt1, Nrf2, SOD-1, and SOD-2 in the MPP$^+$-treated MN9D cells (Supplementary Figure S5). Consistently, as illustrated in Figure 5B, D, MPTP attenuated

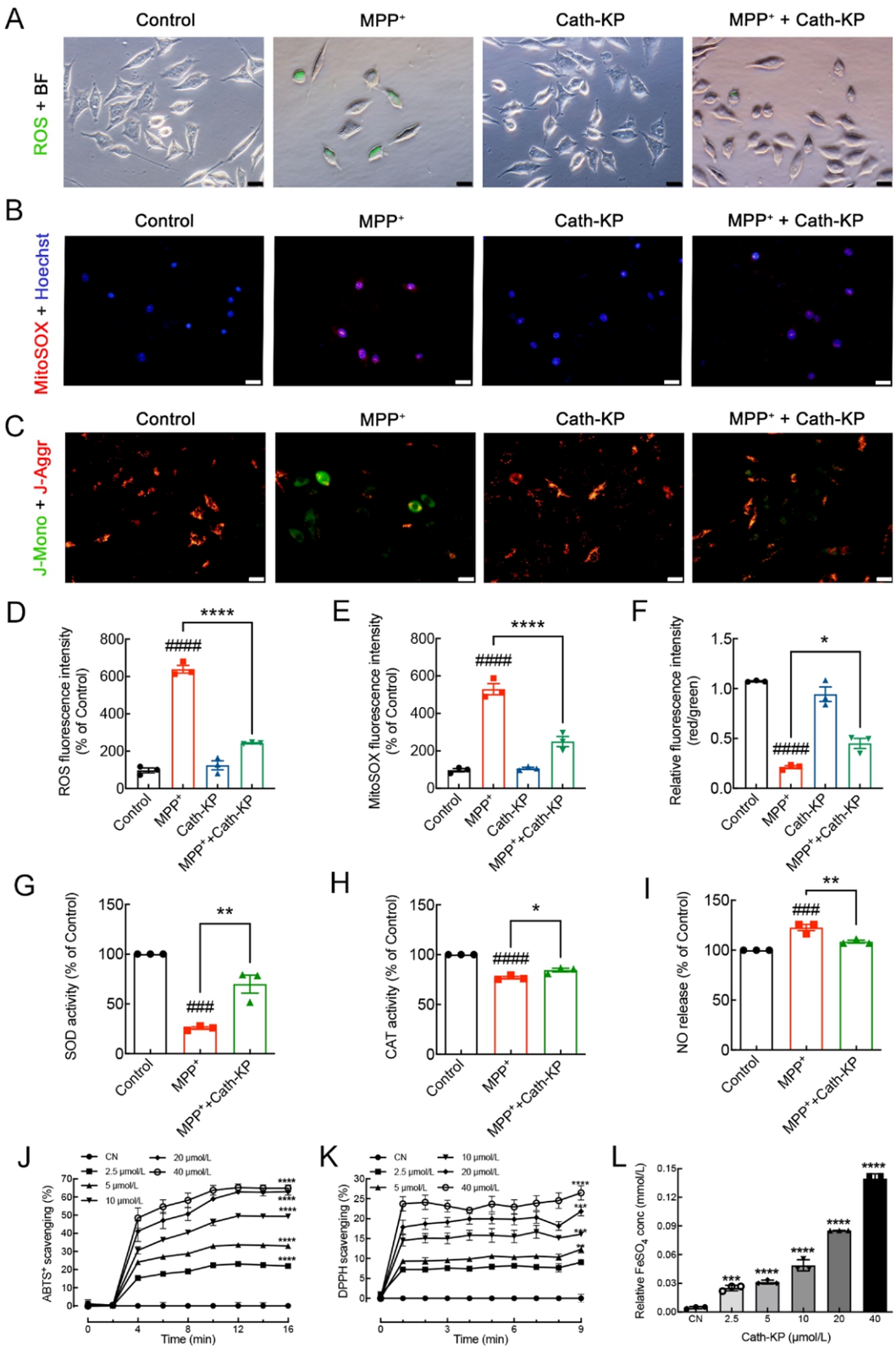


Figure 4 Cath-KP attenuated oxidative stress in MN9D cells based on its antioxidant activity

A, D: Representative fluorescence images (A) and quantitative analysis (D) of intracellular ROS generation in MPP^+ and Cath-KP-treated MN9D cells. Intracellular ROS, green fluorescence; BF: Bright Field. Scale bar: 20 μm. B, E: Representative fluorescence images (B) and quantitative analysis (E) of mitochondrial ROS generation in MPP^+ and Cath-KP-treated MN9D cells. Mitochondrial ROS, red fluorescence; Hoechst, blue fluorescence. Scale bar: 20 μm. C, F: Representative fluorescence images (C) and quantitative analysis (F) of mitochondrial $\Delta\psi m$ in MPP^+ and Cath-KP-treated MN9D cells. J-Mono: J-Monomers, green fluorescence; J-Aggr: J-Aggregates, red fluorescence. Scale bar: 20 μm. G–I: Contents of SOD, CAT, and NO in MPP^+ and Cath-KP-treated MN9D cells. J, K: ABTS and DPPH radical scavenging activity. L: FRAP antioxidant activity. Data are represented as means±SEM (n=3) with ordinary one-way ANOVA. ###: P<0.001 and ####: P<0.0001 vs. Control. *: P<0.05, **: P<0.01, ***: P<0.001, and ****: P<0.0001 vs. PD models.

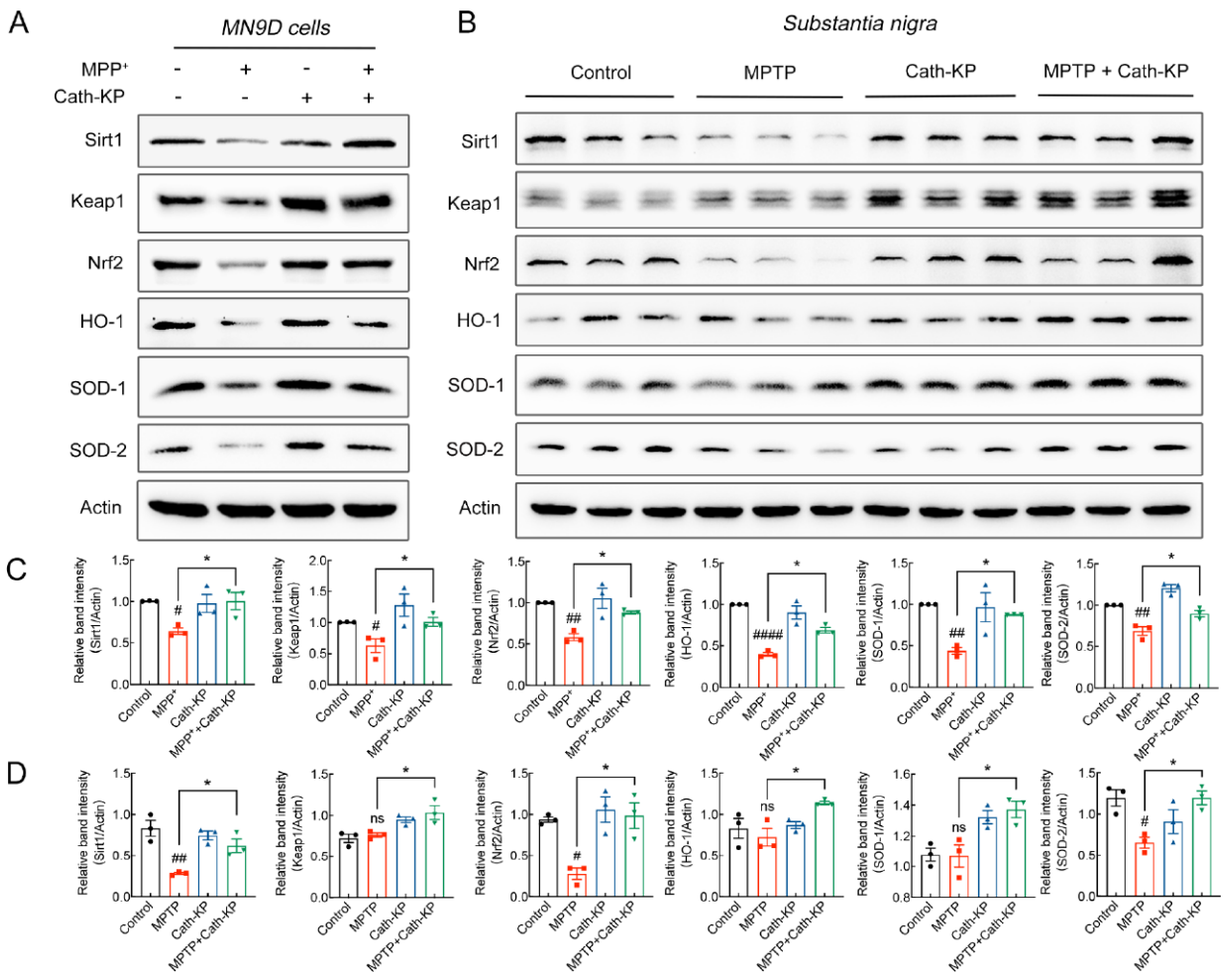


Figure 5 Modification of the Sirt1/Nrf2-mediated pathway by Cath-KP

A, C: Representative images (A) and quantification analysis (C) of western blots, detecting protein expression levels of Sirt1, Keap1, HO-1, Nrf2, SOD-1, and SOD-2 in MN9D cells incubated with MPP^+ and Cath-KP. B, D: Representative images (B) and quantification analysis (D) of western blots, detecting protein expression levels of Sirt1, Keap1, HO-1, Nrf2, SOD-1, and SOD-2 in mice treated with MPTP and Cath-KP. Data are represented as means±SEM (n=3) with ordinary one-way ANOVA. ns: Not significant. $^{\#}$: P<0.05 and $^{\#\#}$: P<0.01 vs. Control. *: P<0.05 vs. PD models.

the expression levels of Sirt1, Nrf2, SOD-2, and HO-1 in the SNpc of mice. These effects were reversed by Cath-KP co-treatment. Interestingly, the expression levels of Keap1 and SOD-1 were not changed under MPTP treatment but were significantly elevated in the peptide-treated group.

Research has indicated that ROS regulates the activation of p38 and FAK (Fão et al., 2019; Qin & Rodrigues, 2012). Therefore, we examined the potential impact of Cath-KP on the expression and phosphorylation of p38 and FAK in MPP^+-treated MN9D cells and in the SNpc of mice. As shown in Figure 6A, C, D, compared to the control, MPP^+ treatment markedly increased the phosphorylation of p38 but had no effect on its total protein expression. In contrast, MPP^+ treatment markedly decreased the phosphorylation of FAK, although also had no effect on its total protein level. However, the effects of MPP^+ on the phosphorylation of p38 and FAK were significantly mitigated by Cath-KP, aligning with observations in the SNpc of mice (Figure 6B, E, F).

Cath-KP promoted nuclear translocation of Nrf2

Typically, upon exposure to oxidative stress, Nrf2 undergoes translocation to the nucleus and interacts with antioxidant response elements (AREs), thus initiating the transcription of multiple antioxidant genes. Analysis of changes in red immunofluorescence intensity co-localized with blue DAPI revealed that MPP^+ treatment decreased nuclear Nrf2 levels and increased cytoplasmic Nrf2 levels (Figure 7A, C). Notably, these changes were reversed by Cath-KP treatment. Western blot analysis confirmed that MPP^+ treatment significantly decreased Nrf2 content in the nucleus while increasing its content in the cytoplasm, indicating suppression of nuclear translocation (Figure 7B, D). In contrast, Cath-KP enhanced the nuclear translocation of Nrf2 in the MPP^+-treated MN9D cells.

Cath-KP was cell-permeable and delivered to deep brain tissues

To ascertain the ability of Cath-KP to penetrate MN9D cells and exert antioxidant protection, the fluorescence emitted from FITC-Cath-KP incubated with MN9D cells was measured. Confocal microscopy (Figure 8A), with TH antibody fluorescence indicating cytoplasmic localization and DAPI fluorescence representing nuclear localization, revealed that FITC-Cath-KP effectively traversed the cytoplasmic membrane and entered the MN9D cells in a concentration-dependent manner relative to the control after 3 h of incubation (Figure 8A). Further examination was conducted to determine the influence of temperature on the internalization

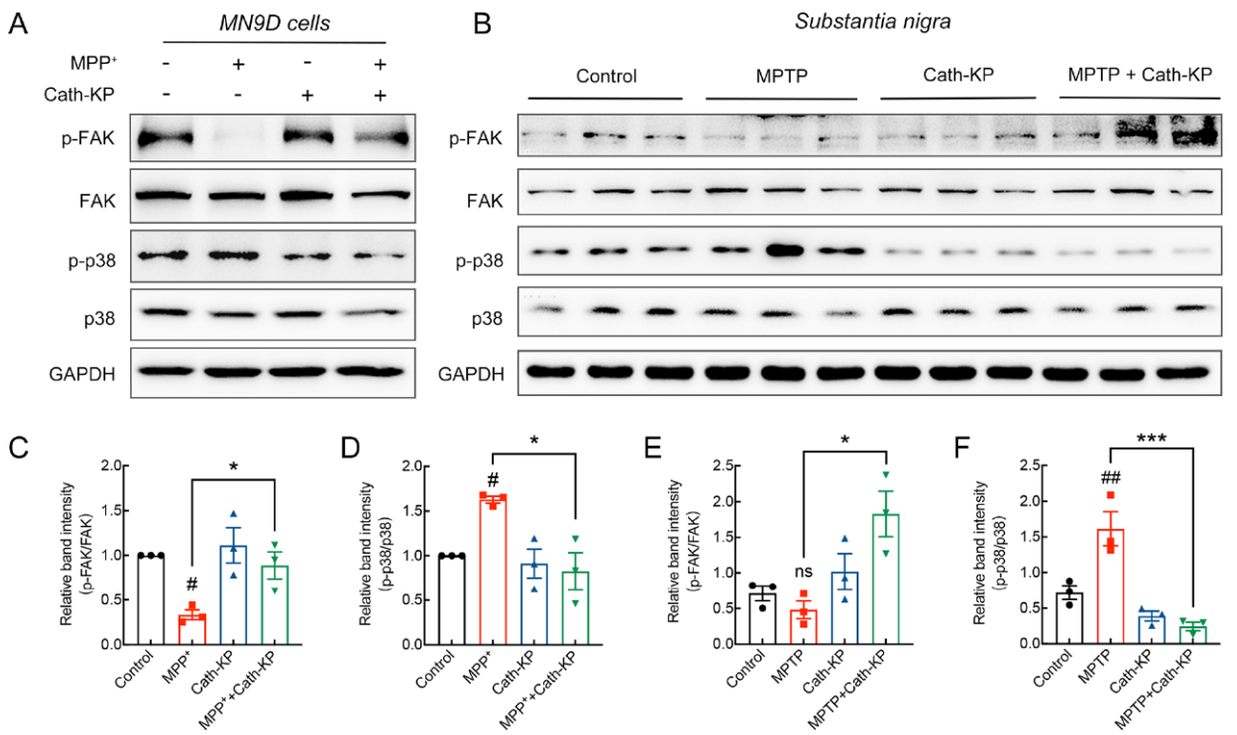


Figure 6 Modification of the FAK/p38-mediated pathway by Cath-KP

A, C, D: Representative images (A) and quantification analysis (C, D) of western blots, detecting protein expression levels of FAK and p38 in MN9D cells treated with MPP^+ and Cath-KP. B, E, F: Representative images (B) and quantification analysis (E, F) of western blots, detecting protein expression levels of FAK and p38 in mice treated with MPTP and Cath-KP. Data are represented as means±SEM (n=3) with ordinary one-way ANOVA. ns: Not significant. $^{\#}$: P<0.05 and $^{\#\#}$: P<0.01 vs. Control. *: P<0.05 and ***: P<0.001 vs. PD models.

of FITC-Cath-KP. As shown in Figure 8B and Supplementary Figure S6A, B, FITC-Cath-KP entered the MN9D cells in a concentration-dependent manner. At 4 °C, cellular uptake was reduced by approximately 30% (2.5 μmol/L) and 39% (5 μmol/L) compared to uptake at 37 °C after 3 h of incubation, suggesting an energy-dependent and thermosensitive transport mechanism for Cath-KP across the MN9D cell membrane. Cell-penetrating peptides often first interact with heparin sulphate, a crucial component of cell membranes (Kim et al., 2015). Thus, we further examined whether heparin sulphate can affect FITC-Cath-KP absorption by cells. Compared to the FITC-Cath-KP only group, the internalization of FITC-Cath-KP was decreased by 39.38% after 30 min with 50 μg/mL heparin pretreatment (Figure 8B; Supplementary Figure S6C). Collectively, these findings suggest that Cath-KP may utilize endocytosis as its primary mechanism of entry into MN9D cells.

To evaluate the potential of Cath-KP to be delivered to brain tissue, brain sections were examined at 2 and 6 h following injections of diluent, FITC only, and FITC-Cath-KP. As shown in Figure 8C, FITC-Cath-KP was distributed into brain tissue and partially co-located with DA neurons, as evidenced by the green and red fluorescence from FITC-Cath-KP and TH antibody, respectively. Moreover, the fluorescence intensity of the FITC-Cath-KP samples at 2 h post-injection was higher than that at 6 h post-injection, indicating that Cath-KP can be non-selectively delivered to deep brain tissues within 2–6 h, with the peak at 2 h.

DISCUSSION

Amphibian skin plays a pivotal role in host defense and contains multiple bioactive peptides with antimicrobial, anti-inflammatory, antitumor, and antioxidant properties (Xu & Lai, 2015). Among frogs belonging to the Microhylidae family, only a limited number of peptides have been identified, namely Brevinin-2KP and two trypsin inhibitors (Conlon & Kim, 2000; Gao et al., 2022; Zhang et al., 2010). In this study, we identified a novel peptide (Cath-KP) with antioxidant activity from the skin of *K. pulchra* frogs. Typically, frog skin bioactive peptide precursors contain an evolutionarily conserved N-terminal signal peptide, an acidic spacer region, and a C-terminal bioactive peptide, which undergo proteolytic cleavage at the KR or RR sites (Xu & Lai, 2015). The precursor of Cath-KP also exhibited this structure. Thus, the sequence following the cathelin domain, ending at the basic RR residues (GCSGRFCNLFNNRRPGRLTLIHRPGGDKRTSTGLIYV), was considered the mature peptide and a new member of the cathelicidin family. Most antimicrobial cathelicidins, including Cath-MH, cathelicidin-PY, and LL-37, contain high amphipathic α-helices in membrane-mimetic environments (Chai et al., 2021; Sancho-Vaello et al., 2017; Wei et al., 2013). In contrast, secondary structure modeling, helix-wheel analysis, and CD spectra results demonstrated that Cath-KP lacked this feature, despite its high net charge of 6 at neutral pH (Figure 1C; Supplementary Figure S7), causing the loss of antimicrobial activity (data not shown). Critical residues such as leucine, proline, phenylalanine, tyrosine, methionine, tryptophan, and free cysteine, capable of donating protons to radicals with low electron density, underpin the functionality of amphibian antioxidant peptides (Chai et al., 2022; Guo et al., 2014; Krishnan et al., 2008; Yang et al., 2009; Zeng et al., 2020a). Here, the primary sequence of Cath-KP consisted of eight hydrophobic amino acids, including four leucine residues, two proline residues, one phenylalanine residue, and one tyrosine residue (Figure 1A). Thus, we hypothesized that

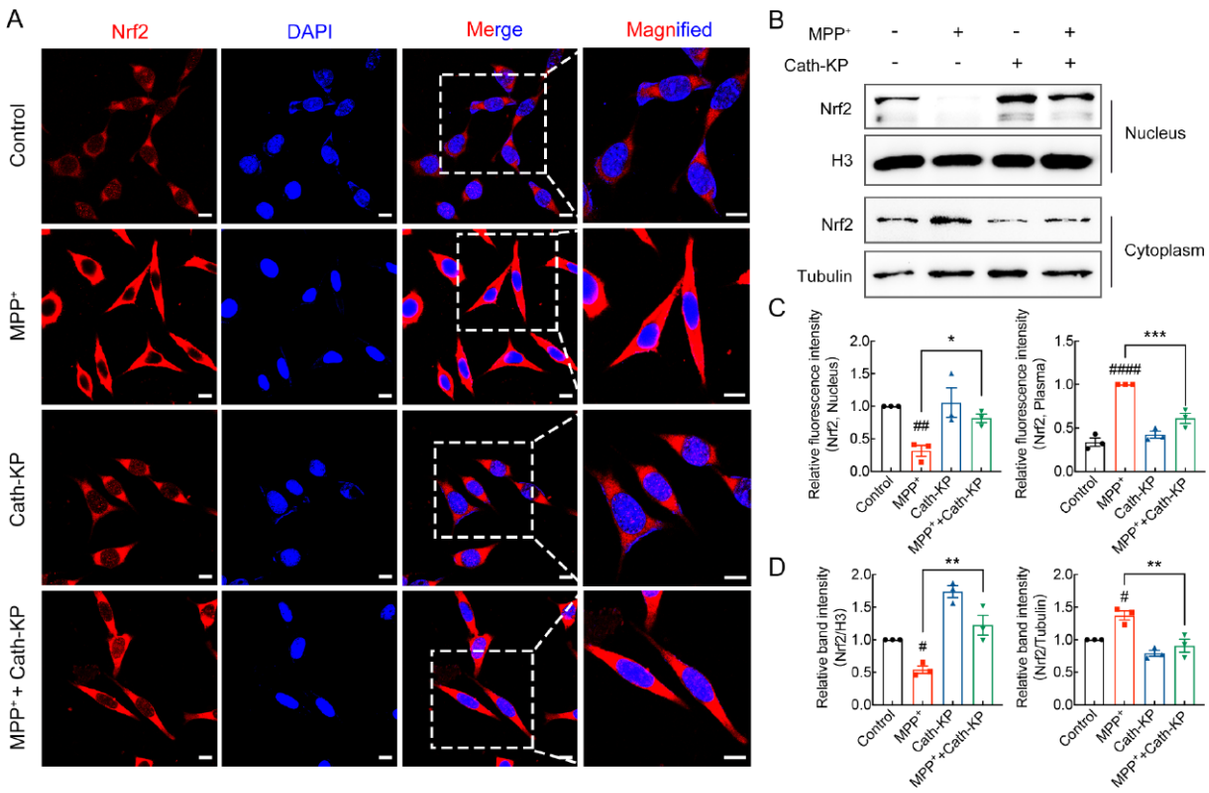


Figure 7 Nuclear translocation of Nrf2 by Cath-KP

A–D: Immunofluorescent staining (A), western blot analysis (B), and quantification (C, D) showing nuclear translocation of Nrf2 in MN9D cells. Nrf2, red fluorescence; DAPI, blue fluorescence. Scale bar: 10 μm for original images and 5 μm for magnified images in (A). Data are represented as means±SEM (n=3) with ordinary one-way ANOVA. $^{\#}$: P<0.05, $^{\#\#}$: P<0.01, and $^{\#\#\#\#}$: P<0.0001 vs. Control. *: P<0.05, **: P<0.01, and ***: P<0.001 vs. PD models.

Cath-KP possessed antioxidant activity, as confirmed by ABTS, DPPH, and FRAP assays and oxidative stress measurement in MPP$^+$-treated cells (Figure 4; Supplementary Figure S4). Interestingly, prolonged exposure to sunlight and strong ultraviolet radiation may contribute to the development and production of antioxidant peptides in amphibians (Guo et al., 2014; Yang et al., 2009). We previously discovered an antioxidant peptide from *Fejervarya limnocharis*, a frog species that shares the same habitat as *K. pulchra* in tropical solar radiation-exposed Guangdong Province in a low-elevation region (N23.12°, E113.28°) (Chai et al., 2022). Therefore, our results provide further evidence that the development of antioxidant peptides is linked to ultraviolet exposure and the presence of antioxidant peptides reflects the adaptation of frog species to a specific environment (Chai et al., 2022; Guo et al., 2014; Liu et al., 2010; Yang et al., 2016).

MPTP/MPP$^+$ models have been widely used to assess pathology, pharmacology, and drug metabolism in PD (Dauer & Przedborski, 2003; Zhang et al., 2017). In this research, MPP$^+$-stimulated MN9D cells and MPTP-induced mice were used as PD models to explore the neuroprotective effects of Cath-KP *in vitro* and *in vivo*. Consistent with previous findings that MPP$^+$ can damage proteins, lipids, and mitochondria, consequently altering the status of brain antioxidant enzymes (Dauer & Przedborski, 2003; Zhang et al., 2017) and inducing Parkinsonian syndrome and behavioral deficits (Tillerson et al., 2002; Watanabe et al., 2008), our results showed a remarkable decrease in MPP$^+$-induced MN9D cell viability, paralleling a decrease in mitochondrial $\Delta\psi m$ and increase in intracellular and mitochondrial ROS levels (Figures 3E–G, 4A–F). Results also showed a concurrent decrease in the levels of antioxidant enzymes, including SOD and CAT (Figure 4G–I). Motor deficits and cognitive impairments in the MPTP-induced PD mice were also observed, as shown by the rotarod, open field, grasping, and pole-climbing tests (Figure 2B–H), which are widely used to assess balance and motor coordination, exploratory activity, locomotor activity, learning and memory ability, and movement disorders related to the basal ganglia (Dauer & Przedborski, 2003; Tillerson et al., 2002; Watanabe et al., 2008; Zhang et al., 2017). Consistent with its antioxidant effects, as shown by the ABTS and DPPH radical scavenging and FRAP antioxidant assays, Cath-KP increased antioxidant enzymes, attenuated intracellular and mitochondrial ROS and total NO accumulation, protected mitochondria, and reversed toxic effects in MPP$^+$-induced MN9D cells (Figures 3, 4). In addition, Cath-KP restored the antioxidant enzyme levels in the mouse brain tissues (Figure 5B, D), reduced oxidative damage from MPTP insult (Figure 2I–P), and improved mouse performance in the behavior tests (Figure 2B–H). Overall, these results demonstrate that Cath-KP can prevent PD damage by enhancing cell survival and reducing oxidative stress due to its antioxidant functions.

DA neurons in the SNpc and striatum diminish over time, resulting in motor impairments, which are the hallmark of PD. The TH enzyme is essential in dopamine production and phenotypic expression (Daubner et al., 2011). Motor function impairment induced by MPTP in PD animal models is typically accompanied by a marked reduction in TH-positive cells in the SNpc (Moreira et al., 2010). Similarly, in the current study, MPTP injection lowered TH content and TH-positive neurons

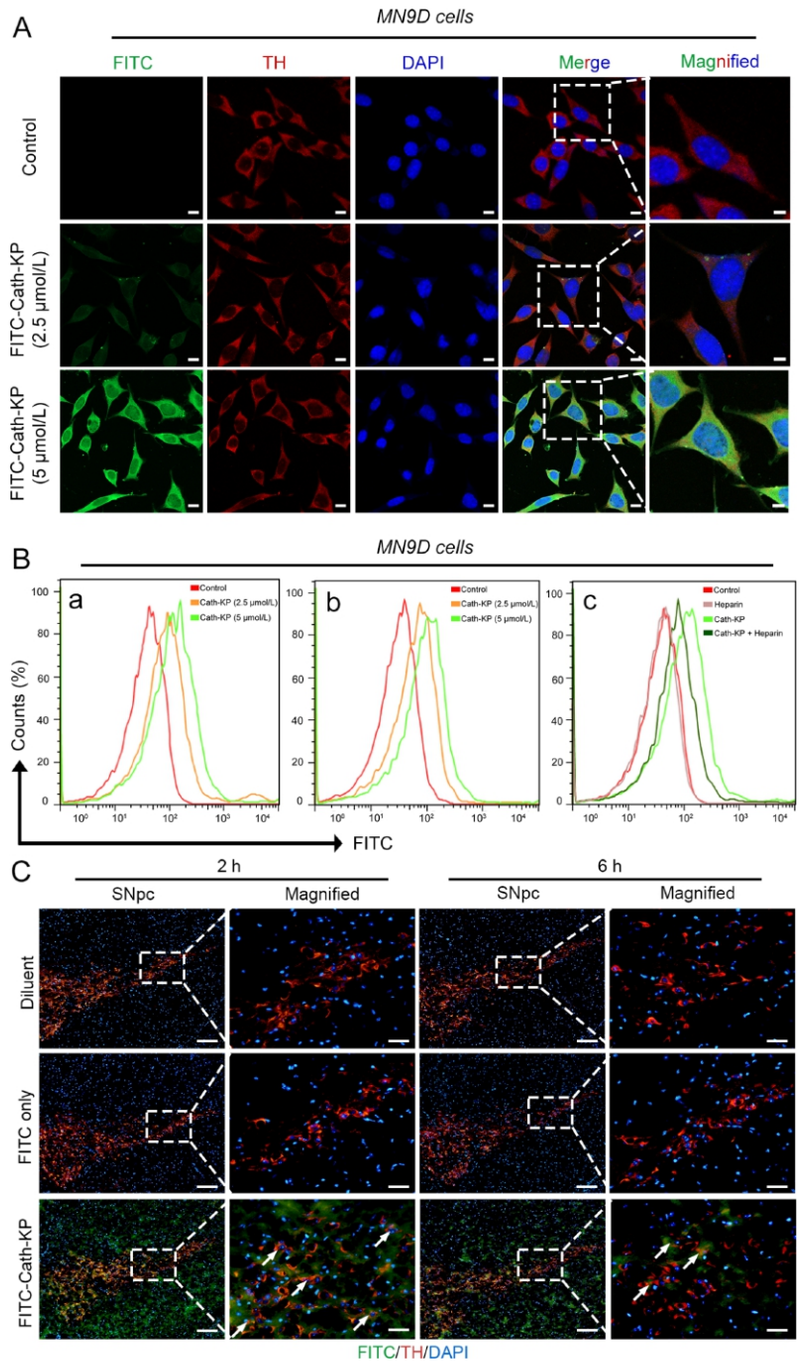


Figure 8 Internalization of Cath-KP into MN9D cells and brain tissue

A: Representative confocal laser scanning microscopy images of FITC-Cath-KP entering MN9D cells. FITC-Cath-KP, green fluorescence; TH, red fluorescence; DAPI, blue fluorescence. Scale bar: 10 μm for original images and 5 μm for magnified images in white square. B: Representative flow cytometry analysis of FITC-Cath-KP entry into MN9D cells under different conditions. Cath-KP (2.5 and 5 μmol/L) was incubated with MN9D cells for 3 h at 37 °C (panel a), 4 °C (panel b), or in the presence of 50 μg/mL heparin (panel c). Fluorescence intensity was then measured by flow cytometry. C: Characteristic TH staining images of SNpc from mice at 2 and 6 h after diluent, FITC only, and FITC-Cath-KP injections. FITC-Cath-KP, green fluorescence; TH, red fluorescence; DAPI, blue fluorescence. White arrows in enlarged area of the right column show the presence of FITC-Cath-KP in DA neurons. Scale bars=100 μm for original images and 20 μm for magnified images.

in the striatum and SNpc (Figure 2I–M). Interestingly, Cath-KP significantly reversed these changes in both the striatum and MN9D cells (Figures 2I–M, 3A–G). Multiple cell death pathways exist, including apoptosis, cell necrosis, and ferroptosis (Callizot et al., 2019; Tang et al., 2023). Endogenous apoptosis (also known as mitochondrial apoptosis), which greatly affects PD pathophysiology, can be triggered by excessive ROS accumulation (Radi et al., 2014). Apoptosis-induced cell death has been demonstrated in PD models *in vivo* (Radi et al., 2014). In line with these studies, our results further demonstrated that apoptosis in MN9D cells and the SNpc was substantially enhanced by MPP^+/MPTP exposure but was inhibited in the presence of Cath-KP, as evidenced by the increasing Bcl-2/Bax ratio (Figures 2N–P, 3H–J), a key factor regulating cell apoptosis (Bar-Am et al., 2005). Thus, the protective effects of Cath-KP on DA neurons

may be associated with its suppression of the mitochondrial apoptotic pathway.

The Sirt1/Nrf2 pathway regulated both inflammatory response and redox homeostasis (Fão et al., 2019). Brain-specific deletion of Sirt1 significantly increases ROS levels, exacerbating mitochondrial dysregulation and promoting cell death. The SNpc experiences more severe DA cell loss in the absence of Nrf2 because Nrf2 activation can up-regulate the expression of antioxidant enzymes/proteins like SOD, HO-1, and glutathione S-transferase, offering protection against oxidative stress (Wang et al., 2018). Consistent with previous research (Dong et al., 2016; Wang et al., 2018), our study showed that MPTP/MPP^+ treatment markedly decreased Sirt1 protein levels. Notably, Cath-KP administration reversed this decrease and subsequently increased the expression and nuclear location of Nrf2 and its downstream effectors (SOD, HO-1, and CAT) in both MN9D cells and brain tissue, thus protecting against MPP^+-induced oxidative damage (Figures 4G–I, 5, 7). Additionally, the effects of Cath-KP were suppressed by the Sirt1-specific inhibitor EX527 (Supplementary Figure S5). These findings suggest that the Sirt1/Nrf2 antioxidant pathway plays an essential role in the neuroprotective effects of Cath-KP in PD.

Nrf2 dissociates from its negative modulator Keap1 in the cytosol, translocates into the nucleus, and binds to AREs, thus promoting the transcription of antioxidant proteins (Fão et al., 2019). Here, compared to the control group, Keap1 expression decreased in the MN9D cells treated with MPP^+ but not significantly in the MPTP-treated tissues (Figure 5). Previous studies have reported that the protein level of Keap1 can be significantly decreased in endothelial cells by H_2O_2 and MPP^+ stimulation (Guo et al., 2021; Zhu et al., 2019), similar to our MN9D cell results. Studies have also shown that irigenin can suppress Keap1 expression, with GABA treatment partially rescuing this decrease (Guo et al., 2021; Zhu et al., 2019), correlating well with our cell experiments (Figure 5). Interestingly, research has also reported that SAP can reduce Keap1 expression in mice, which can be reversed by the antioxidant sitagliptin via Keap1/Nrf2 signaling pathway activation (Kong et al., 2021). However, oxidative stress does not modify the expression of Keap1 in H9c2 cells (Yao et al., 2022); as such, oxidative stress may induce different effects on the Keap1 protein, which warrants further study. Of note, changes in Keap1 expression do not appear to affect Nrf2 expression or its downstream antioxidant molecules, such as SOD and HO-1 (Bian et al., 2021; Choi et al., 2018; Guo et al., 2021; Kong et al., 2021; Pan et al., 2022; Yao et al., 2022; Zhu et al., 2019). In accordance with this conclusion, the expression patterns of other signaling proteins from cells and brain tissues remained consistent, even with variations in Keap1 expression, when compared to the MPTP/MPP^+ treatment groups (Figure 5).

Accumulating evidence suggests that overexpression of Sirt1 can inhibit p38 activation, leading to elevated SOD-1 and Bcl-2 expression, and reduced neuronal and inducible NO synthases governing NO generation (Aquilano et al., 2008; Bloem et al., 2021; Fão et al., 2019; Wang et al., 2018). In addition, p38 may phosphorylate Nrf2, facilitating its association with Keap1 and inhibiting its nuclear translocation (Fão et al., 2019; Karunakaran & Ravindranath, 2009; Newhouse et al., 2004). Thus, various antioxidants, including xanthohumol, ethanol extract of *Centipeda minima* (ECM), eriodictyol, salidroside, and anthocyanin loaded nanoparticles (An-NPs), can exert antioxidant and neuroprotective effects by activating the Nrf2 pathway while dephosphorylating p38 in different cells and animal models (Amin et al., 2017; Fão et al., 2019; Jiao et al., 2020; Jing et al., 2015; Wang et al., 2018, 2019). Consistently, Cath-KP mitigated MPP^+-induced p38 activation (Figure 6). The intracellular tyrosine kinase FAK is thought to control a variety of cellular processes, including cell adhesion, migration, invasion, polarity, proliferation, and survival (Parsons, 2003). MPP^+ treatment in SH-SY5Y cells inhibits FAK phosphorylation, while luteolin, which possesses antioxidative properties, stabilizes phosphorylated FAK levels and exhibits neuroprotective effects (Reudhabibadh et al., 2021). In addition, maclurin, which shows pro-oxidant activity, can activate p38 signaling and inhibit c-JNK, FAK, AKT, and c-Myc signaling in PC3 cells (Lee et al., 2018). Similarly, our results showed that Cath-KP inhibited p38 phosphorylation and induced further suppression of FAK phosphorylation induced by MPTP/MPP^+ treatment (Figure 6). Previous research has indicated that Sirt1/FAK regulates the nuclear translocation of p65 via phosphorylation of p38 (Grbčić et al., 2017; Karunakaran & Ravindranath, 2009; Wan et al., 2019). Thus, Cath-KP likely exerts its neuroprotective effects on PD models by activating the Sirt1/Nrf2 pathway, promoting the expression of antioxidant enzymes, and inhibiting the mitochondrial apoptotic pathway. Furthermore, FAK and p38 appear to be involved in the regulation of this pathway, although the exact regulatory mechanism requires further elucidation.

The BBB impedes the delivery of many therapeutic agents to the central nervous system, complicating treatment approaches for many neurological disorders (Guidotti et al., 2017). In our study, Cath-KP showed potential properties for cell penetration and BBB traversal (Figure 8). According to conventional wisdom, under physiological pH conditions, the guanidine head group of arginine can form bidentate hydrogen bonds with the negatively charged sulphate, carboxylic, and phosphate groups on cell membranes, resulting in cellular internalization of cell-penetrating peptides (Guidotti et al., 2017; Rothbard et al., 2004). Therefore, the transduction capabilities of cell-penetrating peptides can be greatly influenced by the amount and order of amino acids in their sequences, especially arginine. Consistently, Cath-KP contained six arginine residues and heparin sulfate suppressed its cellular uptake (Figures 1A, 8B). Moreover, its cellular internalization was energy-dependent, as evidenced by the decreased uptake rate at 4 °C than at 37 °C (Figure 8B). Thus, the cellular internalization mechanism of Cath-KP functions via energy-dependent endocytosis rather than energy-independent direct penetration. The primary sequence of Cath-KP contained an RTS motif, which is frequently found in disintegrins targeting $\alpha_1\beta_1$ integrin (Calvete et al., 2007). In the mammalian central nervous system, $\alpha_1\beta_1$ integrin is expressed in astrocytes and endothelial cells and plays an important role in maintaining the physiological integrity and stability of the BBB (Peng et al., 2008; Yonezawa et al., 2010). Our molecular docking and antibody competent assays showed that Cath-KP can bind to $\alpha_1\beta_1$ integrin (Supplementary Figures S8, S9). Consequently, we cannot rule out the possibility that the direct translocation of Cath-KP across the BBB is related to its effects on $\alpha_1\beta_1$ integrin in astrocytes. Therefore, more detailed research is required to further explore its uptake process. Nevertheless, the cellular uptake of Cath-KP highlights its value as a drug candidate for

PD therapy. Furthermore, given the role of non-human primates in the development of therapeutics for human diseases (Lei et al., 2015; Li et al., 2021a, 2021b; Su et al., 2015), evaluating the effects of Cath-KP in non-human primate models of PD is essential.

In conclusion, we identified a novel antioxidant peptide (Cath-KP) belonging to the cathelicidin family from the skin of *K. pulchra*. Our results showed that Cath-KP alleviated dyskinesia in MPTP mice by reversing DA neuron loss in the SNpc and decreasing TH levels in the striatum. Mechanistically, Cath-KP penetrated cells and exerted a neuroprotective effect via Sirt1/Nrf2 pathway modulation, further promoting the expression of antioxidant enzymes and alleviating the accumulation of ROS. Additionally, FAK and p38 contributed to this regulatory process. These findings demonstrate the potential of Cath-KP as a therapeutic option for PD that targets oxidative stress.

DATA AVAILABILITY

The raw RNA-seq data were deposited in the NCBI Sequence Read Archive (SRA)(BioProjectID PRJNA1021395), GSA Database (Accession No. CRA013484), and Science Data Bank (https://doi.org/10.57760/sciencedb.j00139.00068).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

H.L. and J.C. performed the major experiments, organized the data, and wrote the original manuscript. S.Q. and X.X. designed the experiments, supervised the study, evaluated the data, and revised the article. Z.X. performed molecular docking. J.W., S.H., H.L., P.H., and X.C. carried out the experiments. X.H. and H.J. revised the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Amin FU, Shah SA, Badshah H, et al. 2017. Anthocyanins encapsulated by PLGA@PEG nanoparticles potentially improved its free radical scavenging capabilities via p38/JNK pathway against $A\beta_{1-42}$-induced oxidative stress. *Journal of Nanobiotechnology*, **15**(1): 12.

Aquilano K, Baldelli S, Rotilio G, et al. 2008. Role of nitric oxide synthases in Parkinson's disease: a review on the antioxidant and anti-inflammatory activity of polyphenols. *Neurochemical Research*, **33**(12): 2416−2426.

Bar-Am O, Weinreb O, Amit T, et al. 2005. Regulation of Bcl-2 family proteins, neurotrophic factors, and APP processing in the neurorescue activity of propargylamine. *The FASEB Journal*, **19**(13): 1899−1901.

Bian YQ, Chen Y, Wang XF, et al. 2021. Oxyphylla A ameliorates cognitive deficits and alleviates neuropathology via the Akt-GSK3β and Nrf2-Keap1-HO-1 pathways in *vitro* and in *vivo* murine models of Alzheimer's disease. *Journal of Advanced Research*, **34**: 1−12.

Bloem BR, Okun MS, Klein C. 2021. Parkinson's disease. *The Lancet*, **397**(10291): 2284−2303.

Callizot N, Combes M, Henriques A, et al. 2019. Necrosis, apoptosis, necroptosis, three modes of action of dopaminergic neuron neurotoxins. *PLoS One*, **14**(4): e0215277.

Calvete JJ, Marcinkiewicz C, Sanz L. 2007. KTS and RTS-disintegrins: anti-angiogenic viper venom peptides specifically targeting the α1β1 integrin. *Current Pharmaceutical Design*, **13**(28): 2853−2859.

Cao XQ, Wang Y, Wu CY, et al. 2018. Cathelicidin-OA1, a novel antioxidant peptide identified from an amphibian, accelerates skin wound healing. *Scientific Reports*, **8**(1): 943.

Chai JW, Chen X, Ye TF, et al. 2021. Characterization and functional analysis of cathelicidin-MH, a novel frog-derived peptide with anti-septicemic properties. *eLife*, **10**: e64411.

Chai JW, Liu JF, Tian ML, et al. 2022. Multiple mechanistic action of Brevinin-1FL peptide against oxidative stress effects in an acute inflammatory model of carrageenan-induced damage. *Oxidative Medicine and Cellular Longevity*, **2022**: 2615178.

Choi JH, Jang M, Lee JI, et al. 2018. Neuroprotective effects of a traditional multi-herbal medicine Kyung-Ok-Ko in an animal model of Parkinson's disease: inhibition of MAPKs and NF-κB pathways and activation of Keap1-Nrf2 pathway. *Frontiers in Pharmacology*, **9**: 1444.

Conlon JM, Kim JB. 2000. A protease inhibitor of the Kunitz family from skin secretions of the tomato frog, *Dyscophus guineti* (Microhylidae). *Biochemical and Biophysical Research Communications*, **279**(3): 961−964.

Daubner SC, Le T, Wang SZ. 2011. Tyrosine hydroxylase and regulation of dopamine synthesis. *Archives of Biochemistry and Biophysics*, **508**(1): 1−12.

Dauer W, Przedborski S. 2003. Parkinson's disease: mechanisms and models. *Neuron*, **39**(6): 889−909.

Dong SY, Guo YJ, Feng Y, et al. 2016. The epigenetic regulation of HIF-1α by SIRT1 in MPP^+ treated SH-SY5Y cells. *Biochemical and Biophysical Research Communications*, **470**(2): 453−459.

Dorsey ER, Constantinescu R, Thompson JP, et al. 2007. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology*, **68**(5): 384−386.

Eiríksdóttir E, Konate K, Langel Ü, et al. 2010. Secondary structure of cell-penetrating peptides controls membrane interaction and insertion. *Biochimica et Biophysica Acta*, **1798**(6): 1119−1128.

Fão L, Mota SI, Rego AC. 2019. Shaping the Nrf2-ARE-related pathways in Alzheimer's and Parkinson's diseases. *Ageing Research Reviews*, **54**: 100942.

Fauna CGOCMA. 1983. China Medicinal Animal Fauna. Tianjin: Tianjin Science and Technology Press. (in Chinese)

Fei L, Ye CY, Jiang JP. 2012. Colored Atlas of Chinese Amphibians and Their Distributions. Chengdu: Sichuan Science and Technology Press. (in Chinese)

Feng GZ, Wei L, Che HL, et al. 2022. Cathelicidin-NV from *Nanorana ventripunctata* effectively protects HaCaT cells, ameliorating ultraviolet B-induced skin photoaging. *Peptides*, **150**: 170712.

Gao YH, Chai JW, Wu JN, et al. 2022. Molecular cloning and characterization of a novel antimicrobial peptide from the skin of *Kaloula pulchra*. *Current Pharmaceutical Biotechnology*, **23**(15): 1873–1882.

Grbčić P, Tomljanović I, Klobučar M, et al. 2017. Dual sphingosine kinase inhibitor SKI-II enhances sensitivity to 5-fluorouracil in hepatocellular carcinoma cells via suppression of osteopontin and FAK/IGF-1R signalling. *Biochemical and Biophysical Research Communications*, **487**(4): 782−788.

Guidotti G, Brambilla L, Rossi D. 2017. Cell-penetrating peptides: from basic research to clinics. *Trends in Pharmacological Sciences*, **38**(4): 406−424.

Guo C, Hu YH, Li J, et al. 2014. Identification of multiple peptides with antioxidant and antimicrobial activities from skin and its secretions of *Hylarana taipehensis*, *Amolops lifanensis*, and *Amolops granulosus*. *Biochimie*, **105**: 192–201.

Guo F, Wang XX, Liu XX. 2021. Protective effects of irigenin against 1-methyl-4-phenylpyridinium-induced neurotoxicity through regulating the Keap1/Nrf2 pathway. *Phytotherapy Research*, **35**(3): 1585−1596.

Hao X, Yang HL, Wei L, et al. 2012. Amphibian cathelicidin fills the evolutionary gap of cathelicidin in vertebrate. *Amino Acids*, **43**(2): 677−685.

Jiao Y, Cao YZ, Lu XY, et al. 2020. Xanthohumol protects neuron from cerebral ischemia injury in experimental stroke. *Molecular Biology Reports*,

47(4): 2417−2425.

Jing X, Shi HY, Zhu XY, et al. 2015. Eriodictyol attenuates β-amyloid 25–35 peptide-induced oxidative cell death in primary cultured neurons by activation of Nrf2. *Neurochemical Research*, **40**(7): 1463−1471.

Karunakaran S, Ravindranath V. 2009. Activation of p38 MAPK in the substantia nigra leads to nuclear translocation of NF-κB in MPTP-treated mice: implication in Parkinson's disease. *Journal of Neurochemistry*, **109**(6): 1791−1799.

Kim HY, Yum SY, Jang G, et al. 2015. Discovery of a non-cationic cell penetrating peptide derived from membrane-interacting human proteins and its potential as a protein delivery carrier. *Scientific Reports*, **5**: 11719.

Kong LM, Deng J, Zhou X, et al. 2021. Sitagliptin activates the p62-Keap1-Nrf2 signalling pathway to alleviate oxidative stress and excessive autophagy in severe acute pancreatitis-related acute lung injury. *Cell Death & Disease*, **12**(10): 928.

Krishnan N, Dickman MB, Becker DF. 2008. Proline modulates the intracellular redox environment and protects mammalian cells against oxidative stress. *Free Radical Biology and Medicine*, **44**(4): 671−681.

Lee YJ, Jung O, Lee J, et al. 2018. Maclurin exerts anti-cancer effects on PC3 human prostate cancer cells via activation of p38 and inhibitions of JNK, FAK, AKT, and c-Myc signaling pathways. *Nutrition Research*, **58**: 62−71.

Lei XG, Li H, Huang BH, et al. 2015. 1-Methyl-4-phenylpyridinium stereotactic infusion completely and specifically ablated the nigrostriatal dopaminergic pathway in rhesus macaque. *PLoS One*, **10**(5): e0127953.

Li H, Su LY, Yang LX, et al. 2021a. A cynomolgus monkey with naturally occurring Parkinson's disease. *National Science Review*, **8**(3): nwaa292.

Li H, Yao YG, Hu XT. 2021b. Biological implications and limitations of a cynomolgus monkey with naturally occurring Parkinson's disease. *Zoological Research*, **42**(2): 138−140.

Li K, Wang M, Huang ZH, et al. 2023. ALOX5 inhibition protects against dopaminergic neurons undergoing ferroptosis. *Pharmacological Research*, **193**: 106779.

Liu CB, Hong J, Yang HL, et al. 2010. Frog skins keep redox homeostasis by antioxidant peptides with rapid radical scavenging ability. *Free Radical Biology and Medicine*, **48**(9): 1173−1181.

Moreira ELG, Rial D, Aguiar Jr AS, et al. 2010. Proanthocyanidin-rich fraction from *Croton celtidifolius* Baill confers neuroprotection in the intranasal 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine rat model of Parkinson's disease. *Journal of Neural Transmission*, **117**(12): 1337−1351.

Mwangi J, Hao X, Lai R, et al. 2019. Antimicrobial peptides: new hope in the war against multidrug resistance. *Zoological Research*, **40**(6): 488−505.

Newhouse K, Hsuan SL, Chang SH, et al. 2004. Rotenone-induced apoptosis is mediated by p38 and JNK MAP kinases in human dopaminergic SH-SY5Y cells. *Toxicological Sciences*, **79**(1): 137−146.

Pan SP, Wei HM, Yuan SS, et al. 2022. Probiotic *Pediococcus pentosaceus* ameliorates MPTP-induced oxidative stress *via* regulating the gut microbiota-gut-brain axis. *Frontiers in Cellular and Infection Microbiology*, **12**: 1022879.

Parsons JT. 2003. Focal adhesion kinase: the first ten years. *Journal of Cell Science*, **116**(8): 1409−1416.

Peng HS, Shah W, Holland P, et al. 2008. Integrins and dystroglycan regulate astrocyte wound healing: the integrin β1 subunit is necessary for process extension and orienting the microtubular network. *Developmental Neurobiology*, **68**(5): 559−574.

Qi RH, Chen Y, Guo ZL, et al. 2019. Identification and characterization of two novel cathelicidins from the frog *Odorrana livida*. *Zoological Research*, **40**(2): 94–101.

Qin SF, Rodrigues GA. 2012. Roles of αvβ5, FAK and MerTK in oxidative stress inhibition of RPE cell phagocytosis. *Experimental Eye Research*, **94**(1): 63−70.

Qu LL, Lin BQ, Zeng WP, et al. 2022. Lysosomal K^+ channel TMEM175 promotes apoptosis and aggravates symptoms of Parkinson's disease. *EMBO Reports*, **23**(9): e53234.

Radi E, Formichi P, Battisti C, et al. 2014. Apoptosis and oxidative stress in neurodegenerative diseases. *Journal of Alzheimer's Disease*, **42 Suppl 3**: S125–S152.

Reudhabibadh R, Binlateh T, Chonpathompikunlert P, et al. 2021. Suppressing Cdk5 activity by luteolin inhibits MPP^+-induced apoptotic of neuroblastoma through ERK/DRP1 and FAK/Akt/GSK3β pathways. *Molecules*, **26**(5): 1307.

Rothbard JB, Jessop TC, Lewis RS, et al. 2004. Role of membrane potential and hydrogen bonding in the mechanism of translocation of guanidinium-rich peptides into cells. *Journal of the American Chemical Society*, **126**(31): 9506−9507.

Sancho-Vaello E, François P, Bonetti EJ, et al. 2017. Structural remodeling and oligomerization of human cathelicidin on membranes suggest fibril-like structures as active species. *Scientific Reports*, **7**(1): 15371.

Su LY, Li H, Lv L, et al. 2015. Melatonin attenuates MPTP-induced neurotoxicity via preventing CDK5-mediated autophagy and SNCA/α-synuclein aggregation. *Autophagy*, **11**(10): 1745−1759.

Subramaniam SR, Chesselet MF. 2013. Mitochondrial dysfunction and oxidative stress in Parkinson's disease. *Progress in Neurobiology*, **106–107**: 17–32.

Tang F, Zhou LY, Li P, et al. 2023. Inhibition of ACSL4 alleviates parkinsonism phenotypes by reduction of lipid reactive oxygen species. *Neurotherapeutics*,doi: 10.1007/s13311-023-01382-4.

Tatton WG, Chalmers-Redman R, Brown D, et al. 2003. Apoptosis in Parkinson's disease: signals for neuronal degradation. *Annals of Neurology*, **53 Suppl 3**: S61–S72.

Tillerson JL, Caudle WM, Reverón ME, et al. 2002. Detection of behavioral impairments correlated to neurochemical deficits in mice treated with moderate doses of 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine. *Experimental Neurology*, **178**(1): 80−90.

Vidoni C, Secomandi E, Castiglioni A, et al. 2018. Resveratrol protects neuronal-like cells expressing mutant Huntingtin from dopamine toxicity by rescuing ATG4-mediated autophagosome formation. *Neurochemistry International*, **117**: 174−187.

Wan XX, Chowdhury IH, Jie ZL, et al. 2019. Origin of monocytes/macrophages contributing to chronic inflammation in chagas disease: SIRT1 inhibition of FAK-NFκB-dependent proliferation and proinflammatory activation of macrophages. *Cells*, **9**(1): 80.

Wang CY, Sun ZN, Wang MX, et al. 2018. SIRT1 mediates salidroside-elicited protective effects against MPP^+-induced apoptosis and oxidative stress in SH-SY5Y cells: involvement in suppressing MAPK pathways. *Cell Biology International*, **42**(1): 84−94.

Wang Y, Ouyang JH, Luo XJ, et al. 2021. Identification and characterization of novel bi-functional cathelicidins from the black-spotted frog (*Pelophylax nigromaculata*) with both anti-infective and antioxidant activities. *Developmental & Comparative Immunology*, **116**: 103928.

Wang YJ, Wang XY, Hao XY, et al. 2019. Ethanol extract of *Centipeda minima* exerts antioxidant and neuroprotective effects via activation of the Nrf2 signaling pathway. *Oxidative Medicine and Cellular Longevity*, **2019**: 9421037.

Watanabe Y, Kato H, Araki T. 2008. Protective action of neuronal nitric oxide synthase inhibitor in the MPTP mouse model of Parkinson's disease. *Metabolic Brain Disease*, **23**(1): 51−69.

Wei L, Yang JJ, He XQ, et al. 2013. Structure and function of a potent lipopolysaccharide-binding antimicrobial and anti-inflammatory peptide. *Journal of Medicinal Chemistry*, **56**(9): 3546−3556.

Wiriyaampaiwong P, Karnmongkol C, Punpad A, et al. 2022. Cathelicidin-HR from *Hoplobatrachus rugulosus*: an antioxidant peptide that performs a

protective effect against UV/H_2O_2 -induced DNA damage. *Journal of Applied Biology & Biotechnology*, **10**(4): 35−44.

Xu XQ, Lai R. 2015. The chemistry and biological activities of peptides from amphibian skin secretions. *Chemical Reviews*, **115**(4): 1760−1846.

Yang HL, Wang X, Liu XH, et al. 2009. Antioxidant peptidomics reveals novel skin antioxidant system. *Molecular & Cellular Proteomics*, **8**(3): 571−583.

Yang XW, Wang Y, Zhang Y, et al. 2016. Rich diversity and potency of skin antioxidant peptides revealed a novel molecular basis for high-altitude adaptation of amphibians. *Scientific Reports*, **6**: 19866.

Yao H, Xie Q, He QM, et al. 2022. Pretreatment with panaxatriol saponin attenuates mitochondrial apoptosis and oxidative stress to facilitate treatment of myocardial ischemia-reperfusion injury via the regulation of Keap1/Nrf2 activity. *Oxidative Medicine and Cellular Longevity*, **2022**: 9626703.

Yonezawa T, Hattori S, Inagaki J, et al. 2010. Type IV collagen induces expression of thrombospondin-1 that is mediated by integrin α1β1 in astrocytes. *Glia*, **58**(7): 755−767.

Zeng BS, Chai JW, Deng ZH, et al. 2018. Functional characterization of a novel lipopolysaccharide-binding antimicrobial and anti-inflammatory peptide in vitro and in vivo. *Journal of Medicinal Chemistry*, **61**(23): 10709−10723.

Zeng QY, Zeng BS, Chai JW, et al. 2020a. Antioxidant properties and neuroprotective effects of Esc-1GN through the regulation of MAPK and AKT signaling. *Life Sciences*, **254**: 117753.

Zeng XK, Xu K, Wang J, et al. 2020b. Pretreatment of ascorbic acid inhibits MPTP-induced astrocytic oxidative stress through suppressing NF-*κ*B signaling. *Neural Plasticity*, **2020**: 8872296.

Zhang QS, Heng Y, Mou Z, et al. 2017. Reassessment of subacute MPTP-treated mice as animal model of Parkinson's disease. *Acta Pharmacologica Sinica*, **38**(10): 1317−1328.

Zhang Y, Wang QQ, Zhao Z, et al. 2021. Animal secretory endolysosome channel discovery. *Zoological Research*, **42**(2): 141−152.

Zhang YX, Wang M, Wei SS. 2010. Isolation and characterization of a trypsin inhibitor from the skin secretions of *Kaloula pulchra hainana*. *Toxicon*, **56**(4): 502–507.

Zhu ZH, Shi ZG, Xie CL, et al. 2019. A novel mechanism of Gamma-aminobutyric acid (GABA) protecting human umbilical vein endothelial cells (HUVECs) against H_2O_2-induced oxidative injury. *Comparative Biochemistry and Physiology Part C:Toxicology & Pharmacology*, **217**: 68−75.

Liang et al. *Zool. Res.* 2024, 45(1): 125−135
https://doi.org/10.24272/j.issn.2095-8137.2023.076

ZR | Zoological Research

Article Open Access

Patterns and drivers of avian taxonomic and phylogenetic beta diversity in China vary across geographical backgrounds and dispersal abilities

Jian-Chao Liang[1,#], Zhi-Feng Ding[1,#], Chun-Lin Li[2], Yi-Ming Hu[1], Zhi-Xin Zhou[1], Gan-Wen Lie[3], Xiao-Nan Niu[4], Wen-Bin Huang[5], Hui-Jian Hu[1,*], Xing-Feng Si[6,*]

[1] *Guangdong Key Laboratory of Animal Conservation and Resource Utilization*, *Guangdong Public Laboratory of Wild Animal Conservation and Utilization*, *Institute of Zoology*, *Guangdong Academy of Sciences*, *Guangzhou*, *Guangdong* 510260, *China*
[2] *School of Resources and Environmental Engineering*, *Anhui University*, *Hefei*, *Anhui* 230601, *China*
[3] *Guangdong Eco-Engineering Polytechnic*, *Guangzhou*, *Guangdong* 510520, *China*
[4] *Huizhou Institute of Forestry (Huizhou Botanical Garden Management Service Center)*, *Huizhou*, *Guangdong* 516001, *China*
[5] *Administration of National Nature Reserve for Marine Ecology of Guangdong Nanpeng Islands*, *Shantou*, *Guangdong* 515900, *China*
[6] *Zhejiang Zhoushan Archipelago Observation and Research Station*, *Institute of Eco-Chongming*, *Zhejiang Tiantong Forest Ecosystem National Observation and Research Station*, *School of Ecological and Environmental Sciences*, *East China Normal University*, *Shanghai* 200241, *China*

ABSTRACT

Geographical background and dispersal ability may strongly influence assemblage dissimilarity; however, these aspects have generally been overlooked in previous large-scale beta diversity studies. Here, we examined whether the patterns and drivers of taxonomic beta diversity (TBD) and phylogenetic beta diversity (PBD) of breeding birds in China vary across (1) regions on both sides of the Hu Line, which demarcates China's topographical, climatic, economic, and social patterns, and (2) species with different dispersal ability. TBD and PBD were calculated and partitioned into turnover and nestedness components using a moving window approach. Variables representing climate, habitat heterogeneity, and habitat quality were employed to evaluate the effects of environmental filtering. Spatial distance was considered to assess the impact of dispersal limitation. Variance partitioning analysis was applied to assess the relative roles of these variables. In general, the values of TBD and PBD were high in mountainous areas and were largely determined by environmental filtering. However, different dominant environmental filters on either side of the Hu Line led to divergent beta diversity patterns. Specifically, climate-driven species turnover and habitat heterogeneity-related species nestedness dominated the regions east and west of the line, respectively. Additionally, bird species with stronger dispersal ability were more susceptible to environmental filtering, resulting in more homogeneous assemblages. Our results indicated that regions with distinctive geographical backgrounds may present different ecological factors that lead to divergent assemblage dissimilarity patterns, and dispersal ability determines the response of assemblages to these ecological factors. Identifying a single universal explanation for the observed pattern without considering these aspects may lead to simplistic or incomplete conclusions. Consequently, a comprehensive understanding of large-scale beta diversity patterns and effective planning of conservation strategies necessitate the consideration of both geographical background and species dispersal ability.

Keywords: Beta diversity; Environmental filtering; Dispersal limitation; Hu Line; Species dispersal ability; Breeding birds

INTRODUCTION

Understanding how and why species composition varies

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 19 June 2023; Accepted: 04 September 2023; Online: 05 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (31901220), Science and Technology Planning Project of Guangdong Province, China (2019B121202004), Guangdong Basic and Applied Basic Research Foundation (2021A1515110744), and Forestry Administration of Guangdong Province (DFGP Project of Fauna of Guangdong-202115)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: 13570909977@139.com; sixf@des.ecnu.edu.cn

spatially is a core area of biogeographic research, given their complex relationships with community assembly (García-Girón et al., 2020; Jiao et al., 2021), biogeographical regionalization (He et al., 2017; Kreft & Jetz, 2010), and biodiversity conservation (Hill et al., 2021; Socolar et al., 2016). Beta diversity, a measure of assemblage dissimilarity among sites, provides a promising approach for addressing these questions. As beta diversity is divided into two distinct and antithetical components, turnover and nestedness, it can provide comprehensive information for assessing assemblage dissimilarity. Turnover signifies the substitution of certain species by others among sites, whereas nestedness denotes differences in species richness caused by species gain or loss among sites (Baselga, 2010, 2012). Generally, turnover can be attributed to factors related to climate or habitat, whereas nestedness can be attributed to factors related to disturbance, productivity, or physical barriers (Alahuhta et al., 2017; Sebastián-González et al., 2020). Therefore, exploring the patterns of beta diversity and its components can provide insights not only into spatial variations in species composition, but also into the mechanisms underlying these variations (Carvalho et al., 2012).

At a broad scale, the mechanisms responsible for beta diversity patterns originate from two ecological processes: environmental filtering and dispersal limitation (König et al., 2017; Qian et al., 2020a; Wang et al., 2012). Environmental filtering implies that environmental factors, such as climate, habitat quality, and habitat heterogeneity, determine species presence or absence; the greater the similarity in the environment between any two sites, the greater the similarity in species assemblages within them (Qian et al., 2020a; Wiescher et al., 2012). In contrast, dispersal limitation proposes that spatial factors, such as geographical distance, influence the probability of species exchange through dispersal; the larger the distance between any two sites, the lower the similarity in species assemblages within them (Nekola & White, 1999). However, as distance increases, species isolation intensifies and environmental dissimilarity increases. Therefore, the challenge in understanding the mechanisms driving beta diversity patterns is to evaluate the relative roles of spatial and environmental factors across different regions (Chase, 2014).

However, growing evidence has indicated that the relative importance of different ecological factors is dependent on both geography and taxonomy (Stuart et al., 2012; Zellweger et al., 2017). Regions with varied geographical characteristics may be influenced by different ecological factors, resulting in diverse beta diversity patterns. For instance, König et al. (2017) reported that vascular plant assemblages in island systems are more influenced by environmental factors than those in mainland systems, resulting in a lower turnover rate among islands globally. Analogously, as the heterogeneous geographical backgrounds of mainland systems are characterized by distinct configurations of environmental conditions, it is reasonable to expect variations in the contributions of both spatial and environmental factors, as well as in the contributions of specific environmental factors across geographical locations. Nevertheless, most large-scale studies on beta diversity have focused on identifying a single universal explanation for the observed patterns, without considering geographical variations in ecological factors (König et al., 2017). Consequently, this has led to simplistic and incomplete conclusions regarding conservation management.

In addition, species with different attributes may respond differently to ecological factors, potentially affecting beta diversity patterns. For example, as species with strong dispersal ability can reach an equilibrium with their environmental niches without being constrained by geographical barriers, they are assumed to form assemblages with homogeneous composition (Nekola & White, 1999). In contrast, species with weak dispersal ability are constrained to a small geographical range due to geographical isolation, and thus tend to form assemblages with heterogeneous composition (Wang et al., 2012). Therefore, comparing species with different dispersal abilities can improve our understanding of broad-scale beta diversity patterns. Although several studies have assessed the influence of dispersal ability on beta diversity, most have relied on comparisons between different taxa (e.g., Peixoto et al., 2017; Zellweger et al., 2017), which may be subjected to bias due to differences in other attributes, such as life history traits or reproductive modes. Therefore, to accurately determine the influence of dispersal ability on beta diversity, it is preferable to assess dispersal-based differences in species within the same taxon.

Additionally, recent developments in beta diversity have emphasized the necessity of incorporating phylogenetic information into the assessment of assemblage dissimilarity (i.e., phylogenetic beta diversity, PBD) (Graham & Fine, 2008). As a supplement to taxonomic beta diversity (TBD), PBD captures the contribution of the evolutionary history of species in assemblages (Jin et al., 2015). Importantly, by facilitating parallel spatial comparisons with TBD, PBD can elucidate the interactions between evolutionary and ecological factors that influence diversity patterns (Peixoto et al., 2017). For instance, PBD may be considerably lower than expected with respect to TBD in regions with strong human disturbance, as heavily modified habitats may favor traits expressed within certain phylogenetic lineages, resulting in phylogenetic clustering between sites (Liang et al., 2019; Weideman et al., 2020).

China has a rich biodiversity and a heterogeneous geography, providing a “natural laboratory” for exploring large-scale patterns of beta diversity. Notably, China harbors over 1 445 bird species, accounting for one-sixth of the total number of bird species worldwide (Zheng, 2017). However, while the geographical patterns of beta diversity have been thoroughly investigated for plants in China (Qian et al., 2020a, 2020b; Wang et al., 2012), similar studies on avian fauna are lacking. China is also characterized by a diverse geographical landscape and broad environmental gradient (Figure 1). Traditionally, the country can be divided into the eastern and western parts based on the Hu Line (also known as the Heihe-Tengchong Line), a geodemographic demarcation line proposed by the geographer Huan-Yong Hu (Hu, 1935). The eastern regions are characterized by warm and humid plains and accommodate over 90% of the national human population (Yao & Zhang, 2020; Yue et al., 2003). Contrastingly, the western regions are dominated by cold and arid plateaus and are inhabited by a relatively small fraction of the national human population. Additionally, an increasing number of studies have demonstrated that the Hu Line may account for differences not only in geographic and demographic patterns but also in economic, sociological, and ecological patterns in China. For instance, Zhang et al. (2022) reported that potential forestation lands in China are mainly located in the eastern

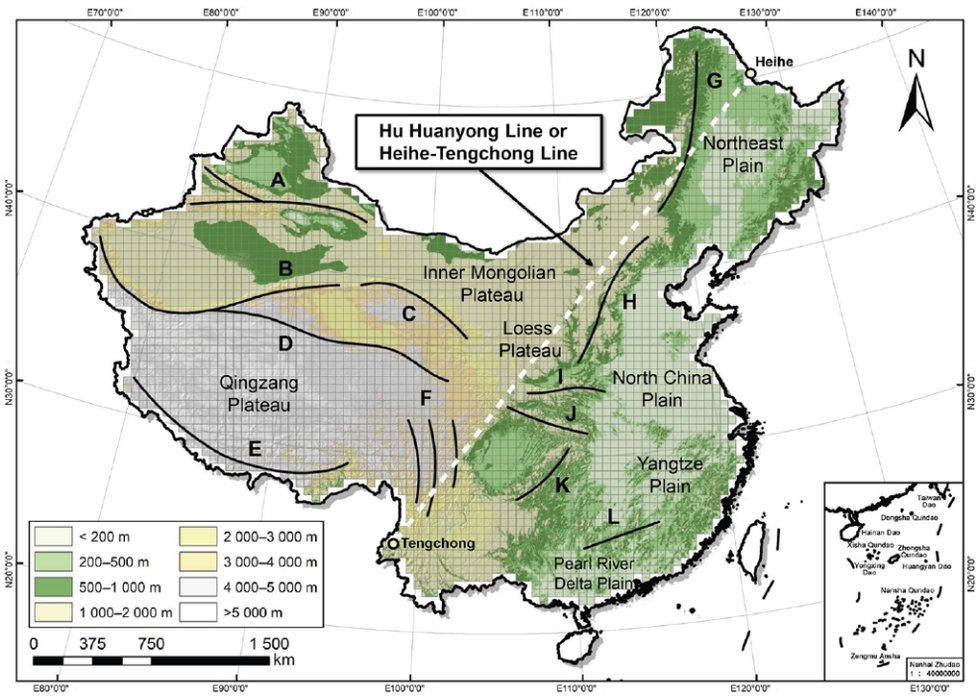


Figure 1 Topographic map of mainland China showing 50 km×50 km equal-area grid system and major mountain ranges

A: Mt. Tianshan; B: Mt. Aerjin; C: Mt. Qilian; D: Mt. Kunlun; E: Mt. Himalyas; F: Mt. Hengduan; G: Mt. Great Khingan; H: Mt. Taihang; I: Mt. Qinling; J: Mt. Daba; K: Mt. Wu; L: Mt. Nanling

regions under both present and future climate scenarios. Kong et al. (2022) found that ecological land changes (e.g., forests, grasslands, lakes) in China from 1995 to 2015 were primarily caused by clustered land degradation in the western regions, but by farmland activities in the eastern regions. Accordingly, it is reasonable to expect that species assemblages will show divergent patterns and mechanisms of compositional dissimilarity on either side of the Hu Line. Previous studies have suggested that assemblages will become more nested and phylogenetically homogeneous in anthropogenic landscapes, owing to increased extinction driven by intense human disturbance (Liang et al., 2019; Wayman et al., 2022; Weideman et al., 2020). However, whether this finding can be applied to the eastern region of the Hu Line, where human disturbance is more intense, remains insufficiently studied.

In the present study, we explored the spatial patterns and determinants of beta diversity in breeding avian assemblages across China, from both a taxonomic and phylogenetic perspective, considering the Hu Line demarcation. We aimed to assess variations in beta diversity and the relative importance of different ecological factors across (1) regions characterized by distinct geographical backgrounds and (2) species delineated by different dispersal abilities. We anticipate that avian assemblage dissimilarity will exhibit divergent patterns and determinants on either side of the Hu Line. Specifically, we predict that in regions east of the Hu Line, nestedness will contribute more substantially to overall dissimilarity, and the divergence between phylogenetic beta diversity and taxonomic beta diversity will be more pronounced. Moreover, we hypothesize that avian assemblages with weak dispersal ability will exhibit higher beta diversity and greater susceptibility to spatial influences, while those with strong dispersal ability will demonstrate lower beta diversity and greater susceptibility to environmental factors.

MATERIALS AND METHODS

Study area

Our study area covered mainland China (hereafter "China", N18°–53° and E73°−134°, Figure 1), covering an area of approximately 9 423 000 km^2. We converted the study area into an equal-area grid system at a spatial resolution of 50 km×50 km, considered an appropriate scale for beta diversity studies in China (e.g., He et al., 2017; Wang et al., 2012; Xu et al., 2014). To assess the robustness of our results in terms of spatial scales, the beta diversity analyses were repeated using 25 km×25 km, 100 km×100 km, and 200 km×200 km spatial resolutions, with the results found to be consistent with those obtained under the 50 km×50 km resolution (Supplementary Figure S1). Therefore, we only reported results obtained using 50 km×50 km resolution. Grids not entirely within the study area were excluded, with 3 486 grids thus retained for subsequent analyses.

Species distribution data

We accessed the previous avian checklists for China compiled by Zheng (2017), MacKinnon et al. (2003), and China Bird Report Checklist of Birds of China v.7.0. We also retrieved avian distribution data from the BirdLife Data Zone (BirdLife International, www.datazone.birdlife.org/). The distribution range of each bird species was categorized into resident, breeding season, non-breeding season, passage, and uncertain. We concentrated on breeding birds, as they exhibit more stable distributions than non-breeding birds, thus yielding more consistent responses to environmental factors. Using GIS software, we extracted the resident and breeding season ranges for each bird species and overlaid these onto the previously established grids covering China. Bird species whose resident and breeding season ranges intersected with any of these grids were considered as breeding birds in China. Conversely, birds whose resident and breeding season ranges

did not intersect with any grid were excluded from the study. Consequently, data for 1 146 species were obtained for subsequent analyses.

Phylogenetic reconstruction

We generated a phylogenetic tree for the 1 146 breeding birds using the "Phylogeny subsets" tool from a global bird phylogeny under the tree option "Hackett All Species: a set of 10 000 trees with 9 993 OTUs each" (Jetz et al., 2012). Subsequently, we sampled 5 000 trees from the pseudo-posterior distributions and calculated a maximum-clade credibility tree based on mean node heights using TreeAnnotator v.1.8.2 in the BEAST package (Drummond & Rambaut, 2007; Si et al., 2017). The resulting consensus tree was used for subsequent calculations of PBD (Supplementary Figure S2).

Environmental variables

Three groups of variables, i.e., climate, habitat heterogeneity, and habitat quality, were used to characterize environmental conditions for each grid (Supplementary Figure S3).

Climate is considered a strong environmental filter for assemblage dissimilarity due to differences in energy availability, physiological tolerance, and speciation (Wayman et al., 2022). For this study, climate variables included mean annual temperature, mean annual precipitation, potential evapotranspiration, mean annual temperature range, temperature seasonality, and precipitation seasonality, which were retrieved from the WorldClim database v.2.0 (www.worldclim.org) at a spatial resolution of 1 km×1 km. The value of each 50 km×50 km grid for all climate variables was calculated as the average of values of all 1 km×1 km grids within it.

Habitat heterogeneity is another important environmental driver of assemblage dissimilarity because the absence or presence of certain habitats may cause the loss or gain of their specialists (Calmé & Desrochers, 1999; Wang et al., 2010). Here, we used habitat diversity and elevational range to represent habitat heterogeneity for each grid. Habitat diversity was calculated by applying the Simpson's diversity index on the percentage values of different land cover categories for 2020 (obtained from the GlobalLand30 database; www.globallandcover.com) within each grid (Meynard et al., 2011). Elevational range was calculated as the difference between maximum and minimum elevations in each grid, thus representing topographic heterogeneity (Kerr & Packer, 1997). Elevation data were derived from the International Scientific & Technical Data Mirror Site, Computer Network Information Center, Chinese Academy of Sciences (www.gscloud.cn).

Both human disturbance and productivity can strongly impact assemblage dissimilarity (Sebastián-González et al., 2020). Here, we grouped human disturbance and productivity into a habitat quality variable group, with the assumption that habitat quality is higher under less human disturbance and/or higher productivity. We used the normalized difference vegetation index (NDVI) as a proxy of primary productivity (Ding et al., 2021; Pettorelli et al., 2011), with the associated data downloaded from the Resource and Environment Science and Data Center, Chinese Academy of Sciences (www.resdc.cn) at a spatial resolution of 1 km×1 km. Average NDVI values from July and October 2019 and January and April 2020 were used. The value for each 50 km×50 km grid was calculated as the average of values for all 1 km×1 km grids within it. We used the area of anthropogenic land cover, including farmlands, residential areas, and construction lands, to represent human disturbance (Zhang et al., 2013). Anthropogenic land cover data were extracted from the GlobalLand30 database and calculated using zonal statistics in GIS for each grid.

Beta diversity calculations

Beta diversity was calculated and partitioned following Baselga (2010) and included three dissimilarity matrices: (1) Sørensen's dissimilarity index (β_{sor}), accounting for the total compositional dissimilarity between assemblages; (2) Simpson's dissimilarity index (β_{sim}), capturing only compositional dissimilarity due to species turnover; and (3) nestedness-resultant dissimilarity, representing the difference between β_{sor} and β_{sim} (β_{sne}). These matrices are defined as follows (Baselga, 2010):

$$\beta_{sor} = \frac{b + c}{2a + b + c} \quad (1)$$

$$\beta_{sim} = \frac{\min(b, c)}{a + \min(b, c)} \quad (2)$$

$$\beta_{sne} = \beta_{sor} - \beta_{sim} = \frac{\max(b, c) - \min(b, c)}{2a + b + c} \times \frac{a}{a + \min(b, c)} \quad (3)$$

When these matrices are applied for TBD analysis, *a* represents the number of species shared by the two grids, *b* represents the number of species unique to one grid, and *c* represents the number of species unique to the other grid (Baselga, 2010). However, when applied for PBD analysis, shared and unique species are replaced with shared and unique branch lengths in the phylogenetic tree, respectively (Leprieur et al., 2012). Total beta diversity and its two components were referred to as $\beta_{tax.sor}$, $\beta_{tax.sim}$, and $\beta_{tax.sne}$ for TBD, and were referred to as $\beta_{phy.sor}$, $\beta_{phy.sim}$, and $\beta_{phy.sne}$ for PBD. The matrices were calculated using the "betapart" package in R v,4.0.5 (Baselga & Orme, 2012).

We applied a moving window approach (Peixoto et al., 2017; Pinto-Ledezma et al., 2018; Qian et al., 2020a; Wang et al., 2012) to estimate the geographical patterns of beta diversity. Specifically, for each focal grid, a moving window, i.e., circle with a specific radius, was generated with its center placed at the centroid of the focal grid (Supplementary Figure S4). The above beta diversity matrices were calculated between the focal grid and each surrounding grid within this moving window, with their average values then assigned to the focal grid. Three different moving window radii, i.e., 100 km, 200 km, and 300 km, were used, which produced consistent results. Thus, we only reported results obtained using the 200 km radius.

Statistical analyses

To assess the contribution of turnover and nestedness components to total beta diversity across China, we divided β_{sne} by β_{sor} to yield the β_{ratio} ($\beta_{tax.ratio}$ and $\beta_{phy.ratio}$ for TBD and PBD, respectively) for each grid. A β_{ratio} value greater than 0.5 indicates that total beta diversity is largely driven by nestedness, whereas a value less than 0.5 indicates that total beta diversity is mainly determined by turnover (Pinto-Ledezma et al., 2018).

To assess the spatial relationship between taxonomic and phylogenetic dissimilarity, we calculated the deviation of PBD from TBD (hereafter β_{dev}) for each grid using the formula:

$$\beta_{dev} = 1(\beta_{phy}/\beta_{tax}) \quad (4)$$

where a greater value of β_{dev} indicates less variation in phylogenetic composition with respect to variations in taxonomic composition between sites, and vice versa (Peixoto et al., 2017; Qian et al., 2020a).

Furthermore, to assess the effects of different ecological factors on the beta diversity of breeding birds, we conducted variance partitioning analysis (VPA) for each grid based on a 200 km radius moving window. In the VPA, the beta diversity matrices between the focal grid and each surrounding grid within the window served as response variables, while spatial and environmental distances between the corresponding grid pair served as predictors. Spatial distance was calculated as the Euclidean distance of each pair of grid centers in a three-dimensional (i.e., latitude, longitude, and altitude) geographic space. Environmental distance was calculated as the Euclidean distance of each pair of grids in a multi-dimensional environmental space with environmental variables as coordinates, and was calculated for the three groups of environmental variables (climate, habitat heterogeneity, and habitat quality), respectively. The VPA results were averaged for all grids across China to yield a universal assessment of the effects of different ecological factors on the beta diversity of breeding birds. Additionally, to explore spatial variation in the relative importance of different ecological factors across China, the predictor with the highest adjusted R^2 value based on VPA for each grid was assigned as the primary determinant. VPA was conducted using the "vegan" package in R v.4.0.5 (Dixon, 2003).

To assess the differences in beta diversity and its determinants among regions with distinct geographical backgrounds, we divided grids into eastern and western regions along the Hu Line, resulting in 1 521 and 1 965 grids, respectively. Subsequently, we counted the proportion of grids dominated by turnover and nestedness components and the proportion of grids dominated by different ecological factors on either side of the line.

To estimate the influence of species dispersal ability on beta diversity and its determinants, we calculated a dispersal ratio for each species by dividing its mean wing length (mm) by the cube root of its mean mass (g), adopted as a proxy for dispersal ability in birds according to previous studies (Fischer & Lindenmayer, 2005; Woinarski, 1989; Zhang et al., 2020). The mass and wing length data were derived based on life history and ecological traits of Chinese birds in Wang et al. (2021). Subsequently, all species were sorted in ascending order based on their dispersal ratio values and trisected into weak, medium, and strong dispersal groups. We repeated beta diversity calculations and VPA for each dispersal group and performed Wilcoxon signed-rank tests to determine any significant differences in beta diversity and its determinants between different groups (Supplementary Figure S5). In addition, to compare the contributions of different dispersal groups to overall patterns, we used the simultaneous autoregressive model to correlate the beta diversity of each group with that of all species, while accounting for spatial autocorrelation (Wang et al., 2012). The Wilcoxon signed-rank test and simultaneous autoregressive model were conducted using the "stats" and "spatialreg" packages in R v.4.0.5.

RESULTS

The spatial patterns of TBD and PBD for breeding birds were highly consistent across China (Figure 2). Areas with high total TBD and PBD, along with their respective turnover and nestedness components, were primarily observed in the mountainous regions, including the Tianshan, Kunlun, Aerjin, Himalaya, Great Khingan, Taihang, and Hengduan mountains (Figure 2).

Contrary to expectations, the turnover component exhibited a much higher contribution than the nestedness component in the eastern regions of China, with $\beta_{tax.ratio}<0.5$ and $\beta_{phy.ratio}<0.5$ grids accounting for 89.2% and 83.8% of the total, respectively. Conversely, nestedness exhibited a higher contribution in the western regions, with $\beta_{tax.ratio}>0.5$ and $\beta_{phy.ratio}>0.5$ grids accounting for 60.8% and 64.2% of the total, respectively (Figure 3; Supplementary Figure S6). Furthermore, across both the eastern and western regions, PBD was markedly lower than the values predicted by TBD, as evidenced by β_{dev} exceeding 0.3 (Figure 2). Similar trends were observed for their turnover and nestedness components.

The VPA results revealed that for both TBD and PBD and their respective turnover and nestedness components, environmental distance alone explained the variations more clearly than spatial distance alone, with independent explained variation range of 0.29−0.37 and 0.03−0.05 and a shared explained variation range of 0.08−0.18 for environmental distance and spatial distance, respectively, (Supplementary Figure S7). In terms of environmental distance, climate and habitat heterogeneity were the predominant explanatory factors in the eastern (Figure 4A) and western (Figure 4B) regions of China, respectively. Comparatively, habitat quality dominated in only a few grids across China. Notably, grids with higher spatial distance explanatory power were mostly located in mountainous regions, such as the Kunlun and Taihang mountains. These trends were observed for both TBD and PBD.

As predicted, species with weaker dispersal abilities exhibited significantly higher levels of TBD, PBD, and their respective components ($P<0.01$), whereas those with stronger dispersal abilities showed significantly lower levels (Figure 5). Moreover, the independent explanatory power of spatial distance increased significantly with decreasing species dispersal ability, while the independent explanatory power of environmental distance and shared explanatory power of spatial and environmental distances decreased significantly with decreasing species dispersal ability (Figure 6). Furthermore, TBD, PBD, and their respective components, when considered across all species, demonstrated the strongest correlations with species possessing strong dispersal capabilities and the weakest correlations with species exhibiting weak dispersal abilities (Supplementary Table S1).

DISCUSSION

General patterns of beta diversity for breeding bird assemblages in China

Elevated TBD and PBD values were identified in mountainous areas, consistent with global trends observed across a wide range of taxa (Melo et al., 2009; Peixoto et al., 2017; Wang et al., 2012). Mountainous areas are typically characterized by strong geographic barriers and environmental gradients, which likely contribute to increased assemblage dissimilarity compared to plains (Graham et al., 2009). Furthermore, China is partitioned into various biogeographical regions by broad

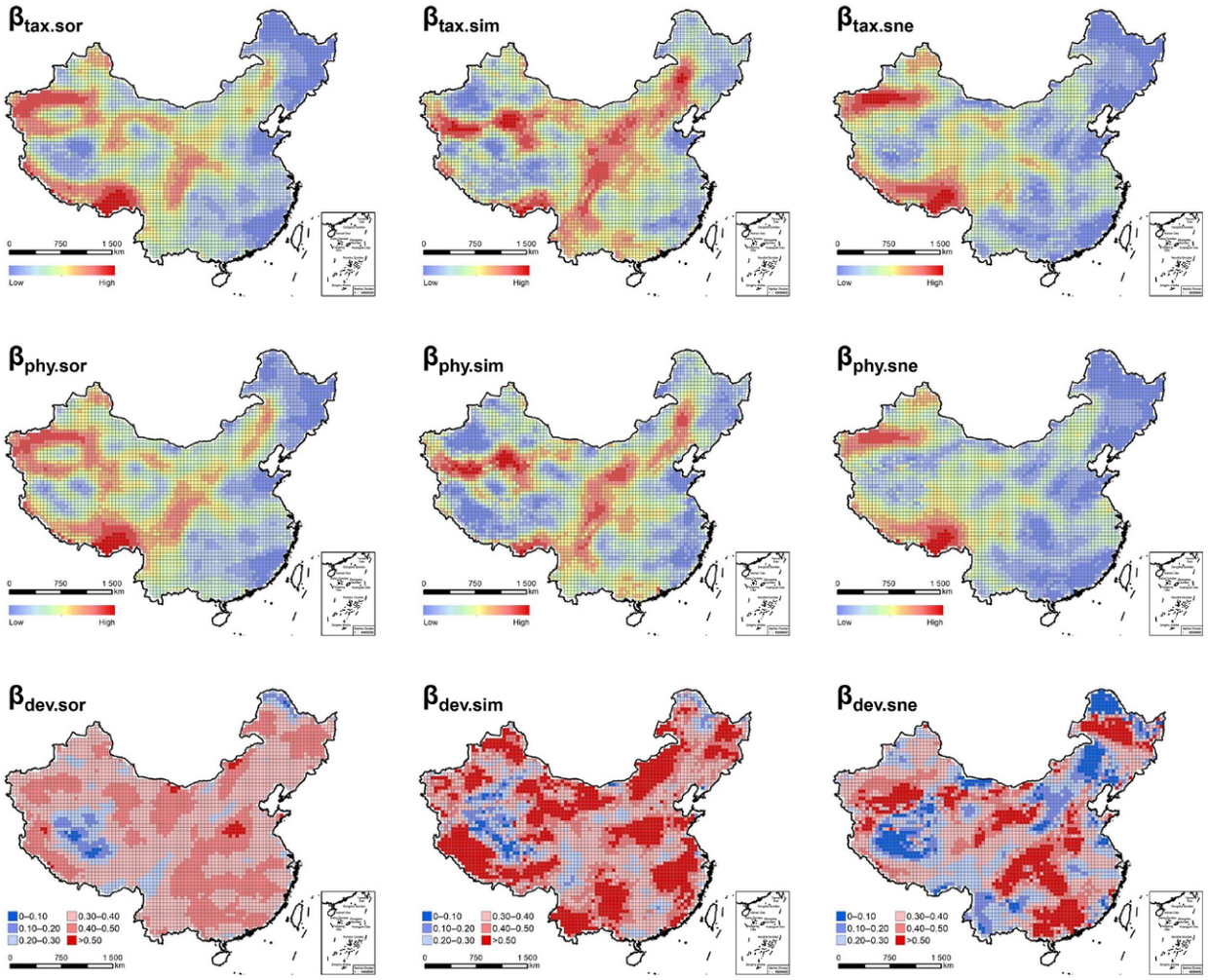


Figure 2 Patterns of taxonomic beta diversity (β_{tax}), phylogenetic beta diversity (β_{phy}), and deviations of β_{phy} from β_{tax} (β_{dev}, calculated as $\beta_{dev}=1-(\beta_{phy}/\beta_{tax})$ for breeding birds in mainland China

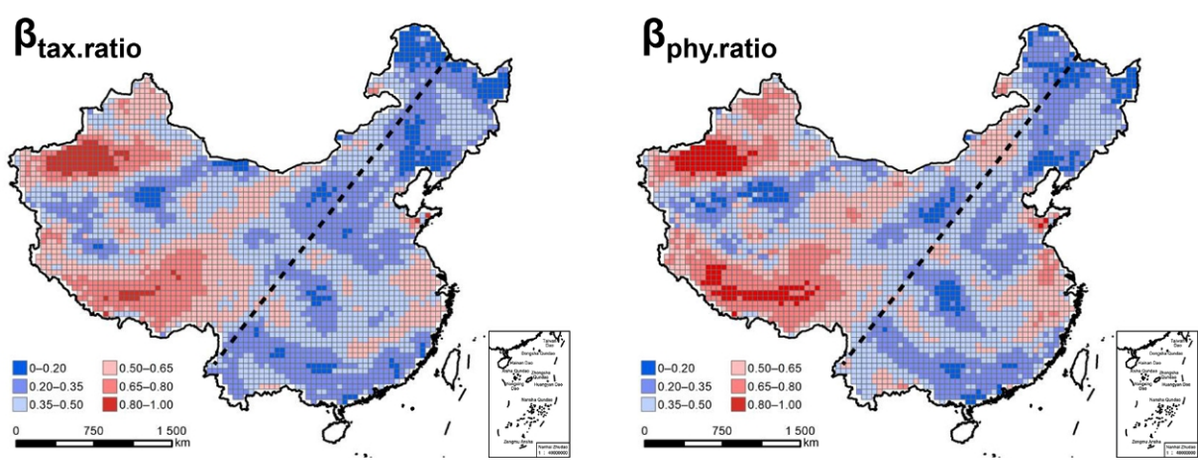


Figure 3 Contribution patterns of turnover and nestedness components to total beta diversity of taxonomic and phylogenetic beta diversity ($\beta_{tax.ratio}$ and $\beta_{phy.ratio}$) for breeding birds in mainland China

β_{ratio} was determined as the ratio of β_{sne} to β_{sor}. Black dashed line represents Hu Line.

mountain ranges (He et al., 2017). These mountain ranges act as ecotones between disparate faunal habitats, thereby facilitating in intense biotic turnover (Lomolino, 2001). For instance, high TBD and PBD turnover rates were observed in the Great Khingan, Taihang, Hengduan mountain ranges, which generally run along the Hu Line. Moreover, certain mountain ranges, such as Himalaya and Tianshan mountains in western China, host multiple endemic species, and thus serve as regional or even global biodiversity hotspots (Wang et al., 2016). Such differences in species richness between mountains and their surrounding habitats may drive the dominance of the nestedness component. These findings

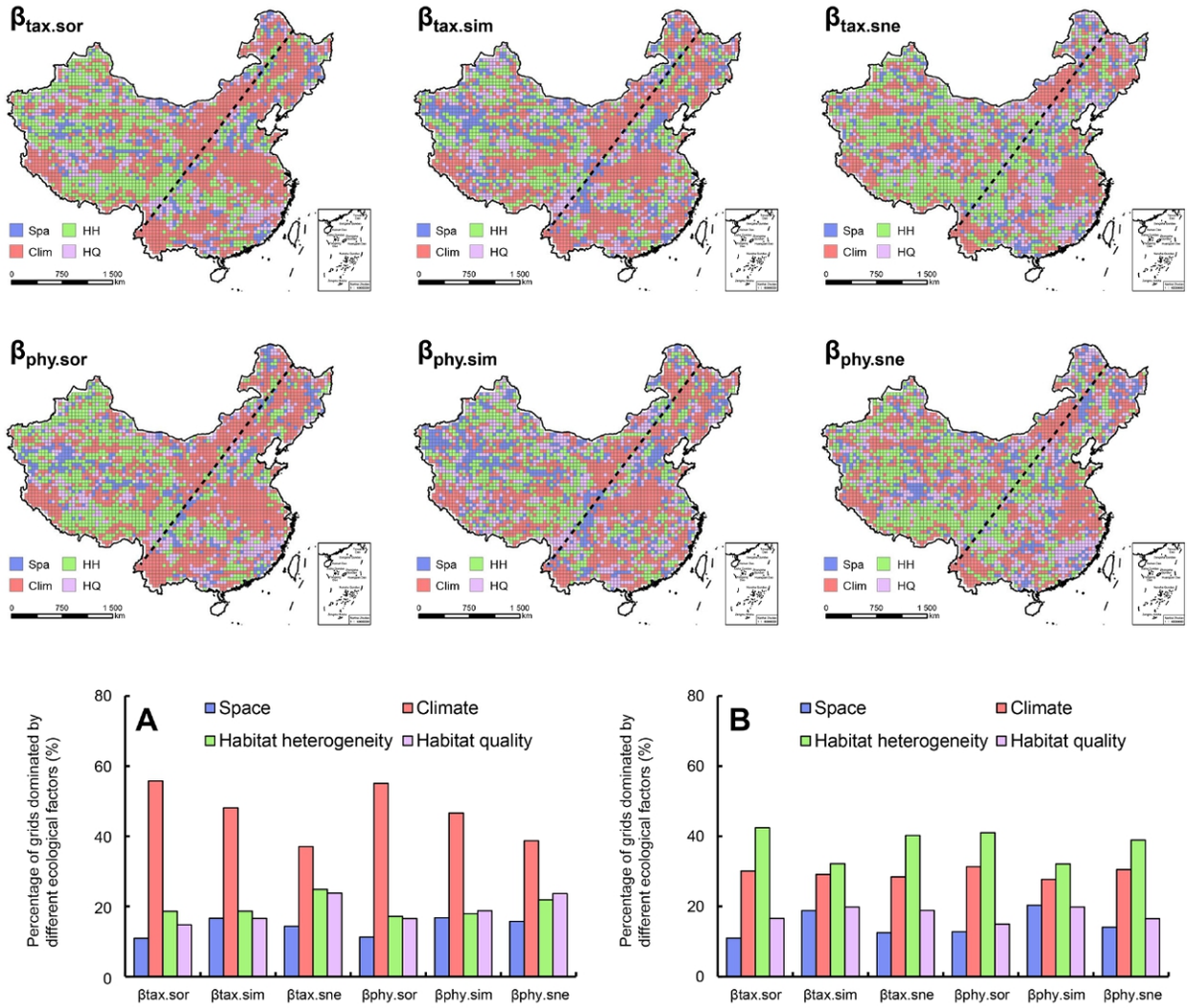


Figure 4 Dominant explanatory factor patterns for breeding birds for taxonomic and phylogenetic beta diversity ($\beta_{tax.sor}$ and $\beta_{phy.sor}$, respectively) and their respective turnover ($\beta_{tax.sim}$ and $\beta_{phy.sim}$) and nestedness ($\beta_{tax.sne}$ and $\beta_{phy.sne}$) components, and proportion of grids dominated by different ecological factors in eastern (a) and western (b) regions of mainland China

Spa: Spatial distance; Clim: Climate distance; HH: Habitat heterogeneity distance; HQ: Habitat quality distance. The black dashed line represents the Hu Line.

suggest that mountain ranges are hotspots not only for alpha diversity but also for beta diversity, thus warranting urgent consideration in the planning and design of large-scale conservation initiatives.

Although the patterns of TBD and PBD were highly consistent, the latter was considerably lower than the former for most regions, irrespective of their geographical location in the eastern or western regions on either side of the Hu Line. These findings suggest that compositional variations in breeding bird assemblages may be more likely to occur in phylogenetically proximate species rather than distant ones. Considering that both TBD and PBD exhibited strong associations with environmental distances, our results support the notion that environmental conditions, whether related to climate or habitat heterogeneity, may selectively favor species with certain niches, leading to phylogenetically homogeneous assemblages according to the phylogenetic niche conservatism hypothesis (Jin et al., 2015).

Contrasting origins and drivers of beta diversity on either side of the Hu Line

Environmental distance predominantly accounted for the patterns in both TBD and PBD, as well as their respective turnover and nestedness components, signifying the crucial role of environmental filtering in shaping breeding bird assemblages in China. However, when geographical variations in ecological factors were considered, different dominant environmental filters and beta diversity components were observed on either side of the Hu Line. In the eastern regions, climate and turnover emerged as the primary drivers and components for both TBD and PBD, suggesting that compositional variations in breeding bird assemblages in these regions are largely driven by species turnover along climatic gradients. This observation contradicts prior assumptions that regions experiencing high levels of human disturbance would display lower assemblage turnover and higher nestedness (Wayman et al., 2022). Conversely, in the western regions, where anthropogenic influence is comparatively weaker, habitat heterogeneity and nestedness became the principal determinants and components of beta diversity, although climate and turnover still showed significant influence. These findings imply that in these regions, sites with fewer habitat types contain only a subset of species present in

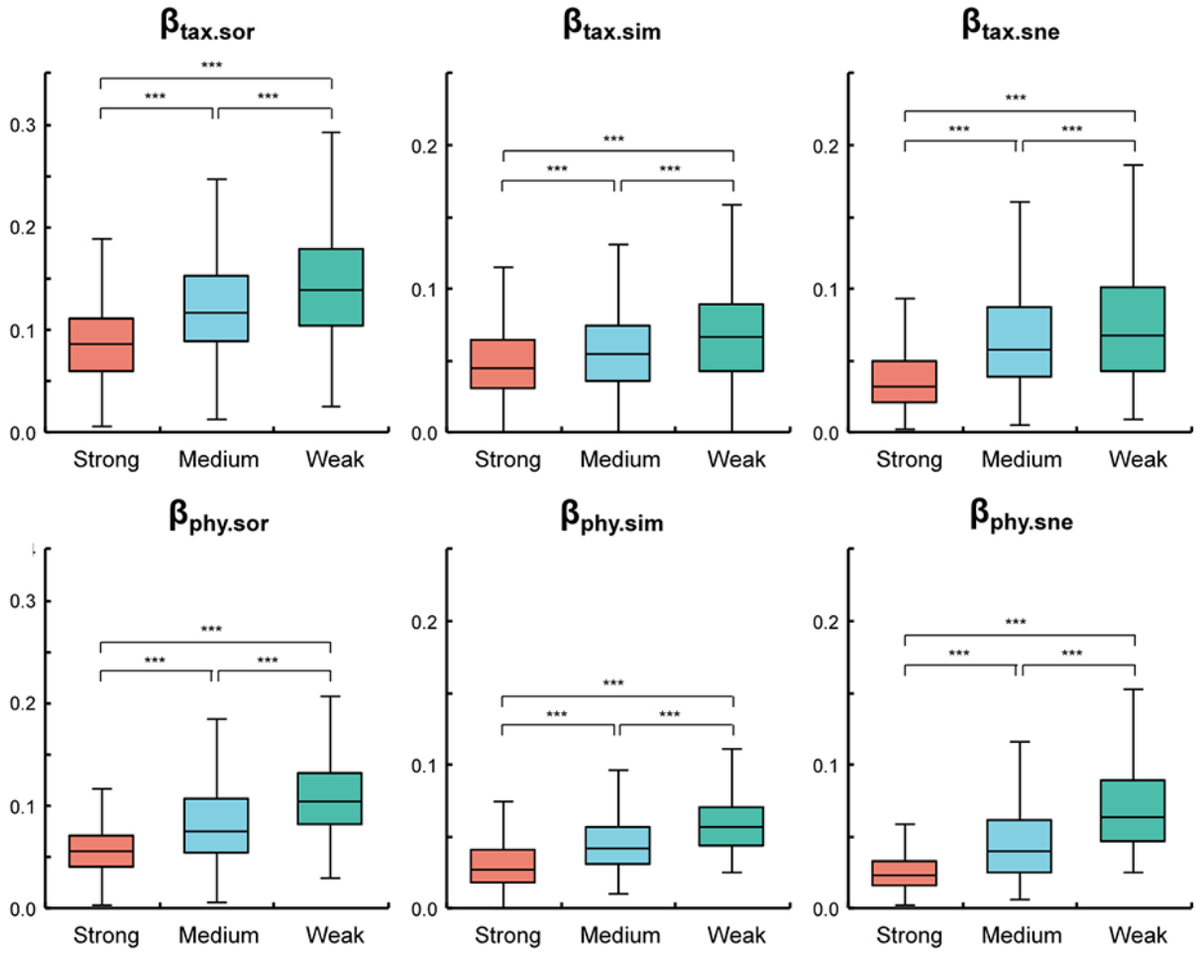


Figure 5 Taxonomic and phylogenetic beta diversity and their components for species with strong, medium, and weak dispersal ability
Significant differences were determined using Wilcoxon signed-rank tests. ***: P<0.001.

sites with greater habitat diversity, leading to a pattern of nestedness in species community composition, as posited by the habitat nestedness hypothesis (Calmé & Desrochers, 1999; Honnay et al., 1999). The harsh environmental conditions present in the western region likely limits the range of available habitat types, thereby promoting a nestedness pattern of species distribution. Moreover, the western region bird assemblages were characterized by species with robust dispersal capabilities (Supplementary Figure S8), which are more adept at adapting to variations in habitat, thereby facilitating nestedness (Cook & Quinn, 1995; de Siqueira Neves et al., 2021).

Our findings broadly illustrated that regions with different geographical backgrounds are subject to different ecological factors, which may lead to divergent patterns of assemblage dissimilarity. Thus, emphasizing a single explanatory factor for the observed patterns (such as climate, in our case) could result in the oversight of other crucial ecological determinants that are closely associated with certain regional properties (such as habitat heterogeneity, in our case). Such bias may not only hinder a comprehensive understanding of the mechanisms underlying large-scale community assemblages, but also misdirect the development of appropriate conservation and management strategies, given that different ecological processes may have divergent management and conservation implications (Angeler, 2013). For instance, our results suggest that prioritizing sites with high habitat heterogeneity could ensure biodiversity conservation in western China, where nestedness is the dominant pattern. Conversely, in eastern regions where turnover is predominant, focusing on multiple sites with diverse climatic environments may be more important, albeit administratively and financially challenging (Socolar et al., 2016).

Effects of species dispersal ability

In addition to geographical background, species dispersal ability also played a profound role in shaping the beta diversity of breeding bird assemblages in China. Generally, species with weak dispersal ability exhibited higher beta diversity, consistent with the findings of Wang et al. (2012), who observed higher beta diversity for woody plants with restricted range sizes within China. Species with weak dispersal ability tend to have narrower geographical ranges compared to those with strong dispersal ability (Nekola & White, 1999; Stevens, 1989). Consequently, they may be less likely to occur uniformly across different assemblages, leading to more dissimilar species composition among assemblages. Nevertheless, our analysis revealed that overall beta diversity (all species combined) was largely driven by species with strong dispersal ability. These results are congruent with previous observations on global turnover patterns in bird assemblages (Buckley & Jetz, 2008) and analogous with the previous finding that species richness patterns are mainly determined by species with wider ranges (Kreft et al. 2006; Lennon et al. 2004). Such evidence implies that loss of

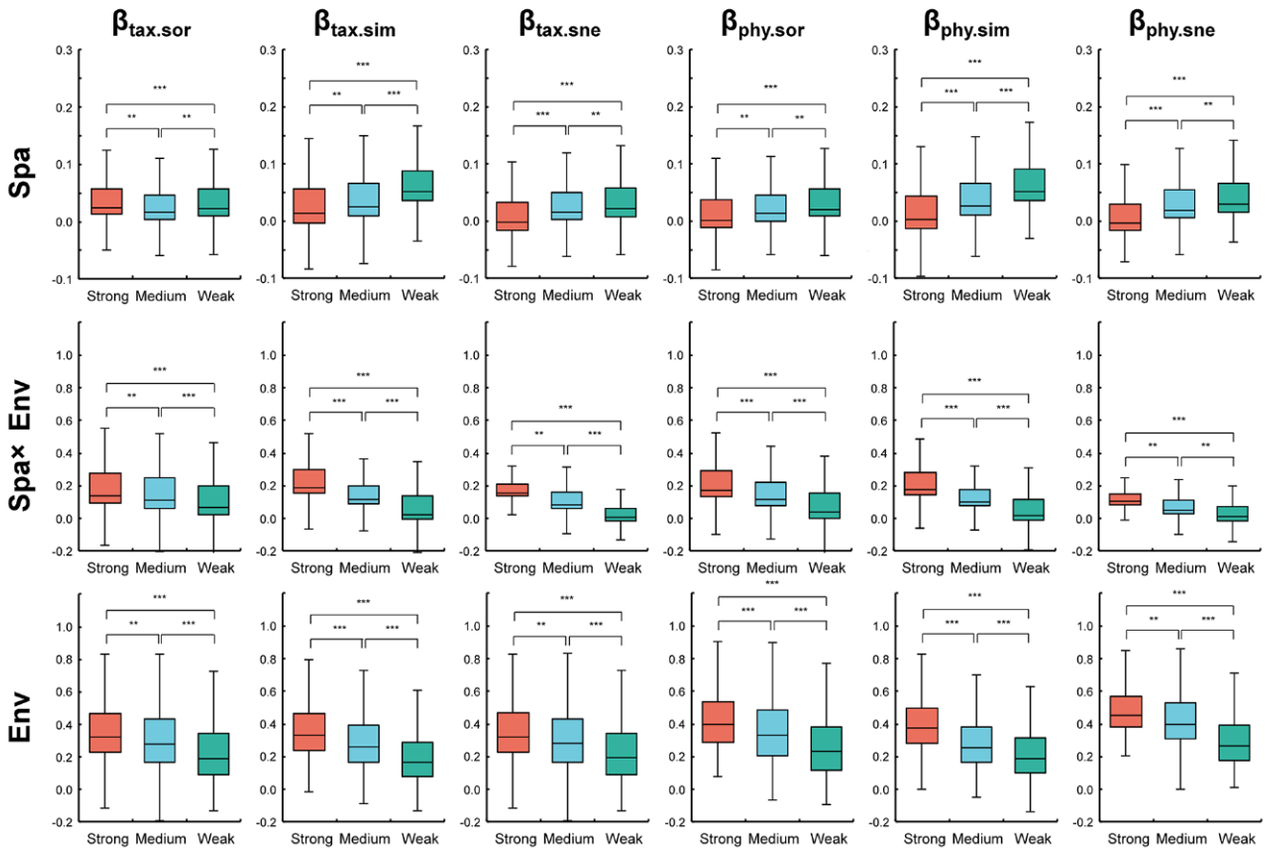


Figure 6 Independent and shared explanatory powers of spatial and environmental distances for taxonomic and phylogenetic beta diversity of species with strong, medium, and weak dispersal ability

Significant differences were determined using Wilcoxon signed-rank tests. **: P<0.01; ***: P<0.001.

species with strong dispersal ability could have profound impact on both alpha and beta diversity, and they should be considered equally important for biodiversity conservation as those with weak dispersal ability, although the latter generally receive greater attention due to their rarity.

The influence of dispersal ability on beta diversity has frequently been attributed to the differential responses of species to different ecological factors (Linares-Palomino & Kessler, 2009; Qian, 2009). It is widely posited that species with strong dispersal ability can access and inhabit most, if not all, environmentally suitable areas, thereby forming more homogeneous assemblages. In contrast, species with weak dispersal ability are restricted to fewer suitable sites, thereby forming more heterogeneous assemblages (Nekola & White, 1999; Wang et al., 2012). In the present study, we found that spatial factors held greater importance for bird species with weaker dispersal abilities. Conversely, both environmental factors alone and the effects of spatial and environmental factors combined were significantly more important for bird species with strong dispersal ability. Given that environmental variables at large scales often exhibit spatial structure, these combined effects are typically considered indirect environmental influences (Mazel et al., 2017; Qian et al., 2020a). Therefore, these results provide strong evidence for the aforementioned hypotheses. Conservation strategies targeting species with strong dispersal ability should focus on maintaining the suitability of available habitats through various measures, such as closely monitoring the impacts of global change. In contrast, conservation efforts targeting species with weak dispersal ability should prioritize the availability of suitable habitats by undertaking measures such as the establishment of ecological corridors.

CONCLUSIONS

Although geographical background and species attributes are important for understanding the mechanisms driving beta diversity patterns, they are frequently neglected in large-scale beta diversity studies. Here, we explored the spatial patterns of beta diversity and its determinants across different geographical regions in mainland China, focusing on breeding birds with different dispersal capabilities. Different components and drivers of beta diversity were prevalent on the two sides of the Hu Line. Moreover, species exhibiting stronger dispersal ability were more susceptible to environmental factors and formed more homogeneous assemblages. These findings underscore the importance of considering both geographical background variation and species dispersal abilities to achieve a comprehensive understanding of the mechanisms underlying large-scale beta diversity. Future research should rigorously examine the potential interactive effects of these factors to provide a suitable reference for implementing conservation strategies tailored to specific local conditions.

DATA AVAILABILITY

The raw data for this work are available at the Dryad Digital Repository (https://datadryad.org/stash/share/c7HFZCksSI6hQLDsaTVqFJvNjTHMuZRC6DDXXDTEhKk).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHOR'S CONTRIBUTION

J.C.L. conceived and designed the study with the help of Z.F.D.; J.C.L., G.W.L., X.N.N., and W.B.H. collected and compiled the data; J.C.L., Y.M.H., and Z.X.Z. analyzed the data; J.C.L. led the writing, which Z.F.D. revised; H.J.H., X.F.S., and C.L.L reviewed the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Alahuhta J, Kosten S, Akasaka M, et al. 2017. Global variation in the beta diversity of lake macrophytes is driven by environmental heterogeneity rather than latitude. *Journal of Biogeography*, **44**(8): 1758−1769.

Angeler DG. 2013. Revealing a conservation challenge through partitioned long-term beta diversity: increasing turnover and decreasing nestedness of boreal lake metacommunities. *Diversity and Distributions*, **19**(7): 772−781.

Baselga A. 2010. Partitioning the turnover and nestedness components of beta diversity. *Global Ecology and Biogeography*, **19**(1): 134−143.

Baselga A. 2012. The relationship between species replacement, dissimilarity derived from nestedness, and nestedness. *Global Ecology and Biogeography*, **21**(12): 1223−1232.

Baselga A, Orme CDL. 2012. betapart: an R package for the study of beta diversity. *Methods in Ecology and Evolution*, **3**(5): 808−812.

Buckley LB, Jetz W. 2008. Linking global turnover of species and environments. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(46): 17836−17841.

Calmé S, Desrochers A. 1999. Nested bird and micro-habitat assemblages in a peatland archipelago. *Oecologia*, **118**(3): 361−370.

Carvalho JC, Cardoso P, Gomes P. 2012. Determining the relative roles of species replacement and species richness differences in generating beta-diversity patterns. *Global Ecology and Biogeography*, **21**(7): 760−771.

Chase JM. 2014. Spatial scale resolves the niche versus neutral theory debate. *Journal of Vegetation Science*, **25**(2): 319−322.

Cook RR, Quinn JF. 1995. The influence of colonization in nested species subsets. *Oecologia*, **102**(4): 413−424.

de Siqueira Neves F, da Silva PG, Solar R, et al. 2021. Habitat generalists drive nestedness in a tropical mountaintop insect metacommunity. *Biological Journal of the Linnean Society*, **133**(2): 577−586.

Ding ZF, Hu HJ, Cadotte MW, et al. 2021. Elevational patterns of bird functional and phylogenetic structure in the central Himalaya. *Ecography*, **44**(9): 1403−1417.

Dixon P. 2003. VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**(6): 927−930.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**(1): 214.

Fischer J, Lindenmayer DB. 2005. Nestedness in fragmented landscapes: a case study on birds, arboreal marsupials and lizards. *Journal of Biogeography*, **32**(10): 1737−1750.

García-Girón J, Heino J, Baastrup-Spohr L, et al. 2020. Global patterns and determinants of lake macrophyte taxonomic, functional and phylogenetic beta diversity. *Science of the Total Environment*, **723**: 138021.

Graham CH, Fine PVA. 2008. Phylogenetic beta diversity: linking ecological and evolutionary processes across space in time. *Ecology Letters*, **11**(12): 1265−1277.

Graham CH, Parra JL, Rahbek C, et al. 2009. Phylogenetic structure in tropical hummingbird communities. *Proceedings of the National Academy of Sciences of the United States of America*, **106**(suppl.2): 19673−19678.

He JK, Kreft H, Gao EH, et al. 2017. Patterns and drivers of zoogeographical regions of terrestrial vertebrates in China. *Journal of Biogeography*, **44**(5): 1172−1184.

Hill MJ, White JC, Biggs J, et al. 2021. Local contributions to beta diversity in urban pond networks: Implications for biodiversity conservation and management. *Diversity and Distributions*, **27**(5): 887−900.

Honnay O, Hermy M, Coppin P. 1999. Nested plant communities in deciduous forest fragments: species relaxation or nested habitats?. *Oikos*, **84**(1): 119−129.

Hu HY. 1935. The Distribution of population in China, with statistics and maps. *Acta Geographica Sinica*, **2**(2): 33−74.

Jetz W, Thomas GH, Joy JB, et al. 2012. The global diversity of birds in space and time. *Nature*, **491**(7424): 444−448.

Jiao S, Chen WM, Wei GH. 2021. Linking phylogenetic niche conservatism to soil archaeal biogeography, community assembly and species coexistence. *Global Ecology and Biogeography*, **30**(7): 1488−1501.

Jin LS, Cadotte MW, Fortin MJ. 2015. Phylogenetic turnover patterns consistent with niche conservatism in montane plant species. *Journal of Ecology*, **103**(3): 742−749.

Kerr JT, Packer L. 1997. Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. *Nature*, **385**(6613): 252−254.

Kong XS, Fu MX, Zhao X, et al. 2022. Ecological effects of land-use change on two sides of the Hu Huanyong Line in China. *Land Use Policy*, **113**: 105895.

König C, Weigelt P, Kreft H. 2017. Dissecting global turnover in vascular plants. *Global Ecology and Biogeography*, **26**(2): 228−242.

Kreft H, Jetz W. 2010. A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, **37**(11): 2029−2053.

Lennon JJ, Koleff P, Greenwood JJD, et al. 2004. Contribution of rarity and commonness to patterns of species richness. *Ecology Letters*, **7**(2): 81−87.

Leprieur F, Albouy C, De Bortoli J, et al. 2012. Quantifying phylogenetic beta diversity: distinguishing between 'True' turnover of lineages and phylogenetic diversity gradients. *PLoS One*, **7**(8): e42760.

Liang CX, Yang GS, Wang N, et al. 2019. Taxonomic, phylogenetic and functional homogenization of bird communities due to land use change. *Biological Conservation*, **236**: 37−43.

Linares-Palomino R, Kessler M. 2009. The role of dispersal ability, climate and spatial separation in shaping biogeographical patterns of phylogenetically distant plant groups in seasonally dry Andean forests of Bolivia. *Journal of Biogeography*, **36**(2): 280−290.

Lomolino MV. 2001. Elevation gradients of species-density historical and prospective views. *Global Ecology and Biogeography*, **10**(1): 3−13.

MacKinnon J, Phillipps K, He FQ. 2003. A Field Guide to the Birds of China. Hunan Education Publishing House.

Mazel F, Wüest RO, Lessard JP, et al. 2017. Global patterns of β-diversity along the phylogenetic time-scale: the role of climate and plate tectonics. *Global Ecology and Biogeography*, **26**(10): 1211−1221.

Melo AS, Fernando T, Rangel LVB, et al. 2009. Environmental drivers of beta-diversity patterns in New-World birds and mammals. *Ecography*, **32**(2): 226−236.

Meynard CN, Devictor V, Mouillot D, et al. 2011. Beyond taxonomic diversity patterns: how do α, β and γ components of bird functional and phylogenetic diversity respond to environmental gradients across France?. *Global Ecology and Biogeography*, **20**(6): 893−903.

Nekola JC, White PS. 1999. The distance decay of similarity in biogeography and ecology. *Journal of Biogeography*, **26**(4): 867−878.

Peixoto FP, Villalobos F, Melo AS, et al. 2017. Geographical patterns of phylogenetic beta-diversity components in terrestrial mammals. *Global Ecology and Biogeography*, **26**(5): 573−583.

Pettorelli N, Ryan S, Mueller T, et al. 2011. The Normalized Difference Vegetation Index (NDVI): unforeseen successes in animal ecology. *Climate Research*, **46**(1): 15−27.

Pinto-Ledezma JN, Larkin DJ, Cavender-Bares J. 2018. Patterns of beta

diversity of vascular plants and their correspondence with biome boundaries across North America. *Frontiers in Ecology and Evolution*, **6**: 194.

Qian H. 2009. Beta diversity in relation to dispersal ability for vascular plants in North America. *Global Ecology and Biogeography*, **18**(3): 327−332.

Qian H, Jin Y, Leprieur F, et al. 2020a. Geographic patterns and environmental correlates of taxonomic and phylogenetic beta diversity for large-scale angiosperm assemblages in China. *Ecography*, **43**(11): 1706−1716.

Qian H, Jin Y, Leprieur F, et al. 2020b. Patterns of phylogenetic beta diversity measured at deep evolutionary histories across geographical and ecological spaces for angiosperms in China. *Journal of Biogeography*, **48**(4): 773−784.

Sebastián-González E, Morales-Reyes Z, Botella F, et al. 2020. Network structure of vertebrate scavenger assemblages at the global scale: drivers and ecosystem functioning implications. *Ecography*, **43**(8): 1143−1155.

Si XF, Cadotte MW, Zeng D, et al. 2017. Functional and phylogenetic structure of island bird communities. *Journal of Animal Ecology*, **86**(3): 532−542.

Socolar JB, Gilroy JJ, Kunin WE, et al. 2016. How should Beta-Diversity inform biodiversity conservation?. *Trends in Ecology & Evolution*, **31**(1): 67−80.

Stevens GC. 1989. The latitudinal gradient in geographical range: how so many species coexist in the tropics. *The American Naturalist*, **133**(2): 240−256.

Stuart YE, Losos JB, Algar AC. 2012. The island–mainland species turnover relationship. *Proceedings of the Royal Society B:Biological Sciences*, **279**(1744): 4071−4077.

Wang Q, Zhang ML, Yin LK. 2016. Phylogeographic structure of a Tethyan relict *Capparis spinosa* (Capparaceae) traces pleistocene geologic and climatic changes in the western himalayas, Tianshan mountains, and adjacent desert regions. *BioMed Research International*, **2016**: 5792708.

Wang YP, Bao YX, Yu MJ, et al. 2010. BIODIVERSITY RESEARCH: Nestedness for different reasons: the distributions of birds, lizards and small mammals on islands of an inundated lake. *Diversity and Distributions*, **16**(5): 862−873.

Wang YP, Song YF, Zhong YX, et al. 2021. A dataset on the life-history and ecological traits of Chinese birds. *Biodiversity Science*, **29**(9): 1149−1153.

Wang ZH, Fang JY, Tang ZY, et al. 2012. Geographical patterns in the beta diversity of China's woody plants: the influence of space, environment and range size. *Ecography*, **35**(12): 1092−1102.

Wayman JP, Sadler JP, Pugh TAM, et al. 2022. Identifying the drivers of spatial taxonomic and functional beta-diversity of British breeding birds. *Frontiers in Ecology and Evolution*, **9**: 620062.

Weideman EA, Slingsby JA, Thomson RL, et al. 2020. Land cover change homogenizes functional and phylogenetic diversity within and among African savanna bird assemblages. *Landscape Ecology*, **35**(1): 145−157.

Wiescher PT, Pearce-Duvet JMC, Feener DH. 2012. Assembling an ant community: species functional traits reflect environmental filtering. *Oecologia*, **169**(4): 1063−1074.

Woinarski JCZ. 1989. Some life history comparisons of small leaf-gleaning bird species of south-eastern Australia. *Corella*, **13**(3): 73−80.

Xu C, Huang ZYX, Chi T, et al. 2014. Can local landscape attributes explain species richness patterns at macroecological scales?. *Global Ecology and Biogeography*, **23**(4): 436−445.

Yao H, Zhang L. 2020. Spatial pattern of talents and the evolution of the situation of Chinese "Heihe-Tengchong Line" under the background of big data. *In*: International Conference on Big Data Analytics for Cyber-Physical-Systems. Shenyang: Springer, 271–281.

Yue TX, Wang YA, Chen SP, et al. 2003. Numerical simulation of population distribution in China. *Population and Environment*, **25**(2): 141−163.

Zellweger F, Roth T, Bugmann H, et al. 2017. Beta diversity of plants, birds and butterflies is closely associated with climate and habitat structure. *Global Ecology and Biogeography*, **26**(8): 898−906.

Zhang J, Kissling WD, He FL. 2013. Local forest structure, climate and human disturbance determine regional distribution of boreal bird species richness in Alberta, Canada. *Journal of Biogeography*, **40**(6): 1131−1142.

Zhang L, Sun PS, Huettmann F, et al. 2022. Where should China practice forestry in a warming world?. *Global Change Biology*, **28**(7): 2461−2475.

Zhang Q, Holyoak M, Chen CW, et al. 2020. Trait-mediated filtering drives contrasting patterns of species richness and functional diversity across montane bird assemblages. *Journal of Biogeography*, **47**(1): 301−312.

Zheng GM. 2017. A Checklist on the Classification and Distribution of the Birds of China. 3rd ed. Beijing: Science Press.

Zhang et al. *Zool. Res.* 2024, 45(1): 136−137
https://doi.org/10.24272/j.issn.2095-8137.2023.378

ZR | Zoological Research

Editorial Open Access

Impact of letters to the editor and publications in 2023

A letter to the editor constitutes a short communication addressing a range of topics pertinent to the readership of a journal (Dkhar, 2018). This format offers several benefits, such as timeliness, accessibility, innovation, and conciseness, thereby serving as an effective means to disseminate cutting-edge scientific ideas. Over the past five years, there has been a considerable increase in the number of letters published, representing the highest growth rate (approximately 20%) observed in the last three decades (Figure 1A). In the field of academic publishing, letters to the editor are typically more concise than typical research papers. However, there is no prescribed standard for their length, which varies substantially across different journals. An analysis of letters published in 2021, based on the data from the Web of Science, indicates that most letters were one to nine pages long (Figure 1B), with the longest letter extending to 72 pages. Similarly, the number of references in these letters ranged from zero to 30 citations (Figure 1C), with the highest number of citations reaching 206.

Zoological Research (*ZR*), paralleling other scientific journals, publishes a relatively high proportion of letters to the editor, constituting approximately 40% of its total publications in 2021 and 2022. This publishing strategy is a response to the consistent increase in manuscript submissions in recent years, with the aim to efficiently disseminate significant yet concise findings through letters to the editor. In addition, by increasing the number of letters to the editor, a journal may elevate its impact factor. This effect occurs because these letters, while contributing to the total citation count, are not included in the denominator for impact factor calculation, which is restricted to review and research articles.

In 2022, despite strict length criteria for letters to the editor published in *ZR*, generally limited to no more than 15 references and five print pages, a considerable number of letters were erroneously reclassified as articles by the Web of Science, despite their correct categorization at the time of indexing. Efforts to appeal to the Web of Science for rectification of these classifications were unsuccessful. This incorrect re-categorization of letters as articles resulted in a sharp decrease in the impact factor of *ZR*, which was released in 2023. Thus, to prevent similar occurrences in the future, the journal has resolved to reduce the number of letters it publishes. It is important to emphasize that *ZR* remains committed to the publication of original and influential research (Yao & Zhang, 2022).

We would like to take this opportunity to express our heartfelt appreciation to all supporters of *ZR*. The editorial team encourages continued submission of high-quality manuscripts, as your contributions are the driving force behind the advancement of our journal. In 2023, *ZR* was honored to publish many prominent research articles. For instance, Liu et al. (2023) provided evidence supporting a scenario of placental radiation characterized by both basal cladogenesis and active interordinal divergences spanning the Late Cretaceous to the Paleogene. Lyu et al. (2023) proposed a new 10-genus classification for the Asian horned toad subfamily Megophryinae based on phylogenetic relationships and morphological differences and further reviewed the taxonomic history of Asian horned toad species in China. Sitthivong et al. (2023) described a new microendemic genus and species of a limestone-dwelling agamid lizard (Squamata: Agamidae: Draconinae) from Laos, which was read more than 16 000 times within three weeks of its online publication, representing the most viewed *ZR* paper published in 2023. Qiu et al. (2023) identified and described eight subspecies of mainland *Apis cerana*, providing insights into their evolutionary history and subspecies boundaries to enable customized conservation strategies for both widespread and endemic honeybee populations, thereby informing colony introduction and breeding strategies.

Many excellent papers related to animal models and disease mechanisms were also published in 2023. For instance, Nadal-Nicolás et al. (2023) thoroughly reviewed and analyzed the specificity of nine markers for the unambiguous identification of all or most RGCs (i.e., pan-RGC markers), a crucial prerequisite for studying their degeneration and neuroprotection, across rats, mice, and macaques. Liu et al. (2023) revealed that monkeys show heightened sensitivity to the facial expressions of conspecifics compared to those of humans, offering new insights into inter-species communication. Li et al. (2023) presented a comprehensive analysis of the genetic and physiological attributes critical for the sustained expansion of undifferentiated spermatogonial stem cells in tree shrews and proposed an effective strategy for extended *in vitro* maintenance. Guo et al. (2023) generated a single-nucleus transcriptomic atlas of all major liver cell types, mapping the multicellular networks that regulate mild to severe mammalian liver fibrosis, at different stages of murine carbon tetrachloride (CCl4)-induced progressive liver fibrosis. Wang et al. (2023) developed a scopolamine injection model to replicate delirium-like phenotypes in mice, offering clues regarding the mechanisms underlying delirium onset, an acute neuropsychiatric syndrome common in the elderly and an independent risk factor for later dementia. Zhang et al. (2023) reported that pre-existing diabetes can exacerbate lung damage in SARS-CoV-2 infection and found that low-molecular-weight heparin may be a valuable therapeutic approach for managing COVID-19 in diabetic patients. In an extensive review, Pan et al. (2023) discussed recent advancements in the cGAS-STING signaling pathway, vital in the innate immune system's defense against external

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

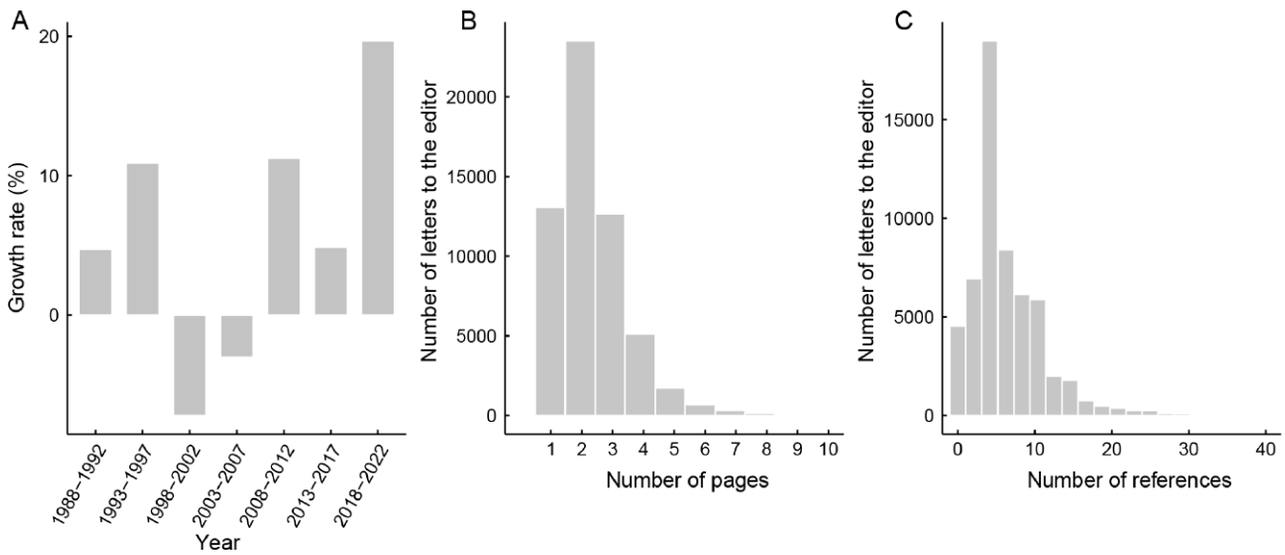


Figure 1 Information on letters to the editor based on Web of Science analysis

A: Growth rate of papers in the letters to the editor category in the past 30 years. The growth rate was calculated by comparing data from each five-year period with the previous five-year period. B, C: Number distribution of print pages (B) and references (C) in each paper in the letters category in 2021. Data used in this analysis were retrieved from the Web of Science.

pathogens and internal damage, and explored new avenues for further study.

While space limitations prevent us from listing all excellent *ZR* papers within this editorial, we remain confident that *ZR* will continue to flourish and expand in the new year. Despite the challenges posed by the misclassification of letters to the editor published in 2021–2022 as articles by Web of Science, which has affected our impact factor, the influence of *ZR* in the field continues to rise. Additionally, we warmly welcome submissions to our new journal *Zoological Research: Diversity and Conservation* (*ZRDC*), which aspires to become a leading publication in conservation and biodiversity.

Xiu-Ping Zhang

Xue-Long Jiang, Executive Editor-in-Chief

Yong-Gang Yao, Editor-in-Chief

Editorial Office of ZR & ZRDC, *Kunming Institute of Zoology*, *Chinese Academy of Sciences*, *Kunming*, *Yunnan* 650223, *China*

REFERENCES

Dkhar SA. 2018. Letter to editor: its importance and drawbacks. *International Journal of Community Medicine and Public Health*, **5**(10): 4634−4636.

Guo PC, Zuo J, Huang KK, et al. 2023. Cell atlas of CCl_4-induced progressive liver fibrosis reveals stage-specific responses. *Zoological Research*, **44**(3): 451−466.

Li C, Bi R, Wang L, et al. 2023. Characterization of long-term *ex vivo* expansion of tree shrew spermatogonial stem cells. *Zoological Research*, **44**(6): 1080−1094.

Liu GM, Pan Q, Du J, et al. 2023a. Improved mammalian family phylogeny using gap-rare multiple sequence alignment: A timetree of extant placentals and marsupials. *Zoological Research*, **44**(6): 1064−1079.

Liu XH, Gan L, Zhang ZT, et al. 2023b. Probing the processing of facial expressions in monkeys via time perception and eye tracking. *Zoological Research*, **44**(5): 882−893.

Lyu ZT, Qi S, Wang J, et al. 2023. Generic classification of Asian horned toads (Anura: Megophryidae: Megophryinae) and monograph of Chinese species. *Zoological Research*, **44**(2): 380−450.

Nadal-Nicolás FM, Galindo-Romero C, Lucas-Ruiz F, et al. 2023. Pan-retinal ganglion cell markers in mice, rats, and rhesus macaques. *Zoological Research*, **44**(1): 226−248.

Pan J, Fei CJ, Hu Y, et al. 2023. Current understanding of the cGAS-STING signaling pathway: Structure, regulatory mechanisms, and related diseases. *Zoological Research*, **44**(1): 183−218.

Qiu LF, Dong JX, Li XA, et al. 2023. Defining honeybee subspecies in an evolutionary context warrants strategized conservation. *Zoological Research*, **44**(3): 483−493.

Sitthivong S, Brakels P, Xayyasith S, et al. 2023. Hiding on jagged karst pinnacles: A new microendemic genus and species of a limestone-dwelling agamid lizard (Squamata: Agamidae: Draconinae) from Khammouan Province, Laos. *Zoological Research*, **44**(6): 1039−1051.

Wang Q, Zhang X, Guo YJ, et al. 2023. Scopolamine causes delirium-like brain network dysfunction and reversible cognitive impairment without neuronal loss. *Zoological Research*, **44**(4): 712−724.

Yao YG, Zhang Y. 2022. *Zoological Research* shines in the East. *Zoological Research*, **43**(1): 1−2.

Zhang ZY, Zhang N, Lu XC, et al. 2023. Anti-infection effects of heparin on SARS-CoV-2 in a diabetic mouse model. *Zoological Research*, **44**(6): 1003−1014.

Jiang et al. *Zool. Res.* 2024, 45(1): 138−151
https://doi.org/10.24272/j.issn.2095-8137.2023.060

ZR | Zoological Research

Article Open Access

Pig H3K4me3, H3K27ac, and gene expression profiles reveal reproductive tissue-specific activity of transposable elements

Tao Jiang[1,#], Zhi-Min Zhou[1,#], Zi-Qi Ling[1], Qing Zhang[1], Zhong-Zi Wu[1], Jia-Wen Yang[1], Si-Yu Yang[1], Bin Yang[1,*], Lu-Sheng Huang[1,*]

[1] *National Key Laboratory of Pig Genetic Improvement and Germplasm Innovation*, *Jiangxi Agricultural University*, *Nanchang*, *Jiangxi* 330045, *China*

ABSTRACT

Regulatory sequences and transposable elements (TEs) account for a large proportion of the genomic sequences of species; however, their roles in gene transcription, especially tissue-specific expression, remain largely unknown. Pigs serve as an excellent animal model for studying genomic sequence biology due to the extensive diversity among their wild and domesticated populations. Here, we conducted an integrated analysis using H3K27ac ChIP-seq, H3K4me3 ChIP-seq, and RNA-seq data from 10 different tissues of seven fetuses and eight closely related adult pigs. We aimed to annotate the regulatory elements and TEs to elucidate their associations with histone modifications and mRNA expression across different tissues and developmental stages. Based on correlation analysis between mRNA expression and H3K27ac and H3K4me3 peak activity, results indicated that H3K27ac exhibited stronger associations with gene expression than H3K4me3. Furthermore, 1.45% of TEs overlapped with either the H3K27ac or H3K4me3 peaks, with the majority displaying tissue-specific activity. Notably, a TE subfamily (LTR4C_SS), containing binding motifs for SIX1 and SIX4, showed specific enrichment in the H3K27ac peaks of the adult and fetal ovaries. RNA-seq analysis also revealed widespread expression of TEs in the exons or promoters of genes, including 4 688 TE-containing transcripts with distinct development stage-specific and tissue-specific expression. Of note, 1 967 TE-containing transcripts were enriched in the testes. We identified a long terminal repeat (LTR), MLT1F1, acting as a testis-specific alternative promoter in *SRPK2* (a cell cycle-related protein kinase) in our pig dataset. This element was also conserved in humans and mice, suggesting either an ancient integration of TEs in genes specifically expressed in the testes or parallel evolutionary patterns. Collectively, our findings demonstrate that TEs are deeply embedded in the genome and exhibit important tissue-specific biological functions, particularly in the reproductive organs.

Keywords: Transposable elements; Porcine; Histone modification; Alternative promoter; TE-containing transcript

INTRODUCTION

Transposable elements (TEs) are a class of mobile DNA sequences that constitute more than half of the DNA in many eukaryotes (Fedoroff, 2012). Based on their transposon mechanism, TEs can be divided into DNA transposons and retrotransposons, which move throughout the genome by direct cut-and-paste DNA sequence and RNA intermediates, respectively. Mammalian retrotransposons are broadly categorized into long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), and short interspersed nuclear elements (SINEs), which are repressed by different epigenetic mechanisms to prevent harmful insertion in the genome (Slotkin & Martienssen, 2007). Various studies have demonstrated that TEs provide binding sites for transcription factors (TFs) and represent an important source of regulatory elements, including promoters and enhancers (Chuong et al., 2017; Fueyo et al., 2022; Rebollo et al., 2012; Sundaram et al., 2017; Sundaram & Wysocka, 2020; Ting et al., 1992). These regulatory elements tend to be restricted to specific species or lineages (Chuong et al., 2016; Jacques et al., 2013; Kunarso et al., 2010). TEs are also expressed as a part of genes, thereby affecting gene expression and alternative splicing (Kapusta et al., 2013; Lee et al., 2022; Modzelewski et al., 2021; Senft & Macfarlan, 2021). Despite considerable progress, the tissue-specific activity and function of TEs remain largely unknown.

The pig (*Sus scrofa*) is an important agricultural animal and a valuable preclinical model for biomedical research (Lunney

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 08 August 2023; Accepted: 04 September 2023; Online: 05 September 2023

Foundation items: This work was supported by the National Natural Science Foundation of China (32160781)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: binyang@live.cn; LushengHuang@hotmail.com

et al., 2021; Wu et al., 2020; Yang et al., 2018; Zhong et al., 2023). The annotation of regulatory elements of the pig genome through profiling of histone modifications provides a basis for analyzing complex biological mechanisms (Zhu et al., 2022), accelerates the progress of medical research (Pan et al., 2021), and promotes the analysis of genetic mechanisms of complex porcine traits (Zhang et al., 2022b; Zhao et al., 2021). Construction of a regulatory element database will also provide a valuable resource for the development of gene editing tools, such as identifying potential safe harbors that enable stable expression of exogenous genes (Zhu et al., 2022). The Cre-loxp-mediated recombination system is a useful tool for tissue-specific gene editing that relies on tissue-specific promoters (Utomo et al., 1999). H3K27ac, which marks active enhancers and active promoters (Creyghton et al., 2010; Duttke et al., 2015), and H3K4me3, which marks active promoters (Barski et al., 2007; Bernstein et al., 2005; Pekowska et al., 2011), are two of the most relevant histone modifications associated with gene transcription. To date, however, few studies have compared the heterogeneity in H3K27ac and H3K4me3 activity or their associations with gene expression across different tissues and developmental stages, with even fewer focusing on the annotation of TEs in relation to their association with tissue-specific histone modifications and mRNA expression in pigs.

In this study, we integrated H3K4me3 ChIP-seq, H3K27ac ChIP-seq, and RNA-seq from 10 tissues of both fetal and adult pigs to annotate the regulatory sequences and TEs in the pig genome. Our results showed that H3K27ac peaks displayed a stronger association with gene expression than H3K4me3. Annotation of TEs with H3K27ac ChIP-seq, H3K4me3 ChIP-seq, and RNA-seq data across different fetal and adult tissues revealed several TE-containing transcripts and TE-derived regulatory elements associated with tissue-specific histone modifications, gene expression, alternative splicing, and alternative promoter usage, particularly in the ovaries and testes. Thus, this study provides a valuable framework for utilizing spatiotemporal H3K4me3 and H3K27ac ChIP-seq and RNA-seq data to explore the biological functions of TEs across different tissues and developmental stages.

MATERIALS AND METHODS

Ethics statement

All procedures involving animals followed the guidelines for the care and use of experimental animals approved by the State Council of the People's Republic of China. This study was approved by the Committee on Animal Biosafety of Jiangxi Agricultural University.

Sample collection

As previously described (Zhu et al., 2022), eight tissues (brain, heart, liver, lungs, muscle, small intestine (SI), ovaries, and testes) from seven fetuses (including two male and two female full-sib Large White pigs on day 75 and two male and one female Bamaxiang pigs on day 74 post-insemination) and 10 tissues (brain, heart, liver, lungs, muscle, pituitary, SI, stomach, ovaries, and testes) from eight adult pigs (two male and two female Large White pigs on day 150 and two male and two female Bamaxiang pigs on days 132–141) were collected in this study. The anatomical sites of sampled tissues were consistent in fetuses and adults. The collected tissues were frozen in liquid nitrogen immediately after slaughter and stored at −80 °C.

ChIP preparation and sequencing

ChIP was performed using the SimpleChIP® Plus Enzymatic Chromatin IP Kit (Magnetic Beads) (Cell Signaling Technology, USA). Briefly, 0.2–0.3 g of each sample was homogenized with 1 mL of phosphate-buffered saline (PBS) and cross-linked with 37% formaldehyde at room temperature for 10 min. Then, 10×glycine was added and mixed thoroughly to end the cross-linking reaction. Buffers A and B were then added for lysing. The DNA was sheared into 100–300 bp segments after sonication and enzyme digestion. ChIP was incubated with 5 μg of H3K27ac antibody (Active Motif, USA) or 3 μg of H3K4me3 antibody (Active Motif, USA) overnight to accumulate H3K27ac-marked regions or H3K4me3-marked regions, respectively. Antibody-bound DNA was collected using ChIP-grade Protein G magnetic beads, then incubated with 2 μL of 20 mg/mL proteinase K and 6 μL of 5 mol/L NaCl at 65 °C to reverse the cross-links. Immunoprecipitated DNA was then purified using a spin column. The steps of the method are binding, washing and elution. Binding of nucleic acid to a silica membrane, washing away particulates and inhibitors that are not bound to the silica membrane, and elution of the nucleic acid, with the end result being purified nucleic acid in an aqueous solution. Single-end 50 bp DNA sequencing was performed using the Illumina HiSeq 2500 platform (USA).

ChIP-seq processing

Clean reads were aligned to the pig genome (*Sus scrofa* 11.1) using Burrows-Wheeler Aligner (BWA) (Abuín et al., 2015), allowing two mismatches. Duplicate alignments were removed using SAMtools markdup (Li et al., 2009). MACS2 v.2.1.1 (Zhang et al., 2008) was used to infer peaks for each sample with the following parameter: "--q 0.01 -B --keep-dup all". Broad domains were identified using the parameter "--broad and --broad-cutoff 0.1". Peaks were merged across all samples using the "merge" function in BEDTools (Quinlan, 2014) and the number of reads in each merged peak was calculated using SAMtools bedcov (Li et al., 2009). Finally, the normalized value (peaks per million) for each merged peak was defined in R program v.3.5.1 (https://www.r-project.org/).

RNA-seq library preparation, sequencing, and processing

Total RNA was extracted from frozen samples using TRIzol reagent (Invitrogen, USA), followed by mRNA collection using magnetic beads attached to poly-T oligos. The collected mRNA was then reverse-transcribed into cDNA. AMPure XP Beads were used for fragment enrichment and polymerase chain reaction (PCR) amplification to obtain strand-specific cDNA libraries. These libraries were then subjected to paired-end 150 (PE150) sequencing on the Illumina HiSeq 4000 platform (USA).

The STAR program (Dobin et al., 2013) was used to map the clean reads to the *Sus scrofa* 11.1 genome using Ensembl GTF (98.111) (Supplementary Table S1). The transcripts were assembled using StringTie (Pertea et al., 2015) and quantified with FeatureCounts (Liao et al., 2014). Genes with transcripts per million (TPM) values less than 1 in 90% of the samples were excluded. In total, 17 095 genes were retained for further analysis.

Predicting expression using ridge regression and peak-gene correlations

A total of 94 samples with two histone modifications

(H3K4me3 and H3K27ac) and corresponding transcriptome data were collected, including data from 63 samples from two previous studies (Supplementary Table S2) and data from 31 samples generated in this study. From Kern et al. (2021), a total of 11 samples were collected from seven tissues (adipose, cerebellum, cerebral cortex, hypothalamus, liver, lungs, and spleen) of a 6-month-old male Yorkshire breed. From Zhao et al. (2021), a total of 52 samples were collected from five tissues (muscle, liver, fat, spleen, and heart) of 6-month-old male MeiShan, Large White, Enshi Black, and Duroc breeds. A uniform ChIP-seq data analysis pipeline was applied to the 63 downloaded datasets. We retained 18 324 genes with TPM>1 in at least 10% of the samples for further analysis. The correlations between genes and peaks were assessed using Spearman correlation analysis. Genes with transcription start sites (TSSs) located within 500 kb of the peak centers were considered. For each peak-gene pair, *P*-values were corrected using the qvalue package, with peak-gene pairs with *Q*-values less than 0.01 retained.

We evaluated three distinct models, each with different predictors, including solely H3K4me3 peaks, solely H3K27ac peaks, or both peaks, to determine their capacity to predict gene expression in three peak-gene distance scenarios, i.e., association of peaks within 50 kb, 200 kb, and 500 kb of corresponding genes. For each prediction model, the whole dataset was divided into training and test data. Ridge regression was used to train a prediction model using the training dataset, after which the model was evaluated using the test data. Pearson correlation coefficients (r^2) between the predicted and measured values of gene expression were calculated as a measure of prediction accuracy.

Identification of tissue-specific promoters

Tissue specificity was defined with a Tau score threshold of 0.95 (Kryuchkova-Mostacci & Robinson-Rechavi, 2017) on the H3K27ac and H3K4me3 ChIP-seq data and RNA-seq data. Tissue-specific peaks within 1 kb upstream of the TSSs of tissue-specific genes and supported by both histone markers were defined as candidate tissue-specific promoters.

Identification of accessible TEs in multiple tissues at different developmental stages

Based on tissues and developmental stages, 84 samples of H3K27ac peaks were divided into 17 stage-tissue groups. TEs in the reference genome *Sus scrofa* 11.1 were identified using RepeatMasker (Tarailo-Graovac & Chen, 2009), as described previously (Chen et al., 2022). TEs with 50% of their sequence overlapping with at least one peak were considered peak overlapping TEs, using intersectBed, as described previously (Judd et al., 2021).

Enrichment analysis of TEs

Analysis of TE enrichment within specific stage-tissue group peaks was conducted using a Perl script (https://github.com/4ureliek/TEanalysis), as described previously (Kapusta et al., 2013), which randomized the location of TEs and maintained the distance to the nearest TSS and number of TEs on each chromosome. This procedure was bootstrapped 1 000 times. TE enrichment in a peak set was deemed significant if the ratio of observed to expected values exceeded 2.

Quantification of transcripts and TE-gene junctions

We remapped 112 RNA-seq datasets to the swine genome assembly (*Sus scrofa* 11.1) using STAR aligner v.2.7.2b (Dobin et al., 2013) with the following parameters: " --outFilterMultimapNmax 500 --alignIntronMax 1000000 --outFilterType BySJout --alignMatesGapMax 1000000". A two-pass alignment strategy was used to increase the detection sensitivity of the spliced RNA-seq reads. For first-pass alignment, the STAR index was installed with the gene annotations provided by RefSeq. The second STAR index was updated using the results from the first alignments. Quantification of TEs and transcripts was performed using TEcount (Jin et al., 2015), as described previously (Modzelewski et al., 2021). Finally, counts of exon model per million mapped reads (CPM) was adopted to normalize raw counts. Stage- and tissue-specific transcript expression analysis was performed using the Tau score, as described previously (Zhu et al., 2022).

For TE-gene junctions, all junctions were assembled and annotated using a previously described pipeline (Modzelewski et al., 2021). In brief, StringTie2 (Pertea et al., 2015) was used for transcript assembly with the following parameters: "-j 2 -s 5 -f 0.05 -c 2". The output was then merged using TACO (Niknafs et al., 2017). Splicing junctions detected in the RNA-seq dataset were then collected. The starting position of the junction was annotated using information on TEs collected from RepeatMasker (Tarailo-Graovac & Chen, 2009). Junctions with at least 10 reads were retained and normalized with transcripts per kilobase of exon model per million mapped reads. Stage- and tissue-specific TE-gene junctions were determined using the Tau score.

Motif analysis

Enrichment of TF-binding motifs in peaks was analyzed using HOMER (Heinz et al., 2010) with default parameters, and the results were verified using MEME (Bailey et al., 2015). FIMO (Grant et al., 2011) was used to predict the locations of potential TF motifs with the database of known vertebrate transcription factor motifs. TF motif heatmap visualization of binding regions was conducting using deepTools (Ramírez et al., 2014).

RESULTS

Samples and data in this study

We performed H3K4me3 and H3K27ac ChIP-seq and RNA-seq on eight tissues from four Large White and three Bamaxiang pig fetuses (prenatal days 74–75) and 10 tissues from four Large White and four Bamaxiang adult pigs (postnatal days 132–150) of both sexes (Figure 1A). In total, we successfully obtained 84 H3K27ac ChIP-seq, 42 H3K4me3 ChIP-seq, and 112 RNA-seq datasets, which passed quality control procedures for follow-up study, including 31 matched datasets for H3K4me3, H3K27ac, and RNA-seq from the same samples (Figure 1B; Supplementary Tables S1, S3, S4). Among these data, 70 H3K27ac ChIP-seq and 62 RNA-seq datasets have been reported previously (Zhu et al., 2022), while the remaining data were obtained in this study for the first time. After removing low-quality reads and duplicates, we generated an average of 23.82 million effective reads per ChIP-seq assay (Supplementary Tables S3, S4). Hierarchical clustering of samples based on the signals of the two histone modifications and gene expression showed that the samples were first grouped by sequencing assay and then by tissue type (Supplementary Figure S1A), similar to the results of Pan et al. (2021). Furthermore, the means of the fraction of reads in peaks (FRiP), normalized strand cross-correlation

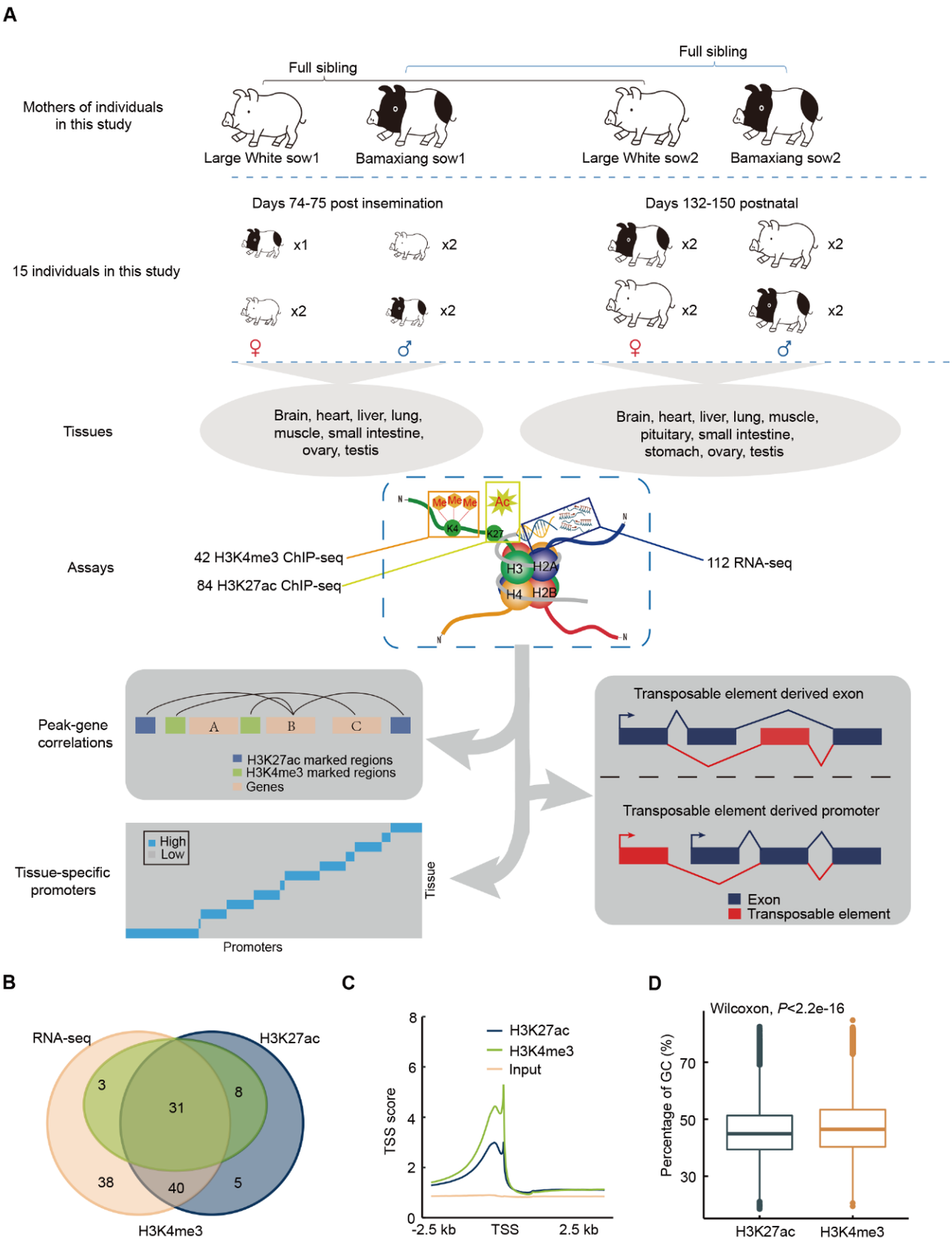


Figure 1 Scheme of samples and experiments in this study

A: Experimental design of study. B: Venn diagram showing overlap of samples by the three assays. Different colors indicate different assays, pink for RNA-seq, blue for H3K4me3 ChIP-seq, and green for H3K27ac ChIP-seq. C: Average read coverages of two histone modifications at 2 500 bp upstream and 2 500 bp downstream regions of genes. D: Box plot showing GC contents in H3K27ac and H3K4me3 peaks. Statistical significance of comparisons between H3K27ac and H3K4me3 peaks was calculated using wilcox.test in the R program.

coefficient (NSC), and relative strand cross-correlation coefficient (RSC) were 0.20, 1.06, and 2.98, respectively (Supplementary Tables S3, S4), meeting all data standards from the ENCODE guidelines and thus supporting the quality of the ChIP-seq data. We identified an average of 24 957 (range: 11 368 to 59 338) H3K4me3 and 38 525 (range: 10 506 to 81 255) H3K27ac peaks per sample (Supplementary Figure S1B, C). After combining peaks across different samples, we obtained 115 181 and 271 376 merged H3K4me3 and H3K27ac peaks, respectively, further enriching the regulatory element catalog for pigs obtained in previous study (Zhu et al., 2022). Of the 115 181 merged H3K4me3 peaks, 98 041 (85.12%) overlapped with at least one of the merged H3K27ac peaks, indicating that H3K27ac tagged most of the active promoter regions exhibiting H3K4me3 activity (Howe et al., 2017; Roth et al., 2001). H3K4me3 peaks exhibited a higher degree of enrichment around TSSs and higher GC content compared to the H3K27ac peaks (Figure 1C, D), indicating that H3K4me3 is more likely to occur at TSSs and be subjected to DNA methylation than H3K27ac (Hughes et al., 2020).

Associations of H3K27ac and H3K4me3 with gene expression

We next compared the H3K4me3 and H3K27ac peaks in terms of their associations with gene expression. To enhance the power of peak-gene association analysis, we downloaded 63 publicly available H3K27ac and H3K4me3 ChIP-seq and RNA-seq datasets (Supplementary Table S2). The data were analyzed using the same analytical pipelines, assembling a dataset of 94 samples with matched H3K27ac and H3K4me3 ChIP-seq and RNA-seq data. In total, 333 867 H3K27ac and 126 651 H3K4me3 merged peaks were identified in the 94 samples. For each gene, we compared the predictive performance of three models (H3K27ac model, H3K4me3 model, and two-modifications model) for histone peaks located within 50 kb, 200 kb, and 500 kb of the TSSs of the corresponding genes. Notably, the H3K27ac model showed higher prediction accuracy of gene expression than the H3K4me3 model in all peak-gene distance scenarios, suggesting a stronger gene regulatory role of H3K27ac than H3K4me3 (Figure 2A). The predictive power of the H3K27ac model was comparable to that of the two-modification model, which may reflect functional redundancy of H3K27ac and H3K4me3 in their regulatory role in gene expression (Karlić et al., 2010).

The integrated data also enabled the identification of potential target genes of the two histone modification markers. We calculated the correlations between the two histone markers and genes within 500 kb and identified 77 989 H3K4me3-gene and 438 615 H3K27ac-gene correlations (*Q*-values<0.01), providing a valuable resource for exploring the links between regulatory elements and genes across pig tissues. For both histone markers, positive peak-gene correlations were enriched around the TSSs of the corresponding genes, whereas negative correlations were under-represented in the TSS regions (Figure 2B; Supplementary Figure S2A). Results showed that 16 939 out of 29 408 H3K4me3 peaks showed stronger associations with distant rather than adjacent genes, e.g., the peak (chr4:90 642 880–90 645 761) showed the highest correlation with *PEA15*, which was located 369 kb from the peak (Supplementary Figure S2B, C), suggesting that some promoters may play enhancer roles, in agreement with observations from Hi-C data (Delaneau et al., 2019).

Identification of tissue-specific promoters

Identification of tissue-specific promoters will facilitate tissue-specific genome editing of genes using the Cre/loxP recombinase system (Utomo et al., 1999). Therefore, we identified H3K27ac and H3K4me3 peaks linked to tissue-specific gene expression. After applying strict filtration procedures (see Methods for more details), we identified 230 tissue-specific promoters simultaneously supported by the tissue-specific activity of H3K27ac, H3K4me3, and mRNA expression (Figure 2C, D; Supplementary Tables S5, S6). Among these genes, we identified the heart-specific gene *TBX20*, a gene required for normal heart function (Kirk et al., 2007; Qian et al., 2008), and kidney-specific gene *PAX2*, which is associated with renal maldevelopment (Fletcher et al., 2005). These tissue-specific promoters are promising candidates for developing tissue-specific editing tools, e.g., Cre/loxP recombinase system.

Enrichment analysis of TEs in H3K27ac peaks

Once considered as "junk DNA" (Biémont, 2010), TEs constitute a rich source of *cis*-regulatory elements involved in the regulation of eukaryotic gene expression via binding to TFs (Bourque et al., 2018). At present, however, the regulatory functions of TEs in pigs remain largely unexplored. First, we investigated the proportion of each TE class in certain functional genomic regions (Supplementary Figure S3A). Results showed that all types of TEs were strongly enriched in intronic and intergenic regions but not in TSSs or exons and were depleted in the two histone modification regions. Despite this, 1.45% (56 763/3 925 013) of the TEs in the pig genome overlapped with H3K27ac or H3K4me3 peaks. To investigate the heterogeneity of the epigenetic states of TEs across stage-tissue groups, we explored the TEs overlapping with the H3K27ac peaks in the 84 H3K27ac ChIP-seq data samples, given that the H3K27ac peaks tagged most of the H3K4me3 peaks (85.12%), and more samples were assayed for H3K27ac ChIP-seq. Results showed that 28–7 031 H3K27ac peaks overlapped with at least one TE in the different stage-tissue groups (Figure 3A). Peaks overlapping with TEs had lower intensities than other peaks across the stage-tissue groups (Supplementary Figure S3B), supporting the notion that TEs tend to be transcriptionally repressed (Colonna Romano & Fanti, 2022). Additionally, over half of the TE-overlapping peaks were active in only a single stage-tissue group (Supplementary Figure S3C), suggesting stage- and tissue-specific regulatory elements (Supplementary Figure S3D). Further analysis was conducted on various TE subfamilies to assess their enrichment in H3K27ac peaks across the different stage-tissue groups using a previously described pipeline (Judd et al., 2021). Three unique TEs, namely, LTR_ERVK_LTR4C_SS, LTR_ERV1_ERV1-2B_SSc-LTR, and LTR_ERVK_LTR6B_SS, were identified as significantly enriched in H3K27ac peaks in at least one stage-tissue group (Figure 3B).

LTR4_SS, with more than 50 observed bound copies, showed greater enrichment in both adult and fetal ovaries than in other tissues (Figure 3B, C). We further performed TF binding motif enrichment analysis on the 34 peaks overlapping the LTR4C_SS element in the ovaries using MEME (Bailey et al., 2015) and HOMER (Heinz et al., 2010). Analyses revealed enrichment of the binding motifs of SIX4 and SIX1,

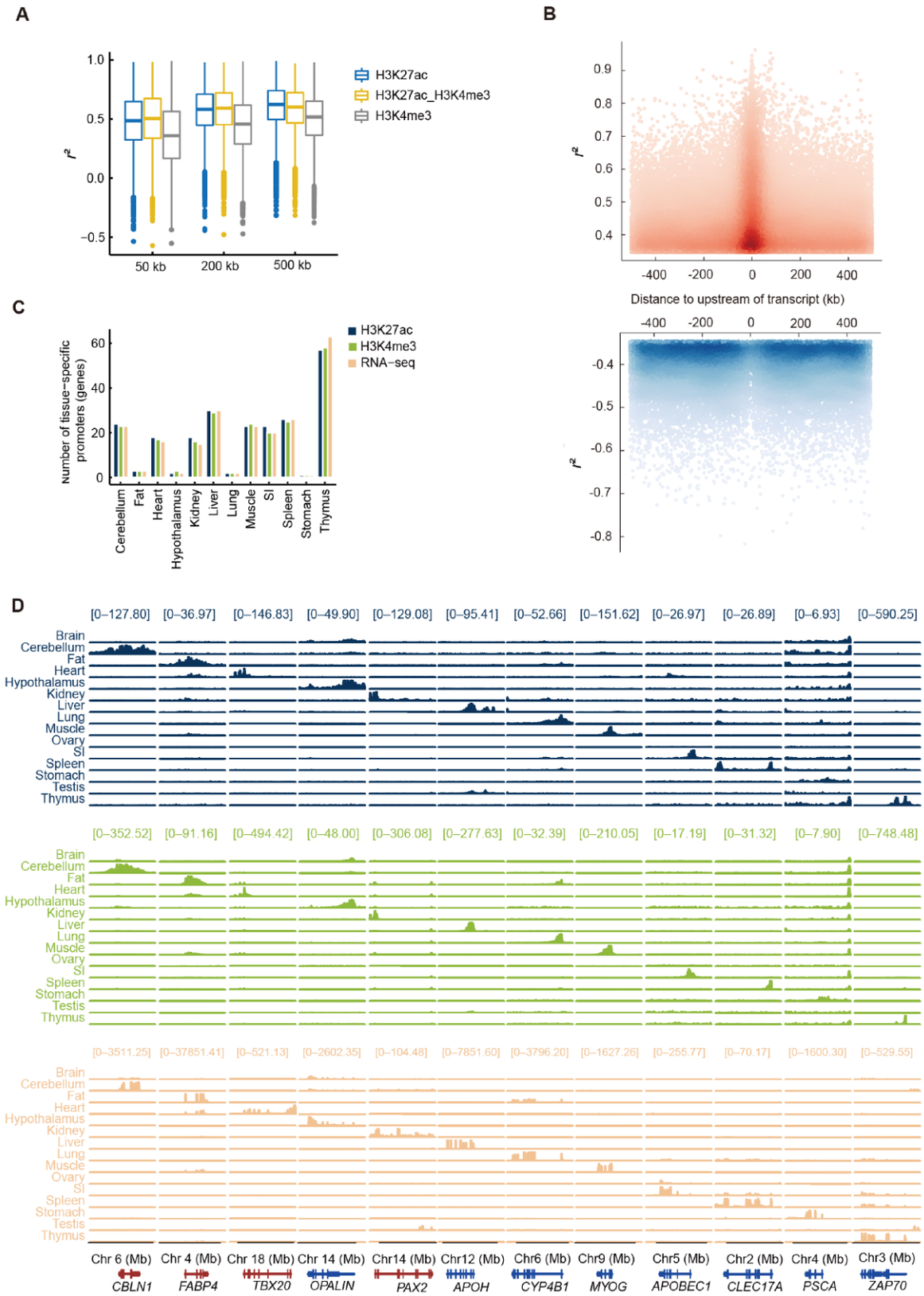


Figure 2 Histone modification levels are predictive for gene expression

A: Pearson correlations between predicted and measured gene expression levels by different models; X-axis corresponds to distance between center of peaks to TSSs of genes. B: Distribution of distance for significant positive peak-gene pairs and negative peak-gene pairs from H3K4me3 are shown in upper and lower panels, respectively. Dark color indicates clustering of peak-gene pairs at this distance. C: Number of tissue-specific peaks (genes) in different tissues. D: Representative examples of tissue-specific promoters. Each track represents average number of base-pairs per million mapped reads in 20 bp bins across all samples in corresponding tissue. Scale of track was consistent for each assay. Different colors indicate different types of assays as described in C.

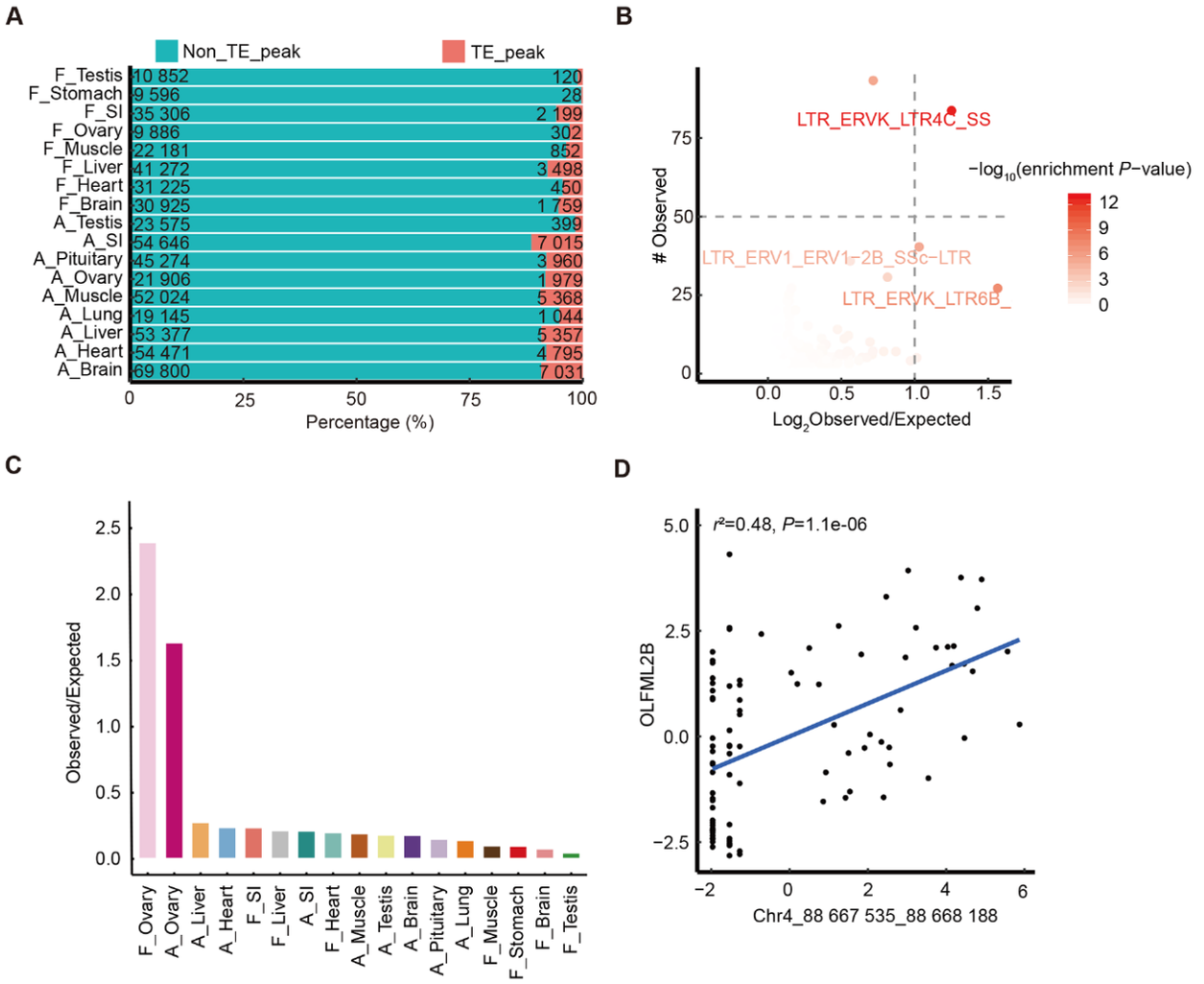


Figure 3 LTR4C_SS elements are highly active peaks enriched in ovaries

A: Bar plot showing numbers and fractions of peaks in stage-tissue groups that overlapped with TEs. B: All TEs detected in ChIP-seq peaks, plotted by number of elements observed compared to ratio of observed to expected occurrences for that particular TE. Expected values were calculated by bootstrapping 1 000 times. Dashed line denotes cutoff of >50 elements observed to filter false positives. C: Ratio of observed to expected occurrences of LTR4C_SS, as in B, for each stage-tissue group. D: Correlation between peak (chr4_88 667 535_88 668 188), which overlapped with LTR4C_SS and target gene (*OLFML2B*). Each dot represents peak intensity and gene expression from each sample.

further supported by the higher ChIP signatures around the binding motifs of SIX1 and SIX4 (Supplementary Figure S3E, F). *SIX1* is implicated in the advance of late-stage ovarian carcinoma by resisting TRAIL-mediated apoptosis (Behbakht et al., 2007), and *SIX4* is known to play an important role in maintaining the resting state of primary follicles and in restricting ovarian growth (Zhang et al., 2016). Overall, these results suggest that LTR4C_SS provides the binding motifs of SIX4 and SIX1, which may be critical for ovarian function in pigs. The overlapping LTR4C_SS peaks had an average distance of 118.87 kb from their closest genes (Supplementary Figure S3G), indicating that the LTR4C_SS element primarily acts as an enhancer. We identified four genes (*OLFML2B*, *SNORA11G*, *NFKBIZ*, and *ENSSSCG00000006350*) associated with the LTR4C_SS overlapping peaks (Figure 3D; Supplementary Figure S3H–J). Among these, *OLFML2B* (*P*=1.1e-6) participates in cell growth, maintenance, cell cycle regulation, apoptosis, and cell communication (Liu et al., 2019). *SNORA11G* (*P*=4.8e-6) encodes a snoRNA class essential for rRNA stability, processing, and fidelity of ribosome assembly and translation (Holley & Topkara, 2011). Dysfunction of the transcription factor *NFKBIZ* (*P*=1.4e-4) is associated with premature ovarian failure-18 (Safran et al., 2010).

Expression of TEs across tissues and developmental stages

Increasing evidence suggests that the expression of TEs in tissues is involved in placental development, immune system, brain development, and other biological functions (Bourque et al., 2018; Cornelis et al., 2017; Joly-Lopez & Bureau, 2018; Naville et al., 2016). Thus, we quantified both TEs and TE-gene junctions based on multiply mapped reads from the RNA-seq data using TEcount and TEtranscript (Jin et al., 2015), respectively (Supplementary Figure S4A). Results showed that TEs accounted for 0.12% to 0.83% of the RNA-seq reads in each stage-tissue group (Supplementary Figure S4B). We identified 8 068 transcribed TE-gene junctions in at least one of the tissues (Supplementary Figure S4C). The majority (89.71%) of the involved TEs were retrotransposons (Supplementary Figure S4C). Additionally, except for the testes, these genes showed higher expression in fetal tissues

than in adult tissues (Supplementary Figure S4C). Most TEs (84.02%) were expressed as parts of exons, while 20.56% of TEs were located in coding sequences, indicating potential in creating novel protein sequences (Supplementary Figure S4C).

Out of the 8 068 transcribed TE-gene junctions, 4 688 stage- and tissue-specific TE-gene junctions were identified with Tau scores greater than 0.85 (Supplementary Figure S4D), involving 483 TE subfamilies, including 123 LINE (25.47%), 37 SINE (7.67%), 191 LTR (39.54%), and 132 DNA transposon (27.33%) subfamilies. Among the 4 688 TE-gene junctions, 3 456 specifically influenced the exons of 2 135 genes. Notably, 57% of these genes (1 220/2 135) showed stage- and tissue-specific expression, including 10 in the fetal SI and 503 in the adult testes (Figure 4A). These genes are involved in pathways associated with the biology of the corresponding tissues (Figure 4A), such as cardiac muscle contraction in the adult heart and carboxylic acid catabolic processes in the adult liver. These findings support the notion that TEs can impact protein coding sequences and that a considerable proportion of TEs expressed in a tissue-specific manner may play important roles in tissue-specific functions.

Notably, more than 41% (1 967/4 688) of the stage-tissue specific TE-exon junctions were found in the adult testes, consistent with the observation that genes expressed in the testes more frequently experience TE insertion events (Bhalla, 2020; Reik et al., 2001). Several corresponding genes, including *ADAM30*, *ADAM32*, and *ZPBP*, are known to be involved in testicular biological functions, while other genes have not previously been shown to act in the testes. For example, *CSNK1A1* displayed high expression across all stage-tissue groups at the gene level (Figure 4C), while the TE (PRE1e)-exon 9 junction in *CSNK1A1* showed adult testis-specific expression (Figure 4B, D), indicating the potential role of PRE1e in the transcription of *CSNK1A1* in adult testes. Interestingly, other casein kinases, such as *CSNK1G2*, are known to suppress necroptosis-promoted testicular aging (Li et al., 2020). Similar patterns were also observed in other stage-tissue groups. For example, MamSINE1 serves as an adult brain-specific expressed exon for *ERMN*, LTR16B2 serves as an adult liver-specific expressed exon for *LECT2*, L1MB3 serves as a fetal brain-specific expressed exon for *STMN4*, MamRTE1 serves as an adult muscle-specific expressed exon for *TMOD4*, L1ME4c serves as a fetal ovary-specific expressed exon for *FIGLA*, and Tigger19a serves as an adult SI-specific expressed exon for *MS4A12* (Supplementary Figure S5A–F).

Stage tissue-specific alternative promoters derived from TEs

Furthermore, 717 junctions were identified where the corresponding TEs functioned as alternative promoters for 600 genes, among which 351 showed stage- and tissue-specific expression (Supplementary Figure S6A). Further screening of these TEs using the two ChIP-seq datasets led to the identification of 97 TE-derived promoters that overlapped with both histone modification regions (Figure 5B; Supplementary Table S7). This group included the promoter-overlapping

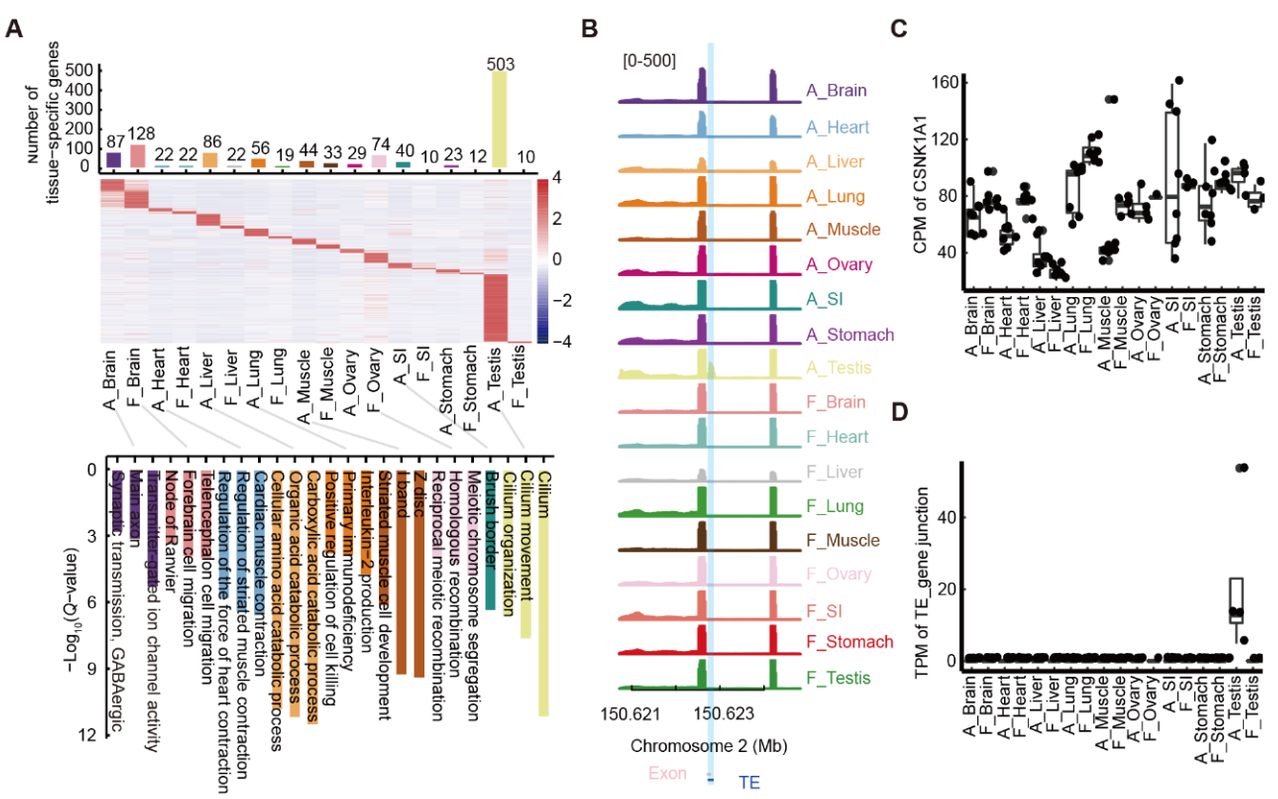


Figure 4 Expression of TE-containing genes

A: Heatmap showing stage-tissue specific gene expression of 3 456 TE-gene junctions, where TEs intersect with exons of corresponding genes. Bar plot above heatmap represents number of stage-tissue specific TE-containing genes. Bar plot below heatmap represents enriched GO terms and KEGG pathways in stage-tissue specific TE-containing genes. B: Tracks of TE-exon expression across stage-tissue groups; Y-axis corresponds to read depths of RNA-seq data. C: Expression of *CSNK1A1* across stage-tissue groups. D: Expression of PRE1e-*CSNK1A1* junction across stage-tissue groups.

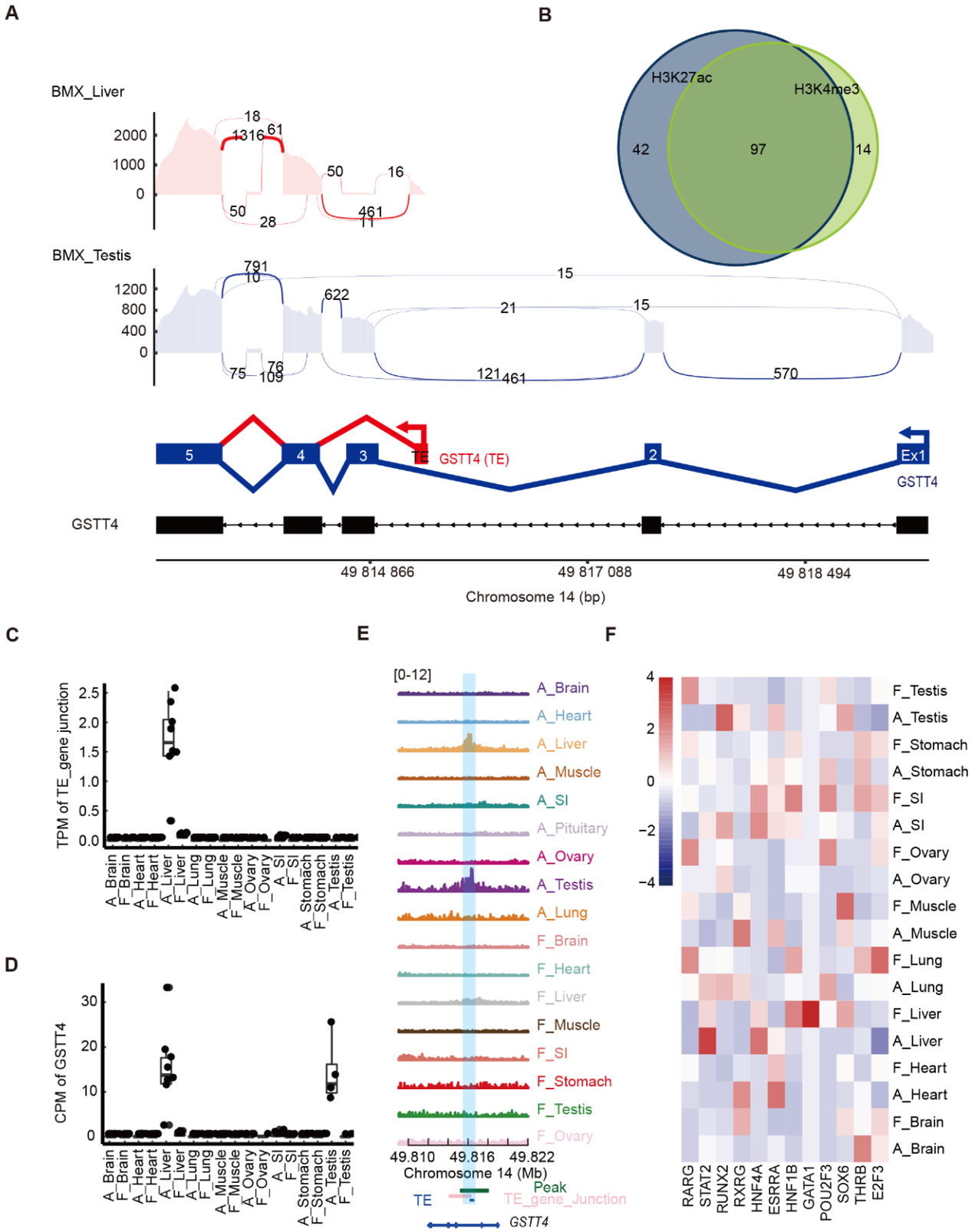


Figure 5 TE-associated tissue-specific alternative promoter usage of *GSTT4* gene

A: Assembled transcripts of *GSTT4* genes in adult BMX liver and adult BMX testis. Top: Sashimi-plot visualization of splicing-junctions from L1MC5-derived promoter. Bottom: Schematic representation of different exon usage. For TE-derived promoter, first exon originated from the TE-derived TSS and was spliced to a fourth exon and again to the fifth exon, skipping the canonical first to third exons. B: Venn diagram showing overlap between TEs involved in stage-tissue specific TE-gene junctions with H3K27ac and H3K4me3 peaks. C: Expression levels of *GSTT4* in all stage-tissue groups. D: Expression levels of L1MC5:GSTT4 junction in all stage-tissue groups. E: H3K27ac tracks in TE-derived *GSTT4* alternative promoter region. F: Heatmap of expression levels of TFs with potential binding motif in L1MC5. Color of each block represents scale value of gene expression level.

junction L1MC5:GSTT4, where the *GSTT4* gene, which was mainly expressed in the adult liver and testes (Figure 5D), is involved in glutathione metabolic processes. The L1MC5 element appeared to facilitate an alternative promoter for *GSTT4* in the liver, resulting in liver-specific expression of exons 4–5 (Figure 5A, C). In comparison, *GSTT4* expressed exons 1–5 in the testes, with low expression in the other tissues (Figure 5A; Supplementary Figure S6B). The liver-specific expression of the L1MC5-GSTT4 junction was associated with liver-specific activity of the two histone markers, while the high activity of the two modifications in the testes was unexpected (Figure 5E). Among all inferred TFs with potential TF-binding motifs in L1MC5 (chr14:49 816 220–49 816 558), *STAT2* showed significantly higher expression in the adult liver than in the other tissues (Figure 5F). *STAT2* is a transcriptional activator that responds to cytokines and growth factors and may drive the liver-specific expression of the L1MC5-GSTT4 junction-containing transcripts. Similar patterns were also observed in the adult testes. For example, MIR serves as an adult testis-specific promoter causing testis-specific expression of the *DLGAP5* transcript (Supplementary Figure S7A–C).

Cross-species conservation of tissue-specific expression of TE-gene junctions

We next examined the cross-species conservation of tissue-specific expression of TE-gene junctions. Using the same analytical pipeline, 8 560 and 3 957 stage- and tissue-specific TE-gene junctions were identified by analyzing published RNA-seq data from humans and mice, respectively (Supplementary Table S8). Notably, 31 genes that displayed conserved tissue-specific expression of TE-gene junctions were identified in the three species (Supplementary Tables S9–S11), among which 29 showed testis-specific expression. For example, the retrotransposon MLT1F1 is reported to drive testis-specific promoter usage of *SRPK2* (SRSF Protein Kinase 2, a gene associated with spermatogenesis and alternative splicing (Giannakouros et al., 2011)) in the three species (Figure 6A–C). Our results indicated that the insertion of MLT1F1 into the genome may precede the divergence of the three species. The MLT1F1 sequences in *SRPK2* from the three species showed 39%–51% sequence identity (Supplementary Figure S8). To identify potential TFs driving conserved alternative promoter use in the three species, FIMO (Grant et al., 2011) was used to predict the locations of TF-binding motifs in the MLT1F1 sequences, revealing a core "GTCATAAAA" sequence shared by all three species, predicted to bind to 14 TFs (Supplementary Figure S8). Among these TFs, HOX proteins are crucial for testicular and epididymal physiological functions and can affect male fertility (Ferguson & Agoulnik, 2013; Topaloğlu et al., 2021; Xian et al., 2015). Species-conserved TE-derived alternative promoters were not limited to the same subfamily of TEs, i.e., different TE subfamilies acted on the same gene across the three species. For example, the MIR element serves as a testis-specific promoter for *SUGP2*, which mediates the alternative splicing regulatory network during spermatogenesis in the testes (Zhan et al., 2021). The subfamilies of MIR were MIRb, MIR, and MIR1_Amn in pigs, mice, and humans, respectively (Supplementary Figure S9).

DISCUSSION

In this study, we performed a comprehensive analysis of H3K4me3 and H3K27ac ChIP-seq and RNA-seq data from 10 tissues of seven fetuses and eight adult pigs of both sexes, then annotated the regulatory sequences of TEs in the pig genome. In total, we identified 115 181 H3K4me3 peaks and 271 376 H3K27ac peaks, further enriching the repository of regulatory elements for pigs. We examined the predictive efficiency of different combinations of histone modifications on gene expression using ridge regression models and found that the activity of H3K27ac exhibited higher associations with gene expression than that of H3K4me3. Additionally, we identified ~230 tissue-specific promoters that may be particularly valuable for tissue-specific gene editing. Together, these results provide novel resources and new insights into the regulation of gene expression.

The colonization of TEs poses a considerable threat to genome integrity (Ardeljan et al., 2017), which may increase the risk of genomic instability (Ayarpadikannan & Kim, 2014). TE fragments embedded in the genome may also have diverse functional consequences (Fueyo et al., 2022). For example, Chang et al. (2022) found that TEs are sources of new protein-coding exons in zebrafish, while Lee et al. (2022) demonstrated that TEs contribute to exons and promoters of genes in zebrafish. Thus, we performed enrichment analysis of TEs using H3K27ac ChIP-seq. A notable finding was the identification of the TE LTR4C_SS, which was significantly enriched in ovarian tissue and harbored binding sites for SIX1 and SIX4 motifs, suggesting involvement in gonadal development.

We then focused on the expression of TEs across stage-tissue groups using RNA-seq to reveal temporal- and spatial-specific patterns of TE expression. In total, we identified 8 068 TE-gene junctions, with the corresponding genes expressed at higher levels in the fetuses than in the adults, except in the testes, reflecting imperfect epigenetic regulatory mechanisms in the fetuses. We also identified 4 688 stage- and tissue-specific TE-gene junctions, most of which were expressed in the adult testes. Remarkably, 3 456 of the 4 688 TE-gene junctions contained TE-derived exons. The corresponding genes (1 220/2 150) displayed stage- and tissue-specific patterns, with 503 genes specifically expressed in the adult testes and enriched in pathways relevant to their corresponding tissues. Notably, PRE1e, located within a *CSNK1A1* exon, demonstrated adult testis-specific expression, while the overall expression of *CSNK1A1* remained consistent across all stage-tissue groups. This gene phosphorylates many proteins and participates in Wnt signaling. Several factors could explain the preferential activation of TEs in adult testes: (1) TEs replicate in germline cells, allowing transportation of novel insertions to the next generation (Cosby et al., 2019); (2) males produce four meiotic products per meiotic division and thus have a higher tolerance for TE insertion (Bhalla, 2020); (3) subtle and brief temperature increases during spermatogenesis can increase TE activity and mobility (Kurhanewicz et al., 2020); (4) chromatin remodeling during spermatogenesis ensures developmental reprogramming, providing favorable conditions for TE activation and mobility (Castañeda et al., 2011); and (5) newly emerged genes tend to exhibit testis-specific expression before becoming more widespread, aligning with the "out of testis" hypothesis (Kaessmann, 2010). Thus, it would be interesting to investigate TE activity in the context of developmental transitions and differentiation trajectories of cells in single-cell data of testes in future work (Zhang et al.,

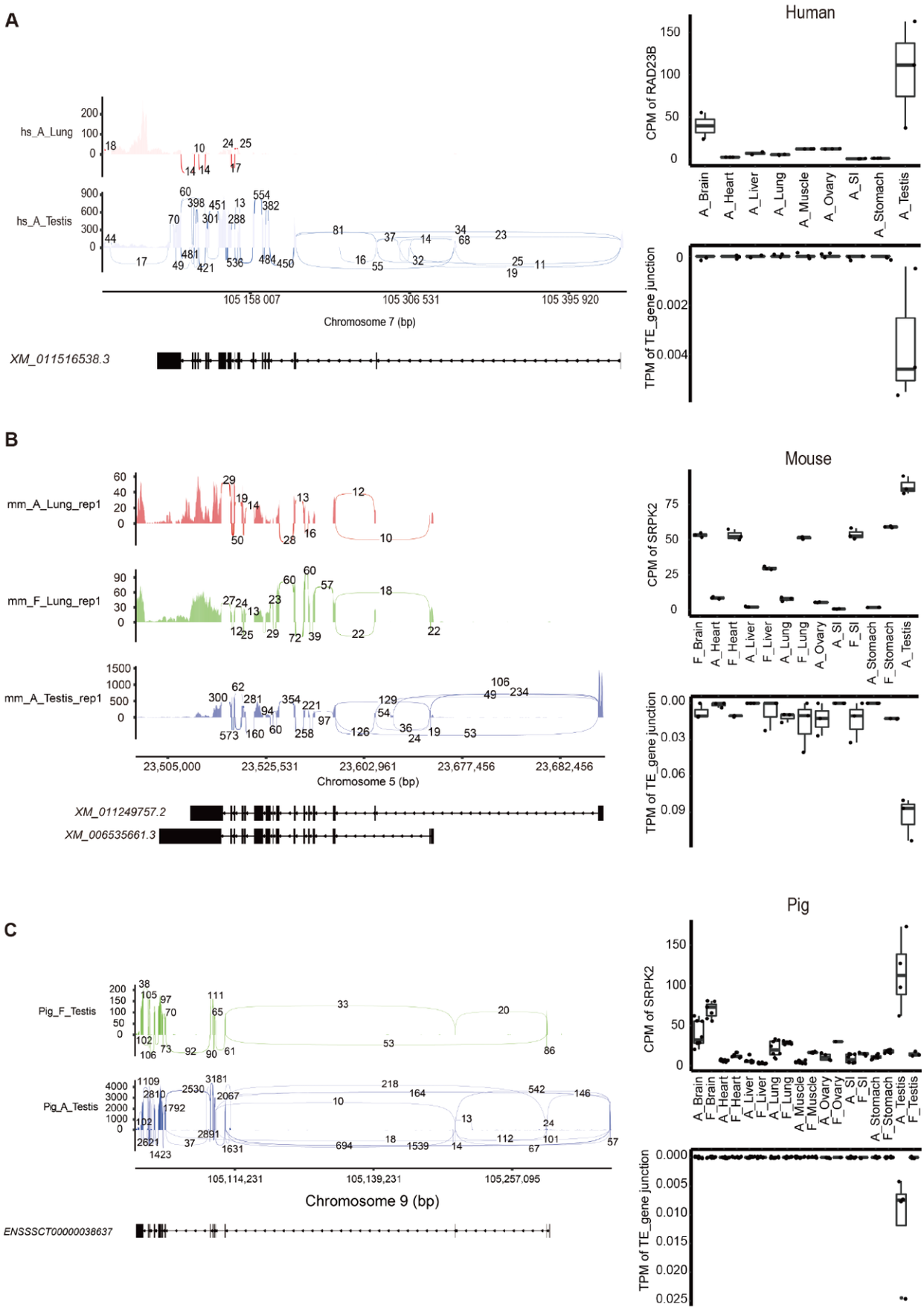


Figure 6 Retrotransposon promoters yield tissue-specific expression of *SRPK2* isoform across pigs, humans, and mice

A–C: Left, Sashimi plot showing RNA-seq reads spanning MLT1F1-SRPK2 junction in testis and lung from three species. MLT1F1 was located in promoter region of *SRPK2* in three species. Right, expression of *SRPK2* and MLT1F1-SRPK2 junction across stage-tissue groups in three species.

2022a).

Multiple lines of evidence suggest that specifically expressed TEs present in the genome can be recruited as promoter elements to reprogram host gene expression (Gardner et al., 2019; Miao et al., 2020). Similarly, we identified 351 stage- and tissue-specific TE-containing genes, with corresponding TEs functioning as alternative promoters. Most of these TE-derived promoter-driven stage- and tissue-specific genes were also expressed in the adult testes. To ensure the authenticity of these promoters, we screened 97 TE-derived alternative promoters supported by both H3K27ac and H3K4me3 peaks. For example, MIR acts as an adult testis-specific promoter, leading to testis-specific expression of the TE-containing *DLGAP5* transcript, a critical gene in cell cycle regulation with testis expression patterns (Liu et al., 2018). We also identified a TE-derived promoter in the adult liver and canonical promoters in the adult testes, both regulating the *GSTT4* gene, which is involved in glutathione metabolism. The observed enrichment of TF-binding motifs indicated that STAT2, an adult liver-specific gene, may also activate this promoter. This supports previous research showing that the same gene can transcribe different transcripts in different tissues depending on specific promoters (Lee et al., 2022). These findings further underscore the extensive recruitment of TEs in the regulation and expression of host genes (Chuong et al., 2017).

Further analysis was conducted to explore the conserved functional roles of TEs among pigs, humans, and mice. This analysis identified 31 homologous genes across these species, demonstrating TE involvement in their expression. Notably, MLT1F1 was found to regulate the expression of *SRPK2* in the adult testes. *SRPK2*, which was highly expressed in the testes, is associated with spermatogenesis (Giannakouros et al., 2011). These results suggest that the insertion of MLT1F1 into the genome likely occurred before the evolutionary divergence of humans, mice, and pigs, evolving into a crucial functional sequence across these species.

DATA AVAILABILITY

All epigenomic sequence data used in this study were submitted to the China National GeneBank DataBase (CNGBdb) with accession code: CNP0001696 (https://db.cngb.org/search/project/CNP0001696/), Genome Sequence Archive (GSA) database (https://ngdc.cncb.ac.cn/gsa/) under accession number CRA013785, Science Data Bank (doi:10.57760/sciencedb.j00139.00086) and NCBI under BioProjectID PRJNA1049515.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

L.S.H. designed and supervised this study and revised the manuscript. B.Y. wrote and revised the manuscript and supervised the data analysis; T.J. and Z.M.Z. performed the experiments, analyzed the data, and wrote the manuscript; Z.Q.L., Q.Z., Z.Z.W., J.W.Y., and S.Y.Y. performed the experiments. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We are grateful to colleagues from the State Key Laboratory of Swine Genetic Improvement and Production Technology, Jiangxi Agricultural University, for sample collection.

REFERENCES

Abuín JM, Pichel JC, Pena TF, et al. 2015. BigBWA: approaching the Burrows-Wheeler aligner to Big Data technologies. *Bioinformatics*, **31**(24): 4003−4005.

Ardeljan D, Taylor MS, Ting DT, et al. 2017. The human long interspersed element-1 retrotransposon: an emerging biomarker of neoplasia. *Clinical Chemistry*, **63**(4): 816−822.

Ayarpadikannan S, Kim HS. 2014. The impact of transposable elements in genome evolution and genetic instability and their implications in various diseases. *Genomics & Informatics*, **12**(3): 98−104.

Bailey TL, Johnson J, Grant CE, et al. 2015. The MEME suite. *Nucleic Acids Research*, **43**(W1): W39−W49.

Barski A, Cuddapah S, Cui KR, et al. 2007. High-resolution profiling of histone methylations in the human genome. *Cell*, **129**(4): 823−837.

Behbakht K, Qamar L, Aldridge CS, et al. 2007. Six1 overexpression in ovarian carcinoma causes resistance to TRAIL-mediated apoptosis and is associated with poor survival. *Cancer Research*, **67**(7): 3036−3042.

Bernstein BE, Kamal M, Lindblad-Toh K, et al. 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*, **120**(2): 169−181.

Bhalla N. 2020. Meiosis: is spermatogenesis stress an opportunity for evolutionary innovation?. *Current Biology*, **30**(24): R1471−R1473.

Biémont C. 2010. A brief history of the status of transposable elements: from junk DNA to major players in evolution. *Genetics*, **186**(4): 1085−1093.

Bourque G, Burns KH, Gehring M, et al. 2018. Ten things you should know about transposable elements. *Genome Biology*, **19**(1): 199.

Castañeda J, Genzor P, Bortvin A. 2011. piRNAs, transposon silencing, and germline genome integrity. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, **714**(1-2): 95−104.

Chang NC, Rovira Q, Wells J, et al. 2022. Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. *Genome Research*, **32**(7): 1408−1423.

Chen JQ, Zhang MP, Tong XK, et al. 2022. Scan of the endogenous retrovirus sequences across the swine genome and survey of their copy number variation and sequence diversity among various Chinese and Western pig breeds. *Zoological Research*, **43**(3): 423−441.

Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*, **351**(6277): 1083−1087.

Chuong EB, Elde NC, Feschotte C. 2017. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*, **18**(2): 71−86.

Colonna Romano N, Fanti L. 2022. Transposable elements: major players in shaping genomic and evolutionary patterns. *Cells*, **11**(6): 1048.

Cornelis G, Funk M, Vernochet C, et al. 2017. An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental *Mabuya* lizard. *Proceedings of the National Academy of Sciences of the United States of America*, **114**(51): E10991−E11000.

Cosby RL, Chang NC, Feschotte C. 2019. Host-transposon interactions: conflict, cooperation, and cooption. *Genes & Development*, **33**(17-18): 1098−1116.

Creyghton MP, Cheng AW, Welstead GG, et al. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(50): 21931−21936.

Delaneau O, Zazhytska M, Borel C, et al. 2019. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science*, **364**(6439): eaat8266.

Dobin A, Davis CA, Schlesinger F, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**(1): 15−21.

Duttke SHC, Lacadie SA, Ibrahim MM, et al. 2015. Human promoters are intrinsically directional. *Molecular Cell*, **57**(4): 674−684.

Fedoroff NV. 2012. Presidential address. Transposable elements, epigenetics, and genome evolution. *Science*, **338**(6108): 758−767.

Ferguson L, Agoulnik AI. 2013. Testicular cancer and cryptorchidism. *Frontiers in Endocrinology*, **4**: 32.

Fletcher J, Hu M, Berman Y, et al. 2005. Multicystic dysplastic kidney and variable phenotype in a family with a novel deletion mutation of *PAX2*. *Journal of the American Society of Nephrology*, **16**(9): 2754–2761.

Fueyo R, Judd J, Feschotte C, et al. 2022. Roles of transposable elements in the regulation of mammalian transcription. *Nature Reviews Molecular Cell Biology*, **23**(7): 481−497.

Gardner EJ, Prigmore E, Gallone G, et al. 2019. Contribution of retrotransposition to developmental disorders. *Nature Communications*, **10**(1): 4630.

Giannakouros T, Nikolakaki E, Mylonis I, et al. 2011. Serine-arginine protein kinases: a small protein kinase family with a large cellular presence. *The FEBS Journal*, **278**(4): 570−586.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**(7): 1017−1018.

Heinz S, Benner C, Spann N, et al. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, **38**(4): 576−589.

Holley CL, Topkara VK. 2011. An introduction to small non-coding RNAs: miRNA and snoRNA. *Cardiovascular Drugs and Therapy*, **25**(2): 151−159.

Howe FS, Fischl H, Murray SC, et al. 2017. Is H3K4me3 instructive for transcription activation?. *BioEssays*, **39**(1): 1−12.

Hughes AL, Kelley JR, Klose RJ. 2020. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation. *Biochimica et Biophysica Acta* (*BBA*) - *Gene Regulatory Mechanisms*, **1863**(8): 194567.

Jacques PÉ, Jeyakani J, Bourque G. 2013. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genetics*, **9**(5): e1003504.

Jin Y, Tam OH, Paniagua E, et al. 2015. TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*, **31**(22): 3593−3599.

Joly-Lopez Z, Bureau TE. 2018. Exaptation of transposable element coding sequences. *Current Opinion in Genetics & Development*, **49**: 34−42.

Judd J, Sanderson H, Feschotte C. 2021. Evolution of mouse circadian enhancers from transposable elements. *Genome Biology*, **22**(1): 193.

Kaessmann H. 2010. Origins, evolution, and phenotypic impact of new genes. *Genome Research*, **20**(10): 1313−1326.

Kapusta A, Kronenberg Z, Lynch VJ, et al. 2013. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genetics*, **9**(4): e1003470.

Karlić R, Chung HR, Lasserre J, et al. 2010. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(7): 2926−2931.

Kern C, Wang Y, Xu XQ, et al. 2021. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nature Communications*, **12**(1): 1821.

Kirk EP, Sunde M, Costa MW, et al. 2007. Mutations in cardiac T-box factor gene *TBX20* are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *The American Journal of Human Genetics*, **81**(2): 280−291.

Kryuchkova-Mostacci N, Robinson-Rechavi M. 2017. A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics*, **18**(2): 205−214.

Kunarso G, Chia NY, Jeyakani J, et al. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genetics*, **42**(7): 631−634.

Kurhanewicz NA, Dinwiddie D, Bush ZD, et al. 2020. Elevated temperatures cause transposon-associated DNA damage in *C. elegans* spermatocytes. *Current Biology*, **30**(24): 5007−5017.e4.

Lee HJ, Hou YR, Maeng JH, et al. 2022. Epigenomic analysis reveals prevalent contribution of transposable elements to *cis*-regulatory elements, tissue-specific expression, and alternative promoters in zebrafish. *Genome Research*, **32**(7): 1424−1436.

Li DR, Ai YW, Guo J, et al. 2020. Casein kinase 1G2 suppresses necroptosis-promoted testis aging by inhibiting receptor-interacting kinase 3. *eLife*, **9**: e61564.

Li H, Handsaker B, Wysoker A, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078−2079.

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, **30**(7): 923−930.

Liu JX, Liu Z, Zhang XZ, et al. 2019. Bioinformatic exploration of OLFML2B overexpression in gastric cancer base on multiple analyzing tools. *BMC Cancer*, **19**(1): 227.

Liu M, Qiu YL, Jin T, et al. 2018. Meta-analysis of microarray datasets identify several chromosome segregation-related cancer/testis genes potentially contributing to anaplastic thyroid carcinoma. *PeerJ*, **6**: e5822.

Lunney JK, Van Goor A, Walker KE, et al. 2021. Importance of the pig as a human biomedical model. *Science Translational Medicine*, **13**(621): eabd5758.

Miao BP, Fu SH, Lyu C, et al. 2020. Tissue-specific usage of transposable element-derived promoters in mouse development. *Genome Biology*, **21**(1): 255.

Modzelewski AJ, Shao WQ, Chen JQ, et al. 2021. A mouse-specific retrotransposon drives a conserved *Cdk2ap1* isoform essential for development. *Cell*, **184**(22): 5541−5558.e22.

Naville M, Warren IA, Haftek-Terreau Z, et al. 2016. Not so bad after all: retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. *Clinical Microbiology and Infection*, **22**(4): 312−323.

Niknafs YS, Pandian B, Iyer HK, et al. 2017. TACO produces robust multisample transcriptome assemblies from RNA-seq. *Nature Methods*, **14**(1): 68−70.

Pan ZY, Yao YL, Yin HW, et al. 2021. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nature Communications*, **12**(1): 5848.

Pekowska A, Benoukraf T, Zacarias-Cabeza J, et al. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *The EMBO Journal*, **30**(20): 4198−4210.

Pertea M, Pertea GM, Antonescu CM, et al. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**(3): 290−295.

Qian L, Mohapatra B, Akasaka T, et al. 2008. Transcription factor neuromancer/TBX20 is required for cardiac function in *Drosophila* with implications for human heart disease. *Proceedings of the National Academy of Sciences of the United States of America*, **105**(50): 19833−19838.

Quinlan AR. 2014. BEDTools: the swiss-army tool for genome feature analysis. *Current Protocols Bioinformatics*, **47**: 11.12.1−11.12.34.

Ramírez F, Dündar F, Diehl S, et al. 2014. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Research*, **42**(W1): W187−W191.

Rebollo R, Romanish MT, Mager DL. 2012. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annual Review of Genetics*, **46**: 21−42.

Reik W, Dean W, Walter J. 2001. Epigenetic reprogramming in mammalian

development. *Science*, **293**(5532): 1089−1093.

Roth SY, Denu JM, Allis CD. 2001. Histone acetyltransferases. *Annual Review of Biochemistry*, **70**: 81−120.

Safran M, Dalah I, Alexander J, et al. 2010. GeneCards Version 3: the human gene integrator. *Database*, **2010**: baq020.

Senft AD, Macfarlan TS. 2021. Transposable elements shape the evolution of mammalian development. *Nature Reviews Genetics*, **22**(11): 691−711.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, **8**(4): 272−285.

Sundaram V, Choudhary MNK, Pehrsson E, et al. 2017. Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nature Communications*, **8**(1): 14550.

Sundaram V, Wysocka J. 2020. Transposable elements as a potent source of diverse *cis*-regulatory sequences in mammalian genomes. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **375**(1795): 20190347.

Tarailo-Graovac M, Chen NS. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*,doi: 10.1002/0471250953.bi0410s25.

Ting CN, Rosenberg MP, Snow CM, et al. 1992. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes & Development*, **6**(8): 1457−1465.

Topaloğlu U, Akbalik ME, Sağsöz H. 2021. Immunolocalization of some HOX proteins in immature and mature feline testes. *Anatomia, Histologia, Embryologia*, **50**(4): 726−735.

Utomo ARH, Nikitin AY, Lee WH. 1999. Temporal, spatial, and cell type-specific control of Cre-mediated DNA recombination in transgenic mice. *Nature Biotechnology*, **17**(11): 1091−1096.

Wu YQ, Zhao H, Li YJ, et al. 2020. Genome-wide identification of imprinted genes in pigs and their different imprinting status compared with other mammals. *Zoological Research*, **41**(6): 721−725.

Xian H, Xian Y, Liu LL, et al. 2015. Expression of β-nerve growth factor and homeobox A10 in experimental cryptorchidism treated with exogenous nerve growth factor. *Molecular Medicine Reports*, **11**(4): 2875−2881.

Yang Y, Adeola AC, Xie HB, et al. 2018. Genomic and transcriptomic analyses reveal selection of genes for puberty in Bama Xiang pigs. *Zoological Research*, **39**(6): 424−430.

Zhan JF, Li JB, Wu YR, et al. 2021. Chromatin-associated protein sugp2 involved in mRNA alternative splicing during mouse spermatogenesis. *Frontiers in Veterinary Science*, **8**: 754021.

Zhang LK, Ma HD, Guo M, et al. 2022a. Dynamic transcriptional atlas of male germ cells during porcine puberty. *Zoological Research*, **43**(4): 600−603.

Zhang Q, Li J, Zhang YF, et al. 2022b. Whole-genome sequence-based association study for immune cells in an eight-breed pig heterogeneous population. *Journal of Genetics and Genomics*, **49**(11): 1068−1071.

Zhang Y, Liu T, Meyer CA, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, **9**(9): R137.

Zhang Y, Zhao B, Roy S, et al. 2016. microRNA-309 targets the Homeobox gene *SIX4* and controls ovarian development in the mosquito *Aedes aegypti*. *Proceedings of the National Academy of Sciences of the United States of America*, **113**(33): E4828–E4836.

Zhao YX, Hou Y, Xu YY, et al. 2021. A compendium and comparative epigenomics analysis of *cis*-regulatory elements in the pig genome. *Nature Communications*, **12**(1): 2217.

Zhong LP, Zheng M, Huang YZ, et al. 2023. An atlas of expression quantitative trait loci of microRNAs in *longissimus* muscle of eight-way crossbred pigs. *Journal of Genetics and Genomics*, **50**(6): 398−409.

Zhu YL, Zhou ZM, Huang T, et al. 2022. Mapping and analysis of a spatiotemporal H3K27ac and gene expression spectrum in pigs. *Science China Life Sciences*, **65**(8): 1517−1534.

Luo & Li. *Zool. Res.* 2024, 45(1): 152−159
https://doi.org/10.24272/j.issn.2095-8137.2023.364

ZR | Zoological Research

Article Open Access

Indian monsoon drove the dispersal of the *thoracica* group of *Scytodes* spitting spiders

Yu-Fa Luo[1], Shu-Qiang Li[2,3,*]

[1] *School of Life and Environmental Sciences*, *Shaoxing University*, *Shaoxing*, *Zhejiang* 312000, *China*
[2] *Key Laboratory of Zoological Systematics and Evolution*, *Institute of Zoology*, *Chinese Academy of Sciences*, *Beijing* 100101, *China*
[3] *Southeast Asia Biodiversity Research Institute*, *Chinese Academy of Sciences*, *Yezin*, *Nay Pyi Taw* 05282, *Myanmar*

ABSTRACT

We examined the global biogeography of the *Scytodes thoracica* group of spitting spiders based on 23 years of sampling at the species level (61 species in the *thoracica* group and 84 species of *Scytodes*) using DNA data from six loci. Our results indicated that the *thoracica* group initially dispersed from Southeast Asia to East Africa between 46.5 and 33.0 million years ago, and dispersal events intensified between Southeast/South Asia and East/South Africa from the early to late Miocene. The timing of these events indicates that Asian-African faunal exchange of the *thoracica* group was driven by the Indian monsoon, and the pattern of dispersal suggests that colonialization took root when the Indian monsoon shifted from a North-South direction to an East-West direction from the middle Eocene.

Keywords: Geological event; Climate change; Ballooning organism; Faunal exchange; Species distribution

INTRODUCTION

The Indian monsoon constitutes an important component of the Asian meteorological system. Extensive research has indicated that it originated during the Paleocene to Eocene (~66–34 million years ago (Ma); Licht et al., 2014; Tardif et al., 2020; Wu et al., 2022; Zhang et al., 2018) and prevailed throughout the Eocene (Huber & Goldner, 2012). Uplift of the Himalayan-Xizang Plateau (HTP) has led to substantial variation in the Indian monsoon, including directional shifts and intensification events (An, 2000; An et al., 2001; Ruddiman & Kutzbach, 1990). These monsoonal changes may have facilitated long-distance dispersal of various organisms, thereby influencing diversification and faunal exchange, including those of spider populations. Here, we investigated whether the emergence and subsequent changes of the Indian monsoon are concordant with biogeographical data pertaining to the *Scytodes thoracica* species group of spitting spiders, which are globally distributed but primarily occur in Asia.

Ballooning dispersal is a well-observed behavior in certain spider species (Morley & Robert, 2018) and is prevalent among various families such as Araneidae, Linyphiidae, Lycosidae, Oxyopidae, Philodromidae, Theridiidae, Tetragnathidae, and Scytodidae (Blandenier, 2009; Dean & Sterling, 1985; Garb & Gillespie, 2009; Greenstone, 1982; Kuntner & Agnarsson, 2011; Lee et al., 2015; Luo et al., 2020; Luo & Li, 2023; Weyman et al., 2002). For example, the golden orb spider (*Nephila*) exhibits the capacity for long-range aerial dispersal, from Africa to islands in the Indian Ocean (Kuntner & Agnarsson, 2011), which is evidently dictated by meteorological conditions, such as wind direction and intensity during monsoon periods. Consequently, this raises questions regarding the influence of monsoonal directional shifts and intensification on the patterns of faunal exchange across continents.

The globally distributed spitting spiders, belonging to the family Scytodidae, encompass four genera and 240 species, including 220 species of *Scytodes*, three species of *Scyloxes*, four species of *Dictis*, and 13 species of *Stedocys* (World Spider Catalog, 2023). The genus *Scytodes* is diverse and broadly distributed across Asia, Europe, Africa, the Americas, and Australia (World Spider Catalog, 2023). Our phylogenetic analyses have identified three distinct species groups within *Scytodes*: the *fusca* group, *thoracica* group, and Himalayas group (Luo & Li, 2022; this study, Supplementary Figure S1). Paleontological evidence suggests that the genus originated during the Paleogene (~58–33.9 Ma; Magalhaes et al., 2020; Wunderlich, 1993, 2011), with considerable diversification in the late Cenozoic (Luo & Li, 2022). Regionally, tectonic events have been important drivers for the diversification of *Stedocys* cave spiders in Southeast Asia (Luo & Li, 2018), as well as the high-elevation Himalayas group of *Scytodes* spiders around the HTP (Luo & Li, 2022). Globally, long-distance dispersal

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 22 November 2023; Accepted: 05 December 2023; Online: 06 December 2023

Foundation items: This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences to S.Q.L. (XDB31000000) and National Natural Sciences Foundation of China to Y.F.L. (NSFC-31860602, 32170463, 31660611)

*Corresponding author, E-mail: lisq@ioz.ac.cn

has played a crucial role in the expansion and diversification of the *Scytodes fusca* species complex (Luo & Li, 2023). However, the spatiotemporal distributions of the cosmopolitan *thoracica* group, as well as the mechanisms driving their distribution, remain unexplored.

In the present study, we used genetic data to generate the phylogenies of the *Scytodes* spitting spiders and dated molecular phylogenies of the *thoracica* group to investigate the composition, phylogenetic relationships, and geographic distribution patterns of the major *Scytodes* lineages over their entire distribution. We also examined global biogeographical patterns of this group and the impact of the Indian monsoon on *Scytodes thoracica* group dispersal during the Cenozoic.

MATERIALS AND METHODS

Taxon sampling and sequence analyses

Scytodes spiders were intensively sampled in Asia, Europe, the Americas, and East Africa between 2007 and 2020 (Supplementary Tables S1–S3). Taxon names, specimen vouchers, localities, longitude-latitude coordinates, and GenBank accession numbers of the *thoracica* group from Asia, Europe, Africa, and the Americas are listed in Supplementary Table S1. Samples were fixed in 95% ethanol and stored at −20 °C. Based on morphological examination, we identified 76 *Scytodes* species, including 61 species belonging to the *thoracica* group. DNA was extracted using a TIANamp Genomic DNA kit (DP304-02). Polymerase chain reaction (PCR) amplification and gene sequencing were conducted following Luo & Li (2018). All primers are provided in Supplementary Table S4. Four nuclear genes (18S, 28S, *H3*, and ITS2) and two mitochondrial genes (*COI* and 16S) were sequenced. Voucher specimens were deposited in the Institute of Zoology, Chinese Academy of Sciences, Beijing, China. Additional sequences were obtained from GenBank.

DNA sequences were checked for quality and edited using BioEdit (Hall, 1999) and SEQUENCHER v.4.1.2 (Gene Codes, USA). Protein-coding gene (*H3* and *COI*) sequences were aligned using CLUSTAL X (Jeanmougin et al., 1998) with default parameters. Alignments of the ribosomal genes (18S, 28S, 16S, and ITS2) were produced in MAFFT v.7 (http://mafft.cbrc.jp/alignment/server/) using the E-INS-i method (Katoh & Standley, 2013). DnaSP v.5 (Librado & Rozas, 2009) was used to identify haplotypes. The concatenated sequences of the six genes were used for phylogenetic analyses, divergence time estimations, and biogeographical reconstructions.

Phylogenetic analysis

Initially, a molecular phylogeny of all globally sampled *Scytodes* spiders was constructed using maximum-likelihood (ML). Subsequently, molecular phylogenies of the *thoracica* group were reconstructed using both ML and Bayesian inference (BI) approaches. Outgroups were based on previous phylogenetic studies (Wheeler et al., 2017). We included the three recognized species of *Scytodes* sister genera, *Stedocys* (*S. ludiyanensis* and *S. shilinensis*) and *Scyloxes* (*S. zhaoi*) and three other species from two closely related families: Dysderidae (*Dysdera silvatica*) and Sicariidae (*Loxosceles laeta* and *L. rufescens*). The ML analyses were conducted using IQ-TREE v.2.0.6 (Nguyen et al., 2015). For each gene partition or codon base, the most suitable evolutionary model across all available models (Supplementary Table S5) (Kalyaanamoorthy et al., 2017) was identified using IQ-TREE, including the FreeRate model (+R; Soubrier et al., 2012), which relaxes the assumption of gamma-distributed rates using the MFP+MERGE function. The Shimodaira-Hasegawa approximate likelihood ratio test (SH-aLRT; Guindon et al., 2010) and 1 000 ultra-fast bootstrap replicates (UFBoot; Hoang et al., 2018) were used to assess branch support. The "-bnni" command was added to improve the search for each UFBoot replicate tree. Transfer bootstrap expectation (TBE; Lemoine et al., 2018) was also conducted in IQ-TREE using the function --tbe following 500 standard Felsenstein bootstraps. Bayesian analysis was performed for the *thoracica* group in MrBayes v.3.1.2 (Ronquist & Huelsenbeck, 2003). The best-fitting substitution model for each partition (Supplementary Table S6) was selected by jModelTest (Posada, 2008) under the Akaike information criterion. The Markov chain Monte Carlo (MCMC) simulation was run for 10 million generations with sampling every 100 generations using unlinked parameters between partitions. Tracer v.1.5 (Rambaut & Drummond, 2009) was employed to verify the effective sample size (ESS) values (>200) for each parameter. A 50% majority rule consensus tree was computed after discarding the first 25% of trees as burn-in.

Divergence time estimation

Molecular dating of the *thoracica* group phylogeny was calculated using BEAST v.1.6.1 (Drummond & Rambaut, 2007). First, the combined DNA sequence data were partitioned by locus and codon position. For each partition, the corresponding evolutionary model (GTR+I+G for 16S, 18S, 28S, ITS2, *COI*-1, *COI*-2, and *H3*-1; GTR+G for *COI*-3 and *H3*-3; JC for *H3*-2) was selected from jModelTest. Divergence time estimation was performed using an uncorrelated lognormal relaxed molecular clock. Analyses with a Birth-death and Yule prior were also repeated to assess the sensitivity of the results to tree prior specification. The Birth-death model outperformed the Yule model according to the marginal likelihood estimated by Stepping-stone sampling. Hence, for subsequent analyses, we focused on the chronogram resulting from the Birth-death prior. Minimum ages based on the fossils of *Scytodes weitschati* (Scytodidae, 33.9 Ma; Wunderlich, 1993, 2011), *Loxosceles defecta* (Sicariidae, 13.82 Ma; Wunderlich, 1988, 2004), *Harpactea communis* (Dysderidae, 35 Ma; Wunderlich, 2004), and *Microsegestria poinari* (Segestriidae, Dysderoidea, 125 Ma; Wunderlich & Milki, 2004) were used as calibration points (Supplementary Table S7; Donoghue & Benton, 2007; Renner, 2005). BEAST analyses were run for 50 million MCMC generations with model parameters and tree sampling every 1 000 generations. The stationarity of runs was examined using Tracer v1.6 with an ESS cutoff of >200 for all parameters. TreeAnnotator v.1.6.1 (in BEAST) was used to produce a maximum clade credibility (MCC) tree after discarding 25% of the trees as burn-in.

Biogeographical reconstruction

The ancestral areas and biogeographical history of the *thoracica* group were inferred using three analytical approaches implemented in RASP v.3.0 (Yu et al., 2010, 2015). Parsimony-based statistical dispersal-vicariance analysis (S-DIVA) (Yu et al., 2010), a modified form of dispersal-vicariance analysis (Ronquist, 1997), accounts for uncertainty in phylogenetic and biogeographical reconstructions (Nylander et al., 2008). The dispersal-

extinction-cladogenesis (DEC) model, which infers dispersal-mediated range expansion and extinction-mediated range contraction (Ree & Smith, 2008), allows for analytical solutions to probabilities of range transitions as a function of time, enabling free model parameters, rates of dispersal, and local extinction to be estimated by ML. The Bayesian binary MCMC (BBM) model is considered the most general and complex model in RASP (Sanmartín et al., 2001). These three analyses were conducted with the maximum area number at each node set to 2. A fossil-calibrated tree, from which the outgroups were trimmed, was employed for biogeographical reconstruction. Based on geographic divisions, climate, and current distributions of *Scytodes*, five geographical areas occupied by the *thoracica* group were delimited, namely the southern Asia/Oriental Realm, Central Asia+Europe/Palearctic, East Africa, South Africa, and the Americas.

RESULTS

Sequence characteristics

A total of 1 174 nuclear (28S: 183, 18S: 238, *H3*: 213 and ITS2: 127) and mitochondrial DNA sequences (*COI*: 223 and 16S: 190) were successfully generated from the global sampling of *Scytodes* (Supplementary Tables S1–S3), including 814 nuclear (28S: 134, 18S: 172, *H3*: 134 and ITS2: 65) and mitochondrial DNA sequences (*COI*: 153 and 16S: 156) of the *thoracica* group from Asia, Europe, the Americas, and Africa (Supplementary Table S1). The aligned sequence data of the DNA haplotypes of the *thoracica* group and outgroups were 3 363 bp long. GenBank accession numbers are provided in Supplementary Tables S1–S3.

Molecular phylogeny

The ML trees for all sampled *Scytodes* spiders are presented in Supplementary Figure S1, and the phylogenetic results are consistent with our previous study (Luo & Li, 2022). The reconstructed tree topologies of the *thoracica* group based on ML and BI analyses were largely congruent (Figure 1). Results indicated that the *thoracica* group comprised four major clades: (1) Clade I (MA), including eight species from the Malay Archipelago, South Africa, and Madagascar; (2) Clade II (EC), composed of five species from Central Asia to Europe and Canada; (3) Clade III (SC), containing eight species from the southern USA, Central America, and China; and (4) Clade IV (CI), including 33 species from China, the Indochina Peninsula, South Asia, the Malay Archipelago, and the Seychelles (Figures 1, 2). Clade IV comprised three subclades: (1) Clade IV-1, including 11 species from Southwest China to Indochina; (2) Clade IV-2, containing three species (including type species *S. thoracica*) distributed from South China eastward to the Seychelles, including Southeast Asia and South Asia; and (3) Clade IV-3, including 19 species distributed in South Asia and Southeast Asia (Figures 1, 2).

Divergence times

The time-calibrated Bayesian phylogeny is shown in Figure 3A. The initial cladogenesis within the *thoracica* group was inferred to have occurred slowly among the four major lineages between ~46.5–12.9 Ma. The diversification of Clade I started ~46.5 Ma, after the split of the most recent common ancestor (MRCA) of the group at ~54.4 Ma. Clade II originated in the early Eocene (~50.3 Ma), and the diversification of this clade occurred ~12.9 Ma across Asia, Europe, and North America, but slowly diversified ~11.2–4.7 Ma during the late Miocene to the present. Clade II (NH, North Himalayas-Xizang) and Clades III and IV (SH, South Himalayas-Xizang) split ~50.3 Ma. The most recent speciation of Clade III occurred in China, with local species emerging ~40.2–9.5 Ma. Within Clade IV, Clade IV-1 diverged from its sister lineages Clades IV-2 and IV-3 ~42.2 Ma.

Biogeography

The current distribution of the *thoracica* group was associated with long-distance dispersal events and *in situ* allopatric speciation (Figure 4). Both the DEC (Figure 4A) and S-DIVA results (Figure 4B) were generally concordant. In the BBM analysis (Figure 4C), nodes 1, 2, 3, 6, and 15 differed from those of DEC (Figure 4A) and S-DIVA (Figure 4B).

Three possible ancestral ranges (Figure 4B) were inferred for the *thoracica* group with the same marginal probability using S-DIVA: (1) southern Asia; (2) southern Asia+Central Asia+Europe, and (3) southern Asia+East Africa+South Africa. The inference from DEC (Figure 4A) favored southern Asia+Central Asia+Europe and southern Asia+East Africa+South Africa as the most likely ancestral area of the *thoracica* group, whereas southern Asia was preferred under BBM (Figure 4C).

The inferred phylogeographic history of the *thoracica* group is shown in Figure 4. The initial burst of diversification that produced the four early-diverging lineages (Clades I–IV) was inferred to have occurred in the ancestral area during the late Paleocene and early Eocene. Long-distance dispersal and geographic isolation were the main drivers of the global distribution of the *thoracica* group. We inferred long-distance overwater dispersals from the southern Malay Peninsula and central Sumatra Island region to Madagascar and South Africa between the mid-Eocene and late Miocene, which resulted in Clade I. Certain ancestors from Southeast Asia expanded northwards to Central Asia, with subsequent geographical isolation between northern and southern Himalayas-Xizang due to the Indian-Eurasian collision and uplift of the HTP, which resulted in Clade II. The Clade II spiders expanded to Europe during the late Miocene and early Pliocene. Within Clade III, we inferred one independent overwater dispersal toward the Americas from China, starting in the mid-Eocene. The Clade IV ancestors diversified mainly in Southeast Asia from the late Eocene to present, while some Clade IV spiders expanded to South Asia and the Seychelles.

DISCUSSION

Encompassing 61 species of the *thoracica* group, this study sought to illuminate the global diversity of the group and phylogenetic relationships among its major clades. Our previous study suggested that the *Scytodes* spiders from Southeast Asia comprise three monophyletic species groups, corresponding to the Himalayas group, *fusca* group, and *thoracica* group (Luo & Li, 2022). In the current study, even after the addition of samples from Africa, Europe, the Americas, and other regions of Asia, our results reaffirmed the same lineage compositions and group-level phylogenetic relationships of *Scytodes* from Southeast Asia, thereby expanding our understanding of their phylogenetic history at a global level.

Phylogeography of *thoracica* group

This study sheds light on the phylogeography of the *thoracica*

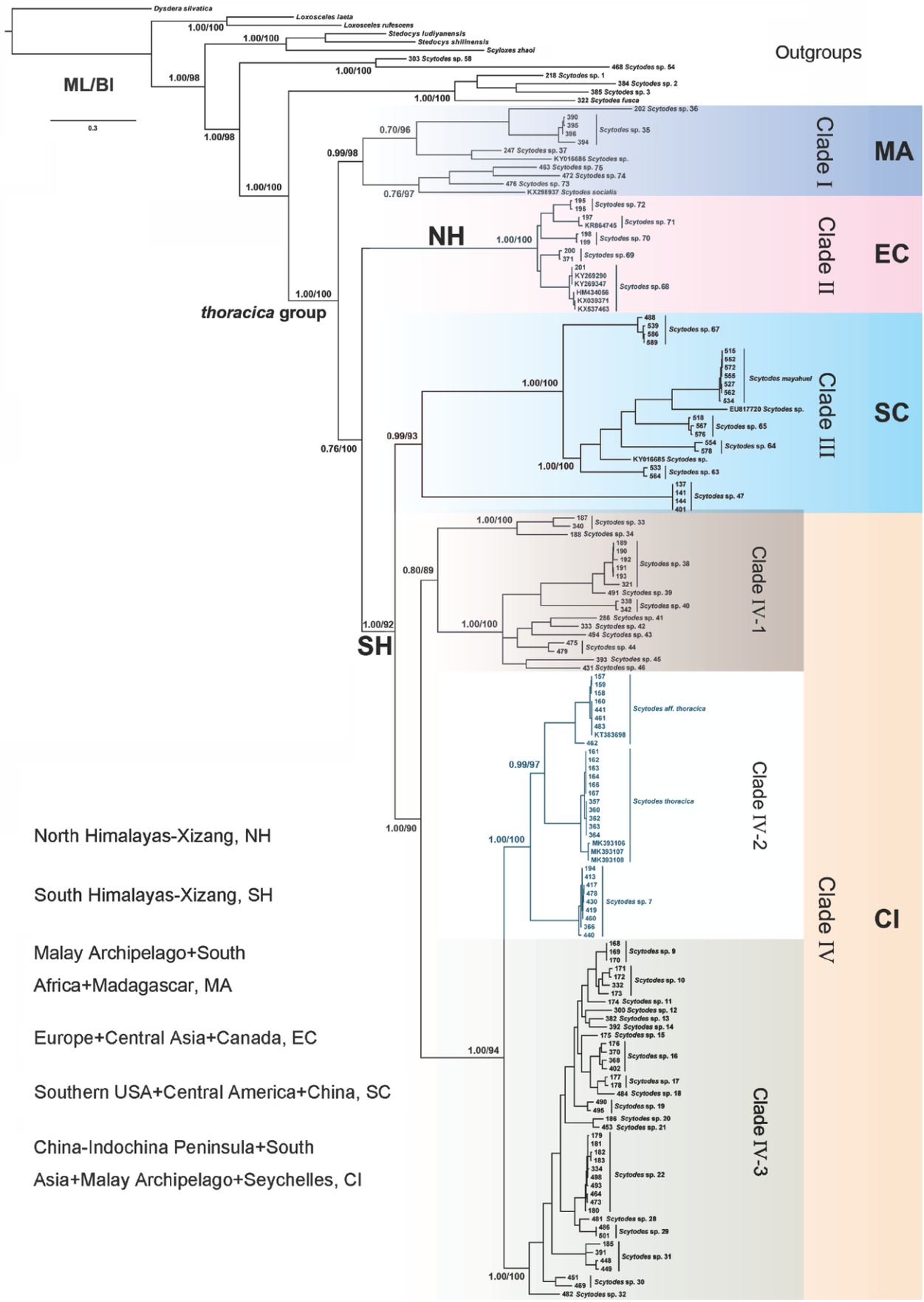


Figure 1 Summary phylogenetic tree based on ML and BI analyses of concatenated mitochondrial (*COI* and 16S) and nuclear (18S, 28S, ITS2, and *H3*) gene datasets

Numbers at nodes represent bootstrap support values/posterior probabilities from both analyses, respectively.

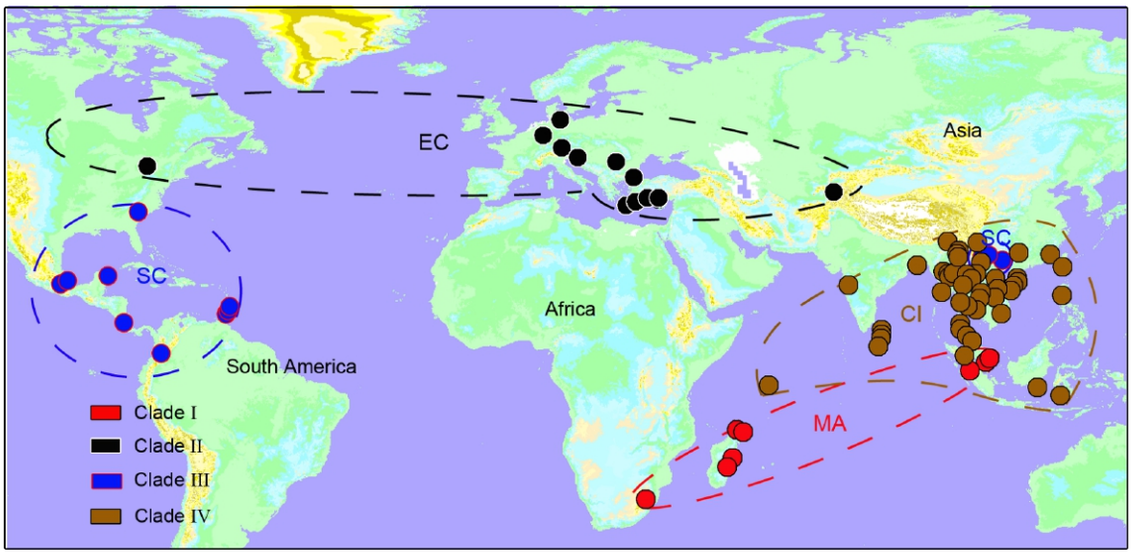


Figure 2 Sampling localities of *thoracica* group

Thoracica group included Clade I (red circles; MA, Malay Archipelago+South Africa+Madagascar), Clade II (black circles; EC, Europe+Central Asia+Canada), Clade III (blue circles; SC, southern USA+Central America+China), and Clade IV (brown circles; CI, China+Indochina peninsula+South Asia+Malay Archipelago+Seychelles) individuals. Detailed sampling information is provided in the Supplementary Materials.

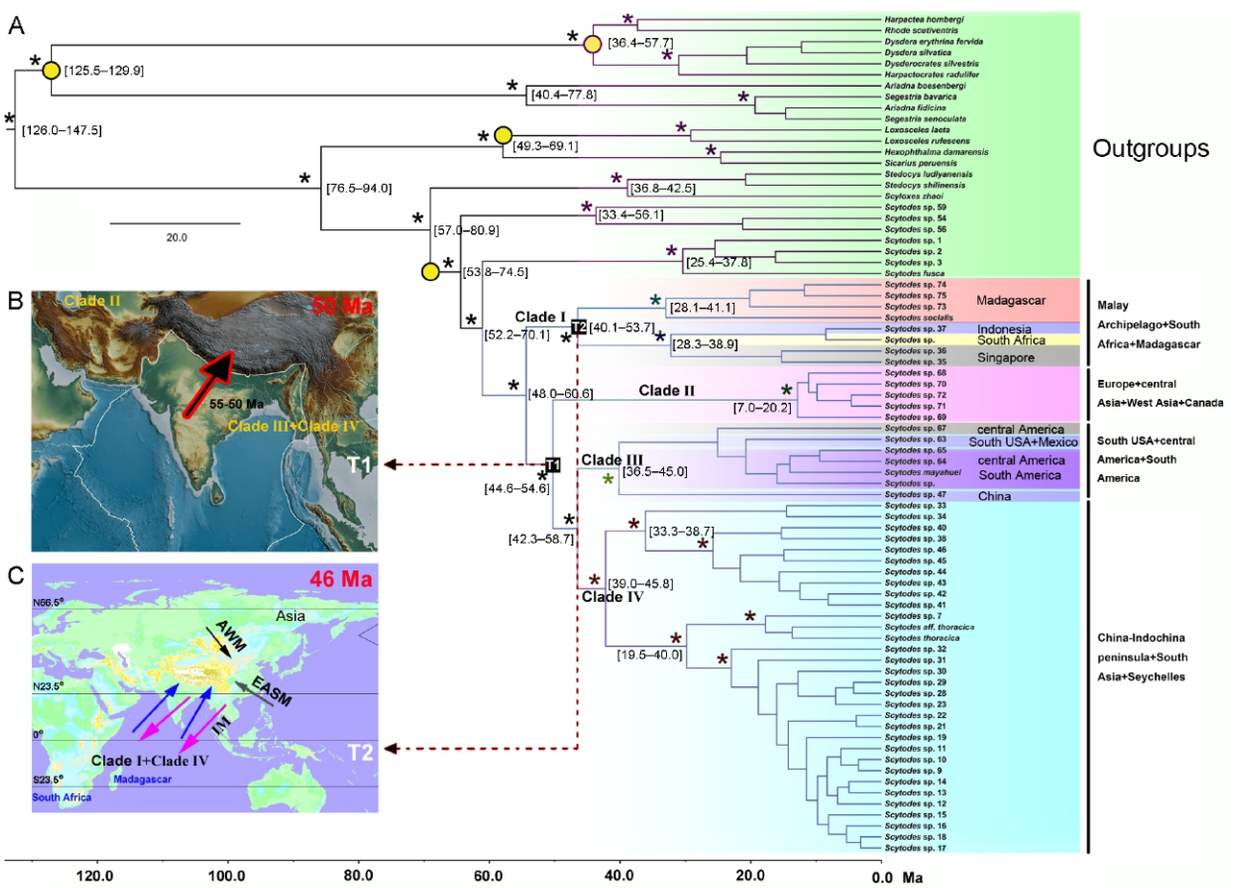


Figure 3 Evolutionary diversification of *thoracica* group with initial Indian-Asian collision and Asian monsoon

A: Time-tree of all sampled *thoracica* group species. "*" indicates stable branches with Bayesian support >0.90. Yellow circles are fossil calibration nodes. "T1" represents Clade II+Clade III+Clade IV; "T2" represents Clade I. Ma: Millions of years ago. B: Evolutionary history of *thoracica* group shown in A corresponds to tectonic hypothesis that initial Indian-Eurasian collision occurred during the early Eocene. C: Evolutionary history of *thoracica* group shown in A corresponds to climate hypothesis that Indian monsoon occurred during the middle Eocene because of the Indian-Eurasian collision. Clade I: MA, Malay Archipelago+South Africa+Madagascar; Clade II: EC, Europe+Central Asia+Canada; Clade III: SC, southern USA+Central America+China; Clade IV: CI, China+Indochina Peninsula+South Asia+Malay Archipelago+Seychelles. AWM: Asian winter monsoon; EASM: East Asian summer monsoon; IM: Indian monsoon. Numbers in "[ ]" at nodes indicate 95% highest posterior density interval for divergence time estimates.

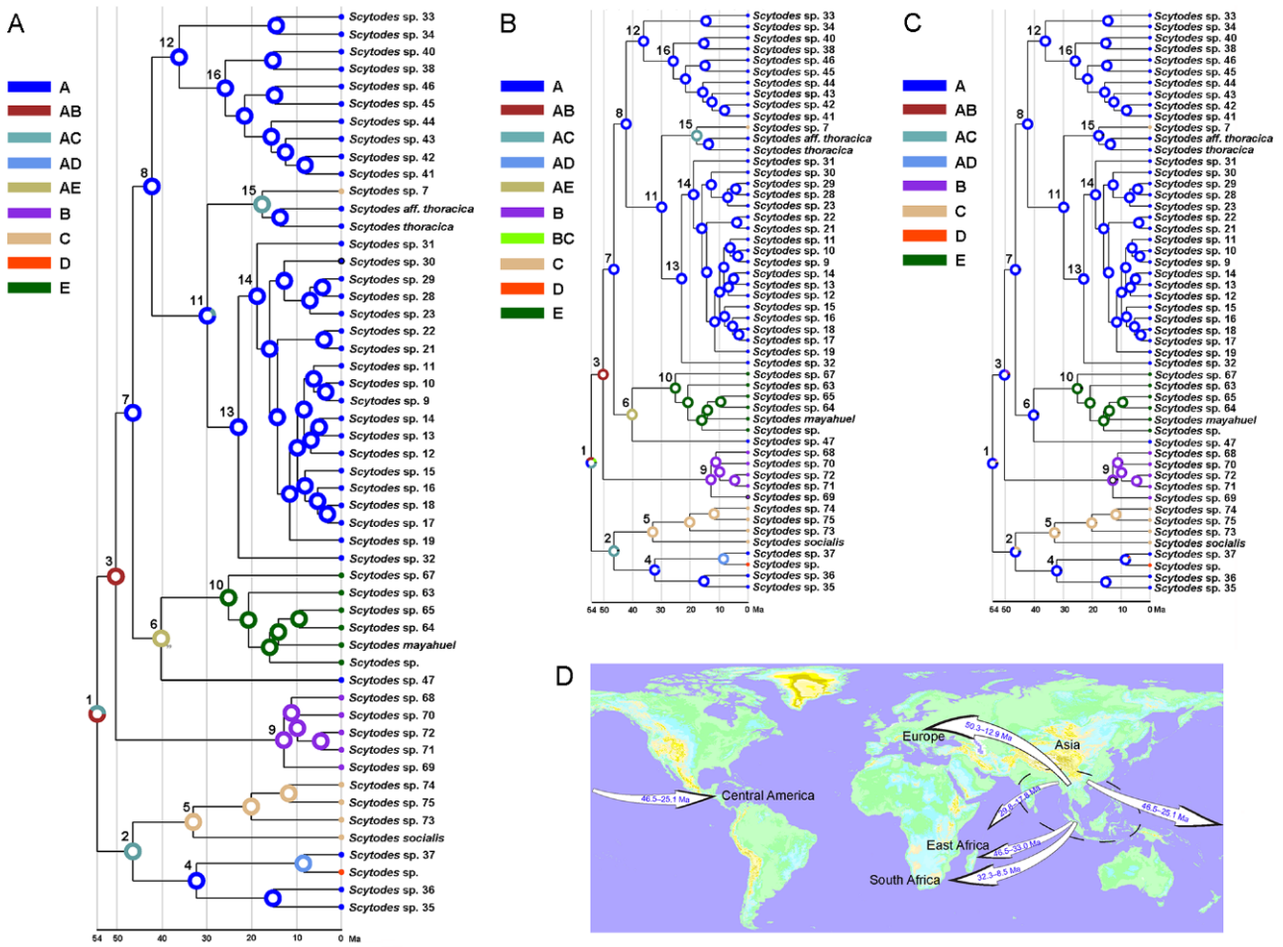


Figure 4 Biogeography of *thoracica* group and probable global dispersion routes

A–C: Reconstruction from Bayesian binary MCMC (BBM), statistical dispersal-vicariance analysis (S-DIVA), and dispersal-extinction-cladogenesis (DEC), respectively, from RASP v3.0. Colors of pie wedges at each node represent geographical areas inferred to have been occupied by ancestral taxa (A: Southern Asia/Oriental Realm; B: Central Asia+Europe/Palearctic; C: East Africa; D: South Africa; E: Americas). D: Global potential dispersion routes (arrows) of *thoracica* group.

group. Our results indicated that the group most likely originated in southern Asia around the Paleocene (~61.1 Ma) when it diverged from the *fusca* group (Figure 4). However, the initial divergence of the extant species within the *thoracica* group began at the Paleocene-Eocene boundary, when four major extant lineages diverged in succession, with rapid diversification occurring mainly in Southeast and South Asia during the Miocene (Figures 3, 4). Furthermore, our results suggested that the initial Indian-Eurasian collision and subsequent uplift of the Himalayas drove the deep lineage divergence of the *thoracica* group between the northern and southern Himalayas. The initial collision event triggered the split between Clade II and Clade III+Clade IV within the pan-Himalayas *Scytodes* group in the early Eocene (~50.3 Ma). These changes are consistent with other studies identifying the Indian-Eurasian collision as a major driving force for the phylogeography, speciation, and diversification of animals (Favre et al., 2015; Guillot & Replumaz, 2013; Licht et al., 2014; Miao et al., 2012).

Concordance of initiation time between *thoracica* group dispersal and development of Indian monsoon into South and East Africa

The phylogeographical analyses conducted in this study indicated that the Indian monsoon drove dispersal of the *thoracica* group, with long-distance dispersal shaping its global distribution patterns. Our results also suggested that after the split of the MRCA of the *thoracica* group in the mid-Miocene (~54.4 Ma), Clade I dispersed by ballooning over the oceans, with the aid of the historical Indian monsoon from the Malay Archipelago to Madagascar, possibly via islands, then diverged between the two regions ~46.5 Ma (Figure 4). The timing (~46.5 Ma) of the initial dispersal of the *thoracica* group to Madagascar is concordant with that of the development of the Indian monsoon into Southeast Africa. This pattern is consistent with previous studies suggesting that the onset of the Indian monsoon during the middle Eocene acted as a significant driver of faunal exchange (Favre et al., 2015; Guillot & Replumaz, 2013; Licht et al., 2014; Miao et al., 2012). In the case of spiders, this was likely aided by ballooning dispersal, which allowed the colonization of oceanic islands and continents far from their native ranges (Kuntner & Agnarsson, 2011; Meijer, 1977; Sugg & Edwards, 1998).

Faunal exchange of the *thoracica* group between Asia and Africa correlates with the shift in Indian monsoon direction

Our results also revealed three dispersal events from Southeast Asia to Africa within the *thoracica* group (Figure 4). The timelines of these dispersals coincided with shifts in the

directionality of the Indian monsoon from North-South to East-West from the middle Eocene due to uplift of the HTP. These findings further enhance our understanding of how the complex interplay of prehistoric geological events (such as uplift of the HTP and development of the Indian monsoon) has shaped the evolution of Asian biodiversity, thus adding a new dimension to the vast body of knowledge already established (Clift & Webb, 2018; He et al., 2022; Licht et al., 2014; Prell & Kutzbach, 1992; Spicer, 2017).

DATA AVAILABILITY

All sequences generated in this study have been deposited in GenBank (accession numbers in Supplementary Tables S1, S2). All information on sampling locations is provided in Supplementary Tables S1–S3. The sequence alignment datasets of phylogeny and molecular clock analyses are publicly available on the Dryad database (online at http://datadryad.org/).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y.F.L. and S.Q.L. designed the study. Y.F.L. carried out laboratory protocols and data analysis. Both authors drafted the manuscript and read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The manuscript benefited greatly from comments by Ingi Agnarsson, Sarah Crews, and Joseph Koh. We thank Ingi Agnarsson, Guo Zheng, Yan-Feng Tong, Jie Liu, Yu-Cheng Lin, Zhi-Yuan Yao, Chun-Xia Wang, Qing-Yuan Zhao, Zhe Zhao, Hui-Feng Zhao, and Francesco Ballarin for specimens.

REFERENCES

An ZS. 2000. The history and variability of the East Asian paleomonsoon climate. *Quaternary Science Reviews*, **19**(1-5): 171−187.

An ZS, Kutzbach JE, Prell WL, et al. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. *Nature*, **411**(6833): 62−66.

Blandenier G. 2009. Ballooning of spiders (Araneae) in Switzerland: general results from an eleven-year survey. *Bulletin of the British Arachnological Society*, **14**(7): 308−316.

Clift PD, Webb AG. 2018. A history of the Asian monsoon and its interactions with solid Earth tectonics in Cenozoic South Asia. *In*: Treloar PJ, Searle MP. Himalayan Tectonics: A Modern Synthesis. London: Geological Society, London, Special Publications.

Dean DA, Sterling WL. 1985. Size and phenology of ballooning spiders at two locations in eastern Texas. *Journal of Arachnology*, **13**(1): 111−120.

Donoghue PCJ, Benton MJ. 2007. Rocks and clocks: calibrating the tree of life using fossils and molecules. *Trends in Ecology & Evolution*, **22**(8): 424−431.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*, **7**: 214.

Favre A, Päckert M, Pauls SU, et al. 2015. The role of the uplift of the Qinghai–Tibetan Plateau for the evolution of Tibetan biotas. *Biological Reviews*, **90**(1): 236−253.

Garb JE, Gillespie RG. 2009. Diversity despite dispersal: colonization history and phylogeography of Hawaiian crab spiders inferred from multilocus genetic data. *Molecular Ecology*, **18**(8): 1746−1764.

Greenstone MH. 1982. Ballooning frequency and habitat predictability in two wolf spider species (Lycosidae: *Pardosa*). *The Florida Entomologist*, **65**(1): 83−89.

Guillot S, Replumaz A. 2013. Importance of continental subductions for the growth of the Tibetan Plateau. *Bulletin de la Société Géologique de France*, **184**(3): 199−223.

Guindon S, Dufayard JF, Lefort V, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**(3): 307−321.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**: 95−98.

He SL, Ding L, Xiong ZY, et al. 2022. A distinctive Eocene Asian monsoon and modern biodiversity resulted from the rise of eastern Tibet. *Science Bulletin*, **67**(21): 2245−2258.

Hoang DT, Chernomor O, von Haeseler A, et al. 2018. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, **35**(2): 518−522.

Huber M, Goldner A. 2012. Eocene monsoons. *Journal of Asian Earth Sciences*, **44**: 3−23.

Jeanmougin F, Thompson JD, Gouy M, et al. 1998. Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences*, **23**(10): 403−405.

Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**(6): 587−589.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, **30**(4): 772−780.

Kuntner M, Agnarsson I. 2011. Phylogeography of a successful aerial disperser: the golden orb spider *Nephila* on Indian Ocean islands. *BMC Evolutionary Biology*, **11**: 119.

Lee VMJ, Kuntner M, Li DQ. 2015. Ballooning behavior in the golden orbweb spider *Nephila pilipes* (Araneae: Nephilidae). *Frontiers in Ecology and Evolution*, **3**: 2.

Lemoine F, Domelevo Entfellner JB, Wilkinson E, et al. 2018. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, **556**(7702): 452−456.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**(11): 1451−1452.

Licht A, van Cappelle M, Abels HA, et al. 2014. Asian monsoons in a late Eocene greenhouse world. *Nature*, **513**(7519): 501−506.

Luo YF, Goh SP, Li DQ, et al. 2020. Global diversification of *Anelosimus* spiders driven by long-distance overwater dispersal and Neogene climate oscillations. *Systematic Biology*, **69**(6): 1122−1136.

Luo YF, Li SQ. 2018. Cave *Stedocys* spitting spiders illuminate the history of the Himalayas and southeast Asia. *Ecography*, **41**(2): 414−423.

Luo YF, Li SQ. 2022. The stepwise Indian-Eurasian collision and uplift of the Himalayan-Tibetan Plateau drove the diversification of high-elevation *Scytodes* spiders. *Cladistics*, **38**(5): 582−594.

Luo YF, Li SQ. 2023. Global expansion of a solitary-social tropical spitting spider shaped by multiple long-distance dispersals. *Ecography*, **2023**(3): e06632.

Magalhaes ILF, Azevedo GHF, Michalik P, et al. 2020. The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic. *Biological Reviews*, **95**(1): 184−217.

Meijer J. 1977. The immigration of spiders (Araneida) into a new polder. *Ecological Entomology*, **2**(1): 81−90.

Miao YF, Herrmann M, Wu FL, et al. 2012. What controlled Mid-Late Miocene long-term aridification in Central Asia? Global cooling or Tibetan Plateau uplift: a review. *Earth-Science Reviews*, **112**(3-4): 155−172.

Morley EL, Robert D. 2018. Electric fields elicit ballooning in spiders. *Current Biology*, **28**(14): 2324−2330.e2.

Nguyen LT, Schmidt HA, von Haeseler A, et al. 2015. IQ-TREE: a fast and

effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**(1): 268−274.

Nylander JAA, Olsson U, Alström P, et al. 2008. Accounting for phylogenetic uncertainty in biogeography: a Bayesian approach to dispersal-vicariance analysis of the thrushes (Aves: *Turdus*). *Systematic Biology*, **57**(2): 257−268.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, **25**(7): 1253−1256.

Prell WL, Kutzbach JE. 1992. Sensitivity of the Indian Monsoon to forcing parameters and implications for its evolution. *Nature*, **360**(6405): 647−652.

Rambaut A, Drummond AJ. 2009. Tracer v1.5. http://beast.bio.ed.ac.uk/Tracer.

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology*, **57**(1): 4−14.

Renner SS. 2005. Relaxed molecular clocks for dating historical plant dispersal events. *Trends in Plant Science*, **10**(11): 550−558.

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. *Systematic Biology*, **46**(1): 195−203.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**(12): 1572−1574.

Ruddiman WF, Kutzbach JE. 1990. Late Cenozoic plateau uplift and climate change. *Transactions of the Royal Society of Edinburgh:Earth Sciences*, **81**(4): 301−314.

Sanmartín I, Enghoff H, Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. *Biological Journal of the Linnean Society*, **73**(4): 345−390.

Soubrier J, Steel M, Lee MSY, et al. 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Molecular Biology and Evolution*, **29**(11): 3345−3358.

Spicer RA. 2017. Tibet, the Himalaya, Asian monsoons and biodiversity–In what ways are they related?. *Plant Diversity*, **39**(5): 233−244.

Sugg PM, Edwards JS. 1998. Pioneer Aeolian community development on pyroclastic flows after the eruption of Mount St. Helens, Washington, U. S. A. *Arctic and Alpine Research*, **30**(4): 400−407.

Tardif D, Fluteau F, Donnadieu Y, et al. 2020. The origin of Asian monsoons: a modelling perspective. *Climate of the Past*, **16**(3): 847−865.

Weyman GS, Sunderland KD, Jepson PC. 2002. A review of the evolution and mechanisms of ballooning by spiders inhabiting arable farmland. *Ethology Ecology & Evolution*, **14**(4): 307−326.

Wheeler WC, Coddington JA, Crowley LM, et al. 2017. The spider tree of life: phylogeny of Araneae based on target-gene analyses from an extensive taxon sampling. *Cladistics*, **33**(6): 574−616.

World Spider Catalog. [2023-08-10]. World spider catalog. version 24.5. Natural History Museum Bern,http://wsc.nmbe.ch.

Wu FL, Fang XM, Yang YB, et al. 2022. Reorganization of Asian climate in relation to Tibetan Plateau uplift. *Nature Reviews Earth & Environment*, **3**(10): 684−700.

Wunderlich J. 1993. Die ersten fossilen Speispinnen (Fam. Scytodidae) im Baltischen Bernstein (Arachnida: Araneae). *Mitteilungen aus dem Geologisch-Palä ontologischen Institute der University Hamburg*, **75**: 243−247.

Wunderlich J. 2011. Some fossil spiders (Araneae) in Eocene European ambers. *Beiträ ge Araneology*, **6**: 472−538.

Yu Y, Harris AJ, Blair C, et al. 2015. RASP (reconstruct ancestral state in phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution*, **87**: 46−49.

Yu Y, Harris AJ, He XJ. 2010. S-DIVA (Statistical Dispersal-Vicariance Analysis): a tool for inferring biogeographic histories. *Molecular Phylogenetics and Evolution*, **56**(2): 848−850.

Zhang R, Jiang DB, Ramstein G, et al. 2018. Changes in Tibetan Plateau latitude as an important factor for understanding East Asian climate since the Eocene: a modeling study. *Earth and Planetary Science Letters*, **484**: 295−308.

Li et al. *Zool. Res.* 2024, 45(1): 160−175
https://doi.org/10.24272/j.issn.2095-8137.2023.169

ZR | Zoological Research

Article Open Access

Comparative analyses of mitogenomes in the social bees with insights into evolution of long inverted repeats in the Meliponini

Yu-Ran Li[1,2,4], Zheng-Wei Wang[2], Richard T. Corlett[1,*], Wen-Bin Yu[1,3,*]

[1] *Yunnan Key Laboratory for the Conservation of Tropical Rainforests and Asian Elephants & Center for Integrative Conservation*, *Xishuangbanna Tropical Botanical Garden*, *Chinese Academy of Sciences*, *Mengla*, *Yunnan* 666303, *China*

[2] *CAS Key Laboratory of Tropical Forest Ecology*, *Xishuangbanna Tropical Botanical Garden*, *Chinese Academy of Sciences*, *Mengla*, *Yunnan* 666303, *China*

[3] *Southeast Asia Biodiversity Research Institute*, *Chinese Academy of Sciences*, *Mengla*, *Yunnan* 666303, *China*

[4] *University of Chinese Academy of Sciences*, *Shijingshan District*, *Beijing* 100049, *China*

ABSTRACT

The insect mitogenome is typically a compact circular molecule with highly conserved gene contents. Nonetheless, mitogenome structural variations have been reported in specific taxa, and gene rearrangements, usually the tRNAs, occur in different lineages. Because synapomorphies of mitogenome organizations can provide information for phylogenetic inferences, comparative analyses of mitogenomes have been given increasing attention. However, most studies use a very few species to represent the whole genus, tribe, family, or even order, overlooking potential variations at lower taxonomic levels, which might lead to some incorrect inferences. To provide new insights into mitogenome organizations and their implications for phylogenetic inference, this study conducted comparative analyses for mitogenomes of three social bee tribes (Meliponini, Bombini, and Apini) based on the phylogenetic framework with denser taxonomic sampling at the species and population levels. Comparative analyses revealed that mitogenomes of Apini and Bombini are the typical type, while those of Meliponini show diverse variations in mitogenome sizes and organizations. Large inverted repeats (IRs) cause significant gene rearrangements of protein coding genes (PCGs) and rRNAs in Indo-Malay/Australian stingless bee species. Molecular evolution analyses showed that the lineage with IRs have lower d_N/d_S ratios for PCGs than lineages without IRs, indicating potential effects of IRs on the evolution of mitochondrial genes. The finding of IRs and different patterns of gene rearrangements suggested that Meliponini is a hotspot in mitogenome evolution. Unlike conserved PCGs and rRNAs whose rearrangements were found only in the mentioned lineages within Meliponini, tRNA rearrangements are common across all three tribes of social bees, and are significant even at the species level, indicating that comprehensive sampling is needed to fully understand the patterns of tRNA rearrangements, and their implications for phylogenetic inference.

Keywords: Social bees; Phylogeny; Mitogenome structure; Gene rearrangement; Inverted repeats

INTRODUCTION

The mitochondria are fundamental organelles in almost all eukaryotic cells with the function of energy conversion (Dowling & Wolff, 2023). As organelles of endosymbiotic origin, they contain their own genomic sequences, the mitogenomes, which are typically maternally inherited (Sato & Sato, 2012). Although mitochondria share the common ancestral origin as Alphaproteobacteria living within ancestral eukaryotes, mitogenome organizations show high diversity across eukaryotic lineages in both genome structures and gene rearrangements (Shtolz & Mishmar, 2023). Six major types of mitogenome structures have been recognized in all living organisms (Kolesnikov & Gerasimov, 2012), and insect mitogenomes correspond to type Ⅰ, with a compact circular molecule of 15–18 kb (Cameron, 2014). The gene content is

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 31 October 2023; Accepted: 18 December 2023; Online: 23 December 2023

Foundation items: This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB31000000), Science and Technology Basic Resources Investigation Program of China (2021FY100200), Yunnan Revitalization Talent Support Program “Young Talent” and "Innovation Team" Projects, and the 14th Five-Year Plan of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Science (XTBG-1450101)

*Corresponding authors, E-mail: corlett@xtbg.org.cn; yuwenbin@xtbg.ac.cn

highly conserved, with 37 unique genes, including 13 protein coding genes (PCGs), two ribosomal RNA genes (rRNAs), and 22 transfer RNA genes (tRNAs) (Boore, 1999). There is also a control region, which is assumed to be the largest non-coding segment within the mitogenome (Boore, 1999). This region is also called the D-loop region or the A+T rich region, and it has been shown to contain regulatory elements for the replication and expression of the mitogenome (Saito et al., 2005). Although containing five conserved elements, the control region is difficult to annotate, because their positions and structures are hypervariable across different lineages, and their lengths are variant as a result of variations of the repeat sequences (Dickey et al., 2015; Quinn & Mindell, 1996; Quinn & Wilson, 1993; Wei et al., 2010; Zhang et al., 1995; Zhang & Hewitt, 1997).

Since the first mitogenome of *Drosophila yakuba* was reported nearly 40 years ago (Clary & Wolstenholme, 1985), the number of insect mitogenomes has increased rapidly to more than 3 256 accessions by 2022 (https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/). Comparative studies of insect mitogenomes have revealed that structural variations and gene arrangements occur across lineages, although their organizations are generally conserved as the "normal" type. Major structural variations have been found in insect mitogenomes of some lineages including: (1) duplication of PCGs in the species *Microchorista philpotti* (Nannochoristidae) (Beckenbach, 2011), or inverted duplication of the nearly whole mitogenome or circular amphimeric mitogenome in two Australian stingless bees of *Tetragonula* (Hymenoptera) (Françoso et al., 2023); (2) multiple fragmented mitogenomes in order Thysanoptera (Dickey et al., 2015), Psocoptera (Shi et al., 2016), and Phthiraptera (Sweet et al., 2022); (3) gene or gene segment translocations or rearrangements in order Hemiptera (Li et al., 2012), Phthiraptera (Sweet et al., 2021), Psocoptera (Li et al., 2013), and Hymenoptera (de Paula Freitas et al., 2020; Oliveira et al., 2008; Xiao et al., 2011; Zheng et al., 2018), with massive rearrangements in *Tetragonula* spp. (Françoso et al., 2023; Wang et al. 2022), *Lepidotrigona* spp. (Wang et al., 2020, 2021). The circular amphimeric genome, two inverted repeats (IRs) separated by two single copy regions (SCs), is the typical feature of the chloroplast genome in land plants (Wicke et al., 2011), but is rather unusual in animal mitogenomes (Rayko, 1997; Kolesnikov & Gerasimov, 2012). Amphimeric genomes have structural isomers, which differ only in the orientation of the SC regions, and high frequency intramolecular recombination between two IRs could mediate structural heteroplasmy in the chloroplast genome of plants (Stein et al., 1986; Wang & Lanfear, 2019).

Hymenoptera is estimated to outnumber Coleoptera as the most species-rich insect order (Forbes et al., 2018), comprising 109 recognized extant families including species commonly known as sawflies, wasps, bees, and ants (Blaimer et al., 2023). Mitogenomes of investigated Hymenopteran lineages show high rates of gene rearrangements, which provide both opportunities and potential pitfalls for exploring homologous rearrangements and their implications in phylogenetic and evolutionary inferences (Dowton & Austin, 1999; Dowton et al., 2009). Most previous comparative analyses of mitogenomes were carried out with one or a few species to represent the whole genus, tribe, family, or even order (Dowton & Austin, 1999; Dowton et al., 2003, 2009; He et al., 2018; Silvestre & Arias, 2006; Wei et al., 2010; Zheng et al., 2018). This kind of sampling strategy could overlook mitogenome variations within family, tribe, genus, or species level, which may lead to some incorrect inferences. In this study, we focused on three social bee tribes, Apini (honeybees), Meliponini (stingless bees) and Bombini (bumblebees), to investigate mitogenome variations and their phylogenetic signature with extensive sampling at the species and population levels.

These three tribes and orchid bees (Euglossini) form a monophyletic group of (Euglossini+(Apini+(Bombini+Meliponini))) in the Apidae (Almeida et al., 2023). Species within this monophyletic group are known as "corbiculate bees" because, except for some cleptobiotic species in Meliponini (Melo, 2020) and Bombini (Lhomme & Hines, 2019), female workers of three social bee tribes and Euglossini females possess corbicula at the metatibia to carry plant resources. Taxa within these three tribes are iconic around the world for their roles in pollination and economic production, as well as their social behavior (Engel, 2023; Grüter, 2020; Michener, 2007). Stingless bees and honeybees are highly eusocial, living in perennial colonies and having two female castes, queens and workers (Roubik, 1989). Bumblebees are also eusocial, but their colonies are smaller and generally last less than a year, with only the newly mated young queen surviving to the next spring by hibernation and starting a new colony on its own (Michener, 2007). Among the three tribes, a large number of mitogenomes from Apini and Bombini have been decoded and show the typical type of mitogenome. By contrast, the mitogenomes of Meliponini have been less investigated (de Paula Freitas et al., 2020; Wang et al., 2020, 2021, 2022). Recent studies showed massive mitogenome variations in gene rearrangements and structural reorganizations in Meliponini. However, these studies only sampled five or fewer taxa of Meliponini for comparative analyses, so the evolutionary trends of mitogenome reorganizations within Meliponni have not been well understood. Bombini have shown pervasive tRNA rearrangements across lineages, which are synapomorphic among bumblebee subgenera, and are also regarded as an ideal group for mitogenome evolution (Gonçalves et al., 2023).

In this study, we *de novo* assembled 73 new mitogenomes and downloaded 81 previously reported mitogenomes (representing 9 honeybees, 40 bumblebees, and 23 stingless bee species) to reconstruct the phylogeny of three tribes of social bees and the orchid bees. Based on the phylogenetic framework, this study aimed to: (1) explore the occurrence of the IR regions in mitogenomic reorganizations within Meliponini, and to gain new insights into the evolutionary pattern; (2) compare the gene rearrangements across the three tribes of social bees to understand their evolutionary patterns at lower taxonomic levels, and their implications for phylogenetic inference.

MATERIALS AND METHODS

Bee samplings, DNA extraction and sequencing

In total, 154 samples were included in this study, including 39 samples of nine Apini species, 50 samples of 42 Bombini species, and 65 samples of 23 Meliponini species (Tables 1, 2). Of those, 40 samples representing eight Meliponini species were newly sampled in Yunnan, Southwest China. Sampling sites are provided in Supplementary Table S1. Morphological and biological information for seven species were described in

Table 1 Information and features of mitogenomes assembled from SRA data or sequencing data by this study

Taxa	Voucher/SRA	Assessment of assembled mitogenomes		Genome size (bp)	IR (bp)	tRNA	GC content (%)
		Matched	Customized error rate				
Tribe Bombini							
Genus *Bombus*							
B. balteatus	SRR13788759	1474.15±229.72	0.0221±.1932	16871		+Met	13.0
B. balteatus	SRR13788762	1608.51±577.79	0.0162±.1437	17017		+Met	13.1
B. bohemicus	SRR12770612	1858.69±5291.58	0.0315±0.4228	20984		-Ser1 -Glu -Arg -Asn +Met	16.7
B. bifarius	SRR11404314	2250.03±934.00	0.0411±.3299	16714		22	14.2
B. bifarius	SRR12691508	1309.23±2389.85	0.0306±0.3136	16883		22	14.3
B. convexus	SRR12770609	3793.03±808.55	0.0315±.3398	16329		22	19.6
B. cullumanus	SRR12528003	285.93±15.82	0.0391±.4256	17418		22	15.7
B. impetuosus	SRR12770615	1356.22±349.93	0.0139±.1911	15767		22	13.8
B. laesus	SRR12770614	1270.73±93.00	0.0079±.0769	15732		22	12.6
B. lantschouensis	SRR12770617	952.42±18.92	0.0022±.0087	16431		+Met	14.0
B. melanopygus	SRR8700077	509.63±31.53	0.0144±.1477	17236		22	13.0
B. polaris	SRR12527998	1435.15±689.73	0.0172±0.1292	16374		-Met -Ile -Ala	12.8
B. sichelii	SRR12770616	1560.37±519.39	0.0017±0.0129	15675		-His	14.7
B. superbus	SRR12528014	2109.69±678.51	0.0060±0.0246	15738		22	15.8
B. sylvicola	SRR12691583	1358.14±06.26	0.0149±.0637	21701		+Met	11.1
B. terricola	SRR7696607	748.41±27.77	0.0123±.0930	16652		-Ala -Met	13.9
B. trifasciatus	SRR12770611	1043.36±28.48	0.0047±.0282	15871		-Met	14.6
B. turneri	SRR12528020	1410.97±63.59	0.0050±.0094	15689		-Ala	12.8
B. vancouverensis nearcticus	SRR11407481	1543.40±23.30	0.0146±.0497	19608		22	14.9
Tribe Meliponini							
Genus *Frieseomelitta*							
F. varia	SRR10065869	880.69±17.88	0.0107±.0167	15260		22	12.1
F. varia	SRR13310985	796.42±88.00	0.0167±.0275	15405		22	11.9
Genus *Nannotrigona*							
N. testaceicornis	SRR5651514	68.07±71.31	0.0443±0.131189202	15257		-Tyr -Lys -Ala -Gln -Met -Ile	13.7
Genus *Melipona*							
M. bicolor	SRR8735007	107.56±2.88	0.0068±.0114	15018		22	13.1
M. quadrifasciata	SRR1945415	742.44±77.06	0.0058±.0475	15246		22	12.0
M. variegatipes	ERR3255838	651.37±82.90	0.0044±.0066	15589		22	12.0
Genus *Heterotrigona*							
H. itama	ERR4276786	802.15±0.14	0.0043±.0041	26714	11275	22	23.7
Genus *Lepidotrigona*							
L. sp.	JHMW2	1199.01±29.44	0.0028±.0086	26566	10991	22	25.6
L. flavibasis	GLTMS1	1311.63±26.24	0.0023±.0050	26609	10821	22	22.3
L. flavibasis	HNBS1	413.78±69.90	0.0060±.0060	26599	10923	22	22.4
L. flavibasis	HNQZ1	620.67±34.07	0.0049±.0043	26588	10911	22	22.3
L. flavibasis	JHDDG4	1273.04±42.81	0.0025±.0036	26596	10920	22	22.3
L. flavibasis	LCCY9	1485.43±29.71	0.0050±.0030	26515	10881	22	22.2
L. flavibasis	MLBB1	884.29±64.70	0.0032±.0041	26600	10905	22	22.4
L. flavibasis	MLML6	2359.40±098.15	0.0018±.0019	26568	10904	22	22.4
L. flavibasis	MLMX13	1019.11±245.34	0.0035±0.0037	26595	10932	22	22.4
L. flavibasis	MLXTBG_R2	538.07±40.88	0.0035±.0048	26606	10938	22	22.4
L. flavibasis	MLYXC4	794.46±84.07	0.0040±.0068	26579	10898	22	22.4
L. flavibasis	MWDGP2	2373.33±93.75	0.0027±.0021	26589	10911	22	22.4
L. flavibasis	MWHTC19	2807.48±460.46	0.0037±.0095	26582	10912	22	22.3
L. terminata	GLTMS3	1346.40±94.36	0.0021±.0079	26626	11026	22	24.3
L. terminata	JHJH1	1134.30±74.75	0.0037±.0069	26628	11027	22	24.4
L. terminata	JHJN	1540.82±27.70	0.0020±.0071	29351	14310	22	24.0
L. terminata	MLMX15	1274.53±516.02	0.0018±0.0071	26630	11016	22	24.3
L. terminata	MLXTBG_L2	691.75±76.59	0.0040±.0079	26628	11020	22	24.3

Continued

Taxa	Voucher/SRA	Assessment of assembled mitogenomes		Genome size (bp)	IR (bp)	tRNA	GC content (%)
		Matched	Customized error rate				
L. terminata	MLYXC3	872.83±08.89	0.0020±.0082	26635	11028	22	24.4
Genus *Lisotrigona*							
L. carpenteri	YXYJ	580.58±76.25	0.0135±.0197	15543		22	11.6
Genus *Plebeina*							
P. hildedrandti	SRR5406061	27.19±4.16	0.0142±.0486	15165		22	13.3
Genus *Tetragonula*							
Subgenus *Tetragonilla*							
T. collina	MAMB2	2937.58±624.05	0.0040±.0198	22329	4681/ 1751	-Glu	20.4
T. collina	MHGDN13	2288.60±363.42	0.0035±.0191	22746	4746/ 1751	-Glu	20.5
T. collina	MHMB1_1	1157.15±73.49	0.0030±.0147	22337	4746/ 1751	-Glu	20.4
T. collina	MHMB1_3	1159.44±433.42	0.0026±0.0117	22337	4746/ 1751	-Glu	20.4
T. collina	MLYXC1_1	1460.65±015.20	0.0032±.0143	22321	4647/ 1750	-Glu	20.5
T. collina	MWHTC14	8323.51±970.48	0.0052±.0171	22335	4746/ 1751	-Gly	20.5
Subgenus *Tetragonula*							
T. carbonaria	SRR10389202	1936.93±620.20	0.0012±0.0357	30639	15309	-Glu	28.0
T. clypearis	SRR10411628	3232.83±340.39	0.0106±0.0163	24802	9373	-Gly	25.8
T. davenporti	SRR10405219	828.91±14.51	0.0030±.0049	30844	15268	-Glu	29.8
T. gressitti	JHMW12	2559.05±1319.9	0.0019±0.0083	24080	8297	-Glu	20.1
T. gressitti	LCCY6	2175.10±211.76	0.0040±.0067	24012	8247	-Gly	20.0
T. gressitti	LCY4	3582.91±177.19	0.0020±.0068	24012	18247	-Gly	20.0
T. gressitti	MWDGP1	822.28±63.19	0.0033±.0087	24077	8312	-Glu	20.1
T. gressitti	PEJC1	917.47±08.55	0.0055±.0069	24080	8093	-Glu	20.1
T. hockingsi	SRR10390715	1899.08±95.12	0.0011±.0018	30410	14992	-Glu	27.3
T. laeviceps	GLTMS2	998.11±01.15	0.0018±.0043	27607	12103	-Glu	25.9
T. laeviceps	GLTMS5	417.79±89.29	0.0018±.0047	29200	13659	-Gly	26.0
T. laeviceps	MLBB4	1044.40±73.04	0.0015±.0047	27699	12097	-Glu	25.9
T. laeviceps	MLML2	455.27±15.46	0.0023±.0047	27699	12097	-Glu	25.9
T. laeviceps	MLMX1	3298.87±230.10	0.0023±.0065	27693	12094	-Glu	25.9
T. laeviceps	MLXTBG1	1353.48±526.10	0.0049±0.0050	27719	12107	-Glu	25.9
T. mellipes	SRR10395358	550.55±7.60	0.0088±.0119	28477	12993	-Gly	27.6
T. mellipes	SRR10426324	1468.56±02.37	0.0063±.006	30828	15284	-Gly	28.0
T. pagdeni	LCCY4	1014.69±00.70	0.0045±.0061	23522	8342	-Glu	25.3
T. pagdeni	LCY2	2418.86±285.90	0.0025±.0077	23522	8342	-Gly	25.3
T. pagdeni	LCZK1	701.59±30.75	0.0051±.0075	23522	8342	-Gly	25.3
Tribe Euglossini							
Genus *Eufriesea*							
E. mexicana	SRR1945065	779.66.59±4523.87	0.0559±1.1037	15532			

There are typically 22 tRNAs which are identified by the one-letter for the corresponding amino acid according to the Rules (1968). 18 tRNAs each specify a single amino acid, while 2 tRNAs (Ser1 and Ser2) specify serine, and 2 tRNAs (Leu1 and Leu2) specify leucine. The "+" and "-" indicate the duplication or loss of tRNAs, and the number inside the brackets indicate the number of duplication or loss of tRNAs. The "IR" column shows the length of a single inverted repeat (IR) region.

a previous study (Li et al., 2021). To extend the sampling size and verify the accuracy of some published mitogenome sequences, we additionally downloaded the high throughput sequencing raw data of 19 Bombini samples (representing 18 species), and 14 Meliponini samples (representing 12 species) from the Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). In addition, we also downloaded 39 Apini (representing nine species), 31 Bombini (representing 26 species), and 11 Meliponini (representing nine species) mitogenome sequences from the National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/).

For the newly sampled Meliponini, genomic DNA was extracted using the standard CTAB protocol (Hunt, 1997) from a single individual after removing the head and abdomen. The purified genomic DNA was used to prepare a sequencing library with an average insert size of 400 bp using Illumina TruSeq™ DNA Sample Prep Kit (USA), then sequenced by the Illumina NovaSeq 6000 platform (USA) to generate 4 Gb of 150 bp paired-end reads.

Mitogenome assembly and annotation

The cleaned raw reads were used to *de novo* assemble mitogenomes using the GetOrganelle toolkit (Jin et al., 2018).

Table 2 Information and features of mitogenomes downloaded from NCBI

Taxa	GenBank accession No.	Genome size (bp)	IR (bp)	tRNA	GC content (%)
Tribe Apini					
Genus *Apis*					
A. andreniformis	AP018490	16694		22	14.3
A. andreniformis	NC039709	16694		22	14.3
A. cerana	AP017314	15916		22	15.9
A. cerana	AP018450	15885		22	16.0
A. cerana japonica	AP017314	15917		22	15.9
A. dorsata	AP018369	15279		22	15.3
A. dorsata	NC037709	15892		22	15.0
A. florea	AP018491	17693		+Ser1(2)	13.7
A. florea	KC170303	15993		+Ser1(2)	14.8
A. florea	NC021401	17694		+Ser1(2)	13.9
A. koschevenikovi	KY348372	16049		+Met(1)	20.9
A. koschevenikovi	AP017643	15277		+Met(2)	16.0
A. laboriosa	NC036155	15510		22	15.0
A. laboriosa	KX908208	15621		22	15.1
A. mellifera adansonii	MN585109	16466		22	15.2
A. mellifera anatoliaca	MT188686	16256		22	15.4
A. mellifera capensis	MG552691	16515		22	17.7
A. mellifera carnica	MN250878	16358		22	17.1
A. mellifera carpatica	AP018403	16336		22	15.1
A. mellifera caucasica	AP018404	16341		22	14.5
A. mellifera caucasica	MN714160	16274		22	15.2
A. mellifera iberiensis	MN585110	16560		22	20.1
A. mellifera intermissa	KY926883	16343		22	14.9
A. mellifera jemenitica	MN714161	16427		22	15.4
A. mellifera lamarckii	KY464958	16589		22	15.1
A. mellifera ligustica	MT859135	16467		22	15.0
A. mellifera meda	KY464957	16248		22	15.5
A. mellifera monticola	MF678581	16343		22	15.2
A. mellifera ruttneri	MN714162	16577		22	15.6
A. mellifera sahariensis	MF351881	16569		22	15.2
A. mellifera scutellata	MG552698	16479		22	15.2
A. mellifera simensis	MN585108	16523		22	15.1
A. mellifera sinisxinyuan	MN733955	16886		22	14.8
A. mellifera syriaca	KY926882	16343		22	15.1
A. mellifera unicolor	MN119925	16373		22	15.4
A. nigrocincta	AP018398	15516		22	15.8
A. nigrocincta	NC038114	15855		22	15.4
A. nuluensis	AP018157	15921		22	15.5
A. nuluensis	MF565375	15843		22	16.1
Tribe Bombini					
Genus *Bombus*					
B. asiaticus	MH998259	15676		22	14.8
B. breviceps	MF478986	16014		22	16.4
B. campestris	HG995151	15573		22	12.8
B. canariensis	MW959771	17300		+Leu2	13.8
B. consobrinus	MF995069	16177		22	14.1
B. difficillimus	MZ352140	19072		+Met(2)	13.8
B. filchnerae	NC082116	18553		22	11.3
B. florilegus	AP018158	15763		+Met	14.4
B. haemorrhoidalis	MZ352141	15334		+Met(2)	14.5
B. hypocrita hypocrita	AP017662	15795		22	14.5
B. hypocrita sapporensis	AP017370	15826		22	14.5
B. hypocrita sapporensis	AP018339	16133		22	14.3
B. hypocrita sapporensis	AP018481	15835		22	14.5

Continued

Taxa	GenBank accession No.	Genome size (bp)	IR (bp)	tRNA	GC content (%)
B. ignitus	DQ870926	15937		22	13.5
B. kashmirensis	MH998261	15931		22	14.9
B. lantschouensis	MW839568	16292		+Met	14.1
B. lapidarius	KT164641	16543		22	14.8
B. longipennis	MW741884	18458		22	12.8
B. opulentus	MZ352142	17262		-Val -Gln	11.2
B. pascuorum	HG995285	15740		22	12.5
B. picipes	MZ352143	18057		+Met +Tyr -Ser2	13.5
B. pratotum	OV884001	21510		+Met -Ile	11.3
B. pyrosoma	MH998260	16262		22	17.3
B. schrencki	ENT01_LC702704	15793		22	12.7
B. schrencki	ENT01_LC702705	15793		22	12.7
B. sibiricus	MH998258	16287		22	15.0
B. soroeensis	MZ352146	17107		+Met(2)	17.0
B. superbus	MZ352147	15768		22	15.9
B. terrestris lusitanicus	MK570128	15888		+Leu2 +Met	14.6
B. terrestris terrestris	NC045179			+Leu2	
B. waltoni	NC045283	16012		22	16.7
Tribe Meliponini					
Genus *Frieseomelitta*					
F. varia	CM022150	15144		22	12.2
F. varia	WNWW01002174	15144		22	12.2
Genus Tetragonisca					
T. angustula	OR030859	15414		22	12.1
Genus *Melipona*					
M. bicolor	AF466146	15001		22	13.1
M. fasciculata	OQ225244	15206		+Trp	12.9
M. scutellaris	KP202303	14862		-Gln	13.2
M. scutellaris	NC026198	14862		-Gln	13.2
Genus *Lepidotrigona*					
L. flavibasis	MN747147	15408		22	21.7
L. terminata	MN737481	15431		22	23.5
Genus *Tetragonula*					
Subgenus *Tetragonula*					
T. carbonaria	OQ918628	30665	15215	-Glu/-Gly	27.9
T. hockingsi	OQ918629	30662	15224	-Glu/-Gly	27.8

Same with Table 1.

The main assembly process was conducted using the command "get_organelle_from_reads.py", which exploits Bowtie2 (Langmead & Salzberg, 2012) to recruit mitogenome-associated reads, then performs *de novo* assembling of all recruited reads using SPAdes (Bankevich et al., 2012). FASTA assembly Graph ("FASTAG") files from the outputs of SPAdes were trimmed by removing non-mitogenomic contigs, then were used to generate the complete mitogenome sequence. For samples whose complete mitogenome sequences could not be automatically exported, their FASTAG file was visualized by Bandage (Wick et al., 2015) and we manually filtered the path to export complete mitogenome sequences. Finally, we assessed all assembled mitogenome sequences using the command "evaluate_assembly_using_mapping.py".

Mitogenome annotation was performed using the MITOS web-server (http://mitos.bioinf.uni-leipzig.de/index.py) (Bernt et al., 2013). MITOS exploits MiTFi (Jühling et al., 2012) to annotate tRNAs. The open reading frames (ORFs) were annotated with the program Geneious Prime 2021.1.1 (Biomatters, New Zealand) to help adjusting start and stop codons manually. For mitogenome sequences downloaded from NCBI, their tRNAs were re-annotated using the same method. The gene maps of annotated mitogenomes were drawn on the online server OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Greiner et al., 2019).

PCR verification of inverted repeats

The FASTAG file in Bandage (Wick et al., 2015) showed that some Meliponini species had amphimeric mitogenomes with one or two pairs of long IRs (Figure 1B). To verify the existence of IRs, we chose mitogenome sequences of six species (*Lepidotrigona flavibasis*, *Lepidotrigona terminata*, *Tetragonula collina*, *Tetragonula laeviceps*, *Tetragonula gressitti*, and *Tetragonula pagdeni*) with multiple samples as template sequences to design primers for PCR amplification and sequencing to verify the boundaries between the IR and SC regions. When these amplicon sequences were consistent with these template sequences, then the existence of the IRs were confirmed.

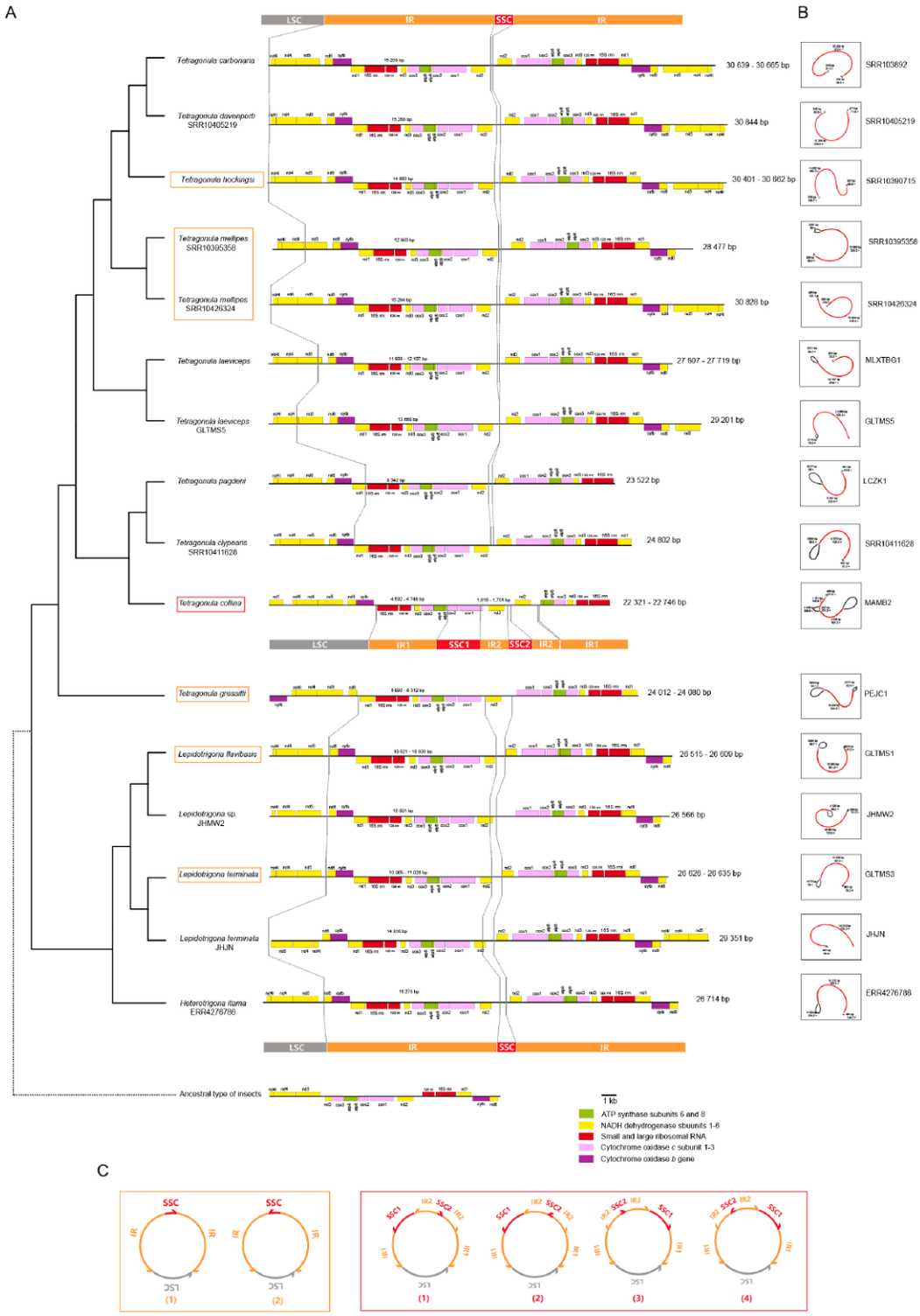


Figure 1 Mitogenome organizations for Meliponini taxa with inverted repeats

A: Linearized mitogenomic maps (in scale) of Meliponini taxa with inverted repeats and the typical type of insects. Gene directions are identified by whether they are above or below the black bar. Only rRNAs and PCGs are displayed in this map. Grey lines indicate inverted repeat boundaries in mitogenome sequences. The mitogenome lengths are given behind each map. For taxa with multiple samples, mitogenomic map of only one sample were drawn as a representative of the other samples which share the same IR boundaries, and the range of mitogenomes lengths are given. B: Target complete assembly graphs of mitogenomes of Meliponini taxa generated by GetOrganelle, with inverted repeated regions indicated by red lines and single copy regions indicated by black lines. The voucher numbers are given near the corresponding graph, and each graph corresponds to the mitogenomic map on the left in A. C: Two structural haplotypes (within the yellow box) detected in *Tetragonula hockingsi*, *Tetragonula mellipes*, *Tetragonula gressitti*, *Lepidotrigona terminata*, and *Lepidotrigona flavibasis* (framed by the yellow box in A), and four structural haplotypes (within the red box) detected in *Tetragonula collina* (framed by the red box in A). The gray regions are long single copy (LSC) regions, the yellow regions are inverted repeat (IR) regions, and the red regions are small single copy (SSC) regions.

When designing primers for the target regions, all mitogenome sequences of the same species were aligned, then we chose manually the highly conserved regions (usually within coding genes, nearly without nucleotide polymorphisms) to design primers with sizes of 20–25 bp. The GC contents (the content percent of guanine and cytosine in the nuclear sequence) of sense and anti-sense primers were designed to approximate to each other, so that they had similar annealing temperature. And adenine (A) was avoided at the 3' end of primers. The designed primers for each species are provided in Supplementary Table S2. For each species, three individuals from different sampling sites were used for PCR verifications. Genomic DNA of eighteen samples was extracted separately from a single individual by removing head and abdomen according to the protocol of the TIANamp Genomic DNA Kit. PCR amplification was performed with the designed primers using the reaction system (20 μL) as follows: ddH_2O (8 μL), 2×Taq PCR Mastermix (Tiangen, China) (10 μL), sense primers (10 P，0.5 μL), anti-sense primers (10 P，0.5 μL), and DNA templates (1 μL). After a series of PCR amplification tests, three PCR amplification programs could successfully amplify all pairs of primers with high quality amplicons. The detailed information for the three programs is as follows: (1) Program1: an initial denaturation step for 2 min at 95 °C, followed by 14 cycles of denaturation (95 °C, 30 s), annealing (39 °C, 20 s), and elongation (72 °C, 20 s), followed by another 25 cycles of denaturation (94 °C, 1 min), annealing (46 °C, 20 s), and elongation (72 °C, 20 s), and a final extension for 10 min at 72 °C. (2) Program2: an initial denaturation step for 2 min at 95 °C, followed by 14 cycles of denaturation (95 °C, 40 s), annealing (42 °C, 40 s), and elongation (72 °C, 2 min), followed by another 25 cycles of denaturation (95 °C, 40 s), annealing (54 °C, 40 s), and elongation (72 °C, 2 min), and a final extension for 10 min at 72 °C. (3) Program3: an initial denaturation step for 2 min at 95 °C, followed by 32 cycles of denaturation (95 °C, 30 s), annealing (temperature ranging from 42 to 54 °C, 30 s), and elongation (72 °C, 40 s), and a final extension for 10 min at 72 °C. All PCR products were preserved at 4 °C before further treatment. The specific PCR amplification program used for each pair of primers is listed in Supplementary Table S3.

All PCR products were evaluated using agarose gel electrophoresis to visualize the quality and length with the size marker. When the size of a high quality amplicon was consistent with the size of the temperate sequence, then the amplicon was selected to sequence. Agarose gel electrophoresis was performed on electrophoresis apparatus. The completely dissolved PCR products were mixed with the 6× Loading Buffer at a ratio of 5:1, and then were added into the gel point. The DL1000 Marker was added simultaneously. After connecting the positive and negative of the electrophoresis tank, the voltage was set at 160 V to perform electrophoresis for 40 min. After electrophoresis, the gel was removed and observed under the UV imaging system and photographed under the gel imager system. PCR products were purified using the Qingke PCR Purification Kit based on the paramagnetic particle method according to the manufacturer's protocol. DNA sequencing was carried out using the BigDye terminator kit on an ABI 3730 XL sequencer (USA).

Phylogenetic analyses

To infer phylogenetic relationships among the three tribes of social bees, 154 mitogenome sequences representing 72 species and one orchid bee *Eufriesea mexicana* were sampled, with the mitogenome of *Ceratina okinawana* (MW281319) from tribe Ceratinini as the outgroup. Thirteen PCGs were extracted from the annotated mitogenome separately and then each PCG aligned as codons using the invertebrate mitochondrial code for translation align. All PCG alignments were concatenated into a supermatrix dataset. The above operations were conducted in Geneious Prime 2021.1.1 (Biomatters, New Zealand). The supermatrix dataset was loaded into PartitionFinder2 (Lanfear et al., 2017) to select the best-fit partitioning schemes and nucleotide substitution models for downstream phylogenetic analyses. The partitioning schemes were based on 39 gene sets of sites, with one set defined separately for the 1st, 2nd and 3rd codon sites in each of 13 PCGs. The model selection and partitioning scheme comparison were performed using the Bayesian Information Criterion (BIC), and the scheme comparison were compared based on a heuristic search algorithm using the set "search=greedy" (Guindon et al., 2010; Lanfear et al., 2012).

Maximum Likelihood (ML) and Bayesian Inference (BI) phylogenies were inferred separately by RAxML-HPC2 v.8.2.12 (Stamatakis, 2014) and MrBayes v.3.2.7a (Ronquist & Huelsenbeck, 2003) with the online server CIPRES (http://www.phylo.org/). For ML analysis, the supermatrix dataset was divided into 12 partitioning schemes found by Partition Finder2. GTR+CAT was chosen as the model for the bootstrapping estimation, and the bootstraps replicates was set at 1 000 to estimate the Bootstrap support values (BS). For BI analysis, the dataset was divided into 12 subsets found by Partition Finder2, with either GTR+Γ+I or GTR+Γ as the nucleotide substitution model. Markov Chain Monte Carlo (MCMC) analyses were run for 4 000 000 generations, and the posterior probabilities (PP) were estimated. The BI analysis started with a random tree and sampled every 100 generations. The BI analysis found that the standard deviation of split frequencies was 0.0041 (<0.01 as requested), the potential scale reduction factor (PSRF) was equal to 1.00, and the estimated sample size (ESS) above 100, which indicated that the two independent runs were convergent. The first 10 000 trees were discarded as burn-in 25%, and the remaining 30 000 trees were used to calculate a majority-rule consensus tree. The phylogenetic trees were viewed and edited with the FigTree v.1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Molecular dating

We used the MCMCTree program of the PAML package (Yang, 1997, 2007) to estimate the species divergence times. We pruned the ML tree to include only one representative for each of the 72 species. We included five time calibrations in the tree based on molecular dating results in previous studies: (1) the most recent common ancestor (MRCA) of the three tribes was set at 80–90 Ma (Cardinal et al., 2018); (2) the MRCA of Bombini and Meliponini was set at 70–80 Ma (Cardinal et al., 2018); (3) the MRCA of the Meliponini clade was set at 70–80 Ma (Rasmussen & Cameron, 2010); (4) the MRCA of the Old World lineage of Meliponini was set at 50–60 Ma (Rasmussen & Cameron, 2010); (5) the MRCA of the Neotropical lineage of Meliponini was set at 30–40 Ma (Rasmussen & Cameron, 2010). The input sequence data was the alignment of 13 PCGs, which was divided into three partitions corresponding to the 1st, 2nd and 3rd codon sites

using the online tools Split Codons (https://www.bioinformatics.org/sms2/split_codons.html) within the Sequence Manipulation Suite (Stothard, 2000). We estimated the divergence times using the approximate likelihood method (Reis & Yang, 2011). We used the independent rates clock model and the GTR for the substitution model. For our analyses, MCMC ran for 400 000 +10×100 000=1 400 000 iterations, which means the program discarded the first 400 000 iterations as burn-in, and then sampled every 10 iterations until it had gathered 100 000 samples.

Reconstruction of ancestral mitogenome organization

We used the *ace* function in the R package *phytools* (Revell, 2012) to estimate the ancestral state of mitogenome organization in the three tribes. We pruned the BI tree to include only one representative for each of the 72 species. We fitted the Equal-Rates (ER) model and estimated ancestral states for three discrete characters of mitochondrial organization: (1) IR and gene rearrangement; (2) normal type; (3) gene rearrangement.

Molecular evolution estimations for the lineage with long inverted repeats

To test whether the nonsynonymous/synonymous rate ratios (d_N/d_S) of mitochondrial PCGs differ between lineages with or without IRs, we conducted codeml analysis (Yang, 1997) using EasyCodeML (Gao et al., 2019). Two phylogenetic trees were used for estimations: a big phylogeny of the three tribes of social bees and a small phylogeny of Meliponini. The BI tree topology was adopted for this analysis, but with one representative for each species. Two models were used for the likelihood analyses: (1) one-ratio model which assumes the same d_N/d_S ratio (ω) for all branches in the phylogeny; (2) two-ratio model which assumes that the branches with long IRs have a different d_N/d_S ratio (ω_1) from the background ratio (ω_0) (Yang, 1998). The lineage with IRs was specified as foreground branches in both big and small phylogenies. Each mitochondrial PCG was extracted from all taxa appearing in the tree to conduct translation alignment in Geneious Prime v.2021.1.1 (Biomatters, New Zealand). Gaps and stop codons were removed from PCG alignments manually. The gene alignment file and labelled tree file were imported into EasyCodeML software and run using the preset branch model. The program conducted the likelihood ratio test to examine whether the two-ratio model fits the data better than the one ratio model.

RESULTS

Mitogenome organization

Mitogenome organizations of Apini and Bombini taxa are the typical type of insects, with only duplications or losses of tRNAs in some taxa (Tables 1, 2). For the Meliponini, mitogenomes are amphimeric circles with a pair of long IRs in *Tetragonula*, *Lepidotrigona*, and *Heterotrigona* species, and two pairs of long IRs in *Tetragonula collina* (Figure 1A, B). Duplications of large gene blocks enlarged mitogenome sizes ranging from 22 kb to 30 kb. The structure of mitogenomes with a single pair of IRs resemble the quadripartite structure of chloroplast genomes in land plants, with a large single copy (LSC) region and a small single copy (SSC) region separated by IRs. Among samples of *T. gressitti*, *T. hockingsi*, *T. mellipes*, *L. terminata*, and *Lepidotrigona flavibasis*, we found two structural isomers, which differ by the orientation of a single SSC region. For *T. collina*, we found four structural isomers formed by the permutation and combination of two opposite directed SSC regions (Figure 1C).

For lineages with IRs, the IR boundaries are variable among species. For example, *Tetragonula* species have varied IR boundaries: IRs of *T. carbonaria*, *T. davenporti*, *T. hockingsi*, and *T. mellipes* were almost completely expanded with nearly all genes duplicated. While, in contrast, IRs were shorter and only a part of all genes were duplicated in the remaining species (Figure 1A). IR boundaries also varied among samples from the same species as shown in the following examples: the enlarged IRs in sample GLTMS5 of *Tetragonula laeviceps* and sample JN of *L. terminata* compared with the other samples of the same species, and the different IR boundaries between samples SRR10426324 and SRR10395358 of *T. mellipes* (Figure 1A). The complete assembly graphs of the mitogenomes produced by the GetOrganelle toolkit showed that read depths of IRs are generally more than twice the read depth of SCs (Figure 1B), and the gel electrophoresis image of PCR products further confirmed the existence of IRs (Figure 2). The amplified sequences were sequenced and can be mapped to the expected regions in the mitogenomes. Thus, the existence of IRs in mitogenomes of Meliponini taxa was confirmed.

Phylogenetic analyses

Topologies of ML and BI trees are almost the same, only with some incongruences in weekly supported nodes within Bombini (Supplementary Figure S1). Both phylogenetic analyses strongly support the Meliponini as sister to Bombini, then the two tribes as sister to Apini (BS=100, PP=1.00), and *Eufriesea mexicana* sister to the three tribes of social bees. The Meliponini is clearly divided into three clades corresponding to their geographical distributions, with the Indo-Malay/Australian and Afrotropical clades forming the sister group to the Neotropical clade (BS=100, PP=1.00). The only exception was *Lisotrigona carpenteri* from Yunnan, China, which falls into the Afrotropical clade (BS=84, PP=0.95) (Supplementary Figure S1). Nine *Tetragonula* species (belonging to two subgenera *Tetragonula* and *Tetragonilla*) are clustered as a monophyletic clade, however two species belonging to subgenus *Tetragonula*, *T. pagdeni* and *T. clypearis*, form a clade which is sister to *T. (Tetragonilla) collina* (BS=53, PP=0.65), instead of clustering with the other species from subgenus *Tetragonula*. Both ML and BI topologies support *T. gressitti* as the sister group to the group consisting of the other species from genus *Tetragonula* (BS=90, PP=0.68) (Supplementary Figure S1). For Bombini, except for subgenera *Mendacibombus* and *Kallobombus*, which diverged early, the other taxa can be classified into the short-faced clade and long-faced clade (Supplementary Figure S1).

Molecular evolution analyses for the lineage with long inverted repeats

The branch model test showed that for the phylogenies of both the three social bee tribes and the Meliponini, the lineage with IRs has lower estimated d_N/d_S ratios for mitochondrial PCGs than lineages without IRs. The differences between the one-ratio model and two-ratio model were significant (likelihood ratio test P-value <0.05) in both phylogenies, indicating that the d_N/d_S ratios differ between lineages with and without IRs (Table 3).

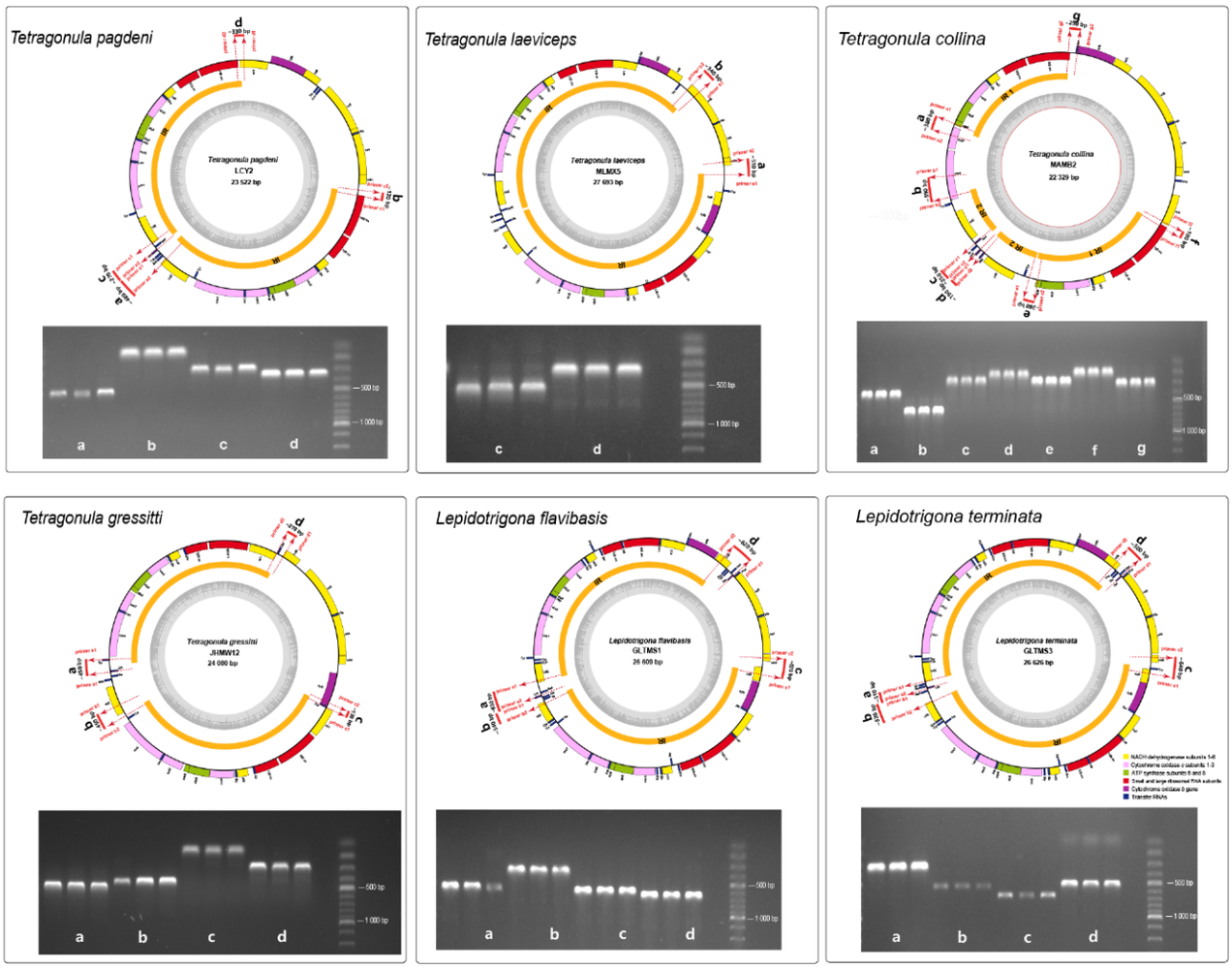


Figure 2 Gel electrophoresis image of PCR products amplified by designed primers and the distribution of primers in the mitogenomes of six Meliponini taxa with inverted repeats

Gene rearrangements in the social bees

In comparison with the typical insect mitogenome, two monophyletic lineages within Meliponini showed extensively reorganized mitogenomes. One is the lineage consisting of species with long IRs in the Indo-Malay/Australian clade. For this lineage, gene block *nd6-cytb-nd1-srRNA-lrRNA was* inverted, and swapped position with gene block *nd2-cox1-cox2-atp8-atp6-cox3-nd3.* Another is the lineage consisting of *Frieseomelitta varia*, *Tetragonisca angustula* and *Nannotrigona testaceicornis* within the Neotropical clade, for which the gene block *nd2-cox1-cox2-atp8-atp6-cox3-nd3* was inverted, and the gene block *nd6-cytb* was reversely transposed, leading to the same direction of all PCGs (Figures 3, 4).

The tRNAs show rearrangements in all three tribes and some hotspots were found. One hotspot is the cluster *trnE-trnS1*. In Apini, *trnE-trnS1* moved from cluster *trnF-trnE-trnS1-trnN* to upstream of *srRNA*. Additionally, *trnE* and *trnS1* swapped positions in *A. mellifera*, and *trnS1* was tandemly duplicated in *A. florea.* In Bombini, *trnE* and *trnS1* swapped positions in most taxa, and *trnE* also moved to downstream of *nd3* in the subgenus *Psithyrus*. In *Bombus polaris,* the *trnS1-trnE-trnN-trnR* block inversed and moved to the downstream of *nd6*. Cluster *trnE-trnS1* has not rearranged in taxa within Meliponini, except for taxa belonging to the two lineages with extensively reorganized mitogenomes (Figure 3). Another hotspot was cluster *trnP-trnT*. In the short-faced clade of Bombini, *trnP* and *trnT* swapped positions, while *trnP* was also tandemly duplicated in *Bombus vancouverensis*, and *trnT* and *trnF* inversed and swapped positions with each other in *B. polaris*. In Meliponini, *trnT* inversed in *L. carpenteri*, *Plebeina hildedrandti*, *Melipona bicolor*, *Melipona scutellaris*, *F. varia*, and *T. angustula*; *trnT* inversed and swapped positions with *trnP* in *Melipona quadrifasciata, Melipona fasciculata* and *Melipona variegatipes* (Figure 3). The last one is cluster *trnD-trnK*. Swapped positions of *trnK* and *trnD* were found across the whole tribe of Apini and Bombini, while transposition of *trnK* was also found in *L. carpenteri*, *P. hildedrandti*, and *Melipona* species in Meliponini (Figure 3). There were also some shared rearrangements among the three tribes. For example, *trnA* moved to the cluster *trnM-trnQ-trnI*, though the gene positions within this cluster varied a lot across lineages (Figure 3).

For taxa in the Indo-Malay/Australian clade of Meliponini, tRNA rearrangements occur with large blocks of genes. These rearrangements were so dramatic that it was nearly impossible to compare them with the ancestral type. One interesting finding is an ambiguous tRNA in genus *Tetragonula* species with multiple samples, including *T. laeviceps*, *T. pagdeni*, *T. gressitti*, and *T. collina*. Some samples within the species presented *trnG* (glycine) between *cox3* and *nd3* genes, while other samples presented *trnE*

Table 3 Estimated d_N/d_S ratios of branches under different models and *P*-value of likelihood ratio test (LRT) of the branch model using EasyCodeml

	Estimated d_N/d_S ratio (omega) for two ratio Model 2		Estimated d_N/d_S ratio for one ratio Model 0	LRT *P*-value
	Omega of foreground branch (ω1)	Omega of background (ω0)	Omega of all branches (ω)	(Model 0 vs. Model 2)
Phylogeny of Meliponini				
atp6	0.00614	0.05818	0.05534	0.000670464**
cox1	0.00082	0.05814	0.05534	0.000690563**
cox 2	0.00300	0.04307	0.03931	0.000000118**
cox 3	0.00498	0.07305	0.03492	0.000000000**
cytb	0.00907	0.03685	0.03492	0.000232998**
nd1	0.00163	0.03708	0.03229	0.000002739**
nd2	0.00209	0.13369	0.11868	0.00000000**0
nd3	0.01195	0.08911	0.08421	0.010207148*
nd4	0.00435	0.03319	0.03212	0.000580657**
nd4L	0.00075	0.03218	0.02988	0.026699323*
nd5	0.00368	0.02748	0.02642	0.000315853**
nd6	0.00498	0.07305	0.06688	0.000000000**
Phylogeny of three tribes				
atp6	0.03803	0.00091	0.03690	0.000004161**
cox1	0.00027	0.00963	0.00916	0.000000000**
cox 2	0.00177	0.02611	0.02503	0.000000004**
cox 3	0.00181	0.05632	0.05456	0.001325546**
cytb	0.00503	0.02007	0.00916	0.000000000**
nd1	0.00177	0.02611	0.02503	0.000000004**
nd2	0.00058	0.08153	0.07673	0.000000000**
nd3	0.00181	0.05632	0.05456	0.001325544**
nd4	0.00081	0.02502	0.02461	0.000001683**
nd4L	0.00101	0.01809	0.04750	0.000000000**
nd5	0.00067	0.02124	0.02069	0.000000405**
nd6	0.00181	0.04130	0.03936	0.001551061**

atp6: ATP synthase subunits 6; *cox1-3*: Cytochrome oxidase subunit Ⅰ, Ⅱ, Ⅲ; *cytb*: Cytochrome *b* apoenzyme; *nd1-6,4L*: NADH dehydrogenase subunits 1-6, 4L. *: $P<0.05$; **: $P<0.01$.

(glutamic acid) between *cox3* and *nd3* genes (Figure 3). The defining difference between the two tRNA is whether the second position of the anticodon is a cytosine (for *trnG*) or a thymine (for *trnE*) (Supplementary Figure S2).

Reconstruction of ancestral mitogenome organizations

We estimated the evolution of mitochondrial organizations among three tribes of social bees. The results showed that long IRs may have occurred once in the common ancestor of the Indo-Malay/Australian clade of Meliponini at 48.36 Ma, along with duplications and rearrangements of genes (Figure 4). The gene rearrangements found in the monophyletic lineage consisting of *F. varia*, *T. angustula* and *N. testaceicornis* may have occurred once in their common ancestor at 22.70 Ma (Figure 4).

DISCUSSION

Significance of mitogenome phylogeny

Meliponini and Apini are the only two highly eusocial groups within Apidae, but all studies, including this one, support a topology of (Meliponini+Bombini)+Apini (Almeida et al., 2023; Cardinal et al., 2018, 2010; Danforth et al., 2013; He et al., 2018; Kapheim et al., 2015; Kawakita et al., 2008). The phylogenetic relationships within Bombini and Apini based on mitochondrial genes reported here are largely consistent with previous studies based on nuclear and/or mitochondrial genes (Arias & Sheppard, 2005; Cameron et al., 2007; Lo et al., 2010; Ramírez et al., 2010), showing that mitochondrial genes can recover a robust phylogeny and evolutionary history for the three tribes of social bees.

Among the three tribes, Meliponini is the largest with 605 described species in 45 extant genera throughout the tropical and subtropical regions (Engel et al., 2023). Phylogenies of Meliponini have been reported in several previous studies (Christy & Roesma Dahelmi, 2019; Costa et al., 2003; Ramírez et al., 2010; Rasmussen & Camargo, 2008; Rasmussen & Cameron, 2007), but the global phylogeny is still not well resolved. The phylogenetic relationships between genera based on mitochondrial genes reported here (Supplementary Figure S1) are largely consistent with the global phylogeny based on five mitochondrial and nuclear gene sequences (Rasmussen & Cameron, 2010). An interesting finding is the phylogenetic placement of *T. gressitti*, which was not included in previous phylogenies. According to the most updated classification, *T. gressitti* belongs to the genus *Tetragonula*, which consists of two subgenera, *Tetragonula* and *Tetragonilla* (Engel et al., 2023). However, our phylogeny strongly supports that *T. gressitti* is sister to the group consisting of the other sampled genus *Tetragonula* species (Supplementary Figure S1). Moreover, workers of *T. gressitti* differ from the other subgenus *Tetragonula* species in several characters: (1) the whole body of *T. gressitti* is black

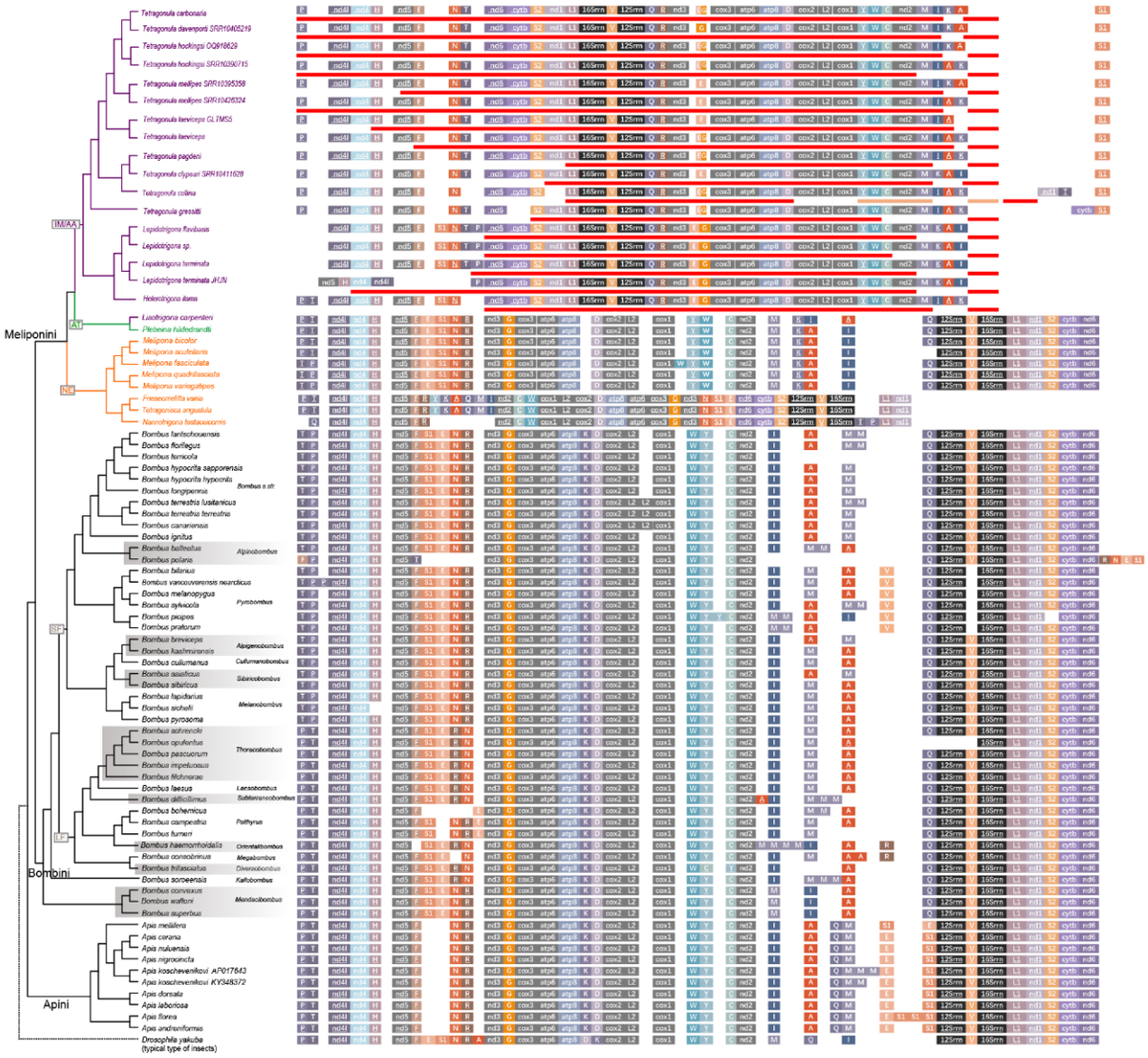


Figure 3 Phylogenetic patterns of mitochondrial gene rearrangements of the three tribes of the social bees

The topology was modified from the Bayesian tree. The colored rectangles represent mitochondrial genes. PCGs and rRNAs are abbreviated as in the text, while tRNAs are identified by the one-letter for the corresponding amino acid according to the Rules (1968). Lines under gene names indicate that the reading directions of the genes are right-to-left. The red lines under the gene block regions indicate the IR regions, with only one set of genes in IRs shown. Abbreviations of geographical distribution for lineages of Meliponini: IM/AA, Indo-Malay/Australasia clade; AT, Afrotropical clade, NE; Neotropical clade. Abbreviations of morphological lineages of Bombini: SF, short-faced clade; LF, long-faced clade.

vs. more or less pale in other species; (2) the head of *T. gressitti* is relatively round vs. more or less transverse in other species (Sakagami, 1978). Both morphology and mitogenomics therefore suggest that *T. gressitti* may represent a new subgenus within genus *Tetragonula* (Engel et al., 2023), but more detailed studies are needed to verify this.

Evolution of long inverted repeats in the mitogenome of Meliponini

In this study, long IRs were found exclusively in the Indo-Malay/Australian clade of Meliponini and are estimated to have originated once in their common ancestor at 48.36 Ma (Figure 4). Assembly graphs displayed the amphimeric mitogenomes with one or two pairs of long IRs, which have been verified by PCR amplification and sequencing. Recently, Françoso et al. (2023) also found long IRs in the mitogenomes of Australian species *T. hockingsi* and *T. carbonaria*, which have been confirmed from different samples of the two same species in this study. Although mitogenomes of *L. terminata* and *T. pagdeni* have long IRs in our assemblies, previous studies failed to identify IRs in *L. terminata* (Wang et al., 2020), *L. flavibasis* (Wang et al. 2021), and *T. pagdeni* (Wang et al., 2022). Firstly, it is only possible to detect IRs using *de novo* assembly. Assembly methods that depend on a reference mitogenome may not produce the real structure of assembled mitogenomes. Secondly, as IRs have not been reported before, mitogenome assembly programs which only take known mitogenome structures into consideration cannot produce complete mitogenome sequences with IRs directly. It is therefore necessary for mitogenome assembly programs to re-recognize the variation of mitogenome organizations, and to bring the uncommon structure reported here into their

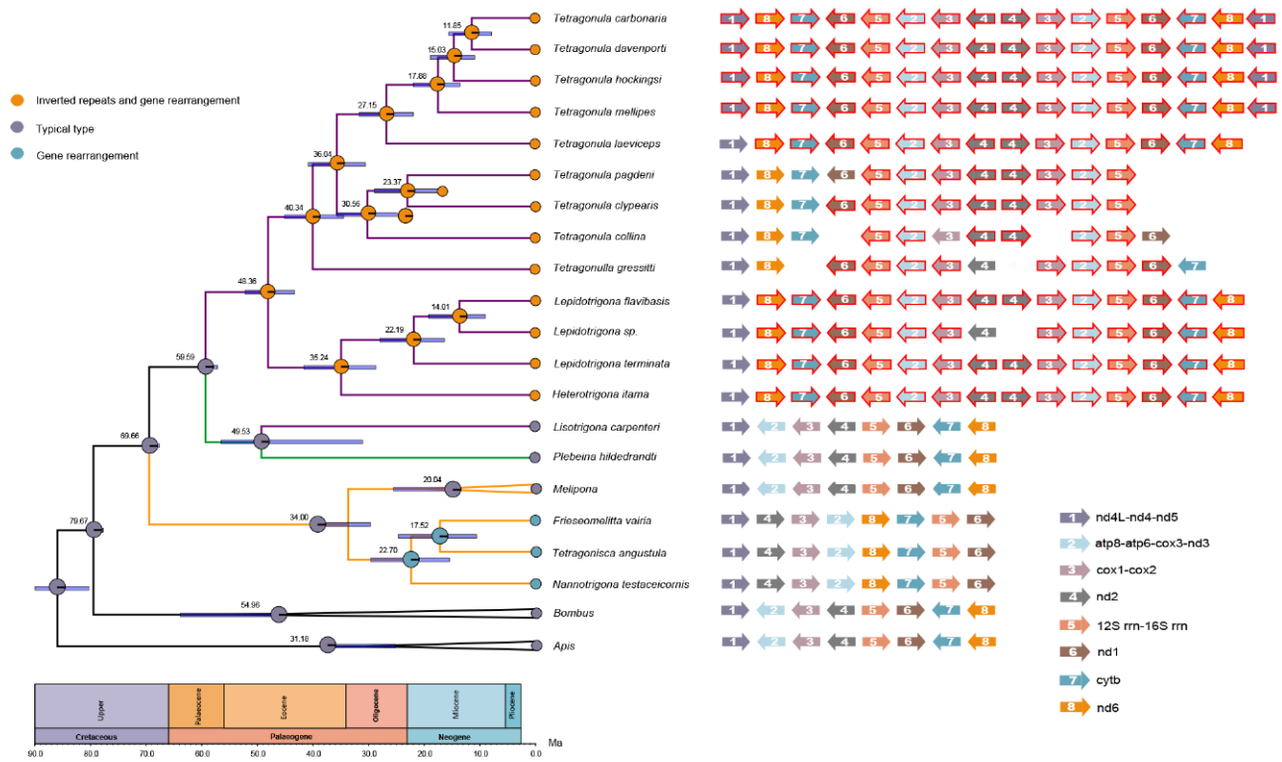


Figure 4 Ancestral reconstruction of mitogenome structures and gene rearrangement of three tribes of social bees

Gene blocks are represented by arrows with numbers, and the directions of arrows indicate the reading direction. Arrows with red boundaries represent gene blocks that are inversely repeated. PCGs and rRNAs are abbreviated as in the text. Abbreviations of geographical distribution for lineages of Meliponini: IM/AA, Indo-Malay/Australasia; AT, Afrotropical; NE, Neotropical.

reference database of animal mitogenomes.

The amphimeric mitogenome is a rare organization in insects, as well as animals (Cameron, 2014; Kolesnikov & Gerasimov, 2012; Rayko, 1997), therefore the evolutionary significance of the long IRs in mitogenomes remains unclear. Françoso et al. (2023) found long IRs in *T. hockingsi* and *T. carbonaria* and compared their mitogenomes with published mitogenomes of another two stingless bee species, *L. flavibasis* (MN747147) and *M. bicolor* (AF466146). This showed that the mitogenome of *L. flavibasis* shares the same specific rearrangements with *T. hockingsi* and *T. carbonaria*, but has no IRs in the sequence. In this study, we assembled the mitogenome of *L. flavibasis* from 12 samples having IRs, and phylogenetic analyses supported *L. flavibasis* as monophyletic including the published mitogenome (Supplementary Figure S1). Based on large-scale taxa samplings and accurate assemblies, ancestral state inference suggested the IRs of amphimeric mitogenomes are more likely to have occurred once in the common ancestor of the Indo-Malay/Australian clade, not only restricted to the Carbonaria species complex (including at least *T. carbonaria*, *T. hockingsi*, *T. davenporti*, and *T. mellipes*) as proposed by Françoso et al. (2023). The lengths of IRs show variations within the genus or species because of IR elongation or constriction, which should cause gene translocations or gene losses.

The results of the branch model test showed that for the phylogenies of both the three tribes and of the Meliponini, the lineages with IRs have lower estimated d_N/d_S ratios for mitochondrial PCGs than the lineages without IRs (Table 3), indicating the potential effect of IRs on gene evolution. There are 98 extant species from 11 extant genera in the Indomalayan, Papuasian, and Australian regions (Engel et al., 2023), while two genera, *Lisotrigona* and *Austroplebeia*, appear in the Afrotropical clade instead of the Indo-Malay/Australian clade (Rasmussen & Cameron, 2010). This study only sampled 13 species from 3 genera in this clade, so more comprehensive sampling is needed to understand the comprehensive pattern of mitogenome structures within the Meliponini and how this structural variation evolved across lineages.

Evolutionary significance of gene rearrangements

There are three types of gene rearrangements: transpositions (changes of gene positions), inversions (changes of gene directions), and inverse transpositions (changes of both directions and positions) (Cameron, 2014). Mechanisms proposed to explain these rearrangements include tandem duplication and random loss model (TDRL) (Moritz et al., 1987), tandem duplication and non-random loss model (TDNR) (Lavrov et al., 2002), and inter/intramolecular recombination (Yokobori et al., 2004). TDRL and TDNR models can explain the transpositions, while recombination can explain the inversions and inverse transpositions (Cameron, 2014).

PCGs and rRNAs are usually conserved within mitogenomes, but in the Meliponini, significant gene rearrangements of PCGs and rRNAs occur alongside IRs for taxa in the Indo-Malay/Australian clade (Figure 4). For this lineage, unknown mechanisms aside from the above-mentioned models might contribute to the formation of IRs and significant gene rearrangement. Another pattern of significant gene rearrangements of PCGs and rRNAs was found in a monophyletic lineage within the Neotropical clade (Figure 4) and a process of inter/intramolecular recombination might

contribute to this. de Paula Freitas et al. (2020) speculated that the Meliponini are a hotspot in mitogenome evolution based on the mitogenome in *F. varia*. In this study, our results strongly support that extensive gene rearrangements occurred once in the common ancestor of the monophyletic lineage within the Neotropical clade, and that the Indo-Malay/ Australian clade also had amphimeric mitogenomes and dramatic gene rearrangements.

Previous studies (Dowton & Austin, 1999; Dowton et al., 2003, 2009; Peters et al., 2017; Silvestre & Arias, 2006; Wang et al., 2020, 2021, 2022; Wei et al., 2010; Zheng et al., 2018) suggest that tRNA rearrangements are prevalent across different taxonomy levels within Hymenoptera, and can provide phylogenetic information. However, all these studies only sampled one or very few species to represent the whole genus or tribe, ignoring the influence of gene rearrangements at lower taxonomic level. Through denser taxonomic sampling across three tribes, this study shows that tRNA rearrangements are significant even at the species level (Figure 3). Some tRNA rearrangements are consistent across all samples as previous studies have reported, such as the inversion of *trnR* reported in Apoidea (Dowton et al., 2009). However, most tRNA rearrangements reported in specific lineages by previous studies are not consistent across all taxa within the lineage. For example, the inverse transposition of *trnK* reported in Meliponini (Cameron, 2014) is invalid for taxa with IRs and *F. varia*. The rearrangement of *trnI-trnQ-trnM* to *trnM-t trnqQ-trnA-trnI* reported in *Apis* is consistent in most species, but for *A. koschevenikovi*, *trnM* is also tandemly duplicated (Cameron, 2014). Additionally, we found an ambiguous tRNA in the *Tetragonula* lineage within Meliponini: either *trnG* or *trnE* presented between *cox3* and *nd3* in samples from the same species. Françoso et al. (2023) have found this in *T. carbonaria* and *T. hockingsi*, and proved the polymorphism even exists within individuals by PCR amplification and sequencing. Thus, without a relatively comprehensive sampling of investigated lineages, the result can be misleading when exploring the phylogenetic implications of tRNA rearrangements.

DATA AVAILABILITY

The newly sequenced and assembled mitogenomes of Meliponini can be downloaded from NCBI (accession numbers in Supplementary Table S4). All the newly assembled mitogenomes can be downloaded from China National Center for Bioinformation (Supplementary Table S5) (PRJCA021965) and Science Data Bank databases (DOI: 10.57760/sciencedb.j00139.00088).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

W.B.Y., R.T.C., and Y.R.L. conceived and designed the research; Z.W.W. and Y.R.L. collected and identified samples; Y.R.L. and W.B.Y. analyzed the data and interpreted results; Y.R.L. and W.B.Y. wrote draft of the manuscript. All authors revised and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We are grateful to Lei Wang, Wen-Bin Yuan and other technicians from the Molecular Center of Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy Sciences for supporting molecular experiments; to Li-Qiong Chen, Xin Yao, Xin Li for their useful suggestions and comments on data analyses; and to the physical support from the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, and the HPC Platform of Science and Technology Information Center, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences.

REFERENCES

Almeida EAB, Bossert S, Danforth BN, et al. 2023. The evolutionary history of bees in time and space. *Current Biology*, **33**(16): 3409−3422.e6.

Arias MC, Sheppard WS. 2005. Phylogenetic relationships of honey bees (Hymenoptera: Apinae: Apini) inferred from nuclear and mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, **37**(1): 25−35.

Bankevich A, Nurk S, Antipov D, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*, **19**(5): 455−477.

Beckenbach AT. 2011. Mitochondrial genome sequences of representatives of three families of scorpionflies (Order Mecoptera) and evolution in a major duplication of coding sequence. Genome, **54**(5): 368–376.

Bernt M, Donath A, Jühling F, et al. 2013. MITOS: improved *de novo* Metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, **69**(2): 313−319.

Blaimer BB, Santos BF, Cruaud A, et al. 2023. Key innovations and the diversification of Hymenoptera. *Nature Communications*, **14**(1): 1212.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Research*, **27**(8): 1767−1780.

Cameron SA, Hines HM, Williams PH. 2007. A comprehensive phylogeny of the bumble bees (*Bombus*). *Biological Journal of the Linnean Society*, **91**(1): 161−188.

Cameron SL. 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. *Annual Review of Entomology*, **59**: 95−117.

Cardinal S, Buchmann SL, Russell AL. 2018. The evolution of floral sonication, a pollen foraging behavior used by bees (Anthophila). *Evolution*, **72**(3): 590−600.

Cardinal S, Straka J, Danforth BN. 2010. Comprehensive phylogeny of apid bees reveals the evolutionary origins and antiquity of cleptoparasitism. *Proceedings of the National Academy of Sciences of the United States of America*, **107**(37): 16207−16211.

Christy BY, Roesma Dahelmi DI. 2019. Phylogenetic analysis of *Tetragonula minangkabau* and other species using cytochrome B gene. *Sciendo*, **63**(1): 117−124.

Clary DO, Wolstenholme DR. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *Journal of Molecular Evolution*, **22**(3): 252−271.

Costa MA, Del Lama MA, Melo GAR, et al. 2003. Molecular phylogeny of the stingless bees (Apidae, Apinae, Meliponini) inferred from mitochondrial 16S rDNA sequences. *Apidologie*, **34**(1): 73−84.

Danforth BN, Cardinal S, Praz C, et al. 2013. The impact of molecular data on our understanding of bee phylogeny and evolution. *Annual Review of Entomology*, **58**: 57−78.

de Paula Freitas FC, Lourenço AP, Nunes FMF, et al. 2020. The nuclear and mitochondrial genomes of *Frieseomelitta varia* – a highly eusocial stingless bee (Meliponini) with a permanently sterile worker caste. *BMC Genomics*, **21**(1): 386.

Dickey AM, Kumar V, Morgan JK, et al. 2015. A novel mitochondrial genome architecture in thrips (Insecta: Thysanoptera): extreme size asymmetry among chromosomes and possible recent control region duplication. *BMC Genomics*, **16**(1): 439.

Dowling DK, Wolff JN. 2023. Evolutionary genetics of the mitochondrial genome: insights from *Drosophila*. *Genetics*, **224**(3): iyad036.

Dowton M, Austin AD. 1999. Evolutionary dynamics of a mitochondrial rearrangement ‘ ‘hot spot’’ in the hymenoptera. *Molecular Biology and Evolution*, **16**(2): 298−309.

Dowton M, Cameron SL, Dowavic JI, et al. 2009. Characterization of 67 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. *Molecular Biology and Evolution*, **26**(7): 1607−1617.

Dowton M, Castro LR, Campbell SL, et al. 2003. Frequent mitochondrial gene rearrangements at the Hymenopteran nad3–nad5 junction. *Journal of Molecular Evolution*, **56**(5): 517−526.

Engel MS, Rasmussen C, Ayala R, et al. 2023. Stingless bee classification and biology (Hymenoptera, Apidae): a review, with an updated key to genera and subgenera. *ZooKeys*, **1172**: 239−312.

Forbes AA, Bagley RK, Beer MA, et al. 2018. Quantifying the unquantifiable: why Hymenoptera, not Coleoptera, is the most speciose animal order. *BMC Ecology*, **18**(1): 21.

Françoso E, Zuntini AR, Ricardo PC, et al. 2023. Rapid evolution, rearrangements and whole mitogenome duplication in the Australian stingless bees *Tetragonula* (Hymenoptera: Apidae): a steppingstone towards understanding mitochondrial function and evolution. *International Journal of Biological Macromolecules*, **242**: 124568.

Gao FL, Chen CJ, Arab DA, et al. 2019. EasyCodeML: a visual tool for analysis of selection using CodeML. *Ecology and Evolution*, **9**(7): 3891−3898.

Gonçalves LT, Françoso E, Deprá M. 2023. Mitochondrial phylogenomics of bumblebees, Bombus (Hymenoptera: Apidae): a tale of structural variation, shifts in selection constraints, and tree discordance. *Zoological Journal of the Linnean Society*, doi: 10.1093/zoolinnean/zlad174.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, **47**(W1): W59−W64.

Grüter C. 2020. Stingless Bees: Their Behaviour, Ecology and Evolution. Cham: Springer.

Guindon S, Dufayard JF, Lefort V, et al. 2010. New algorithms and methods to estimate maximum–likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**(3): 307−321.

He B, Su TJ, Wu YP, et al. 2018. Phylogenetic analysis of the mitochondrial genomes in bees (Hymenoptera: Apoidea: Anthophila). *PLoS One*, **13**(8): e0202187.

Hunt GJ. 1997. Insect DNA extraction protocol. *In*: Micheli MR, Bova R. Fingerprinting Methods Based on Arbitrarily Primed PCR. Berlin: Springer.

Jin JJ, Yu WB, Yang JB, et al. 2018. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv*, **4**: 256479.

Jühling F, Pütz J, Bernt M et al. 2012. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Research*, **40**(7): 2833−2845.

Kapheim KM, Pan HL, Li C, et al. 2015. Genomic signatures of evolutionary transitions from solitary to group living. *Science*, **348**(6239): 1139−1143.

Kawakita A, Ascher JS, Sota T, et al. 2008. Phylogenetic analysis of the corbiculate bee tribes based on 12 nuclear protein–coding genes (Hymenoptera: Apoidea: Apidae). *Apidologie*, **39**(1): 163−175.

Kolesnikov AA, Gerasimov ES. 2012. Diversity of mitochondrial genome organization. *Biochemistry* (*Moscow*), **77**(13): 1424–1435.

Lanfear R, Calcott B, Ho SYW, et al. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution*, **29**(6): 1695−1701.

Lanfear R, Frandsen PB, Wright AM, et al. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution*, **34**(3): 772−773.

Langmead B, Salzberg SLJ. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357−359.

Lavrov DV, Boore JL, Brown WM. 2002. Complete mtDNA sequences of two Millipedes suggest a new model for mitochondrial gene rearrangements: duplication and nonrandom loss. *Molecular Biology and Evolution*, **19**(2): 163−169.

Lhomme P, Hines HM. 2019. Ecology and evolution of cuckoo bumble bees. *Annals of the Entomological Society of America*, **112**(3): 122−140.

Li H, Liu H, Shi AM, et al. 2012. The complete mitochondrial genome and novel gene arrangement of the unique–headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). *PLoS One*, **7**(1): e29419.

Li H, Shao RF, Song F, et al. 2013. Mitochondrial genomes of two barklice, *Psococerastis albimaculata* and *Longivalvus hyalospilus* (Psocoptera: Psocomorpha): contrasting rates in mitochondrial gene rearrangement between major lineages of Psocodea. *PLoS One*, **8**(4): e61685.

Li YR, Wang ZW, Yu ZR, et al. 2021. Species diversity, morphometrics, and nesting biology of Chinese stingless bees (Hymenoptera, Apidae, Meliponini). *Apidologie*, **52**(6): 1239−1255.

Lo N, Gloag RS, Anderson DL, et al. 2010. A molecular phylogeny of the genus *Apis* suggests that the Giant Honey Bee of the Philippines, *A. breviligula* Maa, and the Plains Honey Bee of southern India, *A. indica* Fabricius, are valid species. *Systematic Entomology*, **35**(2): 226−233.

Melo GAR. 2020. Stingless bees (meliponini). *In*: Starr CK. Encyclopedia of Social Insects. Cham: Springer.

Michener CD. 2007. The Bees of the World. 2nd ed. Baltimore: John Hopkins University Press.

Moritz C, Dowling TE, Brown WM. 1987. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, **18**: 269−292.

Oliveira DCSG, Raychoudhury R, Lavrov DV, et al. 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Molecular Biology and Evolution*, **25**(10): 2167−2180.

Peters RS, Krogmann L, Mayer C, et al. 2017. Evolutionary history of the Hymenoptera. *Current Biology*, **27**(7): 1013−1018.

Quinn TW, Mindell DP. 1996. Mitochondrial gene order adjacent to the control region in crocodile, turtle, and tuatara. *Molecular Phylogenetics and Evolution*, **5**(2): 344−351.

Quinn TW, Wilson AC. 1993. Sequence evolution in and around the mitochondrial control region in birds. *Journal of Molecular Evolution*, **37**(4): 417−425.

Ramírez SR, Nieh JC, Quental TB, et al. 2010. A molecular phylogeny of the stingless bee genus *Melipona* (Hymenoptera: Apidae). *Molecular Phylogenetics and Evolution*, **56**(2): 519−525.

Rasmussen C, Camargo JMF. 2008. A molecular phylogeny and the evolution of nest architecture and behavior in *Trigona s. s.* (Hymenoptera: Apidae: Meliponini). *Apidologie*, **39**(1): 102−118.

Rasmussen C, Cameron SA. 2007. A molecular phylogeny of the Old World stingless bees (Hymenoptera: Apidae: Meliponini) and the non-monophyly of the large genus *Trigona*. *Systematic Entomology*, **32**(1): 26–39.

Rasmussen C, Cameron SA. 2010. Global stingless bee phylogeny supports ancient divergence, vicariance, and long distance dispersal. *Biological Journal of the Linnean Society*, **99**(1): 206−232.

Rayko E. 1997. Organization, generation and replication of amphimeric genomes: a review. *Gene*, **199**(1-2): 1−18.

Reis MD, Yang ZH. 2011. Approximate likelihood calculation on a phylogeny for bayesian estimation of divergence times. *Molecular Biology and Evolution*, **28**(7): 2161−2172.

Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, **3**(2):

217−223.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**(12): 1572−1574.

Roubik DW. 1989. Ecology and Natural History of Tropical Bees. Cambridge: Cambridge University Press.

Rules T. 1968. IUPAC-IUB commission on biochemical nomenclature a one-letter notation for amino acid sequences. *Journal of Biological Chemistry*, **243**(13): 3557−3559.

Saito S, Tamura K, Aotsuka T. 2005. Replication origin of mitochondrial DNA in insects. *Genetics*, **171**(4): 1695−1705.

Sakagami SF. 1978. *Tetragonula* Stingless bees of the continental Asia and Sri Lanka (Hymenoptera, Apidae) (with 124 text-figures, 1 plate and 36 tables). *Journal of the Faculty of Science Hokkaido University*, **21**(2): 165−247.

Sato M, Sato K. 2012. Maternal inheritance of mitochondrial DNA: degradation of paternal mitochondria by allogeneic organelle autophagy, allophagy. *Autophagy*, **8**(3): 424−425.

Shi Y, Chu Q, Wei DD, et al. 2016. The mitochondrial genome of booklouse, *Liposcelis sculptilis* (Psocoptera: Liposcelididae) and the evolutionary timescale of *Liposcelis*. *Scientific Reports*, **6**: 30660.

Shtolz N, Mishmar D. 2023. The metazoan landscape of mitochondrial DNA gene order and content is shaped by selection and affects mitochondrial transcription. *Communications Biology*, **6**(1): 93.

Silvestre D, Arias MC. 2006. Mitochondrial tRNA gene translocations in highly eusocial bees. *Genetics and Molecular Biology*, **29**(3): 572−575.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post–analysis of large phylogenies. *Bioinformatics*, **30**(9): 1312−1313.

Stein DB, Palmer JD, Thompson WF. 1986. Structural evolution and flip–flop recombination of chloroplast DNA in the fern genus *Osmunda*. *Current Genetics*, **10**(11): 835–841.

Stothard P. 2000. The sequence manipulation suite: JAVASCRIPT programs for analyzing and formatting protein and DNA sequences. *BioTechniques*, **28**(6): 1102−1104.

Sweet AD, Johnson KP, Cameron SL. 2022. Independent evolution of highly variable, fragmented mitogenomes of parasitic lice. *Communications Biology*, **5**(1): 677.

Sweet AD, Johnson KP, Cao YH, et al. 2021. Structure, gene order, and nucleotide composition of mitochondrial genomes in parasitic lice from Amblycera. *Gene*, **768**: 145312.

Wang CY, Yang PL, Zhao M, et al. 2022. Unusual mitochondrial tRNA rearrangements in stingless bee *Tetragonula pagdeni* and phylogenetic analysis. *Entomological Science*, **25**(4): e12526.

Wang CY, Zhao M, Wang SJ, et al. 2021. The complete mitochondrial genome of *Lepidotrigona flavibasis* (Hymenoptera: Meliponini) and high gene rearrangement in Lepidotrigona mitogenomes. *Journal of Insect Science*, **21**(3): 10.

Wang CY, Zhao M, Xu HL, et al. 2020. Complete mitochondrial genome of the stingless bee *Lepidotrigona terminata* (Hymenoptera: Meliponinae) and phylogenetic analysis. *Mitochondrial DNA Part B*, **5**(1): 752−753.

Wang WW, Lanfear R. 2019. Long–reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biology and Evolution*, **11**(12): 3372−3381.

Wei SJ, Shi M, Sharkey MJ, et al. 2010. Comparative mitogenomics of Braconidae (Insecta: Hymenoptera) and the phylogenetic utility of mitochondrial genomes with special reference to Holometabolous insects. *BMC Genomics*, **11**: 371.

Wick RR, Schultz MB, Zobel J, et al. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics*, **31**(20): 3350−3352.

Wicke S, Schneeweiss GM, DePamphilis CW, et al. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology*, **76**(3-5): 273−297.

Xiao JH, Jia JG, Murphy RW, et al. 2011. Rapid evolution of the mitochondrial genome in Chalcidoid wasps (Hymenoptera: Chalcidoidea) driven by parasitic lifestyles. *PLoS One*, **6**(11): e26645.

Yang ZH. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, **15**(5): 568−573.

Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics*, **13**(5): 555−556.

Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**(7): 1586−1591.

Yokobori SI, Fukuda N, Nakamura M, et al. 2004. Long–term conservation of six duplicated structural genes in cephalopod mitochondrial genomes. *Molecular Biology and Evolution*, **21**(11): 2034−2046.

Zhang DX, Hewitt GM. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*, **25**(2): 99−120.

Zhang DX, Szymura JM, Hewitt GM. 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. *Journal of Molecular Evolution*, **40**(4): 382−391.

Zheng BY, Cao LJ, Tang P, et al. 2018. Gene arrangement and sequence of mitochondrial genomes yield insights into the phylogeny and evolution of bees and sphecid wasps (Hymenoptera: Apoidea). *Molecular Phylogenetics and Evolution*, **124**: 1−9.

Li et al. *Zool. Res.* 2024, 45(1): 176−188
https://doi.org/10.24272/j.issn.2095-8137.2023.032

ZR | Zoological Research

Article Open Access

Endogenous biosynthesis of docosahexaenoic acid (DHA) regulates fish oocyte maturation by promoting pregnenolone production

Yi Li[1,2], Xuehui Li[1,2], Ding Ye[1,2], Ru Zhang[1,2], Chengjie Liu[1,2], Mudan He[1,2], Houpeng Wang[1,2], Wei Hu[1,2,3], Yonghua Sun[1,2,3,*]

[1] *State Key Laboratory of Freshwater Ecology and Biotechnology*, *Institute of Hydrobiology*, *Innovation Academy for Seed Design*, *Chinese Academy of Sciences*, *Wuhan*, *Hubei* 430072, *China*

[2] *College of Advanced Agricultural Sciences*, *University of Chinese Academy of Sciences*, *Beijing* 100049, *China*

[3] *Hubei Hongshan Laboratory*, *Wuhan*, *Hubei* 430070, *China*

ABSTRACT

Omega-3 polyunsaturated fatty acids (n-3 PUFAs), particularly docosahexaenoic acid (22:6n-3, DHA), play crucial roles in the reproductive health of vertebrates, including humans. Nevertheless, the underlying mechanism related to this phenomenon remains largely unknown. In this study, we employed two zebrafish genetic models, i.e., *elovl2*$^{-/-}$ mutant as an endogenous DHA-deficient model and *fat1* (omega-3 desaturase encoding gene) transgenic zebrafish as an endogenous DHA-rich model, to investigate the effects of DHA on oocyte maturation and quality. Results show that the *elovl2*$^{-/-}$ mutants had much lower fecundity and poorer oocyte quality than the wild-type controls, while the *fat1* zebrafish had higher fecundity and better oocyte quality than wild-type controls. DHA deficiency in *elovl2*$^{-/-}$ embryos led to defects in egg activation, poor microtubule stability, and reduced pregnenolone levels. Further study revealed that DHA promoted pregnenolone synthesis by enhancing transcription of *cyp11a1*, which encodes the cholesterol side-chain cleavage enzyme, thereby stabilizing microtubule assembly during oogenesis. In turn, the hypothalamic-pituitary-gonadal axis was enhanced by DHA. In conclusion, using two unique genetic models, our findings demonstrate that endogenously synthesized DHA promotes oocyte maturation and quality by promoting pregnenolone production via transcriptional regulation of *cyp11a1*.

Keywords: Docosahexaenoic acid; Oocyte maturation; Oocyte quality; Pregnenolone; Microtubule

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

INTRODUCTION

Polyunsaturated fatty acids (PUFAs) play essential roles in different biological processes, including neural development (Katakura et al., 2009; Lauritzen et al., 2016) and immune inflammation (Arita et al., 2005; Arita et al., 2007). Among the various PUFAs, omega-3 PUFAs (n-3 PUFAs), especially docosahexaenoic acid (22:6n-3, DHA), are highly beneficial for human health, including the prevention of cardiovascular diseases, fatty liver disease, obesity, and neurodegenerative disorders (Swanson et al., 2012; Zhang et al., 2019a). Therefore, n-3 PUFAs, especially DHA, are considered essential dietary nutrients (Muskiet et al., 2004).

Various studies have suggested that both dietary n-3 PUFAs and DHA exert beneficial effects on animal reproduction and fertility (Li et al., 2004; Vrablik & Watts, 2013; Wathes et al., 2007; Zhu et al., 2019). In humans, supplementation with n-3 PUFAs can delay delivery and reduce the risk of premature birth (Olsen et al., 2019). In mice, diets rich in n-3 PUFAs can prolong the female reproductive lifespan and improve oocyte quality into advanced reproductive age (Nehra et al., 2012). *In vitro* studies also suggest that DHA can promote the proliferation and steroidogenesis of bovine granulosa cells, a steroidogenic cell type found in the ovary (Maillard et al., 2018). Nevertheless, to fully reveal the role and mechanism of n-3 PUFAs and DHA in reproduction, genetic models are essential.

In recent decades, several genetic mutations involving fatty acid desaturase (*fads*) and elongase of very-long-chain fatty acids (*elovl*) have been introduced in mice and *Caenorhabditis elegans*, with subsequent research highlighting the importance of endogenously biosynthesized PUFAs in reproduction

Received: 15 March 2023; Accepted: 08 September 2023; Online: 09 September 2023

Foundation items: This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Precision Seed Design and Breeding, XDA24010108), National Natural Science Foundation of China (31972780 & 31721005), National Key R&D Program of China (2018YFA0801000), and State Key Laboratory of Freshwater Ecology and Biotechnology (2019FBZ05)

*Corresponding author, E-mail: yhsun@ihb.ac.cn

(Vrablik & Watts, 2013). For example, knockout of *elovl2* and mutation of *fads2* in mice result in male sterility (Stoffel et al., 2008; Zadravec et al., 2011), while depletion of *elovl1* and *elovl2* in *C. elegans* results in an increase in saturated fatty acids, a decrease in PUFAs, and disrupted reproduction (Kniazeva et al., 2003). Nonetheless, the precise regulatory mechanism driven by endogenously synthesized PUFAs remains largely unexplored.

Given their developmental and physiological advantages, such as short reproductive cycle, embryo transparency, and high degree of similarity in reproductive regulatory systems with humans (Hoo et al., 2016; Ye et al., 2019; Zhang et al., 2022), zebrafish have emerged as a promising model for studying reproduction and germ cell biology (MacRae & Peterson, 2015; Ye et al., 2022 Wang et al., 2023) as well as the mechanisms underlying fatty acid biosynthesis and lipid metabolism *in vivo* (Monroig et al., 2018; Xu et al., 2023; Zeituni & Farber, 2016). In our recent work, we established two genetic models based on zebrafish, i.e., *fat1* transgenic zebrafish *Tg(CMV:fat1)* as an endogenous DHA-rich model and zebrafish *elovl2*$^{-/-}$ mutants as a DHA biosynthesis-deficient model (Liu et al., 2020; Pang et al., 2014). These models provide the opportunity to investigate the role and mechanism of DHA in reproduction. Zebrafish studies have shown that maternally expressed *cyp11a1* and zygotically expressed *cyp11a2*, encoding cholesterol side-chain enzymes, are critical for the production of pregnenolone (P5), a microtubule-stabilizing reagent and precursor of all steroid hormones (Hsu et al., 2006b; Wang et al., 2022). However, the transcriptional regulation of *cyp11a1* and *cyp11a2* remains to be elucidated.

In this study, we employed the aforementioned zebrafish models to examine the role of DHA in female reproduction. Results showed that endogenous synthesis of DHA enhanced female fecundity and oocyte maturation, while DHA deficiency in *elovl2*$^{-/-}$ females resulted in egg activation defects and impaired microtubule assembly. Mechanistically, endogenously synthesized DHA promoted the transcription of *cyp11a1*, which encodes a cholesterol side-chain enzyme for P5 production, thereby stabilizing microtubule networks in oocytes and promoting egg quality. In conclusion, using the two unique genetic models, our results suggest that endogenously synthesized DHA promotes oocyte maturation and quality by promoting P5 production via transcriptional regulation of *cyp11a1*.

MATERIALS AND METHODS

Zebrafish strains

The zebrafish strains used in this study, including AB (abbreviated WT), *Tg(CMV:Cel.Fat1,CMV:EGFP)*^{ihb22Tg} (abbreviated *fat1*) (Pang et al., 2014), and *elovl2*$^{-/-ihb332}$ (abbreviated *elovl2*$^{-/-}$) (Liu et al., 2020), were maintained and raised at the China Zebrafish Resource Center of the National Aquatic Biological Resource Center (CZRC/NABRC, Wuhan, China, http://zfish.cn). To obtain *elovl2*$^{-/-}$*/Tg(CMV:fat1)* fish (abbreviated *elovl2*$^{-/-}$*/fat1*), *elovl2*$^{-/-}$ mutants were crossed with *fat1* fish to generate *elovl2*$^{+/-}$*/fat1* fish, which were, in turn, incrossed to generate *elovl2*$^{-/-}$*/fat1* fish. The raising density of zebrafish and embryo staging were determined based on previous research (Ye et al., 2022). All experiments involving zebrafish were performed according to the experimental protocols approved by the Institutional Animal Care and Use Committee of the Institute of Hydrobiology, Chinese Academy of Sciences under protocol IHB2015-006.

Experimental diets and feeding

All zebrafish strains received a daily diet of brine shrimp, known to contain only trace amounts of DHA (Pang et al., 2014). Both WT and *fat1* zebrafish were maintained in the same tank from the fertilized egg stage until they were separated at adulthood for experimental procedures. The *elovl2*$^{-/-}$ adult female zebrafish were supplemented with a DHA-rich diet (Nutriera, China) each day for two weeks before mating to ensure an adequate proportion of female offspring, given the extremely low ratio of female offspring under normal feeding conditions. The subsequent feeding conditions of the *elovl2*$^{-/-}$ embryos were the same as those of the other strains. Throughout the whole experiment, an equal number of zebrafish were kept in each tank, consistent per previous study (Pang et al., 2014).

Morphological analysis of ovaries and oocytes

After anesthetization with MS-222 (Sigma, USA), 6-month post-fertilization WT, *elovl2*$^{-/-}$, *fat1*, and *elovl2*$^{-/-}$*/fat1* adult fish were dissected the day after ovulation. The ovaries were collected and weighed to calculate the gonadosomatic index (GSI; gonad weight/body weight×100%). For embryos, oocyte diameter and chorion elevation distance were measured at 15 minutes post-fertilization (mpf) using ImageJ (v.1.53e). Oocyte diameter was the longest distance in the vertical direction of the animal-vegetal axis. Chorion elevation distance was the longest length of the chorion when fully inflated minus the diameter of the oocyte.

Analysis of follicle composition

Freshly dissected zebrafish ovaries were embedded in optimal cutting temperature (OCT) compound, snap-frozen in liquid nitrogen, cryo-sectioned to generate 10 μm thick slices at 100 μm intervals with a frozen microtome (Thermo HM525, USA). The sections were stained with hematoxylin and eosin (H&E), as described previously (Zhang et al., 2020), then photographed using an Olympus BX53 microscope (Japan). Follicles on the sections were staged according to previously reported criteria (Yang et al., 2017). The number of follicles at different stages was counted for each section, and the proportion of each follicle type was calculated. At least three ovaries were examined and analyzed per group.

Germinal vesicle breakdown (GVBD) assay

Dissected ovaries were washed three times with 1×phosphate-buffered saline (PBS) and transferred to oocyte culture medium consisting of 90% Leibovitz's L-15 medium (Gibco, USA) and 10% fetal bovine serum (BI, Israel) with 100 μg/mL penicillin-streptomycin (Gibco, USA). Follicles at different stages were gently separated with tweezers. Dissociated mature follicles were transferred to a new dish of fresh oocyte culture medium. The GVBD ratio was then assessed as described in previous study (Zhang et al., 2020), with some modifications. In brief, mature oocyte-stage follicles were co-cultured with oocyte culture medium containing 17α-20β-dihydroxy-4pregnen-3-one (DHP, Cayman, USA) at 28°C for 2 h. Follicle translucency predicted GVBD success.

Cortical granule exocytosis (CGE) staining

Eggs were squeezed and activated with water, with embryos then collected at 1 min, 5 min, and 15 min after activation, fixed in 4% paraformaldehyde overnight, and stained with

fluorescent dye-conjugated Maclura pomifera agglutinin (MPA, 50 μg/mL, FL-1341, Vector Laboratories, USA) and phalloidin (66 nmol/L) to display cortical granule exocytosis (He et al., 2022). About 20 embryos were measured per group.

Ooplasmic movement analysis

CellTracker™ CM-DiI Dye (100 pg, C7000, Invitrogen, USA) was microinjected into the yolk of newly fertilized embryos near the vegetal pole according to our recent study (He et al., 2022). To record ooplasmic movement, embryos were mounted in small confocal dishes with 1% agarose. For time-lapse analysis, images were captured every 82 s using an inverted microscope (Leica DMI6000B, Germany) with a 10× objective. At least seven embryos were measured per group.

Mature oocyte-stage follicle acquisition

Dissociated mature oocyte-stage follicles were transferred to a new dish of fresh oocyte culture medium. Hair filaments were made into lockets and fixed on the tips of tweezers. Two lockets were gently slid over the follicle surface until the follicular cell layer was peeled off. The follicles were fixed with 4% paraformaldehyde for subsequent microtubule abundance assay.

Immunofluorescence and microtubule arrangement analysis

Microtubule staining of oocytes and embryos was performed according to previously described procedures (Tran et al., 2012). The antibodies used for immunofluorescence were mouse anti-α-tubulin primary antibody (Invitrogen, 1:500, USA) and Alexa 568 mouse Fluor-conjugated secondary antibody (Invitrogen, 1:500, USA). Microtubule arrangement was assessed according to previous research (Hsu et al., 2006b). For embryos, images were acquired using a laser-scanning confocal inverted microscope (SP8, Leica, Germany). For oocytes, images were acquired using a single molecule confocal microscope (STELLARIS 8, Leica, Germany). ImageJ (v.1.53e) was used to measure α-tubulin fluorescence area per unit area, and relative α-tubulin density was calculated by comparing the experimental group area with that of the control group.

Drug treatments

For P5 (739545, Sigma, USA) treatment, dechorionated embryos were cultured in 0.3×Danieau buffer containing 2 μmol/L P5 from one-cell stage to shield stage. For nocodazole (M1404, Sigma, USA) treatment, mature follicles were cultured in oocyte culture medium containing 1 μg/mL nocodazole for 1 h.

Morpholino and mRNA injection

The morpholino (MO) targeting *cyp11a1* was sourced from Gene Tools, LLC (USA), with the sequence: 5'- CCATC ACACTCTCTCTCTCTACTT-3' (Hsu et al., 2006b). One-cell stage embryos were injected with *cyp11a1* MO at a dose of 7.5 ng per embryo, as described previously (Wei et al., 2014). Approximately 300 embryos were injected per group. The embryos were imaged and analyzed at 10 hours post-fertilization (hpf).

The coding sequence (CDS) of zebrafish *cyp11a1* (GenBank: NM_152953.2) and mouse *gpr120* (GenBank: NM_181748) were amplified by polymerase chain reaction (PCR), then subcloned into the pCS2+ vector, as described previously (He et al., 2020). The mRNA was synthesized using a SP6 *In Vitro* Transcription Kit (mMACHINE; Ambion, USA). Each mRNA was injected into one-cell stage embryos at a dose of 0.5 ng per embryo.

Luciferase assay

The genomic sequences of *cyp11a1* were obtained from a public database (http://asia.ensembl.org) (ENSDARG 00000002347). The Promoter 2.0 website was used to predict possible promoter activity regions of *cyp11a1*. Based on analytical results, we cloned the −6 787 to −5 297 nt and −2 150 to −1 nt region upstream of the "ATG" start codon to generate a *cyp11a1: luciferase* reporter system. The predicted promoter regions (−6 787 to −5 297 nt and −2 150 to −1 nt) were amplified from the zebrafish genomic DNA and sub-cloned into the luciferase reporter vector. A reverse PCR method was used to construct a mutant type of the luciferase reporter vector with appropriate primers (listed in Supplementary Table S1).

For the embryo assay, 50 ng of luciferase reporter plasmid and 5 ng of Renilla control plasmid (pRL-TK, Promega, USA) were mixed and injected into one-cell stage embryos (Lu et al., 2014). For the cell culture assay, 100 ng of luciferase reporter plasmid and 2.5 ng of Renilla control plasmid were mixed and co-transfected into 293T cells with 500 ng of pCS2+ −GPR120 plasmid or pCS2+ control plasmid. The injected embryos were collected at 6 hpf and the transfected cells were collected 24 h after transfection. Embryos and cells were fully lysed and centrifuged at 12 000 ×*g* for 15 seconds at room temperature, and luciferase activity of the supernatants was measured using a dual luciferase reporter system (E1910, Promega, USA) according to previous research (Wei et al., 2014). Each experiment was repeated three times.

Fatty acid analysis

Approximately 60 mature oocytes were used as one sample for fatty acid extraction. Fatty acid was extracted and measured by gas chromatography-mass spectrometry (GC-MS) according to previously described protocols (Liu et al., 2020; Pang et al., 2014). Four biological replicates were established for each group. Data were processed as a percentage of total fatty acid methyl esters.

P5 measurement

In total, 100 fertilized eggs at 15 mpf or 100 embryos at 6 hpf from one female were collected as one sample for P5 measurement. For adults, at least four females were measured individually. Fish were weighed after ovulation and quick-frozen in liquid nitrogen for subsequent assay. P5 was extracted and measured according to previous research (Wang et al., 2022).

Reverse-transcription quantitative PCR (RT-qPCR)

Total RNA was extracted from oocytes, embryos, and ovaries using TRIzol reagent (Invitrogen, USA). For brains, total RNA was extracted using an RNA Prep Pure Micro Kit (TIANGEN, China). RT-qPCR was performed as described previously (Wang et al., 2022). All RT-qPCR gene-specific primers are listed in Supplementary Table S2.

Statistical analysis

All results were expressed as mean±standard deviation (SD). Differences between two groups were analyzed by Student's *t*-test (one-tailed) or multiple *t*-tests-one per row using GraphPad v.8.3.0, with $P<0.05$ considered statistically significant. Statistical significance is represented by asterisks

(*: P<0.05; **: P<0.01; ***: P<0.001). Graphs were drawn using GraphPad Prism v.8.3.0.

RESULTS

Endogenously synthesized DHA positively affects female reproduction

In our previous studies, we established two genetic models, *fat1* zebrafish and *elovl2*$^{-/-}$ mutant zebrafish (Liu et al., 2020; Pang et al., 2014), and showed that DHA was significantly enriched in the *fat1* zebrafish, but extremely reduced in the *elovl2*$^{-/-}$ mutants. In the current study, during routine reproduction, the *fat1* females exhibited significantly higher fecundity than the WT controls, whereas the *elovl2*$^{-/-}$ females exhibited significantly lower fecundity than the WT controls (Figure 1A). Dissection of the ovaries showed that the *fat1* ovaries were significantly larger while the *elovl2*$^{-/-}$ ovaries were much smaller than those from WT siblings (Figure 1B). GSI analysis further confirmed the above results (Figure 1C). Therefore, we examined the fatty acid profiles of oocytes from WT, *fat1*, and *elovl2*$^{-/-}$ mutant zebrafish. Among all fatty acids, only DHA content showed a significant difference among the three groups, i.e., *elovl2*$^{-/-}$ mutant oocytes showed a 0.14-fold decrease (1.2% vs. 8.2%; P<0.001) in DHA content, whereas *fat1* oocytes showed a 1.3-fold increase (10.7% vs. 8.2%; P<0.05) compared to the WT group (Figure 1D). These results suggest that DHA content in oocytes may have a positive effect on reproduction in female zebrafish.

To further analyze the effects of DHA on ovarian follicle development, we performed histological analysis of dissected ovaries and counted the number of follicles at each stage. Results showed that follicle development in ovaries was normal in all groups, with all follicular stages evident. However, the number of vitellogenic and mature follicles was higher in *fat1* ovaries (29.0% vs. 18.8%, P<0.05) compared to WT controls, and the number of early follicles at primary and previtellogenic stages was higher in *elovl2*$^{-/-}$ mutant ovaries (81.2% vs. 71.0%, P<0.05) compared to WT controls (Figure 1E, F), suggesting that follicle differentiation and oocyte maturation were accelerated in *fat1* zebrafish but delayed in *elovl2*$^{-/-}$ mutants. The fertilization rate of *elovl2*$^{-/-}$ oocytes was also much lower than that observed in WT oocytes (Figure 1G, H). These findings suggest that endogenous deficiency of DHA reduces fertility in zebrafish females.

DHA promotes oocyte quality

Oocyte quality is the material basis for oocyte maturation, fertilization, and subsequent embryonic development (Jiao et al., 2023; Miao et al., 2020). The lower fertilization rate and delayed oocyte maturation in *elovl2*$^{-/-}$ mutants suggest that DHA may affect oocyte quality. To confirm this hypothesis, we observed the morphology of oocytes after fertilization. The *fat1* oocytes were relatively brighter, more lustrous, and larger than those of the WT controls (Figure 2A, B), whereas the *elovl2*$^{-/-}$ oocytes were slightly opaque after fertilization. At the one-cell stage, the *elovl2*$^{-/-}$ embryos showed shortened chorion elevation distance compared with WT controls (Figure 2C), indicating that the *fat1* eggs were of better quality while the *elovl2*$^{-/-}$ eggs were of poorer quality. Moreover, *in vitro* GVBD assay indicated that the GVBD proportion was lower in *elovl2*$^{-/-}$ follicles (51.5% vs. 72%, P<0.001) but higher in *fat1* follicles (83.5% vs. 72%, P<0.01) (Figure 2D, E) compared to the WT controls. These results suggest that DHA is essential for oocyte maturation and quality.

Given the reduced fertilization rate in *elovl2*$^{-/-}$ mutants, we next asked whether DHA deficiency could affect egg activation and early embryonic development. To test this hypothesis, cortical granule (CG) exocytosis was assessed in activated eggs by directly visualizing CG content during egg activation (He et al., 2022). At 1 min post-activation (mpa), no significant difference was observed in the CG exocytosis process among the three groups. At 5 mpa, CG exocytosis was almost completed in the *fat1* eggs, but persisted in the WT and *elovl2*$^{-/-}$ eggs. At 15 mpa, the CG exocytosis process was completed in WT eggs, while *elovl2*$^{-/-}$ eggs still retained some CGs (Figure 2F, G). We further tracked ooplasmic motility during early embryonic development and found that motility toward the animal pole was more vigorous in *fat1* embryos, with CellTracker reaching the blastoderm faster than in the WT controls. In contrast, *elovl2*$^{-/-}$ embryos showed sluggish ooplasmic movement, with CellTracker arriving at the blastoderm much slower than in the WT controls (Figure 2H, I; Supplementary Movies S1–S3). These results suggest that DHA promotes ooplasmic movement, egg activation, and early embryonic development.

DHA promotes oocyte quality through production of pregnenolone

As CGs are thought to move along parallel microtubule arrays during egg activation (Tran et al., 2012), and microtubules play an important role in intracellular transport (Akhmanova & Steinmetz, 2015), we speculated whether the role of DHA in oocyte quality is related to microtubule function. As such, we examined the microtubule arrangement in follicles and found that the microtubules were more densely arranged on the surface of *fat1* follicles than in WT controls. When the follicles were treated with the microtubule depolymerization agent nocodazole for 1 h, microtubule arrangement became disordered in all three groups (Figure 3A). The relative a-tubulin intensity was significantly higher in *fat1* follicles than in WT controls (Figure 3B), indicating that microtubule arrangement in *fat1* follicles was more resistant to nocodazole treatment. In contrast, the degree of disordered microtubule arrangement was more severe in *elovl2*$^{-/-}$ follicles. Thus, these observations suggest that microtubule arrangement is less stable in *elovl2*$^{-/-}$ follicles and more stable in *fat1* follicles.

The steroid hormone P5 can stabilize microtubule assembly in zebrafish embryos (Hsu et al., 2006b; Weng et al., 2013). Thus, we wondered whether poor microtubule stability in *elovl2*$^{-/-}$ mutants may be due to low P5 content. After measuring P5 content, we found that the *elovl2*$^{-/-}$ eggs had significantly lower P5 content, while the *fat1* eggs had significantly higher P5 content compared with the WT controls (Figure 3C). We then examined the microtubule structure of the embryos during development. Consistent with previous reports (Hsu et al., 2006b), a large array of microtubules oriented along the animal-vegetal axis on the surface of the embryos (Figure 3D). Compared with WT embryos, the microtubule arrangement was slightly denser in *fat1* embryos, although the difference was not significant (Figure 3E), while the length of aligned microtubules was reduced in the *elovl2*$^{-/-}$ mutants. Accordingly, P5 content was significantly higher in *fat1* embryos and lower in *elovl2*$^{-/-}$ embryos compared to the WT group (Figure 3F). Moreover, after incubation with P5, polymerized tubulin length was significantly increased in the

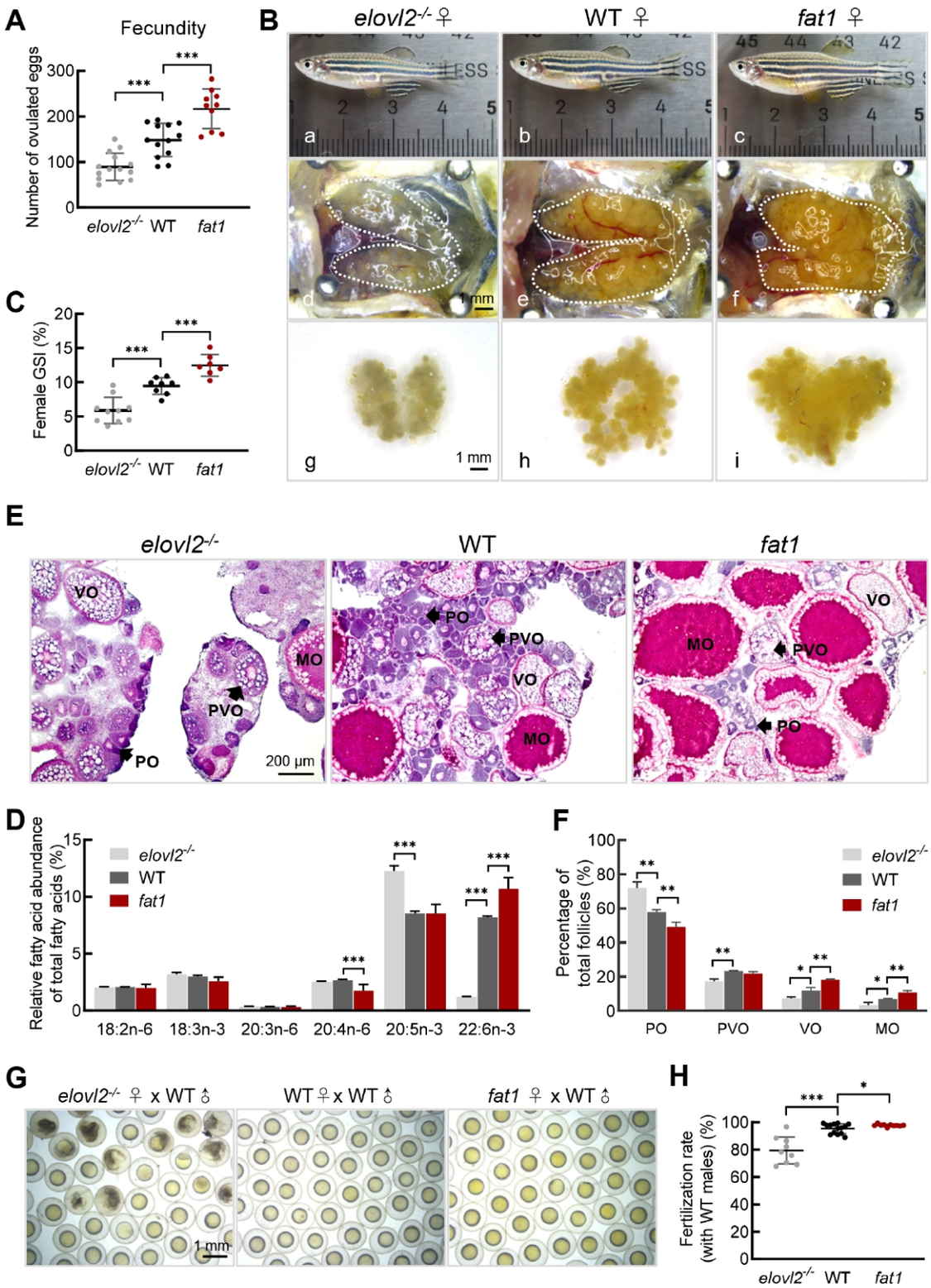


Figure 1 DHA is important for oocyte maturation and successful fertilization

A: Number of eggs laid by WT (*n*=13), *fat1* (*n*=10), and *elovl2*$^{-/-}$ (*n*=14) female zebrafish. B: Morphological and gross anatomical analysis of adult females on day after ovulation at 180 days post-fertilization (dpf). a–c: Morphology of WT, *fat1*, and *elovl2*$^{-/-}$ zebrafish. d–i: Overview of dissected ovaries from WT, *fat1*, and *elovl2*$^{-/-}$ zebrafish. Scale bar: 1 mm. C: GSI of WT (*n*=8), *fat1* (*n*=7), and *elovl2*$^{-/-}$ (*n*=10) zebrafish on day after ovulation. D: Fatty acid composition (molecular percentage) of mature oocytes from WT, *fat1*, and *elovl2*$^{-/-}$, *n*=4. E: H&E staining of WT, *fat1*, and *elovl2*$^{-/-}$ ovaries. Scale bar: 100 μm. F: Follicle quantification of WT, *fat1*, and *elovl2*$^{-/-}$ ovaries on day after ovulation, *n*=3. G: Representative images of embryos from natural mating of WT, *elovl2*$^{-/-}$, and *fat1* females with WT males. Scale bar: 1 mm. H: Comparison of fertilization ratio of groups in panel G. All values are mean±SD. Student's *t*-tests were used in panels A, C, H. Multiple *t*-tests-one per row were used in panels D, F. *: P<0.05; **: P<0.01; ***: P<0.001. Arrows point to follicle cells at different developmental stages. PO: primary oocyte; PVO, previtellogenic oocyte; VO, vitellogenic oocyte; MO, mature oocyte. GSI, gonadosomatic index. H&E, hematoxylin-eosin.

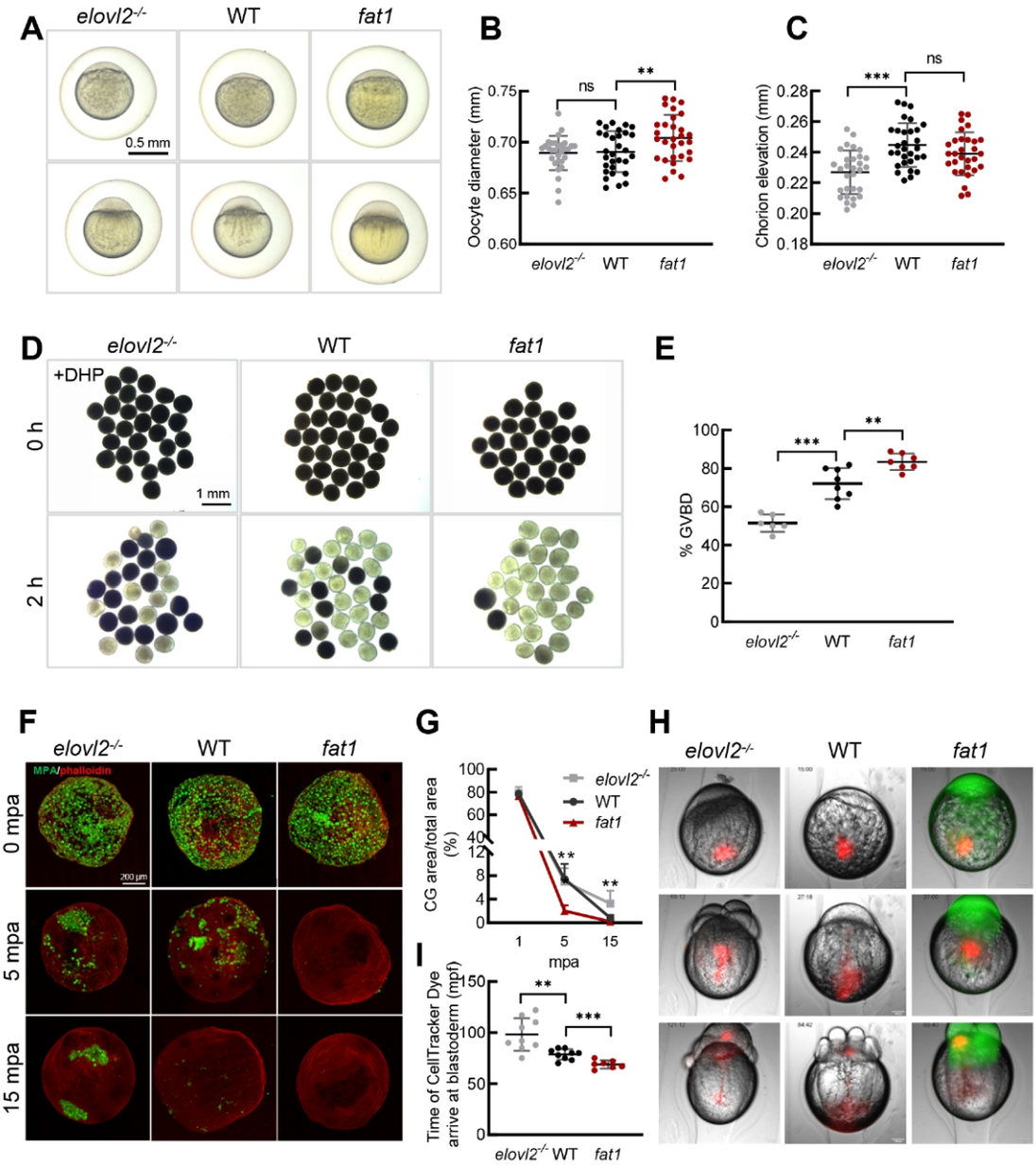


Figure 2 DHA promotes oocyte quality

A: Representative images of WT, *fat1*, and *elovl2*$^{-/-}$ embryos with chorions at 15 and 40 minutes post-fertilization (mpf). Scale bar: 250 μm. B: Measurement of oocyte diameter at 15 mpf. *n*=30. C: Measurement of chorion elevation at 15 mpf. *n*=30. D: Representative images of GVBD assay of WT, *fat1*, and *elovl2*$^{-/-}$ follicles. Scale bar: 1 mm. E: Quantitative results of GVBD assay. F: Maximum z-projection of phalloidin (red) and Maclura pomifera agglutinin (MPA) (green) staining of activated WT, *fat1*, and *elovl2*$^{-/-}$ eggs fixed at 1, 5, and 15 min post-activation (mpa). Scale bar: 50 μm. G: Statistical analysis of total CG positive areas (%) in WT, *fat1*, and *elovl2*$^{-/-}$ eggs. H: Representative images of cytoplasmic movement in WT, *fat1*, and *elovl2*$^{-/-}$ embryos. Images are from time-lapse movies of individual embryos from 15 to 120 mpf injected with CellTracker Dye immediately after fertilization. Scale bar: 100 μm. Number in upper left of image in panel I represents mpf. I: Quantification analysis of cytoplasmic movement in WT, *fat1*, and *elovl2*$^{-/-}$ embryos. Quantification is defined as time of CellTracker Dye arrival in blastoderm. At least seven embryos were measured per group. All values are mean±SD. Student's *t*-tests were used in panels B, C, E, I. Multiple *t*-tests-one per row were used in panels G. ns: No significance; *: P<0.05; **: P<0.01; ***: P<0.001.

elovl2$^{-/-}$ mutant embryos (Figure 3G). Fluorescence intensity analysis further confirmed these rescue effects (Figure 3H). Thus, these results indicate that low P5 content is responsible for the poor microtubule stability in *elovl2*$^{-/-}$ embryos, while sufficient P5 content stabilizes microtubule assembly in *fat1* embryos.

To investigate the mechanism underlying the different P5 levels in the three groups, we first examined the expression of *cyp11a1*, which encodes the cholesterol side-chain cleavage enzyme (Hsu et al., 2009; Hu et al., 2004). Results showed that expression of *cyp11a1* was significantly lower in *elovl2*$^{-/-}$ eggs and embryos, but significantly higher in *fat1* eggs and embryos compared with the WT controls (Figure 3I), suggesting that *cyp11a1* expression may affect P5 content and thus microtubule stability. To confirm this, we knocked down *cyp11a1* expression using a previously validated MO (Hsu et al., 2006b), and carefully compared the phenotypes of 10 hpf embryos injected with low-dose *cyp11a1* MO. After the

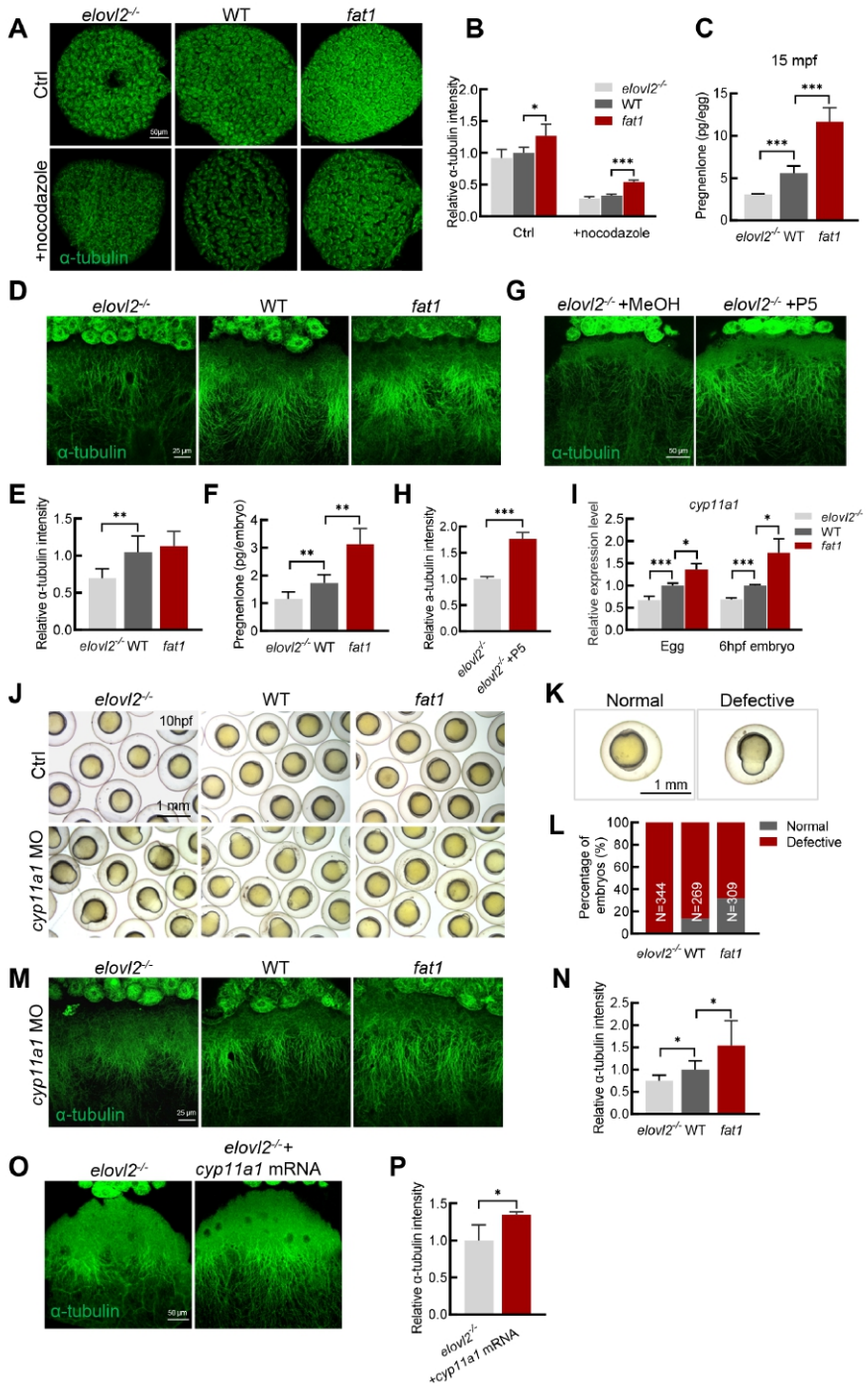


Figure 3 DHA promotes oocyte quality through production of pregnenolone

A: Representative images of anti-α-tubulin staining of mature follicles from WT, *fat1*, and *elovl2*$^{-/-}$ zebrafish treated with methanol (Ctrl) or 1 μg/mL nocodazole. Scale bar: 50 μm. B: Quantification of α-tubulin intensity of groups in panel A. Relative α-tubulin intensity was calculated by comparing fluorescence area of embryos from each experimental group with that of the WT group treated with methanol per unit area. At least three embryos were measured per group. C: Pregnenolone (P5) content in WT, *fat1*, and *elovl2*$^{-/-}$ embryos at 15 mpf. D, E: Representative images and quantitative analysis of anti-α-tubulin staining of WT, *fat1*, and *elovl2*$^{-/-}$ embryos at 6 hpf. Scale bar: 25 μm. At least four embryos were measured per group. F: P5 content in WT, *fat1*, and *elovl2*$^{-/-}$ embryos at 6 hpf. G, H: Representative images and quantitative analysis of anti-α-tubulin staining of *elovl2*$^{-/-}$ embryos incubated with methanol (MeOH, Ctrl) or P5 for 6 hpf. Scale bar: 25 μm. I: Real-time qPCR of *cyp11a1* expression in WT, *fat1*, and *elovl2*$^{-/-}$ eggs and 6 hpf embryos. J–L: Representative images and statistical analysis of WT, *fat1*, and *elovl2*$^{-/-}$ embryos with normal development or knockdown of *cyp11a1* at 10 hpf. Scale bar: 1 mm. Approximately 300 embryos were measured per group. M, N: Representative images and quantitative analysis of anti-α-tubulin staining of *cyp11a1* knockdown embryos at 6 hpf from WT, *fat1*, and *elovl2*$^{-/-}$ zebrafish. Scale bar: 25 μm. At least four embryos were measured per group. O, P: Representative images and quantitative analysis of anti-α-tubulin staining of *elovl2*$^{-/-}$ embryos with normal development or injected with *cyp11a1* mRNA. Scale bar: 25 μm. At least three embryos were measured per group. All values are mean±SD. Student's *t*-tests were used in panels C, E, F, H, L, N, P. Multiple *t*-tests-one per row were used in panels B, I. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$.

injection, 86% of the WT embryos showed epiboly defects, whereas 100% of *elovl2*$^{-/-}$ embryos and only 68% of *fat1* embryos showed defects (Figure 3J-L). At 36 hpf, although uninjected embryos from the three groups (WT, *elovl2*$^{-/-}$, *fat1*) showed a survival rate of nearly 100% (Supplementary Figure S1A, B), after low-dose *cyp11a1* MO injection, none of the *elovl2*$^{-/-}$ embryos survived, while 41.3% of WT embryos and 58.7% of *fat1* embryos survived (Supplementary Figure S1A, C). These observations suggest that DHA-deficient *elovl2*$^{-/-}$ embryos are more sensitive and DHA-rich *fat1* embryos are more resistant to *cyp11a1* knockdown.

We further examined microtubule arrangement after *cyp11a1* knockdown. Disruption of *cyp11a1* function led to a diminished microtubule arrangement in the embryos of all three groups. However, microtubule density was lower in *elovl2*$^{-/-}$ embryos and much higher in *fat1* embryos compared to the WT controls (Figure 3M, N). After injection of *cyp11a1* mRNA, the density of aligned microtubules was significantly increased in the *elovl2*$^{-/-}$ mutants (Figure 3O). These findings were confirmed by fluorescence intensity analysis (Figure 3P), indicating that microtubule assembly defects in *elovl2*$^{-/-}$ embryos are caused by a deficiency of *cyp11a1* and its catalytic product P5. Overall, our results suggest that DHA can improve oocyte quality through stabilizing microtubule arrangement.

DHA promotes P5 production by regulating *cyp11a1* promoter activity

Next, we investigated whether DHA directly regulates *cyp11a1* transcription. After identification of the active *cyp11a1* promoter region, results showed that the −6 787 to −5 297 nt region exhibited promoter activity (*cyp11a1: luc-2*) (Figure 4A, B). We then explored whether the *cyp11a1* promoter responds to DHA regulation. In HEK293T cells, treatment of 10 nmol/L DHA increased luciferase activity of *cyp11a1: luc-2* when co-transfected with a mammalian DHA receptor GPR120 expression construct (Figure 4C), suggesting that DHA may regulate *cyp11a1* transcription through the G protein-coupled signaling pathway. We further examined *cyp11a1* promoter activity in different groups of zebrafish embryos. At 6 hpf, luciferase activity of *cyp11a1: luc-2* in the *elovl2*$^{-/-}$ mutants was much lower than that observed in WT controls, while incubation with DHA significantly increased luciferase activity both in the *elovl2*$^{-/-}$ mutants and WT controls (Figure 4D), indicating that DHA also regulates *cyp11a1* promoter activity in zebrafish embryos.

DHA is reported to bind to GPR120 to stimulate both protein kinase C (PKC) and mitogen-activated protein (MAP) kinase, ultimately acting on the serum response element (SRE) on the genome to regulate gene expression (Oh et al., 2010). Here, we identified a sequence (TCATATTGGG) on the *cyp11a1* promoter very similar to the inner core 10 nucleotide region (CCATATTAGG) of the SRE. To confirm whether the SRE-like sequence is a DHA responsive domain, we deleted the 210 bp sequence covering the SRE-like region (−5 659 to −5 450 nt) on *cyp11a1: luc-2*. Results showed that luciferase activity of the mutated promoter was significantly decreased in all groups compared to the original promoter (Figure 4E), indicating that this sequence on the *cyp11a1* promoter is the region that responds to DHA regulation. In summary, our findings imply that DHA potentially modulates *cyp11a1* transcription by regulating its promoter activity.

Vertebrate reproduction is controlled by steroid hormones and the conversion of cholesterol to P5 constitutes the initial step in the steroidogenesis pathway (Hsu et al., 2006a; Wang et al., 2022). Given their differences in P5 levels, we further examined the expression of genes associated with steroidogenesis in the three groups. Genes upstream of the hypothalamic-pituitary-gonad (HPG) axis, such as *gnrh3, fshb,* and *lhb*, were up-regulated in *fat1* brains, while related genes, such as *cyp19a1a* and *fshr*, were also up-regulated in *fat1* ovaries after ovulation. In contrast, the expression levels of *lhb* and its receptor (LHR) were down-regulated in *elovl2*$^{-/-}$ brains and ovaries, respectively, compared to the WT controls (Figure 4F, G). Consistently, compared to the WT controls, there was a trend toward increased P5 content in *fat1* fish the day after ovulation (Figure 4H). Overall, these results suggest that the HPG axis is enhanced in DHA-rich *fat1* fish but reduced in DHA-deficient *elovl2* mutant, further amplifying the effects of DHA on oocyte maturation.

***fat1* transgene restores fecundity of *elovl2*$^{-/-}$ mutants**

To further confirm the important role of DHA in oocyte maturation and quality, we crossed the *fat1* transgene into *elovl2*$^{-/-}$ mutant zebrafish to explore potential rescue effects. Analysis of fatty acid composition revealed a significant increase in DHA levels in the *elovl2*$^{-/-}$*/fat1* oocytes compared to the *elovl2*$^{-/-}$ mutants (Supplementary Figure S2). Based on fecundity assessment, *elovl2*$^{-/-}$*/fat1* females exhibited significantly higher fecundity than the *elovl2*$^{-/-}$ mutants (Figure 5A). GSI analysis indicated that *elovl2*$^{-/-}$*/fat1* ovaries were significantly larger than *elovl2*$^{-/-}$ ovaries the day after ovulation (Figure 5B, C). Furthermore, histological analysis demonstrated that the *elovl2*$^{-/-}$*/fat1* ovaries contained more vitellogenic and mature-stage follicles (Figure 5D, E), with a higher proportion of GVBD in the mature follicles, compared to the *elovl2*$^{-/-}$ mutants (Figure 5F, G), suggesting that the *fat1* transgene recovers DHA content in *elovl2*$^{-/-}$ mutants and accelerates oocyte maturation. In addition, *cyp11a1* expression was higher in the *elovl2*$^{-/-}$*/fat1* oocytes than in the *elovl2*$^{-/-}$ oocytes (Figure 5H). Furthermore, after *cyp11a1* knockdown, the *elovl2*$^{-/-}$*/fat1* embryos exhibited fewer epiboly defects and longer microtubule lengths compared to the *elovl2*$^{-/-}$ embryos (Figure 5I-K). Dual-luciferase assay also showed that luciferase activity in all *cyp11a1: luc* reporter types was higher in the *elovl2*$^{-/-}$*/fat1* embryos than in the *elovl2*$^{-/-}$ embryos (Figure 5L). These results highlight the ability of the *fat1* transgene to increase DHA content in *elovl2*$^{-/-}$ embryos and thereby improve *elovl2*$^{-/-}$ oocyte quality, thus supporting the conclusion that DHA plays an important role in oocyte maturation and quality.

DISCUSSION

Previous studies have suggested that dietary supplementation of DHA is beneficial for female reproduction in various vertebrates (Cerri et al., 2009; Chiu et al., 2018; Janati Idrissi et al., 2022; Nehra et al., 2012; Wonnacott et al., 2010). However, how and why DHA promotes oocyte maturation and female reproduction remain unknown. Our study utilized two unique genetic models of zebrafish, *elovl2* mutants deficient in DHA and *fat1* transgenics rich in DHA, to demonstrate the role of endogenously synthesized DHA in proper oocyte maturation and female reproduction. Notably, our results showed that endogenously synthesized DHA promoted oocyte maturation and quality by stimulating progesterone production and activating the HPG axis (Figure 6).

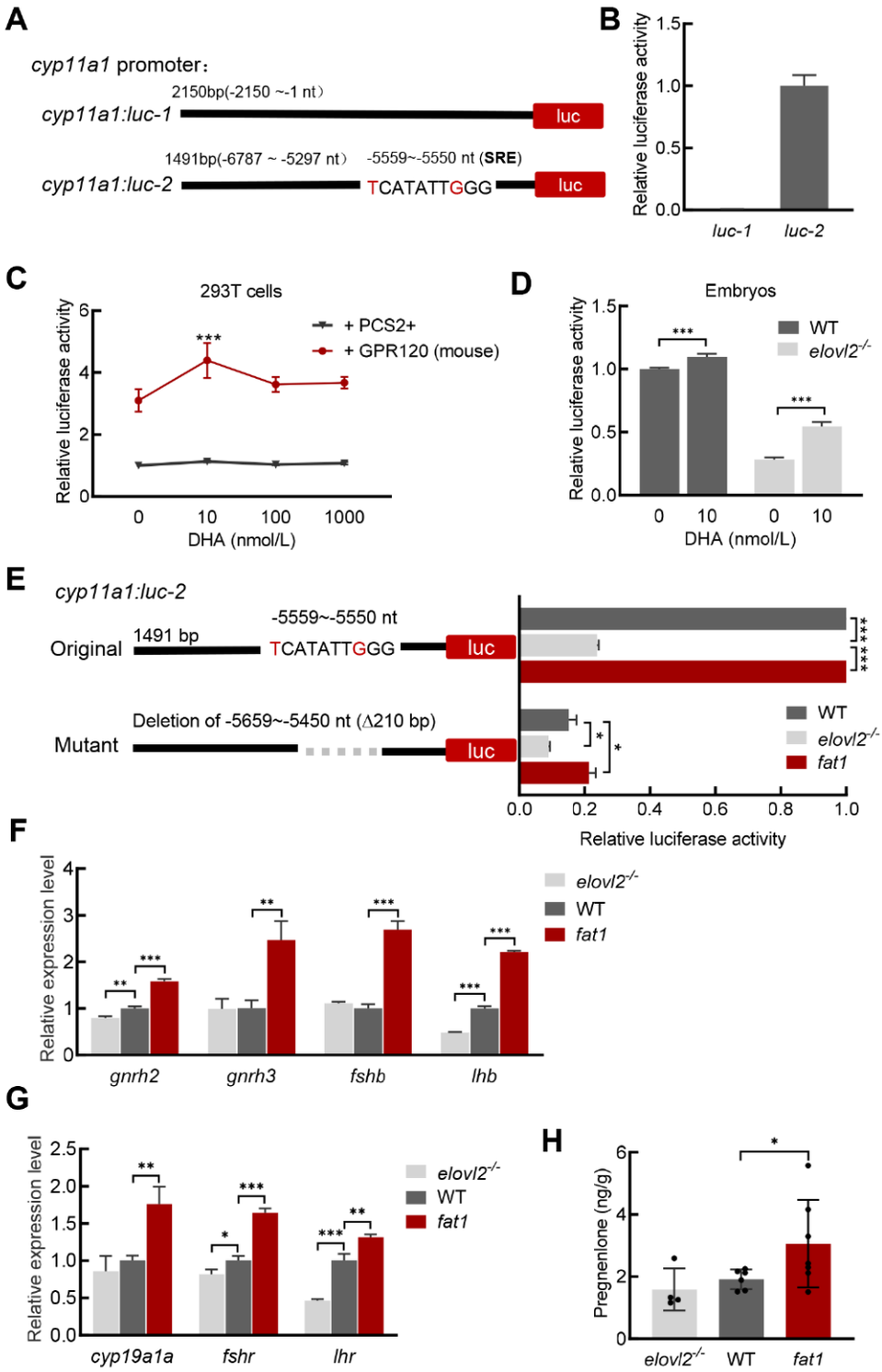


Figure 4 DHA regulates transcription of *cyp11a1* and activates HPG axis

A, B: Schematic of reporter plasmids for *cyp11a1* promoter activity analysis. Luciferase assay showed only *cyp11a1: luc-2* exhibited promoter activity in HEK293T cells. C: Luciferase assay of *cyp11a1* promoter in HEK293T cells. GPR120 increased *cyp11a1* promoter activity and DHA regulated activity of *cyp11a1: luc-2* when co-transfected with GPR120 in HEK293T cells. D: Luciferase assay of *cyp11a1* promoter in WT and *elovl2*$^{-/-}$ embryos at 6 hpf. E: Luciferase assay of different types of *cyp11a1* promoters in WT, *fat1*, and *elovl2*$^{-/-}$ embryos at 6 hpf. F, G: Real-time qPCR of gene expression in HPG axis in brain and gonads of WT, *fat1*, and *elovl2*$^{-/-}$ zebrafish. H: P5 content in WT, *fat1*, and *elovl2*$^{-/-}$ zebrafish females on day after ovulation. At least four fish were measured per group. All values are mean±SD. Student's *t*-tests were used in panels B, D, H. Multiple *t*-tests-one per row were used in panels C, E, F, G. *: $P<0.05$; **: $P<0.01$; ***: $P<0.001$.

Vertebrate reproduction is influenced by many factors, including external factors such as stress (Agarwal et al., 2012) and nutrition (Masoudi et al., 2016), as well as internal factors, such as maternal materials (Gu et al., 2015) and steroids (Krisher, 2013; Ojeda et al., 1979; Zhou, 2013). Previous research using both *in vitro* models and dietary experiments in

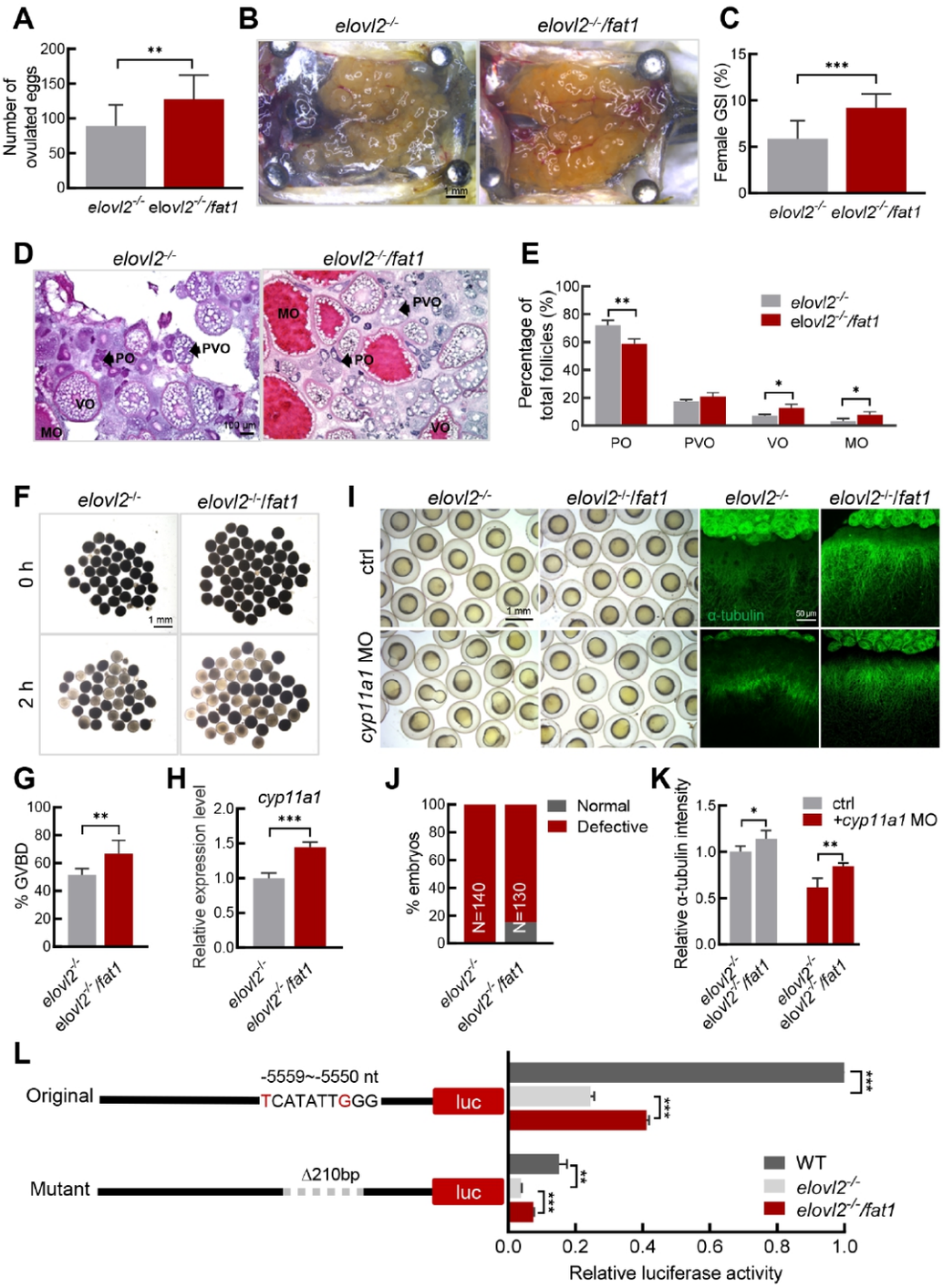


Figure 5 ***fat1*** **transgene restores fecundity of** ***elovl2***$^{-/-}$ **mutant**

A: Number of eggs laid by *elovl2*$^{-/-}$ (*n*=14) and *elovl2*$^{-/-}$*/fat1* (*n*=16) female zebrafish. B: Overview of dissected ovaries from *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* zebrafish on day after ovulation at 180 dpf. Scale bar, 1 mm. C: GSI of *elovl2*$^{-/-}$ (*n*=10) and *elovl2*$^{-/-}$*/fat1* (*n*=7) zebrafish on day after ovulation. D: H&E staining of *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* ovaries. Scale bar: 100 μm. E: Follicle quantification of *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* ovaries on day after ovulation. At least three ovaries were measured per group. F, G: Representative images and quantitative results of GVBD assay of *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* follicles. Scale bar: 1 mm. H: Real-time qPCR of *cyp11a1* expression in *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* eggs. I: Representative images of embryo morphology and anti-α-tubulin staining of *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* embryos with normal development or knockdown of *cyp11a1*. J: Statistical analysis of *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* embryo phenotypes with normal development or knockdown of *cyp11a1*. Approximately 130 embryos were measured per group at 10 hpf. K: Quantitative analysis of anti-α-tubulin staining of *elovl2*$^{-/-}$ and *elovl2*$^{-/-}$*/fat1* embryos with normal development or knockdown of *cyp11a1*. At least five embryos per group were measured at 6 hpf. L: Schematic of reporter plasmids for *cyp11a1* promoter activity analysis and luciferase assay of different types of *cyp11a1* promoters in WT, *elovl2*$^{-/-}$, and *elovl2*$^{-/-}$*/fat1* embryos at 6 hpf. All values are mean±SD. Student's *t*-tests were used in panels A, C, G, H, J. Multiple *t*-tests-one per row were used in panel E, K, L. *: P<0.05; **: P<0.01; ***: P<0.001. Arrows indicate follicle cells at different developmental stages. PO: primary oocyte; PVO, previtellogenic oocyte; VO, vitellogenic oocyte; MO, mature oocyte. GSI, gonadosomatic index; H&E, hematoxylin-eosin.

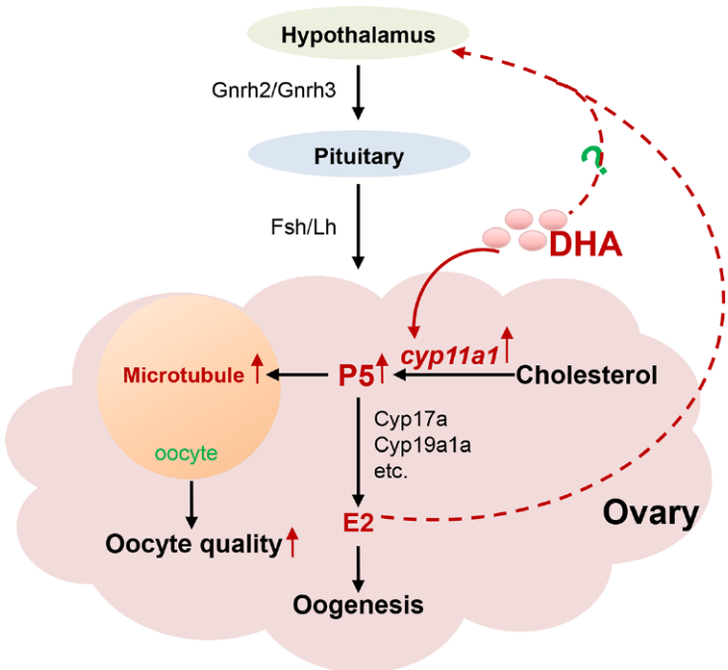


Figure 6 Graphic summary of role of DHA in oocyte maturation and oocyte quality

Endogenously synthesized DHA promotes oocyte maturation and oocyte quality by promoting progesterone production and, in turn, activating the HPG axis.

animals has demonstrated that DHA supplementation can positively regulate reproduction by enhancing HPG axis activation and steroid production (Maillard et al., 2018; Qi et al., 2019; Tocher, 2003; Tran et al., 2016; Wonnacott et al., 2010; Xu et al., 2017). Nevertheless, the underlying mechanism remains poorly elucidated. In the present study, we showed that DHA promoted the synthesis of the precursor of steroid hormones, P5, which, in turn, activated the HPG axis pathway to ensure normal oogenesis. Based on luciferase reporter assays *in vitro*, cultured cells, and *in vivo* embryos, our study further demonstrated that DHA may directly regulate the promoter activity of the *cyp11a1* gene, which encodes the cholesterol side-chain cleavage enzyme (Hsu et al., 2006b; Wang et al., 2022). Cyp11a1 converts cholesterol to P5, which then binds the microtubule plus end-tracking protein, cytoplasmic linker protein 1 (CLIP-170), stabilizing microtubules in early embryos (Hsu et al., 2006b; Weng et al., 2013). In this study, we found that oocyte-originated DHA regulated *cyp11a1* promoter activity during oocyte maturation, thereby promoting oocyte quality. Together with our recent work (Wang et al., 2022), this study reinforces the conclusion that P5 not only promotes early embryonic development but also promotes oocyte development and maturation. Given the recent generation of *fat1* transgenic common carp (*Cyprinus carpio*) (Zhang et al., 2019b), analyzing reproduction and oocyte quality in DHA-rich common carp would be of interest.

DHA has been implicated in the maintenance of spindle shape and size and microtubule morphology stability in vertebrate oocytes during meiosis (Nehra et al., 2012). In addition, DHA is reported to protect neural structural plasticity by prevention of tubulin loss (Loehfelm et al., 2020; Pusceddu et al., 2016; Serrano-García et al., 2018). These observations suggest that DHA plays an important role in stabilizing microtubule structure. The dynamic assembly of microtubules in the yolk is also critical for normal embryonic development in fish (Jesuthasan & Strähle, 1997; Sun et al., 2001; Tran et al., 2012). Our findings indicated that DHA enhances microtubule assembly stability through the up-regulation of *cyp11a1* transcription and P5 synthesis, both pivotal during oogenesis and early embryonic development.

In vitro and *in vivo* experiments have shown that DHA can mediate intracellular signaling in mammals via the G protein-coupled receptor GPR120 (Han et al., 2018; Hilgendorf et al., 2019; Hirasawa et al., 2005; Shin et al., 2019). DHA can bind GPR120 to stimulate PKC and MAP kinase, eventually acting on the genome SRE to regulate gene expression (Oh et al., 2010). Our study suggests that DHA can regulate zebrafish *cyp11a1* promoter activity through a GPR120-like pathway in 293T cells and zebrafish embryos. Future investigations should aim to identify the GPR120-like receptor potentially mediating DHA ligand-receptor signaling in fish.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y.H.S. conceived and designed the study and supervised the analyses. Y.L., X.H.L., D.Y., R.Z., and C.J.L. prepared the data. Y.L. and Y.H.S. analyzed the data. Y.L., M.D.H., and H.P.W. prepared the draft of the manuscript. W.H. provided resources. Y.L. and Y.H.S. revised and finalized the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Agarwal A, Aponte-Mellado A, Premkumar BJ, et al. 2012. The effects of oxidative stress on female reproduction: a review. *Reproductive Biology and Endocrinology*, **10**: 49.

Akhmanova A, Steinmetz MO. 2015. Control of microtubule organization and dynamics: two ends in the limelight. *Nature Reviews Molecular Cell*

Biology, **16**(12): 711−726.

Arita M, Bianchini F, Aliberti J, et al. 2005. Stereochemical assignment, antiinflammatory properties, and receptor for the omega-3 lipid mediator resolvin E1. *Journal of Experimental Medicine*, **201**(5): 713−722.

Arita M, Ohira T, Sun YP, et al. 2007. Resolvin E1 selectively interacts with leukotriene B_4 receptor BLT1 and ChemR23 to regulate inflammation. *The Journal of Immunology*, **178**(6): 3912−3917.

Cerri RLA, Juchem SO, Chebel RC, et al. 2009. Effect of fat source differing in fatty acid profile on metabolic parameters, fertilization, and embryo quality in high-producing dairy cows. *Journal of Dairy Science*, **92**(4): 1520−1531.

Chiu YH, Karmon AE, Gaskins AJ, et al. 2018. Serum omega-3 fatty acids and treatment outcomes among women undergoing assisted reproduction. *Human Reproduction*, **33**(1): 156−165.

Gu L, Liu HL, Gu X, et al. 2015. Metabolic control of oocyte development: linking maternal nutrition and reproductive outcomes. *Cellular and Molecular Life Sciences*, **72**(2): 251−271.

Han LR, Yu J, Chen YY, et al. 2018. Immunomodulatory activity of docosahexenoic acid on RAW264.7 cells activation through GPR120-mediated signaling pathway. *Journal of Agricultural and Food Chemistry*, **66**(4): 926−934.

He MD, Jiao SB, Zhang R, et al. 2022. Translational control by maternal Nanog promotes oogenesis and early embryonic development. *Development*, **149**(24): dev201213.

He MD, Zhang R, Jiao SB, et al. 2020. Nanog safeguards early embryogenesis against global activation of maternal β-catenin activity by interfering with TCF factors. *PLoS Biology*, **18**(7): e3000561.

Hilgendorf KI, Johnson CT, Mezger A, et al. 2019. Omega-3 fatty acids activate ciliary FFAR4 to control adipogenesis. *Cell*, **179**(6): 1289−1305.e21.

Hirasawa A, Tsumaya K, Awaji T, et al. 2005. Free fatty acids regulate gut incretin glucagon-like peptide-1 secretion through GPR120. *Nature Medicine*, **11**(1): 90−94.

Hoo JY, Kumari Y, Shaikh MF, et al. 2016. Zebrafish: a versatile animal model for fertility research. *BioMed Research International*, **2016**: 9732780.

Hsu HJ, Hsu NC, Hu MC, et al. 2006a. Steroidogenesis in zebrafish and mouse models. *Molecular and Cellular Endocrinology*, **248**(1-2): 160−163.

Hsu HJ, Liang MR, Chen CT, et al. 2006b. Pregnenolone stabilizes microtubules and promotes zebrafish embryonic cell movement. *Nature*, **439**(7075): 480−483.

Hsu HJ, Lin JC, Chung BC. 2009. Zebrafish *cyp11a1* and *hsd3b* genes: structure, expression and steroidogenic development during embryogenesis. *Molecular and Cellular Endocrinology*, **312**(1-2): 31−34.

Hu MC, Hsu HJ, Guo IC, et al. 2004. Function of *Cyp11a1* in animal models. *Molecular and Cellular Endocrinology*, **215**(1-2): 95−100.

Janati Idrissi S, Slezec-Frick V, Le Bourhis D, et al. 2022. Effect of DHA on the quality of *in vitro* produced bovine embryos. *Theriogenology*, **187**: 102−111.

Jesuthasan S, Strähle U. 1997. Dynamic microtubules and specification of the zebrafish embryonic axis. *Current Biology*, **7**(1): 31−42.

Jiao SB, He MD, Sun YH. 2023. Research progress and several key scientific questions in studies of fish egg quality. *Acta Hydrobiologica Sinica*, **47**(6): 1007−1024. (in Chinese)

Katakura M, Hashimoto M, Shahdat HM, et al. 2009. Docosahexaenoic acid promotes neuronal differentiation by regulating basic helix-loop-helix transcription factors and cell cycle in neural stem cells. *Neuroscience*, **160**(3): 651−660.

Kniazeva M, Sieber M, McCauley S, et al. 2003. Suppression of the ELO-2 FA elongation activity results in alterations of the fatty acid composition and multiple physiological defects, including abnormal ultradian rhythms, in *Caenorhabditis elegans*. *Genetics*, **163**(1): 159–169.

Krisher RL. 2013. *In vivo* and *in vitro* environmental effects on mammalian oocyte quality. *Annual Review of Animal Biosciences*, **1**: 393−417.

Lauritzen L, Brambilla P, Mazzocchi A, et al. 2016. DHA effects in brain development and function. *Nutrients*, **8**(1): 6.

Li YY, Chen WZ, Sun ZW, et al. 2004. Effects of *n*-3 HUFA content in broodstock diets on reproductive performance and seasonal changes of plasma sex steroids levels in *Plectorhynchus cinctus*. *Zoological Research*, **25**(3): 249–255. (in Chinese)

Liu CJ, Ye D, Wang HP, et al. 2020. Elovl2 but not Elovl5 is essential for the biosynthesis of docosahexaenoic acid (DHA) in zebrafish: Insight from a comparative gene knockout study. *Marine Biotechnology*, **22**(5): 613−619.

Loehfelm A, Elder MK, Boucsein A, et al. 2020. Docosahexaenoic acid prevents palmitate induced insulin dependent impairments of neuronal health. *The FASEB Journal*, **34**(3): 4635−4652.

Lu FI, Sun YH, Wei CY, et al. 2014. Tissue specific derepression of TCF/LEF controls the activity of the Wnt/β-catenin pathway. *Nature Communications*, **5**: 5368.

MacRae CA, Peterson RT. 2015. Zebrafish as tools for drug discovery. *Nature Reviews Drug Discovery*, **14**(10): 721−731.

Maillard V, Desmarchais A, Durcin M, et al. 2018. Docosahexaenoic acid (DHA) effects on proliferation and steroidogenesis of bovine granulosa cells. *Reproductive Biology and Endocrinology*, **16**(1): 40.

Masoudi R, Sharafi M, Zare Shahneh A, et al. 2016. Effect of dietary fish oil supplementation on ram semen freeze ability and fertility using soybean lecithin-and egg yolk-based extenders. *Theriogenology*, **86**(6): 1583−1588.

Miao YL, Cui Z, Gao Q, et al. 2020. Nicotinamide mononucleotide supplementation reverses the declining quality of maternally aged oocytes. *Cell Reports* **32**(5): 107987.

Monroig O, Tocher DR, Castro LFC. 2018. Polyunsaturated fatty acid biosynthesis and metabolism in fish. *In*: Burdge GC. Polyunsaturated Fatty Acid Metabolism. Amsterdam: Elsevier, 31–60.

Muskiet FAJ, Fokkema MR, Schaafsma A, et al. 2004. Is Docosahexaenoic Acid (DHA) Essential? Lessons from DHA status regulation, our ancient diet, epidemiology and randomized controlled trials. *The Journal of Nutrition*, **134**(1): 183−186.

Nehra D, Le HD, Fallon EM, et al. 2012. Prolonging the female reproductive lifespan and improving egg quality with dietary omega-3 fatty acids. *Aging Cell*, **11**(6): 1046−1054.

Oh DY, Talukdar S, Bae EJ, et al. 2010. GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell*, **142**(5): 687−698.

Ojeda SR, Naor Z, Negro-Vilar A. 1979. The role of prostaglandins in the control of gonadotropin and prolactin secretion. *Prostaglandins and Medicine*, **2**(4): 249−275.

Olsen SF, Halldorsson TI, Li M, et al. 2019. Examining the effect of fish oil supplementation in Chinese pregnant women on gestation duration and risk of preterm delivery. *The Journal of Nutrition*, **149**(11): 1942−1951.

Pang SC, Wang HP, Li KY, et al. 2014. Double transgenesis of humanized *fat1* and *fat2* genes promotes omega-3 polyunsaturated fatty acids synthesis in a zebrafish model. *Marine Biotechnology*, **16**(5): 580−593.

Pusceddu MM, Nolan YM, Green HF, et al. 2016. The omega-3 polyunsaturated fatty acid docosahexaenoic acid (DHA) reverses corticosterone-induced changes in cortical neurons. *International Journal of Neuropsychopharmacology*, **19**(6): pyv130.

Qi XL, Shang MY, Chen C, et al. 2019. Dietary supplementation with linseed oil improves semen quality, reproductive hormone, gene and protein expression related to testosterone synthesis in aging layer breeder roosters. *Theriogenology*, **131**: 9−15.

Serrano-García N, Fernández-Valverde F, Luis-Garcia ER, et al. 2018. Docosahexaenoic acid protection in a rotenone induced Parkinson's model: prevention of tubulin and synaptophysin loss, but no association with

mitochondrial function. *Neurochemistry International*, **121**: 26−37.
Shin JI, Jeon YJ, Lee S, et al. 2019. G-protein-coupled receptor 120 mediates DHA-induced apoptosis by regulating IP3R, ROS and, ER stress levels in cisplatin-resistant cancer cells. *Molecules and Cells*, **42**(3): 252−261.
Stoffel W, Holz B, Jenke B, et al. 2008. Δ6-desaturase (FADS2) deficiency unveils the role of ω3- and ω6-polyunsaturated fatty acids. *The EMBO Journal*, **27**(17): 2281−2292.
Sun QY, Lai LX, Park KW, et al. 2001. Dynamic events are differently mediated by microfilaments, microtubules, and mitogen-activated protein kinase during porcine oocyte maturation and fertilization in vitro. *Biology of Reproduction*, **64**(3): 879−889.
Swanson D, Block R, Mousa SA. 2012. Omega-3 fatty acids EPA and DHA: health benefits throughout life. *Advances in Nutrition*, **3**(1): 1−7.
Tocher DR. 2003. Metabolism and functions of lipids and fatty acids in teleost fish. *Reviews in Fisheries Science*, **11**(2): 107−184.
Tran DQ, Ramos EH, Belsham DD. 2016. Induction of *Gnrh* mRNA expression by the ω-3 polyunsaturated fatty acid docosahexaenoic acid and the saturated fatty acid palmitate in a GnRH-synthesizing neuronal cell model, mHypoA-GnRH/GFP. *Molecular and Cellular Endocrinology*, **426**: 125−135.
Tran LD, Hino H, Quach H, et al. 2012. Dynamic microtubules at the vegetal cortex predict the embryonic axis in zebrafish. *Development*, **139**(19): 3644−3652.
Vrablik TL, Watts JL. 2013. Polyunsaturated fatty acid derived signaling in reproduction and development: insights from *Caenorhabditis elegans* and *Drosophila melanogaster*. *Molecular Reproduction & Development*, **80**(4): 244–259.
Wang YQ, Ye D, Zhang FH, et al. 2022. *Cyp11a2* Is essential for oocyte development and spermatogonial stem cell differentiation in zebrafish. *Endocrinology*, **163**(2): bqab258.
Wang X, Zhu J, Wang H, et al. 2023. Induced formation of primordial germ cells from zebrafish blastomeres by germplasm factors. *Nature Communications*, **14**(1): 7918.
Wathes DC, Abayasekara DRE, Aitken RJ. 2007. Polyunsaturated fatty acids in male and female reproduction. *Biology of Reproduction*, **77**(2): 190−201.
Wei CY, Wang HP, Zhu ZY, et al. 2014. Transcriptional factors Smad1 and Smad9 act redundantly to mediate zebrafish ventral specification downstream of Smad5. *Journal of Biological Chemistry*, **289**(10): 6604−6618.
Weng JH, Liang MR, Chen CH, et al. 2013. Pregnenolone activates CLIP-170 to promote microtubule growth and cell migration. *Nature Chemical Biology*, **9**(10): 636−642.
Wonnacott KE, Kwong WY, Hughes J, et al. 2010. Dietary omega-3 and -6 polyunsaturated fatty acids affect the composition and development of sheep granulosa cells, oocytes and embryos. *Reproduction*, **139**(1): 57−69.
Xu HG, Cao L, Wei YL, et al. 2017. Effects of different dietary DHA: EPA ratios on gonadal steroidogenesis in the marine teleost, tongue sole (*Cynoglossus semilaevis*). *British Journal of Nutrition*, **118**(3): 179−188.
Xu SS, Li Y, Wang HP, et al. 2023. Depletion of *stearoyl-CoA desaturase* (*scd*) leads to fatty liver disease and defective mating behavior in zebrafish. *Zoological Research*, **44**(1): 63−77.
Yang YJ, Wang Y, Li Z, et al. 2017. Sequential, divergent, and cooperative requirements of *Foxl2a* and *Foxl2b* in ovary development and maintenance of zebrafish. *Genetics*, **205**(4): 1551−1572.
Ye D, Tu YX, Wang HP, et al. 2022. A landscape of differentiated biological processes involved in the initiation of sex differentiation in zebrafish. *Water Biology and Security*, **1**(3): 100059.
Ye D, Zhu L, Zhang QF, et al. 2019. Abundance of early embryonic primordial germ cells promotes zebrafish female differentiation as revealed by lifetime labeling of germline. *Marine Biotechnology*, **21**(2): 217−228.
Zadravec D, Tvrdik P, Guillou H, et al. 2011. ELOVL2 controls the level of n-6 28: 5 and 30: 5 fatty acids in testis, a prerequisite for male fertility and sperm maturation in mice. *Journal of Lipid Research*, **52**(2): 245−255.
Zeituni EM, Farber SA. 2016. Studying lipid metabolism and transport during zebrafish development. *In*: Kawakami K, Patton E, Orger M. Zebrafish: Methods and Protocols. New York: Humana Press, 237–255.
Zhang QF, Ye D, Wang HP, et al. 2020. Zebrafish *cyp11c1* knockout reveals the roles of 11-ketotestosterone and Cortisol in sexual development and reproduction. *Endocrinology*, **161**(6): bqaa048.
Zhang R, Tu YX, Ye D, et al. 2022. A germline-specific regulator of mitochondrial fusion is required for maintenance and differentiation of germline stem and progenitor cells. *Advanced Science*, **9**(36): 2203631.
Zhang TT, Xu J, Wang YM, et al. 2019a. Health benefits of dietary marine DHA/EPA-enriched glycerophospholipids. *Progress in Lipid Research*, **75**: 100997.
Zhang XF, Pang SC, Liu CJ, et al. 2019b. A novel dietary source of EPA and DHA: metabolic engineering of an important freshwater species-common carp by *fat1*-transgenesis. *Marine Biotechnology*, **21**(2): 171−185.
Zhou Q. 2013. Advances in the study of neuroendocrinological regulation of kisspeptin in fish reproduction. *Zoological Research*, **34**(5): 519−530.
Zhu YH, Tan QS, Zhang LS, et al. 2019. The migration of docosahexenoic acid (DHA) to the developing ovary of female zebrafish (*Danio rerio*). *Comparative Biochemistry and Physiology Part A:Molecular & Integrative Physiology*, **233**: 97−105.

Tong et al. *Zool. Res.* 2024, 45(1): 189−200
https://doi.org/10.24272/j.issn.2095-8137.2023.012

ZR | Zoological Research

Article Open Access

Core and variable antimicrobial resistance genes in the gut microbiomes of Chinese and European pigs

Cui-Hong Tong[1,2,3], Zhi-Peng Huo[1,2,3], Lu Diao[1,2,3], Dan-Yu Xiao[1,2,3], Ruo-Nan Zhao[1,2,3], Zhen-Ling Zeng[1,2,3,4,*], Wen-Guang Xiong[1,2,3,4,*]

[1] *Guangdong Provincial Key Laboratory of Veterinary Pharmaceutics Development and Safety Evaluation*, *South China Agricultural University*, *Guangzhou*, *Guangdong* 510642, *China*

[2] *National Risk Assessment Laboratory for Antimicrobial Resistance of Animal Original Bacteria*, *College of Veterinary Medicine*, *South China Agricultural University*, *Guangzhou*, *Guangdong* 510642, *China*

[3] *National Laboratory of Safety Evaluation (Environmental Assessment) of Veterinary Drugs*, *South China Agricultural University*, *Guangzhou*, *Guangdong* 510642, *China*

[4] *National Reference Laboratory of Veterinary Drug Residues*, *South China Agricultural University*, *Guangzhou*, *Guangdong* 510642, *China*

ABSTRACT

Monitoring the prevalence of antimicrobial resistance genes (ARGs) is vital for addressing the global crisis of antibiotic-resistant bacterial infections. Despite its importance, the characterization of ARGs and microbiome structures, as well as the identification of indicators for routine ARG monitoring in pig farms, are still lacking, particularly concerning variations in antimicrobial exposure in different countries or regions. Here, metagenomics and random forest machine learning were used to elucidate the ARG profiles, microbiome structures, and ARG contamination indicators in pig manure under different antimicrobial pressures between China and Europe. Results showed that Chinese pigs exposed to high-level antimicrobials exhibited higher total and plasmid-mediated ARG abundances compared to those in European pigs (P<0.05). *ANT(6)-Ib*, *APH(3')-IIIa*, and *tet*(40) were identified as shared core ARGs between the two pig populations. Furthermore, the core ARGs identified in pig populations were correlated with those found in human populations within the same geographical regions. *Lactobacillus* and *Prevotella* were identified as the dominant genera in the core microbiomes of Chinese and European pigs, respectively. Forty ARG markers and 43 biomarkers were able to differentiate between the Chinese and European pig manure samples with accuracies of 100% and 98.7%, respectively. Indicators for assessing ARG contamination in Chinese and European pigs also achieved high accuracy (r=0.72–0.88). *Escherichia flexneri* in both Chinese and European pig populations carried between 21 and 37 ARGs. The results of this study emphasize the importance of global collaboration in reducing antimicrobial resistance risk and provide validated indicators for evaluating the risk of ARG contamination in pig farms.

Keywords: Metagenomic; Pig manure; Antimicrobial pressure; Antimicrobial resistance genes; Microbiome

INTRODUCTION

The World Health Organization (WHO) identifies antimicrobial resistance (AMR) as a significant global threat to human and animal health (He et al., 2016). Projections for 2030 estimate a 200% increase in global antimicrobial consumption compared to the period from 2000 to 2015, with approximately 70% of this consumption anticipated to be for growth promotion and disease prevention in animal agriculture (Klein et al., 2018). Under antimicrobial pressure, antimicrobial resistance genes (ARGs) can disseminate across bacterial communities in different habitats via horizontal gene transfer (HGT), facilitated by mobile genetic elements (MGEs) such as plasmids, integrons, and insertion sequences (ISs) (Zhao et al., 2020). Bacteria, especially pathogenic bacteria, are the main factors involved in ARG transmission (Han et al., 2018). Under selective antimicrobial pressure, non-drug-resistant pathogens exhibit a propensity to acquire ARGs, while drug-resistant pathogens can persist and proliferate across various

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 19 April 2023; Accepted: 08 September 2023; Online: 09 September 2023

Foundation items: This work was supported by the Foundation for the National Key R&D Program (2022YFD1800400), Innovative Research Groups of the National Natural Science Foundation of China (32121004), and Natural Science Foundation of Guangdong Province of China (2021A1515011159)

*Corresponding authors, E-mail: zlzeng@scau.edu.cn; xiongwg@scau.edu.cn

habitats (Liang et al., 2020). For example, *Escherichia coli*, a zoonotic pathogen, exhibits extensive multidrug resistance and can thrive and circulate in many different environments (Wang et al., 2017; Zhang et al., 2017).

To address the public health implications of ARGs, the European Union (EU) prohibited the use of antimicrobials as growth promoters in 2006. Similarly, China banned the inclusion of antibiotic additives in feed in 2020 (Ma et al., 2021). Pig manure is recognized as a substantial repository of ARGs (Duan et al., 2019), and its ARG and microbiome structures have been extensively studied. For example, ARGs have been identified in pigs and poultry from nine European countries, revealing compositional differences between these two animal groups (Munk et al., 2018). Additionally, studies have explored the microbiome and ARG structures in pigs from various industrialized feedlots in China (Wang et al., 2019), while a reference gene catalog for the pig gut microbiome has been established based on pig manure samples from France, Denmark, and China (Xiao et al., 2016). Although antimicrobial usage on pig farms is subject to regional and national variation, most previous studies have not compared ARG and microbiome structures in pigs across different countries or regions. A comprehensive understanding of ARGs and microbiome distributions in swine at a country/regional scale could more effectively assist policymakers in developing strategies to prevent and control ARG transmission.

Monitoring ARGs is an essential step in addressing AMR. ARGs originating from pig manure present a potential threat to human health worldwide (He et al., 2016). Various techniques, such as metagenomic analysis and quantitative polymerase chain reaction (qPCR), have been employed to assess the risk of ARG dissemination in pig farms and more extensive ARG contamination (Fang et al., 2018; Li et al., 2020; Pu et al., 2018; Yang et al., 2022b). Although metagenomics can offer comprehensive insights, its prohibitive costs limit its applicability for routine ARG surveillance in pig farms; therefore, qPCR remains the preferred approach. Previous studies have identified indicators for total ARG abundance in the pig gut microbiome, although they relied on a limited number of samples (16–24) (Joyce et al., 2019; Qian et al., 2018; Zhao et al., 2018). Furthermore, a recent extensive metagenomic analysis based on 425 Chinese pig manure samples identified 10 highly accurate indicators of AMR in pigs (Zhou et al., 2022). Nevertheless, these indicators do not adequately capture the variations in ARG profiles between Chinese and European pigs, indicating the pressing need for a targeted ARG subset that can assess regional ARG contamination and variation (Liu et al., 2019). Analytical tools used in prior studies for identifying indicative microbiome/ARG features have primarily relied on linear discriminant analysis (LDA) (Goecks et al., 2010). In contrast, the random forest (RF) machine learning algorithm excels in classification, prediction, and variable selection tasks, with an error rate much lower than that of LDA (Verikas et al., 2011). Recently, the application of RF has expanded to include the prediction of clinical diseases, prediction of drug resistance phenotypes of bacteria, and detection and prediction of global sewage AMR (Hendriksen et al., 2019; Parisot et al., 2018; Su et al., 2019).

In the current study, we utilized a publicly available reference gene catalogue for pigs, as well as pig metagenomic data from our laboratory, including well-characterized patterns of antimicrobial usage (Tong et al., 2022; Xiao et al., 2016). We aimed to: (i) decipher the association between the microbiome and ARG in pig manure under different antimicrobial pressure (different antimicrobial monitoring policies) based on metagenomics; (ii) identify a specific subset of ARG indicators in pig farms via the RF algorithm; and (iii) explore potential associations between ARGs and hosts through metagenomic assembled genomes (MAGs) obtained from samples.

MATERIALS AND METHODS

Sample collection

Publicly available pig manure metagenomic datasets from China, Denmark, and France were collected, including PRJEB11755 (*n*=287) and PRJNA750156 (*n*=35), the latter of which was provided by our laboratory. The patterns of antimicrobial use in these pigs have been well-characterized in a previous study (Supplementary Table S1) (Xiao et al., 2016). Based on national and regional policies governing antimicrobial regulation, pigs from France and Denmark were grouped as European, while pigs from China with high antimicrobial exposure were classified as Chinese. Additionally, metagenomic datasets from healthy individuals in Denmark (PRJEB4336, *n*=140) and China (PRJNA422434, *n*=109; PRJNA750156, *n*=4) were collected for comparative analysis of ARG profiles with pigs from their corresponding regions (Supplementary Table S2).

Metagenomic analysis

Raw sequence reads were filtered, then quality trimmed using Sickle (quality scores<20 or length<20 bp) (Joshi & Fass, 2011). Clean reads were *de novo* assembled based on *k*-mer size using CLC Genomics Workbench (v.10.0.1, CLC Bio, Denmark). Open reading frames (ORFs) of the assembled contigs were predicted using Prodigal (v.2.6.3) (Hyatt et al., 2010). ARG-like ORFs were searched against the Comprehensive Antibiotic Resistance Database (CARD) using BLASTN with an E-value threshold of $\leq 1\times10^{-5}$. Only ARG-like ORFs with an identity ≥80% and query coverage ≥70% were retained. The coverage of these ARG-like ORFs was determined by mapping them to the contigs with a minimum length coverage of 95% at 95% similarity using CLC Genomics Workbench. The number of bacterial cells in each metagenome dataset was generated using ARGs-OAP (v.2.0) (Yin et al., 2018). ARG abundance was normalized to the number of cells (copies/cell) in each metagenome dataset, calculated as follows (Yin et al., 2018):

$$\text{Coverage} = \sum_{1}^{n} \frac{N \times 150/L}{C} \quad (1)$$

Where N represents the number of reads mapped to ARG-like ORFs, 150 signifies the read length, L is the length of the target ARG-like ORF sequence, n denotes the number of ARG-like ORFs, and C is the bacterial cell number per metagenome dataset.

Plasmids, integrons, and ISs were identified using PlasmidFinder (Carattoli & Hasman, 2020), INTEGRALL database (Moura et al., 2009), and ISfinder (Siguier et al., 2006), respectively (accessed on 6 July 2020). The methods used for the normalization and calculation of MGE abundance were consistent with those used for the ARGs. Plasmid, chromosomal, and unclassified sequences from ARG-carrying contigs (ACCs) were predicted using PlasFlow (v.1.1.0)

(default parameters) (Krawczyk et al., 2018). The composition and relative abundance of the microbiome community were profiled using MetaPhlAn3 (v.3.0.2) (Beghini et al., 2021).

Metagenomic binning

MAGs were produced with the Binning_refiner module (parameter: --metabat2) in MetaWRAP (Song & Thomas, 2017; Uritskiy et al., 2018). Estimates of MAG completeness and contamination were performed using CheckM (v.1.1.2), with only those MAGs with completeness ≥80% and contamination ≤10% retained (Parks et al., 2015). The dRep program (v.3.2.2) was used to dereplicate the retained MAGs in both the Chinese and European pigs (parameters: -sa 0.99 and -pa 0.95) (Olm et al., 2017). Taxonomic affiliation for each MAG was performed using GTDB-Tk (v.1.7.0) based on the Genome Taxonomic Database (GTDB) with default parameters. Genomes were classified as novel strains if they exhibited an average nucleotide identity (ANI) of less than 99%, and as novel species based on an ANI of less than 95% or if no ANI was output by GTDB-Tk (Glendinning et al., 2020). Of note, *Shigella flexneri* was renamed as *Escherichia flexneri* as proposed in GTDB taxonomy (Sanford et al., 2021). BLASTN was used to screen the MAGs for acquired ARGs against the CARD database. PhyloPhlAn (v.3.0.2) was used to reconstruct the phylogenetic trees with 400 universal markers (Asnicar et al., 2020). The phylogenetic tree was visualized using the "ggtreeExtra" package (Xu et al., 2021).

Core resistome and microbiome

To identify ARGs consistently present in pig manure samples across China and Europe, the core resistome was defined based on occurrence in at least 95% of the metagenome datasets (Munk et al., 2018). Genera present in fewer than 95% of the metagenome datasets were excluded from the core microbiome. Changing trends in the total abundance of the core resistome were analyzed in both pig and human samples originating from the corresponding countries or regions.

ARG markers, biomarkers, and indicators

To identify the main discriminant features of ARGs (ARG markers) and microbiomes (biomarkers) between Chinese and European pigs, the relative abundances of ARGs and bacterial taxa (phylum, class, order, family, genus, and species level) were classified in the pig manure samples using the "randomForest" package (v.4.7-1.1) (Liaw & Wiener, 2002). Cross-validation was performed to select significant features, while varImpPlot was used to illustrate the importance of features in the classification. The "rfPermute" package (v.2.5.1) was used to assess the significance of important features, while the "ggplot2" package (v.3.3.6) was used to visualize important features and cross-validation curves. In the RF models, the significant features served as indicators for ARG contamination on pig farms. The "linkET" package (v.0.0.5) was used to estimate the importance and interpretability of the significant features in the ARG profiles of both the Chinese and European pigs to facilitate identification of key indicators.

To evaluate the ability of indicators to predict ARG abundance in Chinese and European pig manure, abundance matrices of resistance genes were downloaded from Chinese and European pig manure studies (Munk et al., 2018; Zhou et al., 2022). To reduce data noise, the low-antibiotic exposure group was excluded from the ARG abundance matrix of Chinese pigs. The abundance of ARGs within each sample was summed at the ARG subtype level. Spearman correlation analysis was then performed to assess the relationship between the abundance of indicators and total ARG abundance. The ARGs included in the selected indicators but not detected in the downloaded dataset were excluded from prediction analysis.

Statistical analysis

Richness and Shannon diversity indices were computed using the "picante" package. Principal component analysis (PCA) and permutational multivariate analysis (Adonis) were performed to reflect the variation in ARG profiles and bacterial community between Chinese and European pigs using the "vegan" package. Spearman correlation analysis was used to explore the correlations among core ARGs, ARG markers, and indicators with ARG profiles. The "ggplot2" package was used to construct violin plots, line charts, stacked plots, and scatter diagrams. Pie charts and Sankey diagrams were established using the "scatterpie" and "networkD3" packages, respectively. Nonparametric Kruskal-Wallis and Mann-Whitney tests were performed to compare differences in data, with $P<0.05$ considered statistically significant following false discovery rate (FDR) corrections.

RESULTS

ARG and microbiome response in Chinese and European pigs

We detected 19 ARG types and 291 ARG subtypes in the Chinese and European pigs (Supplementary Table S3). Both ARG abundance and diversity were significantly higher ($P<0.05$) in Chinese pigs than in European pigs (Supplementary Figure S1A). Notably, ARG abundance ranged from 1.57 to 7.73 copies/cell in Chinese pigs and from 0.31 to 2.36 copies/cell in European pigs. Similar to ARG distribution, MGE abundance and diversity were higher in the Chinese pigs ($P<0.05$; Supplementary Figure S1B), with MGE abundance ranging from 0.33 to 9.76 copies/cell in Chinese pigs and from 0.03 to 6.88 copies/cell in European pigs. ARG abundance on the chromosomes was markedly higher than that on the plasmids in both Chinese and European pigs ($P<0.05$; Figure 1A; Supplementary Table S4). Furthermore, ARG abundance on the plasmids was significantly higher in Chinese pigs than in European pigs ($P<0.05$). Genes conferring resistance to beta-lactam (45.4%–68.7%), diaminopyrimidine (56.7%–74.2%), and phenicol (65.2%–68.8%) were more prevalent on plasmids, while genes conferring resistance to fluoroquinolone (60.5%–76.2%), nucleoside (60.0%–86.0%), and peptide (72.4%–74.3%) were more frequently encoded on chromosomes in Chinese and European pigs (Figure 1B). Moreover, PCA showed that Chinese and European pigs could be well separated based on ARG subtype composition and bacterial structure (genera) (Adonis, $P<0.05$; Figure 1C, D). Aminoglycoside, tetracycline, and macrolide-lincosamide-streptogramin (MLS) were the dominant ARG types in Chinese pigs, while beta-lactam, tetracycline, and MLS were the dominant ARG types in European pigs (Supplementary Figure S1C). Bacteroidetes and Firmicutes were the dominant phyla in all samples (Supplementary Figure S1D).

Core resistome and microbiome in Chinese and European pigs

To identify the core ARGs and microbial species in the

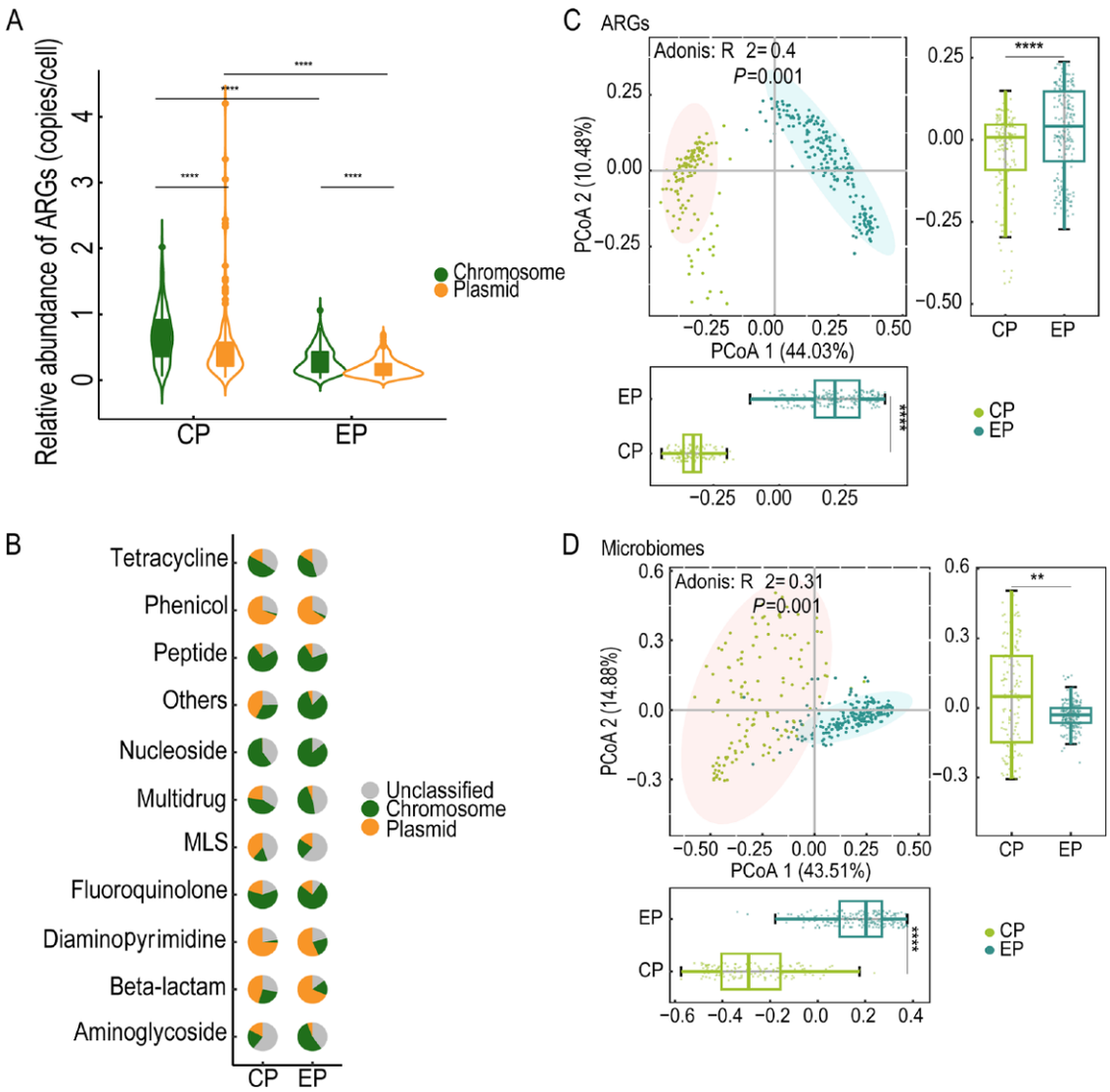


Figure 1 Location of ACCs and differences in ARG subtypes and microbiomes between groups

A: Abundance of ACCs located on plasmids and chromosomes (Mann-Whitney tests, *: P<0.05; **: P<0.01; ***: P<0.001; ****: P<0.0001). B: Percentages of various ARG types located on plasmids, chromosomes, and unclassified sequences. MLS: macrolide-lincosamide-streptogramin. C, D: PCA and Adonis analysis based on ARGs (C) and microbiomes (D) in Chinese and European pigs. Ellipses were drawn with a 0.95 probability. CP: Chinese pig; EP: European pig.

Chinese and European pigs, we analyzed ARGs and species consistently present in at least 95% of samples across the different locations. In total, seven core ARGs were identified in Chinese pigs, including *ACI-1*, *ANT(6)-Ib*, *APH(3')-IIIa*, *CfxA2*, *ErmB*, *lsaE*, and *tet*(40), and four core ARGs were identified in European pigs, including *ANT(6)-Ib*, *APH(3')-IIIa*, *lnuC*, and *tet*(40) (Figure 2A, B). Although these core ARGs constituted a relatively small proportion of the overall ARG types, they were highly abundant, representing 53.5% and 48.0% of total ARG abundance in Chinese and European pigs, respectively (Figure 2B; Supplementary Figure S2A). A robust correlation existed between core resistome and total ARG abundance in European pigs (Spearman ρ=0.91, P<0.001; Figure 2C), while a significant but moderate correlation was observed in Chinese pigs (Spearman ρ=0.52, P<0.001; Figure 2C).

To elucidate the genetic composition of the core resistome, we assessed the presence of the ARGs on either plasmid or chromosomal contigs, as well as their co-occurrence with bacterial genera (Figure 2D; Supplementary Figure S2B). Excluding unclassified categories, a higher proportion of core ARGs targeting beta-lactam (62.3%) and MLS (98.0%) was found on the plasmids, while a higher proportion of core ARGs targeting tetracycline (100%), aminoglycoside (86.8%), and multidrug (80.1%) was found on the chromosomes in Chinese pigs. In contrast, in European pigs, aminoglycoside and tetracycline were more often found on chromosomes, while MLS was relatively evenly distributed on both plasmids (51.9%) and chromosomes (48.1%). The hosts for the core resistome were identified based on the recovered MAGs. Notably, Bacteroidota and Firmicutes_A served as the principal hosts for the core resistome in both the Chinese and European pig populations.

Interestingly, the observed trend in the abundance of the core resistome in Chinese pigs closely paralleled that in the Chinese human population, with the exception of *APH(3')−IIIa*. Similarly, the observed trend in the abundance of the core resistome in European pigs closely paralleled that in the European human population, with the exception of *tet*(40) (Figure 2F). Notably, these core antibiotic resistance genes were detected at a lower frequency in human samples

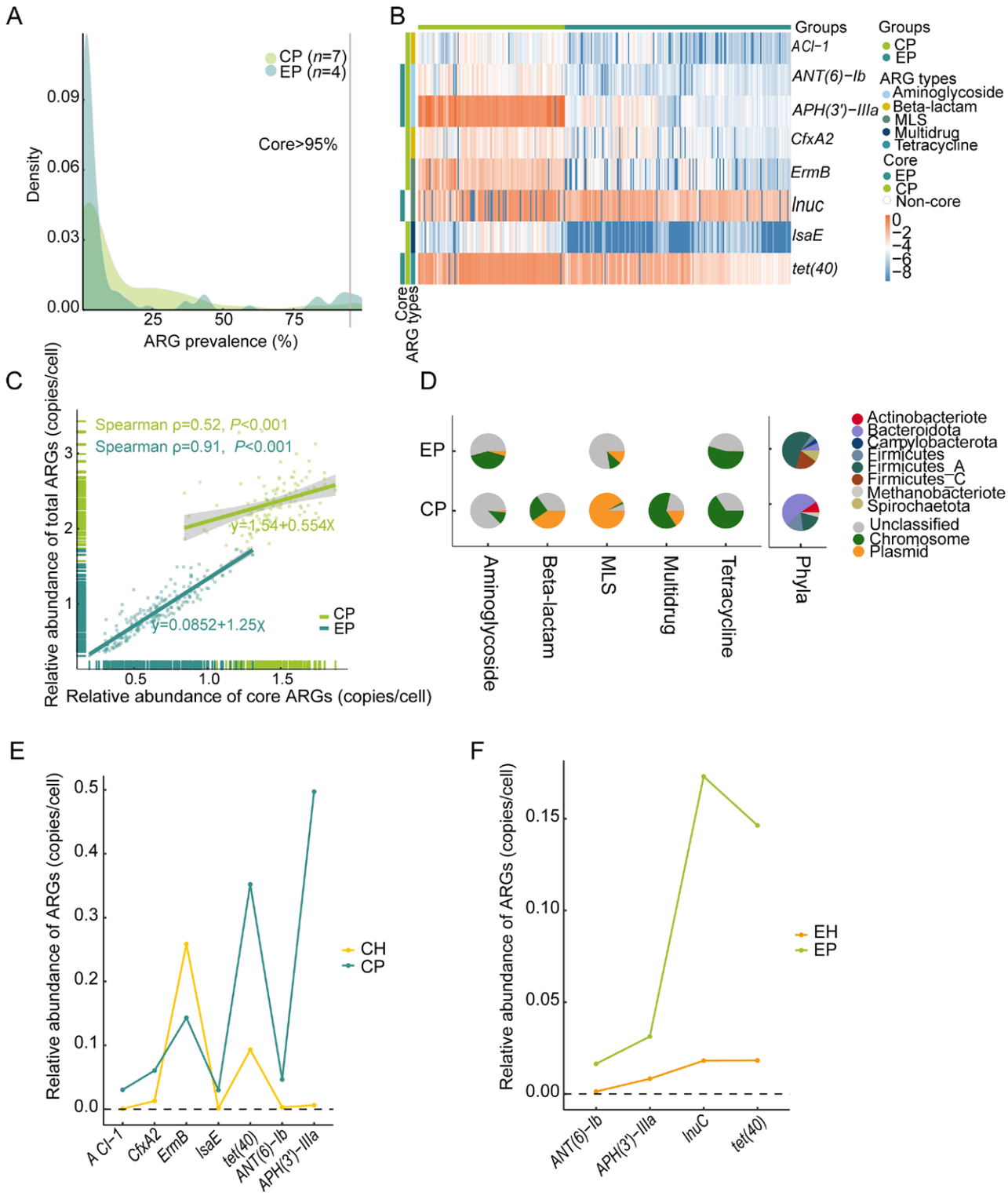


Figure 2 Core resistome of Chinese and European pigs

A: Distribution of ARG subtype prevalence in all manure samples from Chinese and European pigs. Gray lines show defined cutoffs. B: Quantitative profile of core resistome (coverage, ×/cell, $\log_2$ transferred). C: Correlation analysis of total ARG abundance and core resistome abundance. D: Proportion of locations and potential hosts in core resistome. E, F: Abundance of core ARGs in pigs and humans from China (E) and (F) Europe, respectively. Core ARGs were identified in pigs from China and Europe, respectively.

(Supplementary Figure S2C, D).

Using the same thresholds established for the core resistome, we identified 11 core microbiomes (Supplementary Figure S3A), including seven from European pigs and three from Chinese pigs. *Lactobacillus* and *Prevotella* emerged as the dominant genera in the core microbiomes of Chinese and European pigs, respectively, while *Prevotella_sp_P3_122* was shared between both (Supplementary Figure S3C). The core microbiomes were present in high abundance in samples from the corresponding groups (Supplementary Figure S3B, C). No ARGs were detected in the core microbiomes.

Potential drivers for distinguishing Chinese and European pigs

Using RF machine learning, we identified key ARGs and microbial species that may serve as potential discriminators between Chinese and European pigs. Ten-fold cross-validation, repeated five times, was employed to evaluate the importance of each ARG subtype. The cross-validation error rate reached stability when the 40 most significant ARG subtypes, out of a total of 291, were included in the analysis; these 40 ARGs were consequently designated as marker ARGs (Figure 3A). Not all ARG subtypes contributed equally to the distinction between Chinese and European pigs, with *APH(3')-IIIa*, *fexA*, *optrA*, *SAT-4*, and *ErmB* emerging as the most important variables distinguishing between Chinese and European pigs (P<0.05; Figure 3A; Supplementary Table S5). The initial classification model demonstrated exceptional accuracy in differentiating between Chinese and European pigs, with an error rate of 0% and a recall rate of 100% (Supplementary Table S5). To simplify the classification model, the top 40 important ARG subtypes were selected for modeling instead of all ARG subtypes in the original dataset. The simplified classification model showed that the two pig populations could be accurately distinguished based on the 40 most important ARG subtypes, maintaining an error rate of 0% and a recall rate of 100% (Adonis, P<0.05; Figure 3B; Supplementary Table S6). These marker ARGs were highly abundant in both pig populations (Figure 3B, C), contributing 79.3% and 72.0% of the total ARG abundance in Chinese and European pigs, respectively (Supplementary Figure S4). Thus, these marker ARGs could potentially serve as indicators of ARG contamination in both pig populations (Spearman ρ=0.96, P<0.05) (Figure 3D).

RF analysis was used to identify Chinese and European pigs at the phylum, class, order, family, genus, and species levels. Our model showed 98.7% accuracy at the family level, the highest among all taxonomic levels (Supplementary Table S7). Ten-fold cross-validation, repeated five times, was also carried out to evaluate the importance of indicator bacterial families. The cross-validation error curve stabilized when the 43 (43/80) most relevant families were used (Supplementary Figure S5A). The contributions of bacterial families to differentiating between Chinese and European pigs are reported in Supplementary Table S7. Similarly, upon simplification of the classification model, Chinese and European pigs could be accurately differentiated, exhibiting recall rates of 97.4% and 98.4%, respectively, with an error rate of 1.30% (Adonis, P<0.05; Supplementary Figure S5B; Supplementary Table S8). Therefore, these findings indicate that the 43 most relevant families could be used as biomarkers to distinguish between Chinese and European pigs. Notably, these biomarkers accounted for 99.3% and 99.8% of the total microbiome abundance in Chinese and European pigs, respectively (Supplementary Figure S5C).

Indicators for ARG contamination in Chinese and European pigs

To identify the ARGs that affected dissimilarities in ARG profiles between Chinese and European pigs, multiple regression models were used to determine the correlation between total ARG abundance of Chinese and European pigs and the identified ARG markers (Figure 4A). Analysis revealed that the ARG markers accounted for 73.7% and 76.3% of the variation in ARG profiles of the Chinese and European pigs, respectively. *ErmB*, *AAC(6')-Ie-APH(2'')-Ia*, *APH(3')-Ia*, *sul3*, and *tet*(A) were the most significant variables in predicting the variance in total ARG abundance in Chinese pigs, with importance values exceeding 7 (P<0.05; Figure 4A; Supplementary Table S9), while *tet*(40), *ErmF*, *tet*(X10), *CfxA6*, and *ErmG* were the most significant predictors of variance in European pigs, with importance values exceeding 13 (P<0.05; Figure 4A; Supplementary Table S10). Thus, these significantly important ARG markers (P<0.05) could serve as indicators of total ARG abundance in both pig populations, as corroborated by Spearman correlation analysis (Spearman ρ=0.72, P<0.05, Chinese pigs, Figure 4B; Spearman ρ=0.88, P<0.05, European pigs, Figure 4B). Notably, these indicators accounted for 41.2% and 68.5% of the total ARG abundance in Chinese and European pigs, respectively (Supplementary Figure S6A). Regarding proportions of indicators, 25.0% in Chinese pigs and 23.0% in European pigs were assigned to plasmids (Figure 4C). Among the ARG classes, aminoglycoside (57.6%) was dominant in Chinese pigs, while MLS (46.5%) was dominant in European pigs. Furthermore, hosts for the ARG indicators were identified across nine bacterial phyla. In the Chinese pig population, Spirochaetota and Bacteroidota constituted the dominant host phyla, with respective proportions of 25.5% and 20.2%, while *E. flexneri* (phylum Proteobacteria) harbored the greatest number of ARG indicators. In contrast, in the European pig population, Bacteroidota and Firmicutes_A emerged as the principal hosts, accounting for 40.2% and 38.2% of the host range, respectively, with *Streptococcus alactolyticus* (phylum Cyanobacteria) identified as a potential host for the ARG indicator *tet*(X10) (Figure 4C; Supplementary Figure S6B).

To corroborate the predictive capacity of the selected ARGs in regard to total ARG abundance in both the Chinese and European pig populations, validation datasets were employed, specifically ARG datasets from nine European countries and an expanded gene catalog for Chinese pigs (Munk et al., 2018; Zhou et al., 2022). In the Chinese validation dataset, only 12 out of 18 indicators (i.e., *AAC(6')-Ie-APH(2'')-Ia*, *aadA2*, *APH(3')-Ia*, *dfrA12*, *emr*B, *Erm*B, *flo*R, *OXA-347*, *qacH*, *QnrS1*, *sul*1, and *sul*3) were found to be predictive of total ARG abundance, with a strong correlation observed between their abundance and total ARG abundance (Spearman ρ=0.61–0.80, P<0.001; Supplementary Figure S7A). In the European validation dataset, only five out of 13 indicators (*CfxA2*, *CfxA6*, *lnuC*, *tet*(40), and *ACI-1*) were identified, with a strong correlation observed between their abundance and total ARG abundance (Spearman ρ=0.89, P<0.001; Supplementary Figure S7B).

MAGs recovered from Chinese and European pigs

To investigate the microbiome structure, including uncultured taxa, we reconstructed 25 466 MAGs from 322 manure samples collected from Chinese and European pigs. Subsequent to dereplication and stringent quality control measures (completeness ≥80%, contamination ≤10%), 782 MAGs from Chinese pigs and 1 245 MAGs from European pigs were retained for downstream analysis (Supplementary Figure S8A; Supplementary Table S11). Based on taxonomic assignment using GTDB-Tk, we identified 18 phyla among the 2 027 retained MAGs (Supplementary Table S12). Firmicutes_A (n=338/387) and Bacteroidota (n=197/452) were the dominant phyla in Chinese and European pigs. Notably, Planctomycetota was endemic to Chinese pigs, while

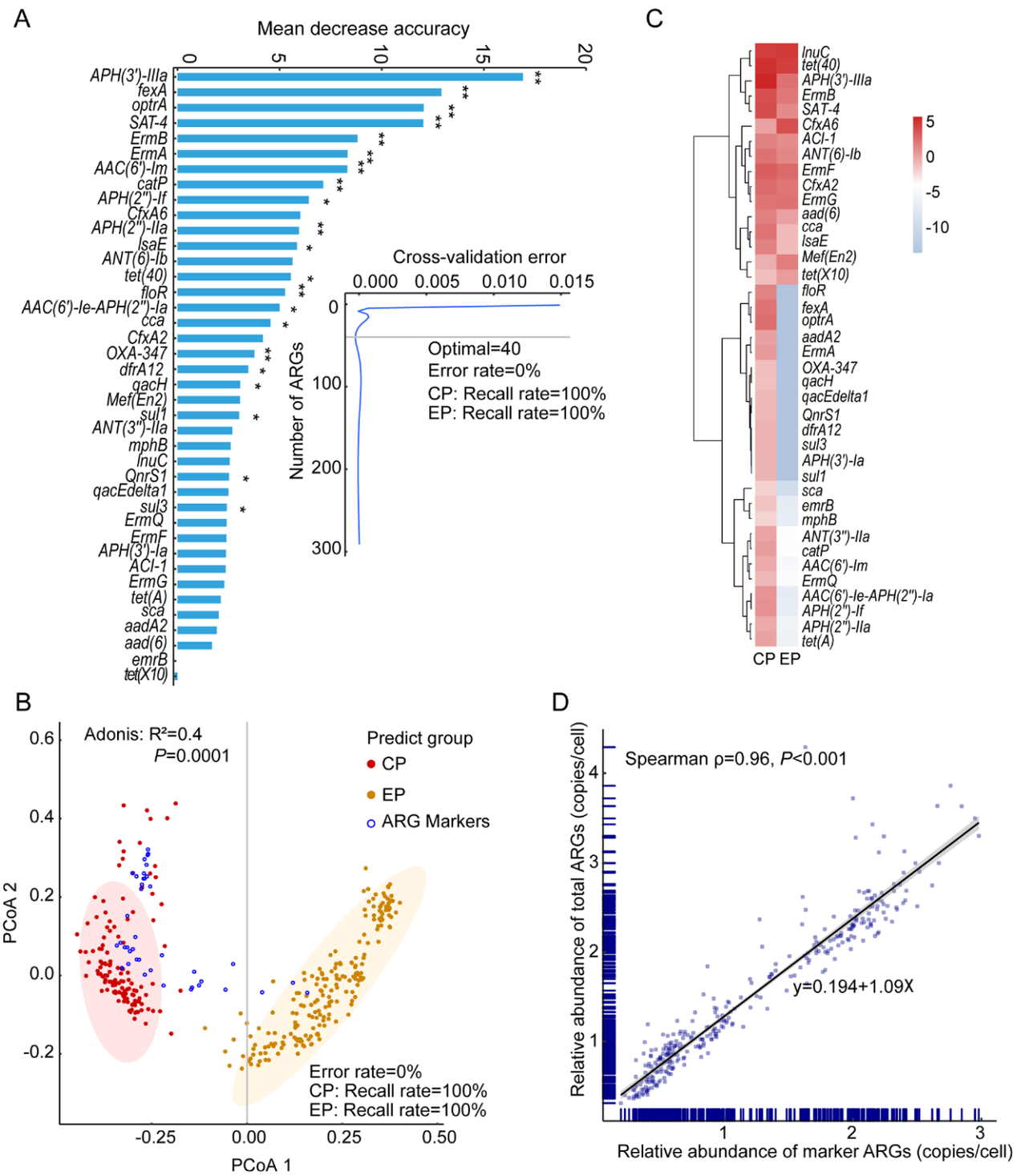


Figure 3 ARG markers between Chinese and European pigs

A: Top 40 ARG subtypes identified by RF classification of relative abundance of ARGs in Chinese and European pigs, ranked in descending order of importance to the accuracy of the model. Inset represents ten-fold cross-validation error for number of input ARGs that differentiate Chinese and European pigs. Higher mean decrease accuracy implies more important predictors. Significance levels: *: P<0.05; **: P<0.01. B: PCA and Adonis analysis were based on 40 ARG markers in Chinese and European pigs. C: Quantitative profile of 40 ARG markers in Chinese and European pigs (coverage, ×/cell, $\log_2$ transferred). cca: *Campylobacter_coli*_chloramphenicol_acetyltransferase, sca: *Streptococcus_suis*_chloramphenicol_acetyltransferase. D: Correlation analysis of total ARG abundance and ARG marker abundance.

Thermoplasmatota and Deferribacterota were endemic to European pigs. Among the bacterial genomes, 294 species-level MAGs from Chinese pigs and 373 species-level MAGs from European pigs were inferred as novel species, predominantly assigned to Firmicutes_A (Figure 5A; Supplementary Table S13). Among the relatively rare archaeal genomes, four putative new species (4/9) were identified in the Chinese pigs, whereas no new archaeal species were identified in the European pigs. Among the retained MAGs, 176 MAGs carried ARGs, including 104 MAGs (104/782) from Chinese pigs belonging to 10 phyla and 72 MAGs (72/1 245) from European pigs belonging to nine phyla (Supplementary

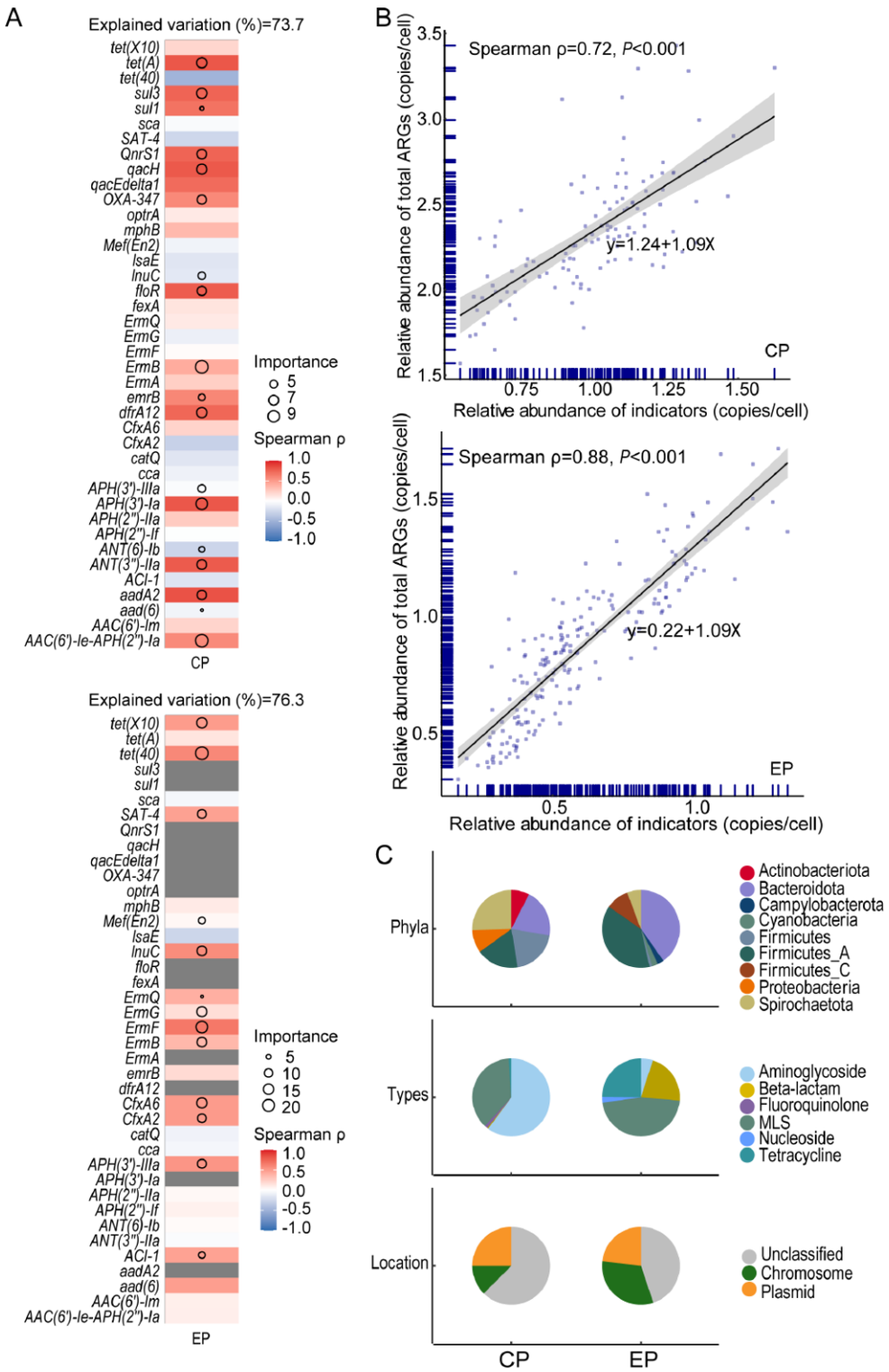


Figure 4 Indicators of ARG contamination in Chinese and European pigs

A: Potential contributions of ARG markers to ARG profiles and best multiple regression model for ARG markers in Chinese and European pigs, respectively. Proportion of explained variation was calculated via multiple regression modeling and variance decomposition analysis. Circle size represents importance of variable (P<0.05), with significantly important variable defined as indicators. Color represents Spearman correlation coefficient. cca: *Campylobacter_coli*_chloramphenicol_acetyltransferase, sca: *Streptococcus_suis*_chloramphenicol_acetyltransferase. B: Correlation analysis of total ARG abundance and indicator ARG abundance. C: Proportion of potential hosts, ARG types, and locations in indicator ARGs.

Figure S8B, C and Table S14). Firmicutes_A was the dominant phylum hosting ARGs in both populations, with 28 MAGs from Chinese pigs and 27 MAGs from European pigs. Phylogenetic analysis revealed a conspicuous presence of ARGs carried by *E. flexneri* (Proteobacteria) in the Chinese and European pigs (Figure 5A). Specifically, 47 ARGs were detected across 17 *E. flexneri* genomes, with the number of carried ARGs ranging from 22 to 35 in Chinese pigs and from

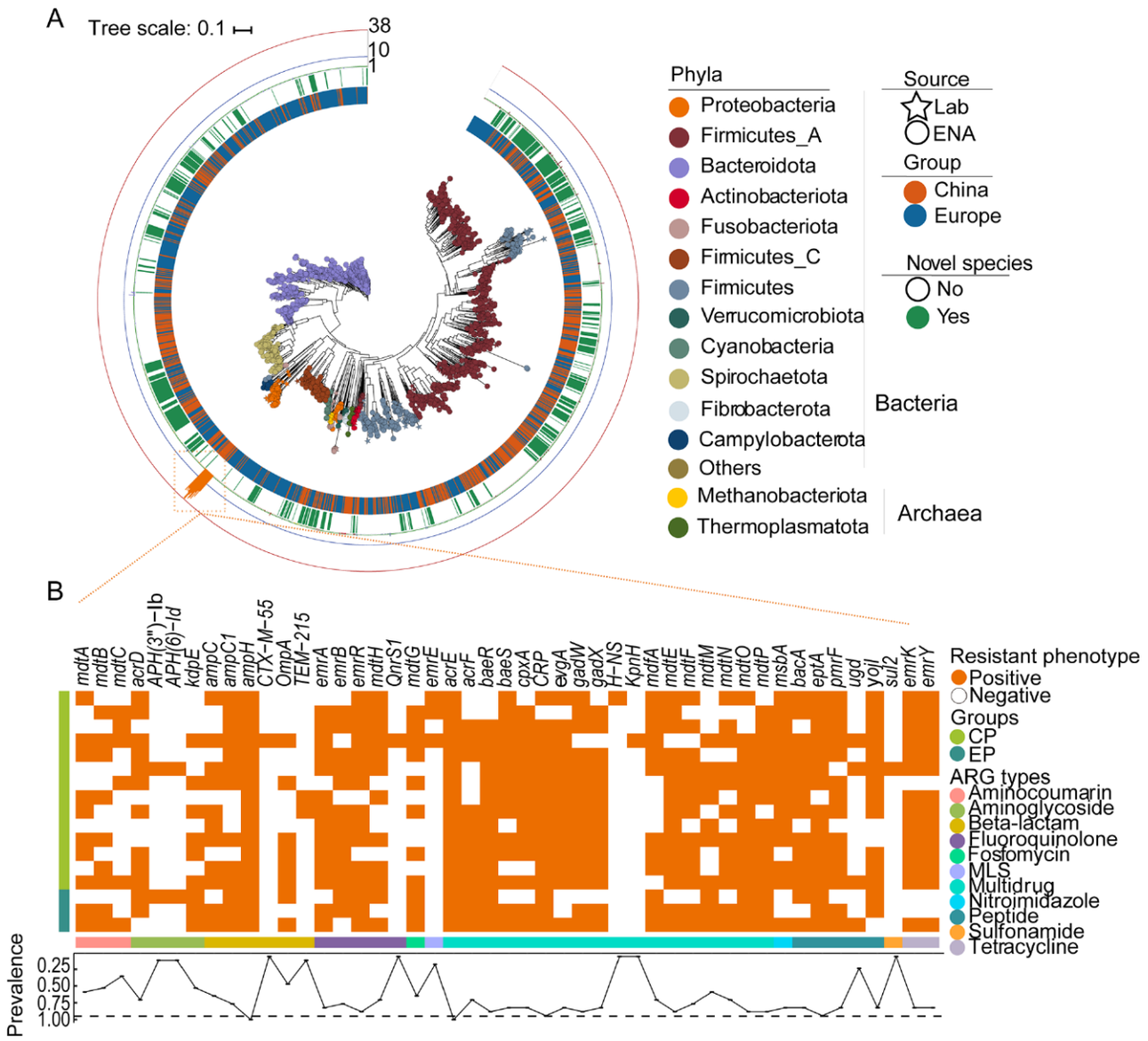


Figure 5 Metagenome-assembled nonredundant genomes (2 027) from pig manure microbiomes (322)

A: Phylogenetic tree of 2 015 bacterial and 12 archaeal species. From inner to outer rings: colors and shapes in first ring represent phyla and source, respectively; second ring represents groups; third ring represents novel species; and fourth ring represents number of ARGs carried by each species. B: ARGs detected in *E. flexneri* genomes. Line chart illustrates ARG prevalence in each *E. flexneri* genome.

21 to 37 in European pigs (Figure 5B). Notably, *acrE* (conferring multidrug resistance) and *ampH* (conferring beta-lactam resistance) occurred in all *E. flexneri* genomes. Multidrug resistance genes were highly prevalent in the *E. flexneri* genomes (18/47).

DISCUSSION

The association between the abundance of ARGs, MGEs, and microbiome structure was strongly influenced by the degree of antibiotic utilization. Our results showed that Chinese pigs exhibited greater diversity and abundance of ARGs and MGEs compared to European pigs. Previous studies have indicated that total levels of AMR are correlated with the use of antimicrobial agents in livestock and geographic regions with similar usage patterns display comparable resistance profiles (Munk et al., 2018). The abundance of ARGs on plasmids was much higher in Chinese pigs than in European pigs. Under selective antibiotic pressure, plasmids can indicate potential mobility and may effectively contribute to the spread of antibiotic resistance (Zhao et al., 2020). Recent studies have reported a significant association between ARG and MGE abundance (Li et al., 2022; Liu et al., 2018; Xie et al., 2022). Consequently, the higher levels of antimicrobial exposure in Chinese pigs, particularly in the absence of rigorous monitoring, may pose a greater risk to public health than in European pigs, where antibiotic use is more tightly regulated. Although ARGs were primarily located on chromosomes across all breeding modes (excluding those unclassified), the HGT mechanism can incorporate ARGs into chromosomes. Furthermore, free DNA released from the chromosomes can then be captured by other bacterial cells, thereby contributing to the spread of resistance (Croucher et al., 2016).

To determine potential response targets of ARGs and microbiomes in pigs under different antimicrobial exposure, we identified core ARGs in both Chinese and European pig populations. Of note, only eight core ARGs were identified in the two populations, implying that the majority of ARGs were

metastatic rather than constitutive in nature (Zhou et al., 2022). *ANT(6)−Ib*, *APH(3')−IIIa*, and *tet*(40) were identified as shared core ARGs across the different antimicrobial exposure regimes. These shared core ARGs were associated with commonly used classes of antibiotics, such as aminoglycosides and tetracyclines, which are the most widely used antibiotics for therapeutic and growth promotion in livestock worldwide (He et al., 2020). Notably, a correlation was found in the abundance of identified core ARGs between pig and human populations within corresponding geographic regions. Previous research has indicated that ARGs can be present in pigs even in the absence of antibiotic administration, highlighting the risk of ARG transmission from animal feces to both environmental reservoirs and human populations (Joyce et al., 2019). Moreover, bacteria harboring ARGs in manure have the potential for dissemination via multiple vectors, including rivers, groundwater, air dust, and agricultural products, thereby exacerbating ARG contamination on both regional and global scales (Gao et al., 2020a, 2020b; Smillie et al., 2011).

Our study revealed marked dissimilarities in both the ARG profiles and microbiome structures between the Chinese and European pigs. Using high-resolution machine learning models, we identified specific ARG markers and biomarker taxa, capable of distinguishing between these two pig populations. Antimicrobial exposure is known to significantly affect gut bacterial composition (Looft et al., 2014; Sun et al., 2020). As ARGs and microbiomes in pig manure are subject to distinct antibiotic exposure profiles, these factors provide a basis for tracking ARG contamination in pig farming operations. The environmental persistence and transmission of ARGs further support the need to establish ARG indicators for monitoring in pig farms (Li et al., 2019). Future ARG monitoring strategies could thus be based on recognized ARG indicators (Ishii, 2020). In this context, we identified specific high-resolution indicators of ARG contamination in Chinese and European pigs, pinpointing Spirochaetota, Bacteroidota, and Firmicutes_A as potential preferred hosts. Notably, *tet*(X10), which confers resistance to tigecycline, emerged as a significant indicator of ARG contamination in European pigs. Despite the EU's 2006 ban on the use of antibiotics as growth promoters, tetracycline usage in European livestock remained high at 41.2% between 2015 and 2017, compared to 31.2% in Asia (Fang et al., 2020). Furthermore, studies have reported higher detection rates of *tet*(X) variants in Europe than in Asia (Zhang et al., 2021). Antibiotic resistant bacteria may serve as potential vectors for the transfer of ARGs between environmental reservoirs and human or animal hosts when they persist across different geographical locations (Nnadozie & Odume, 2019). Therefore, the identification of indicators of ARG contamination and their respective hosts may provide a potential avenue for tracing the sources of ARG contamination in the environment and assessing ARG prevalence in pig farms.

The recovery of genomes from metagenomic datasets facilitates detailed strain classification and analysis of characteristics related to ARGs carried by bacteria (Almeida et al., 2021). In the current study, we examined the features of ARGs carried by the microbiome in both Chinese and European pigs through the examination of recovered medium-quality MAGs. Results showed that Firmicutes_A and Bacteroidota were prevalent in pig manure samples from both regions, consistent with previous research on pig and chicken gene catalogs showing Firmicutes_A and Bacteroidota dominance (Chen et al., 2021; Feng et al., 2021). Furthermore, Firmicutes_A was identified as the dominant host of ARGs in both Chinese and European pigs, in accordance with previous studies (Wang et al., 2021). Notably, *E. flexneri*, belonging to the phylum Proteobacteria, was found to harbor an extensive array of ARGs (ranging from 21 to 37), primarily constituting multidrug resistance genes, in both Chinese and European pigs. The average number of ARGs carried by animal-derived *E. coli* in China has risen from 8.05 to 16.85 over the last five decades, leading to excessive and extensive use of antimicrobial agents (Yang et al., 2022a). The indiscriminate use of antimicrobials has been linked to elevated levels of multidrug resistance in *E. coli* released into the environment near pig farms (Park et al., 2017). Of note, E. flexneri in European pigs, despite being subject to monitored antimicrobial administration, harbored ARG counts comparable to those in Chinese pigs, which are under unmonitored regimes. This underscores the role of *E. flexneri* as a zoonotic pathogen and a critical vector for ARG transmission (Zhang et al., 2020), as well as the urgent need for effective wastewater treatment strategies in pig farms worldwide (Cheng et al., 2020). The WHO has emphasized the necessity for a globally coordinated effort to address the escalating issue of AMR (Antimicrobial Resistance Collaborators, 2022).

In this study, we utilized metagenomic analyses and RF models to identify indicators of ARG contamination in both Chinese and European pig populations. However, we acknowledge that further improvements in analytical accuracy and an expanded sample set are required for a more comprehensive evaluation of ARG contamination within the pig industry. Consequently, the results presented here may contain some degree of uncertainty.

CONCLUSIONS

We identified ARGs and microbiome structures in pig manure across different countries and regions, revealing variations determined by antimicrobial usage.. Elevated antimicrobial pressure was found to promote ARG enrichment and potential migration. Aminoglycoside and tetracycline resistance genes were stable across the Chinese and European pig manure samples. Notably, the abundance of core ARGs identified in pigs may have implications for similar changes in humans. The core microbiome in pig manure exhibited geographical dependence. Utilizing high-resolution machine learning models, we identified specific ARG markers and biomarker taxa capable of discriminating between Chinese and European pig manure samples. We additionally identified indicators for accurately assessing AMR burden in both Chinese and European pig populations. MAG-based analysis demonstrated no specific host-association resistance pattern under different antimicrobial pressures and identified *E. flexneri* as exhibiting a multiple resistance pattern. These results emphasize the importance of collective global action aimed at reducing AMR risk and provide a reference subset of indicator ARGs for future routine AMR assessments in pig farms.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

C.H.T.: Visualization, Writing - Original draft preparation. Z.P.H. and L.D.: Data curation and Investigation. D.Y.X. and R.N.Z.: Conceptualization and Methodology. Z.L.Z. and W.G.X.: Conceived the study. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

The authors thank the anonymous reviewers for their insightful suggestions and comments.

REFERENCES

Almeida A, Nayfach S, Boland M, et al. 2021. A unified catalog of 204, 938 reference genomes from the human gut microbiome. *Nature Biotechnology*, **39**(1): 105−114.

Antimicrobial Resistance Collaborators. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, **399**(10325): 629−655.

Asnicar F, Thomas AM, Beghini F, et al. 2020. Precise phylogenetic analysis of microbial isolates and genomes from metagenomes using PhyloPhlAn 3.0. *Nature Communications*, **11**(1): 2500.

Beghini F, McIver LJ, Blanco-Míguez A, et al. 2021. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. *eLife*, **10**: e65088.

Carattoli A, Hasman H. 2020. PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS). *In*: de la Cruz F (eds). Horizontal Gene Transfer. Methods in Molecular Biology, Vol 2075. New York, NY: Humana, 285–294.https://doi.org/10.1007/978-1-4939-9877-7_20.

Chen CY, Zhou YY, Fu H, et al. 2021. Expanded catalog of microbial genes and metagenome-assembled genomes from the pig gut microbiome. *Nature Communications*, **12**(1): 1106.

Cheng DL, Ngo HH, Guo WS, et al. 2020. A critical review on antibiotics and hormones in swine wastewater: water pollution problems and control approaches. *Journal of Hazardous Materials*, **387**(5): 121682.

Croucher NJ, Mostowy R, Wymant C, et al. 2016. Horizontal DNA transfer mechanisms of bacteria as weapons of intragenomic conflict. *PLoS Biology*, **14**(3): e1002394.

Duan ML, Gu J, Wang XJ, et al. 2019. Factors that affect the occurrence and distribution of antibiotic resistance genes in soils from livestock and poultry farms. *Ecotoxicology and Environmental Safety*, **180**: 114−122.

Fang H, Han LX, Zhang HP, et al. 2018. Dissemination of antibiotic resistance genes and human pathogenic bacteria from a pig feedlot to the surrounding stream and agricultural soils. *Journal of Hazardous Materials*, **357**: 53−62.

Fang LX, Chen C, Cui CY, et al. 2020. Emerging high-level tigecycline resistance: novel tetracycline destructases spread via the mobile Tet(X). *BioEssays*, **42**(8): 2000014.

Feng YQ, Wang YN, Zhu BL, et al. 2021. Metagenome-assembled genomes and gene catalog from the chicken gut microbiome aid in deciphering antibiotic resistomes. *Communications Biology*, **4**(1): 1305.

Gao FZ, He LY, He LX, et al. 2020a. Untreated swine wastes changed antibiotic resistance and microbial community in the soils and impacted abundances of antibiotic resistance genes in the vegetables. *Science of the Total Environment*, **741**: 140482.

Gao FZ, Zou HY, Wu DL, et al. 2020b. Swine farming elevated the proliferation of *Acinetobacter* with the prevalence of antibiotic resistance genes in the groundwater. *Environment International*, **136**: 105484.

Glendinning L, Stewart RD, Pallen MJ, et al. 2020. Assembly of hundreds of novel bacterial genomes from the chicken caecum. *Genome Biology*, **21**(1): 34.

Goecks J, Nekrutenko A, Taylor J, et al. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, **11**(8): R86.

Han XM, Hu HW, Chen QL, et al. 2018. Antibiotic resistance genes and associated bacterial communities in agricultural soils amended with different sources of animal manures. *Soil Biology and Biochemistry*, **126**: 91−102.

He LY, Ying GG, Liu YS, et al. 2016. Discharge of swine wastes risks water quality and food safety: antibiotics and antibiotic resistance genes from swine sources to the receiving environments. *Environment International*, **92–93**: 210–219.

He Y, Yuan QB, Mathieu J, et al. 2020. Antibiotic resistance genes from livestock waste: occurrence, dissemination, and treatment. *npj Clean Water*, **3**(1): 4.

Hendriksen RS, Munk P, Njage P, et al. 2019. Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature Communications*, **10**(1): 1124.

Hyatt D, Chen GL, Locascio PF, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, **11**: 119.

Ishii S. 2020. Quantification of antibiotic resistance genes for environmental monitoring: current methods and future directions. *Current Opinion in Environmental Science & Health*, **16**: 47−53.

Joshi NA, Fass JN. 2011. Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. https://github.com/najoshi/sickle.

Joyce A, McCarthy CGP, Murphy S, et al. 2019. Antibiotic resistomes of healthy pig faecal metagenomes. *Microbial Genomics*, **5**(5): e000272.

Klein EY, Van Boeckel TP, Martinez EM, et al. 2018. Global increase and geographic convergence in antibiotic consumption between 2000 and 2015. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(15): E3463−E3470.

Krawczyk PS, Lipinski L, Dziembowski A. 2018. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. *Nucleic Acids Research*, **46**(6): e35.

Li LG, Huang Q, Yin XL, et al. 2020. Source tracking of antibiotic resistance genes in the environment—Challenges, progress, and prospects. *Water Research*, **185**: 116127.

Li N, Zhu CX, Liu C, et al. 2019. The persistence of antimicrobial resistance and related environmental factors in abandoned and working swine feedlots. *Environmental Pollution*, **255**: 113116.

Li XJ, Rensing C, Vestergaard G, et al. 2022. Metagenomic evidence for co-occurrence of antibiotic, biocide and metal resistance genes in pigs. *Environment International*, **158**: 106899.

Liang JS, Mao GN, Yin XL, et al. 2020. Identification and quantification of bacterial genomes carrying antibiotic resistance genes and virulence factor genes for aquatic microbiological risk assessment. *Water Research*, **168**: 115160.

Liaw A, Wiener M. 2002. Classification and regression by randomForest. *R News*, **2–3**: 18–22.

Liu LM, Su JQ, Guo YY, et al. 2018. Large-scale biogeographical patterns of bacterial antibiotic resistome in the waterbodies of China. *Environment International*, **117**: 292−299.

Liu X, Xiao P, Guo YY, et al. 2019. The impacts of different high-throughput profiling approaches on the understanding of bacterial antibiotic resistance genes in a freshwater reservoir. *Science of the Total Environment*, **693**: 133585.

Looft T, Allen HK, Cantarel BL, et al. 2014. Bacteria, phages and pigs: the effects of in-feed antibiotics on the microbiome at different gut locations. *The ISME Journal*, **8**(8): 1566−1576.

Ma X, Yang ZR, Xu TT, et al. 2021. Chlortetracycline alters microbiota of

gut or faeces in pigs and leads to accumulation and migration of antibiotic resistance genes. *Science of the Total Environment*, **796**: 148976.

Moura A, Soares M, Pereira C, et al. 2009. INTEGRALL: a database and search engine for integrons, integrases and gene cassettes. *Bioinformatics*, **25**(8): 1096−1098.

Munk P, Knudsen BE, Lukjancenko O, et al. 2018. Abundance and diversity of the faecal resistome in slaughter pigs and broilers in nine European countries. *Nature Microbiology*, **3**(8): 898−908.

Nnadozie CF, Odume ON. 2019. Freshwater environments as reservoirs of antibiotic resistant bacteria and their role in the dissemination of antibiotic resistance genes. *Environmental Pollution*, **254**: 113067.

Olm MR, Brown CT, Brooks B, et al. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *The ISME Journal*, **11**(12): 2864−2868.

Parisot S, Ktena SI, Ferrante E, et al. 2018. Disease prediction using graph convolutional networks: application to Autism Spectrum Disorder and Alzheimer's disease. *Medical Image Analysis*, **48**: 117−130.

Park J, Gasparrini AJ, Reck MR, et al. 2017. Plasticity, dynamics, and inhibition of emerging tetracycline resistance enzymes. *Nature Chemical Biology*, **13**(7): 730−736.

Parks DH, Imelfort M, Skennerton CT, et al. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, **25**(7): 1043−1055.

Pu CJ, Liu H, Ding GC, et al. 2018. Impact of direct application of biogas slurry and residue in fields: in situ analysis of antibiotic resistance genes from pig manure to fields. *Journal of Hazardous Materials*, **344**: 441−449.

Qian X, Gu J, Sun W, et al. 2018. Diversity, abundance, and persistence of antibiotic resistance genes in various types of animal manure following industrial composting. *Journal of Hazardous Materials*, **344**: 716−722.

Sanford RA, Lloyd KG, Konstantinidis KT, et al. 2021. Microbial taxonomy run amok. *Trends in Microbiology*, **29**(5): 394−404.

Siguier P, Perochon J, Lestrade L, et al. 2006. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research*, **34**(S1): D32−D36.

Smillie CS, Smith MB, Friedman J, et al. 2011. Ecology drives a global network of gene exchange connecting the human microbiome. *Nature*, **480**(7376): 241−244.

Song WZ, Thomas T. 2017. Binning_refiner: improving genome bins through the combination of different binning programs. *Bioinformatics*, **33**(12): 1873−1875.

Su M, Satola SW, Read TD. 2019. Genome-based prediction of bacterial antibiotic resistance. *Journal of Clinical Microbiology*, **57**(3): e01405−18.

Sun J, Liao XP, D'Souza AW, et al. 2020. Environmental remodeling of human gut microbiota and antibiotic resistome in livestock farms. *Nature Communications*, **11**(1): 1427.

Tong CH, Xiao DY, Xie LF, et al. 2022. Swine manure facilitates the spread of antibiotic resistome including tigecycline-resistant *tet*(X) variants to farm workers and receiving environment. *Science of the Total Environment*, **808**: 152157.

Uritskiy GV, Diruggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*, **6**(1): 1−13.

Verikas A, Gelzinis A, Bacauskiene M. 2011. Mining data with random forests: a survey and results of new tests. *Pattern Recognition*, **44**(2): 330−349.

Wang CL, Li P, Yan QL, et al. 2019. Characterization of the pig gut microbiome and antibiotic resistome in industrialized feedlots in China. *mSystems*, **4**(6): e00206−19.

Wang J, Gu J, Wang XJ, et al. 2021. Enhanced removal of antibiotic resistance genes and mobile genetic elements during swine manure composting inoculated with mature compost. *Journal of Hazardous Materials*, **411**: 125135.

Wang Y, Zhang RM, Li JY, et al. 2017. Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. *Nature Microbiology*, **2**(4): 16260.

Xiao L, Estellé J, Kiilerich P, et al. 2016. A reference gene catalogue of the pig gut microbiome. *Nature Microbiology*, **1**(12): 16161.

Xie WY, Wang YT, Yuan J, et al. 2022. Prevalent and highly mobile antibiotic resistance genes in commercial organic fertilizers. *Environment International*, **162**: 107157.

Xu SB, Dai ZH, Guo PF, et al. 2021. ggtreeExtra: compact visualization of richly annotated phylogenetic data. *Molecular Biology and Evolution*, **38**(9): 4039−4042.

Yang L, Shen YB, Jiang JY, et al. 2022a. Distinct increase in antimicrobial resistance genes among *Escherichia coli* during 50 years of antimicrobial use in livestock production in China. *Nature Food*, **3**(3): 197−205.

Yang YW, Li LF, Huang F, et al. 2022b. The fate of antibiotic resistance genes and their association with bacterial and archaeal communities during advanced treatment of pig farm wastewater. *Science of the Total Environment*, **851**: 158364.

Yin XL, Jiang XT, Chai BL, et al. 2018. ARGs-OAP v2.0 with an expanded SARG database and Hidden Markov Models for enhancement characterization and quantification of antibiotic resistance genes in environmental metagenomes. *Bioinformatics*, **34**(13): 2263−2270.

Zhang P, Shen ZQ, Zhang CP, et al. 2017. Surveillance of antimicrobial resistance among *Escherichia coli* from chicken and swine, China, 2008–2015. *Veterinary Microbiology*, **203**: 49−55.

Zhang RM, Sun J, Sun RY, et al. 2021. Source tracking and global distribution of the tigecycline non-susceptible *tet*(X). *Microbiology Spectrum*, **9**(3): e0116421.

Zhang SQ, Abbas M, Rehman MU, et al. 2020. Dissemination of antibiotic resistance genes (ARGs) via integrons in *Escherichia coli*: a risk to human health. *Environmental Pollution*, **266**: 115260.

Zhao RX, Yu K, Zhang JY, et al. 2020. Deciphering the mobility and bacterial hosts of antibiotic resistance genes under antibiotic selection pressure by metagenomic assembly and binning approaches. *Water Research*, **186**: 116318.

Zhao Y, Su JQ, An XL, et al. 2018. Feed additives shift gut microbiota and enrich antibiotic resistance in swine gut. *Science of the Total Environment*, **621**: 1224−1232.

Zhou YY, Fu H, Yang H, et al. 2022. Extensive metagenomic analysis of the porcine gut resistome to identify indicators reflecting antimicrobial resistance. *Microbiome*, **10**(1): 39.

Wu et al. *Zool. Res.* 2024, 45(1): 201−214
https://doi.org/10.24272/j.issn.2095-8137.2023.021

ZR | Zoological Research

Article Open Access

Combined ATAC-seq, RNA-seq, and GWAS analysis reveals glycogen metabolism regulatory network in Jinjiang oyster (*Crassostrea ariakensis*)

Biao Wu[1,2,#], Xi Chen[1,2,#], Jie Hu[3,#], Zhen-Yuan Wang[1,4], Yan Wang[1], Da-You Xu[3], Hao-Bing Guo[3], Chang-Wei Shao[1,2], Li-Qing Zhou[1,2], Xiu-Jun Sun[1,2], Tao Yu[5], Xiao-Mei Wang[5], Yan-Xin Zheng[5], Guang-Yi Fan[3,6,*], Zhi-Hong Liu[1,2,*]

[1] *State Key Laboratory of Mariculture Biobreeding and Sustainable Goods*, *Yellow Sea Fisheries Research Institute*, *Chinese Academy of Fishery Sciences*, *Qingdao*, *Shandong* 266071, *China*
[2] *Laboratory for Marine Fisheries Science and Food Production Processes*, *Laoshan Laboratory*, *Qingdao*, *Shandong* 266071, *China*
[3] *BGI-Qingdao*, *BGI-Shenzhen*, *Qingdao*, *Shandong* 266426, *China*
[4] *National Demonstration Center for Experimental Fisheries Science Education*, *Shanghai Ocean University*, *Shanghai* 201306, *China*
[5] *Changdao Enhancement and Experiment Station*, *Chinese Academy of Fishery Sciences*, *Yantai*, *Shandong* 265800, *China*
[6] *State Key Laboratory of Agricultural Genomics*, *BGI- Shenzhen*, *Shenzhen*, *Guangdong* 518083, *China*

ABSTRACT

Glycogen serves as the principal energy reserve for metabolic processes in aquatic shellfish and substantially contributes to the flavor and quality of oysters. The Jinjiang oyster (*Crassostrea ariakensis*) is an economically and ecologically important species in China. In the present study, RNA sequencing (RNA-seq) and assay for transposase-accessible chromatin using sequencing (ATAC-seq) were performed to investigate gene expression and chromatin accessibility variations in oysters with different glycogen contents. Analysis identified 9 483 differentially expressed genes (DEGs) and 7 215 genes with significantly differential chromatin accessibility (DCAGs) were obtained, with an overlap of 2 600 genes between them. Notably, a significant proportion of these genes were enriched in pathways related to glycogen metabolism, including “Glycogen metabolic process” and “Starch and sucrose metabolism”. In addition, genome-wide association study (GWAS) identified 526 single nucleotide polymorphism (SNP) loci associated with glycogen content. These loci corresponded to 241 genes, 63 of which were categorized as both DEGs and DCAGs. This study enriches basic research data and provides insights into the molecular mechanisms underlying the regulation of glycogen metabolism in *C. ariakensis*.

Keywords: *Crassostrea ariakensis*; Glycogen; Transcriptome; ATAC; GWAS

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

INTRODUCTION

Glycogen, a branched glucose polysaccharide, is ubiquitous across various taxa, including animals and fungi (Liu et al., 2021). It serves multifaceted functions in biological processes, such as energy storage, cell differentiation, signal transduction, redox regulation, and stem cell formation (Zhang et al., 2021). As the primary form of carbohydrate storage in animals, glycogen has garnered considerable attention regarding its role in energy storage and supply. In shellfish, glycogen functions as the most direct and effective energy reserve for metabolism (Liu et al., 2020; Ojea et al., 2004; Smolders et al., 2004). Beyond its biological functions, glycogen content has also been found to affect the flavor, fatness, and color of oyster meat, thereby affecting consumer preferences (Liu et al., 2020). Glycogen content has also been implicated in stress tolerance in oysters (Li et al., 2017). As such, it is considered a critical quality trait in the selection of new oyster varieties, such as *Crassostrea gigas* (Zhang et al., 2020). Overall, glycogen is critical for both the biological mechanisms of oysters and the economic viability of oyster aquaculture.

In 2021, oyster production in China reached 5.82 million tons, accounting for 37.6% of total shellfish production (Ministry of Agriculture and Rural Affairs of the People's Republic of China, 2022). The Jinjiang oyster (*Crassostrea*

Received: 10 May 2023; Accepted: 08 September 2023; Online: 09 September 2023
Foundation items: This work was supported by the National Key R&D Program of China (2022YFD2400105, 2018YFD0900104), Central Public-interest Scientific Institution Basal Research Fund, CAFS (2021XT0102, 2023TD30), Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (2021QNLM050103), Key Research and Development Project of Shandong Province (2021LZGC028), and National Marine Genetic Resource Center
[#]Authors contributed equally to this work
*Corresponding authors, E-mail: fanguangyi@genomics.cn; liuzh@ysfri.ac.cn

ariakensis) exhibits a broad geographic distribution in China, extending from the northern boundary with North Korea to the southern estuarine regions bordering Vietnam. Demonstrating adaptability to a broad spectrum of salinities and temperatures, *C. ariakensis* is well-suited for the aquaculture industry (Li et al., 2021; Wu et al., 2022). It is one of the five major oyster varieties currently cultured in China and is expected to undergo larger-scale cultivation in the near future (Peng et al., 2021; Zhang et al., 2020). Glycogen content in oysters can be influenced by environmental conditions and is closely associated with gonadal development stages (Figure 1A, C). Recent evidence suggests that glycogen, protein, and amino acid levels are subject to genetic control and inheritance (Meng et al., 2019). Given the widely varied and elevated glycogen content in oysters compared to other aquatic species (Figure 1B; Supplementary Table S1), they represent an ideal model for studying glycogen metabolism. However, the molecular regulatory mechanisms underlying glycogen metabolism in oysters remain inadequately understood. As such, more comprehensive research is needed to elucidate the regulation of glycogen metabolism in *C. ariakensis* given its biological and economic importance.

Over the past decade, rapid advances in sequencing technologies have enabled large-scale identification of candidate genes associated with different phenotypes (Swain et al., 2021). Multi-omics techniques, such as whole-genome sequencing, simplified genome sequencing, transcriptomics, proteomics, and metabolomics, have been widely applied to study genetic regulatory mechanisms in shellfish, including preliminary studies on the regulation of glycogen metabolism in *C. gigas*. For example, using transcriptomics, Wang et al. (2021b) identified several non-coding RNAs (ncRNAs) in the gonads of *C. gigas* postulated to indirectly regulate the expression of genes involved in glycogen metabolism, including protein phosphatase 1 regulatory subunit 3b (*ppp1r3b*), isocitrate dehydrogenase 1 (*idh1*), and β-glucosidase. In addition, based on transcriptomics and metabolomics, Chen et al. (2021a) identified several genes related to glycogen metabolism, including glycogen synthase (*gys*), solute carrier family 2 member 4 (*slc2a4*), and glycerol-3-phosphate acyltransferase 4 (*gpat4*). These findings highlight the feasibility and efficiency of multi-omics technologies in screening genes related to quality traits of aquatic products.

Multi-omics analyses can enhance the accuracy and efficiency of gene screening. In this study, an integrated approach combining RNA sequencing (RNA-seq), assay for transposase-accessible chromatin using sequencing (ATAC-

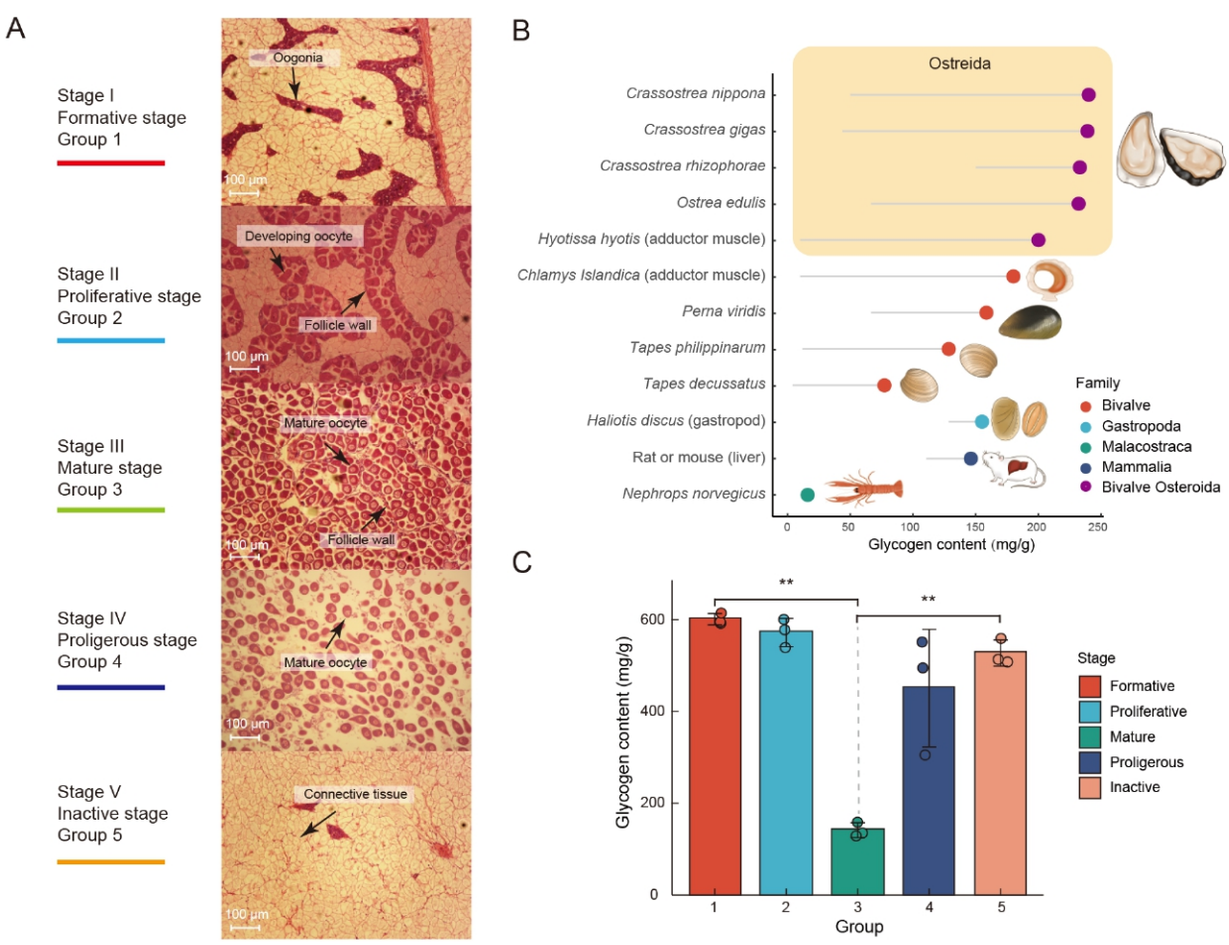


Figure 1 Developmental stages and glycogen content of *Crassostrea ariakensis* gonads

A: Sections of female *C. ariakensis* gonads at different developmental stages: Formative stage (Stage I), Proliferative stage (Stage II), Mature stage (Stage III), Proligerous stage (Stage IV), and Inactive stage (Stage V). B: Lollipop chart of glycogen content of different species (mg/g, dry weight). Head of lollipop represents highest value of glycogen content, and line represents range of glycogen content in certain species. Unless otherwise specified (e.g., specific tissue is mentioned in parentheses for some species), values indicate total glycogen content in all soft tissues of the species (source literature in Supplementary Table S1). C: Column chart of glycogen content in different oyster groups. **: $P<0.01$.

seq), and genome-wide association study (GWAS) was applied to screen for genes implicated in glycogen metabolism in *C. ariakensis*. To the best of our knowledge, this is the first study to employ ATAC-seq to examine the correlation between chromatin accessibility and gene expression related to glycogen metabolism in oysters. More importantly, the concurrent use of ATAC-seq and RNA-seq enabled analysis of whether differentially expressed genes (DEGs) exhibit significantly different chromatin accessibility around their transcription start sites (TSSs), and whether DEGs are regulated by transcription factors associated with specific motifs or footprints within open chromatin regions (Buenrostro et al., 2013). This multifaceted approach also revealed insights into the regulation of glycogen metabolism and cellular heterogeneity among different oyster groups from an epigenetic perspective (Yan et al., 2020). In addition, GWAS enabled the exploration and identification of genes and loci at the genome level, further facilitating analysis of the mechanisms underlying glycogen metabolism. These results contribute to the existing body of basic research data and provide novel insights into the molecular mechanisms governing glycogen metabolism regulation in *C. ariakensis*.

MATERIALS AND METHODS

Experimental oysters

The *C. ariakensis* samples were collected from five locations in China, including Binzhou and Dongying (Shandong Province), Luchaogang (Shanghai), Guangzhou (Guangdong Province), and Beihai (Guangxi Province). A total of 271 oysters were collected, with 250 used for genome resequencing, 15 used for RNA-seq and ATAC-seq, and six used for quantitative real-time polymerase chain reaction (qPCR) experiments. Six tissues, including the mantle, gill, lip, gonad, hepatopancreas and adductor muscle, were sequentially dissected from each oyster, rapidly frozen in liquid nitrogen, and stored at −80°C for DNA/RNA extraction and glycogen content determination. The gonads were also collected for tissue section analysis (see below).

Glycogen content in *C. ariakensis* is closely related to gonadal development (Liu et al., 2021). For RNA-seq analysis, a total of 15 gonads from Dongying oysters were sampled, corresponding to five distinct gonadal development stages characterized by varying glycogen levels. These stages included the formative stage (group 1), proliferative stage (group 2), mature stage (group 3), proligerous stage (group 4), and inactive stage (group 5), with each represented by three biological replicates. Groups 1, 2, and 3 were further selected for ATAC-seq analysis. Detailed sample information is provided in Supplementary Table S2. For GWAS analysis, a total of 250 oysters were collected from the above five locations. Resequencing data for these samples can be found in our previous study (Wu et al., 2022), with additional information shown in Supplementary Table S3. Oysters exhibiting significant differences in glycogen content were collected for qPCR to verify the expression of key candidate genes (Supplementary Table S2).

Determination of gonadal development stages and glycogen content

Tissue sectioning was used to accurately determine the developmental stages of the oyster gonads. Specifically, gonadal tissue (1 cm^3) was dissected and fixed in Bouin's solution for 24 h, then transferred to a 70% alcohol solution for cleaning and preservation. The samples were subsequently dehydrated using ethanol solutions at different concentrations (80%, 90%, 95%, 100%), immersed in wax, cut into sections (4 μm thick), and stained with hematoxylin and eosin (H&E). A light microscope LEICA DM4000 (Leica Microsystems GmbH, Germany) was used to observe the sex and gonad stage of each sample.

Glycogen content in all tissues from 250 oysters was detected using the improved anthrone method (Chen et al., 2021b). In brief, fresh tissues were first lyophilized for 48 h, then ground into powder under freezing conditions. Each sample (20±1 mg) was transferred to a 10 mL test tube and digested for 45 min with 2 mL of 30% potassium hydroxide solution (w/w) under boiling conditions. Total volume was adjusted to 5 mL with distilled water, with centrifugation at 4 000 ×*g* for 8 min at room temperature subsequently performed to collect the supernatant. After being diluted 10 times, 100 μL of the supernatant was mixed with 200 μL of 0.2% (w/v) anthrone (Tokyo Chemical Industry, Japan) sulfuric acid solution on ice. The mixture was then heated in a boiling water bath for 10 min, and finally cooled with tap water. After 30 min at room temperature, 200 μL of the mixture was transferred to a microplate for the determination of optical density (OD) at 620 nm using a microplate reader.

A standard curve was established using glucose solutions with serial concentrations ranging from 0.01 mg/mL to 0.2 mg/mL. Glycogen content in each sample was then calculated using the formula $C=C_0\times(5\ mL/W)\times10\times0.9$, where C is the glycogen content of the sample (mg/g), C_0 is the glucose content in the test solution calculated according to the standard curve (mg/mL), W is the weight of the sample to be tested (g), 5 mL is the constant volume, 10 is the dilution factor, and 0.9 is the conversion factor between glucose and glycogen. The glycogen content in each oyster is shown in Supplementary Table S3.

RNA-seq analysis

Comparative transcriptome analysis was carried out to explore differential gene expression in oysters characterized by different gonadal development stages and glycogen levels. Total RNA was extracted using TRIzol reagent (Thermo Fisher Scientific, USA) following the manufacturer's instructions. Total RNA quality was assessed using an Agilent 2100 Bioanalyzer; subsequently, Oligo (dT) magnetic beads were employed for the enrichment of eukaryotic mRNA. The RNA integrity number of the 15 samples ranged from 8.0 to 9.2 (Supplementary Table S2). The obtained mRNA was fragmented and used as a template for first-strand cDNA synthesis. The cDNA was then used as a template for the generation of double-stranded cDNA, which was subsequently purified and recovered. Following sticky-end repair, an "A" base was added to the 3' end of the cDNA for subsequent adapter ligation. The cDNA was then used for library construction and short-read sequencing. A 2×150 bp paired-end library was successfully constructed and sequenced on the BGISEQ-500 platform (MGI Tech, China).

SOAPnuke v1.4.0 (Chen et al., 2018) was used to remove low-quality reads, joint contaminants, and reads with a high proportion of unknown base N from the initial dataset. After obtaining clean reads, HISAT v2.1.0 (Kim et al., 2019) was used to compare clean reads with the *C. ariakensis* genome (GenBank assembly accession: GCA_020458035.1) to predict

new transcripts, while Bowtie2 v2.2.5 (Langmead & Salzberg, 2012) was used to compare clean reads to the reference sequence for gene alignment. RSEM v1.3.3 (Li & Dewey, 2011) was then applied to calculate the expression levels of genes and transcripts to obtain fragments per kilobase of transcript per million mapped reads (FPKM) (Supplementary Table S4). Principal component analysis (PCA) and hierarchical clustering analysis were respectively conducted using the princomp and hclust functions (Müllner, 2013) in R. DESeq2 v1.30.1 (Love et al., 2014) was used to detect DEGs among different samples, screened based on fold-change>=2 and corrected *P*<=0.05 (DESeq2 parameters: fold-change>=2.00 and adjusted *P*<=0.05).

GSEA v4.0.0.0 (Subramanian et al., 2005) was used for gene set enrichment analysis (GSEA) of the MSigDB "c2.cp.kegg.7.5.1.symbols.gmt", "c2.cp.reactome.v7.5.1.symbols.gmt", and "c5.go.7.5.1.symbols.gmt" gene sets (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). To facilitate analysis, the *C. ariakensis* genes were named according to the HGNC database nomenclature (https://www.genenames.org/) (Tweedie et al., 2020).

ATAC-seq analysis

Gonadal cells were prepared according to previously published protocols (Xu et al., 2021), with subsequent transposition reaction, purification, library construction, and quantification conducted using established methods (Buenrostro et al., 2015). Nine samples were sequenced using 50 bp paired-end reads on the BGISEQ-500 platform (MGI Tech, China). The quality of the sequencing data was assessed using FastQC v0.11.9 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Contaminating primer and adapter sequences were removed, followed by the filtration of low-quality reads to obtain high-quality clean reads. Bowtie2 v2.2.5 (Langmead & Salzberg, 2012) was subsequently used to align the clean reads to the *C. ariakensis* genome (GenBank assembly accession: GCA_020458035.1). Read depth distributions within gene regions, as well as regions located 3 kb upstream and downstream in the genome, were then calculated.

Peak calling was performed using MACS2 v2.1.0 (https://pypi.python.org/pypi/MACS2/) to obtain peak summit, peak cumulative density, standardized *P*-values, and fold enrichment information. A region was considered a peak at a *P*-value<1×10^{-5}. Peaks were classified into different gene functional elements, including TTS, exon, intron, intergenic, and promoter elements, and the proportions of peaks in those regions were calculated according to peak location information. All peaks were associated with genes according to the positional relationship between peak and genes (*C. ariakensis* gene annotation files are available in the CNGBdb with the accession number CNA0022698.). Using DESeq2 v1.30.1 (Love et al., 2014), different peaks among samples were screened, and related genes were subjected to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis.

GWAS

GWAS was conducted to explore the single nucleotide polymorphisms (SNPs) associated with the regulation of glycogen metabolism. Data related to genome resequencing and SNP loci of 250 individuals were obtained from our previous study (Wu et al., 2022), and glycogen content was detected using the abovementioned method. To reduce false positive associations, SNP quality control was carried out using VCFtools with parameters "--max-missing 0.99 --maf 0.05 --minDP 4 --minGQ 10 --minQ 30 --min-meanDP 10". GWAS analysis was carried out using EMMAX (Kang et al., 2010). PLINK v1.9 was used for population stratification assessment and PCA (Purcell et al., 2007). Considering the factors that influence glycogen content, PCA values associated with varying sample collection dates and gonadal developmental stages in oysters were used as covariates to adjust for population structure and the kinship matrix. Considering tissue-specific regulation of glycogen metabolism, GWAS was conducted for each tissue separately. The genomic inflation value (λ) was calculated using the R package "qqman".

Dimensionality reduction via PCA was performed on the glycogen content data across six different tissues, with the first principal component (PC1) used as phenotypic data for GWAS analysis (Ning et al., 2019). Prior to that, missing data were completed using the missMDA package in R (Supplementary Table S5). To screen as many target candidate genes related to glycogen content as possible, the threshold for GWAS was set to *P*<1×10^{-5} (He et al., 2021; Li et al., 2019). Upon identifying significant SNPs, standard practice dictates their automatic mapping to proximal genes (Brodie et al., 2016).

Prediction of potential transcription factors of target genes was carried out using ChEA3 (https://maayanlab.cloud/chea3/) (Keenan et al., 2019). Gene functional enrichment analysis in the context of GWAS was carried out using Metascape (https://metascape.org/gp/index.html) (Zhou et al., 2019). Prior to this, the annotated gene names of *C. ariakensis* were standardized using HGNC database nomenclature (Tweedie et al., 2020).

Combination and visualization of multi-omics analysis results

Visualization of gene structure, RNA-seq, ATAC-seq, and SNPs was performed using IGV v2.14.0 (Thorvaldsdóttir et al., 2013). Subsequent color matching and integration were carried out using Adobe Illustrator (https://adobe.com/products/illustrator).

Motif enrichment analysis of peak sequences within the ATAC-seq data was performed using the SEA module of the MEME suite v5.4.1 (Bailey & Grant, 2021). Based on the enrichment results and annotations in the HOCOMOCO human (v11 FULL) motif database with default parameters, potential transcription factor-binding sites and transcription factor types in the peak sequences were identified.

Heatmaps depicting gene expression and box plots representing glycogen content were generated using Hiplot (Li et al., 2022). Linkage disequilibrium (LD) blocks for significant genes were plotted using LDBlockShow (Dong et al., 2021). The glycogen metabolism pathway diagram was created using Servier Medical Art image resources (https://smart.servier.com/).

Gene validation using qPCR

Five genes screened through joint RNA-seq and ATAC-seq analysis and four genes identified from GWAS were selected for qPCR validation. The oysters were divided into two groups (high and low glycogen content, respectively) for analysis. Three biological replicates for each group (Supplementary Table S2) and three technical replicates for each biological

replicate were utilized. In brief, cDNA was synthesized from 1 μg of total RNA using HiScript III RT Super Mix for qPCR (Vazyme, China), with qPCR then performed using an ABI StepOnePlus Real-Time PCR System with Tower (Applied Biosystems, USA) and ChamQ SYBR Color qPCR Master Mix (Vazyme, China) following the manufacturers' instructions. Melting curve analysis was performed at the end of the total number of cycles to confirm amplification specificity. The β-actin (GenBank Number: AF026063) gene served as the internal control and the primers used for qPCR analysis are listed in Supplementary Table S6.

RESULTS

Oysters with different glycogen content showed significant transcriptomic differences

A total of 15 individuals, categorized into five groups based on glycogen levels (group 1: 601 mg/g, group 2: 572 mg/g, group 3: 141.3 mg/g, group 4: 450.6 mg/g, group 5: 527.3 mg/g), were subjected to sequencing on the BGISEQ-500 platform, yielding an average of 6.11 Gb of data per sample. The average alignment rate of the sample genome was 73.58%, indicative of high-quality sequencing (Supplementary Table S7). Overall, 27 266 genes, including 25 327 known genes and 1 950 predicted new genes, were identified. Among these genes, 22 589 were present in all samples (Figure 2C; Supplementary Table S8).

The PCA results revealed that the 15 samples were relatively closely related based on PC1 (abscissa) but could be divided into two large categories based on PC2 (ordinate): groups 3 and 4 belonged to one category, while groups 1, 2, and 5 belonged to the other (Figure 2A). This division was consistent with the observed patterns of oyster gonadal development.

The DEG analysis results were consistent with the PCA findings (Figure 2B, left side of dotted line). While a small number of DEGs were observed between groups 1 and 2, 1 and 5, and 3 and 4, with counts ranging from 45 to 341, a large number of DEGs were identified between all other pairwise comparisons, ranging from 3 578 to 7 296 (Figure 2B, right side of dotted line). Notably, groups 3 and 5, which exhibited significant differences in glycogen content, had the highest number of DEGs (Figure 2C). Relative similarity in gene expression was found among groups 1, 2, and 5, and

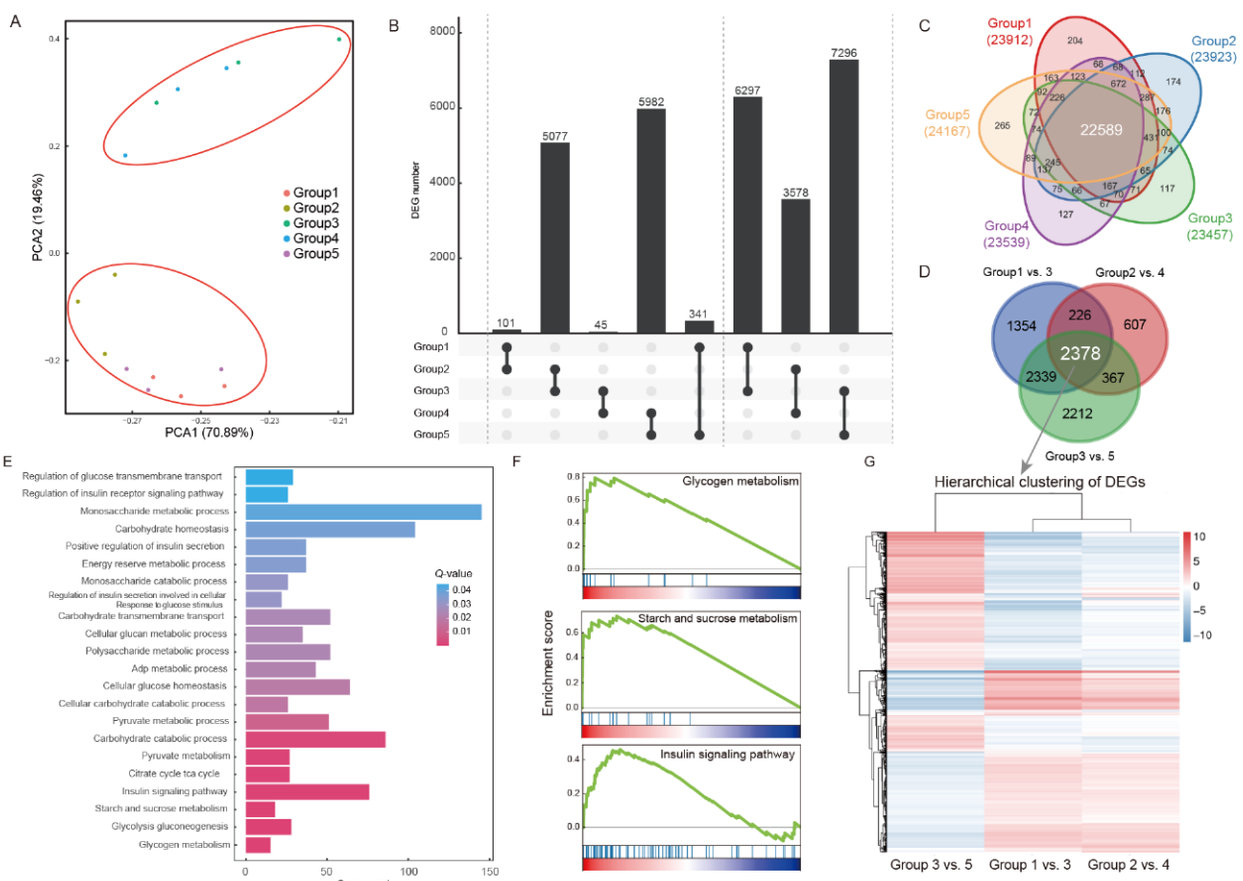


Figure 2 Transcriptomic analysis of *Crassostrea ariakensis* gonad tissue at different developmental stages

A: PCA of *C. ariakensis*. B: DEG number between different groups of *C. ariakensis*. C: Venn diagram of DEGs among five groups of *C. ariakensis*. D: Venn diagram of DEGs between groups 1 and 3, groups 2 and 4, and groups 3 and 5. E: Summary of pathways and entries closely related to glycogen metabolism in GSEA. F: Enrichment plot: Glycogen metabolism, Insulin signaling pathway, and Starch and sucrose metabolism. Top part of each graph shows enrichment score (ES) of each gene. In the middle part of each graph, each vertical line represents a gene in the gene set and its rank position. Bottom part of each graph shows gene-phenotype association matrix. Red indicates positive correlation with first phenotype (stage I or group 1) and up-regulation of gene expression in group 1, while blue indicates positive correlation with second phenotype (stage III or group 3) and up-regulation of gene expression in group 3. G: Hierarchical cluster analysis of DEGs between groups 1 and 3, groups 2 and 4, and groups 3 and 5.

between groups 3 and 4. Comparative analysis between groups 1 and 3, 2 and 4, and 3 and 5 yielded a total of 9 483 DEGs, 2 378 of which were shared across these group comparisons (Figure 2D). Moreover, these shared DEGs demonstrated analogous expression patterns in groups 1 and 3 and in groups 2 and 4, in contrast to the patterns observed in groups 3 and 5 (Figure 2G). This observation is in alignment with the distinct phases of oyster gonadal development, notably diverging in the mature and proligerous stages relative to other developmental stages (Figure 1A), implying that these DEGs are intricately linked to the process of gonadal development. As gonadal development has been shown to correlate with glycogen content in oysters (Chen et al., 2021a; Liu et al., 2020; Qin et al., 2021), the observed expression patterns also correlated with the glycogen phenotypes of these groups, whereby glycogen content was higher in groups 1, 2, and 5 than in groups 3 and 4. Consequently, the identified DEGs are likely associated with glycogen content in oysters.

GSEA identified pathways closely related to glycogen metabolism

GSEA was performed on the entire set of 27 266 identified genes to ascertain differential pathways. Given that groups 1 and 3 exhibited a considerable number of DEGs (6 297) (Figure 2C), along with marked differences in glycogen content (Figure 1C), these two groups were selected for GSEA as well as subsequent ATAC-seq analysis. Among the differential pathways identified, 22 were selected as potential pathways related to glycogen metabolism (Figure 2E; Supplementary Table S9).

Enrichment analysis based on the GO gene set revealed that all genes could be classified into 3 633 gene sets, 1 875 of which were up-regulated in group 1 (Supplementary Table S10) and 1 758 of which were up-regulated in group 3 (Supplementary Table S11). Based on GO enrichment analysis, two gene sets were enriched in glycogen metabolism-related terms, including "Energy Reserve Metabolic Process" and "Cellular Glucan Metabolic Process". Within the Reactome gene set, 15 genes were found to be involved in the oyster-specific "Glycogen Metabolism Pathway". Notably, within this pathway, 12 genes were up-regulated in group 1, characterized by high glycogen content (Supplementary Figure S1). Additional enrichment analyses identified other gene sets enriched in glycogen metabolism-related terms, including "Regulation of Glucose Transmembrane Transport" (Figure 2F; Supplementary Figure S2).

Upon conducting KEGG enrichment analysis, 142 gene sets were identified, with 92 gene sets up-regulated in group 1 and 50 in group 3 (Supplementary Tables S12–S13). Several of these gene sets were enriched in glycogen metabolism-related pathways, including "Starch and Sucrose Metabolism" and "Insulin Signaling Pathway" (Figure 2F). Notably, 18 genes were enriched in the "Starch and Sucrose Metabolism" pathway, 17 of which exhibited high expression in group 1, while only one showed high expression in group 3 (Supplementary Figure S1). These findings suggest the potential activation of the "Starch and Sucrose Metabolism" and "Insulin Signaling Pathway" in group 1.

The genes screened by RNA-seq analysis were taken as important candidate genes related to glycogen metabolism in further studies.

Analysis of ATAC-seq data demonstrated high-sequencing quality

ATAC-seq analysis performed on groups 1, 2, and 3 generated between 7.64 and 12.72 Gb of clean reads per sample, with an average size of 9.55 Gb (Supplementary Table S14). Comparison of these clean reads to the reference genome revealed read mapping rates ranging from 85.47% to 92.92%, indicative of high-quality sequencing (Supplementary Tables S15–S16).

The distribution of read depths across gene regions, as well as 3 kb upstream and downstream, is shown in Figure 3A and Supplementary Figure S3. Average sequencing depths in those regions were lower than those near the TSS and transcription end site (TES) of each gene. Peak proportions in the TTS and promoter regions were lower than those in the intron intergenic and exon regions (Figure 3B; Supplementary Tables S16, S17). These results align with the expected characteristics of the ATAC-seq data. The fragment size distribution of the acquired DNA spanned both the nucleosome and nucleosome-free regions (NFR). The curve patterns observed across the nine samples aligned with the characteristics of the ATAC-seq data (Figure 3C; Supplementary Figure S4), thereby reaffirming the high quality of the sequencing results.

As the peak regions could be either protein-binding domains or genomic regions regulated subject to histone modification, correlations between peak regions and genomic genes were analyzed. Among the three groups analyzed, 21 133 genes were identified as being associated with peaks, 18 612 of which were shared by all three groups (Figure 3D).

Genes associated with differential peaks were enriched in glycogen metabolism-related pathways

Due to the substantial number of DEGs found in groups 1, 2, and 3, ATAC-seq analysis was conducted on these three groups. Analysis identified many differential peaks corresponding to 7 215 genes. Groups 1 and 3 exhibited the largest number of differential peaks, associated with 6 841 genes. In contrast, groups 1 and 2 displayed the fewest differential peaks, associated with only 129 genes (Figure 3E). The differential peaks observed between groups 2 and 3 were related to 3 021 genes, 90.8% (2 743) of which were also observed between groups 1 and 3 (Figure 3D). In addition, a total of 2 600 genes overlapped with those identified in RNA-seq analysis (Figure 3D). Based on KEGG enrichment analysis, 21 genes between groups 1 and 3 were enriched in the "Starch and sucrose metabolism" pathway, suggesting direct involvement in oyster glycogen metabolism (Figure 3F; Supplementary Table S18).

Combined RNA-seq and ATAC-seq analysis

Based on GSEA and ATAC-seq data, five KEGG pathways, one Reactome pathway, and 16 GO terms potentially relevant to glycogen metabolism were identified (Figure 2E). Specifically, the "Glycogen metabolic process" and "Starch and sucrose metabolism" pathways are directly implicated to glycogen metabolism. These pathways included 67 genes covering 54 functional categories, highlighting their possible importance in glycogen metabolism (Figure 4A; Supplementary Table S19). Of these 67 genes, 49 were identified either through RNA-seq or ATAC-seq analyses, with 21 detected in both. Subsequent Metascape enrichment

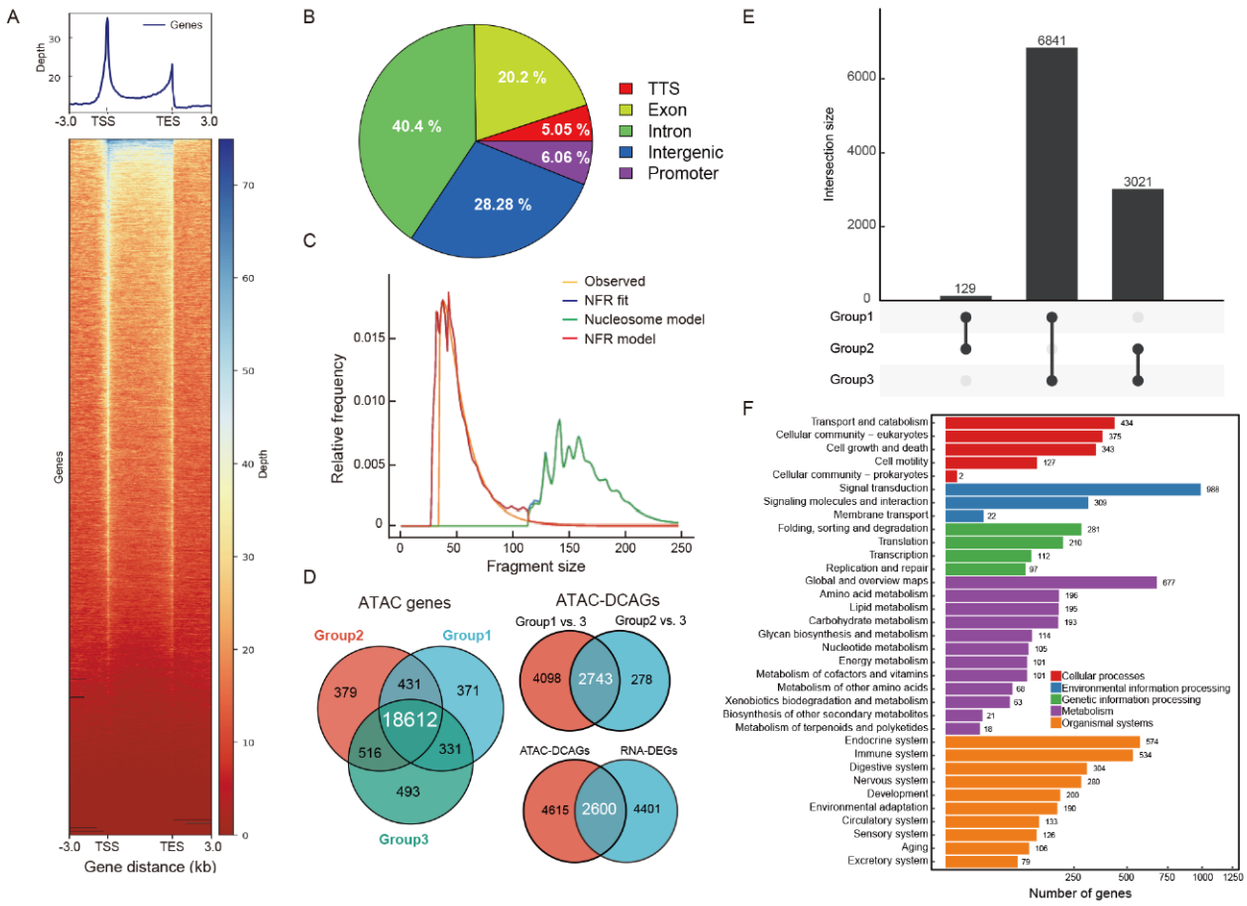


Figure 3 Comprehensive analysis of ATAC-seq data

A: Sequencing depth distribution map of promoter (up 3 kb) region of the sample gene. Abscissa represents sequence length and ordinate represents sequencing depth. B: Distribution map of peaks in gene functional elements. C: Results of nucleosome localization prediction. D: Venn diagram of genes related to peaks in different groups revealed by ATAC-seq (left), Venn diagram of genes associated with differential peaks between different groups revealed by ATAC-seq (right upper), and Venn diagram of DEGs revealed by ATAC-seq and RNA-seq (right lower). E: Number of genes related to differential peaks between different groups. F: KEGG enrichment analysis of DEGs between groups 1 and 3.

analysis of these 67 genes revealed the top six enriched pathways, including “Glycogen metabolism process”, “Starch and sucrose metabolism”, “Insulin signaling pathway”, “Glycogen synthesis and degradation”, “Galactose metabolic”, and “Glucagon signaling pathway” (Figure 4C; Supplementary Table S20). Furthermore, these terms and pathways were closely related (Figure 4D), suggesting that the identified genes may be involved in multiple metabolic pathways potentially relevant to oyster glycogen metabolism.

To verify the accuracy of our multi-omics analysis, five well-known genes closely related to glycogen metabolism were selected for further analysis, including glycogen debranching enzyme (*gde*), glycogen synthase (*gys*), glycogen phosphorylase (*pyg*), UTP-glucose-1-phosphate uridylyltransferase (*ugpa*), and phosphoglucomutase-1 (*pgm1*). RNA-seq analysis indicated significant disparities in gene expression among the various oyster groups, while distinct differences in chromatin accessibility were observed for *pyg*, *ugpa*, and *pgm1* across these groups. The qPCR results confirmed that the five genes were highly expressed in the high-glycogen group, corroborating the activation of glycogen metabolism-related pathways in this group as indicated by the multi-omics data (Figure 4B).

GWAS of glycogen phenotypes yielded 394 SNPs

Glycogen content (dry weight) was measured in six different tissue types across 250 individuals collected from five geographic locations (Supplementary Table S3). Results showed that glycogen content varied widely, from 5.6 mg/g to 128.9 mg/g in the adductor muscle, 21.0 mg/g to 464.2 mg/g in the gill, 28.31 mg/g to 706.9 mg/g in the mantle, 62.8 mg/g to 626.5 mg/g in the hepatopancreas, 24.72 mg/g to 674.36 mg/g in the gonads, and 103.62 mg/g to 749.64 mg/g in the labial palps (Figure 5A). Average glycogen content in the gonads (366.3 mg/g) and labial palps (399.0 mg/g) was significantly higher than that in the other four tissues (P<0.01). Compared to other tissues, glycogen content in the adductor muscle was relatively stable, consistent with previous studies (Liu et al., 2020; Qin et al., 2021). Glycogen content in the hepatopancreas, labial palps, and gonads showed a relatively normal distribution pattern, whereas that in the other tissues showed a skewed distribution (Figure 5B), potentially due to the small sampling size.

A total of 4.8 billion paired-end reads were generated from the 250 individuals, yielding an average sequencing depth of approximately 22.43× per oyster (Supplementary Table S3). Following reference genome alignment and SNP calling, a total of 9 653 216 SNPs were identified. Post-filtration, 1 918 741 high-quality SNPs were retained for subsequent GWAS analysis.

Manhattan and quantile-quantile (QQ) plots are shown in

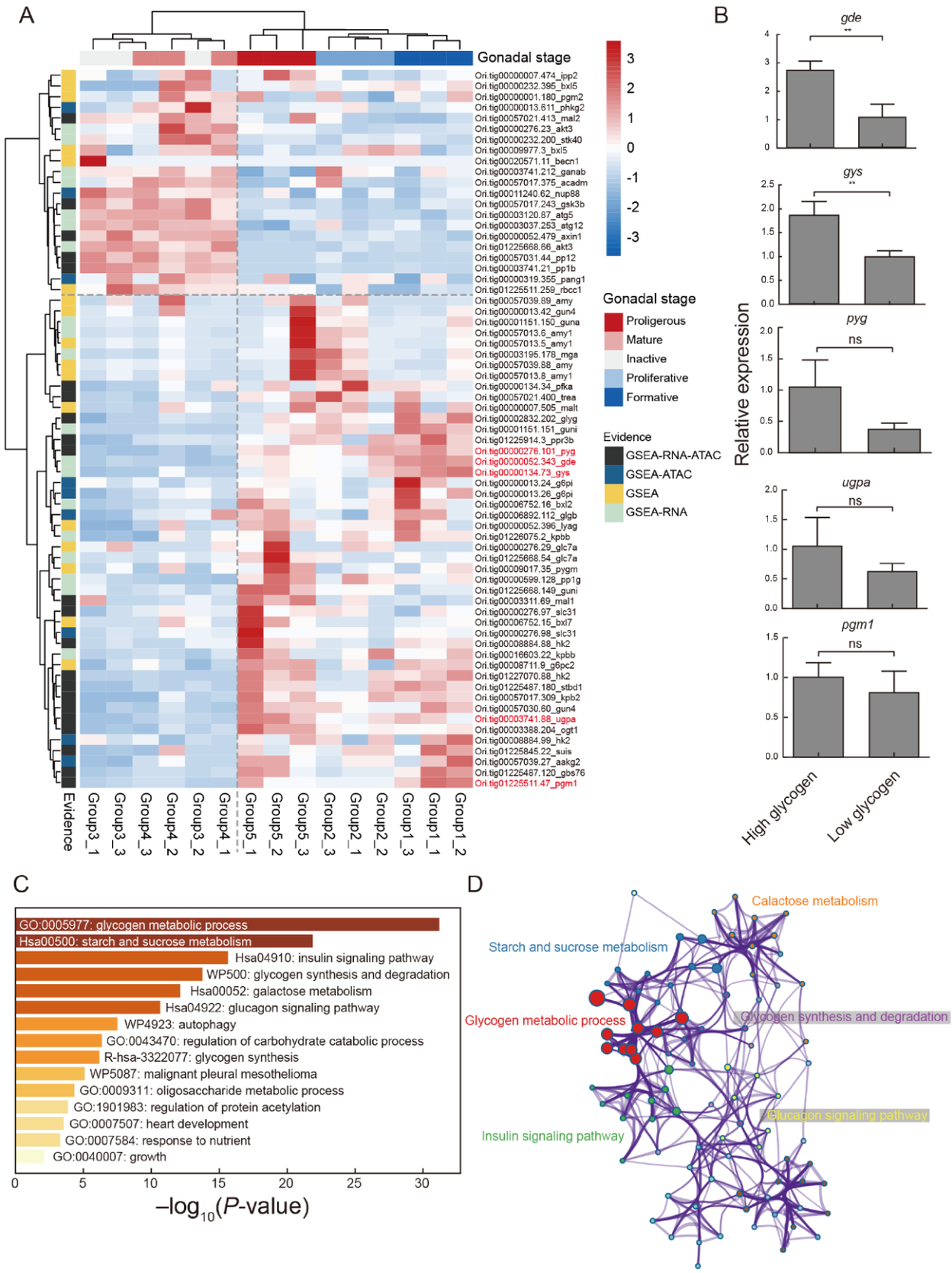


Figure 4 Summary analysis of genes related to glycogen metabolism

A: Heatmap of expression of genes related to glycogen metabolism. “Evidence” bar means that different analyses, including GSEA, ATAC-seq and RNA-seq, were used to reveal genes related to glycogen metabolism. Genes in red are validated by qPCR in B. B: qPCR verification of five genes, including *gde*, *gys*, *pyg*, *ugpa*, and *pgm1*. Vertical axis of bar graph represents relative expression level. ns: Not significant; **: P<0.01. C: Metascape enrichment analysis of genes in A. D: Correlation diagram among individual pathways in C.

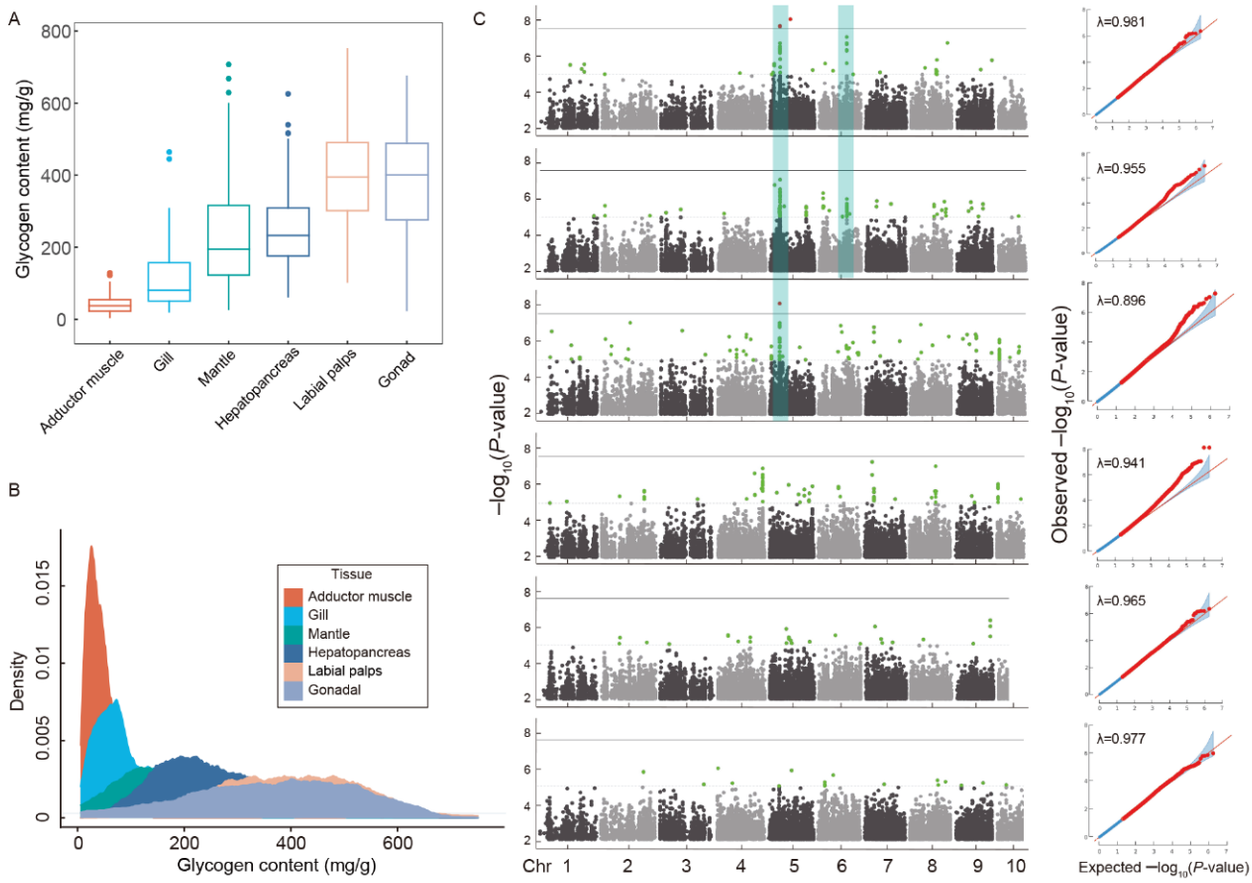


Figure 5 Glycogen contents in six different tissues of *Crassostrea ariakensis* and GWAS analysis

A: Box plot of glycogen content in six tissues from 250 oysters. B: Distribution of glycogen content in different tissues among individuals. C: GWAS analysis of glycogen content in six tissues in *C. ariakensis*. For each tissue, a Manhattan plot and corresponding QQ-plot (upper right) are shown. λ represents associated genomic inflation value. Dotted line corresponds to an ordinate value of 5 ($-\log_{10}P$) and solid line corresponds to an ordinate value of 7.58, calculated from $-\log_{10}$ (0.05/1 918 741).

Figure 5C. In total, 394 SNPs potentially associated with glycogen content were screened (Supplementary Table S21). Of these, 15 SNP loci, each located within a distinct gene, were detected in at least two tissues. Moreover, several significant loci (highlighted in blue boxes) located in two common chromosome regions (chromosomes 5 and 6) were detected in the gonad, mantle, and gill tissues. These genes warrant further study. The standard QQ plot confirmed the reliability of these SNPs across tissues.

GWAS after PCA dimensionality reduction of glycogen data obtained 132 significant SNPs

PCA of glycogen contents showed that PC1 accounted for 41.86% of the total phenotypic variance, indicating the feasibility of PC1 representing the measured glycogen content trait (Supplementary Table S22). Therefore, GWAS was further carried out using PC1 as a compound glycogen content trait to obtain more comprehensive SNPs (Figure 6A). Results yielded 132 SNP loci associated with the target trait, more than those yielded on a tissue basis (Supplementary Table S23). The *P*-values of nine loci met the Bonferroni test threshold criteria ($P<2.61\times10^{-8}$), as shown by the red dots in Figure 6A, and they were directly related to six genes. Among these six genes, growth factor independent 1 (*gfi1*) and Brachyury protein homolog A (T-box, *tbx5*), corresponding to the two SNPs, chr5_15152293 and chr5_15219936 (15 152 293rd and 15 219 936th base loci on chromosome 5, respectively), were identified as transcription factors ($-\log_{10}P>7.58$), which are involved in the regulation of many genes (Kazanjian et al., 2006; Smith, 1997) (Figure 6D). In addition, two SNPs associated with hexokinase type 2 (*hk2*) and glucose transporter type 1 (*glut1*) were also found on chromosomes 1 and 9, respectively (Figure 6A), which may be involved in glycogen metabolism (Bramer et al., 2017; Fajardo et al., 2021). SNPs falling within the *hk2* (chr9_37174195) gene were divided into three genotypes, containing CC, AC, and AA. Glycogen content corresponding to the three genotypes (represented by the first column of data after PCA dimensionality reduction) is shown in a violin diagram (Figure 7C), which suggested that the three genotypes were characterized by significantly different glycogen contents ($P<0.01$). These results further verified that this SNP was correlated with glycogen content.

The qPCR results shown in Figure 6C indicated that *hk2*, Ori.tig00057017.233, and *glut1* were differentially expressed between the high- and low-glycogen oysters. Notably, *hk2* was highly expressed in the high-glycogen oysters, while Ori.tig00057017.233 and *glut1* were highly expressed in the low-glycogen oysters. The *gfi1* gene was slightly more up-regulated in the low-glycogen group compared to the high-glycogen group. These expression patterns were consistent with the transcriptome data (Supplementary Figure S5).

Multi-omics analysis

Combining these 132 SNPs with the above 394 SNPs, we obtained 460 SNP loci belonging to 241 genes

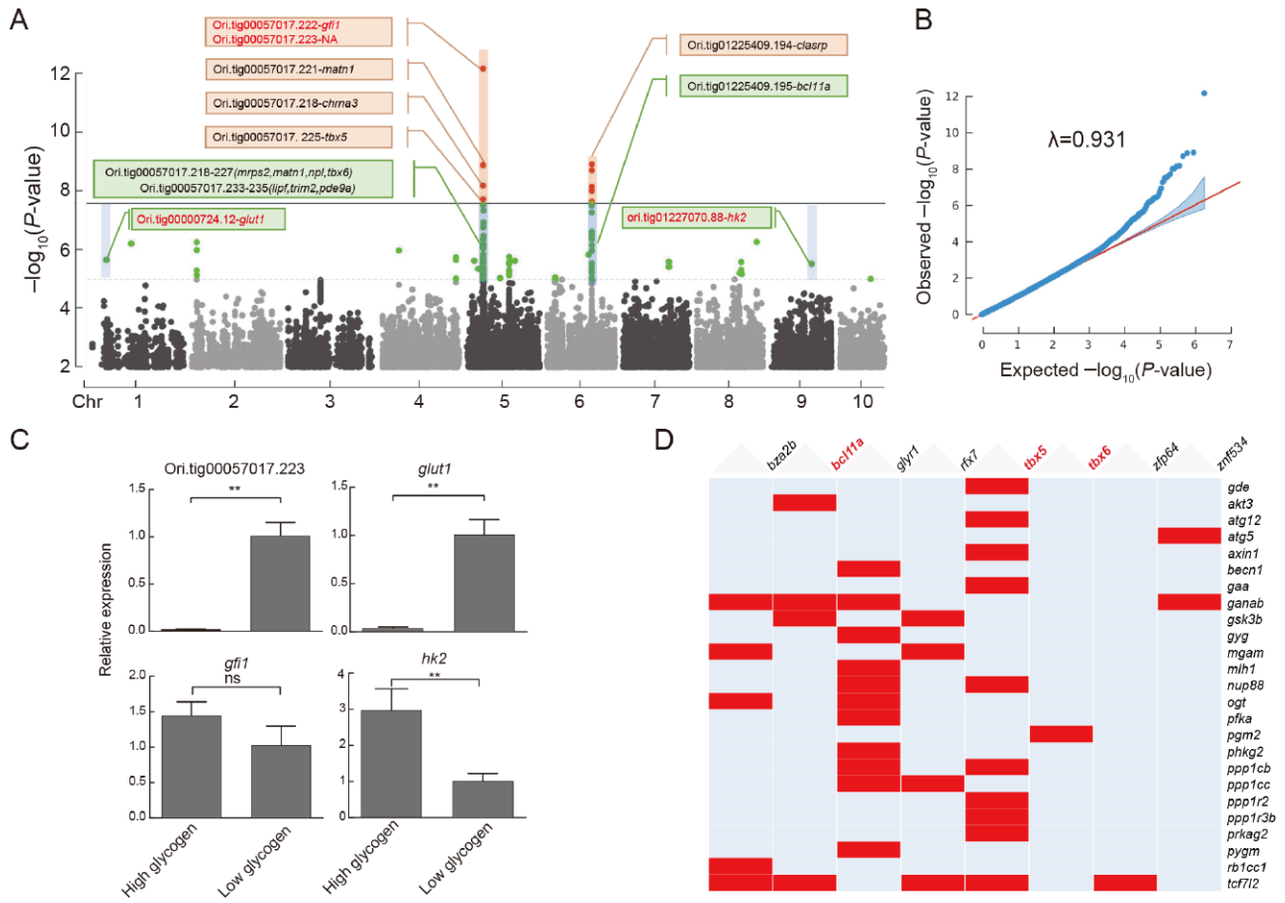


Figure 6 Analysis of genes associated with significant SNPs

A: Manhattan plot obtained by GWAS analysis of glycogen content data after dimensionality reduction using GEMMA software. Dotted line corresponds to an ordinate value of 5 ($-\log_{10}P$) and solid line corresponds to an ordinate value of 7.58, calculated from $-\log_{10}$ (0.05/1 918 741). B: QQ-plot corresponding to Manhattan plot, λ represents associated genomic inflation value. C: qPCR verification of four genes in oysters, including *gfi1*, Ori.tig00057017.233, *glut1*, and *hk2*. Vertical axis of bar graph represents relative expression level. ns: Not significant; **: P<0.01. D: Bar graph of enriched terms across input gene lists, colored by P-values. E: Glycogen metabolism-related genes corresponding to transcription factors obtained by GWAS.

(Supplementary Tables S21–S23). Further combining with RNA-seq results, 120 of the 241 genes showed significantly different expression among oysters with different glycogen contents (Supplementary Figure S6), while 63 of the 120 genes were characterized by different chromatin accessibility in oysters with different glycogen contents. We determined that these genes were probably related to oyster glycogen metabolism. Notably, 52 of the 241 genes were transcription factors, which could be classified into 26 transcription factor types (Supplementary Table S24). Through target gene prediction, eight transcription factors were involved in regulating glycogen metabolism, including BAF chromatin remodeling complex subunit Bcl11A (*bcl11a*), *tbx5*, and T-box transcription factor 6 (*tbx6*) (Figure 6D; Supplementary Table S25).

Through RNA-seq, ATAC-seq, and GWAS analysis, several important genes were identified, including *hk2*. The *hk2* gene plays an important role in glycogen metabolism, as it can convert glucose into glucose 6 phosphate, a precursor of glycogen synthesis (Blaha et al., 2022). We also found that higher chromatin accessibility of the *hk2* transcription initiation region was associated with higher gene expression (Figure 8A). The LD-block of all SNP loci in this gene showed strong linkage imbalance among them (Figure 7B). The SNP (chr9_37174195) was exactly localized to the peak (Chr9:37173892~37174226) of the *hk2* gene intron. This peak region showed higher chromatin accessibility in groups 1 and 2 than in group 3 (Figure 7A), suggesting this region may be closely related to *hk2* transcription. Further analysis predicted a motif binding to the forkhead box O1 (*foxo1*) transcription factor in this peak region (Supplementary Tables S26, S27). Therefore, this SNP site may affect the binding of transcription factor *foxo1* in this region, thereby affecting *hk2* gene expression.

We finally integrated all multi-omics results and selected 12 key genes to explore the regulation of glycogen metabolism in *C. ariakensis*, with differences in transcriptional expression and chromatin accessibility observed between the high- and low-glycogen groups (Figure 8). Specifically, seven genes (*glut1*, *hk2*, *pgm1*, *ugp*, *glyg*, *pyg*, and *foxo1*) exhibited significant differences in both mRNA expression and chromatin accessibility, while two genes (*gys* and *agl*) exhibited significant differences in mRNA expression. Excluding *glut1*, the remaining eight genes were highly expressed in the high-glycogen group. Additionally, insulin (*ins*) showed differences in chromatin accessibility but not in mRNA expression. Similar to the seven DCAGs described above, *ins* showed high chromatin accessibility in the high-glycogen group.

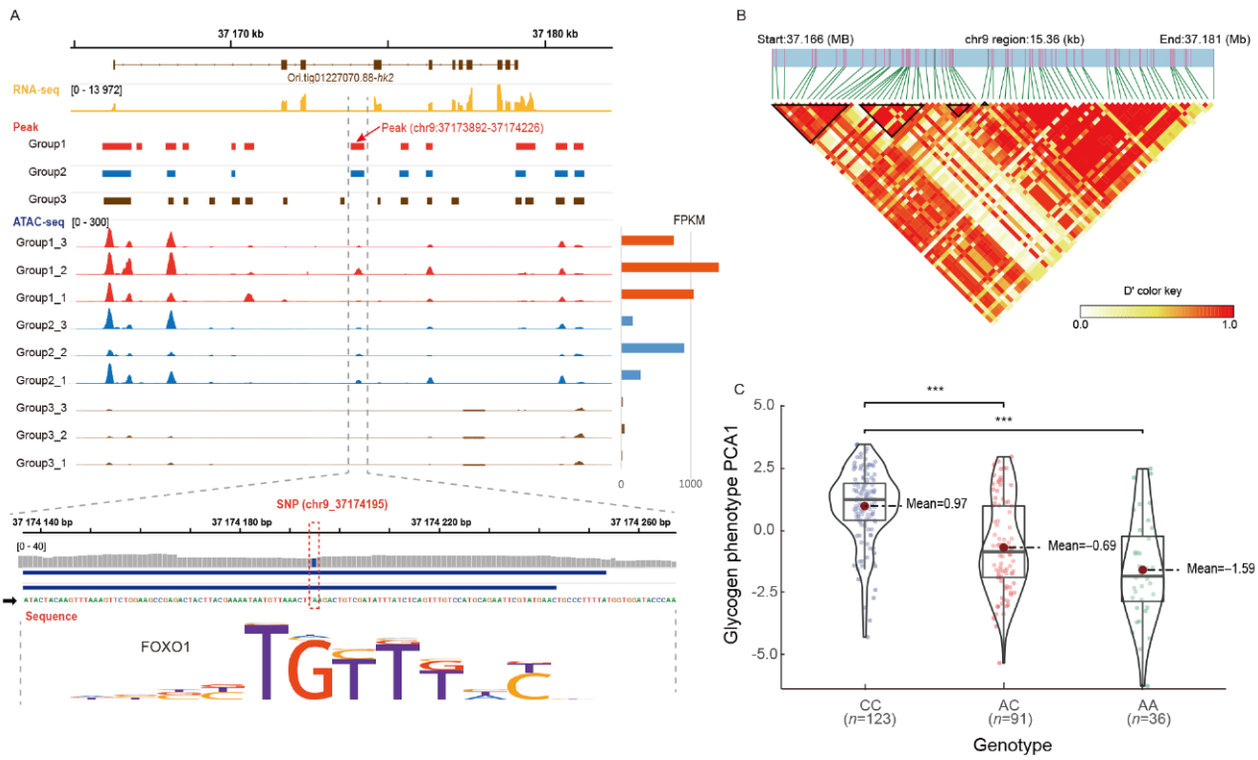


Figure 7 Multi-omics analysis of *hk2*

A: Visualization of ATAC-seq and RNA-seq data of *hk2*. Column chart is expression level of *hk2* in oyster gonads at different developmental stages, expressed in FPKM; bottom is predicted binding transcription factor and motif in this peak. B: LD-block diagram of *hk2* gene. C: Violin diagram showing glycogen content in oysters with different genotypes. Glycogen content was obtained after PCA dimensionality reduction. ***: P<0.001.

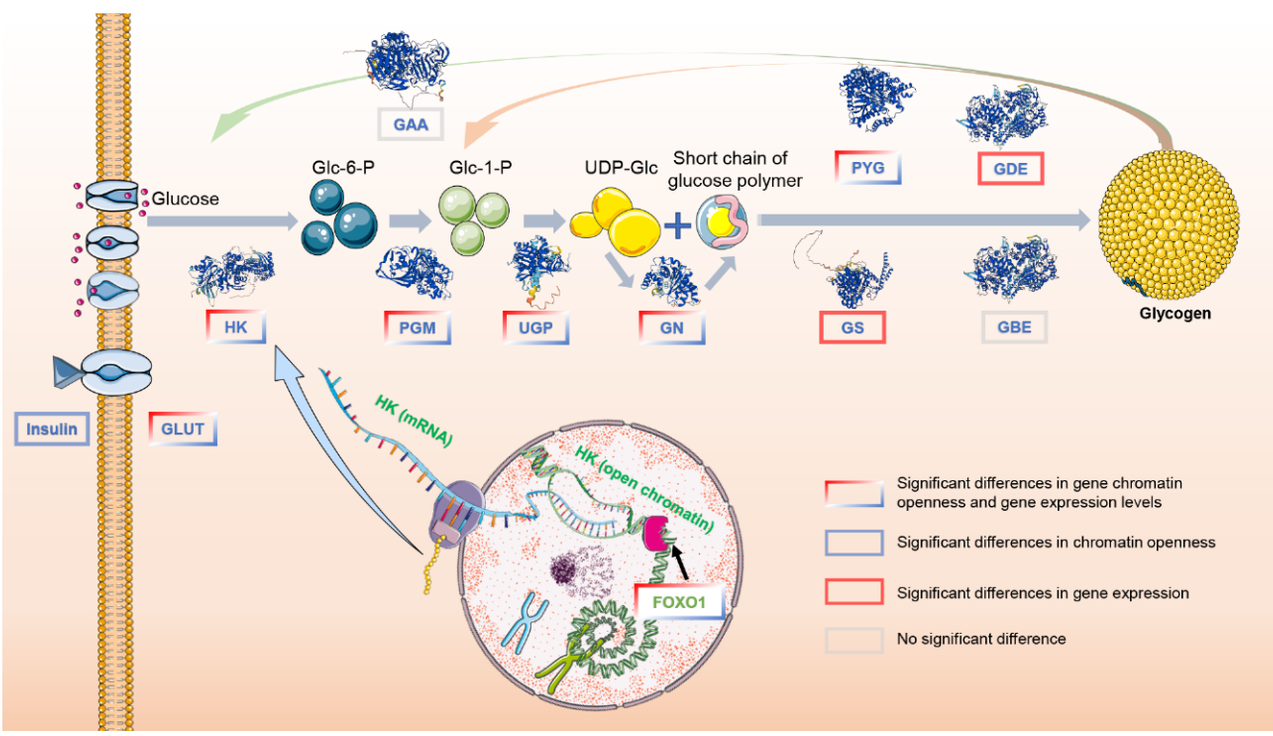


Figure 8 Diagram of glycogen metabolism pathway

Glycogen metabolism-related enzymes are shown in blue, and substance molecules involved in glycogen metabolism process are in black. Arrows indicate reaction direction. Three-dimensional structure of enzyme is quoted from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/). GLUT, glucose transporter; HK, hexokinase; PGM, phosphoglucomutase; UGP, UDP-glucose pyrophosphatase; GN, glycogen protein; GS, glycogen synthase; GBE, glycogen branching enzyme; PYG, glycogen phosphorylase; GDE, glycogen debranching enzyme; GAA, lysosomal alpha-glucosidase; FOXO1, forkhead box protein O1; Glc-6-P, Glucose-6-Phosphate; Glc-1-P, Glucose-1-Phosphate; UDP-Glc, Uridine Diphosphate Glucose.

DISCUSSION

GSEA revealed expression patterns of oyster glycogen metabolism-related gene sets

Various investigations have explored oyster glycogen metabolism based on transcriptomics (Chen et al., 2021a; Li et al., 2017; Wang et al., 2021a, 2021b). While certain DEGs and pathways related to glycogen metabolism have been identified, the comprehensive expression patterns of specific pathways have not yet been fully elucidated. In the current study, a multitude of gene sets were identified, including 22 directly related to glycogen metabolism. These important gene sets were notably up-regulated in oysters exhibiting high-glycogen content, a finding not reported in prior studies. Our approach facilitated the identification of diverse gene sets, thereby providing a more in-depth understanding of the regulatory network underlying oyster glycogen metabolism.

ATAC-seq revealed relationship between chromatin accessibility and expression of key genes

Despite a decade since its initial introduction (Buenrostro et al., 2013), ATAC-seq technology has only been applied in a single mollusk-based study to date (Xu et al., 2021). Notably, using both ATAC-seq and ChIP-seq, Xu et al. (2021) demonstrated that the upstream regulatory element of the oyster insulin gene (*cgILP*) can bind to the Pdx transcription factor, a finding congruent with the interaction between the vertebrate insulin-like gene and Pdx1 transcription factor. In vertebrate systems, the insulin gene plays a pivotal role in glycogen synthesis by inhibiting glycogen phosphorylase and concurrently activating glycogen synthase (Adeva-Andany et al., 2016). Consequently, we hypothesize that insulin-like genes in oysters may similarly be integral to glycogen synthesis.

In this study, ATAC-seq was employed for the first time to investigate changes in chromatin accessibility in *C. ariakensis* with varying levels of glycogen content. Contrary to traditional transcriptomic analysis, ATAC-seq not only yielded genome-wide data on chromatin accessibility but also enabled a more nuanced understanding of the epigenetic mechanisms regulating oyster glycogen metabolism. Among the identified chromatin open regions with significant accessibility differences, many genes involved in oyster glycogen metabolism were identified, including the insulin-like gene, *hk2*, *gys*, and *pyg*. Beyond the direct identification of genes related to differential chromatin accessibility, we also used the open chromatin sequences associated with 67 glycogen metabolism genes to predict 91 motifs and the transcription factors that may bind to them.

GWAS identified multiple genes and transcription factors associated with oyster glycogen content

In oysters, glycogen content exhibits marked variability among individuals and tissues (Meng et al., 2019; Shi et al., 2020), influenced by both genetic and environmental factors, such as season, nutrients, salinity, and temperature (Liu et al., 2020). To identify the principal factors affecting glycogen content, we utilized wild *C. ariakensis* oysters sourced from an extensive geographical range for GWAS analysis. Population structure, sampling time, and gonadal development stage were incorporated as the first, second, and third covariates, respectively, to eliminate potential environmental impact.

Limited GWAS analyses have been conducted on glycogen phenotypic traits in oysters. Using GWAS, Shi et al. (2020) identified the citrate lyase subunit beta-like (*clybl*) mitochondrial protein, which may enhance glycogen synthesis by promoting the accumulation of precursor substances. Similarly, Meng et al. (2019) mapped 17 SNP loci associated with 24 genes based on GWAS; notably, cytochrome P450 17A1 (*cyp17a1*) contained a non-synonymous SNP, displayed elevated expression in individuals with higher glycogen content, and potentially enhanced gluconeogenesis by transcriptionally regulating the expression of phosphoenolpyruvate carboxykinase (*pepck*) and glucose 6-phosphatase (*g6p*).

Previous GWAS studies on oyster glycogen content have focused on whole soft tissue analysis. Nevertheless, mammalian studies suggest that genes regulating glycogen metabolism may vary across different tissues, such as the brain, liver, and muscle (Adeva-Andany et al., 2016). Therefore, we conducted GWAS analysis of glycogen content in six individual tissues. The resultant Manhattan diagrams for the mantle, gill, and gonad were very similar, while that for the adductor muscle differed markedly. This divergence aligns with findings showing that genes regulating glycogen metabolism in mammalian muscle are distinct from those in other tissues (Adeva-Andany et al., 2016). However, further evidence is needed to ascertain whether glycogen metabolism regulation in oyster muscle differs from that in other tissues.

To better understand the relationship between overall glycogen content and SNPs, PCA was employed to reduce the dimensionality of the glycogen content data, with PC1 used as the phenotype for GWAS analysis. This approach yielded more significant SNP loci, the regions of which were consistent with our additional GWAS analysis, thereby validating the reliability of this method.

As an important enzyme in the glycogen synthesis pathway, hk2 mediates the initial step of glycolysis by catalyzing the phosphorylation of D-glucose to D-glucose 6-phosphate. Bramer et al. (2017) suggested that *hk2* serves as a key regulator of glycogen accumulation during equine diestrum and pregnancy. Glucose functions as an essential energy source for almost all life and is the only raw material for glycogen synthesis. *Glut1* is a member of the glucose transporter protein family, a diverse group of membrane proteins. Fajardo et al. (2021) showed that overexpression of *glut1* in mouse hearts can lead to increased glucose uptake and glycogen accumulation, highlighting the regulatory role of *glut1* in glycogen metabolism.

Multi-omics analysis revealed oyster glycogen regulatory network

Obtaining a complete regulatory network through single omics analysis is a challenge. While transcriptomic analysis can effectively identify DEGs related to glycogen metabolism, it is restricted to the RNA level (Wang et al., 2021b). Similarly, a single GWAS can only isolate SNPs associated with genes implicated in glycogen metabolism based on SNP locations. Therefore, for the first time, we combined RNA-seq, ATAC-seq, and GWAS analyses for a comprehensive exploration of glycogen metabolism regulation. We analyzed the open chromatin sequences in oyster gonads to elucidate the regulatory mechanisms of gene expression after transcription factor binding, thus establishing a regulatory network from DNA to RNA. The GWAS findings further clarified the important roles of candidate genes in glycogen metabolism at the DNA level. Notably, the multi-omics data could be cross-

validated, and further analysis on gene function and experimental phenotype provided a clear depiction of the biological process of epigenetic regulation-gene expression-gene function-phenotype.

In the present study, *hk2*, a critical gene implicated in glycogen metabolism, was identified through RNA-seq, ATAC-seq, and GWAS analysis. We hypothesized that the SNP site identified by GWAS was associated with the binding domain of the transcription factor *foxo1* (Sekine et al., 2007; Singh et al., 2017). Furthermore, co-expression between *foxo1* (Ori. tig00008711.64) and *hk2* (Ori. tig01227070.88) was observed in the transcriptomic data, suggesting that *foxo1* may serve as an enhancer regulating expression of the *hk2* gene.

Based on multi-omics data analysis, we identified several key genes related to glycogen metabolism. Notably, many key genes implicated in vertebrate glycogen metabolism were also identified in the oyster genome. These findings suggest that glycogen metabolism regulation in oysters and vertebrates may be evolutionarily conserved, although further evidence is needed to substantiate this claim. Future research should clarify the transcription factors regulating glycogen metabolism-related genes in oysters and explore the differences and similarities between mollusk and mammalian glycogen metabolic networks through multi-dimensional comparisons.

In conclusion, the present study employed multi-omics approaches to comprehensively investigate the relationship between chromatin accessibility and expression levels of key genes related to glycogen metabolism in *C. ariakensis*, and further revealed the underlying regulatory network for glycogen metabolism. This study provides valuable data and establishes a robust foundation for subsequent research.

CONCLUSIONS

Through GSEA and transcriptomic sequencing of oysters with different glycogen phenotypes, we observed the activation of specific pathways, including “Starch and sucrose metabolism”, “Insulin signaling pathway”, and “Glycogen metabolic process”, in oysters with high-glycogen content. Furthermore, genes in these pathways showed higher expression levels. Concurrently, ATAC-seq analysis revealed that genes with differential chromatin accessibility were also enriched in these pathways, thereby suggesting a direct relationship between chromatin accessibility and expression in genes. Furthermore, GWAS analysis indicated that transcription factors play a crucial role in the regulation of glycogen metabolism. In addition to SNPs directly related to transcription factors, several SNPs were found in key genes implicated in glycogen metabolism, such as *hk2*. Within open chromatin regions, we hypothesized that these SNP loci may influence the binding affinity of transcription factors, thereby modulating gene expression. Subsequently, a glycogen metabolism network in oysters was delineated based on these pathways. Remarkably, most of these key genes are also present in mammals, suggesting a high degree of evolutionary conservation in glycogen metabolism networks across diverse species.

DATA AVAILABILITY

The *C. ariakensis* RNA-seq data and ATAC-seq data generated in this study were submitted to the China National GeneBank Database (CNGBdb; https://db.cngb.org/; accession number CNP0003045), National Center for Biotechnology Information (NCBI SRA; https://www.ncbi.nlm.nih.gov/sra; accession number PRJNA999955), Genome Sequence Archive (GSA; https://ngdc.cncb.ac.cn/gsa; accession number PRJCA018671), and Science Data Bank (SDB; https://www.scidb.cn/en; Data DOI: 10.57760/sciencedb.09809). The *C. ariakensis* assembled genome and whole-genome resequencing datasets that support the findings of this study are available at CNGBdb (doi.org/10.1111/1755-0998.13556, accession number CNP0001149).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS’ CONTRIBUTIONS

B.W., X.C., and G.Y.F. designed the research. X.C., J.H., D.Y.X., and H.B.G. analyzed the data. B.W. and X.C. wrote the manuscript. Y.W. and Z.Y.W. performed histological sectioning and validation experiments. C.W.S., L.Q.Z., X.J.S., T.Y., X.M.W., and Y.X.Z. revised the manuscript. Z.H.L. guided revision of the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank the China National GeneBank, National Center for Biotechnology Information, Genome Sequence Archive, and Science Data Bank for technical support in data service. Parts of Figure 8 were drawn using images from Servier Medical Art, licensed under a Creative Commons Attribution 3.0 Unported License (https://creativecommons.org/licenses/by/3.0/).

REFERENCES

Adeva-Andany MM, González-Lucán M, Donapetry-García C, et al. 2016. Glycogen metabolism in humans. *BBA Clinical*, **5**: 85−100.

Bailey TL, Grant CE. 2021. SEA: simple enrichment analysis of motifs. *bioRxiv*, doi:https://doi.org/10.1101/2021.08.23.457422.

Blaha CS, Ramakrishnan G, Jeon SM, et al. 2022. A non-catalytic scaffolding activity of hexokinase 2 contributes to EMT and metastasis. *Nature Communications*, **13**(1): 899.

Bramer SA, Macedo A, Klein C. 2017. Hexokinase 2 drives glycogen accumulation in equine endometrium at day 12 of diestrus and pregnancy. *Reproductive Biology and Endocrinology*, **15**(1): 4.

Brodie A, Azaria JR, Ofran Y. 2016. How far from the SNP may the causative genes be. *Nucleic Acids Research*, **44**(13): 6046−6054.

Buenrostro JD, Giresi PG, Zaba LC, et al. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, **10**(12): 1213−1218.

Buenrostro JD, Wu BJ, Chang HY, et al. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, **109**: 21.29.1−21.29.9.

Chen C, Yu H, Li Q. 2021a. Integrated proteomic and transcriptomic analysis of gonads reveal disruption of germ cell proliferation and division, and energy storage in glycogen in sterile triploid pacific oysters (*Crassostrea gigas*). *Cells*, **10**(10): 2668.

Chen X, Wu B, Wang Y, et al. 2021b. Establishment and optimization of micro-reaction system for determination of oyster glycogen content. *South China Fisheries Science*, **17**(4): 126−132. (in Chinese)

Chen YX, Chen YS, Shi CM, et al. 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience*, **7**(1): 1−6.

Dong SS, He WM, Ji JJ, et al. 2021. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on

variant call format files. *Briefings in Bioinformatics*, **22**(4): bbaa227.
Fajardo VM, Feng I, Chen BY, et al. 2021. GLUT1 overexpression enhances glucose metabolism and promotes neonatal heart regeneration. *Scientific Reports*, **11**(1): 8669.
He X, Li CY, Qi HG, et al. 2021. A genome-wide association study to identify the genes associated with shell growth and shape-related traits in Crassostrea gigas. *Aquaculture*, **543**: 736926.
Kang HM, Sul JH, Service SK, et al. 2010. Variance component model to account for sample structure in genome-wide association studies. *Nature Genetics*, **42**(4): 348−354.
Kazanjian A, Gross EA, Grimes HL. 2006. The growth factor independence-1 transcription factor: New functions and new insights. *Critical Reviews in Oncology/Hematology*, **59**(2): 85−97.
Keenan AB, Torre D, Lachmann A, et al. 2019. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. *Nucleic Acids Research*, **47**(W1): W212−W224.
Kim D, Paggi JM, Park C, et al. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, **37**(8): 907−915.
Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357−359.
Li A, Dai H, Guo XM, et al. 2021. Genome of the estuarine oyster provides insights into climate impact and adaptive plasticity. *Communications Biology*, **4**(1): 1287.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**(1): 323.
Li BS, Song K, Meng J, et al. 2017. Integrated application of transcriptomics and metabolomics provides insights into glycogen content regulation in the Pacific oyster Crassostrea gigas. *BMC Genomics*, **18**(1): 713.
Li JF, Miao BB, Wang SX, et al. 2022. Hiplot: a comprehensive and easy-to-use web service for boosting publication-ready biomedical data visualization. *Briefings in Bioinformatics*, **23**(4): bbac261.
Li Y, Cao K, Zhu GR, et al. 2019. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biology*, **20**(1): 36.
Liu QH, Tang JW, Wen PB, et al. 2021. From prokaryotes to eukaryotes: insights into the molecular structure of glycogen particles. *Frontiers in Molecular Biosciences*, **8**: 673315.
Liu S, Li L, Wang W, et al. 2020. Characterization, fluctuation and tissue differences in nutrient content in the Pacific oyster (*Crassostrea gigas*) in Qingdao, northern China. *Aquaculture Research*, **51**(4): 1353−1364.
Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**(12): 550.
Meng J, Song K, Li CY, et al. 2019. Genome-wide association analysis of nutrient traits in the oyster *Crassostrea gigas*: genetic effect and interaction network. *BMC Genomics*, **20**(1): 625.
Ministry of Agriculture and Rural Affairs of the People's Republic of China, National Fisheries Technology Extension Center, China Society of Fisheries. 2022. China Fishery Statistical Yearbook. Beijing: China Agriculture Press. (in Chinese)
Müllner D. 2013. fastcluster: fast hierarchical, agglomerative clustering routines for R and python. *Journal of Statistical Software*, **53**(9): 1−18.
Ning XH, Li X, Wang J, et al. 2019. Genome-wide association study reveals *E2F3* as the candidate gene for scallop growth. *Aquaculture*, **511**: 734216.
Ojea J, Pazos AJ, Martínez D, et al. 2004. Seasonal variation in weight and biochemical composition of the tissues of Ruditapes decussatus in relation to the gametogenic cycle. *Aquaculture*, **238**(1-4): 451−468.
Peng DM, Zhang SC, Zhang HZ, et al. 2021. The oyster fishery in China: trend, concerns and solutions. *Marine Policy*, **129**: 104524.
Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**(3): 559−575.
Qin YP, Li XY, Li J, et al. 2021. Seasonal variations in biochemical composition and nutritional quality of Crassostrea hongkongensis, in relation to the gametogenic cycle. *Food Chemistry*, **356**: 129736.
Sekine K, Chen YR, Kojima N, et al. 2007. Foxo1 links insulin signaling to C/EBPα and regulates gluconeogenesis during liver development. *The EMBO Journal*, **26**(15): 3607−3615.
Shi RH, Li CY, Qi HG, et al. 2020. Construction of a high-resolution genetic map of Crassostrea gigas: QTL mapping and GWAS applications revealed candidate genes controlling nutritional traits. *Aquaculture*, **527**: 735427.
Singh P, Han EH, Endrizzi JA, et al. 2017. Crystal structures reveal a new and novel FoxO1 binding site within the human glucose-6-phosphatase catalytic subunit 1 gene promoter. *Journal of Structural Biology*, **198**(1): 54−64.
Smith J. 1997. Brachyury and the T-box genes. *Current Opinion in Genetics & Development*, **7**(4): 474−480.
Smolders R, Bervoets L, De Coen W, et al. 2004. Cellular energy allocation in zebra mussels exposed along a pollution gradient: linking cellular effects to higher levels of biological organization. *Environmental Pollution*, **129**(1): 99−112.
Subramanian A, Tamayo P, Mootha VK, et al. 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, **102**(43): 15545−15550.
Swain P P, Sahoo L, Kumar R, et al. 2021. Applications of next-generation sequencing in aquaculture and fisheries. *In*: Pandey PK, Parhi J. Advances in Fisheries Biotechnology. Singapore: Springer, 41–64.
Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, **14**(2): 178−192.
Tweedie S, Braschi B, Gray K, et al. 2020. Genenames. org: the HGNC and VGNC resources in 2021. *Nucleic Acids Research*, **49**(D1): D939−D946.
Wang CG, Li A, Wang W, et al. 2021a. Integrated application of transcriptomics and metabolomics reveals the energy allocation-mediated mechanisms of growth-defense trade-offs in *Crassostrea gigas* and *Crassostrea angulata*. *Frontiers in Marine Science*, **8**: 744626.
Wang X, Wang WJ, Li Z, et al. 2021b. Comprehensive analysis of differentially expressed ncRNA, mRNA, and their ceRNA networks in the regulation of glycogen content in the Pacific oyster. *Crassostrea gigas. Aquaculture*, **531**: 735895.
Wu B, Chen X, Yu MJ, et al. 2022. Chromosome-level genome and population genomic analysis provide insights into the evolution and environmental adaptation of Jinjiang oyster Crassostrea ariakensis. *Molecular Ecology Resources*, **22**(4): 1529−1544.
Xu F, Marlétaz F, Gavriouchkina D, et al. 2021. Evidence from oyster suggests an ancient role for Pdx in regulating insulin gene expression in animals. *Nature Communications*, **12**(1): 3117.
Yan F, Powell DR, Curtis DJ, et al. 2020. From reads to insight: a hitchhiker' s guide to ATAC-seq data analysis. *Genome Biology*, **21**(1): 22.
Zhang GF, Li L, Que HY. 2020. An evolution of oyster mariculture industry in China: new knowledge, variety and product. *Oceanologia et Limnologia Sinica*, **51**(4): 740−749. (in Chinese)
Zhang HF, Ma JW, Tang K, et al. 2021. Beyond energy storage: roles of glycogen metabolism in health and disease. *The FEBS Journal*, **288**(12): 3772−3783.
Zhou YY, Zhou B, Pache L, et al. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, **10**(1): 1523.

Wei et al. *Zool. Res.* 2024, 45(1): 215−225
https://doi.org/10.24272/j.issn.2095-8137.2023.039

ZR | Zoological Research

Article Open Access

Molecular phylogenetic relationships based on mitochondrial genomes of novel deep-sea corals (Octocorallia: Alcyonacea): Insights into slow evolution and adaptation to extreme deep-sea environments

Zhan-Fei Wei[1,2,#], Kai-Wen Ta[3,#], Nan-Nan Zhang[2,#], Shan-Shan Liu[2], Liang Meng[2], Kai-Qiang Liu[1], Chong-Yang Cai[2], Xiao-Tong Peng[3,*], Chang-Wei Shao[1,*]

[1] *National Key Laboratory of Mariculture Biobreeding and Sustainable Goods*, *Yellow Sea Fisheries Research Institute*, *Chinese Academy of Fishery Sciences*, *Qingdao*, *Shandong* 266071, *China*

[2] *BGI-Qingdao*, *BGI-Shenzhen*, *Qingdao*, *Shandong* 266555, *China*

[3] *Institute of Deep-Sea Science and Engineering*, *Chinese Academy of Sciences*, *Sanya*, *Hainan* 572000, *China*

ABSTRACT

A total of 10 specimens of Alcyonacea corals were collected at depths ranging from 905 m to 1 633 m by the manned submersible Shenhai Yongshi during two cruises in the South China Sea (SCS). Based on mitochondrial genomic characteristics, morphological examination, and sclerite scanning electron microscopy, the samples were categorized into four suborders (Calcaxonia, Holaxonia, Scleraxonia, and Stolonifera), and identified as 9 possible new cold-water coral species. Assessments of GC-skew dissimilarity, phylogenetic distance, and average nucleotide identity (ANI) revealed a slow evolutionary rate for the octocoral mitochondrial sequences. The nonsynonymous (*Ka*) to synonymous (*Ks*) substitution ratio (*Ka*/*Ks*) suggested that the 14 protein-coding genes (PCGs) were under purifying selection, likely due to specific deep-sea environmental pressures. Correlation analysis of the median *Ka*/*Ks* values of five gene families and environmental factors indicated that the genes encoding cytochrome *b* (cyt *b*) and DNA mismatch repair protein (*mutS*) may be influenced by environmental factors in the context of deep-sea species formation. This study highlights the slow evolutionary pace and adaptive mechanisms of deep-sea corals.

Keywords: Mitochondrial genome; Alcyonacea; *Ka*/*Ks* evolution; Environmental factors

INTRODUCTION

The Octocorallia subclass of anthozoans, commonly referred to as sea fans, sea whips, sea pens, or soft corals, provides critical habitat for many aquatic organisms and microorganisms. These corals are widely distributed across a range of water environments, from shallow tropical regions to the deep sea (Baillon et al., 2012; Fryer et al., 2020; Roberts et al., 2006; Van De Water et al., 2018). The Octocorallia subclass comprises three orders, with more than 3 500 species in 378 genera and 55 families (McFadden et al., 2010). Within the Octocorallia subclass, the Alcyonacea order has undergone significant taxonomic revision since the late 20th century. This order encompasses all soft corals and provides habitat for biota in deep-sea ecosystems (McFadden et al., 2010; Zhou et al., 2021). Corals of the Alcyonacea order are well-adapted to the harsh conditions present in deep oceans, such as darkness, low temperatures, and high hydrostatic pressure. These cold-water corals act as ecosystem engineers in extreme deep-sea environments (Hebbeln et al., 2020; Roberts et al., 2006) and are noted for their long lifespans and slow growth rates (Sun et al., 2010; Thierens et al., 2013). Cold-water coral forests consisting of bamboo corals and gorgonian corals from the Isididae and Gorgoniidae families, respectively, have been documented in the South China Sea (SCS) at depths ranging from 1 200 to 1 380 m (Li & Wang, 2019).

Mitochondria are essential for cellular energy generation and play critical roles in the differentiation, apoptosis, growth,

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 27 March 2023; Accepted: 11 September 2023; Online: 12 September 2023

Foundation items: This study was supported by the Marine S&T Fund of Shandong Province for Pilot National Laboratory for Marine Science and Technology (Qingdao) (No. 2022QNLM030004), Hainan Science and Technology Department (ZDKJ2019011), Open Project Fund of Key Laboratory of Sustainable Development of Polar Fisheries, Ministry of Agriculture and Rural Affairs of PRC (2022OPF02), State Key R&D Project (2021YFF0502500), and Qingdao Postdoctoral Applied Research Project (JZ2223j06100)

[#]Authors contributed equally to this work

*Corresponding authors, E-mail: xtpeng@idsse.ac.cn; shaocw@ysfri.ac.cn

and division of eukaryotic cells (Kroemer & Reed, 2000; Taanman, 1999). Mitochondrial genomes, characterized by their small size, maternal inheritance, compact gene arrangement, high conservation, and simplified structure, serve as powerful molecular markers in studies exploring phylogenetic relationships and biomolecular evolution (Li et al., 2021; McFadden et al., 2010; Taanman, 1999; Yan et al., 2019). Within the Cnidaria phylum, which includes the anthozoans, the nucleotide sequences of mitochondrial genes exhibit relatively low mutation rates and high stability (Hellberg, 2006; Shearer et al., 2002). Mitochondrial genomes are typically circular and consist of protein-coding genes (PCGs), ribosomal RNA (rRNA) genes, and transfer RNA (tRNA) genes (Boore, 1999). Interestingly, the DNA mismatch repair protein (*mutS*) homolog, which is implicated in DNA mismatch repair and intramolecular recombination, is present in the mitochondria of all octocorals, with studies suggesting it may be horizontally transferred from the epsilonproteobacteria or DNA viruses (Bilewitch & Degnan, 2011; Brugler & France, 2008; Ogata et al., 2011). The presence of *mutS* in octocoral mitochondrial genomes may contribute to the slower nucleotide substitution rates and participate in gene rearrangement events during evolution (Brockman & McFadden, 2012; Muthye et al., 2022). Species-specific conserved DNA fragments, such as *mutS* and cytochrome c oxidase subunit 1 (*cox1*), are often employed for the identification of coral taxon (Xu et al., 2021).

As key organelles involved in energy and redox homeostasis, mitochondria are sensitive to environmental changes in animals (Sokolova, 2018). Variations in mitochondrial nucleotide sequences can improve the fitness and metabolism of organisms within extreme environments (Carapelli et al., 2019). For example, certain mitochondrial genes in alvinocaridid shrimps have undergone positive selection to facilitate adaptation to the harsh conditions of hydrothermal vents (Wang et al., 2017). Similarly, mitochondrial genes under positive selection pressure may contribute to the survival of anemones in the deep-sea environment (Zhang et al., 2017). Natural selection not only acts on phenotypic traits but also reduces mitochondrial dysfunction by inhibiting deleterious mutations in mitochondrial genes (Shtolz & Mishmar, 2019). Environmental factors, acting as selective forces, can also induce gene mutations that influence evolutionary processes, potentially manifesting as variations in mitochondrial genes.

In the present study, morphological and mitochondrial genome analyses led to the identification of 9 novel cold-water coral species belonging to the Alcyonacea order. The characteristics of the mitochondrial genomes were analyzed using various bioinformatics approaches. Furthermore, based on a comparative assessment of mitochondrial genomes, variation and conservation across the Octocorallia subclasses were evaluated. The contributions of environmental factors to the evolution of mitochondrial genes were also assessed, providing potential evidence for the adaptation of cold-water corals to extreme deep-sea environments. Our study delineated the characteristics of the mitochondrial genomes of deep-sea corals and the impact of environmental factors on the evolution of these genes.

MATERIALS AND METHODS

Ethics statement

Ethics approval for all deep-sea coral experiments was granted by the Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences following the recommendations of the Regulations for the Administration of Affairs Concerning Experimental Animals of China (Qingdao, China) (Approval No.: YSFRI-2022016).

Sampling, DNA extraction, and sequencing

A total of 10 deep-sea coral colonies were collected from the SCS during two cruises between August and November 2020 (Figure 1; Table 1). The samples were captured by a manned submersible Shenhai Yongshi equipped with mechanical arms, then stored in a bio-box. Once the manned submersible returned, the coral samples were immediately rinsed twice with deionized water, then stored at −80 °C in alcohol for subsequent experiments in the shipboard laboratory. A conductivity, temperature, and depth (CTD) sensor (Sea-Bird, USA) on the manned submersible was employed to monitor variations in deep-sea environmental factors during sampling, including temperature, salinity, and density. Sampling site information (i.e., longitude and latitude) was also collected by the manned submersible.

The coral samples were soaked in liquid nitrogen, crushed using sterilized mortars, and then subjected to total DNA extraction using phenol-chloroform with some modification. CTAB lysis buffer (3 mL) containing 5 μL of β-mercaptoethanol and 20 μL of proteinase (20 mg/mL) was added to the crushed sample, followed by incubation at 55 °C for 1 h to disrupt the cells. An equal volume of phenol/chloroform/isopentyl alcohol (25:24:1) was added and fully mixed. The DNA in the supernatant was extracted using an equal volume of chloroform/isopentyl alcohol (24:1). DNA was precipitated using a 0.7×volume of isopropyl alcohol and diluted with 50 μL of deionized water.

The RNA fragments were removed with a final concentration of 1 mg/μL RNaseA (Takara, Dalian, China) at 37°C for 10 min. The quality and quantity of the extracted DNA were checked using gel electrophoresis and a Qubit 2.0 Fluorometer (Life, USA). Approximately 1 μg of DNA was fragmented into 500 bp fragments by ultrasonication to prepare paired-end short-read libraries. The DNA libraries were constructed using a MGIEasy Universal DNA Library Prep Kit (MGI, China) and sequenced on the MGISEQ-2000 platform with 100 bp paired-end reads to obtain approximately 6 Gb per sample.

Morphological identification

To identify the taxa of the 10 deep-sea coral samples, their general morphology and anatomy were studied according to previous studies, with some modification (Li et al., 2020). Parts of individuals, including tentacles, were incubated in sodium hypochlorite for 15 min, then centrifuged at 3 000 r/min for 2 min at room temperature. The precipitates were subsequently washed with distilled water and 70% ethanol to collect the sclerites, which were air-dried, mounted on carbon double adhesive tape, and coated for scanning electron microscopy (SEM, FEI Apreo, USA) with a voltage of 2 kV and working distance of 10 mm.

Assembly and annotation of mitochondrial genomes

The raw data were confirmed by Fastqc (v.0.11.9) with default settings (Andrews, 2014). Low-quality reads and adaptors were removed using Fastp (v.0.22.0) ("-f 5 -F 5 -w 24 -c -q 20 -g -W 5 -3 -l 50") (Chen et al., 2018). MEGAHIT (v.1.1.2) was used to assemble the quality-filtered reads to contigs with

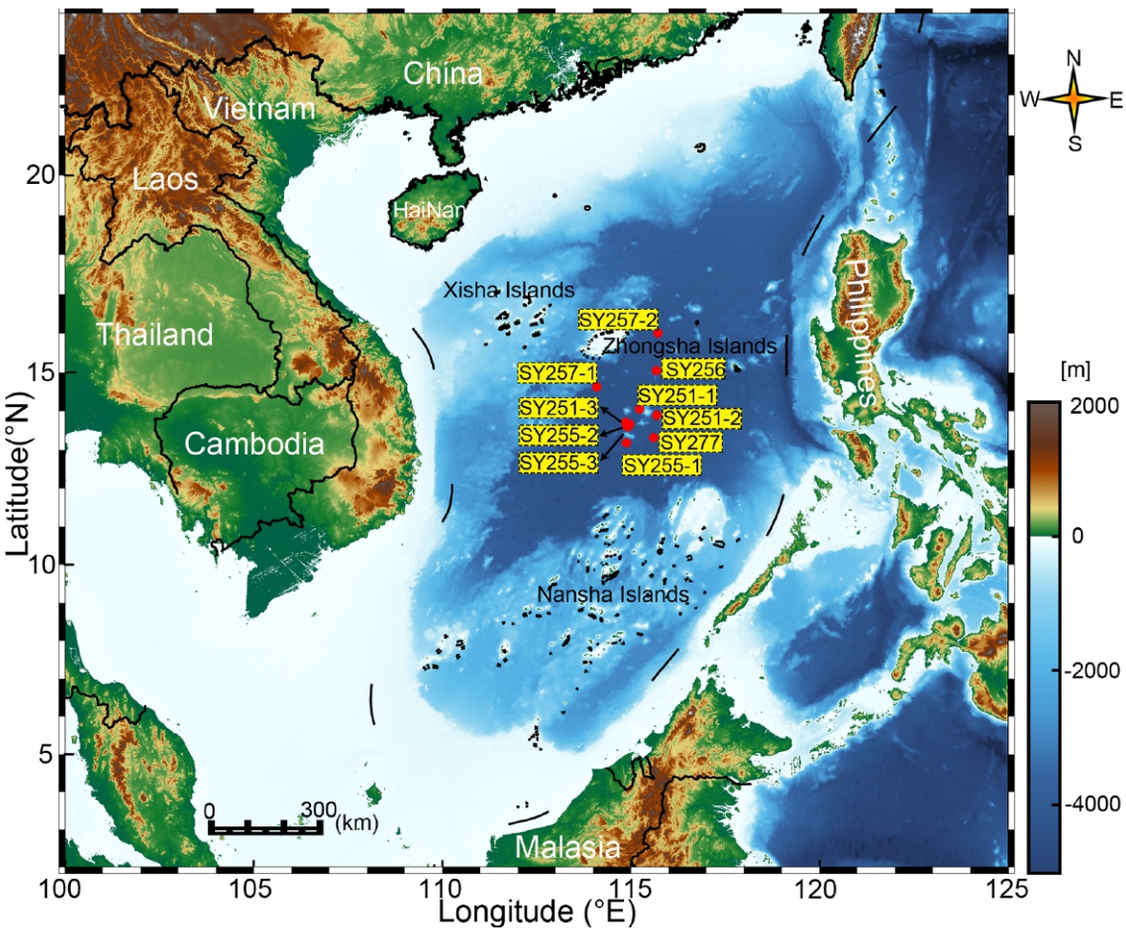


Figure 1 Sites of cold-water coral collection in SCS

Red circles denote samples collected at different depths during two cruises. Detailed information is shown in Table 1.

Table 1 Information on 10 mitochondrial genomes from this study

Sample ID	Species	Genome size (bp)	GC content	No. PCGs	No. tRNA	No. rRNA	Temperature (°C)	Salinity (ppt)	Density (kg/m³)	Longitude (°E)	Latitude (°N)	Depth (m)
SY251-1	*Sibogagorgia dennisgordoni* **sp. nov.**	19 530	0.36	14	1	2	2.75	34.53	1 034.91	115.22	14.16	1 623
SY251-2	*Victorgorgia eminens* **sp. nov.**	18 663	0.37	14	1	2	3.31	34.51	1 033.59	115.67	13.99	1 329
SY251-3	*Acanthogorgia hirsuta* **sp. nov.**	19 044	0.37	14	1	2	3.07	34.52	1 034.13	114.85	13.81	1 440
SY255-1	*Candidella helminthophora* **sp. nov.**	18 872	0.38	14	1	2	4.99	34.44	1 031.4	114.88	13.3	905
SY255-2	*Hemicorallium guttatum* **sp. nov.**	19 104	0.39	14	1	2	4.86	34.45	1 031.5	114.97	13.77	925
SY255-3	*Isidella* **sp. nov.**	19 481	0.38	14	1	2	3.14	34.51	1 033.74	114.91	13.7	1 356
SY256	*Calyptrophora wyvillei* **sp. nov.**	19 270	0.38	14	1	2	2.75	34.53	1 035.07	115.66	15.13	1 633
SY257-1	*Hemicorallium guttatum* **sp. nov.**	19 049	0.39	14	1	2	3.19	34.51	1 033.57	114.1	14.72	1 320
SY257-2	*Lepidisis* **sp. nov.**	19 223	0.38	14	1	2	2.95	34.52	1 034.66	115.7	16.08	1 552
SY277	*Victorgorgia macrocalyx* **sp. nov.**	13 067	0.37	9	0	2	4.54	34.46	1 031.73	115.59	13.42	964

the parameter "--min-contig-len 1000 --k-list 21,33,55,77,99,127" (Li et al., 2014). Additionally, 134 reference octocoral mitochondrial genomes (number of PCGs >7 and length >10 kbp) from the NCBI database were used to identify potential mitochondrial contigs using BLASTn with an e-value of 1e-5 (Supplementary Table S1). The quality-filtered data were compared with the potential mitochondrial contigs using bowtie2 (v2.4.4) ("--very-fast-local") to extract potential mitochondrial reads (Langmead & Salzberg, 2012). Subsequently, the reads were reassembled into mitochondrial genomes using SPAdes (v.3.6.2) with "--careful -k 21,33,55,77,99" (Nurk et al., 2013).

The Mitos2 and Ge-SEQ webservers (http://mitos2.bioinf.uni-leipzig.de/index.py and https://chlorobox.mpimp-golm.mpg.de/geseq.html) were used to annotate the mitochondrial assemblies with a gene identity of 85% (Donath et al., 2019; Tillich et al., 2017), while the NR database was also used to verify the PCGs. Online tRNAscan-SE2.0

software (http://lowelab.ucsc.edu/tRNAscan-SE) was used to identify the tRNA genes (Lowe & Chan, 2016). The output GenBank format files were manually confirmed and modified based on the annotation results, which were used to draw the circular map using Organellar Genome DRAW (OGDRAW) on the MPI-MP CHLOROBX website (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) (Greiner et al., 2019).

Mitochondrial assembly structural analysis

Genome size and GC content of the mitochondrial genomes were calculated using in-house scripts. Average nucleotide identity (ANI) values of the 134 reference octocoral genomes and our 10 mitochondrial genomes were determined through pairwise comparisons using fastANI with the parameter "--minFraction 0.8" (Table 1; Supplementary Table S1) (Richter et al., 2016). Tandem repeats were identified using Tandem Repeats Finder (v.4.09) with default settings (https://tandem.bu.edu/trf/trf.html) (Benson, 1999). Composition skewness of each PCG in the mitochondrial genomes was calculated using the following formulas: AT-skew=(A−T)/(G+C), GC-skew=(G−C)/(G+C). Additionally, 125 of the 134 reference octocoral genomes and nine of our mitochondrial genomes were complete and contained adenosine triphosphate (ATP) synthase F0 subunit 6 (*atp6*), ATP synthase F0 subunit 8 (*atp8*), cytochrome *b* (cyt *b*), *cox1*, cytochrome c oxidase subunit 2 (*cox2*), cytochrome c oxidase subunit 3 (*cox3*), NADH dehydrogenase subunit 1 (*nad1*), NADH dehydrogenase subunit 2 (*nad2*), NADH dehydrogenase subunit 3 (*nad3*), NADH dehydrogenase subunit 4 (*nad4*), NADH dehydrogenase subunit 4L (*nad4l*), NADH dehydrogenase subunit 5 (*nad5*), NADH dehydrogenase subunit 6 (*nad6*), and *mutS*. These genes were selected to calculate dissimilarity of GC-skew (method=Euclidean) with the vegdist function and principal coordinates analysis (PCoA) in R (v.4.3.1).

Phylogenetic analysis

A phylogenetic tree was constructed using the 134 reference octocoral genomes, three reference hexacoral genomes, and 10 mitochondrial genomes from this study (Table 1; Supplementary Table S1). Within the mitochondrial genomes, 14 conserved PCGs were extracted using an in-house script and aligned using MAFFT L-INS-i (v7.294b), respectively (Katoh & Toh, 2010). The aligned protein sequences for each PCG were trimmed using trimAl (v.1.4) with the parameter "-automated1" to remove poorly aligned regions (Capella-Gutiérrez et al., 2009). Subsequently, the trimmed protein sequences were concatenated using an in-house script. A maximum-likelihood phylogenetic tree was then generated using iq-tree (v.1.6.12) with "MFP" to select the optimal model ("-m MFP -bb 1000 -alrt 1000") (Nguyen et al., 2015). Three *Hexacorallia* reference genomes were designated as the outgroup for analysis. Additionally, the resultant treefile, after removal of the outgroup, was used to compute phylogenetic distances with the ape package in R (v.4.3.1).

***Ka*/*Ks* ratio analysis**

To elucidate the selective pressure on the mitochondrial genes of cold-water corals, the complete mitochondrial genomes were also selected to analyze the *Ka*/*Ks* ratio using KaKs_Calculator2 with the averaging model (MA) and coelenterate mitochondrial genetic code (Table 1; Supplementary Table S1) (Wang et al., 2010). In addition, the median *Ka*/*Ks* values of the 14 PCGs were used to illustrate the genetic evolutionary heterogeneity among our nine complete mitochondrial genomes using the vegdist function (method=Euclidean) in R (v.4.3.1). Additionally, the median *Ka*/*Ks* values of the *atp*, cyt *b*, *mutS*, *cox*, and *nad* gene families were co-analyzed with environmental factors to assess the importance of environmental factors in molecular evolution using the Mantel test in R (v.4.3.1). The relationship between environmental factors was determined using Spearman correlation analysis with the corrplot package in R (v.4.3.1). Pairwise correlation analyses of GC-skew dissimilarity, phylogenetic distance, and ANI values from the complete mitochondrial genomes described above were conducted using MRM in R (v.4.3.1).

RESULTS

Genomic structure

Ten deep-sea coral samples were collected at depths ranging from 905 m to 1 633 m by a manned submersible (Shenhai Yongshi, China) during two cruises in the SCS (Figure 1; Table 1). The count of paired-end reads for potential mitochondrial genomes was 211 303 (Supplementary Table S2). Ten mitochondrial genomes were assembled, with genome sizes ranging from 13 067 bp to 19 530 bp and GC content ranging from 0.36 to 0.39 (Table 1). Circular structure maps with 14 PCGs were obtained for all mitochondrial genomes, except for SY277, which only contained nine PCGs (Table 1). Moreover, two rRNA genes (*rrnS* and *rrnL*) and one tRNA (tRNA-Met) were present in the nine complete mitochondrial genomes. The *mutS* gene was identified in all 10 mitochondrial genomes (Figure 2).

The complete mitochondrial genomes exhibited different gene orders and arrangements (Figure 2; Supplementary Table S3). Among the complete mitochondrial genomes, the genomic structures of SY256, SY255-1, SY251-3, and SY251-2 exhibited identical genomic structures. In these genomes, *cox1*, *rrnS*, *nad1*, cyt *b*, *nad6*, *nad3*, *nad4l*, *mutS*, *rrnL*, *nad2*, *nad5*, and *nad4* were on the heavy strand, and *trnM*, *cox3*, *atp6*, *atp8*, and *cox2* were on the light strand (Figure 2). The genomic structures of SY257-2 and SY255-3 were found to be identical, in which *cox1*, *rrnS*, *nad1*, cyt *b*, *nad6*, *nad3*, and *nad4l* were on the heavy strand, and *nad4*, *nad5*, *nad2*, *rrnL*, *mutS*, *trnM*, *cox3*, *atp6*, *atp8*, and *cox2* were on the light strand (Figure 2). The SY257-1 and SY255-2 were same species, with *nad6*, *nad3*, *nad4l*, *mutS*, *rrnL*, *nad2*, *nad5*, and *nad4* on the heavy strand, and *trnM*, *cox3*, *atp6*, *atp8*, *cox2*, cyt *b*, *nad1*, *rrnS*, and *cox1* on the light strand (Figure 2). The SY251-1 mitochondrial genome had a unique genomic structure, with *cox1*, *rrnS*, *nad1*, cyt *b*, *mutS*, *rrnL*, *nad2*, *nad5*, and *nad4* on the heavy strand, and *trnM*, *cox3*, *atp6*, *atp8*, *cox2*, *nad4l*, *nad3*, and *nad6* on the light strand (Figure 2). The genomic structure of the partial SY277 genome may be the same as that of SY256, SY255-1, SY251-3, and SY251-2 (Figure 2).

Except for *nad4l*, which utilized GTG as its start codon in SY251-1, ATG was the conserved start codon for all 14 PCGs across the mitochondrial genomes in this study. Furthermore, TAA (31.29%) and TAG (68.71%) were the primary termination codons for the 14 PCGs across the 10 mitochondrial genomes (Supplementary Figure S1). Tandem repeats were only identified in three mitochondrial genomes. One tandem repeat was found in SY251-1 with 94% sequence identity and a length of 35 bp. Two tandem repeats were

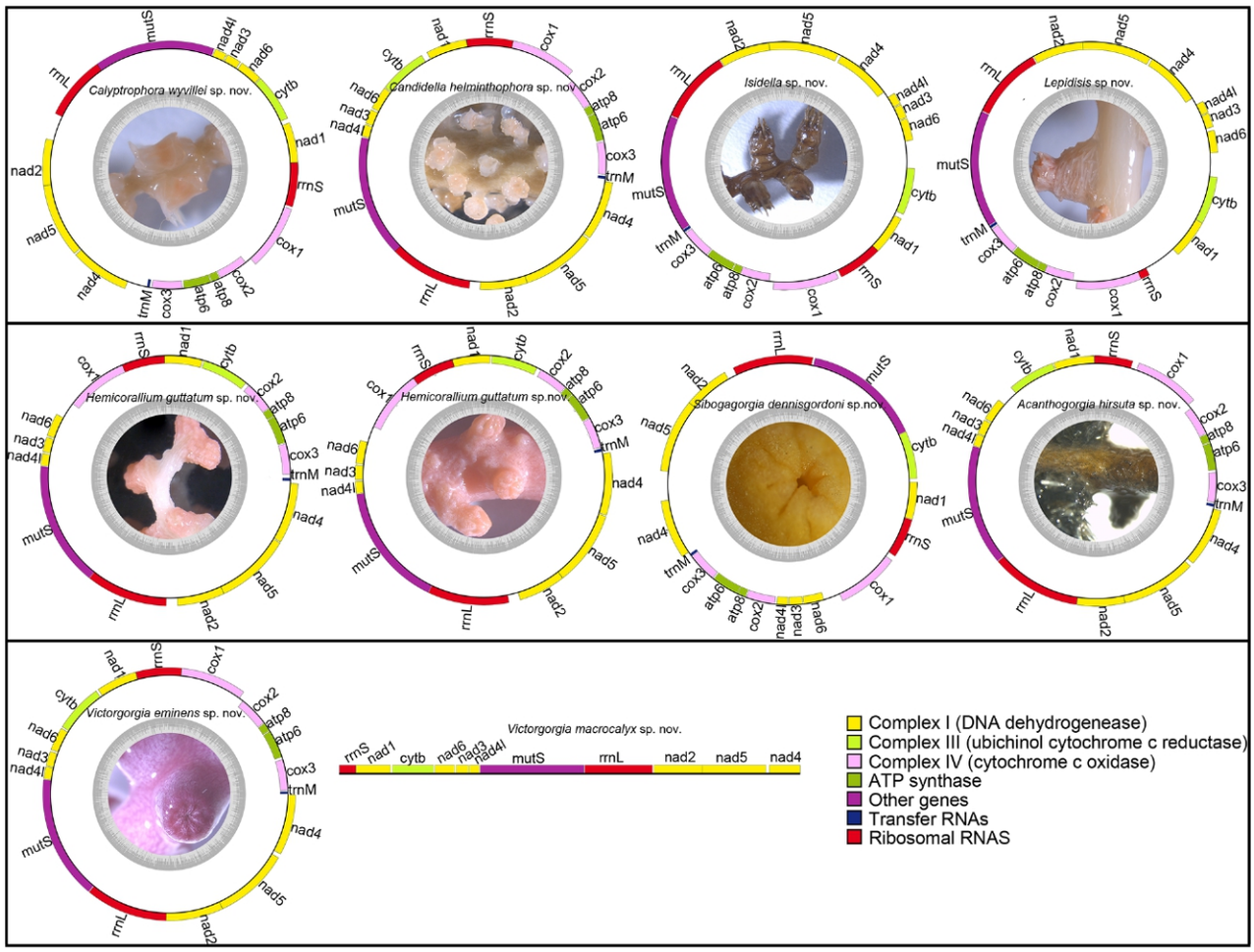


Figure 2 Circular maps of 10 cold-water coral samples

Gene map showing arrangement of 14 PCGs and one tRNA gene, and their positions on heavy or light strands. Gene arrangements are shown in Supplementary Table S3.

identified in SY255-3 with sequence identities of 88% and 95%, and lengths of 41 bp and 34 bp. Three tandem repeats were found in SY257-2 with sequence identities ranging from 89% to 100%, and lengths ranging from 30 bp to 61 bp.

Phylogenetic tree and morphological identification

The 10 mitochondrial genomes were assigned into four suborders and represented 9 new deep-sea species based on phylogenetic and morphological analysis (Figure 3; Supplementary Figures S2–S7). Among the genomes, SY257-2 and SY255-3 were clustered in the Isididae family with articulated skeletons (Figure 3; Supplementary Figure S2). Based on its whip-shaped unbranched colony and sclerite morphology, SY257-2 may represent a new species within the *Lepidisis* genus (*Lepidisis* **sp. nov.**) (Supplementary Figure S2A, B) (Heestand Saucier et al., 2021). While SY255-3 had similar sclerites and microcrystal ultra-structure to the *Lepidisis* genus, differences in morphology suggested that it may represent a novel species within the *Isidella* genus (*Isidella* **sp. nov.**) (Supplementary Figure S2C , D) (Heestand Saucier et al., 2021). Both SY256 and SY255-1 were clustered within the Primnoidae family and displayed different sclerite morphologies, potentially representing two new species in the *Calyptrophora* genus (*Calyptrophora wyvillei* **sp. nov.**) (Supplementary Figure S3A, B) and *Candidella* genus (*Candidella helminthophora* **sp. nov.**) (Supplementary Figure S3C, D) (Cairns, 2009, 2012). All four of the above deep-sea species belonged to the Calcaxonia suborder. Both SY257-1 and SY255-2 showed similar appearances and were placed in the *Hemicorallium* genus of the Coralliidae family (Figure 3; Supplementary Figure S4). Given their yellowish-white color, pale vermilion red colonies, and multiple sclerites, they may represent the same new species of *Hemicorallium guttatum* **sp. nov.** (Supplementary Figure S4) (Tu et al., 2016). SY251-1 was clustered in the Paragorgiidae family with an independent branch (Figure 3), and exhibited similar morphology and sclerite characteristics as those in the *Sibogagorgia* genus, inferring it may represent a novel species within that genus (*Sibogagorgia dennisgordoni* **sp. nov.**) (Supplementary Figure S5) (Sánchez, 2005). SY277 and SY251-2 were grouped together within the Clavulariidae family (Figure 3). Given their respective purple to pink and black polyps and coenenchyma and straight to slightly curved tuberculate spindles on their sclerites, they may represent two novel species within the *Victorgorgia* genus (*Victorgorgia macrocalyx* **sp. nov.** and *Victorgorgia eminens* **sp. nov.**) (Supplementary Figure S6) (Li et al., 2020; Moore et al., 2017). Although SY251-3 exhibited characteristics similar to those of the *Echinogorgia* genus in the Plexauridae family (Figure 3), its morphology and sclerites were more similar to the *Acanthogorgia* genus, suggesting that it may represent a novel species within that genus (*Acanthogorgia hirsuta* **sp. nov.**) (Supplementary Figure S7) (Horvath, 2019). Moreover,

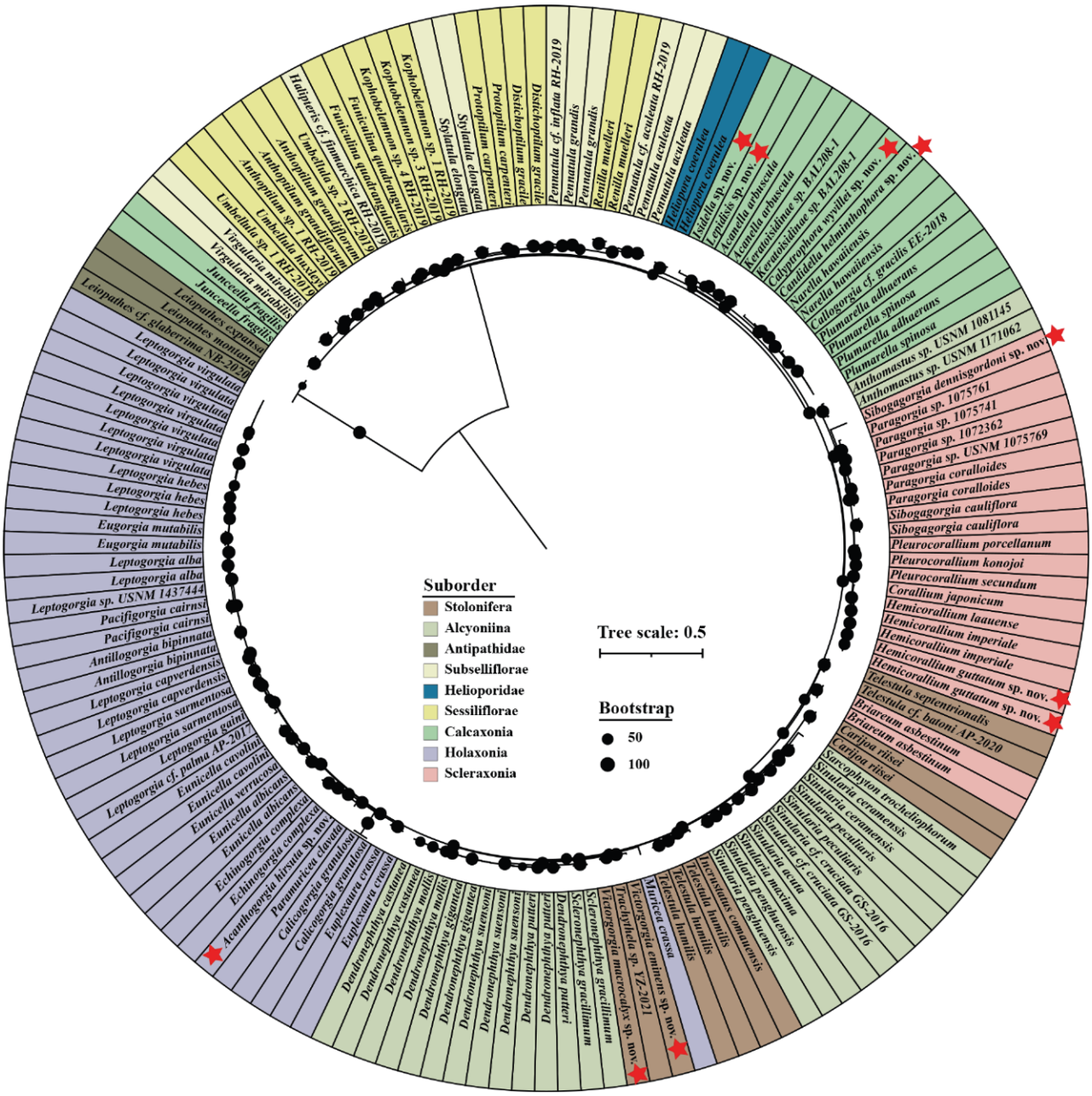


Figure 3 Maximum-likelihood phylogenetic tree of octocoral mitochondrial genomes

Tree was generated based on concatenated proteins of 14 PCGs by iq-tree. Three hexacoral mitochondrial genomes were used as the outgroup. Black dots on branches denote bootstrap support, with only those >50% depicted. Mitochondrial genomes in our study are marked with red stars. Reference mitochondrial genomes were collected from the NCBI database and are shown in Supplementary Table S1.

it was noteworthy that the species within the same family exhibited closer phylogenetic distances, and closer phylogenetic distances were also observed between those from different families in different suborders (Supplementary Figure S8).

ANI analysis

ANI pairwise comparisons for all octocoral mitochondrial genomes further demonstrated that our mitochondrial genomes represented 9 novel species assigned to four suborders (Calcaxonia, Holaxonia, Scleraxonia, and Stolonifera) (Figure 4A). Results showed that *Hemicorallium guttatum* **sp. nov.** was closely related to *Hemicorallium imperiale* (KC782352.1), with 99.66% and 99.74% sequence similarities, respectively (Figure 4Ba). *Acanthogorgia hirsuta* **sp. nov.** shared 97.43% sequence identity with *Echinogorgia complexa* (NC_020457.1 and HQ694727.1) (Figure 4Bb). *Sibogagorgia dennisgordoni* **sp. nov.** displayed 93.40% sequence similarity with *Sibogagorgia cauliflora* (NC_026193.1 and KM015354.1) (Figure 4Bc). *Candidella helminthophora* **sp. nov.** and *Calyptrophora wyvillei* **sp. nov.** shared 99.31% and 98.05% sequence identity with *Narella hawaiiensis* (NC_026192.1 and KM015351.1), respectively, and exhibited a sequence similarity of 98.34% to each other (Figure 4Bd). *Victorgorgia eminens* **sp. nov.** and *Victorgorgia macrocalyx* **sp. nov.** showed 98.40% and 98.59% sequence identities, respectively, with *Trachythela sp. YZ-2021* (MW238423.1) from the *Trachythela* genus, and 98.62% identity with each other (Figure 4Be). In addition, *Isidella* **sp. nov.** and *Lepidisis* **sp. nov.** showed 98.33% and 98.30%

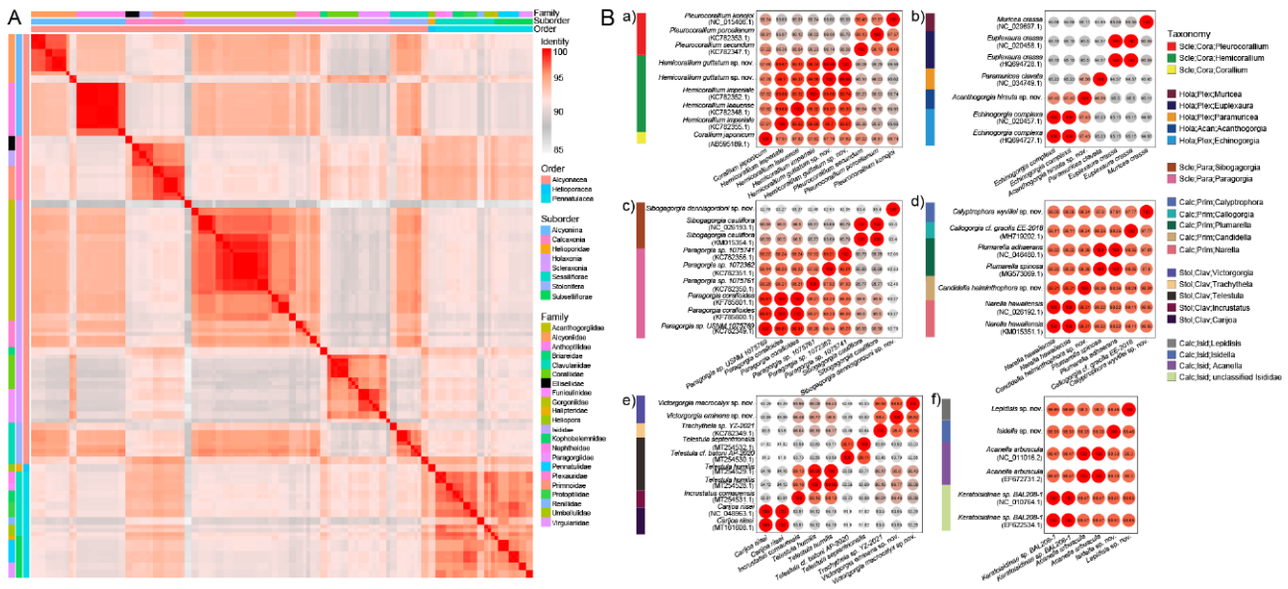


Figure 4 ANI plots of octocoral mitochondrial genomes

Octocoral mitochondrial genomes were used to calculate ANI values through pairwise comparison. A: ANI values of all mitochondrial genomes were sorted by taxonomy based on phylogenetic tree. B: ANI values of our 10 mitochondrial genomes. Reference mitochondrial genomes were collected from NCBI database and are shown in Supplementary Table S1. Suborder: Scle: Scleraxonia; Hola: Holaxonia; Clac: Calcaxonia; Stol: Stolonifera. Family: Cora: Coralliidae; Plex: Plexauridae; Acan: Acanthogorgiidae; Prim: Primnoidae; Clav: Clavulariidae; Isid: Isididae.

sequence similarities, respectively, with *Acanella arbuscula* (NC_011016.2 and EF672731.2) from the *Acanella* genus, and 98.46% sequence similarity to each other (Figure 4Bf). Interestingly, higher ANI values were observed in corals from different families, which may help in deep-sea species identification.

Positive selection analysis

All *Ka*/*Ks* ratios calculated from the 14 PCGs were examined to evaluate selective pressure on mitochondrial genes (Figure 5). However, all *Ka*/*Ks* ratios of the 14 PCGs from nine of our mitochondrial genomes were less than 1. Notably, *atp8* had the highest median *Ka*/*Ks* value in *Sibogagorgia dennisgordoni* **sp. nov.**, *Candidella helminthophora* **sp. nov.**, *Hemicorallium guttatum* **sp. nov.**, and *Calyptrophora wyvillei* **sp. nov.** (Kruskal-Wallis test, P<0.01), while *mutS* had the highest median *Ka*/*Ks* value in *Victorgorgia eminens* **sp. nov.**, *Acanthogorgia hirsuta* **sp. nov.**, *Isidella* **sp. nov.**, and *Lepidisis* **sp. nov.** (Kruskal-Wallis test, P<0.01). Conversely, *nad4l* had the lowest median *Ka*/*Ks* value (Kruskal-Wallis test, P<0.01), except for *Sibogagorgia dennisgordoni* **sp. nov.**, in which *cox1* was the lowest (Figure 5).

In addition, species with closer phylogenetic relationships, closer phylogenetic distances, and higher ANI values exhibited lower dissimilarities in median *Ka*/*Ks* values across the 14 PCGs (Supplementary Figure S9). Specifically, *Isidella* **sp. nov.** and *Lepidisis* **sp. nov.** displayed reduced mitochondrial evolutionary heterogeneity among the nine complete mitochondrial genomes based on median *Ka*/*Ks* values. Low genetic evolutionary heterogeneity was also observed between *Candidella helminthophora* **sp. nov.** and *Calyptrophora wyvillei* **sp. nov.** Furthermore, the dissimilarities among the remaining three novel species were substantial, consistent with the phylogenetic tree and ANI analyses.

Genetic distance analysis

The dissimilarities among the eight suborders were assessed using GC-skew calculations of the 14 PCGs (Supplementary Figure S10). Although the suborders explained 34.4% of the variance, the PCoA results did not reveal clear separations based on suborder classifications. This observation suggests that corals within the same suborder may not necessarily possess close genetic distances and may intermix with other suborders. The GC-skew-based dissimilarities aligned with both the phylogenetic distance and ANI findings. Furthermore, pairwise correlation analyses demonstrated a positive correlation between GC-skew dissimilarity and phylogenetic distance (R^2=0.37, P<0.001), a negative correlation between GC-skew dissimilarity and ANI (R^2=0.37, P<0.001), and a negative correlation between ANI and phylogenetic distance (R^2=0.88, P<0.001) (Supplementary Figure S11).

Environmental factor analysis

Various environmental factors, including temperature, salinity, density, depth, longitude, and latitude, were collected during sampling (Table 1). The water temperature ranged from 2.75°C to 4.99°C in the deep-sea area. Both salinity and density were stable during sampling, with an average of 34.50±0.03 ppt and 1 033.43±1.33 kg/m³, respectively. Correlation analysis of environmental factors showed that temperature was strongly negatively correlated with depth, salinity, and density (R<−0.97) (Supplementary Figure S12). Additionally, a strong positive correlation was observed between salinity and density (R=0.98) (Supplementary Figure S12). Furthermore, an increase in salinity and density was observed with increasing depth (R>0.97) (Supplementary Figure S12). The correlation between latitude and other environmental factors was higher than that between longitude and other environmental factors. Co-analysis of environmental factors and median *Ka*/*Ks* values of five gene families revealed that the cyt *b* gene was positively correlated with all environmental factors, except for latitude, longitude, and salinity, while *mutS* was correlated with all environmental factors, except for latitude and longitude. No significant correlations between other genes and environmental factors

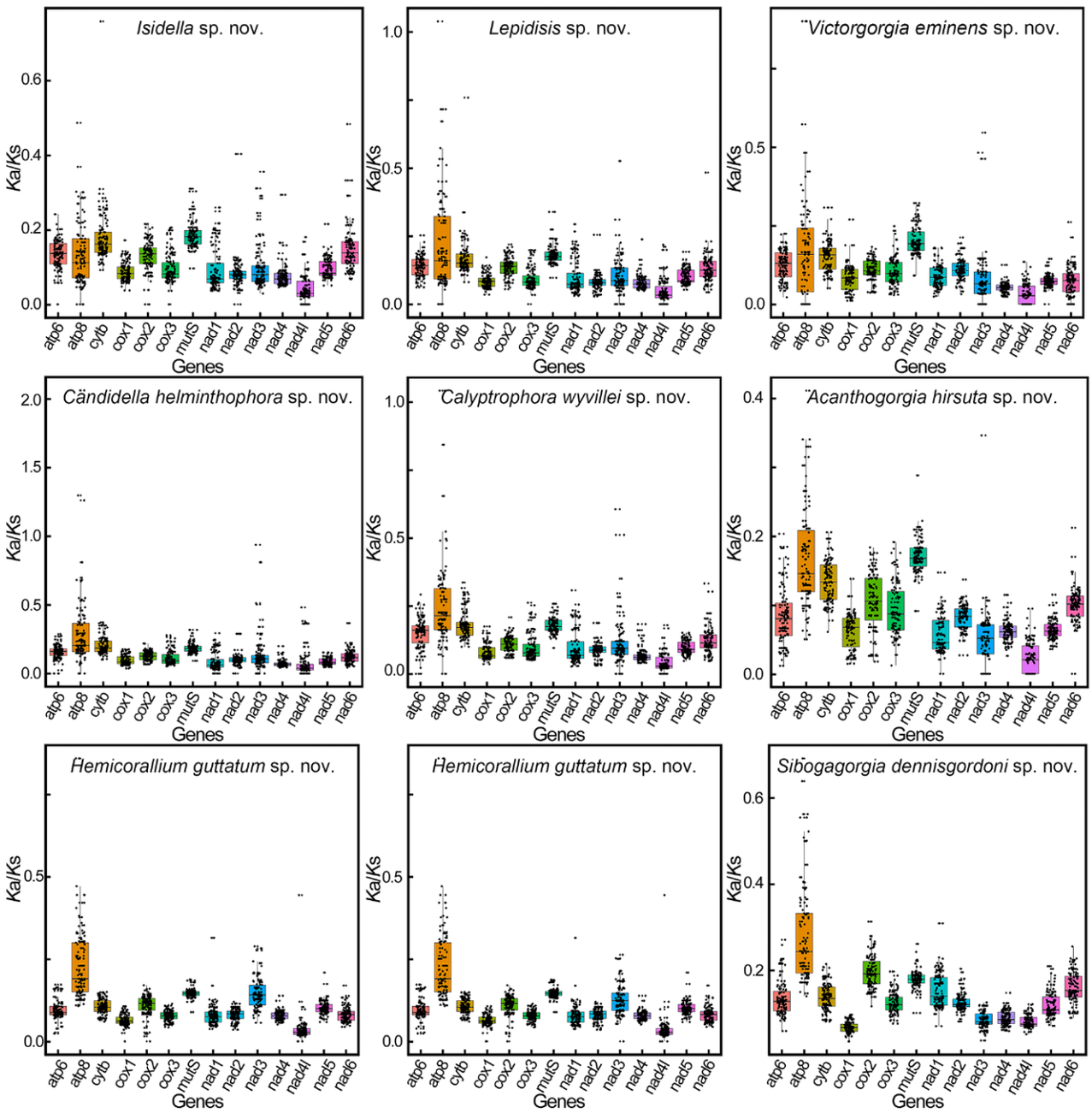


Figure 5 Boxplots of *Ka*/*Ks* ratios of 14 PCGs from nine complete mitochondrial genomes in this study

Complete octocoral mitochondrial genomes were used to calculate *Ka*/*Ks* ratios, as shown in Supplementary Table S1.

were found.

DISCUSSION

In the present study, 10 cold-water coral samples were collected at depths ranging from 905 m to 1 633 m in the SCS. Based on morphological identification and mitochondrial genome structure, the samples were classified within the Alcyonacea order of the Octocorallia subclass, and subsequently divided into four suborders (Calcaxonia, Holaxonia, Scleraxonia, and Stolonifera) and identified as 9 possible novel cold-water coral species. Phylogenetic analysis of all publicly available mitochondrial genomes from the Octocorallia subclass revealed that different suborders may have overlapping positions on the phylogenetic tree. ANI values within the same family were lower than those between families. Conversely, analysis of phylogenetic distance yielded the opposite results, while GC-skew-based dissimilarities did not demonstrate clear separation by suborder. These results support the hypothesis that mitochondrial genomes in the Octocorallia subclass exhibit a lower evolutionary rate compared to that of other animal species (Shearer et al., 2002).

The GC content and genome size of our cold-water samples, which ranged from 0.36–0.39 and 13 067–19 530 bp, respectively, were the same as those in the Alcyonacea order. No obvious genome expansions or contractions were found in any of the 10 mitochondrial genomes, indicating minimal gene loss or acquisition events during the diversification process in deep-sea corals. The stable genomic architecture in terms of both size and GC content within the Alcyonacea order may be associated with the slow rates of nucleotide substitution in anthozoans (Hellberg, 2006; Shearer et al., 2002).

Although gene number was highly conserved in the Alcyonacea order, significant differences in gene order and position on heavy and light strands were observed in the different families. Five different gene arrangements have been described in the Octocorallia subclass (Beaton et al., 1998; Brockman & McFadden, 2012; Brugler & France, 2008; Park et al., 2012; Uda et al., 2011). In the current study, we observed four different gene arrangements, among which *Calyptrophora wyvillei* **sp. nov.**, *Candidella helminthophora* **sp. nov.**, *Acanthogorgia hirsuta* **sp. nov.**, and *Victorgorgia eminens* **sp. nov.** exhibited ancestral arrangements in octocorals (McFadden et al., 2006; Uda et al., 2011). *Sibogagorgia dennisgordoni* **sp. nov.** displayed the "konojoi" arrangement discovered in *Corallium konojoi* (Uda et al., 2011). The *Hemicorallium guttatum* **sp. nov.** showed the "japonicum" arrangement consistent with *Paracorallium japonicum* (Uda et al., 2011). Additionally, the gene arrangements of *Isidella* **sp. nov.** and *Lepidisis* **sp. nov.** were the same as that of Keratoisidinae (Van Der Ham et al., 2009). Based on mechanisms such as tandem duplication via slipped-strand mispairing, followed by random deletion of genes and gene inversion via intramitochondrial recombination (Brugler & France, 2008; Mueller & Boore, 2005), we hypothesized that at least three gene rearrangement events may have occurred during cold-water coral evolution in the deep-sea environment, with one gene rearrangement event occurring during the evolution of the Calcaxonia suborder and two gene rearrangement events occurring in the Scleraxonia suborder.

Typically, either TAG or TAA serves as the termination codon formed by post-transcriptional polyadenylation, a common feature in many metazoan mitochondrial genomes (Brugler & France, 2008). Consistent with other octocoral mitochondrial genomes, only a single tRNA was identified, suggesting extensive loss of tRNA genes during evolution, as also observed in protozoans and angiosperms (Small et al., 1992). Tandem repeats are widely found in eukaryotic genomes as well as in some prokaryotes (Gao & Kong, 2005). The presence of tandem repeats in three of the mitochondrial genomes may be indicative of gene rearrangements during evolutionary adaptation to deep-sea environments.

The *Ka*/*Ks* ratio is a common indicator used to assess selective pressure and evolutionary relationships among species in molecular studies (Zhu et al., 2017). Here, we analyzed the *Ka*/*Ks* ratios of the 14 PCGs in the mitochondrial genomes to provide a comprehensive understanding of cold-water coral evolution in the deep-sea environment. Strong purifying selection may be a major driving force influencing mitochondrial genome diversity in deep-sea corals (Noll et al., 2022). Environmental conditions contribute substantially to the basic patterns of genome evolution driven by random genetic drift and mutation pressure (Schaack et al., 2020). Moreover, environmental factors can regulate the expression levels of PCGs, thereby influencing energy regulation, reproduction, and immune behaviors (Yan et al., 2019). Therefore, a species may undergo positive selection during adaptation to a new environmental niche, leading to the formation of novel functions (Nielsen, 2005). The purifying selection of cold-water corals indicated that the deep-sea environment variation might be insufficient to induce alterations in genetic functionality. In addition, the *atp8* gene, which plays an important role in ATP synthesis, exhibited the highest median *Ka*/*Ks* value in selected cold-water coral species, consistent with previous studies in other organisms (Kumar et al., 2020; Lv et al., 2018; Niu et al., 2020). Based on their higher *Ka*/*Ks* ratios and increased amino acid changes, the *atp8* and *mutS* genes may have evolved more rapidly, experiencing more relaxed selective constraints and accumulating more mutations.

In this study, environmental factors were initially co-analyzed with gene selective pressure. Significant positive relationships were observed between environmental factors and the evolution of the cyt *b* and *mutS* genes. The cyt *b* gene encodes cytochrome *b* and participates in the respiratory chain to produce electrochemical potential coupled with ATP synthesis, while the *mutS* gene participates in DNA mismatch repair and mediation of intramolecular recombination (Hsieh et al., 2001; Muthye et al., 2022). These results suggest that environmental factors may have influenced the evolutionary trajectory of deep-sea corals by affecting mutations in genes related to energy metabolism.

CONCLUSIONS

Cold-water corals provide both organic food and benthic habitats for many organisms, thereby contributing to stable ecosystems in deep-ocean regions characterized by darkness, low temperature, oligotrophy, and high hydrostatic pressure. As such, exploring the origins and evolutionary pathways of these deep-sea corals is of considerable importance. In this study, we conducted a systematic analysis of mitochondrial genomes from cold-water corals, classified into four suborders belonging to the Alcyonacea order and Octocorallia subclass. Our analysis contributes to a better understanding of the taxonomic status, species diversity, distribution, mitochondrial characteristics, and genetic plasticity of these deep-sea corals. Notably, our findings suggest that extreme deep-sea environmental factors may drive coral evolution by inducing mutations in genes related to energy metabolism, which warrants further validation through the collection of more environmental data in future studies.

DATA AVAILABILITY

The raw data and mitochondrial genomes in this study were submitted to NCBI (PRJNA951866), Science Data Bank (10.57760/sciencedb.j00139.00051), and GSA (PRJCA016049).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

The project was conceived by S.S.L., X.T.P., and C.W.S.. Sample collection, DNA extraction, and DNA library preparation were carried out by N.N.Z., C.Y.C. and L.M. Morphological examination and sclerite scanning were carried out by K.W.T. Pipeline assembly, including data cleaning, genome assembly, annotation, and all downstream analyses, was carried out by Z.F.W. and K.Q.L. The manuscript was written by Z.F.W. with contributions from all authors. All authors edited the manuscript before submission. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We are grateful to the crew of the scientific research vessel *Tansuo Yihao* for sample collection and to the China National GeneBank for their technical support in library construction and sequencing.

REFERENCES

Andrews S. 2014. FastQC a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Baillon S, Hamel JF, Wareham Hayes V, et al. 2012. Deep cold-water corals as nurseries for fish larvae. *Frontiers in Ecology and the Environment*, **10**(7): 351−356.

Beaton MJ, Roger AJ, Cavalier-Smith T. 1998. Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum*: conserved gene order among octocorals. *Journal of Molecular Evolution*, **47**(6): 697−708.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, **27**(2): 573−580.

Bilewitch JP, Degnan SM. 2011. A unique horizontal gene transfer event has provided the octocoral mitochondrial genome with an active mismatch repair gene that has potential for an unusual self-contained function. *BMC Evolutionary Biology*, **11**(1): 228.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Research*, **27**(8): 1767−1780.

Brockman S, McFadden C. 2012. The mitochondrial genome of *Paraminabea aldersladei* (Cnidaria: Anthozoa: Octocorallia) supports intramolecular recombination as the primary mechanism of gene rearrangement in octocoral mitochondrial genomes. *Genome Biology and Evolution*, **4**(9): 994−1006.

Brugler MR, France SC. 2008. The mitochondrial genome of a deep-sea bamboo coral (Cnidaria, Anthozoa, Octocorallia, Isididae): genome structure and putative origins of replication are not conserved among octocorals. *Journal of Molecular Evolution*, **67**(2): 125−136.

Cairns SD. 2009. Review of octocorallia (Cnidaria: Anthozoa) from Hawai'i and adjacent seamounts. Part 2: Genera *Paracalyptrophora*; *Candidella*; and *Calyptrophora*. *Pacific Science*, **63**(3): 413–448.

Cairns SD. 2012. The Marine Fauna of New Zealand: New Zealand Primnoidae (Anthozoa, Alcyonacea) - Part 1. Genera *Narella, Narelloides, Metanarella, Calyptrophora*, and *Helicoprimnoa*. Wellington: NIWA.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**(15): 1972−1973.

Carapelli A, Fanciulli P P, Frati F, et al. 2019. Mitogenomic data to study the taxonomy of Antarctic springtail species (Hexapoda: Collembola) and their adaptation to extreme environments. *Polar Biology*, **42**(4): 715−732.

Chen SF, Zhou YQ, Chen Y, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884−i890.

Donath A, Jühling F, Al-Arab M, et al. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Research*, **47**(20): 10543−10552.

Fryer P, Wheat CG, Williams T, et al. 2020. Mariana serpentinite mud volcanism exhumes subducted seamount materials: implications for the origin of life. *Philosophical Transactions of the Royal Society A:Mathematical, Physical and Engineering Sciences*, **378**(2165): 20180425.

Gao H, Kong J. 2005. Distribution characteristics and biological function of tandem repeat sequences in the genomes of different organisms. *Zoological Research*, **26**(5): 555−564. (in Chinese)

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3. 1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, **47**(W1): W59−W64.

Hebbeln D, Wienberg C, Dullo WC, et al. 2020. Cold-water coral reefs thriving under hypoxia. *Coral Reefs*, **39**(4): 853−859.

Heestand Saucier E, France SC, Watling L. 2021. Toward a revision of the bamboo corals: part 3, deconstructing the Family Isididae. *Zootaxa*, **5047**(3): 247−272.

Hellberg ME. 2006. No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. *BMC Evolutionary Biology*, **6**: 24.

Horvath EA. 2019. A review of gorgonian coral species (Cnidaria, Octocorallia, Alcyonacea) held in the Santa Barbara Museum of Natural History research collection: focus on species from Scleraxonia, Holaxonia, and Calcaxonia - Part I: introduction, species of Scleraxonia and Holaxonia (Family Acanthogorgiidae). *ZooKeys*, **860**: 1−66.

Hsieh HM, Chiang HL, Tsai LC, et al. 2001. Cytochrome b gene for species identification of the conservation animals. *Forensic Science International*, **122**(1): 7−18.

Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics*, **26**(15): 1899−1900.

Kroemer G, Reed JC. 2000. Mitochondrial control of cell death. *Nature Medicine*, **6**(5): 513−519.

Kumar V, Tyagi K, Chakraborty R, et al. 2020. The complete mitochondrial genome of endemic giant tarantula, *Lyrognathus crotalus* (Araneae: Theraphosidae) and comparative analysis. *Scientific Reports*, **10**(1): 74.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357−359.

Li DH, Liu CM, Luo RB, et al. 2014. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics*, **31**(10): 1674−1676.

Li F, Lv YY, Wen ZY, et al. 2021. The complete mitochondrial genome of the intertidal spider (*Desis jiaxiangi*) provides novel insights into the adaptive evolution of the mitogenome and the evolution of spiders. *BMC Ecology and Evolution*, **21**(1): 72.

Li JR, Wang PX. 2019. Discovery of deep-water bamboo coral forest in the South China Sea. *Scientific Reports*, **9**(1): 15453.

Li Y, Zhan ZF, Xu KD. 2020. Morphology and molecular phylogenetic analysis of deep-sea purple gorgonians (Octocorallia: Victorgorgiidae) from seamounts in the Tropical Western Pacific, with description of three new species. *Frontiers in Marine Science*, **7**: 701.

Lowe TM, Chan PP. 2016. TRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Research*, **44**(W1): W54−W57.

Lv YY, Li YP, Ruan ZQ, et al. 2018. The complete mitochondrial genome of *Glyptothorax macromaculatus* provides a well-resolved molecular phylogeny of the Chinese sisorid catfishes. *Genes*, **9**(6): 282.

McFadden CS, France SC, Sánchez JA, et al. 2006. A molecular phylogenetic analysis of the Octocorallia (Cnidaria: Anthozoa) based on mitochondrial protein-coding sequences. *Molecular Phylogenetics and Evolution*, **41**(3): 513−527.

McFadden CS, Sánchez JA, France SC. 2010. Molecular phylogenetic insights into the evolution of octocorallia: a review. *Integrative and Comparative Biology*, **50**(3): 389−410.

Moore KM, Alderslade P, Miller KJ. 2017. A taxonomic revision of *Anthothela* (Octocorallia: Scleraxonia: Anthothelidae) and related genera, with the addition of new taxa, using morphological and molecular data. *Zootaxa*, **4304**(1): 1.

Mueller RL, Boore JL. 2005. Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. *Molecular Biology and Evolution*, **22**(10): 2104−2112.

Muthye V, Mackereth CD, Stewart JB, et al. 2022. Large dataset of octocoral mitochondrial genomes provides new insights into *mt-mutS* evolution and function. *DNA Repair*, **110**: 103273.

Nguyen LT, Schmidt HA, Von Haeseler A, et al. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, **32**(1): 268−274.

Nielsen R. 2005. Molecular signatures of natural selection. *Annual Review of Genetics*, **39**: 197−218.

Niu WT, Xiao JG, Tian P, et al. 2020. Characterization of the complete mitochondrial genome sequences of three Merulinidae corals and novel insights into the phylogenetics. *PeerJ*, **8**: e8455.

Noll D, Leon F, Brandt D, et al. 2022. Positive selection over the mitochondrial genome and its role in the diversification of gentoo penguins in response to adaptation in isolation. *Scientific Reports*, **12**(1): 3767.

Nurk S, Bankevich A, Antipov D, et al. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *Journal of Computational Biology*, **20**(10): 714−737.

Ogata H, Ray J, Toyoda K, et al. 2011. Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. *The ISME Journal*, **5**(7): 1143−1151.

Park E, Hwang DS, Lee JS, et al. 2012. Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Molecular Phylogenetics and Evolution*, **62**(1): 329−345.

Richter M, Rosselló-Móra R, Oliver Glöckner F, et al. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*, **32**(6): 929−931.

Roberts JM, Wheeler AJ, Freiwald A. 2006. Reefs of the deep: the biology and geology of cold-water coral ecosystems. *Science*, **312**(5773): 543−547.

Sánchez JA. 2005. Systematics of the bubblegum corals (Cnidaria: Octocorallia: Paragorgiidae) with description of new species from New Zealand and the Eastern Pacific. *Zootaxa*, **1014**(1): 1−72.

Schaack S, Ho EKH, Macrae F. 2020. Disentangling the intertwined roles of mutation, selection and drift in the mitochondrial genome. *Philosophical Transactions of the Royal Society B:Biological Sciences*, **375**(1790): 20190173.

Shearer TL, Van Oppen MJH, Romano S, L et al. 2002. Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). *Molecular Ecology*, **11**(12): 2475−2487.

Shtolz N, Mishmar D. 2019. The mitochondrial genome-on selective constraints and signatures at the organism, cell, and single mitochondrion levels. *Frontiers in Ecology and Evolution*, **7**: 342.

Small I, Maréchal-Drouard L, Masson J, et al. 1992. In vivo import of a normal or mutagenized heterologous transfer RNA into the mitochondria of transgenic plants: Towards novel ways of influencing mitochondrial gene expression?. *The EMBO Journal*, **11**(4): 1291−1296.

Sokolova I. 2018. Mitochondrial adaptations to variable environments and their role in Animals' stress tolerance. *Integrative and Comparative Biology*, **58**(3): 519−531.

Sun Z, Hamel JF, Edinger E, et al. 2010. Reproductive biology of the deep-sea octocoral *Drifa glomerata* in the Northwest Atlantic. *Marine Biology*, **157**(4): 863−873.

Taanman JW. 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, **1410**(2): 103−123.

Thierens M, Browning E, Pirlet H, et al. 2013. Cold-water coral carbonate mounds as unique palaeo-archives: the plio-pleistocene challenger Mound record (NE Atlantic). *Quaternary Science Reviews*, **73**: 14−30.

Tillich M, Lehwark P, Pellizzer T, et al. 2017. GeSeq-Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, **45**(W1): W6−W11.

Tu TH, Dai CF, Jeng MS. 2016. Taxonomic revision of Coralliidae with descriptions of new species from New Caledonia and the Hawaiian Archipelago. *Marine Biology Research*, **12**(10): 1003−1038.

Uda K, Komeda Y, Koyama H, et al. 2011. Complete mitochondrial genomes of two Japanese precious corals, *Paracorallium japonicum* and *Corallium konojoi* (Cnidaria, Octocorallia, Coralliidae): Notable differences in gene arrangement. *Gene*, **476**(1-2): 27−37.

Van De Water JAJM, Allemand D, FerrierPagès C. 2018. Host-microbe interactions in octocoral holobionts - recent advances and perspectives. *Microbiome*, **6**(1): 64.

Van Der Ham JL, Brugler MR, France SC. 2009. Exploring the utility of an indel-rich, mitochondrial intergenic region as a molecular barcode for bamboo corals (Octocorallia: Isididae). *Marine Genomics*, **2**(3-4): 183−192.

Wang DP, Zhang YB, Zhang Z, et al. 2010. KaKs_Calculator 2.0: a toolkit incorporating Gamma-series methods and sliding window strategies. *Genomics, Proteomics & Bioinformatics*, **8**(1): 77−80.

Wang ZF, Shi XJ, Sun LX, et al. 2017. Evolution of mitochondrial energy metabolism genes associated with hydrothermal vent adaption of Alvinocaridid shrimps. *Genes & Genomics*, **39**(12): 1367−1376.

Xu Y, Zhan ZF, Xu KD. 2021. Morphological and molecular characterization of five species including three new species of golden gorgonians (Cnidaria: Octocorallia) from seamounts in the Western Pacific. *Biology*, **10**(7): 588.

Yan CJ, Duanmu XY, Zeng X, et al. 2019. Mitochondrial DNA: distribution, mutations, and elimination. *Cells*, **8**(4): 379.

Zhang B, Zhang YH, Wang X, et al. 2017. The mitochondrial genome of a sea anemone *Bolocera* sp. exhibits novel genetic structures potentially involved in adaptation to the deep-sea environment. *Ecology and Evolution*, **7**(13): 4951−4962.

Zhou Y, Feng CG, Pu YJ, et al. 2021. The first draft genome of a cold-water coral *Trachythela* sp. (Alcyonacea: Stolonifera: Clavulariidae). *Genome Biology and Evolution*, **13**(2): evaa265.

Zhu KC, Liang YY, Wu N, et al. 2017. Sequencing and characterization of the complete mitochondrial genome of Japanese Swellshark (*Cephalloscyllium umbratile*). *Scientific Reports*, **7**(1): 15299.

Zhang et al. *Zool. Res.* 2024, 45(1): 226−229
https://doi.org/10.24272/j.issn.2095-8137.2023.292

ZR | Zoological Research

Letter to the editor Open Access

Ancient DNA unravels species identification from Laosicheng site, Hunan Province, China, and provides insights into maternal genetic history of East Asian leopards

DEAR EDITOR,

The Laosicheng site in Yongshun County, Hunan Province, China, plays an important role in elucidating the animal population composition of the region prior to the middle Qing Dynasty (approximately 300 years ago). This site has yielded a substantial number of animal bones, 11 960 of which have been identified to the species level, serving as direct evidence of ancient animal populations (Hunan Provincial Institute of Cultural Relics and Archaeology, 2018). Notably, two mammalian skulls from this site, designated as Laosicheng1: 10YLG10 (101) and Laosicheng2: 10YLG10 (134), have garnered particular attention, with Laosicheng2 selected as the cover image for the above publication. Initially, these skulls were classified as Canids, with Laosicheng1 identified as a dog and Laosicheng2 identified as a wolf (Hunan Provincial Institute of Cultural Relics and Archaeology, 2018). However, the accuracy of these classifications has been called into question. In this context, a comprehensive analysis was undertaken, involving precise cranial morphological examination and mitochondrial genome (mitogenome) sequencing of the two Laosicheng samples using micro-computed tomography (micro-CT) and ancient DNA (aDNA) techniques to facilitate taxonomic reassessment. The results revealed that both skulls are attributable to leopards (*Panthera pardus japonensis*). Furthermore, phylogenetic analysis using data from this study and global leopard datasets, with a focus on East Asia, strongly suggested that East Asian leopards (*P. p. orientalis*, *P. p. japonensis*, and *P. p. delacouri*) belong to three distinct populations. This insight significantly advances our understanding of the maternal genetic history of East Asian leopards.

Initially, micro-CT scans (Institute of High Energy Physics, Chinese Academy of Sciences, China) were performed on the two samples to obtain precise cranial morphological data for species identification. This process also ensured the preservation of the original morphological characteristics of the skull before proceeding to genetic analysis. To describe the skull morphology, 22 skull measurements (Supplementary Figure S1 and Table S1) were taken from 3D models, with the average values and standard deviations (SD) of these measurements shown in Supplementary Table S1. Skull size was generally similar to that of *P. pardus* and *P. uncia*, but conspicuously larger than that of *Neofelis* and Felini species. Notably, the auditory bulla of the skulls was inflated, with a relatively small ectotympanic chamber, a feature distinct from *P. uncia*. Furthermore, the mastoid process was relatively weak, and the distant postorbital constriction supported a closer affinity to *P. pardus* than to *P. uncia*. For accurate species identification, photographs displaying the dorsal view of skulls from 23 extant Felidae species, including 15 Felini species, six Pantherini species, and the two Laosicheng samples, were collected and used for principal component analysis (PCA) employing a variance-covariance matrix of eight landmark measurements (Supplementary Tables S1–S3). Results demonstrated that the Laosicheng samples formed a loose cluster, positioned between the Felini and Pantherini species (Supplementary Figure S2), without providing a definitive species identification. Although the morphological evidence suggested that the two samples belonged to the family Felidae, characterized by distinctive features (e.g., auditory bulla and mastoid process) (Supplementary Figure S3), the damaged state of the skulls presented challenges in specific species determination. Therefore, we attempted to recover endogenous aDNA sequences for a more accurate identification.

Previous studies have established the effectiveness of aDNA analysis in the taxonomic identification of ancient animals, with genetic data also allowing for a more comprehensive understanding of population genetic history (Zhu et al., 2022). Here, high-quality mitogenome sequences were successfully obtained for the Laosicheng samples (Supplementary Figures S4). Utilizing a dataset combining the two Laosicheng samples with 114 other individuals representing all genera of the family Felidae (Supplementary Table S4), pairwise genetic distance analysis indicated that the Laosicheng samples share a closer genetic affiliation with the *Panthera* genus than any other (F_{ST}<0.01) (Supplementary Figures S5, S6 and Table S5). Further genetic distance comparisons between the two Laosicheng samples and 79 individuals within the *Panthera* genus (including *P. pardus* (*n*=50), *P. leo* (*n*=5), *P. uncia* (*n*=15), *P. onca* (*n*=4), and *P. tigris* (*n*=5)) (Supplementary Figures S5, S6 and Table S5) placed the Laosicheng samples closest to

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Received: 03 November 2023; Accepted: 09 January 2024; Online: 10 January 2024

Foundation items: This work was supported by the National Natural Science Foundation of China (41925009, 32100488), National Key R&D Program of China (2022YFE0203800), Chinese Academy of Sciences (YSBR-019), and New Cornerstone Science Foundation through the XPLORER PRIZE

P. pardus (Supplementary Figures S5, S6 and Table S5), with both Laosicheng individuals being genetically closest to *P. p. japonensis* than any other subspecies (Figure 1B; Supplementary Figure S6 and Table S5). To verify species identification and gain further insights into the maternal genetic history of East Asian leopards, the same dataset was used to construct phylogenetic trees (Supplementary Figures S7, S8 and Table S4). The resulting Bayesian and maximum-likelihood trees were consistent with the genetic distance results. Based on these findings, we propose the classification of the two Laosicheng samples as leopards, specifically *P. p. japonensis*.

All *P. pardus* species individuals were categorized into two groups (Eurasia and Africa), revealing a deep bifurcation (~769 thousand years ago (ka), 95% CI: 518–1 000 ka) (Figure 1C) between the Eurasian and African mitochondrial lineages, similar to the findings of Paijmans et al. (2021) and supporting Africa as the most likely place of origin (Supplementary Figures S7–S9). The Asian population of *P. pardus* was divided into three main lineages, including the North group represented by *P. p. orientalis*, Central group represented by *P. p. japonensis*, and South group represented by *P. p. delacouri*. The Central group included one sample (GenBank accession number: MK043027) sourced from a zoo at Daejeon O-World amusement park in Korea that was taxonomically specified as *P. p. orientalis*, but was found to contain approximately 30% genetic admixture from the North-Chinese leopard (*P. p. japonensis*) (marked in Figure 1C with **P. p. orientalis*_Korea) (Kim et al., 2016). Additionally, another sample (GenBank accession number: MG932393) collected from Shanxi Province, China, was only identified as *P. pardus*, with no further subspecies information, but which clustered with *P. p. japonensis* in the phylogenetic tree (marked in Figure 1C with **P.p. japonensis*_Shanxi) (Ren et al., 2019).

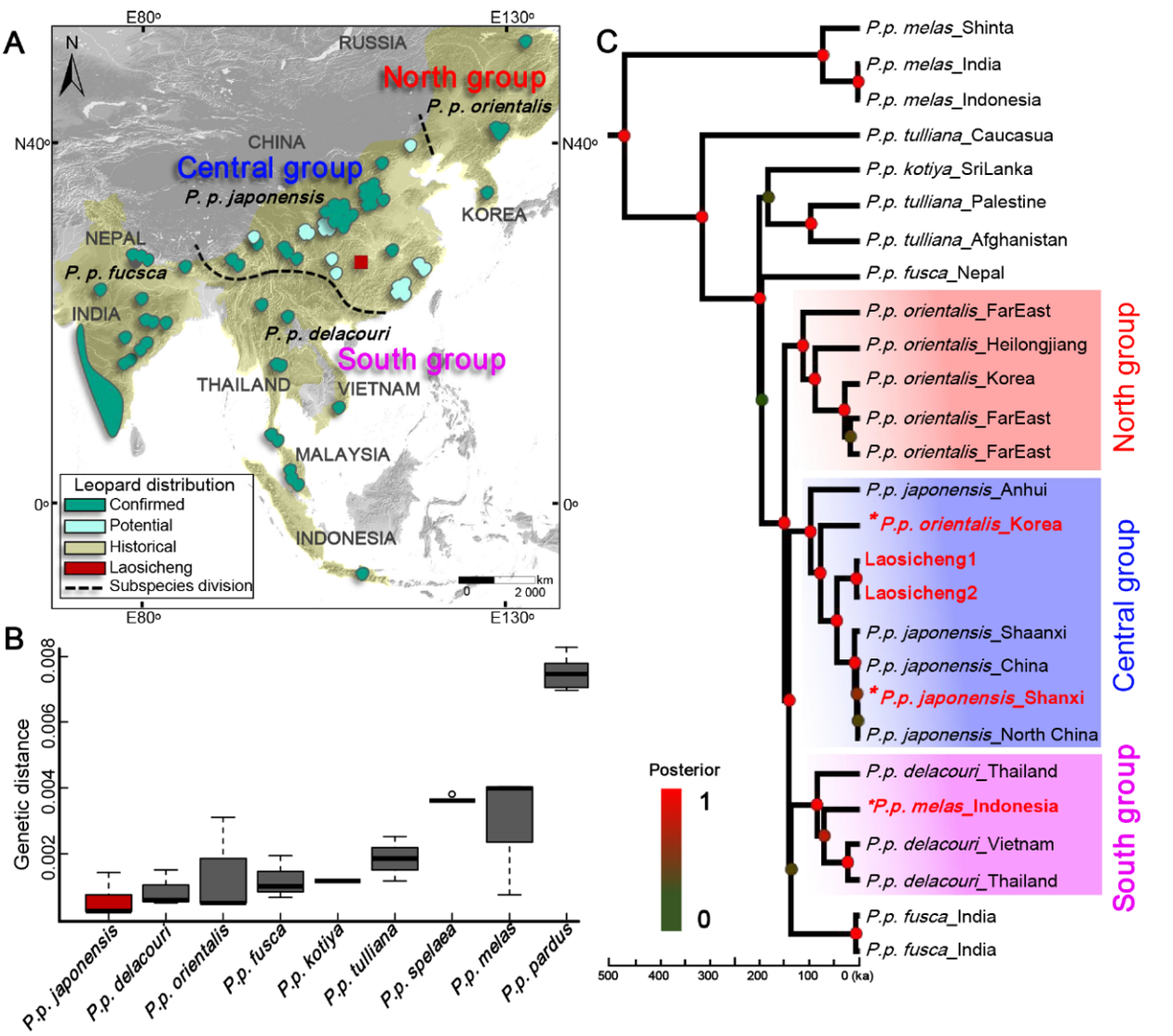


Figure 1 Morphological and genetic analysis of Laosicheng individuals (*Panthera pardus*)

A: Geographic distribution of three primary leopard populations in East Asia based on Laguardia et al. (2017), Kitchener et al. (2017), and this study. Dark green shaded area represents confirmed distribution of leopards, light green shaded area represents potential distribution area. Black dashed line represents boundary between three groups (North, Central, and South). Laosicheng is marked with red square. B: Boxplot of genetic distances across 50 individuals covering nine subspecies of *P. pardus*. C: Partial tree is based on selected individuals from global populations (full tree shown in Supplementary Figure S7). Laosicheng samples and three samples of interest are shown in red. Posterior values are marked with dots in varying colors, ranging from green to red, reflecting posterior support values from 0 to 1.

The South group included an Indonesian sample (GenBank accession number: MH588627) (marked in Figure 1C with *P. p. melas*_Indonesia) closely related to *P. p. delacouri*, which may be attributed to historical exchanges of live animals or acquisition of hunting trophies between the Southeast Asian islands and the mainland (Paijmans et al., 2018) (Figure 1A, C). The two Laosicheng samples exhibited the highest genetic correlation with each other and were grouped with individuals from the Central group (Figure 1C), aligning with previous results showing similarity with *P. p. japonensis*.

Three East Eurasian populations of leopard (*P. p. orientalis*, *P. p. japonensis*, and *P. p. delacouri*) exhibited a continuous distribution, with significant correlations between genetic patterns and geographical locations (Figure 1A, C; Supplementary Figures S7–S9 and Table S6). Given the lack of clear biogeographical barriers and minimal molecular differences between continental Asian leopards and Javan leopards (Wilting et al., 2016), the International Union for the Conservation of Nature (IUCN) Cat Specialist Group, comprising 22 experts from the Cat Classification Task Force, has proposed that *P. p. japonensis* be merged with *P. p. orientalis*, and possibly *P. p. delacouri* with *P. p. orientalis*, based on molecular evidence from only 724 bp concatenated mitochondrial sequences (Kitchener et al., 2017; Wilting et al., 2016). However, complete mitogenomes are significantly more informative and reliable than partial sequences (Zhu et al., 2022). Notably, based on complete mitogenomes, our phylogenetic trees distinctly separated *P. p. orientalis*, *P. p. japonensis*, and *P. p. delacouri* into North, Central, and South groups, respectively (posterior value=1) (Figure 1C; Supplementary Figures S7, S8). This finding was further corroborated by molecular variance analysis (AMOVA) (*F*ct=37.72%, *P*=0.100) and phylogenetic network analysis (Supplementary Figure S9 and Table S6). Divergence of these three clades was estimated to have occurred approximately 155 ka (95% CI: 102–215 ka) (Figure 1C). Previous microsatellite analysis has also indicated significant genetic divergence between *P. p. orientalis* and *P. p. japonensis* (Wang, 2019). Thus, despite the absence of a clear biogeographical barrier, we conclude that they can be accurately considered as three distinct (sub)populations.

The two Laosicheng samples clustered with individuals from the Central group, consistent with their place of origin. More than 19 leopards have been found at the Laosicheng site (including the two identified in this study) (Hunan Provincial Institute of Cultural Relics and Archaeology, 2018), indicating that approximately 300–600 years ago, there was a relatively significant leopard population in the surrounding area. This phenomenon differs greatly from the current distribution of leopards in central China (Laguardia et al., 2017). Bayesian skyline plot analysis indicated that the effective population size of leopards has consistently declined over the past 30 000 years (Supplementary Figure S10). However, we speculate that intensified human activities over the past few centuries have greatly contributed to the decline of leopard populations. As concluded in previous research, major threats to leopard populations include human-wildlife conflict, prey decrease, habitat loss and degradation, and hunting/poaching (Laguardia et al., 2017).

In summary, our study resolved the controversial identification of mammalian remains from the Laosicheng site, determining that the two skulls in question should be recognized as leopards (*P. p. japonensis*). This finding not only clarifies the species identification but also provides novel insights into the maternal genetic history of East Asian leopards, markedly diverging from the existing genetic framework. The skull measurements obtained from the two Laosicheng specimens acquired in this study have bridged a gap in the digital morphological model and genetic data of ancient leopards in China, exemplifying the successful integration of morphological and molecular data for the accurate identification of mammalian remains from archaeological sites. Animal remains are essential for regional zoological research, and for the validation of taxonomic and antiquity attributions. Our research illustrates the efficacy of genetic methodologies in resolving identification challenges associated with mammalian remains from archaeological sites and highlights the potency of these approaches in elucidating the evolutionary history of animals.

DATA AVAILABILITY

The data reported in this paper were deposited in GenBase of the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences/China National Center for Bioinformation under accession numbers C_AA055152.1 and C_AA055153.1 (https://ngdc.cncb.ac.cn/genbase), in the NCBI GenBank database under accession numbers OR871436 and OR871437, and in the Science Data Bank under https://doi.org/10.57760/sciencedb.14194.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

M.Z., Q.M.F., and L.H.M. designed the study. L.H.M. conducted the archaeological excavations. M.Z. collected and drilled samples. M.Z., Y.X.Z., Y.M.H., and Q.G.J.Z. performed CT scanning, CT data collection, and morphological character discussion. M.Z., Q.Y.D., R.W.Y., F.L., X.T.F., and P.C. performed the DNA experiments. M.Z., C.H.W., and Y.X.Z. performed data processing and analysis and wrote the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank E. Andrew Bennett (Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences) for reviewing and revising our manuscript. The manuscript also benefited greatly from comments by anonymous reviewers.

Ming Zhang[1,2,#], Cai-Hui Wang[1,#], Yu-Xin Zheng[3,4,#], Qi-Gao Jiangzuo[4], Ye-Mao Hou[4], Peng Cao[4], Qing-Yan Dai[4], Ruo-Wei Yang[4], Feng Liu[4], Xiao-Tian Feng[4], Lin-Heng Mo[5,*], Qiao-Mei Fu[4,2,6,*]

[1] *China-Central Asia "the Belt and Road" Joint Laboratory on Human and Environment Research*, *Key Laboratory of Cultural Heritage Research and Conservation*, *School of Culture Heritage*, *Northwest University*, *Xi'an*, *Shaanxi* 710127, *China*

[2] *Shanghai Qi Zhi Institute*, *Shanghai* 200232, *China*

[3] *College of Life Sciences*, *Northwest University*, *Xi'an*, *Shaanxi* 710127, *China*

[4] *Key Laboratory of Vertebrate Evolution and Human Origins*, *Institute of Vertebrate Paleontology and Paleoanthropology*, *Center for Excellence in Life and Paleoenvironment*, *Chinese Academy of Sciences*, *Beijing* 100044, *China*

[5] *Cultural Relics and Archaeology Institute of Hunan*, *Changsha*, *Hunan* 410083, *China*
[6] *University of the Chinese Academy of Sciences*, *Beijing* 100049, *China*
[#]Authors contributed equally to this work
*Corresponding authors, E-mail: 718481364@qq.com; fuqiaomei@ivpp.ac.cn

REFERENCES

Hunan Provincial Institute of Cultural Relics and Archaeology. 2018. Animal Remains Unearthed from Laosicheng, Yongshun. Beijing: Science Press, 1–512. (in Chinese)

Kim S, Cho YS, Kim HM, et al. 2016. Comparison of carnivore, omnivore, and herbivore mammalian genomes with a new leopard assembly. *Genome Biology*, **17**(1): 211.

Kitchener AC, Breitenmoser-Würsten C, Eizirik E, et al. 2017. A revised taxonomy of the Felidae. The final report of the Cat Classification Task Force of the IUCN/SSC Cat Specialist Group. Cat News Special Issue 11, IUCN, SSC, 80.

Laguardia A, Kamler JF, Li S, et al. 2017. The current distribution and status of leopards *Panthera pardus* in China. *Oryx*, **51**(1): 153−159.

Paijmans JLA, Barlow A, Becker MS, et al. 2021. African and Asian leopards are highly differentiated at the genomic level. *Current Biology*, **31**(9): 1872−1882.e5.

Paijmans JLA, Barlow A, Förster DW, et al. 2018. Historical biogeography of the leopard (*Panthera pardus*) and its extinct Eurasian populations. *BMC Evolutionary Biology*, **18**(1): 156.

Ren ZM, Niu XF, Lv T, et al. 2019. The complete mitochondrial genome of *Panthera pardus* (Felidae: Pantheriinae), a first-class national-protected wild animal from China. *Conservation Genetics Resources*, **11**(4): 389−392.

Wang M. 2019. Genetic diversity and genetic structure of Amur Leopard (*Panthera pardus orientalis*) and North China Leopard (*Panthera pardus fontanierii*) in China. Master thesis, Northeast Forestry University, Harbin. (in Chinese)

Wilting A, Patel R, Pfestorf H, et al. 2016. Evolutionary history and conservation significance of the Javan leopard *Panthera pardus melas*. *Journal of Zoology*, **299**(4): 239–250.

Zhu SL, Chen ZH, Hu SM, et al. 2022. Ancient DNA traces a Chinese 5400-year-old cat specimen as leopard cat (*Prionailurus bengalensis*). *Journal of Genetics and Genomics*, **49**(11): 1076−1079.

Lee & Lai. *Zool. Res.* 2024, 45(1): 230−232
https://doi.org/10.24272/j.issn.2095-8137.2023.415

ZR | Zoological Research

Obituary Open Access

Obituary: Prof. Yun Zhang (1963–2023) – A scientist focused on toxins and their underlying mechanisms to decipher human diseases

Wenhui Lee, Ren Lai

Prof. Yun Zhang was born on 9 July 1963 in Kunming, Yunnan, China, during a tumultuous period which he often referenced. Throughout his life, he harbored a steadfast belief in using knowledge to unravel the mysteries of human diseases. His educational journey was marked by frequent changes in schools due to his parents' occupational relocations. However, despite these challenges, he consistently displayed diligence and was admitted to the East China University of Science and Technology, Shanghai, after completing high school in 1980. He remained an active and loyal member of the School of Biotechnology at the university. Upon completing his bachelor's degree in 1984, Yun Zhang began his master's studies in biochemistry, focusing on snake toxin research, at the Kunming Institute of Zoology, Chinese Academy of Sciences (CAS) under the supervision of Prof. Yu-Liang Xiong in 1987. He later progressed to a dual doctoral program in science at the Kunming Institute of Zoology and Pasteur Institute at Paris, mentored by both Prof. Li-Ming Shi and Dr. Cassian Bon. Yun Zhang obtained his PhD in September 1992, followed by a tenure at the Pasteur Institute from December 1993 to September 1995. Subsequently, he conducted postdoctoral research at the Pasteur Institute at unite de venom, under the direction of Dr. Cassian Bon. In February 1998, Dr. Yun Zhang joined the Kunming Institute of Zoology, CAS, as a researcher, a role that proved pivotal in his career. He dedicated himself to the study of venoms, with a particular focus on snake venom and active proteins from the amphibian species *Bombina maxima*. His PhD thesis, which primarily involved the isolation, characterization, and expression of a plasminogen activator termed TSV-PA — the first plasminogen activator isolated from snake venom (Zhang et al., 1995) — earned the Special Award of the Chinese Academy of Sciences Dean's Reward. Later, the crystal structure of the first snake venom serine protease, represented by TSV-PA, was successfully elucidated (Parry et al., 1998) and Dr. Yun Zhang identified that the unique Phe 193 feature in TSV-PA is crucial for its high substrate specificity, rendering it incapable of forming stable complexes with bovine pancreatic trypsin inhibitors and other extended substrates and inhibitors (Zhang et al., 1997). The study of TSV-PA significantly advanced our understanding of the structure-function relationships of human plasminogen activators, specifically tissue-type (tPA) and urokinase-type (uPA), and provided an excellent template for modeling other homologous family members. Dr. Yun Zhang and his research team also discovered a host-derived, pore-forming toxin-like protein and trefoil factor complex, termed βγ-CAT (Gao et al., 2011; Liu et al., 2008), which plays a crucial role in protecting hosts against microbial infection (Xiang et al., 2014) and tissue injury (Gao et al., 2019), mediated via activation of inflammasomes. Subsequent research indicated that βγ-CAT also neutralizes the acidification of endocytic organelles to counter intracellular pathogen invasion (Li et al., 2017), acts directly on the herpes simplex virus type 1 (HSV-1) envelope to form a channel in the viral membrane to block infection (Liu et al., 2021), and augments antigen presentation by modulating cellular endolysosomes (Deng et al., 2020). Their original discovery helped clarify the potential involvement and mechanisms of aerolysin-like proteins and/or trefoil factors, widely present in vertebrates, in defending hosts against pathogens. Additionally, Dr. Yun Zhang and his research team elucidated the role of βγ-CAT in stress response and nutrient acquisition (Shi et al., 2022), with the complex shown to drive the host to scavenge extracellular proteins for nutrient supply and survival (Liu et al., 2023) and to acquire and maintain water via macropinocytosis induction and exocytosis modulation (Zhao et al., 2022). He also led the team in systematically exploring the mechanistic actions of βγ-CAT, including identification of its receptors and regulatory mechanisms (Bian et al., 2023; Guo et al., 2019; Wang et al., 2020). As a consequence of his important contributions to the field, Prof. Yun Zhang served as the director of the Academic Committee of the Kunming Institute of Zoology; deputy

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2024 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Drs. W Lee and R Lai are from the Key Laboratory of Bioactive Peptides of Yunnan Province, Engineering Laboratory of Peptides of Chinese Academy of Sciences, KIZ-CUHK Joint Laboratory of Bioresources and Molecular Research in Common Diseases, National Resource Centre for Non-Human Primates, Kunming Primate Research Centre, National Research Facility for Phenotypic & Genetic Analysis of Model Animals (Primate Facility), Sino-African Joint Research Centre, New Cornerstone Science Institute, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan 650107, China

E-mail: leewh@mail.kiz.ac.cn; rlai@mail.kiz.ac.cn.

director of the Academic Committee of the Key Laboratory of Animal Models and Human Disease Mechanisms of the Chinese Academy of Sciences; council member of the Asia-Pacific region of the International Society on Toxinology; executive director of the Chinese Society of Toxicology; director of the Bio-toxins Professional Committee; editorial board member of the international journal Toxicon, associate editor of the international journal *Zoological Research* (*ZR*); and a drug evaluation expert for the National Medical Products Administration.

Throughout his life, Prof. Yun Zhang exemplified dedication, meticulousness, innovation, and a quest for truth in his scholarly endeavors. He consistently engaged in cutting-edge scientific research, focusing on the molecular biodiversity, structure, and function of animal toxin proteins and peptides, and significantly advanced our understanding of biological adaptation mechanisms, functional genomic evolutionary characteristics, and their molecular mechanisms. His contributions were pivotal in enhancing the scientific research capabilities of the institute. Prof. Yun Zhang dedicated his life to cultivating virtue and nurturing talent. He was a mentor and guide, fostering the younger generation and selflessly dedicating himself to the training of over 70 outstanding graduate students. His loyalty to the party and unwavering commitment to his profession were evident throughout his career. Even during his illness and treatment, he remained concerned about the development of the institute, continued to mentor his students, and stayed true to his original aspirations, striving forward until his final moments.

Scientific journals play a crucial role in disseminating knowledge, facilitating scientific exploration, and nurturing talent (Yao & Zhang, 2022). Owing to his exceptional academic expertise and great enthusiasm, Prof. Yun Zhang made enduring and generous contributions to the evolution of *ZR*. He served as an editorial board member from 1998–2014 and as an associate editor-in-chief from 2014 until his final days. Despite facing serious health challenges, he continued to fulfill his duties with utmost dedication. His significant contributions to enhancing the influence of *ZR* encompassed developing journal policies, offering insightful feedback for manuscript revisions, and authoring influential articles for *ZR*, among other pivotal roles. His notable publications in *ZR* include, but are not limited to, the following: (1) Comprehensive review papers summarizing current knowledge on venoms and their biological implications (Zhang, 2015) and on animal secretory endolysosome channel discovery (Zhang et al., 2021), with both featuring as highly cited papers in *ZR*. (2) Together with his colleagues, he found that amphibian pore-forming protein βγ-CAT drives metabolite release from small extracellular vesicles through channel formation (Wang et al., 2023), characterized protein composition of extracellular vesicles from skin secretions of the amphibian *Bombina maxima* (Wei et al., 2022), and performed a comparative analysis of diverse toxins from a new pharmaceutical centipede, *Scolopendra mojiangica* (Liu et al., 2020); (3) He also identified king cobra peptide OH-CATH30 as a potential candidate drug through clinical drug-resistant isolates (Zhao et al., 2018). Prof. Yun Zhang's positive and far-reaching impact on *ZR* will always be remembered and valued.

Despite his passing, the legacy of his spirit and demeanor persists. His relentless pursuit of truth, rigorous academic approach, dedication to selfless research, and commitment to nurturing and supporting future generations remain enduringly worthy of our emulation and remembrance. Prof. Yun Zhang will forever reside in our hearts.



REFERENCES

Bian XL, Si ZR, Wang QQ, et al. 2023. IgG Fc-binding protein positively regulates the assembly of pore-forming protein complex βγ-CAT evolved to drive cell vesicular delivery and transport. *Journal of Biological Chemistry*, **299**(6): 104717.

Deng CJ, Liu L, Liu LZ, et al. 2020. A secreted pore-forming protein modulates cellular endolysosomes to augment antigen presentation. *The FASEB Journal*, **34**(10): 13609−13625.

Gao Q, Xiang Y, Zeng L, et al. 2011. Characterization of the βγ-crystallin domains of βγ-CAT, a non-lens βγ-crystallin and trefoil factor complex, from the skin of the toad *Bombina maxima*. *Biochimie*, **93**(10): 1865–1872.

Gao ZH, Deng CJ, Xie YY, et al. 2019. Pore-forming toxin-like protein complex expressed by frog promotes tissue repair. *The FASEB Journal*, **33**(1): 782−795.

Guo XL, Liu LZ, Wang QQ, et al. 2019. Endogenous pore-forming protein complex targets acidic glycosphingolipids in lipid rafts to initiate endolysosome regulation. *Communications Biology*, **2**: 59.

Li SA, Liu L, Guo XL, et al. 2017. Host pore-forming protein complex neutralizes the acidification of endocytic organelles to counteract intracellular pathogens. *The Journal of Infectious Diseases*, **215**(11): 1753−1763.

Liu L, Deng CJ, Duan YL, et al. 2021. An aerolysin-like pore-forming protein complex targets viral envelope to inactivate herpes simplex virus type 1. *The Journal of Immunology*, **207**(3): 888−901.

Liu LZ, Liu L, Shi ZH, et al. 2023. Amphibian pore-forming protein βγ-CAT drives extracellular nutrient scavenging under cell nutrient deficiency. *iScience*, **26**(5): 106598.

Liu SB, He YY, Zhang Y, et al. 2008. A novel non-lens *βγ*-crystallin and trefoil factor complex from amphibian skin and its functional implications. *PLoS One*, **3**(3): e1770.

Liu ZC, Liang JY, Lan XQ, et al. 2020. Comparative analysis of diverse toxins from a new pharmaceutical centipede *Scolopendra mojiangica*. *Zoological Research*, **41**(2): 138−147.

Parry MA, Jacob U, Huber R, et al. 1998. The crystal structure of the novel snake venom plasminogen activator TSV-PA: a prototype structure for

snake venom serine proteinases. *Structure*, **6**(9): 1195−1206.

Shi ZH, Zhao Z, Liu LZ, et al. 2022. Pore-forming protein βγ-CAT promptly responses to fasting with capacity to deliver macromolecular nutrients. *The FASEB Journal*, **36**(10): e22533.

Wang QQ, Bian XL, Zeng L, et al. 2020. A cellular endolysosome-modulating pore-forming protein from a toad is negatively regulated by its paralog under oxidizing conditions. *Journal of Biological Chemistry*, **295**(30): 10293−10306.

Wang QQ, Lan XQ, Wei XS, et al. 2023. Amphibian pore-forming protein βγ-CAT drives metabolite release from small extracellular vesicles through channel formation. *Zoological Research*, **44**(4): 739−742.

Wei XS, Liu LZ, Bian XL, et al. 2022. Protein composition of extracellular vesicles from skin secretions of the amphibian *Bombina maxima*. *Zoological Research*, **43**(4): 687–690.

Xiang Y, Yan C, Guo XL, et al. 2014. Host-derived, pore-forming toxin–like protein and trefoil factor complex protects the host against microbial infection. *Proceedings of the National Academy of Sciences of the United States of America*, **111**(18): 6702−6707.

Yao YG, Zhang Y. 2022. *Zoological Research* shines in the East. *Zoological Research*, **43**(1): 1−2.

Zhang Y. 2015. Why do we study animal toxins?. *Zoological Research*, **36**(4): 183−222.

Zhang Y, Wang QQ, Zhao Z, et al. 2021. Animal secretory endolysosome channel discovery. *Zoological Research*, **42**(2): 141−152.

Zhang Y, Wisner A, Maroun RC, et al. 1997. *Trimeresurus stejnegeri* snake venom plasminogen activator: site-directed mutagenesis and molecular modeling. *Journal of Biological Chemistry*, **272**(33): 20531−20537.

Zhang Y, Wisner A, Xiong YL, et al. 1995. A novel plasminogen activator from snake venom: purification, characterization, and molecular cloning. *Journal of Biological Chemistry*, **270**(17): 10246−10255.

Zhao F, Lan XQ, Du Y, et al. 2018. King cobra peptide OH-CATH30 as a potential candidate drug through clinic drug-resistant isolates. *Zoological Research*, **39**(2): 87−96.

Zhao Z, Shi ZH, Ye CJ, et al. 2022. A pore-forming protein drives macropinocytosis to facilitate toad water maintaining. *Communications Biology*, **5**(1): 730.

Zoological Research Editorial Board (2025)

EDITOR-IN-CHIEF:

Yong-Gang Yao — Kunming Institute of Zoology, CAS, China

ASSOCIATE EDITORS-IN-CHIEF:

Wai-Yee Chan — The Chinese University of Hong Kong, China
Xue-Long Jiang — Kunming Institute of Zoology, CAS, China
Yong-Tang Zheng — Kunming Institute of Zoology, CAS, China

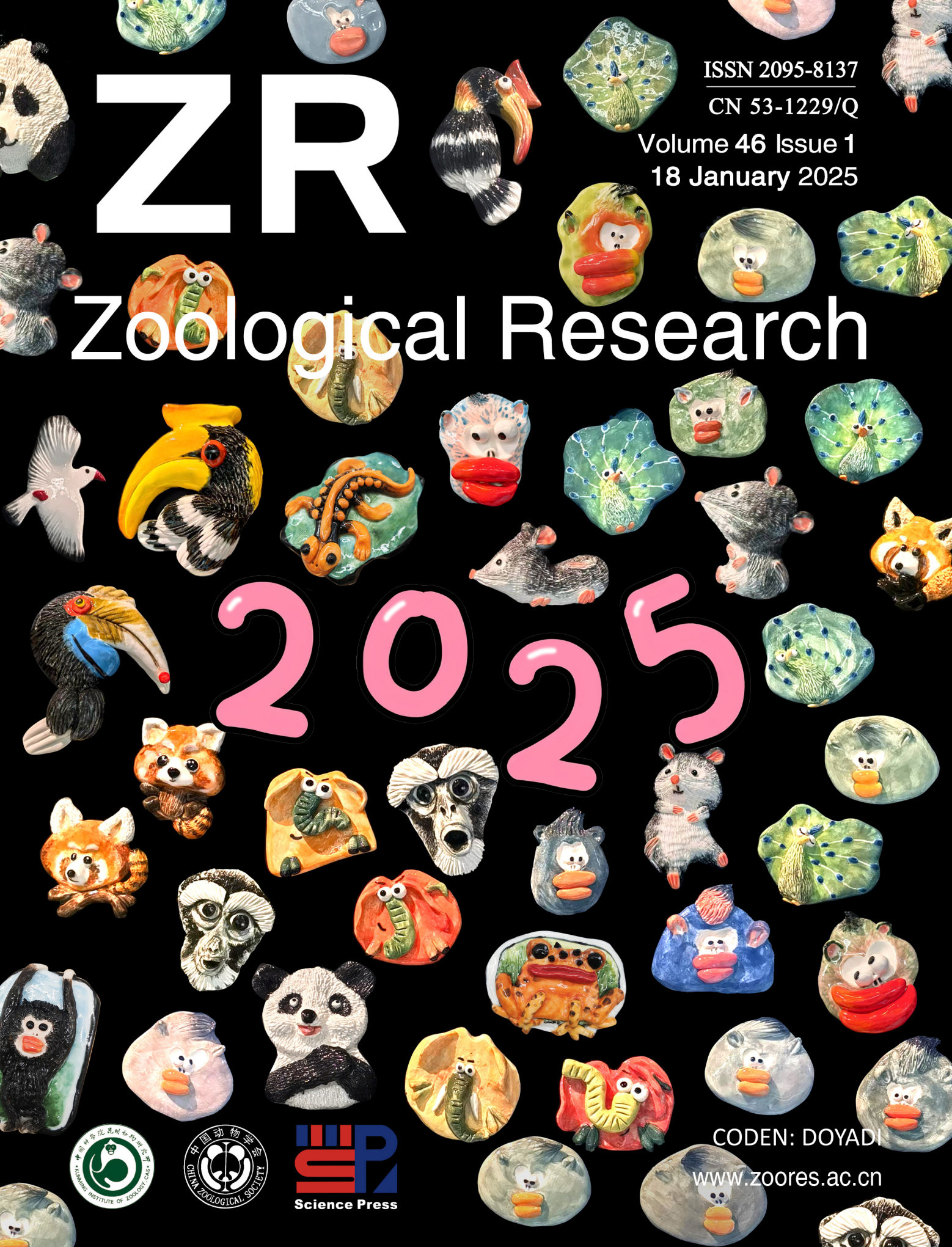
ASSOCIATE EDITORS:

Jiong Chen — Ningbo University, China
Peng-Fei Fan — Sun Yat-sen University, China
David Irwin — University of Toronto, Canada
Jia-Tang Li — Chengdu Institute of Biology, CAS, China
Shu-Qiang Li — Institute of Zoology, CAS, China

MEMBERS:

Amir Ardeshir — University of California, Davis, USA
Yong-Li Bao — Northeast Normal University, China
Yu-Hai Bi — Institute of Microbiology, CAS, China
Chao Bian — Shenzhen University, China
Jean Philippe Boubli — University of Salford, UK
Wen-Jun Bu — Nankai University, China
Kevin L. Campbell — University of Manitoba, Canada
Gang Cao — Huazhong Agricultural University, China
Colin Chapman — McGill University, Canada
Ce-Shi Chen — Kunming Medical University, China
Jia-Yu Chen — Tongji University, China
Liang-Biao Chen — Shanghai Ocean University, China
Luo-Nan Chen — Center for Excellence in Molecular Cell Science, CAS, China
Xiang-Jun Chen — Zhejiang University, China
Xue-Xin Chen — Zhejiang University, China
Yong-Chang Chen — Kunming University of Science and Technology, China
Dao-Jun Cheng — Southwest University, China
Ying Cheng — Yunnan University, China
Sandra Chiu — University of Science and Technology of China, China
Yu-Peng Cun — Chongqing Medical University, China
Ji Dai — Shenzhen Institute of Advanced Technology, CAS, China
Zhi-Fang Dong — Chongqing Medical University, China
Jiu-Lin Du — Center for Excellence in Brain Science and Intelligence Technology, CAS, China
Li-Na Du — Guangxi Normal University, China
Zhen-Yu Du — East China Normal University, China
Miguel Angel Esteban — BGI-Shenzhen & Guangzhou Institutes of Biomedicine and Health, CAS, China
Guang-Yi Fan — BGI-Shenzhen, China
He Zhi Fang — Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, China
Ji-Feng Fei — Southern Medical University, China
Ze-Xia Gao — Huazhong Agricultural University, China
Chu-Tian Ge — Zhejiang Wanli University, China
De-Yan Ge — Institute of Zoology, CAS, China
Shao-Yu Ge — SUNY at Stony Brook, USA
Cyril C. Grueter — The University of Western Australia, Australia
Jian-Fang Gui — Institute of Hydrobiology, CAS, China
Chang-Jun Guo — Sun Yat-sen University, China
Xian-Guang Guo — Chengdu Institute of Biology, CAS, China
Hong-Xuan He — Institute of Zoology, CAS, China
Shun-Ping He — Institute of Hydrobiology, CAS, China
Yong-Han He — Kunming Institute of Zoology, CAS, China
David Hillis — University of Texas at Austin, USA
Chun-Hui Hou — Kunming Institute of Zoology, CAS, China
Ji Hu — ShanghaiTech University, China
Zhong-Hua Hu — Xiangya Hospital, Central South University, China
Xiao-Xiang Hu — China Agricultural University, China
Jin-Lian Hua — Northwest A&F University, China
Lu-Sheng Huang — Jiangxi Agricultural University, China
You-Hua Huang — South China Agricultural University, China
Andrew Hutchins — Southern University of Science and Technology, China
William R. Jeffery — University of Maryland, USA
Xiang Ji — Wenzhou University, China
Jian-Ping Jiang — Chengdu Institute of Biology, CAS, China
Yu Jiang — Northwest A&F University, China
Le Kang — Hebei University, China
Yun Hak Kim — Pusan National University, Republic of Korea
Esther N. Kioko — National Museums of Kenya, Kenya
Randall C. Kyes — University of Washington, USA
Liang-Xue Lai — Guangzhou Institutes of Biomedicine and Health, CAS, China
Ren Lai — Kunming Institute of Zoology, CAS, China
Xiao-Jian Lai — Jimei University, China
Tian-Ming Lan — Northeast Forestry University, China
Kai Lei — Westlake University, China
Peng Lei — Sichuan University, China
Cheng-Hua Li — Ningbo University, China
Da-Li Li — East China Normal University, China
Dai-Qin Li — National University of Singapore, Singapore
Di-Yan Li — Chengdu University, China
Dong-Ming Li — Hebei Normal University, China
Gang Li — Shaanxi Normal University, China
Guo-Gang Li — Qinghai Normal University, China
Hai-Peng Li — Shanghai Institute of Nutrition and Health, CAS, China
Hong Li — Nanjing Normal University, China
Jin-Song Li — Center for Excellence in Molecular Cell Science, CAS, China
Jing Li — Sichuan University, China
Lei Li — Shenzhen Institute of Advanced Technology, CAS, China
Ming Li — Institute of Zoology, CAS, China
Ming Li — Kunming Institute of Zoology, CAS, China
Ming Li — University of Nebraska-Lincoln, USA
Ming-Zhou Li — Sichuan Agricultural University, China
Wei Li — Guangzhou Medical University, China
Wei-Dong Li — Shanghai Jiao Tong University, China
Xiao-Jiang Li — Jinan University, China
Xue-Ling Li — Inner Mongolia University, China
Xue-Yan Li — Kunming Institute of Zoology, CAS, China
Zhan-Jun Li — Jilin University, China
Wei Liang — Hainan Normal University, China
Qiang Lin — South China Sea Institute of Oceanology, CAS, China
Wei-Qiang Lin — Zhejiang University, China
Bing Liu — Fifth Medical Center of Chinese PLA General Hospital, China
Chun-Yu Liu — Shanghai Jiao Tong University, China
Ci-Rong Liu — Center for Excellence in Brain Science and Intelligence Technology, CAS, China
Huan-Zhang Liu — Institute of Hydrobiology, CAS, China
Jian-Hua Liu — South China Agricultural University, China
Lei Liu — Ningbo University, China
Sheng Liu — Sun Yat-sen University, China

Xing-Yue Liu	China Agricultural University, China
Zhen Liu	Institute of Neuroscience, CAS, China
Zhen Liu	Kunming Institute of Zoology, CAS, China
Jia-Hong Lu	University of Macau, China
Li Lu	Kunming Institute of Zoology, CAS, China
Anders Pape Møller	Université Paris-Saclay, France
Li Ma	Kunming Institute of Zoology, CAS, China
Ya-Fei Mao	Shanghai Jiao Tong University, China
Zi-Ning Meng	Sun Yat-sen University, China
Changjong Moon	Chonnam National University, Republic of Korea
Masaharu Motokawa	Kyoto University Museum, Japan
Yong-Gang Nie	Institute of Zoology, CAS, China
Sang-Bing Ong	The Chinese University of Hong Kong, China
Hong-Bo Pan	Shanghai Ocean University, China
Hyun Park	Korea University, Republic of Korea
Julian Kerbis Peterhans	Roosevelt University, USA
Nikolay A. Poyarkov, jr.	Lomonosov Moscow State University, Russia
Xiao-Guang Qi	Northwest University, China
Jian-Wen Qiu	Hong Kong Baptist University, China
Qiang Qiu	Northwestern Polytechnical University, China
Jing Qu	Institute of Zoology, CAS, China
Ming-Qiang Rong	Hunan Normal University, China
Christian Roos	Leibniz-Institute for Primate Research, Germany
Chang-Wei Shao	Yellow Sea Fisheries Research Institute, CAFS, China
De-Li Shi	Sorbonne University, France
Qing-Hua Shi	University of Science and Technology of China, China
Qiong Shi	Shenzhen University, China
Wei-Feng Shi	Shanghai Jiao Tong University, China
Xing-Feng Si	East China Normal University, China
Ning Song	Qingdao University, China
Bing Su	Kunming Institute of Zoology, CAS, China
Jian-Guo Su	Huazhong Agricultural University, China
Qiang Sun	Center for Excellence in Brain Science and Intelligence Technology, CAS, China
Yong-Hua Sun	Institute of Hydrobiology, CAS, China
Yu-Hua Sun	Institute of Hydrobiology, CAS, China
Chatmongkon Suwannapoom	University of Phayao, Thailand
Yi-Gang Tong	Beijing University of Chemical Technology, China
Christoph W. Turck	Max Planck Institute of Psychiatry, Germany
Guo-Hao Wang	National Institutes of Health, USA
Ding Wang	Institute of Hydrobiology, CAS, China
Kai Wang	Kunming Institute of Zoology, CAS, China
Kun Wang	Northwestern Polytechnical University, China
Ming-Shan Wang	Kunming Institute of Zoology, CAS, China
Wen Wang	Northwestern Polytechnical University, China
Xiao-Hui Wang	Changchun Institute of Applied Chemistry, CAS, China
Xiao-Qun Wang	Beijing Normal University, China
Yan-Jiang Wang	Third Military Medical University, China
Ying-Yong Wang	Sun Yat-sen University, China
Bin Wei	Shanghai University, China
Fu-Wen Wei	Jiangxi Agricultural University, China
Dong-Dong Wu	Kunming Institute of Zoology, CAS, China
Jun-Hong Xia	Sun Yat-sen University, China
Hui Xiang	South China Normal University, China
Zuo-Fu Xiang	Central South University of Forestry & Technology, China
Shi-Jun Xiao	Jilin Agricultural University, China
Xiao Xiao	Fudan University, China
Can Xie	Zhejiang University, China
Song-Guang Xie	Hainan University, China
Yu-Yun Xing	Jiangxi Agricultural University, China
Fu-Qiang Xu	Shenzhen Institute of Advanced Technology, CAS, China
Luo-Hao Xu	Southwest University, China
Shi-Qing Xu	Soochow University, China
Qing-Pi Yan	Jimei University, China
Sen Yan	Jinan University, China
Da-Hai Yang	East China University Of Science and Technology, China
Guang Yang	Nanjing Normal University, China
Jia-Long Yang	East China Normal University, China
Lian-Dong Yang	Institute of Hydrobiology, CAS, China
Shi-Hua Yang	South China Agricultural University, China
Wei-Li Yang	Jinan University, China
Xin-Wang Yang	Kunming Medical University, China
Xing-Lou Yang	Kunming Institute of Zoology, CAS, China
Yun-Peng Yang	Yangzhou University, China
Zhi-Yuan Yao	Shenyang Normal University, China
Bing Ye	University of Michigan, USA
Wen-Hua Yu	Guangzhou University, China
Ti-Fei Yuan	Shanghai Mental Health Center, China
Vitaliy Zablotskyy	Institute of Physics of the Academy of Sciences, the Czech Republic
Jian-Xiong Zeng	Shanghai Jiao Tong University, China
Bao-Wei Zhang	Anhui University, China
Bo Zhang	Southern Ocean Science and Engineering Guangdong Laboratory-Zhanjiang, China
Chi-Yu Zhang	Shanghai Public Health Clinical Center, Fudan University, China
Guo-Jie Zhang	Zhejiang University, China
Jing-Song Zhang	Naval Medical University, China
Shi-Chang Zhang	Hubei University, China
Xin Zhang	Hefei Institutes of Physical Science, CAS, China
Ya-Ping Zhang	Kunming Institute of Zoology, CAS, China
Yan Zhang	Peking University, China
Yun Zhang	Nanjing Drum Tower Hospital, Nanjing University, China
Yun-Wu Zhang	Xiamen University, China
Zhi Zhang	University of Science and Technology of China, China
Zhi-Jun Zhang	Southeast University, China
Bo Zhao	Kunming Institute of Zoology, CAS, China
Hua-Bin Zhao	Wuhan University, China
Hui Zhao	The Chinese University of Hong Kong, China
Jian-Guo Zhao	Institute of Zoology, CAS, China
Qian-Jun Zhao	Institute of Animal Science, CAAS, China
Tian Zhao	Southwest University, China
Wen-Ming Zhao	Beijing Institute of Genomics, CAS, China
Ya-Hui Zhao	Institute of Zoology, CAS, China
Ping Zheng	Fudan University, China
Gui-Sheng Zhong	ShanghaiTech University, China
Jiang Zhou	Guizhou Normal University, China
Peng Zhou	Guangzhou National Laboratory, China
Qi Zhou	Zhejiang University, China
Wu Zhou	The University of Mississippi, USA
Xin Zhou	China Agricultural University, China
Xu-Ming Zhou	Institute of Zoology, CAS, China
Yang Zhou	Huazhong Agricultural University, China
Xian-Jun Zhu	Sichuan Provincial People's Hospital, China
Kang Zou	Nanjing Agricultural University, China



ISSN 2095-8137
CN 53-1229/Q
Volume 46 Issue 1
18 January 2025

KUNMING INSTITUTE OF ZOOLOGY, CAS
CHINA ZOOLOGICAL SOCIETY
Science Press

CODEN: DOYADI
www.zoores.ac.cn

ZOOLOGICAL RESEARCH
动物学研究
Bimonthly, Since 1980
Volume 46 Issue 1, 18 January, 2025

中国最具国际影响力期刊
中国科技核心期刊
中国百强报刊

Editor-in-Chief: Yong-Gang Yao

Executive Editor-in-Chief: Xue-Long Jiang

Editors: Su-Qing Liu Long Nie Rong-Fu Wang

Edited by Editorial Office of Zoological Research

(Kunming Institute of Zoology, Chinese Academy of Sciences, 17 Longxin Road, Kunming, Yunnan, Post Code: 650201 Tel: +86 871 65199026 E-mail: zoores@mail.kiz.ac.cn)

Sponsored by Kunming Institute of Zoology, Chinese Academy of Sciences; China Zoological Society

Supervised by Chinese Academy of Sciences

Published by Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences©

Printed by Kunming Xiaosong Plate Making & Printing Co, Ltd

Domestic distribution by Yunnan Post and all local post offices in China

International distribution by China International Book Trading Corporation (Guoji Shudian) P.O.BOX 399, Beijing 100044, China

Overseas Code: BM358

Domestic Postal Issue No.: 64-20

Price: 25.0 USD/150.0 CNY

ISSN 2095-8137

CN 53-1229/Q

ISSN 2095-8137