Molecular basis of reproductive plasticity of termites under environmental stresses and prediction of colony fecundity using machine learning

Chen-Xu Ye [1], Wen-Xiu Zhang [2,3] , Xin-Yi Liu [2,3], Tian-Ge Zhang [2,3], Yuan-Yan Tian [2,3], Yan-Li Yang[2,3], Noor-Us Saba[2,3], Lian-Xi Xing[2,3], Sheng Li[1], Xiao-Hong Su[2,3,*]

[1] Guangdong Provincial Key Laboratory of Insect Developmental Biology and Applied Technology, Institute of Insect Science and Technology & School of Life Sciences, South China Normal University, Guangzhou, Guangdong 510631, China

[2] Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, Northwest University, Xi'an, Shaanxi 710069, China

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

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

Foundation items: This work was supported by the National Natural Science Foundation of China (31870389, 32471563), Natural Science Basic Research Program of Shaanxi (2019JM356)

ABSTRACT

The reproductive plasticity of social insects provides colonies with tremendous flexibility to respond to environmental changes and adapt to new habitats. However, understanding how gene networks regulate this plasticity to generate complex social phenotypes remains a challenge. Here, we propose an "inhibition-activation-reinhibition" model for the reproductive transformation of female workers using three species of *Reticulitermes* termites. Through transcriptomic analysis of female workers during this dynamic cycle, we identified functional genes involved in their reproductive transformation. We found that gene expression profiles in female workers respond strongly to queen loss. Furthermore, we confirmed that under severe environmental stress that threatened colony survival, the workers did not activate the gene regulatory networks associated with reproductive transformation. We also demonstrated that *piggyBac* and *tigger* transposable element-derived genes in the worker brains cooperatively promoted reproductive transformation. Metabolomic changes in female workers were linked to stress signals from queens and nest environment, primarily involving fatty acid and amino acid metabolism pathways. Finally, we constructed seven machine learning models to elucidate gene expression profiles related to reproductive plasticity and to predict the presence and loss of female reproductives (queens, an indicator of colony fecundity). Among these, the deep neural network models showed the highest accuracy and practicability for analyzing gene regulatory networks and predicting caste. We propose that female workers, as indicators, carry extensive colony-specific information at the transcriptomic level. These findings provide valuable insights and innovative methodologies for understanding reproductive plasticity in social insects under environmental stress.

Keywords: Termite; Reproductive plasticity; Transcriptomic profiling; Metabolome profiling; Machine learning

INTRODUCTION

Eusociality is an important step in animal evolution that increases the level of biological complexity, in which the reproductive division of labour is the most significant evolutionary transitions of eusociality

(Shigenobu et al., 2022). Reproductive plasticity provides colonies with tremendous flexibility to respond to environmental and social pressures, forming the basis for their evolutionary and ecological success (Ashton et al., 2019; Harrison et al., 2018). The lower termite *Reticulitermes* is an excellent model system to study the mechanisms of reproductive plasticity. Workers, unlike the soldier and reproductive castes, are pluripotent individuals. A worker can develop in one of three ways: remain a worker, differentiate into a sterile soldier, or become a neotenic reproductive that develops in the absence of reproductives for colony growth (Elliott & Stay, 2008; Korb & Hartfelder, 2008; Su et al., 2017). Interestingly, female workers are unable to produce mature eggs unless they moult into conspicuous reproductive females, whereas male workers can copulate with queens (Fujita & Watanabe, 2010; Wu et al., 2013; Su et al., 2017). Matsuura et al. (2018) reported that female neotenic reproductives derived from workers were observed, but no male neotenic reproductive derived from workers was found in field colonies of *R. speratus*. The workers actively exploit new food resources outside the natal nest and move to a new nest site when food supplies are low (Korb & Hartfelder, 2008), triggering the differentiation of neotenic reproductives in the new nest site. *Reticulitermes* can utilize budding dispersal by worker groups. A budding group of 15 *R. santonensis* workers can survive and establish a new colony (Krajewski, 2023). This is a highly efficient way to establish new colonies and invade new or marginal habitats (Figure1a).

The transformation of female workers to reproductives is maintained via a dynamic cycle of "inhibition-activation-reinhibition" due to the influence of pheromones from female reproductives (queens). Normally, the transformation of female workers into reproductives is inhibited by the queens. After a queen dies (or is absent), the reproductive transformation of female workers is activated. Subsequently, this transformation is inhibited again after a new female reproductive emerges in the colony (Elliott & Stay, 2008; Su et al., 2017). Caste differentiation to form distinct phenotypes is accomplished through the expression of different gene sets in response to environmental stimuli (Matsuura et al., 2018; Saiki et al., 2022; Takata et al., 2025). The pattern of the expression of genes involved in the differentiation of workers into reproductives was revealed by comparing the transcriptomes of female workers and neotenic reproductives of *R. labralis* (Ye et al., 2019). A module of co-expressed genes that characterizes queens compared to workers has been identified, supporting the hypothesis of a conserved genetic toolkit for termite queens with gradual linear development (Lin et al., 2021; Lin et al., 2024). Juvenile hormones (JH), insulin/insulin-like growth factor (IGF) signaling, and nutrition are related to different gene expression profiles during caste differentiation. Lower and higher JH synthesis induce the differentiation of workers into reproductives and soldiers, respectively (Elliott & Stay, 2008; Korb, 2015; Oguchi & Miura, 2024; Rau et al., 2023; Saiki et al., 2022; Veenstra, 2023a). In American cockroach, insulin/IGF signaling regulates JH biosynthesis to promote vitellogenesis, and different neuroendocrine cell types in the brains produce their own specific IGF-related peptides (Zhu et al., 2020; Veenstra, 2023b). Recently, in termites, the efficient regulation of transposable elements (TEs) was found to promote reproductive fitness even at advanced ages, and their movement has been linked to eusocial adaptation through changes in gene expression in different castes (Berger et al., 2022; Elsner et al., 2018; Post et al., 2023). TE-derived genes facilitate developmental adaptation to rapid environmental changes (Baduel & Quadrana, 2021; Chen et al., 2020; Etchegaray et al., 2021). However, whether TE-derived genes promoting reproductive plasticity in workers remains unclear. Moreover, shifts in the levels of specific metabolites have been associated with phenotypic plasticity, and the production of secondary metabolites is controlled by the transcriptional regulatory network in response to genetic and environmental changes. Metabolome-transcriptome analysis revealed that larvae capable of anhydrobiosis adopt a unique metabolic strategy to cope with extreme desiccation. This strategy is particularly important for enabling successful recovery upon rehydration (Liska et al., 2023; Ryabova et al.,

2020; Jönsson et al., 2025). Currently, research on phenotypic plasticity is shifting from an early focus on single-gene pathways towards a systemic exploration of large-scale genomic and network-based regulation (Taylor et al., 2021; Wyatt et al., 2023). However, the gene regulatory networks underlying reproductive plasticity remains largely unclear. Analyses of the molecular foundations of phenotypic plasticity are challenging, especially in the case of complex social phenotypes (Taylor et al., 2021).

Obtaining information from the inside of social insect colonies, such as colony size, caste structure, and the presence or absence of a queen, has always been challenging. Current methods for studying termite colonies primarily rely on physical signals, mathematical modeling, or nest dissection (Nanda et al., 2019; Patel et al., 2020; Takata et al., 2023; Thorne et al., 1996). However, molecular approaches for predicting castes and colony development remain scarce. Recently, based on brain transcriptomes of social wasps, a machine-learning algorithm - support vector machine model (SVM) has been constructed to predict castes and analyse gene expression profiles in workers following the queen loss (Taylor et al., 2021; Wyatt et al., 2023). Gordon, an ant ecologist and behaviorist, proposed that there are clear similarities between a brain and social insect colonies. In both systems, local information is exchanged between simple units, leading to complex results at the level of the entire system (Gordon et al., 1992), and collective behaviour, like any other phenotypic trait, evolves in a dynamic environment (Gordon, 2019). For termite colonies, the absence or scarcity of reproductives represents a critical environmental change that requires rapid resolution. However, the number of isolated workers affects the speed of neotenic reproductive differentiation (Elliott & Stay, 2008). This paradox arises because as the number of isolated workers decreases, indicating a more severe crisis for the colony, there is an urgent need for supplementary reproductives. Despite this urgency, the speed at which neotenic reproductives are produced is slow, raising further questions about the underlying mechanisms. This phenomenon bears a striking resemblance to neural networks; the fewer neurons a neural network contains, the simpler the problems it can process. This implies that, in eusocial insects, individuals exist as neurons within the “brain of the colony”. The functions performed by a colony are analogous to the principles of operation of neural networks (Pinter-Wollman et al., 2012; Razin, 2013; Waters & Fewell, 2012). Indeed, processes such as sensory processing, learning, motor control and social behaviour can be simulated using neural networks (Cowley et al., 2024; Rappa & Nawrot, 2020). In a eusocial paper wasp, changes in gene expression may occur in workers without outwardly detectable phenotypic alterations, such as in external morphology, ovarian development or behavior (Taylor et al., 2021). Since transcriptomic changes can occur before phenotype appearance, transcriptomic analysis can be used to predict phenotype changes. Utilizing machine learning models to predict and analyze gene regulatory networks in response to environmental changes will further deepen our understanding of how gene expression regulates phenotypic plasticity (Taylor et al., 2021; Wyatt et al., 2023; Jönsson et al., 2025).

Of the nearly 3 000 species of termites, the genus *Reticulitermes* (Rhinotermitidae) occupies an important evolutionary position, exhibiting intermediate social complexity between higher (Termitidae) and lower (all other families) termites, and *Reticulitermes* is a common termite group worldwide, with ecological, and economic relevance (Shigenobu et al., 2022). Natural and wild colonies of *R. labralis*, *R. chinensis* and *R. aculabialis* can produce female reproductives that are transformed from workers. Here, the differentiation of female reproductives was induced in groups of isolated workers of *R. labralis* and *R. chinenesis* about 4 weeks under laboratory conditions. However, the isolated workers of *R. aculabialis* did not transform into reproductives under the same experimental conditions. In this species isolated workers died after 16 weeks. Based on our findings, we established an “inhibition-activation-reinhibition” model of female worker reproductive transformation using the three

species, and employed transcriptomic profiling (RNA-sequencing), metabolome profiling, RNA interference (RNAi) and machine learning modelling using female workers from natal colonies with female reproductives (W), isolated female workers (IW), and isolated female workers after female neotenic reproductives emerged in isolated groups (NRW) from the three *Reticulitermes* species. We identified the functional genes involved in the transformation process and found that the transformation of workers into reproductives in *R. aculabialis* was not triggered under the same isolation conditions. We found that TE-related genes were associated with reproductive plasticity of female workers. Moreover, analyses of metabolome changes revealed key components and metabolic adaptations. More importantly, we confirmed that when female workers experience severe environmental stresses that affect their survival, they do not initiate reproductive transformation, even if there is no female reproductives in their colonies. Finally, based on the gene expression profiles, we constructed seven machine learning models, including the random forest, support vector machine (SVM), k-nearest neighbour (KNN), linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), gradient boosting machine (GBM), and deep neural network models, to demonstrate how numerous genes co-regulate the reproductive plasticity of workers and to estimate female reproductives (queen, an indicator of colony fecundity) . We found that among these models, the deep neural network models exhibited the highest accuracy and practicability. By analyzing an assembly of multiple GO functions through tensor reshaping, we discovered that each molecular function used to predict castes was not independent but involved spatial interactions. Deep neural network models provide a powerful methodology for revealing complex gene regulatory networks underlying caste differentiation, reproductive plasticity, and interactions among different castes in social insects.

MATERIALS AND METHODS

Differentiation of isolated workers into neotenic reproductives

R. labralis, *R. chinensis* and *R. aculabialis* are widely distributed in China and are well-studied termite species**.** Mature colonies of three species (three colonies per species) were collected from the wild in Xi'an and Chengdu, China. The nest wood was brought to the laboratory and covered with wet soil to maintain the natal colonies. To induce the differentiation of female workers into neotenic reproductives (NRs) and establish new colonies, groups of workers were isolated from their natal colonies. Our previous studies demonstrated that each late-instar female worker retains the potential to develop into a reproductive (Su et al., 2017). Each isolated group consisted of fifty sixth-instar (late-instar) workers and two soldiers. The groups were placed in Petri dishes containing moist sawdust and maintained in darkness at room temperature (Figure 1b, c, Supplementary Figure S1). Termite sex was identified based on the morphology of the seventh sternite. After about 4 weeks, NRs appeared in the isolated groups of *R. labralis* and *R. chinensis*. The transformation from workers to neotenic reproductives occurs through molting, and there are significant morphological differences between workers and neotenic reproductives.

RNA extraction, library construction and sequencing

Total RNA of the heads of female workers from natal colony (W), isolated female workers for 21 days (IW) and isolated female workers after female reproductives emerged in isolated groups (NRW) (25 female worker heads per sample; three biological replicate**s** per species per condition) was extracted using Trizol reagent kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol, respectively. RNA quality was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and checked using RNase-free agarose gel electrophoresis. After total RNA was extracted, eukaryotic mRNA was enriched by Oligo (dT) beads (Epicentre, Madison, WI, USA). Then the enriched mRNA was fragmented into short fragments using fragmentation buffer and reverse transcribed into cDNA with

random primers. Second-strand cDNA was synthesized by DNA polymerase I, RNase H, dNTP and buffer. Then the cDNA fragments were purified with QiaQuick PCR extraction kit (Qiagen, Venlo, The Netherlands), end repaired, dA added, and ligated to Illumina sequencing adapters. The ligation products were size-selected by agarose gel electrophoresis, PCR amplified, and sequenced using Illumina HiSeq^{TM} 4000 by Gene Denovo Biotechnology Co. (Guangzhou, China).

Filtering of clean reads and denovo assembly: Reads obtained from the sequencing machines included raw reads containing adapters or low quality bases which would affect the following assembly and analysis. Thus, to get high quality clean reads, reads were further filtered by fastp (version 0.18.0) with parameters: quality score cutoff = 20 (-q 20), maximum unqualified read percentage = 50% (-u 50), maximum allowed N bases = 15 (-N 15), and minimum read length = 50 bp (-l 50) (Chen et al. 2018). Transcriptome denovo assembly was carried out with short reads assembling program-Trinity (version 2.8.4) using KMER_SIZE=31, min_kmer_cov=36, and min_glue=36 (Grabherr et al., 2011). Assembly quality assessment was performed via BUSCO (v3) to evaluate completeness against lineage-specific benchmarks (Simão et al., 2015).

Unigene expression analysis: The unigene expression was calculated and normalized to RPKM (reads per kb per million reads). Gene expression quantification (raw counts and RPKM) was calculated using RSEM (v1.2.19) (Li & Dewey, 2011). Differential gene expression analysis was implemented in DESeq (v1.20.0) (Love et al., 2014) with significance thresholds set at |log2(fold change)|>1 and false discovery rate (FDR)<0.05.

Annotation of unigenes: Annotation of unigenes includes protein functional annotation, pathway annotation, COG/KOG functional annotation and Gene Ontology (GO) annotation. To annotate the unigenes, we used the BLAST+ software (version 2.6.0+) with the BLASTx program (http://www.ncbi.nlm.nih.gov/BLAST/) against the following databases under an E-value threshold of 1e-5: the NCBI non-redundant protein (Nr) database (version 20200514; http://www.ncbi.nlm.nih.gov), the Swiss-Prot protein database (version 20200514; http://www.expasy.ch/sprot), the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000) database (release 93.0; http://www. genome. jp/ kegg), the Clusters of Orthologous Groups (COG) database (version 20150724; http://www.ncbi.nlm.nih.gov/COG), the Eukaryotic Orthologous Groups (KOG) database (version 20150724; http://www.ncbi.nlm.nih.gov/COG), and the Gene Ontology (GO) database (version 20160301; http://geneontology.org) (Conesa et al., 2005; Ashburner et al., 2000). Protein functional annotations were assigned based on the best alignment results.

Protein coding sequence (CDS) prediction: Unigenes were aligned by BLASTx (evalue<0.00001) to protein databases in priority of nr, Swissprot, KEGG and COG/KOG. The best alignment results were chosen to decide the sequence direction of unigenes. When a unigene could not be aligned to any of these protein databases, protein coding sequence and sequence direction would be confirmed using TransDecoder program (version 5.5.0).

Gene set enrichment analysis (GSEA): We performed gene set enrichment analysis using software GSEA (version 2.2.4) and the Molecular Signatures Database (MSigDB) to identify whether a set of genes in specific GO terms/pathways terms shows significant differences in W-IW and IW-NRW. We input gene expression matrix and rank genes by SinaltoNoise normalization method. Enrichment scores and p value was calculated in default parameters.

Quantitative real-time PCR (qPCR)

For gene expression analysis of 56 TE-related genes for the female workers in natal colony (W), isolated female workers (IW), isolated female workers after the female reproductives emerged in isolated groups

(NRW) and female reproductives in isolated groups (NR). The total RNA of the heads of female workers (25 heads per sample) was extracted using RNAiso Plus reagent (Takara Biomedical Technology (Beijing) Co., Ltd., Beijing, China) and quantified using a NanoReady F-1100 (Life Real, China). First-strand cDNA was synthesized from total RNA using the PrimeScript™ RT Master Mix (Takara Biomedical Technology (Beijing) Co., Ltd., Beijing, China). The gene-specific primers were designed by Primer 5.0. The primer sequences of *beta-actin* were obtained from *R. flavipes* (Scharf et al., 2005; Zhou et al., 2006). The relative quantitative of transcripts was performed using a 2×TB Green Premix Ex Taq^{TM} II (Tli RNaseH Plus) (Takara Biomedical Technology (Beijing) Co., Ltd., Beijing, China) with the QuantGene 9600 Fluorescent Quantitative Detection System (Bioer Technology, Zhejiang, China). The relative gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method. All qPCR experiments were repeated in three biological and three technical replications. The qPCR primer sequences for 56 TE-related genes were shown in Supplementary Data (Supplementary Table S19).

RNAi bioassays

To synthesize double-stranded RNA (dsRNA), we designed the primers using the Primer Premier 5.0 program and a T7 RNA polymerase promoter was added to the partial cDNA sequences using PCR by including the T7 promoter sequence at the 5'-end of either of the gene-specific primers, based on the cDNA sequences of *TIGD1*, *PGBD4*, *TC3a* and *GFP* (Supplementary Table S20). The *GFP* gene was used as a control dsRNA. *TIGD1*, *PGBD4* and *Tc3a* were identified in both *R. labralis* and *R. chinensis*, and qPCR results showed that the expression levels of *TIGD1*, *PGBD4* and *Tc3a* were significantly higher in IW compared to W, NRW and NR in both *R. labralis* and *R. chinensis*. The dsRNAs were synthesized from PCR product templates with the T7 RiboMAX™ Express RNAi System (Promega (Beijing) Biotech Co., Ltd, Beijing, China). The dsRNA quality was examined by electrophoresis in a 1% native agarose gel, and the concentration was quantitated by measuring absorbance at a wavelength of 260 nm with a NanoReady F-1100 (Life Real, Zhejiang, China).

dsRNA coloured with Fast Green FCF was fed to workers that were isolated for 19 days by using a PLI-100A Pico-Injector (Warner Instruments, MA, USA) equipped with a custom-pulled borosilicate glass needle. A worker was fixed upright with Blu-Tack. After dispensing a droplet of known volume and letting it hang on the tip of the needle, the worker was forced to ingest the droplet. On day 19 of isolation, each worker was force-fed with 200 ng of either ds*TIGD1*, ds*PGBD4*, ds*TC3a,* ds*GFP* or dsCocktail that targeted the three transcripts (*TIGD1* + *PGBD4* + *TC3a*) and put back to the colony. After two days, female workers were picked out to extract total RNA of heads. cDNA was synthesized from the total RNA employing a PrimeScript™ RT Master Mix (Takara Biomedical Technology (Beijing) Co., Ltd., Beijing, China), and was used as the template for qPCR. For morphological observation, we used the same RNAi method and recorded the total number of days required to produce neotenic reproductives in the isolated colonies. Three replicates were performed per treatment.

Metabolomic profiling

We performed untargeted metabolomic profiling of the W and IW heads of *R. labralis* using LC-tandem mass spectrometry (LC-MS/MS). Thirty heads of female workers were used for each sample (six biological replicates).

Metabolites extraction: The samples were ground with liquid nitrogen, the homogenate was resuspended in pre-chilled 80% methanol by vortexing. The samples were centrifuged at 15 000×g, 4°C for 20 min. Some of the supernatant was diluted to a final concentration of 53% in methanol by LC-MS grade water. The samples were subsequently transferred to a fresh Eppendorf tube and then were centrifuged at

15 000×g, 4°C for 20 min. Finally, the supernatant was injected into the LC-MS/MS system analysis (Want et al., 2013).

Ultra-high performance (UHP)LC-MS/MS analysis: UHPLC-MS/MS analyses were performed using a Vanquish UHPLC system (Thermo Fisher Scientific, Germany) coupled with an Orbitrap Q Exactive™ HF mass spectrometer or an Orbitrap Q Exactive™HF-X mass spectrometer (Thermo Fisher Scientific) at Novogene Co., Ltd. (Beijing, China). Samples were injected onto a Hypersil Gold column (100 mm × 2.1 mm, 1.9 μm) using a 12-min linear gradient at a flow rate of 0.2 mL/min. Eluents A (0.1% FA in water) and B (methanol) were used for the positive polarity mode. The eluents used for the negative polarity mode were eluents A (5 mM ammonium acetate, pH9.0) and B (methanol). The solvent gradient was set as follows: 2% B, 1.5 min; 2%85% B, 3 min; 85%-100% B, 10 min; 100%-2% B, 10.1 min; 2% B, 12 min. Q Exactive™ HF mass spectrometer was operated in positive/negative polarity mode with spray voltage of 3.5 kV, capillary temperature of 320°C, sheath gas flow rate of 35 psi and aux gas flow rate of 10 L/min, S-lens RF level of 60, Aux gas heater temperature of 350°C.

Data processing and metabolite identification: The raw data files generated via UHPLC-MS/MS were processed using Compound Discoverer 3.3 (CD3.3, ThermoFisher) to perform peak alignment, peak picking, and quantitation for each metabolite. The main parameters were set as follows: peak area corrected with the first QC, actual mass tolerance, 5 ppm; signal intensity tolerance of 30%, and minimum intensity. The normalised data were then used to predict the molecular formula based on additive ions, molecular ion peaks, and fragment ions. And then peaks were matched with the mzCloud (https://www.mzcloud.org/), mzVault and Masslist database to obtain the accurate qualitative and relative quantitative results. Statistical analyses were performed using the statistical software R (R version R-3.4.3), Python (Python 2.7.6 version) and CentOS (CentOS release 6.6). When data were not normally distributed, they were standardized according to the formula: sample raw quantitation value / (the sum of sample metabolite quantitation value / the sum of QC1 sample metabolite quantitation value) to obtain relative peak areas; compounds whose CVs of relative peak areas in QC samples were greater than 30% were removed, and finally the metabolites' identification and relative quantification results were obtained.

The metabolites were annotated using the KEGG (https://www.genome.jp/kegg/pathway.html), HMDB (https://hmdb.ca/metabolites), and LIPIDMaps (http://www.lipidmaps.org/) databases. Principal components analysis (PCA) and partial least squares discriminant analysis were performed using metaX (flexible and comprehensive software for processing metabolomic data) (Wen et al., 2017). We applied univariate analysis (*t*-test) to calculate the statistical significance (*P*-value). The metabolites with $VIP>1$ and $P\text{-value}<0.05$ and $FC\geq 2$ or ≤ 0.5 were considered to be differential metabolites. The correlation between differential metabolites was analysed using cor() in the R language (method=Pearson). Statistically significant correlations between differential metabolites were calculated using the cor.mtest() in R language. Statistical significance was set at $P<0.05$, and correlation plots were plotted using the corrplot package in R language. The functions of these metabolites and metabolic pathways were studied using the KEGG database, and metabolic pathway enrichment of differential metabolites was performed; when ratios were satisfied by x/n>y/N (where x=the number of significant metabolites observed in a specific metabolic pathway; n=the total number of metabolites in that metabolic pathway; y=the number of significant metabolites observed in all metabolites; N=the total number of all metabolites.), metabolic pathways were considered as enriched, and were considered as statistically significant enrichment.

Immunohistochemical localization

Paraffin sections (7 μm) of heads of IW were dewaxed using xylene, rehydrated through a graded alcohol series, and endogenous peroxidase activity was suppressed using hydrogen peroxide block at room

temperature. For antigen retrieval, the slides were treated with 0.01 M citrate buffer. Following the Streptavidin-Biotin Complex (SABC) immunohistochemistry protocol, the slides were treated with 5% BSA blocking solution for 30 min at 37°C. The primary antibody, PGBD4 (ab122675, Abcam, UK) and TIGD1 (NBP2-47606, Novus Biologicals, USA) rabbit polyclonal antibody, were applied to the sections for 60 min at 37°C, respectively. A biotinylated goat anti-rabbit IgG secondary antibody was applied for 20 min at 37°C, and then SABC (Boster Biological Technology Co., Ltd, China) was applied for 30 min at 37°C. After the DAB chromogenic reaction, slices were counterstained with haematoxylin. For the negative control, the primary antibody was replaced with PBS.

Data processing for machine learning models

Transcriptome denovo assembly was carried out with the short-read assembling program Trinity (version 2.12.0). Protein Coding Sequence (CDS) was confirmed using the TransDecoder program (version 5.5.0). Bowtie2 (version 2.4.4) and Samtools (version 1.13) were used to obtain the raw counts of the CDS of the three species. Gene Ontology (GO) annotation was processed by using the Emapper and the eggNOG database.

We used the rpkm() function of the edgeR package (version 3.34.0) in R to calculate RPKM values, and the clusterProfiler package (version 4.0.5) to perform GSEA. With the obtained GO results, in order to avoid the situation where a GO contained only a few unigenes and thus affected the generalization of the model, the GO entries with unigene numbers<20 were removed.

To construct the dataset, a dataframe was created, with the number of rows equal to the number of samples that were to be generated, and the number of columns equal to the number of GOs that had been selected. A random unigene RPKM value was taken from each GO and combined as a sample value. This implied that if 160 GOs were selected to form the dataset, and each GO contained at least 20 unigenes, then at least 20^{160} different samples could be generated. In this study, 600 000 random samples were generated as the total dataset.

Machine learning model design

A variety of statistical learning models were constructed using different methodologies. Initially, the randomForest model was built employing the randomForest() function from the randomForest package (version 4.6-14), setting the number of trees to 500 (Breiman, 2001; Liaw & Wiener, 2022). Subsequently, the SVM model was developed with the e1071 package (version 1.7-8), where the tune.svm function was utilized to identify the optimal gamma value (1e-3) and cost value (5), followed by the construction of the model using the svm function (Cortes & Vapnik, 1995; Meyer, 2021). Furthermore, the KNN model was formulated using the caret package (version 6.0-88), with the train function aiding in the parameter search to ascertain the value of k (k=33), and the knn3 function employed for model building (Cover & Hart, 1967; Kuhn, 2008). Additionally, LDA and QDA models were generated respectively using the lda and qda functions from the MASS package (version 7.3-54) (Fisher, 1936; Hastie, 2009; Venables & Ripley, 2022). A GBM model was also assembled with the H2O package (version 3.32.1.3), utilizing the h2o.grid() function to optimize learn rate (0.25), max depth (8), and ntrees (100), followed by the h2o.gbm() function for model creation (Friedman, 2001; LeDell et al., 2021). Lastly, deep neural network models were assembled utilizing the keras package (version 2.9.0) (Allaire, 2022).

RESULTS

Research model of reproductive plasticity of female workers

Based on the experiments of the transformation of female workers into neotenic reproductives (NRs) using

three *Reticulitermes* species, a model of "inhibition-activation-reinhibition" of reproductive plasticity of workers was established (Figure 1f; Supplementary Figure S1). The assay to induce reproductive differentiation was performed using isolated workers of *R. labralis*, *R. chinensis* and *R. aculabialis* (a group of fifty workers and two soldiers). In a isolated groups of *R. labralis* and *R. chinensis*, the first female reproductives appeared about four weeks, and then the presence of female reproductive inhibited other female workers developing into reproductives. Obviously, the transformation of workers into reproductives was in a dynamic cycle of "inhibition- activation-reinhibition". Surprisingly, after 16 weeks, no female reproductives appeared in the isolated groups of *R. aculabialis* and the isolated workers eventually died. Although it is common for female workers to transform into reproductives in natural colonies of *R. aculabialis* (Figure 1b), the isolated workers did not initiate reproductive transformation under the experimental conditions. After the isolated workers of *R. labralis* and *R. chinensis* developed into female reproductives, oocyte development was activated (Figure 1c-e), indicating that these workers are highly adaptable to new environments, and they can actively initiate caste reconstruction, which has great significance for colony survival and expansion. This finding provides us with a perfect optimization model, "*R. labralis* and *R. chinensis*" vs *R. aculabialis*, for analyzing the molecular mechanisms of reproductive plasticity of workers.

Transcriptomic pattern for "inhibition-activation-reinhibition" in the transformation of workers to reproductives

The transcriptomes of W, IW, and NRW of the three *Reticulitermes* species were sequenced and analysed. Based on their clean reads, 112 721 unigenes in *R. labralis*, 102 226 unigenes in *R. chinensis* and 105 516 unigenes in *R. aculabialis* were assembled. The average length was 10 561 073 and 1 060 bp, respectively. The N50 length was 2 603, 2 631 and 2 812 bp, respectively. In the W-IW process, of the differentially expressed genes (DEGs), 551 and 316 genes were upregulated and downregulated in *R. labralis*, respectively. 773 and 937 genes were upregulated and downregulated in *R. chinensis*, respectively. 1 180 and 2 084 genes were upregulated and downregulated in *R. aculabialis*, respectively. In the IW-NRW process, There were 2 289 and 758 DEGs in *R. labralis* and *R. chinensis*, respectively, indicating that the responses of the female workers of two species to new environmental stresses are different after the emergence of female reproductives. To investigate changes in biological function associated with the transition from "W-IW-NRW" in *R. labralis* and *R. chinensis*, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses using gene set enrichment analysis (GSEA). During the W-IW transition in both species, 89 and 100 of the top 100 significantly enriched GO terms and 92 and 82 of the top 100 significant KEGG pathways were upregulated in *R. labralis* and *R. chinensis*, respectively, a pattern opposite to that observed in *R. aculabialis* (Figure 2a; Supplementary Tables S1-S4). In the IW-NRW transition, 100 and 94 of the top 100 GO terms and 100 and 96 of the top 100 KEGG pathways were downregulated, respectively (Figure 2a; Supplementary Tables S5-S8). These results indicate that, in *R. labralis* and *R. chinensis*, the majority of the most significantly affected GO terms and KEGG pathways in female workers were upregulated in the absence of reproductives within the colony. Conversely, when reproductives were present, these same functional categories tended to be downregulated.

In *R. labralis* and *R. chinensis*, when considering all the GO functions and KEGG pathways identified by GSEA, 94.8% and 92.6% of the GO functions upregulated in W-IW were downregulated in NRW (Figure 2e), and 91.7% and 88.7% of the KEGG pathways upregulated in W-IW were downregulated in NRW, respectively (Figure 2c). These findings indicate that during the W-IW-NRW process, GO functions

and KEGG pathways were upregulated in IW and downregulated in NRW from the same batch. The upregulation of these molecular functions promotes the transformation of IW into reproductives. After NRs appeared in the isolated groups, the inhibition of the transformation of remaining workers (NRW) into reproductives was induced, forming a dynamic cycle of molecular regulation.

Molecular functions of reproductive plasticity of workers in *R. labralis* and *R. chinensis* were highly consistent

Analysis of the combined transcriptome data from *R. labralis* and *R. chinensis* showed that during the W-IW process, 65.5% of the upregulated GO functions in *R. chinensis* were consistent with those in *R. labralis*. Subsequently, during the IW-NRW process, 77.2% of the downregulated GO functions in *R. chinensis* were consistent with those in *R. labralis* (Figure 2b). Notably, nine significantly enriched GO functions were identified from the four datasets (false discovery rate; *FDR*<0.05) (Figure 2d; Supplementary Table S9). This indicates that during the W-IW-NRW process in both species, these GO functions were significantly upregulated in IW, and then downregulated in NRW. Among these, three belonged to cellular component (extracellular matrix, proteinaceous extracellular matrix, and extracellular matrix component), two to molecular function (oxidoreductase activity and peptidase inhibitor activity), and four to the biological process (embryonic heart tube morphogenesis, long-chain fatty acid metabolic process, steroid metabolic process, and hormone metabolic process).

In the W-IW process, 72.7% of the upregulated KEGG pathways in *R. chinensis* were consistent with those in *R. labralis*, and then in the IW-NRW process, 84.7% of the downregulated KEGG pathways in *R. chinensis* were consistent with those in *R. labralis* (Figure 2c). Of the 87 KEGG pathways intersected by the four datasets, 42 KEGG pathways were upregulated in the W-IW process (*NES* > 1, *FDR* <0.05) and downregulated in the NRW (*NES*<−1, *FDR*<0.05) of *R. labralis* and *R. chinensis* (Figure 2e; Supplementary Table S10). Of the 42 KEGG pathways, five pathways were related to signal transduction, including ECM-receptor interaction, Hedgehog signalling pathway, AMPK signalling pathway, FoxO signalling pathway, and TGF-beta signalling pathway, and eight pathways were related to metabolism, including biosynthesis of unsaturated fatty acids, fatty acid metabolism, tryptophan metabolism, fatty acid degradation, oxidative phosphorylation, arginine and proline metabolism, valine, leucine, and isoleucine degradation, and glutathione metabolism. These results demonstrate that the molecular regulatory mechanism of the transformation from female workers to reproductives may be the same in different species of *Reticulitermes*, and this confirms the reproducibility in different species.

Moreover, the expression levels of genes in the FoxO, AMPK, and fatty acid metabolism pathways changed dynamically during the W-IW-NRW process. A total of 22 and 13 genes were screened from the FoxO signalling pathways of *R. labralis* and *R. chinensis*, respectively, including *CAMK/CAMKL/AMPK protein kinase*, *stress-activated protein kinase JNK*, *epidermal growth factor receptor*, and *cell division control protein2*. We screened 36 and 23 genes from the AMPK signalling pathway, including *fatty acid synthase*, *CAMK/CAMKL/AMPK protein kinase*, *translation elongation factor2*, and *stearoyl-CoA 9-desaturase*. A total of 23 and 35 genes were screened from their fatty acid metabolism pathways, respectively, mainly involving *fatty acid synthase*, *acyl-CoA desaturase*, and *elongation of very longchain fatty acids protein* (Supplementary Figure S2; Supplementary Table S11-S16).

We identified two *hemolymph Juvenile hormone binding protein* (*JHBPs*) sequences in *R. chinensis* and three in *R. labralis*. All these *JHBP* genes exhibited similar differential expression patterns: they were downregulated in IW compared with W. JHBPs are carrier proteins that bind to juvenile hormone (JH), and then transport JH to target organs to regulate insect growth and development (Liu et al., 2024). Compared with W, the expression levels of the two *JHBPs* in *R. chinensis* decreased by

40% and 57% in IW, respectively, and the expression levels of the three *JHBPs* in *R. chinensis* decreased by 68%, 71 and 78% in IW, respectively.

Molecular regulation of the transformation of isolated workers to reproductives was not triggered in *R. aculabialis*

To resolve why the isolated workers of *R. aculabialis* did not transform into reproductives, we compared the expression of GO and KEGG pathways in the W-IW of *R. aculabialis* with those of *R. labralis* and *R. chinensis* based on the GSEA results. The numbers of GO terms significantly enriched in the W-IW of *R. labralis* and *R. chinensis* were 107 and 236, respectively, and most of the GO functions were significantly upregulated (*FDR*<0.05, *NES*>0) (Supplementary Table S1, S3). In contrast, our analysis showed that only four GO terms in the W-IW of *R. aculabialis* were significantly enriched and significantly downregulated (*FDR*<0.05, *NES*<0) in the W-IW (Supplementary Table S17). The four enriched GO functions included structural constituents of chitin-based cuticles, structural constituents of cuticles, ligase activity, formation of carbon-oxygen bonds, and cuticle development, indicating a decrease in the structural integrity and function of the cuticle in isolated workers of *R. aculabialis*. The expression levels of cuticle protein genes (*cuticle protein 19 isoform X1*, *cuticle protein 19*, *cuticle protein 19.8*, *cuticle protein 18.6*, *larval cuticle protein A2B*, *cuticle protein 7*, and *cuticle protein 14.6*) and endocuticle structural glycoprotein genes (*SgAbd-5*, *SgAbd-2*, *SgAbd-3*, *SgAbd-8*, and *SgAbd-9*) were significantly lower in IW than in W. The insect cuticle provides protection against mechanical injury, pathogens, and parasites, as well as facilitating sensory perception, muscle attachment points, and other physiological functions (Noh et al., 2017; Ye et al., 2021).

According to GSEA results, the numbers of KEGG pathways significantly enriched in the W-IW of *R. labralis* and *R. chinensis* were 49 and 22 (*FDR*<0.05), respectively (Supplementary Table S2, S4). In contrast, only 9 KEGG pathways were significantly enriched in the W-IW of *R. aculabialis* (*FDR*<0.05, *NES*>0) (Supplementary Table S18). Two of those were associated with signal transduction (Hippo signalling pathway and neuroactive ligand-receptor interaction), and three with metabolism (insect hormone biosynthesis, valine, leucine, and isoleucine degradation, and fatty acid degradation). Through comparative analysis of the three termite species, we found that the GO functions and KEGG pathways associated with the transformation of isolated workers to reproductives in *R. labralis* and *R. chinensis* were not activated in *R. aculabialis* under identical experimental conditions. This supports our finding that the isolated workers of *R. aculabialis* cannot transform into reproductives under the experimental conditions. In addition, in contrast to *R. labralis* and *R. chinensis*, two *hemolymph Juvenile hormone binding protein* (*JHBPs*) sequences in *R. aculabialis* were significantly upregulated in IW compared to W. The expression levels of $JHBP_1$ and $JHBP_2$ in IW were 2.5-fold and 2.6-fold higher than those in W, respectively.

Metabolome change associated with reproductive plasticity of workers

Eight metabolic pathways from the transcriptome analysis were associated with the transition of workers into reproductives; thus, we investigated metabolomic differences between W and IW in *R. labralis* using liquid chromatography coupled to a high-resolution mass spectrometer (LC-MS). Principal component analysis (PCA) was used to determine the variability in the changes in metabolite content in W-IW, showing clear separation between W and IW with high reproducibility among replicates (Figure 3a). In total, 1 065 metabolites (509 positive and 556 negative ions) were identified. Significant changes occurred in 330 metabolite, including 163 positive ions (83 upregulated and 80 downregulated) and 167 negative ions (91 upregulated and 76 downregulated) (fold change *FC*>1.5 or *FC*<0.667, *P*<0.05) (Figure 3c).

The upregulated metabolites belonged to seven categories: (i) organic acids and derivatives, (ii) lipids and lipid-like molecules, (iii) benzenoids, (iv) organic oxygen compounds, (v) nucleosides, nucleotides, and analogues, (vi) phenylpropanoids and polyketides, (vii) organoheterocyclic compounds. Lipids and lipid-like molecules (67 metabolites) were the most abundant category. Thus, these lipids likely play a key role in the IW-to-reproductive transition. Among the upregulated metabolites, the levels of 19 metabolites in IW were more than 5-fold higher than those in W (*FC*>5). Notably, N1-prop-2-ynyl- 4- chlorobenzene-1 –sulfonamide increased 180-fold in IW (*FC*=180.81) (Figure 3b). However, this metabolite lacks classification or functional annotations in databases. Its drastic accumulation suggests this metabolite may represent a previously uncharacterized regulator of insect reproductive physiology.

In W-IW, the top five metabolites downregulated included N'-(4-chlorobenzoyl)-5-[4-(1,3-dithiolan-2 -yl)phenoxy]-2-furohydrazide (*FC*=0.006), D-glutamine (Organic acids and derivatives class, *FC*=0.032), 5-methyluridine (pyrimidine nucleosides class, *FC*=0.020), N-phenylacetylglutamine (amino acids, peptides, and analogues class, *FC*=0.038), gemifloxacin (diazanaphthalenes class, *FC*=0.068). In particular, the level of N'-(4-chlorobenzoyl)-5-[4-(1,3-dithiolan-2-yl)phenoxy]-2-furohydrazide in W was 151-fold higher than that in IW， indicating a drastic decline in IW (Figure 3b). However, the database did not show the classification or biological information of this metabolite, suggesting that it may be a previously unknown functional molecule with physiological functions in animals. Different metabolites had the same or opposite change patterns, showing positive or negative correlations with each other (Figure 3e).

The main biological functions of differential metabolites were determined by KEGG pathway enrichment analysis (Figure 3d). Combined analysis of metabolome and transcriptome revealed that four metabolic pathways were clearly related to the transformation of IW to reproductives: biosynthesis of unsaturated fatty acids, arginine and proline metabolism, tryptophan metabolism, and glutathione metabolism. In addition, fatty acid metabolism was one of the nine GO functions identified in the transcriptome analysis. During the biosynthesis of unsaturated fatty acids, erucic acid was upregulated in IW. It is a monounsaturated fatty acid that belongs to the omega-9 fatty acid family. Erucic acid has been known to produce significant behavioural improvement in the Y-maze test of rats; improve proinflammatory markers, oxidative stress, and memory function; and protect against neuronal damage (Sayyed et al., 2022). In the arginine and proline metabolism pathways, N-acetylputrescine was upregulated and L-ornithine and spermidine were downregulated. N-formylkynurenine and xanthurenic acid in the tryptophan metabolism pathway were downregulated. In glutathione metabolism, cadaverine was upregulated and spermidine and L-ornithine were downregulated. Thus, the metabolic levels of W and IW are associated with changes in stress responses caused by female reproduction and the nest environment. The combined metabolome and transcriptome analyses showed that fatty acid and amino acid metabolism were the most prominent.

Role of transposable element derived genes in the transformation of workers into reproductives

We detected dynamic changes in the expression levels of 56 TE-related genes at the four stages of *R. labralis* and *R. chinensis* using qPCR (Supplementary Figure S3, S4). Based on transcriptome analysis and qPCR results, 44 TE-related genes were significantly upregulated in W-IW, and 18 of the 44 genes were significantly downregulated in IW-NRW (Figure 4a). The expression profile of these genes showed a trend of upregulation in IW and then downregulation in NRW, indicating that they were related to an “inhibition-activation-reinhibition” process during the transition of workers to reproductives. We identified four TE-derived gene sequences: *tigger transposable element-derived 1* (*TIGD1*), *TIGD6*, *piggyBac transposable element-derived 4* (*PGBD4*) and *PGBD3*. Among them, *TIGD1* and *PGBD4* were present in

both *R. labralis* and *R. chinensis*, qPCR results showed that the expression levels of *TIGD1* and *PGBD4* were significantly higher in IW compared to W, NRW, and NR. *Tigger* and *piggyBac*, as DNA transposons, are more common in organisms ranging from fungi to humans. TE-derived genes represent instances of genes derived from TE sequences (Kolacsek et al., 2022; Raskó et al., 2022). Moreover, *Tc3 transposase* was identified. Forced expression of the *Tc3 transposase* led to excision and transposition of *Tc* elements (Colloms et al., 1994). Thus, we consider that these TE-derived genes, representing both new additions and novel functions within the gene repertoire, can give rise to reproductive plasticity in workers.

We were interested in how *PGBD*, *TIGD* and *Tc3a* affect the reproductive plasticity of female workers. Here, we used RNAi to knock down *TIGD1*, *PGBD4* and *Tc3a* in isolated female workers of *R. labralis* for 19 days. Double-stranded RNAs (ds*TIGD1*, ds*PGBD4* and ds*Tc3a*) were used for quantitative feeding of the isolated workers; ds*GFP* was used as a control. The expression levels of *TIGD1*, *PGBD4*, and *Tc3a* decreased significantly after 48 h ($P<0.01$) (Figure 4b), resulting in a delay in the transformation time from an average of 24 days (control group) to 39 days (ds*TIGD1* group), 42 days (ds*PGBD4* group), and 33 days (ds*Tc3a* group) (Figure 4c). Furthermore, after interfering with isolated female workers using dsCocktail, a mixture of ds*TIGD1*, ds*PGBD4* and ds*Tc3a,* the workers took an average of 55 days to transform into reproductives (Figure 4d).

The termite brain consists of three zones: procerebrum, deutocerebrum and tritocerebrum (Figure 4d). To determine the localization of PGBD and TIGD proteins in the brain, we performed immunohistochemical localization of PGBD4 and TIGD1 in the brain sections of isolated female workers of *R. labralis*. Strong immunoreactivity was detected in nerve cell nuclei of the forebrain (Figure 4e, f), indicating that TEs in the brains of workers are associated with the transformation of workers into reproductive, and the forebrains of workers may be the regulatory centre of reproductive plasticity.

Constructing datasets based on three *Reticulitermes* species to reflect the key GO functions associated with reproductive transformation of workers

In our approach to predict the presence of reproductives in a colony based on the key molecular functions of female workers, the binary classification is applied. W or NRW indicate that the colony contains female reproductives (fertility) and is assigned to class 0 as its label. In contrast, IW denotes colonies that lack female reproductives (sterility) and is assigned to class 1 as its label. We utilised raw transcriptome data from *R. labralis*, *R. chinensis*, and *R. aculabialis* to construct a dataset. To ensure the reliability, consistency, and generalisability of the dataset, we rigorously filtered and reassembled the *R. labralis* transcriptome raw data and obtained annotations, raw counts, reads per kb per million reads (RPKM), and GSEA results. The raw counts and RPKM values of *R. chinensis* and *R. aculabialis* were based on the assembly results for *R. labralis*, ensuring unigene consistency across the three datasets (Figure 5a, b). The GO enrichment results obtained from GSEA were used as input data features. This method has three advantages: 1. As each GO function corresponds to multiple unigenes, each input can have multiple selections, allowing for a large dataset. 2. We have more flexible options when using the model for prediction, avoiding the inability to use models because some genes are difficult to express in certain species. 3. We can choose multiple unigenes for certain features (GO functions) to obtain comprehensive results and improved prediction accuracy by adopting the mode of the predicted results.

In the dataset generation part, based on the GO enrichment results from GSEA, we selected 160 GO functions containing unigenes≥20 and obtained the RPKM values of the corresponding unigenes in *R. labralis*, *R. chinensis*, and *R. aculabialis* (Figure 5c). We generated 2×10^5 samples for each species through random combinations and mixed the three datasets to obtain a total of 6×10^5 samples. For the 160 GO

functions, with each function having unigenes≥20, there are theoretically at least 20^{160} combinations, which is a vast number compared to the total dataset of 6×10^5. To simulate this, the sample sizes for the training, validation, and test datasets were divided into 5×10^4, 5×10^4, and 5×10^5. Additionally, to ensure consistent performance across different data sources and scales (such as qPCR and RNA-seq results) and thus ensure high generalisability, we scaled the test dataset.

Random forest, SVM, KNN, LDA, QDA, and GBM models lack generalization

Initially, we employed a variety of machine learning models, including random forest, SVM, KNN, LDA, QDA, and GBM. After training, these models achieved a prediction accuracy of over 70% on the validation dataset, with the random forest and GBM models reaching 92.728% and 99.796% accuracy, respectively (Figure 6a). However, when predictions were made on the scaled test dataset, all classic machine learning models performed poorly, with an accuracy of approximately 50%. Therefore, these models cannot maintain consistent performance across data at various scales, resulting in poor generalisation.

Deep neural network models exhibit strong generalization and robustness

Based on the aforementioned results and the fact that a tensor with a size of 160 can be reshaped into various shapes, we focused on the performance of deep learning to address this issue. We initially constructed six different deep neural network models, each focusing on a specific type of layer: a fully connected neural network (FCNN), one-dimensional convolutional neural network (1D-CNN), two-dimensional convolutional neural network (2D-CNN), depthwise separable convolutional neural network (DS-CNN), gated recurrent unit (GRU), and long short-term memory (LSTM). This approach allowed us to explore the type of layer that was most suitable for extracting information from the expression levels of multiple GO functions. After training, all six deep neural network models achieved over 80% accuracy for both the validation and scaled test datasets, indicating their ability to maintain a stable performance across data of different scales (Figure 6a; Supplementary Figure S5). Surprisingly, on the test dataset, the FCNN achieved the lowest prediction accuracy among the six deep neural network models (84.691%), while the 2D-CNN, typically used for processing images, achieved a high accuracy of 90.176%. The 1D-CNN, LSTM, and GRU, which are adept at handling text and sequences, also reached prediction accuracies of 91.029%, 90.694%, and 90.023%, respectively.

Subsequently, we selected the best-performing 2D-CNN, 1D-CNN, and LSTM to construct a multi-branch neural network (MBNN). We anticipated that a combination of these three types of layers would leverage their respective strengths and mitigate their weaknesses, thereby enhancing their predictive capabilities. After training, MBNN achieved the highest prediction accuracy on the test dataset, reaching 91.037% (Figure 6a, e). Thus, we identified an optimal neural network model using 160 GOs terms as input features.

Light neural network models with 32 GO functions as input features based on SHapley Additive exPlanations (SHAP) values

Although MBNN models have high predictive accuracy, considering the practicality of the model, using 160 GO functions as inputs seemed cumbersome and costly. After confirming the feasibility of the neural network models, we reduced the input size and reconstructed a lightweight model. In terms of inputs, we reduced the feature size from 160 to 32, and in terms of model construction, we employed a light multi-branch neural network (LMBNN). With a significant reduction in input size, we discarded the LSTM and GRU, and each branch used FCNN, 2D-CNN, and 1D-CNN (Figure 5c). We employed a game-theoretic approach to calculate the contribution of each input feature to the MBNN model (Lundberg

& Lee, 2017), ranked them based on their SHAP values, and selected the top 32 GO functions as input features (Figure 6b, c). The LMBNN-1 model achieved an accuracy of 79.561% for the test dataset after training (Figure 6a, e). By randomly selecting two unigenes from the GO terms with the highest SHAP values for the combination and using the mode of all results as the final outcome, accuracy was improved. We conducted experiments with two unigenes selected from one to seven GO terms and obtained the highest accuracy of 80.3% using five GO terms (Figure 6d).

The LMBNN architecture was employed to test its generalisability across different species. We continued to use the aforementioned 32 GO terms as input but utilised the species that workers could transform under experimental conditions, designated as *R. labralis* and *R. chinensis*, for the training and validation datasets, and the species that workers could not transform, designated as *R. aculabialis*, for the test dataset. The LMBNN-2 model achieved accuracies of 77.263% and 76.594% for the validation and test datasets, respectively (Figure 6a, e). When the two random unigenes corresponding to the GOs with the highest SHAP values were combined and the mode of all outputs was taken as the final result, the accuracy was improved. The highest accuracy was 78.45% when three GO functions were selected (Figure 6d). These results indicated that the structure of the model performed consistently when dealing with different species.

DISCUSSION

We found that the female workers of three species exhibited significantly different responses to environmental stress. This finding has great significance for understanding the mechanism of termite reproductive plasticity as well as the colony survival strategies across species. Significant reproductive differences exist among these species; for example, *R. aculabialis* can undergo parthenogenesis, whereas *R. labralis* and *R. chinensis* cannot. In *R. labralis*, adultoid reproductives with floppy wings develop from nymphs and serve as secondary reproductives. However, adultoid reproductives are absent in *R. aculabialis* (Su et al. 2015; Tan et al. 2016). Transcriptomic signatures associated with reproductive plasticity and social behaviour may also be influenced by the level of social complexity (Sun et al., 2019; Gospocic et al., 2021; Oguchi et al., 2022; Wyatt et al., 2023).

In this study, we conducted a detailed analysis of the relationship between gene expression and reproductive plasticity, revealing broad-scale shifts in caste-associated transcriptomic profiles. The "inhibition- activation-reinhibition" cycle is regulated not by a small number of genes, but by numerous genes involved in multiple functional pathways. Notably, the molecular functions and pathways underlying this process were highly consistent between *R. labralis* and *R. chinensis*. Metabolites serve as intermediates in biochemical pathways and regulate both enzymatic and non-enzymatic proteins. Alterations in the levels of specific metabolites have been linked to phenotypic plasticity (Liska et al., 2023; Ryabova et al., 2020). In this study, four metabolome-enriched pathways were also found to be enriched in the transcriptome (biosynthesis of unsaturated fatty acids, arginine and proline metabolism, tryptophan metabolism, and glutathione metabolism). A recent study revealed sex-specificity in the *Caenorhabditis elegans* metabolome, demonstrating the incorporation of building blocks from nucleoside, carbohydrate, lipid, and amino acid metabolism. It also identified a male germline-dependent metabolite that increases food consumption, reduced lifespan, and accelerates larval development in hermaphrodites (Burkhardt et al., 2023). Elsner et al. (2021) found that storage proteins underlie task division among similarly aged termite workers. Therefore, we propose that fatty acid and amino acid metabolism play important regulatory roles

in the transformation of workers into reproductives. Moreover, previous studies indicate that reduced JH synthesis induces the differentiation of isolated workers into neotenic reproductives (Elliott & Stay, 2008). The blood-brain barrier regulates the entry of JH_3 into the brain, where it reprograms the brain transcriptome (Gospocic et al., 2017 and 2021; Ju et al., 2023). JH is transported by carrier proteins (JHBPs) to target organs (Liu et al., 2024). In this study, we found that *JHBP* expression was significantly downregulated in isolated workers of *R. labralis* and *R. chinensis*, but significantly upregulated in *R. aculabialis*. These results are consistent with previous reports and also confirm that the isolated workers of *R. aculabialis* did not initiate reproductive transformation.

We confirm that the absence of female reproductives (queens) is a necessary but certainly not sufficient condition for female workers to develop into reproductives. When female workers are subjected to severe environmental stress that threatens their survival, they delay or do not initiate reproductive transformation, even in the absence of queens in the nests. We were interested in whether the isolated workers of *R. aculabialis* might exhibit delayed transformation. However, our experiments showed that the isolated workers of *R. aculabialis* did not transform into reproductives even after 16 weeks, and all isolated groups eventually died. Previous studies reported that in groups of 24 and 12 isolated workers of *R. flavipes*, the first reproductives appeared after 10 and 13 weeks, respectively (Elliott & Stay, 2008), suggesting that larger group sizes facilitate the transformation of workers into reproductives. Interestingly, although female reproductives derived from workers are commonly observed in natural colonies of *R. aculabialis*, it appears difficult to activate the molecular regulatory mechanisms associated with reproductive development in isolated workers under experimental conditions. We suggest that the workers' limited ability to adapt to environmental stresses may prevent them from triggering the molecular mechanisms associated with transformation into reproductives, even in the absence of queens.

Reproductive plasticity represents a key stress response of workers to changes in colony fecundity. This study identifies the involvement of TE-derived genes in regulating this reproductive plasticity in female workers. Previous research has shown that various stressors, such as cold and heat stress, ultraviolet irradiation, salinity stress, and pollution, can alter TE activity, and that such activation is often adaptive under challenging environmental conditions (Payer & Burns, 2019; Mombach et al., 2022). Post et al. (2023) observed substantially higher TE activity in termite workers compared to reproductives. Here, we confirmed that, in *R. labralis* and *R. chinensis*, 18 genes were related to an "inhibition-activation-reinhibition" process during the transition of workers to reproductives, and confirmed that *piggyBac* and *tigger* transposable element-derived genes in the worker brains promoted the reproductive transformation. *TIGDs* and *PGBDs* can promote the proliferation, migration, and invasion of tumor cells in humans (Henssen et al., 2017; Yin et al., 2019). When cells were exposed to various stress conditions, including hypoxia, oxidative or UV stress, the expression profiles of *PGBDs* showed distinct changes (Kolacsek et al., 2022). However, their exact molecular mechanisms and biological roles remain to be elucidated. Our findings demonstrate that TE activity is associated with the reproductive plasticity of female workers. The differentiation of female workers into reproductives can be triggered to prevent colony extinction and facilitate rapid adaptation to environmental stress.

Our approach utilizes the findings of the reproductive plasticity regulation mechanism of female workers to construct random forest, SVM, KNN, LDA, QDA, GBM and deep neural network models. These models employ key Gene Ontology (GO) term expression levels during reproductive transformation of workers to predict caste and colony fecundity. The transition from workers to reproductives is not driven by individual molecular functions acting in isolation, but rather results from complex interactions among multiple molecular functions. We leveraged both recurrent neural networks and convolutional neural

networks in our approach. According to Frege's principle, a complex expression is determined by its constituent expressions and the rules used to combine them (Hintikk, 1984). In the context of reproductive transformation in workers, these key GOs terms function act more like pixels in an image or words in a sentence, conveying information about the presence or absence of reproductives within a colony. These findings indicate that colony fecundity can be accurately predicted using molecular data from workers. Furthermore, the findings also demonstrate that workers, the only caste within the colony exhibiting reproductive plasticity, carry substantial molecular information about the colony's reproductive status. This offers a novel perspective and methodology for future research on the survival strategies of termite colonies and other eusocial insects.

The molecular functions associated with the reproductive plasticity of workers not only regulate their transformation into NRs, but also reflect the absence of reproductives within the colony, that is, whether the colony has reproductive capacity. The accuracy of an MBNN model using 160 GO functions as inputs was greater than that of an LMBNN using 32 GOs, suggesting that the reflection of colony reproduction is the result of multiple interacting molecular functions. When 32 GO functions are used as inputs, selecting multiple corresponding unigenes from the GOs for combination and using the mode of all outputs as the final result can improve accuracy and allow for more flexible use of the model. It implies that the absence of reproductives in the colony may not only be reflected in one combination of various molecular function expression levels but also would correspond to multiple combinations. These features could represent the robustness of the mapping relationship between the molecular level of social insects and the colony state.

This study brings a new perspective to the analysis of termite gene functions; instead of treating multiple genes as multiple independent vectors, they are reshaped into tensors of different dimensions, presented in the form of images or sentences. In this analogy, each function serves as a pixel in an image or character in a sentence. This method allowed us to understand the differences at the molecular level between individual termites in various states within a colony in a more abstract but effective manner. After reshaping, the inputs yield a higher accuracy in predictions made using convolutional neural networks and recurrent neural networks compared with those made using fully connected neural networks using inputs without reshaping. This finding suggests that when the genes regulating the reproductive plasticity of female workers are reshaped into image-like or sequence-like formats, the model can extract more information about colony fecundity (the presence or absence of queens) from the expression levels of these functional genes. Thus, the molecular functions that regulate the transformation of female workers interact with each other, and these functions, along with their "contextual relationships", jointly regulate the reproductive plasticity of the female workers and reflect the absence of reproductives in the colony.

DATA AVAILABILITY

RNA sequence data were deposited in NCBI Database SRA (accession numbers SRP559985, SRP559966, SRP559824), GSA database (accession numbers PRJCA061431), and Science Date Bank database (https://doi.org/10.57760/sciencedb.j00139.00275). All raw LC-MS/MS data have been deposited to the iProX system (http://www.iprox.org/index) with the accession number: IPX0010997000.

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.Y., Conceptualization, Data curation, Formal analysis, Methodology, Investigation, Project administration, Writing–original draft. W.X.Z., X.Y.L., T.G.Z., Y.Y.T., Y.L.Y., and N.U.S., Investigation, Formal analysis. L.X.X., and S.L., Methodology, Resources, Supervision, Funding acquisition. X.H.S., Resources, Supervision, Funding acquisition, Writing – review and editing. All authors read and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank Dr. Hui-Ping Wang and Ling-Fang Yin for the culture of termites. This study was supported by the National Natural Science Foundation of China (31870389, 32471563), Natural Science Basic Research Program of Shaanxi (2019JM356).

REFERENCES

Allaire JJ, Chollet F, Tang Y. 2022. keras: R Interface to 'Keras'. R package version 2.9.0.

Ashburner M, Ball CA, Blake JA, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*, **25**: 25–29.

Ashton LA, Griffiths HM, Parr CL, et al. 2019. Termites mitigate the effects of drought in tropical rainforest. *Science*, **363**: 174–177.

Baduel P, Quadrana L. 2021. Jumpstarting evolution: how transposition can facilitate adaptation to rapid environmental changes. *Current Opinion in Plant Biology*, **61**: 102043.

Berger J, Legendre F, Zelosko KM, et al. 2022. Eusocial transition in Blattodea: a transposable elements and shifts of gene expression. *Genes*, **13**: 1948.

Breiman L. 2001. Random Forests. *Machine Learning*, **45**: 5–32.

Burkhardt RN, Artyukhin AB, Aprison EZ, et al. 2023. Sex-specificity of the *C. elegans* metabolome. *Nature Communications*, **14**:320.

Chen Q, Luo W, Veach RA, et al. 2020. Structural basis of seamless excision and specific targeting by piggyBac transposase. *Nature Communications*, **11**: 3446.

Chen S, Zhou Y, Chen Y, et al. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**: 884–890.

Colloms SD, Luenen HGAM, et al. 1994. DNA binding activities of the *Caenorhabditis elegans* Tc3 transposase. *Nucleice Acids Resesrch*, **22**: 5548–5554.

Conesa A, Götz S, García-Gómez JM, et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, **21**: 3674–3676.

Cortes C, Vapnik V. 1995. Support-Vector Networks. *Machine Learning*, **20**: 273–297.

Cover T, Hart P. 1967. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, **13**: 21–27

Cowley BR, Calhoun AJ, Rangarajan N, Ireland E, Turner MH, Pillow JW, Murthy M. 2024. Mapping model units to visual neurons reveals population code for social behaviour. *Nature*, **629**: 1100–1108.

Elliott KL, Stay B. 2008. Changes in juvenile hormone synthesis in the termite *Reticulitermes flavipes* during development of soldiers and neotenic reproductives from groups of isolated workers. *Journal of Insect Physiology*, **54**: 492–500.

Elsner D, Meusemann K, Korb J. 2018. Longevity and transposon defense, the case of termite reproductives. *Proc Natl Acad Sci USA*, **115**: 5504–5509.

Elsner D, Hartfelder K, Korb J. 2021. Molecular underpinnings of division of labour among workers in a socially complex termite. *Scientific Reports*, **11**:18269.

Etchegaray E, Naville M, Volff JN, et al. 2021. Transposable element-derived sequences in vertebrate development. *Mobile DNA*, **12**:1.

Fisher RA. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**: 179–188.

Friedman JH. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, **29**: 1189–1232.

Fujita A, Watanabe H. 2010. Inconspicuous matured males of worker form are produced in orphaned colonies of *Reticulitermes speratus* (Isoptera: rhinotermitidae) and participate in Reproduction. *Journal of Insect Physiology*, **56**:1510–1515.

Gordon DM, Goodwin BC, Trainor LEH. 1992. A parallel distributed model of the behaviour of ant colonies. *Journal of Theoretical Biology*, **156**: 293–307.

Gordon DM. 2019. The Ecology of Collective Behavior in Ants. *Annual Review of Entomology*, **64**: 35–50.

Gospocic J, Glastad KM, Sheng L, et al. 2021. Kr-h1 maintains distinct caste-specific neurotranscriptomes in response to socially regulated hormones. *Cell*, **184**: 5807–5823.

Gospocic J, Shields EJ, Glastad KM, et al. 2017. The neuropeptide corazonin controls social behavior and caste identity in ants. *Cell*, **170**: 748–759.

Grabherr MG, Haas BJ, Yassour M, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, **29**: 644–652.

Harrison MC, Jongepier E, Robertson HM, et al. 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution*, **2**: 557–566.

Hastie T, Tibshirani R, Friedman J. 2009. The elements of statistical learning (2nd ed., Sec. 4.3). Springer.

Henssen AG, Koche R, Zhuang J, et al. 2017. PGBD5 promotes site-specific oncogenic mutations in human tumors. *Nature Genetics*, **49**: 1005–1014.

Hintikk JA. 1984. A hundred years later: The rise and fall of Frege's influence in language theory. *Synthese*, **59**: 27–49.

Jönsson M, Sigrist R, Gren T, et al. 2025. Machine learning uncovers the transcriptional regulatory network for the production host *Streptomyces albidoflavus*. *Cell Reports*, **44**: 115392.

Ju L, Glastad KM, Sheng L, et al. 2023. Hormonal gatekeeping via the blood-brain barrier governs caste-specific behavior in ants. *Cell*, **186**: 4289–4309.

Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, **28**: 27–30.

Kolacsek O, Wachtl G, Fothi A, et al. 2022. Functional indications for transposase domestications – Characterization of the human piggyBac transposase derived (PGBD) activities. *Gene*, **834**: 146609.

Korb J, Hartfelder K (2008). Life history and development-a framework for understanding developmental plasticity in lower termites. *Biol. Rev*, 83: 295–313.

Korb J. 2015. Juvenile Hormone: A Central Regulator of Termite Caste. Polyphenism. *Genomics, Physiology and Behaviour of Social Insects*, **48**: 131–161.

Krajewski A. 2023. The importance of the minimal number of workers required in the dispersal of *Reticulitermes santonensis* Feytaud, 1924 by budding. *Sylwan*, **167**:101–109.

Kuhn M. 2008. Building predictive models in R using the caret package. *Journal of Statistical Software*, **28**: 1-26.

LeDell E, Parmar V, Candel A, et al. 2021. h2o: R Interface for the H2O Scalable Machine Learning Platform (Version 3.32.1.3) [Computer software]. CRAN. https://CRAN.R-project.org/package=h2o.

Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, **12**: 323.

Liaw A, Wiener M. 2022. Classification and regression by random forest. *R News*, **2**: 18–22.

Lin S, Elsner D, Ams L, et al. 2024. A genetic toolkit underlying the queen phenotype in termites with totipotent workers. *Scientific Reports*, **14**: 2214.

Lin S, Werle J, Korb J. 2021. Transcriptomic analyses of the termite, *Cryptotermes secundus*, reveal a gene network underlying a long lifespan and high fecundity. *Commun. Biol.*, **4**: 384.

Liu Y, Zou KL, Wang TH, et al. 2024. Genome-wide identification and analysis of family members with juvenile hormone

binding protein domains in *Spodoptera frugiperda*. *Insects*, **15**: 573.

Liska O, Borossb G, Rocabert C, et al. 2023. Principles of metabolome conservation in animals. *Proc Natl Acad Sci USA*, **120**: e2302147120.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*., **15**: 550.

Lundberg SM, Lee S. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, **30**: 4768–4777.

Matsuura K, Mizumoto N, Kobayashi K, et al. 2018. A genomic imprinting model of termite caste determination: not genetic but epigenetic inheritance influences offspring caste fate. *The American Naturalist*, **191**: 677–690.

Meyer D, Dimitriadou E, Hornik K, et al. 2021. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071) R Package Version, 1:7–8.

Mombach DM, Gomes TMFF, Loreto ELS. 2022. Stress does not induce a general transcription of transposable elements in *Drosophila*. *Molecular Biology Reports*, 9033–9040.

Nanda MA, Seminar KB, Nandika D, et al. 2019. Development of termite detection system based on acoustic and temperature signals. *Measurement*, **147**: 106902.

Oguchi K, Koshikawa S, Miura T. 2022. Hormone-related genes heterochronically and modularly regulate neotenic differentiation in termites. *Developmental Biology*, **485**: 70–79.

Oguchi K, Miura T. 2024. Body part-specific development in termite caste differentiation: crosstalk between hormonal actions and developmental toolkit genes. *Current Opinion in Insect Science*, **63**: 101183.

Patel JS, Lee SB, Chouvenc T, et al. 2020. Inferring termite colony size using wood consumption in subterranean termites (Blattodea: Rhinotermitidae) in laboratory-rearing conditions. *Journal of Economic Entomology*, **113**: 905–910.

Payer LM, Burns KH. 2019. Transposable elements in human genetic disease. *Nature Reviews Genetics*, **20**: 760–772.

Pinter-Wollman N, Hubler J, Holley JA, et al. 2012. How is activity distributed among and within tasks in Temnothorax ants? *Behav. Ecol. Sociobiol*, **66**: 1407–1420.

Post F, E Bornberg-Bauer, Vasseur-Cognet M, et al. 2023. More effective transposon regulation in fertile, long-lived termite queens than in sterile workers. Molecular Ecology 32: 369–380.

Rappa H, Nawrot MP. 2020. A spiking neural program for sensorimotor control during foraging in flying insects. *Proc. Natl Acad. Sci. USA*, **117**: 28412–28421.

Rasheed H, Ye C, Meng Y, et al. 2019. Comparative transcriptomic analysis and endocuticular protein gene expression of alate adults, workers and soldiers of the termite *Reticulitermes aculabialis*. *BMC Genomics*, **20**: 742.

Raskó T, Pande A, Radscheit K, et al. 2022. A novel gene controls a new structure: PiggyBac transposable element-derived 1, unique to mammals, controls mammal-specific neuronal paraspeckles. *Mol. Biol. Evol*., **39**: msac175.

Rau V, Flatt T, Korb J. 2023. The remoulding of dietary effects on the fecundity / longevity trade-off in a social insect. *BMC Genomics*, **24**: 244.

Razin N. 2013. Desert ants achieve reliable recruitment across noisy interactions. *Journal of the Royal Society Interface*, **10**: 20130079.

Ryabova A, Cornette R, Cherkasov A, et al. 2020. Combined metabolome and transcriptome analysis reveals key components of complete desiccation tolerance in an anhydrobiotic insect. *Proc Natl Acad Sci USA*, **117**:19209-19220.

Saiki R, Hayashi Y, Toga1 K, et al. 2022. Comparison of gene expression profiles among caste differentiations in the termite *Reticulitermes speratus*. *Scientific Reports*, **12**, 11947.

Sayyed N, Hafeez A, Al-Abbasi FA, et al. 2022. Erucic acid ameliorates the lipopolysaccharide-induced memory deficit in rats through inhibited inflammation cytokines expression/caspase 3/NF-κB pathways. *European Journal of Lipid Science and Technology*, **125**: 202200205.

Scharf ME, Scharf DW, Zhou X, et al. 2005. Gene expression profiles among immature and adult reproductive castes of the termite *R. flavipes*. *Insect Mol. Biol.*, **14**: 31–44.

Shigenobu S, Hayashi Y, Watanabe D, et al. 2022. Genomic and transcriptomic analyses of the subterranean termite *Reticulitermes speratus*: Gene duplication facilitates social evolution. *Proc. Natl Acad. Sci. USA*, **119**: 3 e2110361119.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**: 3210–3212.

Su X, Yang X, Li J, et al. 2017. The transition path from female workers to neotenic reproductives in the termite *Reticulitermes labralis*. *Evolution & Development*, **19**: 218–226.

Su XH, Xue W, Liu H, et al. 2015. The Development of adultoid reproductives and brachypterous neotenic reproductives from the last instar nymphs in *Reticulitermes labralis* (Isoptera: Rhinotermitidae): A comparative study. *Journal of Insect Science*, **15**: 147.

Sun P, Li G, Jian J, et al. 2019. Transcriptomic and functional analyses of phenotypic plasticity in a higher termite, *Macrotermes barneyi* Light. *Frontiers Genetics* **10**: 964.

Takata M, Takahashi M, Ishibashi T, et al. 2025. Transgenerational epigenetic effect of kings' aging on offspring's caste fate mediated by sperm DNA methylation in termites. *Proc. Natl Acad. Sci. USA*, **122**: e2509506122.

Takata M, Yabe K, Noro T, et al. 2023. A method for estimating colony size using queen fecundity in termites under feld conditions. *Science of Nature*, **110**: 35.

Tan YL, Dang YL, Zhang HG, et al. 2016. A comparative study of embryonic development between gamogenesis and parthenogenesis in two *Reticulitermes* termites (Isoptera:Rhinotermitidae). *Acta Entomologica Sinica*, **59**: 438–445.

Taylor BA, Cini A, Wyatt CDR, et al. 2021. The molecular basis of socially mediated phenotypic plasticity in a eusocial paper wasp. *Nature Communications*, **12**:775.

Thorne BL, Russek-Cohen E, Forschler BT, et al. 1996. Evaluation of mark-release-recapture methods for estimating forager population size of subterranean termite (Isoptera: Rhinotermitidae) colonies. *Environmental Entomology*, **25**: 938–951.

Veenstra JA. 2023 a. Differential expression of some termite neuropeptides and insulin/IGF-related hormones and their plausible functions in growth, reproduction and caste determination. *PEER J*, **11**: e15259.

Veenstra J. A.2023 b. Different neuroendocrine cell types in the pars intercerebralis of *Periplaneta americana* produce their own specific IGF-related peptides. *General and Comparative Endocrinology*, **335**: 114233.

Venables WN, Ripley BD. 2022. Modern Applied Statistics with S (4th ed.). Springer.

Want EJ, Masson P, Michopoulos F, et al. 2013. Global metabolic profiling of animal and humantissues via UPLC-MS. *Nature Protocols*, **8**:17–32.

Waters JS, Fewell JH. 2012. Information processing in social insect networks. *PLoS One*, **7**: e40337.

Wen B, Mei Z, Zeng C, et al. 2017. metaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinformatics*, **18**: 183.

Wu J, Su X, Kong X, et al. 2013. Multiple male and female reproductive strategies and the presence of a polyandric mating system in the termite *Reticulitermes labralis* (Isoptera: Rhinotermitidae). *Sociobiology* **60**: 459–465.

Wyatt CDR, Bentley MA, Taylor D, et al. 2023. Social complexity, life-history and lineage influence the molecular basis of castes in vespid wasps. *Nature Communications*, **14**:1046.

Zhou X, Oi FM, Scharf ME. 2006. Social exploitation of hexamerin: RNAi reveals a major caste-regulatory factor in termites. *Proc. Natl. Acad. Sci. USA*, **103**, 4499–4504.

Ye C, Rasheed H, Ran Y, et al. 2019. Transcriptome changes reveal the genetic mechanisms of the reproductive plasticity of workers in lower termites. *BMC Genomics*, **20**:702.

Ye CX, Song ZZ, Wu TY, et al. 2021. Endocuticle is involved in caste differentiation of the lower termite. *Current Zoology*, **67**: 489–499.

Yin L, Yan J, Wang Y, et al. 2019. TIGD1, a gene of unknown function, involves cell-cycle progression and correlates with poor prognosis in human cancer. *Journal of Cellular Biochemistry*, **120**: 9758–9767.

Zhu SM, Liu FF, Zeng HC, et al. 2020. Insulin/IGF signaling and TORC1 promote vitellogenesis via inducing juvenile hormone biosynthesis in the American cockroach. *Development*, **147**: dev188805.

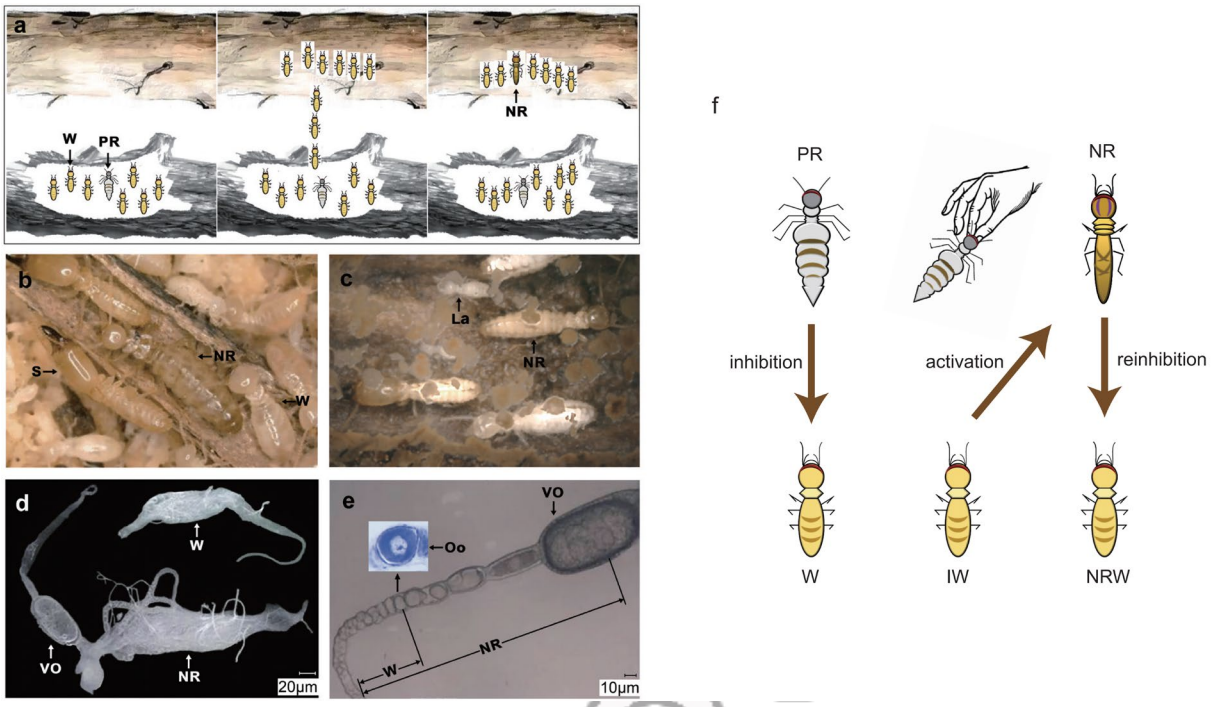

Figure 1 The workers of *Reticulitermes* created new nests

a: In the field, the workers of *Reticulitermes* exploited new food resources outside the natal nest and moved to a new nest site. The differentiation of female workers into neotenic reproductives was triggered in the new nest site. b: Female neotenic reproductive of *R. aculabialis* were transformed from workers in wild colonies. c: In the laboratory, a group of isolated workers of *R. labralis* created a new colony. The differentiation of female neotenic reproductive was induced in isolated workers, and then the neotenic reproductive produced offspring. d: The ovaries of workers and neotenic reproductives. e: The process of oogenesis in an ovariole of neotenic reproductives. The oocyte development in workers was inhibited and stopped at oocyte growth stage, vitellogenesis occurred in the ovarioles of the neotenic reproductives. f: The transition of female workers into reproductives is influenced by the presence or absence of the reproductives (queens), resulting in a dynamic cycle pattern of "inhibition-activation-reinhibition". W, workers; PR, female primary reproductives developed from alate adults; NR, female neotenic reproductives transformed from workers; NRW, isolated female workers after female reproductives emerged in isolated groups; S, soldiers; La, larvae; Oo, oocytes at growth stage; VO, vitellogenic oocytes at late vitellogenesis stage.

Science Press
ZR Zoological Research

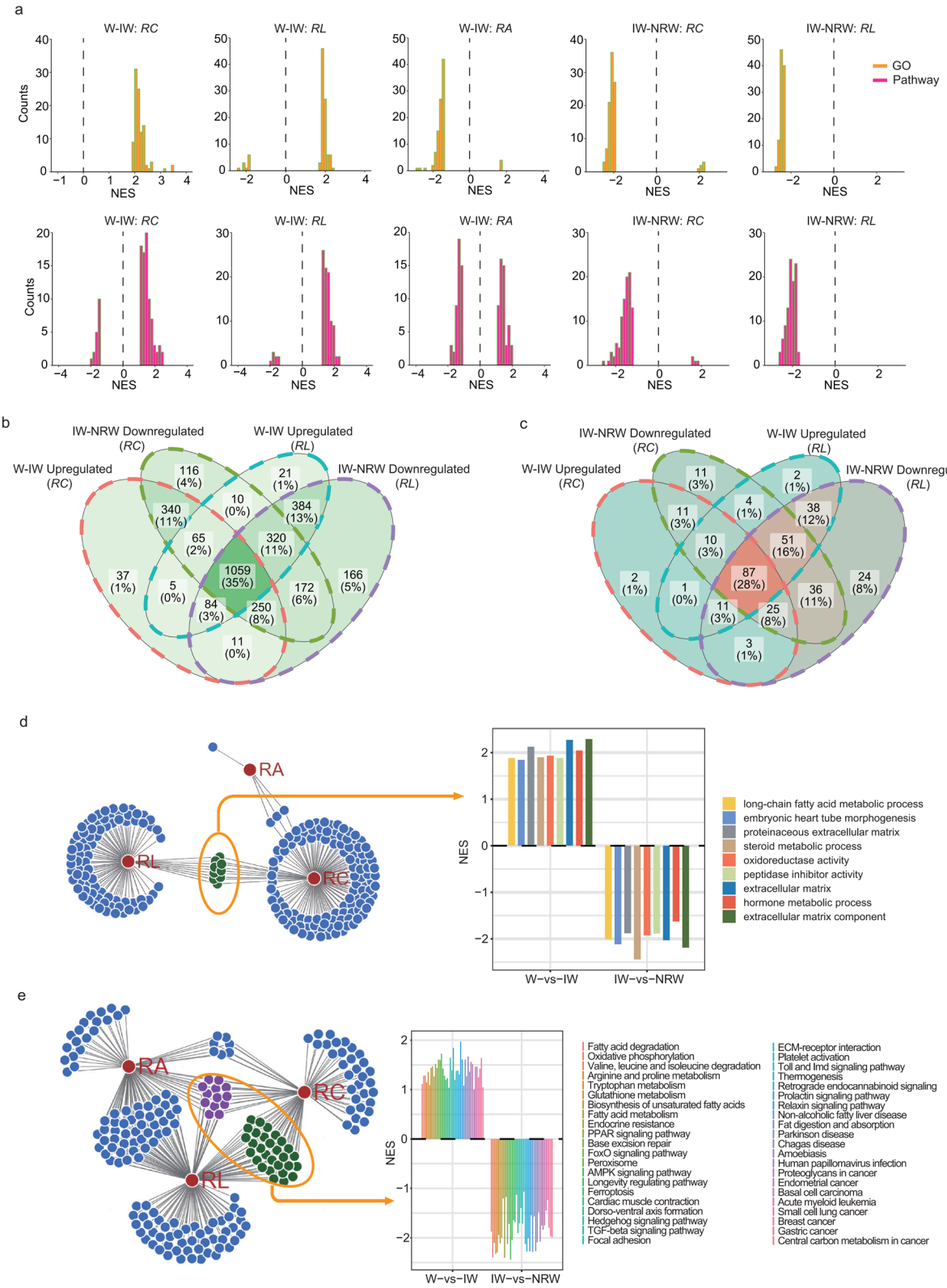

Figure 2 Transcriptomic pattern for "inhibition-activation-reinhibition" in the transformation of workers to reproductives in "*R. labralis* and *R. chinensis*" vs. *R. aculabialis*

a: NES of the top 100 significant GO functions and KEGG pathways during W-IW and IW-NRW from GSEA results in three *Reticulitermes* species, respectively. b & c: Venn diagrams of the intersection GO functions (b) and KEGG pathways (c) of

Science Press
ZR Zoological Research

RC-up, RC-down, RL-up and RL-down datasets. d: The network of the "inhibition-activation-reinhibition" GO functions (*NES*>0 in W-IW; *NES*<0 in IW-NRW; *FDR*<0.05) in RL and RC, respectively, and the GO functions significantly regulated during W-IW in RA (*FDR*<0.05). The NES of the intersected 9 GO functions (green and purple dots) was shown using the GSEA result of RL. e: The network of the "inhibition-activation-reinhibition" KEGG pathways *(NES*>1 in W-IW; *NES*<-1 in IW-NRW) in RL and RC, respectively, and the KEGG pathways upregulated during W-IW in RA (*NES*>1). The *NES* of the intersected 42 KEGG pathways (green dots) was shown using the GSEA result of RL. RL, *R. labralis*; RC, *R. chinensis*; RA, *R. aculabialis*; RL-up, The functions upregulated during W-IW in *R. labralis*; RL-down, The functions downregulated during IW-NRW in *R. labralis*; RC-up, The functions upregulated during W-IW in *R. chinensis*; RC-down, The functions downregulated during IW-NRW in *R. chinensis*.

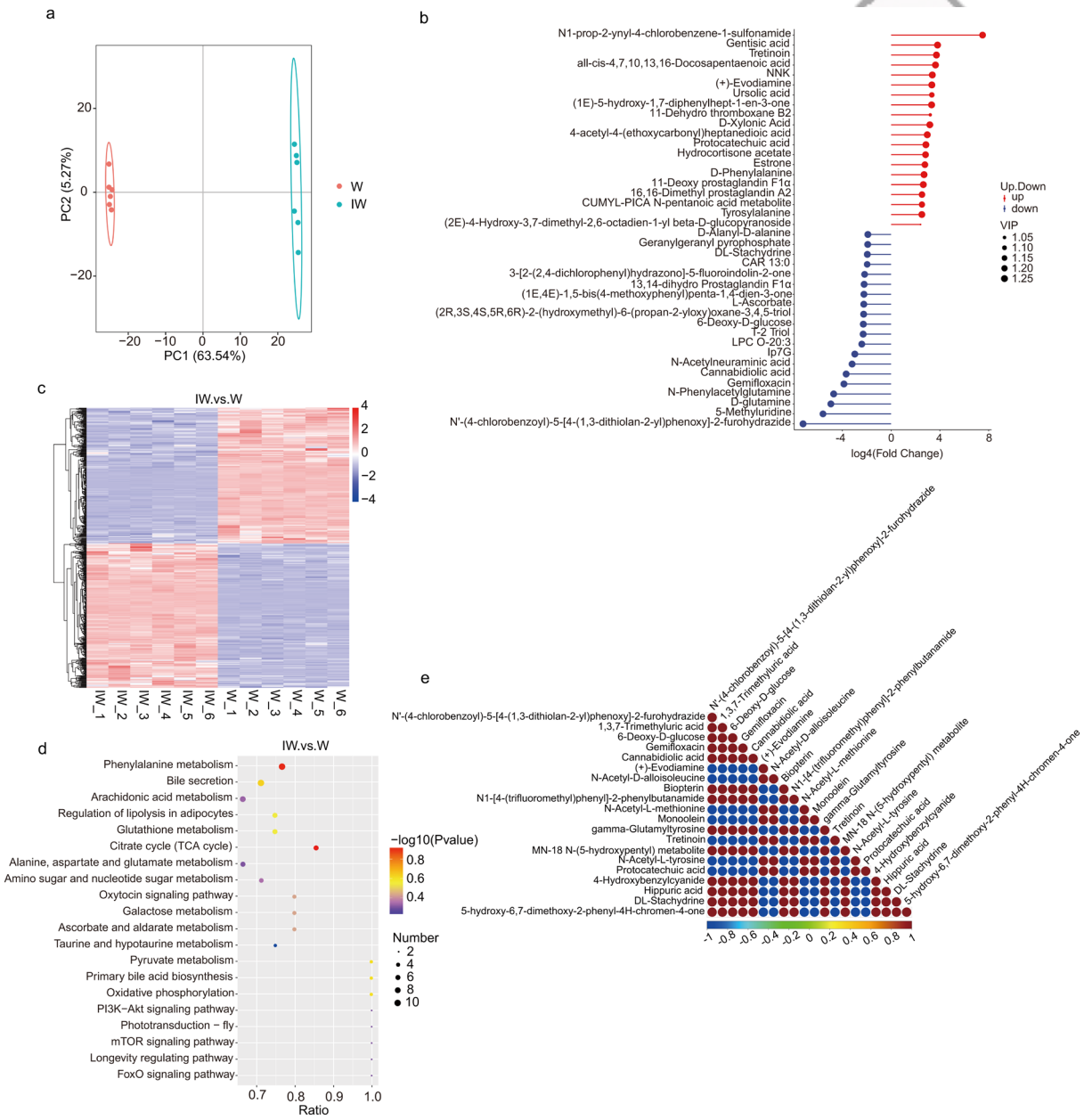

Figure 3 Metabolic changes in "inhibition-activation" of the transformation of female workers to reproductive

a: Results of principal component analysis (PCA). Six replicates for female workers in natal colonies (W) and isolated female workers (IW) were shown as dots in different colors. PC1 and PC2 accounted for 63.54 and 5.27% of variables, respectively. b: Top 20 up- and down-regulated metabolites in W-IW. The length of the rod represented the size of $\log_2$ (Fold Change). c:

Science Press
ZR Zoological Research

Heat map of the differential metabolite cluster analysis. d: Top 20 of KEGG pathway enrichment of differential metabolites. e: Correlations among the top 20 differential metabolites ranked by ascending *P*-value. a correlation coefficient of 1 indicates a perfect positive correlation (red), while a coefficient of –1 represents a perfect negative correlation (blue). Uncolored areas indicate *P*-value>0.05.

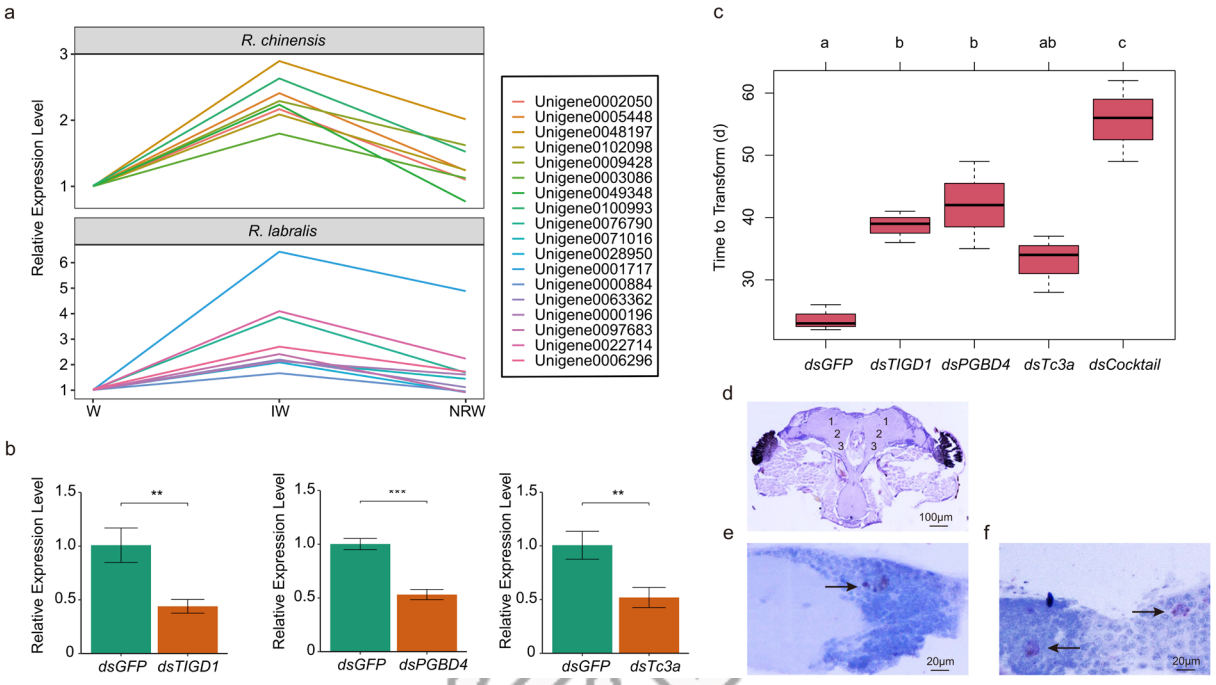

Figure 4 Role of *TIGD1, PGBD4* and *Tc3a* in the transformation of female workers into neotenic reproductives

a: Relative transcript level of 8 and 10 TE-related genes significantly upregulated during W-IW and downregulated during IW-NRW in *R. chinensis* and *R. labralis*, respectively. b: Relative transcript levels of *TIGD1*, *PGBD4* and *Tc3a* 48 h after RNAi. c: The time required for IW to transform into NR after delivering *TIGD1*, *PGBD4, Tc3a* and dsCocktail that targeted the three transcripts (*TIGD1* + *PGBD4* + *Tc3a*). d: HE staining showed that the brains of termites included procerebrum (1), deutocerebrum (2) and tritocerebrum (3). e: PGBD4 immunopositive nerve cells (→) were located in the procerebrum of IW. f: TIGD1 immunopositive nerve cells (→) were located in the procerebrum of IW. **, *P*<0.01; ***, *P*<0.001. W, female workers from natal colonies; IW, isolated female workers.

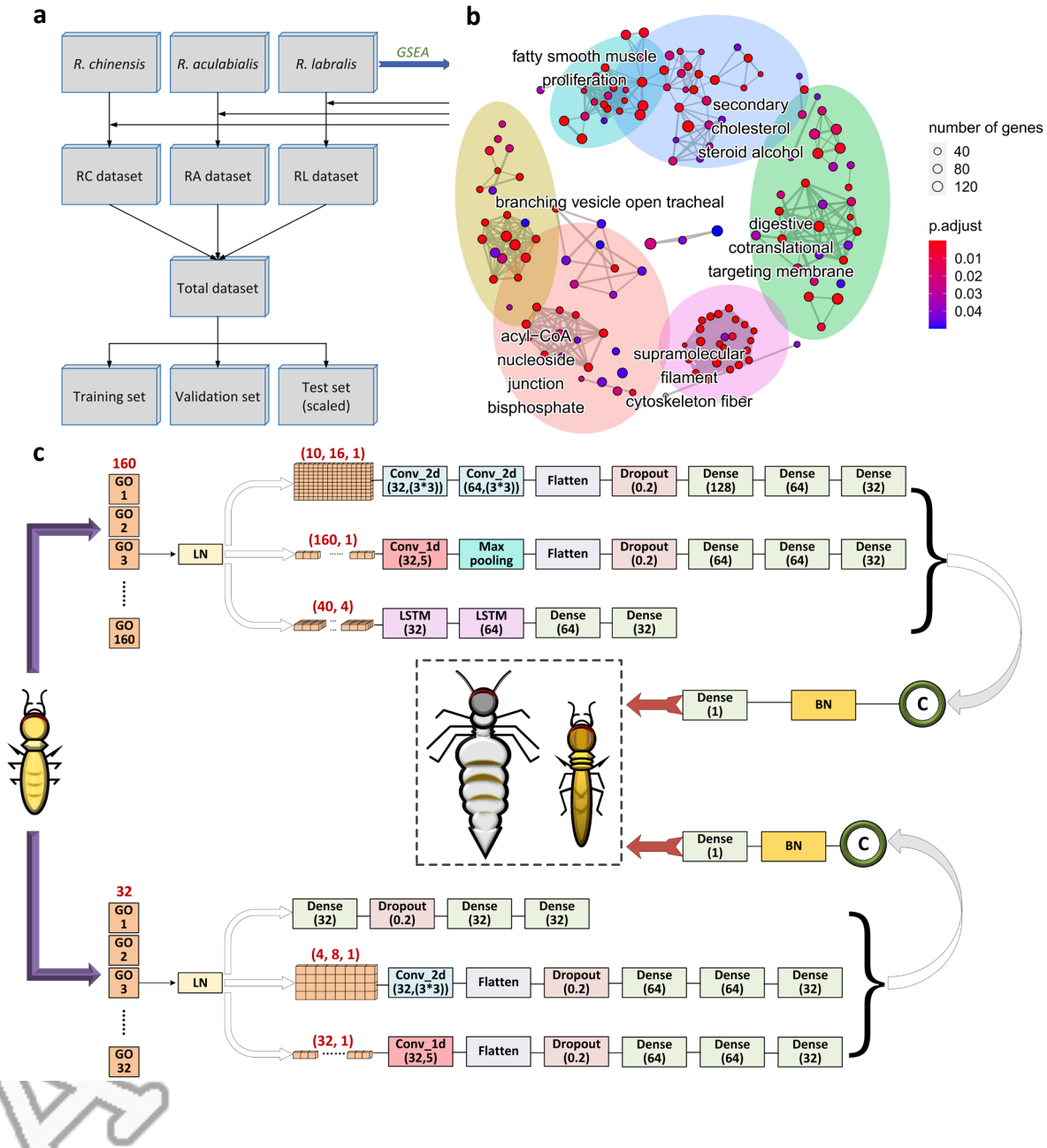

Figure 5 Datasets construction based on the GO functions associated with reproductive plasticity regulation mechanism of female workers and the design of the artificial deep neural networks

a & b: Raw transcriptome data from *R. labralis*, *R. chinensis*, and *R. aculabialis* were utilised to construct the dataset (a). 160 GO functions obtained from GSEA result in *R. labralis* were used as input features. The Jaccard correlation coefficient was used to calculate the correlation between the GO terms (b). For each GO function, a corresponding unigene was randomly selected, and the RPKM value was utilised to quantify the expression level of the GO function. Test dataset was scaled to test the generalization of the models. c: The structures of MBNN and LMBNN models. LN, layer normalization; BN, batch normalization; C, layer concatenation; Worker, the individual on the left side; Reproductives, two individuals in the middle.

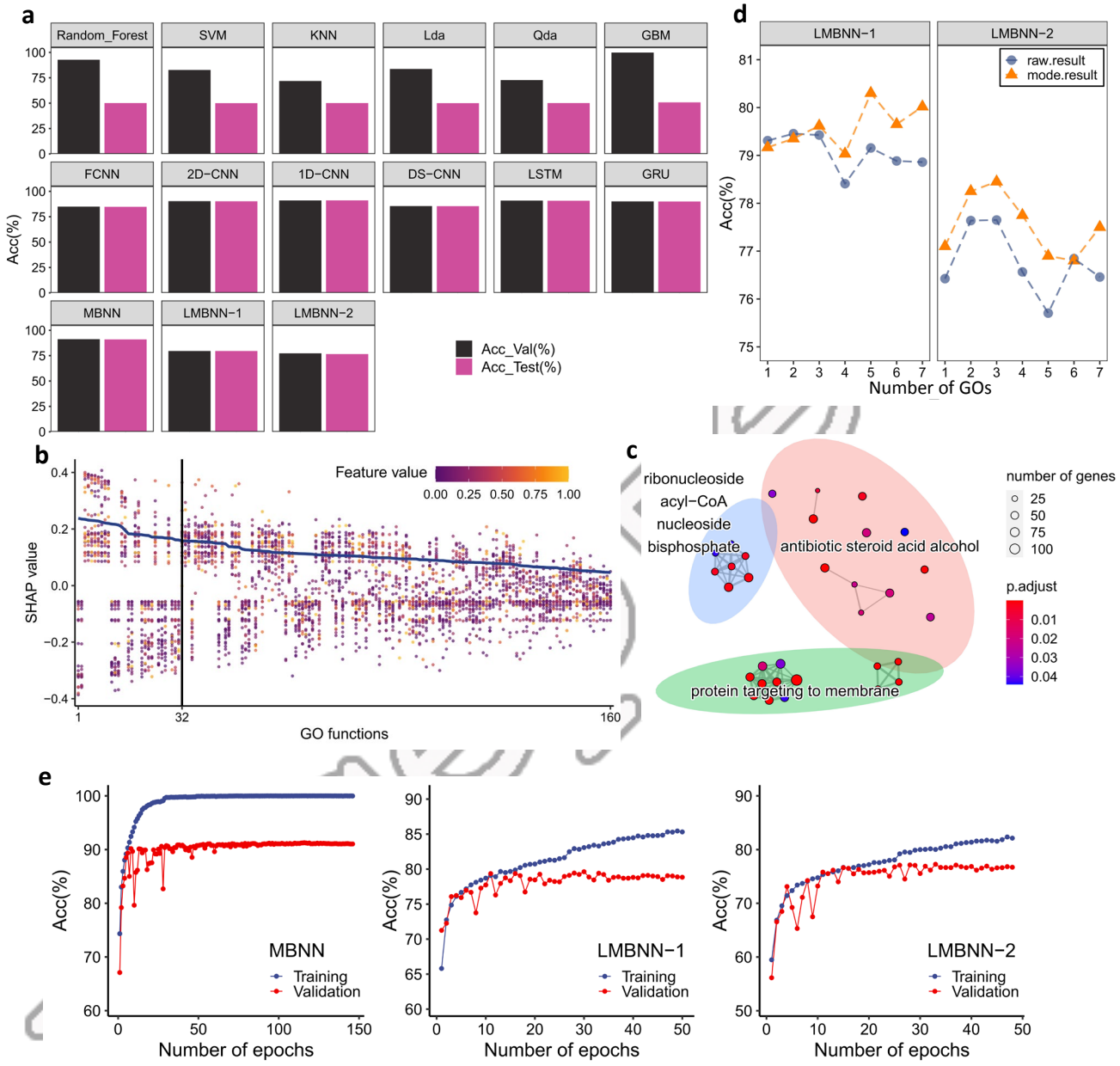

Figure 6 Trained deep neural network models predict the presence of reproductives (fertility) in a colony

a: Prediction performance on the validation and test datasets of all the machine learning models. b&c: SHAP values of the 160 GO functions calculated on 20 ramdom samples and the average absolute SHAP value per feature (b). The top 32 GO functions were selected as the input features of LMBNN and the Jaccard correlation coefficient was used to calculate the correlation between the GO terms (c). d: Prediction performance of LMBNN-1 and LMBNN-2 when randomly selecting two unigenes from the 1-7 GO terms with the highest SHAP values for the combination and using the mode of all results as the final outcome. e: Training and validation accuracy as a function of training epoch.

环境胁迫下白蚁生殖可塑性的分子基础及基于机器学习的群体繁殖力预测

叶晨旭[1]，张文秀[2,3]，刘鑫忆[2,3]，张天戈[2,3]，田圆炎[2,3]，杨艳丽[2,3]，Noor-Us Saba[2,3]，邢连喜[2,3]，李胜[1]，苏晓红[2,3,*]

[1]华南师范大学广东省昆虫发育生物学与应用技术重点实验室，昆虫科学与技术研究所和生命科学学院，广州，广东 510631，中国

[2]西北大学西部资源与生物技术教育部重点实验室，西安，陕西 710069，中国

[3]西北大学生命科学学院，西安，陕西 710069，中国

*通讯作者: sxhnwu@nwu.edu.cn

摘要：

社会性昆虫的生殖可塑性为其巢群提供了巨大的灵活性以应对环境变化和适应新栖息地。然而理解基因如何调控这种可塑性以形成复杂的社会表型仍然是一个挑战。本研究利用三种散白蚁（*Reticulitermes*）建立了雌性工蚁向生殖蚁转化的“抑制–激活–再抑制”模型。通过对该动态周期中雌性工蚁的转录组学分析，我们确定了参与其生殖转化的功能基因，发现雌性工蚁的基因表达谱对蚁后缺失具有强烈响应。进一步研究表明在白蚁巢群受到环境胁迫难以存活时，工蚁不激活与生殖转化相关的基因调控网络。研究还证实了工蚁脑部由 *piggyBac* 和 *tigger* 转座子衍生而来的基因协同促进生殖转化。雌性工蚁的代谢组变化与蚁后及巢内环境的胁迫信号相关，主要涉及脂肪酸和氨基酸代谢通路。最后我们构建了七种机器学习模型，以阐明与生殖可塑性相关的基因表达谱，并预测雌性生殖蚁（蚁后可作为巢群繁殖力指标）的存在与缺失。在这些模型中，深度神经网络模型在分析基因调控网络和预测品级方面展现出最高的准确性与实用性。我们提出雌性工蚁在转录组水平携带了丰富的巢群特异性信息，可以作为巢内信息的指示者。这些发现为理解环境胁迫下社会性昆虫的生殖可塑性提供了重要见解和新的方法。

关键词：白蚁；生殖可塑性；转录组分析；代谢组分析；机器学习

Science Press ZR Zoological Research