Sun et al. *Zool. Res.* 2025, 46(3): 505−517
https://doi.org/10.24272/j.issn.2095-8137.2025.094
https://cstr.cn/32248.14.zr.2025.094

ZR Zoological Research

Article Open Access

Convergent musk biosynthesis across host and microbiota in musk deer and muskrat

Yi-Shan Sun[1,#], Lei Zhao[1,#], Cheng-Li Zheng[2,#], Xiao-Ting Yan[1], Ye Li[1], Xue-Li Gao[1], Ting-Feng Xue[1], Yi-Ming Zhang[1], Zhi-Peng Li[3], Rasmus Heller[4], Chen-Guang Feng[1], Chao Xu[3,*], Kun Wang[1,*], Qiang Qiu[1,*]

[1] *Shaanxi Key Laboratory of Qinling Ecological Intelligent Monitoring and Protection*, *School of Ecology and Environment*, *Northwestern Polytechnical University*, *Xi'an*, *Shaanxi* 710072, *China*

[2] *Sichuan Institute of Musk Deer Breeding*, *Sichuan Institute for Drug Control*, *Chengdu*, *Sichuan* 611845, *China*

[3] *College of Animal Science and Technology*, *Jilin Agricultural University*, *Changchun*, *Jilin* 130118, *China*

[4] *Section for Computational and RNA Biology*, *Department of Biology*, *University of Copenhagen*, *Copenhagen DK*-2200, *Denmark*

ABSTRACT

Mammalian scent glands mediate species-specific chemical communication, yet the mechanistic basis for convergent musk production remain incompletely understood. Forest musk deer and muskrat have independently evolved specialized musk-secreting glands, representing a striking case of convergent evolution. Through an integrated multi-omics approach, this study identified cyclopentadecanone as a shared key metabolic precursor in musk from both forest musk deer and muskrat, although downstream metabolite profiles diverged between the two lineages. Single-cell RNA sequencing revealed that these specialized apocrine glands possessed unique secretory architecture and exhibited transcriptional profiles associated with periodic musk production, distinct from those in conventional apocrine glands. Convergent features were evident at the cellular level, where acinar, ductal, and basal epithelial subtypes showed parallel molecular signatures across both taxa. Notably, acinar cells in both species expressed common genes involved in fatty acid and glycerolipid metabolism (e.g., *ACSBG1, HSD17B12*, *HACD2*, and *HADHA*), suggesting a conserved molecular framework for musk precursor biosynthesis. Metagenomic analysis of musk samples further revealed parallel microbial community structures dominated by *Corynebacterium* and enriched in lipid metabolic pathways. These findings suggest multi-level convergence in musk biosynthesis, from molecular pathways to microbial communities, providing novel insights into mammalian chemical signaling and artificial musk production.

Keywords: Convergence; Musk biosynthesis; Musk gland; Microbiota; Multi-omics approach

INTRODUCTION

Chemical communication mediated by specialized scent glands serves as an important mechanism for social and sexual signaling in various mammalian species (Brennan & Zufall, 2006; Buchinger & Li, 2023). These exocrine structures exhibit considerable anatomical and functional diversity, including preorbital glands (Bartoš et al., 2005), anogenital glands (Zhou et al., 2021; Zhou et al., 2019), and subcaudal pouches (Theis et al., 2013). Despite this diversity, various phylogenetically distant lineages have independently evolved morphologically distinct but functionally analogous chemical signaling organs. Notably, forest musk deer (*Moschus berezovskii*) and muskrat (*Ondatra zibethicus*) have each developed specialized musk-secreting glands capable of producing complex semiochemicals with comparable olfactory characteristics. Musk derived from forest musk deer has long been utilized in perfumery and pharmaceutical applications (Ahmed et al., 2018; Lv et al., 2022; McGinty et al., 2011; Shirasu et al., 2014; Wang et al., 2020), while musk secreted from muskrat shares key volatile constituents and a similar scent profile (Shi et al., 2023; Sun et al., 2021).

In both taxa, the musk gland is a male-specific ventral structure located proximate to reproductive organs (Bi et al., 1993; Chen et al., 2007), and its activity is tightly regulated by the reproductive cycle (Fan et al., 2018; Jiang et al., 2022). Studies on the musk gland in forest musk deer have identified several key genes associated with steroidogenesis and lipid metabolic pathways (Feng et al., 2023; Yang et al., 2021), alongside two molecularly distinct acinar cell populations contributing to musk biosynthesis (Liu et al., 2023, 2024a,

---

Received: 02 March 2024; Accepted: 14 March 2025; Online: 15 March 2025

Foundation items: This work was supported by the National Natural Science Foundation of China (32225009, 32370452, 82274046) and Jilin Agricultural University High-level Talent Introduction Fund (202020218)

[#] Authors contributed equally to this work

*Corresponding authors, E-mail: xcjlau@163.com; wangkun@nwpu.edu.cn; qiuqiang@nwpu.edu.cn

2024b). Previous microbial profiling of forest musk deer musk using 16S rRNA sequencing identified bacterial taxa potentially involved in the biochemical maturation of musk compounds, including Firmicutes, Proteobacteria, and Actinobacteria phyla (Li et al., 2018; Xu et al., 2024). However, a systematic cross-species analysis of cellular architecture, transcriptional landscapes, and microbiota composition in the musk glands of forest musk deer and muskrat has not yet been conducted. This gap in comparative data hinders a deeper understanding of the molecular and microbial convergence underlying musk biosynthesis and limits evaluation of the degree of similarity between the chemical end products.

To address this, an integrated multi-omics approach, combining metabolomics, proteomics, single-nucleus RNA sequencing (snRNA-seq), genomics, and metagenomics, was employed. This integrative framework enabled the systematic characterization and high-resolution comparison of musk glands from forest musk deer and muskrat across molecular, cellular, and microbial dimensions. Our findings revealed the mechanisms behind convergent musk biosynthesis, highlighting coordinated host-microbiota interactions within musk glands and providing a valuable framework for advancing our understanding of mammalian chemical signal synthesis.

MATERIALS AND METHODS

Ethics statement and processing of samples

Adult forest musk deer and muskrats were obtained from the Sichuan Institute of Musk Deer Breeding (Dujiangyan and Maerkang, China) and the Shahezi Xinke Breeding Farm (Wuchang City, Heilongjiang Province, China), respectively. All experimental procedures were conducted in accordance with guidelines approved by the Ethics Committee Institutional Review Board of Northwestern Polytechnic University (approval number: 202401193).

For tissue collection, nine-year-old male forest musk deer in the active phase of musk secretion were sedated using Lu-Mian-Ning, a xylazine hydrochloride-based anesthetic (100 mg/mL; Huamu Animal Health Product Company, Jilin, China). A surgical incision was made adjacent to the musk gland orifice, permitting the extraction of approximately 50 mg of glandular tissue. Post-operative care included wound disinfection, meticulous suturing, and administration of anti-inflammatory agents. In parallel, systematic anatomical dissections were performed on adult male muskrats (n=3), enabling the collection of 12 distinct tissue types: musk gland, muscle, thyroid, adrenal gland, testis, spleen, heart, liver, lung, kidney, brain, and bone. All tissues were thoroughly rinsed with phosphate-buffered saline (PBS, Sigma-Aldrich, USA). Additionally, mature musk samples were obtained from three individuals per species during musk maturation period. All collected specimens were immediately transferred to cryogenic vials, snap-frozen in liquid nitrogen, and allocated for downstream multi-omics analyses.

Genome sequencing, assembly, and annotation

Genome sequencing was conducted exclusively for the muskrat, as a high-quality reference genome for musk deer was already available (Chen et al., 2022), whereas no suitable reference genome existed for muskrat at the time of analysis. Fresh muscle tissue was collected from adult muskrat and subjected to high-throughput sequencing for genome assembly. Genomic DNA was extracted using a DNA purification kit (Qiagen, China) and sequenced using both Illumina HiSeq2000 (USA) and Oxford Nanopore GridIOn platforms (ONT, UK), generating 300 Gb and 120 Gb of data, respectively. For Hi-C library construction, 300 Gb of data was obtained from the Illumina HiSeq X Ten platform (USA). The resulting muskrat genome assembly spanned 2.43 Gb across 27 chromosomes, with 99.68% of contigs successfully anchored. Assembly continuity was high, with contig and chromosome N50 values of 67.5 Mb and 90.6 Mb, respectively. The assembled genome exhibited excellent completeness, with over 99% of paired-end reads mapped across more than 99% of the genome and an average recovery of 96.1% of BUSCO orthologs from the mammalia_odb10 database (Supplementary Figure S1 and Table S1).

Genome size was estimated to be approximately 2.5 Gb based on k-mer frequency analysis using Jellyfish (v.2.2.10) (Marçais & Kingsford, 2011) and GenomeScope (Vurture et al., 2017). Long-read sequences from the Nanopore platform were self-corrected and assembled into contigs using NextDenovo (v.2.4.0; https://github.com/Nextomics/NextDenovo), followed by polishing with NextPolish (v.1.3.0) (Hu et al., 2020). Hi-C short reads were aligned to the contigs using Juicer (v.1.6) (Durand et al., 2016), and the contigs were anchored using 3D-DNA (v.180419) (Dudchenko et al., 2017). Manual refinement of the assembly was performed using Juicebox Assembly Tools (https://github.com/theaidenlab/juicebox). Assembly quality was further assessed using BUSCO (v.4.1.2) (Manni et al., 2021) against the mammalia_odb10 lineage database.

Protein-coding gene annotation was performed using a comprehensive strategy that combined *ab initio* predictions, homology-based inference, and transcriptome-guided evidence. *Ab initio* predictions were generated using Augustus (v.3.2.1) (Stanke & Waack, 2003), incorporating species-specific gene models. Homology-based predictions were derived from protein sequences of 11 species, including human (*Homo sapiens*; GRCh38), mouse (*Mus musculus*, GRCm39), Arctic ground squirrel (*Urocitellus parryii*, ASM342692v1), cow (*Bos taurus*, ARS-UCD1.3), American mink (*Neovison vison*; NNQGG.v01), pig (*Sus scrofa*; Sscrofa11.1), rat (*Rattus norvegicus*, mRatBN7.2), red fox (*Vulpes vulpes*; VulVul2.2), prairie vole (*Microtus ochrogaster*; MicOch1.0), alpine marmot (*Marmota marmota marmota*; marMar2.1), and forest musk deer (*Moschus berezovskii*; NCBI GCF_022376915.1). Transcriptome-based gene models were obtained from 19 RNA-seq datasets representing 11 tissues (musk gland, thyroid, adrenal gland, testis, spleen, heart, liver, lung, kidney, brain, and bone) extracted from three muskrats. RNA was sequenced using the Illumina HiSeq 2000 platform (USA), with raw reads filtered using Fastp (v.0.20) (Chen et al., 2018) and assembled without reference using SPAdes (v.3.15.3) (Bushmanova et al., 2019) under default parameters. Coding regions were predicted using TransDecoder (v.5.5.0; RRID:SCR_017647). These proteins and transcriptome-predicted protein sequences were aligned to the genome using default parameters in BLAT (v.35.1) (Kent, 2002). Final gene model refinement was conducted using GeneWise (v.2.4.1) (Birney et al., 2004), followed by manual curation to eliminate duplicate annotations.

The integrated gene set was translated into amino acid sequences using InterProScan (v.5.73) (Jones et al., 2014) to annotate motifs and domains in protein sequences by searching against publicly available databases (including Pfam, PRINTS, PANTHER, ProDom, and SMART). Functional classification of genes was further performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

Metabolomic profiling of musk secretions

For gas chromatography-mass spectrometry (GC-MS), musk samples (50±1 mg) were homogenized in 500 μL of methanol-chloroform solution (3:1, v/v) containing the internal standard adonitol. Homogenates were subjected to vortexing, sonication, and centrifugation (15 min at 12 000 r/min, 4℃) to isolate the supernatant. A pooled quality control (QC) sample was prepared by combining equal volumes from all individual extracts. The extracts were dried and derivatized with methoxyamine hydrochloride (TCI, Japan) and N,O-bis(trimethylsilyl)trifluoroacetamide (BSTFA) (containing 1% trimethylchlorosilane (TMCS) (REGIS Technologies, USA) at 80℃ and 70℃, respectively. Fatty acid methyl esters (FAMEs) were added to the QC sample to serve as retention index markers. GC-TOF-MS analysis was performed using an Agilent 7890 gas chromatograph system coupled with a Pegasus HT time-of-flight mass spectrometer in Novogene (China). Raw data analysis, including peak extraction, baseline adjustment, deconvolution, alignment, and integration, was conducted using Chroma TOF (v.4.3x; Laboratory Equipment Corporation (LECO), USA) (Kind et al., 2009). Metabolite identification was achieved through spectral and retention index matching against the LECO-Fiehn Rtx5 database. Relative metabolite abundance was calculated based on normalized peak areas, with each characteristic peak corresponding to a specific compound.

For liquid chromatography-mass spectrometry (LC-MS)-based lipidomics, lipids were extracted from musk samples (100 μL or 100 mg) using a modified Folch method with methanol, methyl tert-butyl ether (MTBE), and water. The organic phase was isolated, dried, and reconstituted in isopropanol for analysis. Lipid profiling was performed using a Vanquish UHPLC system coupled to an Orbitrap Q Exactive HF mass spectrometer (Thermo Fisher, Germany). Chromatographic separation was achieved on a Thermo Accucore C30 column (150 mm×2.1 mm, 2.6 μm; Germany) using a 20-min gradient of acetonitrile/water and acetonitrile/isopropanol buffers, both containing ammonium acetate and formic acid. The Q Exactive HF was operated in both positive and negative ionization modes. Key parameters included: spray voltage 3.5 kV, capillary temperature 320℃, heater temperature 350℃, scan range 114–1 700 m/z, and normalized collision energies of 22, 24, and 28 eV. Dynamic exclusion duration was set to 6 s. Raw UHPLC-MS/MS data were processed using Lipidsearch software, including peak detection, alignment, and quantitation. The main parameters included: actual mass tolerance, 5 ppm; retention time tolerance, 0.05 min; signal/noise ratio, 3. Lipid annotation was performed using spectral libraries from LipidBlast (Kind et al., 2013) and LIPID MAPS (http://www.lipidmaps.org/), covering eight major lipid classes and associated subclasses. Lipid abundance was quantified based on normalized peak areas, with each feature reflecting the relative concentration of a specific lipid species.

Proteomic profiling of musk secretions

Musk samples were cryogenically pulverized under liquid nitrogen and lysed in PASP buffer (100 mmol/L NH4HCO3, 8 mol/L urea, pH 8.0; Sigma-Aldrich, USA), followed by sonication and centrifugation (15 min at 10 927 r/min, 4℃). The supernatant was reduced using DL-Dithiothreitol (DTT) (Sigma-Aldrich, USA) and alkylated with iodoacetamide (IAM) (Sigma-Aldrich, USA), then precipitated using cold acetone. The resulting pellet was dissolved in a buffer containing 8 mol/L urea and 100 mmol/L triethylammonium bicarbonate (TEAB, pH 8.5) (Sigma-Aldrich, USA). Total protein concentration was determined using the Bradford assay (Thermo Fisher, Germany). For proteomic analysis, 20 μg of protein from each sample was separated by 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by Coomassie Brilliant Blue R-250 staining. Tryptic peptides were analyzed by LC-MS/MS. Spectral data were searched against the species-specific (forest musk deer or muskrat) protein database using Proteome Discoverer (v.2.4) (Thermo Fisher Scientific, USA). Search parameters included: 10 ppm precursor and 0.02 Da fragment ion mass tolerances; carbamidomethylation as a fixed modification; oxidation of methionine and N-terminal modifications (acetylation, Met-loss, Met-loss+Acetyl) as variable modifications; and maximum of two missed cleavages. Peptide spectrum matches (PSMs) were filtered for high confidence (>99%), and proteins were required to contain at least one unique peptide. The false discovery rate (FDR) was set at ≤1.0% for both peptides and proteins.

Comparative proteomic analysis of human exocrine secretions —encompassing milk (Martin Carli et al., 2023), sweat (Burat et al., 2021), prostatic fluid (Principe et al., 2012), tears (Beisel et al., 2024), and saliva (Wang et al., 2018) —revealed that the protein composition of musk secretions from both species most closely resembled that of human milk.

SnRNA-seq

To ensure consistent gene nomenclature across species included in this study, all gene identifiers were converted to human orthologs based on sequence homology. One-to-one orthologs were determined using the Reciprocal Best Hits (RBH) method implemented in BLASTP (v.2.9.0) (Zhang et al., 2000). Genes lacking a one-to-one homologous match were retained with their original species-specific annotations.

For musk gland transcriptomic profiling, raw snRNA-seq data generated via the 10x Genomics platform were aligned to species-specific reference genomes. The muskrat genome assembled in this study (NCBI: PRJNA1150420) and the forest musk deer genome (NCBI: GCF_004010195.1) were used for alignment using Cell Ranger (https://github.com/10XGenomics/cellranger) with default parameters. Following previously established methods (Yuan et al., 2022). Gene expression matrices were then generated for downstream processing. Data preprocessing and analysis were conducted using the Seurat package (v.3.1.1) (Stuart et al., 2019) in R. Expression matrices were imported using the Read10X function and converted into Seurat objects with “CreateSeuratObject”, filtering out low-quality cells (expressing fewer than 200 genes) and low-abundance genes (expressed in fewer than three cells). Doublets and ambient RNA contamination were removed using Scrublet (v.2.0) (Wolock et al., 2019) and SoupX (v.1.5.0) (Young & Behjati, 2020), respectively. Data normalization was conducted using the NormalizeData function, and the top 2 000 highly variable genes were identified with the FindVariableFeatures function.

Subsequently, the data were scaled using the ScaleData function, and principal component analysis (PCA) was performed for dimensionality reduction using the RunPCA function. Cell clustering was carried out using the FindNeighbors and FindClusters functions, followed by dimensionality reduction mapping of cell clusters using RunUMAP.

To integrate snRNA-seq data across both species, canonical correlation analysis (CCA) was employed. Integration anchors representing shared features were identified using the FindIntegrationAnchors function in Seurat with “dims=1:30”, and the datasets were merged using IntegrateData. The integrated dataset was then processed using the ScaleData, RunPCA, FindNeighbors, and FindClusters (dim=30) functions. Cell clusters were annotated using canonical marker genes. Finally, RunUMAP was used to obtain a uniform manifold approximation and projection (UMAP) representation of the cell expression profiles.

To investigate conserved secretory programs, musk gland transcriptomes were compared with publicly available single-cell datasets from human and mouse exocrine tissues. Human datasets included salivary glands (GSE199209) (Horeth et al., 2023), mammary glands (E-MTAB-9841) (Twigger et al., 2022), pancreas (GSE84133) (Baron et al., 2016), and prostate (GSE150692) (Crowley et al., 2020). Mouse datasets included salivary glands (GSE175649) (Horeth et al., 2021), lacrimal glands (GSE232146) (Delcroix et al., 2023), mammary glands (GSE176063) (Fei et al., 2022), pancreas (GSE84133) (Suzuki & Shimodaira, 2006), and prostate (GSE150692) (Crowley et al., 2020). Each sample underwent quality control, filtering, normalization, and integration using CCA. Tissue-level gene expression patterns were compared using Spearman correlation analysis, followed by phylogenetic tree construction based on the correlation matrix. Cluster stability was evaluated using multiscale bootstrap resampling implemented in the pvclust R package (v.2.2.0) (Suzuki & Shimodaira, 2006), which calculated bootstrap support values for each dendrogram node. Analysis indicated that musk gland transcriptomes from both species showed the highest similarity to mammary gland profiles. Focusing on the epithelial compartment, highly expressed genes in the musk and mammary gland epithelial cells were identified using the FindAllMarkers function (parameters set to “logfc.threshold= 0.25, min.pct=0.25, and *P*-adjust<0.05”). A subset of 354 genes exhibiting high expression in musk gland epithelial cells from both species—but minimal or no expression in mammary epithelial cells —was selected for functional enrichment analysis using KOBAS (v.3.0) (http://bioinfo.org/kobas/).

Identification of highly expressed genes and functional signatures in musk gland epithelial cell subtypes

Given the critical role of epithelial cells in musk biosynthesis and secretion, transcriptomic profiles of musk gland epithelial cells from both forest musk deer and muskrat were subjected to dimensionality reduction and cell-type annotation. Genes expressed in various cell subtypes were identified using the FindAllMarkers function (parameters set to “logfc.threshold=0.25 and min.pct=0.25”). The top 200 marker genes per subtype were selected for functional enrichment analysis using the org.Hs.eg.db (Panwar et al., 2016) and clusterProfiler (Wu et al., 2021) packages in R. Focused analysis of acinar cells was conducted to characterize conserved transcriptional features across species. Highly expressed genes significantly enriched in acinar cells from each species were identified using the FindAllMarkers function (parameters set to “logfc.threshold=0.25, min.pct=0.25, and *P*-adjust<0.05”). To assess the significance of overlap between the two acinar gene sets, Fisher’s exact test was applied. The background consisted of 14 712 orthologous genes shared between the forest musk deer and muskrat. A 2×2 contingency table was constructed to compare the observed overlap against the null hypothesis of random overlap. Statistical significance was determined at $P<0.05$. The resulting intersecting genes were further subjected to pathway and functional annotation using KOBAS (v.3.0).

To detect positively selected genes (PSGs) in the musk deer and muskrat, a phylogenetic tree was constructed using RAxML (v.8.2.4) (Stamatakis, 2014) with 11 species: forest musk deer (*Moschus berezovskii*; GCF_022376915.1), muskrat (*Ondatra zibethicus*; generated in this study, NCBI: PRJNA1150420), cow (*Bos taurus*; ARS-UCD1.3), Yarkand deer (*Cervus hanglu yarkandensis*; CEY_v1), mouse (*Mus musculus*; GRCm39), human (*Homo sapiens*; GRCh38), giraffe (*Giraffa camelopardalis*; GCA_017591445.1), pig (*Sus scrofa*; Sscrofa11.1), dog (*Canis lupus familiaris*; ROS_Cfam_1.0), rabbit (*Oryctolagus cuniculus*; OryCun2.0), lesser Egyptian jerboa (*Jaculus jaculus*; JacJac1.0), and opossum (*Monodelphis domestica*; ASM229v1). Positive selection analyses were conducted using the adaptive branch-site random effects likelihood (ABSREL) model (Murrell et al., 2015) of HyPhy (v.2.5.15) (Pond et al., 2005) with the parameters “code Universal -branches -tree -output”. Genes were considered positively selected when their likelihood ratio tests (LRTs) showed adjusted *P*-values less than 0.05.

DNA extraction and metagenomic sequencing

Total genomic DNA was extracted from 200 mg of each sample following previously established protocols (Yu & Morrison, 2004) based on repeated bead-beating using a mini-bead beater (Biospec Products, Bartlesville, USA). DNA integrity was assessed by electrophoresis on 1% agarose gels, and concentration and quality were determined using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific, Wilmington, USA). For each sample, 1 μg of high-quality DNA was used as input material for library construction. Sequencing libraries were generated using a NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB, USA), following the manufacturer’s recommendations. Indexing barcodes were added to assign sequences to individual samples. DNA was fragmented by sonication to 350 bp, followed by end-repair, A-tailing, and ligation of Illumina-compatible adapters. Adapter-ligated fragments were amplified by polymerase chain reaction (PCR) and purified using the AMPure XP system (Beckman Coulter, USA). Library quality was verified by fragment size analysis on an Agilent 2100 Bioanalyzer (Agilent, USA), and library concentrations were quantified via real-time PCR. Clustering of the index-coded samples was performed using the cBot Cluster Generation System according to the manufacturer’s instructions. After cluster generation, library preparations were sequenced on the Illumina HiSeq platform (USA) and paired-end reads were generated.

Metagenomic profiling of musk secretions

To elucidate the microbial composition and functional potential of musk secretions, metagenomic sequencing was performed on musk samples of musk glands from the forest musk deer and muskrat (*n*=3 per species), generating approximately 30 Gb of high-quality data per sample. Additionally, publicly available fecal metagenomic datasets from the forest musk deer (D_fece) were obtained from the NCBI database (accession numbers SRR18763203, SRR18763217, and SRR18763216). Raw sequencing reads were first subjected to quality filtering using Trimmomatic (v.0.36) (Bolger et al., 2014) to remove adapters, low-quality bases (quality scores<25), and reads with lengths <50 bp. Host-derived sequences were excluded by aligning filtered reads against the reference genomes of forest musk deer and muskrat using BWA-MEM (v.0.7.17) (Li & Durbin, 2009). The filtered reads from each sample were *de novo* assembled using MEGAHIT (v.1.1.1) (Li et al., 2015) with the option min-contig-len 500. Open reading frames (ORFs) from each sample were predicted using GeneMarkS (v.2.7) (Besemer et al., 2001) with the parameter: gmhmmp-m MetaGeneMark_v1.mod. CD-HIT (v.4.6.6) (Li & Godzik, 2006) was used to construct a non-redundant gene catalog with 95% cutoff sequencing identity and 90% coverage. Kraken 2 (v.2.1.3) (Wood et al., 2019) was used to identify taxonomic information based on the nonredundant gene catalog. Community structure and functional variation were visualized through principal coordinates analysis (PCoA) based on the Bray-Curtis dissimilarity matrices. Statistical evaluation of group-wise differences was performed using analysis of similarities (ANOSIM). Functional annotations were assigned by aligning translated protein sequences from the integrated gene catalog against the protein or domain sequences in the KEGG database (v.90.0) (Kanehisa et al., 2016) using DIAMOND (v.0.8.24.86) (Buchfink et al., 2015) with a cutoff e-value<1e-5. KEGG Orthology (KO) identifiers were assigned based on the highest-scoring annotated hit. High-quality reads from each sample were aligned against the gene catalogs using BWA-MEM (v.0.7.17) (Li & Durbin, 2009), and abundance profiles of genes (alignment length≥50 bp and sequence identity>95%) were calculated in transcripts per million (TPM), with corrections for variations in gene length and mapped reads per sample. KO-level relative abundances were calculated from the abundances of annotated genes, while pathway-level relative abundances were calculated from the summation of the relative abundances of its contained KOs. Comparative statistical analyses between the two groups were tested using the Wilcoxon rank sum test, with $P<0.05$ indicating significance, and FDR was calculated based on Benjamini and Hochberg correction.

RESULTS

Comparative metabolomic profiling of musk secretions

To determine the chemical basis underlying the shared olfactory characteristics of musk deer and muskrat secretions, comprehensive metabolomic analyses were conducted using GC-MS. A total of 112 and 89 volatile compounds were detected in musk samples from forest musk deer and muskrat, respectively. Among these, macrocyclic ketones containing 15 to 17 carbons constituted the principal odor-active compounds in both species. Muscone was the most abundant compound in musk deer, accounting for 42.22% of total GC-MS peak area. In contrast, cycloheptadecanone (18.36%) and cyclopentadecanone (7.54%) were the primary macrocyclic ketones identified in muskrat musk (Figure 1A; Supplementary Table S2). Given the limited availability and high cost of natural musk (traditionally obtained from musk deer), cyclopentadecanone has been widely adopted as a key precursor in synthetic musk production (Liu et al., 2016). Notably, this bioactive component was present in musk secretions from both species (Figure 1A; Supplementary Table S2), highlighting convergence in bioactive metabolite

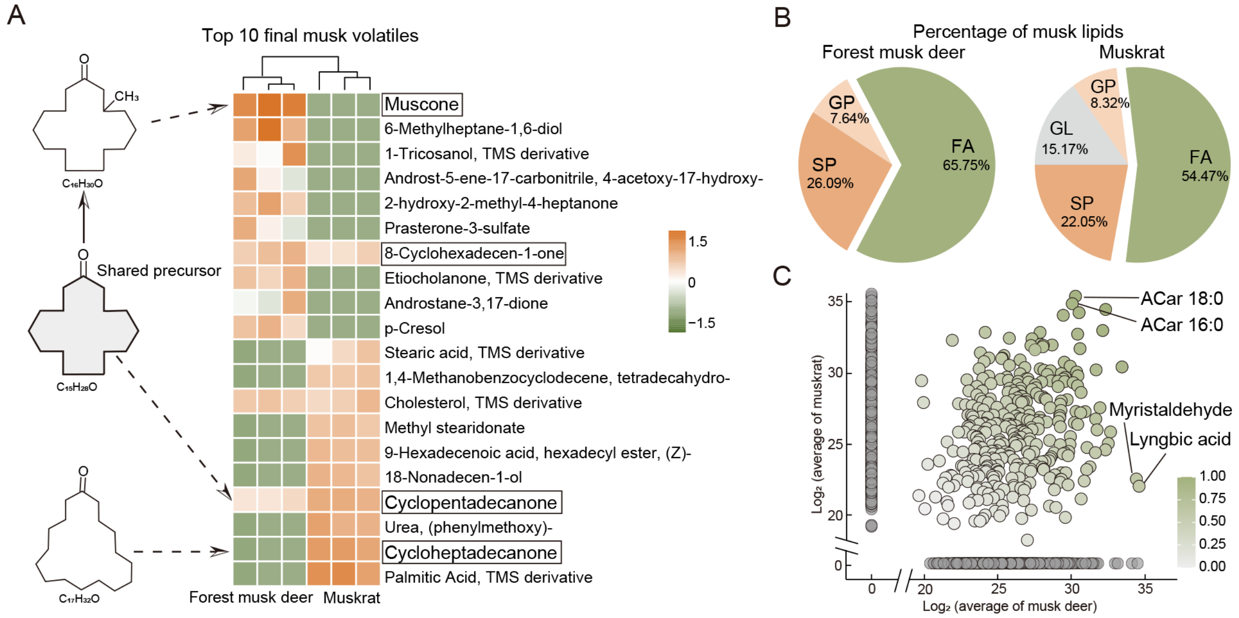

Figure 1 Comparative analysis of chemical compositions and lipid profiles of musk secretions between both species

A: Heatmap showing the top 10 most abundant volatile musk metabolites in musk deer and muskrat, determined by GC-MS analysis. Color scale represents $\log_2$-transformed normalized relative abundance of metabolites. B: Pie charts depicting percentage of lipid categories between two species, determined by LC-MS analysis. C: Scatter plot showing 347 conserved lipids between both species (green), with color intensity representing scaled sum of lipid relative abundance in both species. FA, fatty acyls, GL, glycerolipids; GP, glycerophospholipids; SP, sphingolipids; SL, saccharolipids; Acar, acylcarnitines.

production and reinforcing its relevance as a functional scent molecule.

Previous studies have suggested that macrocyclic ketones may be synthesized through lipid-derived intermediates, particularly long-chain fatty acids (Erickson & Hix, 1948; Meng et al., 2019; Ward & van Dorp, 1981). To investigate upstream biosynthetic precursors of the above macrocyclic ketones, LC-MS-based lipidomic analysis was conducted. A total of 974 and 1 007 lipid species were identified in musk from forest musk deer and muskrat, respectively. Lipids were classified into five categories, including fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), and saccharolipids (SL). Fatty acyls represented the predominant category, comprising 65.75% and 54.47% of total musk lipids in musk deer and muskrat, respectively (Figure 1B; Supplementary Table S3). A total of 347 lipids were shared between the two datasets, including acyl carnitine (ACar) 18:0, myristaldehyde and lyngbic acid (Figure 1C; Supplementary Table S4). These findings demonstrate a fundamentally similar lipid composition of musk dominated by fatty acyls in both species, suggesting similar secretory mechanisms in musk glands despite their distinct evolutionary origins.

Proteomic and transcriptomic signatures reveal apocrine specialization of musk glands

To elucidate the secretory architecture of musk glands in both muskrat and musk deer, a multi-omics approach integrating proteomics, transcriptomics, and high-quality reference genomes was employed. Proteomic analysis of musk gland secretions, anchored by the newly assembled chromosome-level genome for muskrat (Supplementary Figure S1 and Table S1) and the publicly available genome for musk deer (Chen et al., 2022), identified 1 954 proteins in musk deer and 2 207 in muskrat. Of these, 1 142 proteins were conserved across both species, with functional enrichment in metabolic processes and cellular organization (Supplementary Figure S2A, B and Table S5). Comparative profiling against human exocrine secretions revealed a striking similarity between musk secretions from both species and milk proteomes (Figure 2A), the latter representing a canonical apocrine secretion (Farkaš, 2015; Khan et al., 2022). The shared musk-milk proteins (*n*=780) were enriched in cellular organelle components and metabolic activities (Supplementary Figure S2C, D and Table S6), consistent with apocrine secretion involving the extrusion of organelle-containing apical

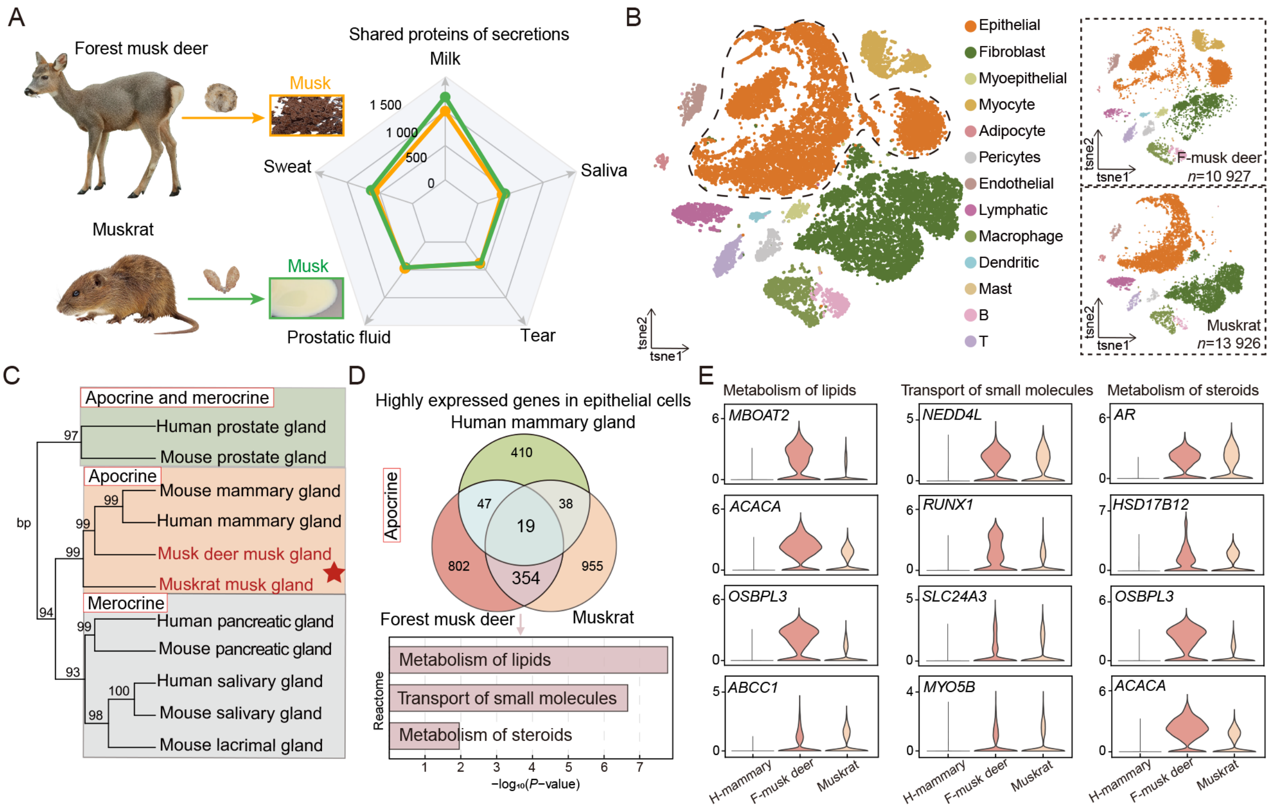

Figure 2 Cellular and transcriptional analyses of musk glands as specialized apocrine glands

A: Left panels show forest musk deer (upper) and muskrat (lower) graphs. Radar chart (right) showing number of shared proteins between musk gland secretions and other accessible exocrine glands secretions, highlighting maximum similarity between musk glands and mammary gland. B: t-distributed stochastic neighbor embedding (t-SNE) plots showing integrated musk gland cells from both species, with individual species plots displayed on the right. C: Dendrogram illustrating gene expression pattern similarities between musk glands and various accessible exocrine glands (including mammary, prostate, pancreatic, lacrimal, and salivary glands from humans and mice). Bootstrap values indicate statistical robustness of clustering groups, demonstrating that musk glands from both species most closely resemble mammary glands. D: Venn diagram (upper) of highly expressed genes in epithelial cells from musk glands and human mammary glands. Bar plot (lower) showing reactome functional enrichment pathways of 354 shared and uniquely highly expressed genes in musk gland epithelial cells of musk deer and muskrat but absent in human mammary glands. E: Violin plot depicting gene expression patterns for three enriched pathways identified in Figure 2D. Y-axis depicts epithelial cells from three apocrine glands: human mammary gland (H-mammary), forest musk deer musk gland (F-musk deer), and muskrat musk gland (Muskrat).

cytoplasm (Farkaš, 2015). These findings suggest that musk glands employ an apocrine-like mode of secretion.

To further investigate the cellular and transcriptional underpinnings of musk gland secretion, snRNA-seq was performed on musk gland tissue from both species. This approach yielded 10 927 nuclei from musk deer and 13 926 nuclei from muskrat. CCA revealed conserved cell populations, including epithelial cells, endothelial cells, lymphatic endothelial cells, fibroblasts, adipocytes, pericytes, and immune cells (Figure 2B; Supplementary Figure S3A). Epithelial cells represented the dominant population, comprising 44.60% and 47.35% of total cells in forest musk deer and muskrat, respectively (Supplementary Figure S3B and Table S7).

Cross-species transcriptomic comparisons with publicly available single-cell datasets from human and mouse exocrine glands —including apocrine mammary, merocrine salivary, pancreatic, and lacrimal glands, as well as mixed-mode prostate glands—revealed that musk gland expression profiles most closely resembled those of mammary glands (Figure 2C). As the primary functional cells responsible for synthesis and secretion in exocrine glands (Khan et al., 2022), epithelial cells were examined in greater detail. A conserved set of genes was found to be highly and consistently expressed in epithelial cells of both musk and mammary glands (Figure 2D; Supplementary Table S8). Among these, *ACLY* and *PLIN2* are closely associated with lipid synthesis and droplet formation (Chong et al., 2011; Novak & Innis, 2011), while *EHF* is involved in alveolar cell differentiation (Nightingale et al., 2024). The presence of these transcripts in both tissues underscores a shared regulatory program underpinning lipid- and protein-rich secretory activity, reinforcing the hypothesis of an apocrine-based mechanism in musk production.

Beyond this conserved core, 354 genes were selectively and robustly expressed in musk gland epithelial cells from both species while being minimally or not expressed in mammary epithelial cells (Figure 2D, E). These molecular markers underscore the functional specialization of musk glands as a distinct apocrine organ. These genes were enriched in pathways related to lipid metabolism (*MBOAT2*, *ACACA*, *OSBPL3*, and *ABCC1*), small-molecule transport (*NEDD4L*, *RUNX1*, *SLC24A3*, and *MYO5B*), and steroid metabolism (*AR*, *HSD17B12*, *OSBPL3*, and *ACACA*) (Figure 2D, E; Supplementary Table S9). Collectively, these findings demonstrate that musk glands are specialized apocrine glands adapted for steroid-regulated synthesis and secretion of lipid-based musk compounds.

Comparative cellular and molecular profiles of musk biosynthesis

Single-nucleus transcriptomic analysis of epithelial cells in forest musk deer and muskrat identified three major cell types (acinar, ductal, and basal) and five conserved cellular subtypes (acinar_1, basal_1, basal_2, ductal_1, and ductal_2) with similar molecular signatures (Figure 3A, B; Supplementary Figure S4A, B and Table S10). Gene Ontology (GO) enrichment analysis identified acinar_1 cells as the dominant site for musk biosynthesis, characterized by high expression of genes involved in steroid and fatty acid metabolism (*AR*, *PLIN2*, and *ACSBG1*) (Figure 3B; Supplementary Figure S4C and Table S11). In contrast, ductal subtypes (ductal_1 and ductal_2) expressed genes linked to epidermal differentiation (*KRT6A* and *KRT79*) and epithelial cell proliferation (*AREG* and *EHF*), while basal subtypes (basal_1 and basal_2) showed high expression of genes related to cell proliferation, adhesion, and junction assembly (*TP63*, *CADM1*, and *TNS1*) (Figure 3B; Supplementary Figure S4C; Supplementary Table S11).

In forest musk deer, two additional epithelial subtypes, acinar_2 and ductal_3, emerged as species-specific clusters (Figure 3A, B; Supplementary Figure S4A, B). Ductal_3 cells exhibited enriched expression of genes involved in mammary gland alveolus development (*RUNX1*, *TNFRSF11A*, and *ERBB4*), while acinar_2 cells showed high expression of *AR*, *OLAH*, *ABCC11*, *ACLY*, and *SLC7A8,* genes involved in steroid and lipid metabolism. (Figure 3B; Supplementary Figure S4C and Table S11). Correlation analysis (R=0.59, P<2.2e−16) and functional assessments suggested that acinar_2 shared many steroid and fatty acid metabolic processes with acinar_1 (Supplementary Figure S4C, D and Table S11), potentially enhancing musk biosynthetic output in forest musk deer.

Further analysis of acinar cells identified 1 042 and 992 highly expressed genes in forest musk deer and muskrat, respectively, with 188 genes forming a conserved core set (P=3.856963×10^{-38}), primarily involved in fatty acid metabolism (*ACSBG1*, *HSD17B12*, *HACD2*, *HADHB*, *ACOX1*, *HADHA*, and *ACAA2*) and glycerolipid metabolism (*MBOAT2* and *LPIN2*) (Figure 3C, D; Supplementary Table S12). Evolutionary analysis revealed signatures of positive selection acting on *ACSBG1* in both muskrat (P-adjust=9.74×10^{-6}) and musk deer (P-adjust=0.031), while *HSD17B12*, *HACD2* and *HADHA* showed signatures of positive selection uniquely in forest musk deer (P-adjust<0.05), highlighting the potential role of fatty acid metabolism genes in adaptive evolution related to musk biosynthesis. Specifically, *ACSBG1* activates long-chain fatty acids into acyl-CoA, a fundamental step enabling fatty acids to enter metabolic pathways and regulate cellular lipid processing and signal transduction (Watkins et al., 2007; Ye et al., 2024). *HSD17B12* drives key oxidation-reduction reactions involving fatty acids and sterols, participating in fatty acid chain elongation that mediates polyunsaturated fatty acid synthesis (Heikelä et al., 2020; Moon & Horton, 2003). These findings suggest a shared molecular framework for musk lipid biosynthesis between these two species.

Building upon this shared metabolic scaffold, species-specific expression signatures revealed distinct biosynthetic strategies. In the forest musk deer, acinar cells showed elevated expression of genes associated with fatty acid biosynthesis (*ACSL1*, *ACACA*, *ELOVL4*, *FASN*, *EHHADH*, *OLAH*, and *ACACB*) and glycerolipid metabolism (*ALDH7A1*, *MGLL*, and *MOGAT2*) (Figure 3C, D; Supplementary Table S13). In contrast, muskrat-specific genes were enriched in glycerolipid (*GPAM*, *PLPP4*, *ALDH3A2*, *DGKI*, *DGKB*, *AKR1B1*, *DGKG*, and *LPIN1*) and fatty acid metabolism (*ACSL5*, *HADH*, and *ELOVL6*) (Figure 3C, D; Supplementary Table S14). Together, these patterns reveal a conserved molecular scaffolding for musk lipid production, complemented by species-specific metabolic pathways that refine and diversify musk biosynthesis in each species.

Convergent microbial lipid metabolism underpins musk production

Emerging evidence suggests that microbial communities may

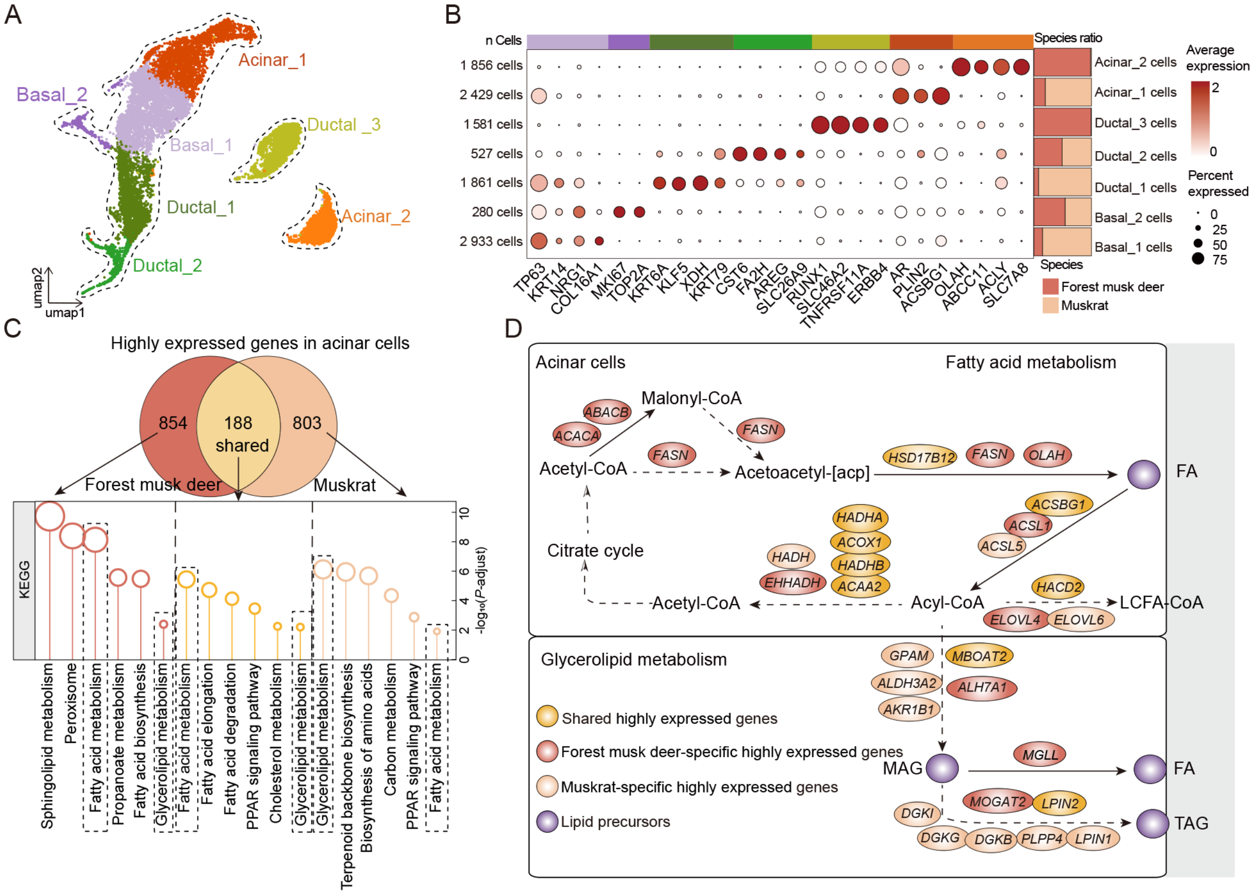

Figure 3 Shared cellular and expression characteristics of epithelial cell subtypes between both species

A: UMAP plot showing integrated epithelial cells of musk glands from two species. B: Bubble plot showing marker gene expression in epithelial cell subtypes. C: Venn diagram (upper) showing highly expressed genes in acinar cells between both species. Lollipop plot (bottom) illustrating KEGG pathway enrichment among highly expressed acinar cell genes, including forest musk deer-specific genes (854), common genes (188), and muskrat-specific genes (803). D: Schematic illustrating key genes involved in fatty acid metabolism and glycerolipid metabolism. Bubbles in red represent lipid precursors of FA, MAG, and TAG. Genes showing different highly expressed patterns in acinar cells are labeled in red (musk deer-specific genes), light orange (muskrat-specific genes), and yellow (shared genes between musk deer and muskrat), respectively. FA, fatty acyls; MAG, monoacylglycerols; TAG, triacylglycerols.

contribute to musk formation, potentially influencing the production of muscone, fatty acids, and other musk compounds (Xu et al., 2024). To investigate this potential microbial involvement, shotgun metagenomic sequencing was performed on musk gland secretions from forest musk deer and muskrat. This effort yielded 180 Gb of high-quality reads across six samples, which were assembled into 155 663 contigs with an average N50 of 18 235 bp. A total of 392 487 open reading frames (ORFs) were predicted across these samples (Supplementary Table S15). Taxonomic classification revealed that bacteria dominated the musk microbiomes, accounting for 99.06% of the total microbial community, while eukaryotes (0.44%) and archaea (0.08%) represented a minimal fraction (Supplementary Table S16).

To determine the distinctiveness of bacterial communities in musk secretions, publicly available fecal metagenomic data from forest musk deer were analyzed alongside musk microbiomes. Taxonomic comparisons demonstrated clear divergence in bacterial community compositions between musk and fecal samples at both the phylum (Figure 4A) and genus levels (Figure 4B). PCoA and hierarchical clustering further confirmed that musk microbiomes from both species were more similar to each other than to fecal microbiomes (ANOSIM, *P*=0.001, *R*=0.9269; Supplementary Figure S5A, B). Notably, Actinobacteria was enriched in musk samples (*P*-adjust<0.01), whereas fecal communities were dominated by Firmicutes and Bacteroidetes (*P*-adjust<0.01) (Supplementary Table S17). A dominant genus across both musk datasets was the lipid-metabolizing genus *Corynebacterium*. Of the 123 microbial species shared across the musk microbiomes, 106 belonged to *Corynebacterium* (Supplementary Table S18), suggesting taxonomic and functional convergence in microbial composition.

Functional metagenomic comparison based on KEGG pathway analysis revealed extensive divergence between musk and fecal microbiomes. In musk gland communities, 123 and 186 pathways were significantly enriched in muskrat and musk deer, respectively (Figure 4C). Lipid-related pathways, such as fatty acid degradation, synthesis and degradation of ketone bodies, alpha-linolenic acid metabolism, and unsaturated fatty acid biosynthesis, were consistently up-regulated in musk microbiomes (Figure 4C; Supplementary Table S19). This lipid metabolic specialization aligns with the biosynthetic demands of musk, where ketones and other lipophilic volatiles are central to fragrance production.

Further enzymatic profiling of the musk microbiomes

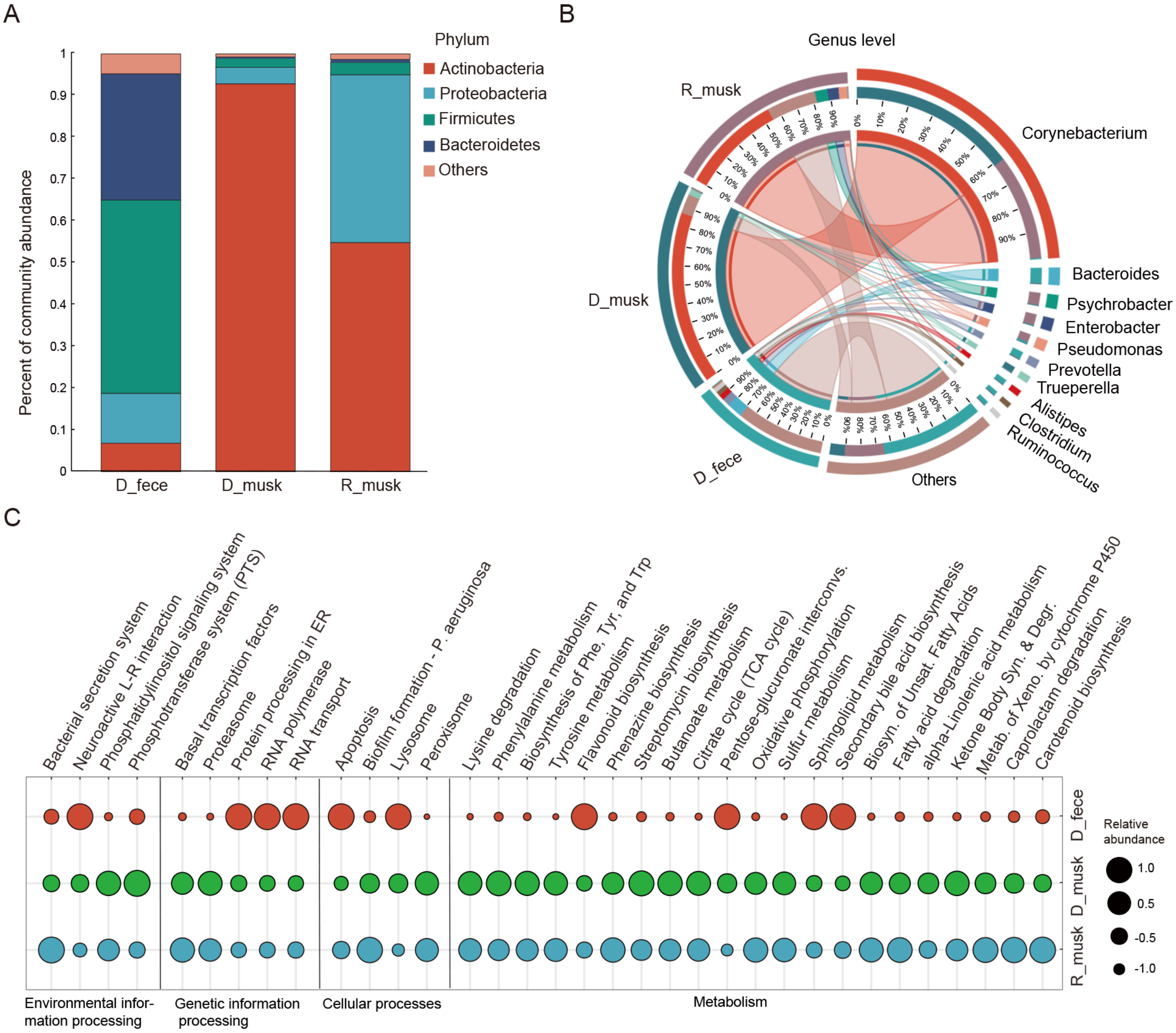

Figure 4 Convergent microbial community structure and functions of musk from both species

A: Phylum-level microbial composition of musk from musk deer and muskrat (D_musk and R_musk) and feces from forest musk deer (D_fece). B: Genus-level microbial composition of musk from both species and feces from musk deer. C: Shared differentially expressed functions (partial) obtained from KEGG pathway comparison between musk and fecal microbiota (circle size represents relative abundance of each pathway). Ketone Body Syn. & Degr, synthesis and degradation of ketone bodies; Biosyn. of Unsat. Fatty Acids, unsaturated fatty acid biosynthesis.

revealed 35 highly abundant lipid metabolism-related enzymes conserved across both species (out of 123 shared enzymes; Supplementary Table S20). These enzymes span multiple interconnected lipid metabolism pathways (Figure 5), processing triacylglycerols (TAGs) and primary fatty acids generated from musk glands as key precursors of musk odorants. In the glycerolipid metabolism pathway (ko00561), triacylglycerol lipase (EC: 3.1.1.3) and related enzymes hydrolyze TAGs into glycerol and fatty acids. The generation of ketones, major musk odorants, occurs through two distinct microbial pathways. In the fatty acid degradation pathway (ko00071), fatty acids are first converted to aldehydes and alcohols by aldehyde dehydrogenase (ALDH), alcohol dehydrogenase (E1.1.1.1), and alcohol dehydrogenase propanol-preferring (adhP) (Figure 5, A → B). Subsequently, long-chain acyl-CoA synthetase (ACSL) activates fatty acids to fatty acyl-CoA, which is subsequently catabolized to acetyl-CoA via acyl-CoA oxidase (E1.3.3.6), enoyl-CoA hydratase (paaF), and acetyl-CoA acyltransferase (fadA) (Figure 5, B → C). Within the ketone body synthesis pathway (ko00072), fadA and acetyl-CoA C-acetyltransferase (ACAT) convert acetyl-CoA to acetoacetyl-CoA, while 3-oxoacid CoA-transferase subunit A (scoA) plays a crucial role in ketone biosynthesis (Figure 5, C → D).

Acetyl-CoA also serves as a key precursor in *de novo* fatty acid biosynthesis (ko00061 and ko01040), facilitated by 11 conserved enzymes, including 3-oxoacyl-[acyl-carrier protein] reductase (fabG), synthase II (fabF), and acyl carrier protein (acpP) (Figure 5C; Supplementary Table S20). The presence of these highly abundant enzymes across both species underscores their essential role in the synthesis of musk volatile compounds.

Despite these shared features, species-specific differences in lipid-related enzyme profiles were evident. Forest musk deer microbiomes exhibited 42 unique lipid-metabolic enzymes, compared to 16 in muskrat. These enzymes were functionally associated with fatty acid degradation (HADH, HADHA, CPT1A, CPT1B, CPT2, ACADVL, and ALDH7A1) and unsaturated fatty acid biosynthesis (ELOVL6, ELOVL1, ELOVL4, ELOVL7, and HSD17B4) (Figure 5; Supplementary Tables S21, S22). The presence of species-specific enzymatic pathways, in conjunction with a conserved metabolic core,

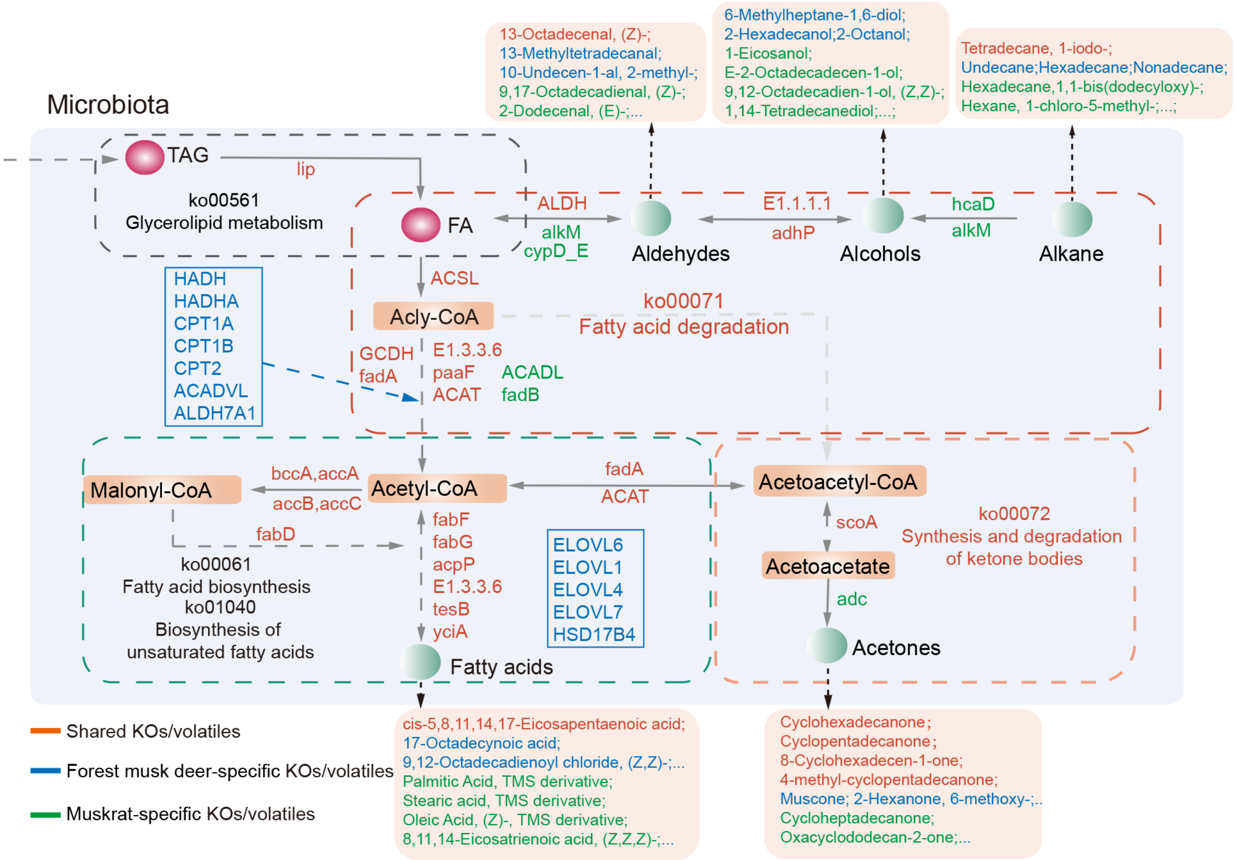

Figure 5 Coordinated framework of microbial-mediated lipid metabolism pathways in musk volatile compound biosynthesis

Different colored dashed boxes in metabolic framework represent distinct key pathways of glycerolipid metabolism (A: ko00561, gray), fatty acid degradation (B: ko00071, red), fatty acid biosynthesis (C: ko00061, ko01040, green), and synthesis and degradation of ketone bodies (D: ko00072, orange). Arrows in framework indicate directionality of various biochemical reactions. Key enzymes or volatile compounds involved in these reactions are coded in blue (forest musk deer-specific KOs/volatiles), green (muskrat-specific KOs/volatiles), and red (shared KOs/volatiles between forest musk deer and muskrat), respectively.

underscores the complex biochemical landscape that drives musk production in each species.

DISCUSSION

This study identified a collaborative musk biosynthesis process orchestrated by the host glands and associated microbial communities in both forest musk deer and muskrat. Despite their distant phylogenetic backgrounds, both species exhibited a shared, stepwise “ladder-like” architecture for musk production, involving host-synthesized lipid precursors and subsequent microbial metabolism to generate volatile end products (Figure 6).

The initial phase of musk production is driven by the host musk gland, which functions as a specialized exocrine organ for lipid precursor secretion. Proteomic and single-cell transcriptomic analyses revealed that both musk glands exhibited apocrine characteristics, a finding supported by their unique gene expression patterns and secreted protein compositions (Figure 2A, C). These results are congruent with previous histological studies (Hong et al., 2023; Zheng et al., 2021) and further supported by the high expression of genes implicated in androgen signaling and lipid biosynthesis (*AR*, *HSD17B12*, *OSBPL1A*, and *MBOAT2*), indicating an androgen-driven mechanism adapted for the production of lipid-rich musk secretions (Figure 2D, E).

Comparative single-cell analyses pinpointed acinar cells as the primary site of musk biosynthesis in both species (Figure 3C, D), with a conserved set of 188 highly expressed genes forming the molecular foundation for lipid precursor production. These genes are involved in fatty acid metabolism (*HSD17B12*, *ACSBG1*, *HADHB*, and *ACAA2*) and glycerolipid metabolism (*MBOAT2* and *LPIN2*), underscoring a shared biochemical scaffold. Species-specific genes with elevated expression profiles further highlighted divergent but complementary strategies for generating diverse lipid precursors. Consistent with this, lipidomic analyses revealed a high degree of compositional overlap (R=0.7), with fatty acyl species dominating in both taxa (Figure 1B, C).

The second phase of musk production features microbial metabolism of host-derived lipid precursors into final volatile compounds. The anatomical positioning of the musk gland near the abdominal region adjacent to the reproductive organs (Supplementary Figure S6A, B) facilitates exposure to external microbial communities, fostering a stable colonization niche. Microbiome profiling revealed marked divergence between musk and fecal communities (Figure 4A, B), sustained by specialized semi-anaerobic microenvironments and unique glandular structures. At the genus level, the musk glands of both species harbored *Corynebacterium*-dominant communities (R=0.8) with shared lipid-metabolizing capabilities, although further research is needed to establish whether this similarity exceeds expectations for homologous organs across species. This finding aligns with recent

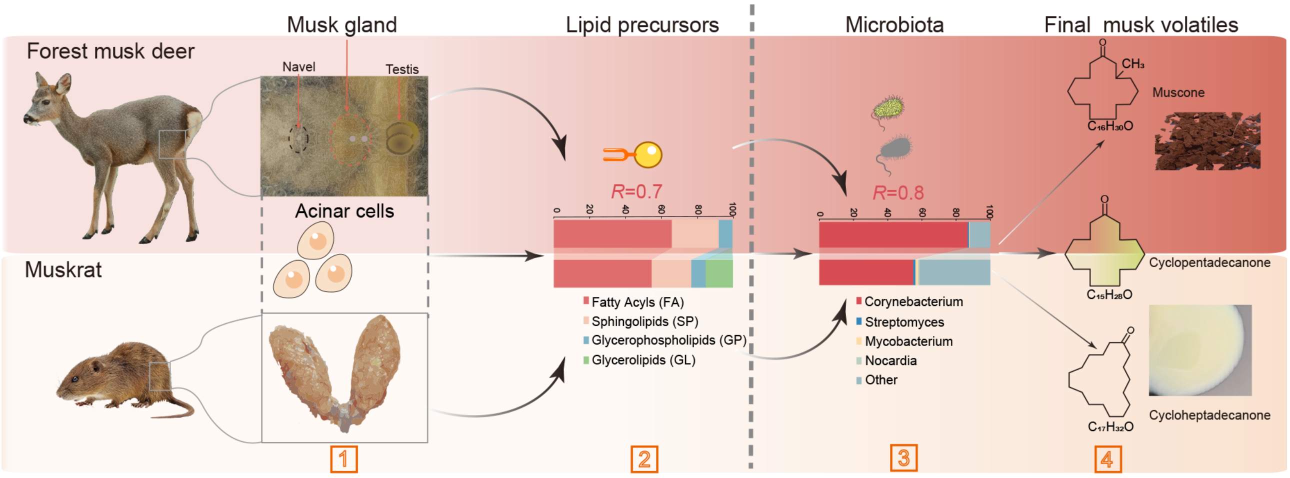

Figure 6 Schematic representation of ladder-like similarities in musk production processes between forest musk deer and muskrat

Left panel illustrates species-specific models for forest musk deer (upper) and muskrat (lower). Central gray dashed line demarcates two sequential stages of musk production: initial stage of lipid precursor generation by host musk glands, followed by the second stage involving microbial metabolism for final musk volatile compound production. Red numbers (1–4) indicate four key aspects of musk synthesis between two species: (1) cellular composition of musk glands, particularly key pathways and highly expressed genes in acinar cells for musk lipid precursor synthesis; (2) lipid composition of musk, predominantly FA; (3) microbiome composition, dominated by *Corynebacterium*; and (4) final musk volatile compounds, primarily consisting of macrocyclic ketones, with cyclopentadecanone as a shared intermediate compound. Increasing Pearson correlation coefficients across these aspects demonstrate the ladder-like similarities in musk biosynthesis between forest musk deer and muskrat.

research demonstrating dynamic relationships between testosterone levels, musk composition, and microbiota across various secretion phases, with Actinobacteria (including *Corynebacterium*) shown to play a crucial role in musk biosynthesis (Xu et al., 2024).

Metagenomic and pathway-level analyses revealed that musk-associated microbial communities harbor a suite of lipid-degrading enzymes capable of converting host-derived lipids into volatile musk compounds (Figure 5). Central to this microbial transformation are enzymes involved in fatty acid degradation (ACSL, E1.3.3.6, paaF, and ACAT), as well as key components of ketone body metabolism (fadA and scoA). These enzymes collectively mediate the biochemical conversion of fatty acids into macrocyclic ketones, including cyclopentadecanone, a core constituent of both musk profiles. Despite this shared enzymatic architecture, species-specific metabolic outcomes were evident. Cyclopentadecanone was abundant in both species, but muscone predominated in musk deer, whereas cycloheptadecanone was more abundant in muskrat (Figure 1A; Supplementary Table S1). This divergence may reflect species-specific enzymatic capabilities. In particular, musk deer microbiomes exhibited a broader complement of genes associated with enhanced fatty acid degradation (HADH, HADHA, CPT1A, CPT1B, CPT2, ACADVL, ALDH7A1) and unsaturated fatty acid biosynthesis (ELOVL6, ELOVL1, ELOVL4, ELOVL7, HSD17B4). These findings provide compelling evidence for the role of microbiota in musk biosynthesis and suggest promising directions for future research. Building upon the temporal dynamics observed in previous studies (Xu et al., 2024), comprehensive metagenomic and metabolomic analyses across different secretion phases could reveal the complex relationships between microbial community composition, enzymatic activities, and musk compound production. Such investigations would not only enhance our understanding of microbiomes in musk biosynthesis but may also inform strategies for optimizing natural musk production.

These findings extend recent genomic advances on the importance of lipid metabolism and cell specialization in musk secretion evolution (Wang et al., 2025). Through integrated multi-omics analyses, this study establishes a comprehensive molecular framework for musk biosynthesis, spanning gene expression, metabolite composition, and microbiome function. Despite the distant phylogenetic separation between forest musk deer and muskrat, both species exhibit parallel adaptations in genomic features, molecular pathways, cellular specialization, and microbial functions, suggesting potential convergent mechanisms that may underlie musk production.

Collectively, these results underscore a convergent yet species-tailored framework of musk production, wherein host apocrine gland activity and specialized microbial transformations synergize to produce a distinctive roster of volatile compounds. This shared biochemical machinery provides a foundation for understanding and harnessing musk biosynthesis across diverse mammalian taxa.

DATA AVAILABILITY

All sequencing data were deposited in the National Center for Biotechnology Information (NCBI) database (BioProjectID PRJNA1150420), Genome Sequence Archive (GSA) database (https://ngdc.cncb.ac.cn/gsa/) (accession number PRJCA037262), and Science Data Bank (doi: 10.57760/sciencedb. j00139.00182).

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Q.Q., Z.P.L., C.X., K.W., and S.Y.S., designed and coordinated the project and its research aspects; S.Y.S., Y.L., C.L.Z., X.L.G., C.G.F., and C.X. collected samples; S.Y.S., L.Z., X.T.Y, and Y.L. performed data analysis; S.Y.S., L.Z., X.T.Y., and K.W. contributed to figure design; S.Y.S., L.Z., R.H., and K.W. wrote the manuscript. Q.Q., T.F.X., Y.M.Z., C.L.Z., and K.W. amended the manuscript. All authors read and approved the final version of the manuscript.

REFERENCES

Ahmed L, Zhang YT, Block E, et al. 2018. Molecular mechanism of activation of human musk receptors OR5AN1 and OR1A1 by (*R*)-muscone and diverse other musk-smelling compounds. *Proceedings of the National Academy of Sciences of the United States of America*, **115**(17): E3950–E3958.

Baron M, Veres A, Wolock SL, et al. 2016. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. *Cell Systems*, **3**(4): 346–360. E4.

Bartoš L, Víchová J, Lancingerová J. 2005. Preorbital gland opening in red deer (*Cervus elaphus*) calves: signal of hunger?. *Journal of Animal Science*, **83**(1): 124–129.

Beisel A, Jones G, Glass J, et al. 2024. Comparative analysis of human tear fluid and aqueous humor proteomes. *The Ocular Surface*, **33**: 16–22.

Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Research*, **29**(12): 2607–2618.

Bi SZ, Jia LZ, Guan Q, et al. 1993. Ultrastructures of musk glandular sacs in musk deer's annual period and secretion and formation of musk. *Chinese Pharmaceutical Journal*, **28**(11): 653–657. (in Chinese)

Birney E, Clamp M, Durbin R. 2004. GeneWise and genomewise. *Genome Research*, **14**(5): 988–995.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**(15): 2114–2120.

Brennan PA, Zufall F. 2006. Pheromonal communication in vertebrates. *Nature*, **444**(7117): 308–315.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, **12**(1): 59–60.

Buchinger TJ, Li WM. 2023. Chemical communication and its role in sexual selection across Animalia. *Communications Biology*, **6**(1): 1178.

Burat B, Reynaerts A, Baiwir D, et al. 2021. Characterization of the human eccrine sweat proteome—a focus on the biological variability of individual sweat protein profiles. *International Journal of Molecular Sciences*, **22**(19): 10871.

Bushmanova E, Antipov D, Lapidus A, et al. 2019. rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data. *GigaScience*, **8**(9): giz100.

Chen CY, Yin Y, Li HR, et al. 2022. Ruminant-specific genes identified using high-quality genome data and their roles in rumen evolution. *Science Bulletin*, **67**(8): 825–835.

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

Chen YS, Zhao WG, Zhao M, et al. 2007. Histological observation on musk-secreting scented gland in muskrat. *Chinese Journal of Zoology*, **42**(2): 91–95. (in Chinese)

Chong BM, Reigan P, Mayle-Combs KD, et al. 2011. Determinants of adipophilin function in milk lipid formation and secretion. *Trends in Endocrinology & Metabolism*, **22**(6): 211–217.

Crowley L, Cambuli F, Aparicio L, et al. 2020. A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *eLife*, **9**: e59465.

Delcroix V, Mauduit O, Lee HS, et al. 2023. The first transcriptomic atlas of the adult lacrimal gland reveals epithelial complexity and identifies novel progenitor cells in mice. *Cells*, **12**(10): 1435.

Dudchenko O, Batra SS, Omer AD, et al. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**(6333): 92–95.

Durand NC, Shamim MS, Machol I, et al. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Systems*, **3**(1): 95–98.

Erickson JLE, Hix HB. 1948. Fatty acids from the scent glands of the Louisiana muskrat. *Journal of the American Oil Chemists' Society*, **25**(12): 447–453.

Fan MY, Zhang MS, Shi MH, et al. 2018. Sex hormones play roles in determining musk composition during the early stages of musk secretion by musk deer (*Moschus berezovskii*). *Endocrine Journal*, **65**(11): 1111–1120.

Farkaš R. 2015. Apocrine secretion: new insights into an old phenomenon. *Biochimica et Biophysica Acta (BBA) - General Subjects*, **1850**(9): 1740–1750.

Fei LJ, Chen HD, Ma LF, et al. 2022. Systematic identification of cell-fate regulatory programs using a single-cell atlas of mouse development. *Nature Genetics*, **54**(7): 1051–1061.

Feng H, Feng TY, Mo YD, et al. 2023. Integrated multi-omics analysis reveals insights into Chinese forest musk deer (*Moschus berezovskii*) genome evolution and musk synthesis. *Frontiers in Cell and Developmental Biology*, **11**: 1156138.

Heikelä H, Ruohonen ST, Adam M, et al. 2020. Hydroxysteroid (17β) dehydrogenase 12 is essential for metabolic homeostasis in adult mice. *American Journal of Physiology-Endocrinology and Metabolism*, **319**(3): E494–E508.

Hong TT, Liu CM, Wang SH, et al. 2023. The histostructures of musk gland at different ages and periods in forest musk deer. *Chinese Journal of Zoology*, **58**(5): 742–750. (in Chinese)

Horeth E, Bard J, Che M, et al. 2023. High-resolution transcriptomic landscape of the human submandibular gland. *Journal of Dental Research*, **102**(5): 525–535.

Horeth E, Oyelakin A, Song EAC, et al. 2021. Transcriptomic and single-cell analysis reveals regulatory networks and cellular heterogeneity in mouse primary Sjögren's syndrome salivary glands. *Frontiers in Immunology*, **12**: 729040.

Hu J, Fan JP, Sun ZY, et al. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, **36**(7): 2253–2255.

Jiang YL, Han XY, Feng NN, et al. 2022. Androgen plays an important role in regulating the synthesis of pheromone in the scent gland of muskrat. *The Journal of Steroid Biochemistry and Molecular Biology*, **217**: 106026.

Jones P, Binns D, Chang HY, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, **30**(9): 1236–1240.

Kanehisa M, Sato Y, Kawashima M, et al. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, **44**(D1): D457–D462.

Kent WJ. 2002. BLAT–the BLAST-like alignment tool. *Genome Research*, **12**(4): 656–664.

Khan S, Fitch S, Knox S, et al. 2022. Exocrine gland structure-function relationships. *Development*, **149**(1): dev197657.

Kind T, Liu KH, Lee DY, et al. 2013. LipidBlast *in silico* tandem mass spectrometry database for lipid identification. *Nature Methods*, **10**(8): 755–758.

Kind T, Wohlgemuth G, Lee DY, et al. 2009. FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Analytical Chemistry*, **81**(24): 10038–10048.

Li DH, Liu CM, Luo RB, et al. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph. *Bioinformatics*, **31**(10): 1674–1676.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.

Li WZ, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**(13): 1658–1659.

Li YM, Zhang TX, Qi L, et al. 2018. Microbiota changes in the musk gland of male forest musk deer during musk maturation. *Frontiers in Microbiology*, **9**: 3048.

Liu CM, Hong TT, Yu L, et al. 2023. Single-nucleus RNA and ATAC sequencing uncovers the molecular and cellular characteristics in the musk gland of Chinese forest musk deer (*Moschus berezovskii*). *FASEB Journal*, **37**(2): e22742.

Liu CM, Hong TT, Yu L, et al. 2024a. Single-nucleus multiomics unravels the genetic mechanisms underlying musk secretion in Chinese forest musk deer (*Moschus berezovskii*). *International Journal of Biological Macromolecules*, **279**: 135050.

Liu CM, Hong TT, Zhao CC, et al. 2024b. Single-nucleus transcriptomics and chromatin accessibility analysis of musk gland development in Chinese forest musk deer (*Moschus berezovskii*). *Integrative Zoology*, **19**(5): 955–974.

Liu JY, Hu YR, Yang FK. 2016. Synthesis and optimization of musk ketone. *Journal of Qingdao University of Science and Technology: Natural Science Edition*, **37**(6): 609–612,618. (in Chinese)

Lv SQ, Lei ZX, Yan G, et al. 2022. Chemical compositions and pharmacological activities of natural musk (*Moschus*) and artificial musk: a review. *Journal of Ethnopharmacology*, **284**: 114799.

Manni M, Berkeley MR, Seppey M, et al. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, **38**(10): 4647–4654.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, **27**(6): 764–770.

Martin Carli JF, Dzieciatkowska M, Hernandez TL, et al. 2023. Comparative proteomic analysis of human milk fat globules and paired membranes and mouse milk fat globules identifies core cellular systems contributing to mammary lipid trafficking and secretion. *Frontiers in Molecular Biosciences*, **10**: 1259047.

McGinty D, Letizia CS, Api AM. 2011. Fragrance material review on 3-methyl-1-cyclopentadecanone. *Food and Chemical Toxicology*, **49**(S2): S120–S125.

Meng SQ, Guo J, Li ZY, et al. 2019. Enzymatic cascade biosynthesis reaction of musky macrolactones from fatty acids. *Enzyme and Microbial Technology*, **131**: 109417.

Moon YA, Horton JD. 2003. Identification of two mammalian reductases involved in the two-carbon fatty acyl elongation cascade. *Journal of Biological Chemistry*, **278**(9): 7335–7343.

Murrell B, Weaver S, Smith MD, et al. 2015. Gene-wide identification of episodic selection. *Molecular Biology and Evolution*, **32**(5): 1365–1371.

Nightingale R, Reehorst CM, Vukelic N, et al. 2024. *Ehf* controls mammary alveolar lineage differentiation and is a putative suppressor of breast tumorigenesis. *Developmental Cell*, **59**(15): 1988–2004. e11.

Novak EM, Innis SM. 2011. Impact of maternal dietary n-3 and n-6 fatty acids on milk medium-chain fatty acids and the implications for neonatal liver metabolism. *American Journal of Physiology-Endocrinology and Metabolism*, **301**(5): E807–E817.

Panwar B, Menon R, Eksi R, et al. 2016. Genome-wide functional annotation of human protein-coding splice variants using multiple instance learning. *Journal of Proteome Research*, **15**(6): 1747–1753.

Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, **21**(5): 676–679.

Principe S, Kim Y, Fontana S, et al. 2012. Identification of prostate-enriched proteins by in-depth proteomic analyses of expressed prostatic secretions in urine. *Journal of Proteome Research*, **11**(4): 2386–2396.

Shi X, Zeng DJ, Zhao GJ, et al. 2023. Correlation analysis between muskrat (*Ondatra zibethicus*) musk and traditional musk. *Animals*, **13**(10): 1678.

Shirasu M, Yoshikawa K, Takai Y, et al. 2014. Olfactory receptor and neural pathway responsible for highly selective sensing of musk odors. *Neuron*, **81**(1): 165–178.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**(9): 1312–1313.

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, **19**(S2): ii215–ii225.

Stuart T, Butler A, Hoffman P, et al. 2019. Comprehensive integration of single-cell data. *Cell*, **177**(7): 1888–1902. E21.

Sun FF, Zhang ZL, Liang D, et al. 2021. GC-MS analysis of the differences in the chemical components of muskrat and musk. *Natural Product Research and Development*, **33**(10): 1643–1648,1712. (in Chinese)

Suzuki R, Shimodaira H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, **22**(12): 1540–1542.

Theis KR, Venkataraman A, Dycus JA, et al. 2013. Symbiotic bacteria appear to mediate hyena social odors. *Proceedings of the National Academy of Sciences of the United States of America*, **110**(49): 19832–19837.

Twigger AJ, Engelbrecht LK, Bach K, et al. 2022. Transcriptional changes in the mammary gland during lactation revealed by single cell sequencing of cells from human milk. *Nature Communications*, **13**(1): 562.

Vurture GW, Sedlazeck FJ, Nattestad M, et al. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics*, **33**(14): 2202–2204.

Wang J, Xing H, Qin XM, et al. 2020. Pharmacological effects and mechanisms of muscone. *Journal of Ethnopharmacology*, **262**: 113120.

Wang K, Wang XQ, Zheng SN, et al. 2018. iTRAQ-based quantitative analysis of age-specific variations in salivary proteome of caries-susceptible individuals. *Journal of Translational Medicine*, **16**(1): 293.

Wang T, Yang MS, Shi X, et al. 2025. Multiomics analysis provides insights into musk secretion in muskrat and musk deer. *GigaScience*, **14**: giaf006.

Ward JP, van Dorp DA. 1981. The animal musks and a comment of their biogenesis. *Experientia*, **37**(9): 917–922.

Watkins PA, Maiguel D, Jia ZZ, et al. 2007. Evidence for 26 distinct acyl-coenzyme A synthetase genes in the human genome. *Journal of Lipid Research*, **48**(12): 2736–2750.

Wolock SL, Lopez R, Klein AM. 2019. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Systems*, **8**(4): 281–291. e9.

Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biology*, **20**(1): 257.

Wu TZ, Hu EQ, Xu SB, et al. 2021. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *The Innovation*, **2**(3): 100141.

Xu ZX, Li F, Liu Q, et al. 2024. Chemical composition and microbiota changes across musk secretion stages of forest musk deer. *Frontiers in Microbiology*, **15**: 1322316.

Yang JM, Peng GF, Shu F, et al. 2021. Characteristics of steroidogenesis-related factors in the musk gland of Chinese forest musk deer (*Moschus berezovskii*). *The Journal of Steroid Biochemistry and Molecular Biology*, **212**: 105916.

Ye XL, Li YY, González-Lamuño D, et al. 2024. Role of ACSBG1 in brain lipid metabolism and X-linked adrenoleukodystrophy pathogenesis: insights from a knockout mouse model. *Cells*, **13**(20): 1687.

Young MD, Behjati S. 2020. SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. *GigaScience*, **9**(12): giaa151.

Yu ZT, Morrison M. 2004. Improved extraction of PCR-quality community DNA from digesta and fecal samples. *BioTechniques*, **36**(5): 808–812.

Yuan Y, Sun DM, Qin T, et al. 2022. Single-cell transcriptomic landscape of the sheep rumen provides insights into physiological programming development and adaptation of digestive strategies. *Zoological Research*, **43**(4): 634–647.

Zhang Z, Schwartz S, Wagner L, et al. 2000. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, **7**(1-2): 203–214.

Zheng CL, Jie H, Zhou L, et al. 2021. Tissue structure and developmental characteristics analysis of sachet of *Moschus berezouskii. Heilongjiang Animal Science and Veterinary Medicine*, (16): 126–129, 154–156. (in Chinese)

Zhou WL, Nie YG, Hu YB, et al. 2019. Seasonal and reproductive variation in chemical constituents of scent signals in wild giant pandas. *Science China Life Sciences*, **62**(5): 648–660.

Zhou WL, Qi DW, Swaisgood RR, et al. 2021. Symbiotic bacteria mediate volatile chemical signal synthesis in a large solitary mammal species. *The ISME Journal*, **15**(7): 2070–2080.