

Differences in gene expression patterns between parthenogenetically and sexually produced offspring during early development of *Reticulitermes speratus*

Running title: Gene expression in sexual and asexual termite offspring

Riho Kamiyama¹, Michihiko Takahashi^{1,2}, Mamoru Takata¹, and Kenji Matsuura¹

¹Department of Applied Biosciences, Graduate School of Agriculture, Kyoto University, Kyoto, Japan

²Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan

ACKNOWLEDGEMENTS

We thank Tomoki Ishibashi for his assistance in analyzing. This work was supported by JSPS KAKENHI Grant Numbers 20K20380, 23H00332, JP22H04925 (PAGS).

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest with the contents of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are archived in publicly accessible repositories. Specifically, the data are available in DDBJ under the accession numbers LC850934, PRJDB19121 and PRJDB16588. Analysis code is provided as supplementary material with this manuscript.

Correspondence

Kenji Matsuura, Graduate School of Agriculture, Kyoto University, Kyoto, Japan. Email: matsuura.kenji.6s@kyoto-u.ac.jp

Funding information

This work was supported by JSPS KAKENHI Grant Numbers 20K20380, 23H00332, JP22H04925 (PAGS).

Abstract

Social insects exhibit reproductive division of labor, governed by both external and internal factors influencing caste determination. In termites with a unique reproductive system known as asexual queen succession (AQS), queens produce neotenic queens via parthenogenesis, while workers and alates arise through sexual reproduction. This inherent caste differentiation bias may have resulted from differences in gene expression potentially influenced by the parent-of-origin effect, as parthenogenetic daughters inherit only maternal genomes, while sexually produced daughters inherit both paternal and maternal genomes. Here, we show that gene expression patterns in developing embryos of the termite *Reticulitermes speratus* differ significantly between parthenogenetic and sexually produced offspring. However, SNP analysis indicated that these differences were not attributable to the parent-of-origin effect. Through RNA-seq analysis of female embryos post-katatrepsis, we identified 21 genes, including *jhbp*, *nlk*, and *wge*, which are known to be involved in caste differentiation and morphogenesis, with significant expression differences between parthenogenetic and sexually produced daughters. SNP analysis of sexually produced embryos did not reveal any parent-of-origin biased expression except for mitochondrial genes, though twelve genes exhibited colony-specific expression patterns. These findings suggested that early developmental gene expression partly explained caste differentiation biases. Further research is essential to elucidate the molecular mechanisms behind these transgenerational effects, providing insight into the evolution of AQS and complex caste determination in social insects from a gene expression perspective.

KEY WORDS

genomic imprinting, parthenogenesis, asexual queen succession, gene expression, caste determination

INTRODUCTION

Social insects are characterized by reproductive division of labor, and individuals undergo different developmental processes to grow into castes with distinct morphologies (Wilson, 1971). The factors involved in this caste determination are diverse and can be broadly categorized into exogenous factors and endogenous factors. For example, in some ant species, whether an individual becomes a queen or a worker is determined by the amount

of nutrition provided during the larval stage (Emery, 1896; Wilson, 1971; Michener, 1974; Oster & Wilson, 1978; Wheeler, 1986), which is classified as exogenous caste determination. In other species, caste is determined by genotype (Helms Cahan et al., 2002; Julian et al., 2002; Volny & Gordon, 2002; Helms Cahan & Vinson, 2003; Cahan et al., 2004), which is classified as endogenous caste determination. It is becoming clear that caste determination is influenced by the interaction of these two factors.

Some termite species have a reproductive system called asexual queen succession (AQS), in which workers, soldiers, and alates are produced by sexual reproduction, while neotenic queens are produced by parthenogenesis (Matsuura et al., 2009; Vargo, Labadie, & Matsuura, 2012; Luchetti et al., 2013). In *Reticulitermes speratus* (Heterotermitidae), unfertilized eggs restore diploidy through automixis with terminal fusion (Matsuura, Fujimoto, & Goka, 2004), and the embryos develop into parthenogenetic daughters (Matsuura & Kobayashi, 2007). The evolution of asexual queen succession (AQS) requires not only the capacity for parthenogenetic reproduction but also an innate developmental predisposition for parthenogenetic daughters to differentiate into neotenic queens (Nozaki et al., 2018; Matsuura, 2020). In the non-AQS species *R. okinawanus*, daughters produced via tytoparthenogenesis (accidental parthenogenesis) are more likely to develop into nymphs than their sexually produced counterparts (Nozaki, Yashiro, & Matsuura, 2018). This suggests that the predisposition for offspring with only maternal chromosomes to develop into neotenic queens evolved prior to the origin of AQS. Such a predisposition is also observed in parthenogenetic daughters produced by female–female (FF) founder pairs in *R. speratus*. When raised in incipient colonies by their founding queens, these parthenogenetic daughters differentiate into workers (Matsuura & Nishida, 2001). However, when the same type of eggs are reared by workers from a mature colony, they develop into nymphs (Tamaki, Takata, & Matsuura, 2021). In contrast, sexually produced offspring from male–female (MF) founder pairs develop into workers regardless of whether they are reared by founders or by mature workers (Matsuura & Nishida, 2001; Tamaki, Takata, & Matsuura, 2021). These findings indicate that parthenogenetic daughters possess a latent tendency to differentiate into nymphs. However, this developmental bias is not absolute and can be redirected toward worker differentiation under different socioenvironmental factors.

As explained above, it is evident that some heritable factors influence caste differentiation in termites. In the termite *R. speratus*, parental phenotypes influence the caste fate of offspring, which was once explained by a genetic caste determination model (Hayashi et al., 2007). However, subsequent empirical and theoretical studies have shown

that this phenomenon cannot be explained genetically and can only be accounted for by epigenetic influences from the parents (Matsuura et al., 2018). The genomic imprinting model explains how parental phenotypes influence offspring caste fate through the asymmetry of epigenetic marks inherited from the father and mother (Matsuura et al., 2018). Genomic imprinting can also cause differences in gene expression between parthenogenetic daughters and sexually produced offspring, as parthenogenetic daughters that inherit only maternal genomes may receive different epigenetic marks from their parents compared to sexually produced offspring.

In this study, we investigated whether gene expression differs between parthenogenetically produced daughters and sexually produced offspring at the embryonic stage—prior to the influence of exogenous environmental factors—in the AQS species *R. speratus*, and whether these differences are attributable to genomic imprinting. Embryos at the stage immediately after katasprepsis were chosen because this is the earliest developmental stage at which both RNA and genomic DNA are available in sufficient quantities, enabling both gene expression analysis and sex identification using genetic markers (Table S1). Sexually produced eggs were separated into male and female using a sex-specific genetic marker, and RNA-seq was conducted on three groups: parthenogenetically produced female embryos (PFE), sexually produced female embryos (SFE), and sexually produced male embryos (SME). While pairwise comparison between PFE and SFE is critical to uncover expression differences attributable to reproductive mode, the inclusion of SME allows deeper insight into biological interpretation (Fig. 1a). PFE differ from both SFE and SME in terms of reproductive origin, whereas SME differ from PFE and SFE in terms of sex. Furthermore, under identical rearing conditions (i.e., when raised by workers), only PFE are predisposed to differentiate into nymphs, while both SFE and SME typically develop into workers. Thus, this three-way comparison enables us to disentangle gene expression differences associated with reproductive origin, sex, and caste fate tendency. To explore the potential contribution of parent-of-origin effects to these expression differences, we also performed SNP-based allele-specific expression analysis. For this, reciprocal crosses between colonies collected from Kyoto and Tsukuba were established to maximize SNP coverage. Whole-genome sequencing of parents and RNA-seq of offspring allowed us to distinguish between paternal and maternal alleles and to compare their expression levels.

RESULTS

Genes with differential expression among PFE, SFE, and SME

We identified 54 genes with significant differences among the three groups: parthenogenetically produced female embryos (PFE), sexually produced female embryos (SFE), and sexually produced male embryos (SME) (FDR < 0.05, ANOVA; Table S5). Among the 21 genes with significant differences in expression levels between PFE and SFE (FDR < 0.05, Tukey's HSD; Table S5), six genes (*jhbpl*, *tretl*, *top1*, *ccha2-r*, *med26*, and *cora*) showed higher expression in PFE (Fig. 2a, Table S5), whereas 15 genes including *nlk* and *wge* exhibited lower expression (Fig. 2b; 2c, Table S5). We identified 16 genes with significantly higher expression in males compared to females (PFE and SFE), including *dsx* (Fig. 3a, Table S5). Among sexually produced embryos, nine genes had significantly higher expression levels in SFE than in SME (Fig. 3bc, Table S5). BLAST homology searches revealed that 48 out of the 54 genes showed differential expression among PFE, SFE, and SME (Table S6, S7).

To investigate whether alternative splicing contributes to gene expression differences among PFE, SFE, and SME, we performed a splicing variant analysis. A total of 96 genes exhibited alternative splicing isoforms (Supplementary Table S8). However, none of the 54 differentially expressed genes (DEGs) among PFE, SFE, and SME showed splicing variants. This suggests that the transcriptional differences observed are not attributable to alternative splicing but rather to changes in gene expression levels. Nonetheless, some genes with strong splicing variation across pairwise comparisons (e.g., ANN12100, ANN28563, ANN28610) may play roles in sex- or reproductive mode-specific developmental pathways, warranting further functional investigation.

Genes with allele-specific expression

There were 8,953 SNPs with homozygous bases differing between the father and mother, which can be used to distinguish paternal or maternal origin. Of these, 2,641 SNPs were located within 847 genes, either in introns or exons, while the remaining SNPs were located in non-coding regions. We found that 13 genes exhibited maternally biased expression, all of which were related to mitochondria (Fig. 4b, Table S9, S10). Therefore, this SNP analysis successfully detects genes, if any, with maternally or paternally biased expression. However, apart from mitochondria-related genes inherited maternally, no genes showed maternally or paternally biased expression (Fig. 4b). Among the 54 differentially expressed genes (DEGs) between PFE, SFE, and SME, four genes were suitable for comparing maternal and paternal expression through SNP analysis; however, none of these showed a parent-of-origin effect (Fig. 4c). Although no genes exhibited a

parent-of-origin effect, 12 genes showed expression significantly dependent on the natal colony of the parent, with five showing higher expression of alleles from Kyoto and seven showing higher expression of alleles from Tsukuba (Fig. 5; Table S11). Six out of the 12 genes with colony-specific allele expression, had high homology with known genes, while the remaining genes did not show significant hits in BLAST searches and could not be translated into long amino acid sequences (Table S12, S13).

Whole mitochondrial genome assembly

The complete mitochondrial genome was 15,549 base pairs (bp) in length and contained a total of 37 genes, including 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, and 2 ribosomal RNA (rRNA) genes. The data are available in DDBJ under the accession number LC850934.

DISCUSSION

Our results demonstrate that parthenogenetically and sexually produced embryos in *R. speratus* exhibit significant differences in gene expression prior to hatching. Under identical rearing conditions (i.e., when raised by workers from a mature colony), parthenogenetic daughters tend to differentiate into nymphs, whereas sexually produced offspring develop into workers (Tamaki, Takata, & Matsuura, 2021). These observations suggest that caste fate differences between the two reproductive modes are, at least in part, programmed during embryogenesis. In this study, we identified 21 genes with significantly different expression levels between parthenogenetically produced female embryos and sexually produced female embryos. These genes are likely involved in regulating caste fate at an early stage, before exogenous socioenvironmental factors come into play.

The origin of AQS is thought to require the simultaneous evolution of parthenogenetic ability and a developmental predisposition for parthenogenetic daughters to become neotenic queens (Nozaki et al., 2018; Matsuura, 2020). Previous studies have addressed the molecular basis of parthenogenesis by comparing early-stage gene expression between AQS and non-AQS species (Peng et al., 2023). Our study expands on this framework by focusing on a later embryonic stage—immediately after katatrepsis—when organogenesis has begun, enabling us to capture caste fate-related gene expression patterns with greater resolution. Moreover, by incorporating sex identification at the egg

stage, our analysis avoids confounding effects and allows precise comparisons not only between parthenogenetic and sexually produced daughters, but also between sexes.

Among the genes specifically expressed in parthenogenetically produced embryos, *Rsjhbp*, *Rsnlk*, and *Rswge* were particularly noteworthy. *Rsjhbp* which showed higher expression in parthenogenetically produced embryos, encodes a protein that binds to juvenile hormone (JH) (Kramer et al., 1974). Because JH is lipophilic, the water-soluble protein JHBP enables its transport through the aqueous hemolymph by binding to it (Gilbert, Granger, & Roe, 2000). JH is known to regulate molting in insects (Kramer et al., 1974; Truman & Riddiford, 1999), but it also plays a role in caste differentiation in social insects. In *Reticulitermes* species, JH levels increase during differentiation into presoldiers and soldiers (Hrdy, 1985; Okot-Kotber et al., 1993; Mao et al., 2005; Park & Raina, 2003, 2004, 2005; Scharf et al., 2005; Elliott & Stay, 2008). Furthermore, the model proposed by Nijhout and Wheeler (1982) states that JH titers during the three sensitive periods between two molts determine whether an individual molts into a worker, a neotenic, or a presoldier/soldier (Nijhout & Wheeler, 1982). Additionally, JH promotes ovarian development (Scharf et al., 2003). In other termite genera, fluctuations in JH levels also influence caste differentiation (Zhou, Tarver, & Scharf, 2007; Cornette et al., 2008). Although the effects of JH on caste differentiation during embryonic development have not been investigated in termites, an increase in JH levels during embryogenesis induces differentiation into the queen caste in ants (Nijhout, 2021). Therefore, it is plausible that JH levels during embryogenesis also influence caste differentiation in *R. speratus*, and the variation in *Rsjhbp* expression observed in this study may be related to this process. It remains unknown whether JH influences caste determination during embryonic development, and it is highly intriguing to investigate how JH levels change throughout the embryonic development in parthenogenetic and sexually produced individuals of termites. *Rsnlk* and *Rswge* are also important genes: *nemo*, a homolog of *nlk* in *Drosophila*, is involved in wing formation (Brott, Pinsky, & Erikson, 1998; Katsuyama et al., 2005; Zeng et al., 2007), and *wge* is associated with eye formation (MacKay et al., 2003). In termite, nymphs have wings and eyes, while workers lack these structures (Maekawa et al., 2008). Therefore, the differences in the expression of *Rsnlk* and *Rswge* may be involved in the future morphological differences between parthenogenetic and sexually produced embryos. It is necessary to investigate how the expression of these genes changes and how eyes and wings are formed from hatching to the third instar, when clear morphological differences between castes emerge.

The genetic marker used for sex identification of sexually produced embryos was originally developed by Hayashi et al. (2017) and has been validated for sexing individuals at the first instar or later stages (Hayashi et al., 2017; Takata et al., 2020). In this study, we applied this marker to early-stage embryos for the first time, utilizing residual genomic DNA present in total RNA extracts prior to DNase treatment. The RNA-seq analysis revealed substantial and consistent differences in gene expression profiles between embryos classified as male and female, strongly supporting the overall reliability of our sex identification method. Nevertheless, we acknowledge that faint male-specific bands were detected in a few samples identified as female (Fig. 1b), indicating the possibility of low-level contamination or incomplete exclusion of the opposite sex. While such contamination, if present, is likely to be minimal, it may have contributed to underestimating the number of sex-biased genes due to signal dilution in pooled samples. Despite this limitation, the observed gene expression patterns showed clear, biologically consistent sex differences, suggesting that the sexing method was sufficiently robust for differential gene expression analysis at the embryo pool level. For future studies, single-embryo RNA-seq would offer even greater resolution and allow precise quantification of sex-specific expression patterns at the individual level.

In the comparison of gene expression among the three groups (PFE, SFE, and SME), 32 genes showed sex-specific expression, with a notable example being *Rsdsex*. The *dsx* is well-known for its role in sex determination in many insects (Verhulst, van de Zande, & Beukeboom, 2010), and in *R. speratus*, this gene consists of a single exon containing a DM domain (lacking a conserved OD sequence) and is transcribed only in males, but not in females (Miyazaki et al., 2021). Our study corroborates this, as *Rsdsex* showed high expression in males, while in females, both parthenogenetically produced and sexually produced embryos exhibited very low expression levels. This indicates that *Rsdsex* is already showing male-specific expression at the embryonic stage.

Among other genes showing differential expression between the three groups, 17 were involved in regulating the expression of other genes, such as those related to histone methylation, splicing, and transcription factors (Table S7). These genes likely act as upstream regulators of caste differentiation or sex development and may contribute to organ formation post-katatrespsis and caste- or sex-specific gene expression during post-hatching development. Additionally, five genes associated with cell adhesion were identified, which may be involved in caste- or sex-specific organ development. Other genes included five related to energy storage and metabolism, and four involved in stress response. In ants, caste-specific differences in energy metabolism and immune responses

to stress have been observed (Quque et al., 2021), and similar patterns in *R. speratus* might explain the differential expression of metabolism and stress-related genes between castes. Furthermore, the eight genes exhibiting specific expression in SFE may play roles in suppressing female sexual development. This suppression is a critical function in social insects that exhibit reproductive division of labor, and these genes could be key players in the evolution of this trait.

Such differences in gene expression between parthenogenetic daughters and sexually produced offspring are likely to be influenced by epigenetics. The transmission of epigenetic marks from parents to offspring is well-documented (Jablonka & Raz, 2009; Molaro et al., 2011; Trerotola et al., 2015). In sexual reproduction, offspring inherit epigenetic marks from both parents, whereas parthenogenetic daughters of *R. speratus* inherit marks only from the mother. Therefore, if there is asymmetry in the epigenetic marks inherited from the father and mother, it could lead to differences in the epigenetic profiles between parthenogenetic daughters and sexually produced offspring, potentially explaining the observed differences in gene expression. In termites, it is known that parental phenotypes influence the caste fate of offspring (Hayashi et al., 2007), and the genomic imprinting model has been proposed to explain this phenomenon (Matsuura et al., 2018). This model posits that asymmetry in the epigenetic marks inherited from the parents leads to differences in caste fate. Thus, genomic imprinting is likely to be involved in the differences in caste fate between parthenogenetic daughters and sexually produced offspring.

In the allele-specific expression analysis using SNPs, no nuclear genes exhibited biased expression towards either the paternal or maternal allele. However, mitochondrial genes showed maternal-biased expression, suggesting that the allele-specific expression analysis was conducted reliably. This suggests that our allele-specific expression analysis was reliable. There are two possible interpretations of these results: either termites do not exhibit genomic imprinting, or imprinting does exist but was not detected with the methods used in this study. It is premature to conclude the former. In this experiment, we analyzed allele-specific expression in only 847 genes, representing only a few percent of the approximately 20,000 genes in termites. It is possible that genes not included in this analysis exhibit allele-specific expression. Furthermore, the embryos used in this study were collected immediately after katrepsis, representing only a snapshot of gene expression during the 35-day embryonic development period in termites (Matsuura & Kobayashi, 2007). Therefore, to definitively determine whether genomic imprinting exists in termites, future studies should aim to increase the number of SNPs analyzed to

encompass all genes and examine gene expression throughout the entire embryonic development period. Once genes subject to genomic imprinting are identified, knockout or knockdown experiments will be necessary to investigate their role in termite caste differentiation. Furthermore, molecular techniques such as bisulfite sequencing should be employed to determine whether differential methylation occurs between alleles, which could provide insights into the molecular mechanisms underlying genomic imprinting.

On the other hand, we identified several genes showing colony-specific allele expression, with five genes exhibiting higher expression of the Kyoto allele and seven genes showing higher expression of the Tsukuba allele. While nearly 90% of the genes showing differential expression between PFE, SFE, and SME could be functionally annotated using BLAST (48 out of 54 genes), only half of the colony-specific allele expression genes could be annotated (6 out of 12 genes). The transcripts with unknown functions did not show significant hits in BLAST searches and could not be translated into long amino acid sequences, suggesting that they may be novel non-coding RNAs. Non-coding RNAs encompass a diverse range of groups, including long non-coding RNAs (lncRNAs), microRNAs, and small interfering RNAs (Rinn & Chang, 2012; Ulitsky & Bartel, 2013). In particular, lncRNAs have been found to play diverse roles in various insects, with studies showing that in honeybees, the expression of lncRNAs changes in response to heat stress (Zafar et al., 2023; Zhang et al., 2024). It is possible that, in termites, non-coding RNA expression may differ between colonies as part of an adaptation to different environments. While it remains unclear whether the colony-specific allele expression observed in this experiment is due to genetic differences or the asymmetry of epigenetic inheritance, the results suggest that non-coding regions may be involved in adaptation to different environmental conditions.

This study has provided important insights into the evolutionary process of AQS and the complex caste determination mechanisms in termites by revealing innate differences in gene expression between parthenogenetically produced and sexually produced embryos.

EXPERIMENTAL PROCEDURES

Termite collection

In April 2023, four colonies (A, B, D, and F) of *Reticulitermes speratus* containing nymphs were collected in Kyoto Prefecture, and two colonies (C and E) containing nymphs were collected in Ibaraki Prefecture, Japan (Table S2). These colonies, together

with nest woods, were transported to the laboratory and kept at 20°C in darkness until the nymphs molted into alates.

The woods containing alates were transferred to a room maintained at 25°C under constant light to induce swarming. Following swarming, alates were collected, their wings were removed, and they were sexed in preparation for the next step of the experiment. These individuals with their wings removed are called dealates.

Colony foundation and sample preparation for gene expression analysis

All subsequent experimental procedures were also performed separately for each colony. Approximately 50 female-female pairs and male-female pairs were formed using dealates and placed in petri dishes (ca. 40 × 11 mm) containing brown-rotted pinewood mixed (BPC) medium blocks (Mitaka, Akino, & Matsuura, 2023) and stored at 25°C in darkness. The nest woods were dissected after swarming, and workers were extracted. Randomly selected 50 workers were placed in petri dishes (ca. 40 × 11 mm) containing a BPC block and stored at 25°C in darkness.

Eggs laid within one day were collected from both female-female and male-female pairs and transferred to petri dishes containing BPC blocks and 50 workers for fostering, with parthenogenetically produced eggs being obtained from female-female pairs and sexually produced eggs from male-female pairs. Using nestmate pairs allowed us to eliminate potential confounding effects arising from genetic differences between founders or colony-of-origin effects. This experimental design ensured that observed differences in gene expression could be attributed specifically to reproductive origin rather than to variation in genetic background or compatibility. Previous research has demonstrated that nestmate and non-nestmate pairs do not differ significantly in terms of offspring survival rate, total number of progeny, or number of hatching larvae (Matsuura, Fujimoto, & Goka, 2004). Fifteen days after egg-laying, the eggs were checked daily under a stereomicroscope, and those immediately after katatrepsis were collected for RNA extraction. Worker-laid eggs were not mixed among these samples, as it takes 27.6 ± 1.2 days for female workers to differentiate into ergatoids (Miyata, Furuichi, & Kitade, 2004), and the experimental period was shorter than this duration.

Total RNA was extracted from individual sexually produced eggs and from parthenogenetically produced eggs using the RNAqueous-Micro Kit (Thermo Fisher Scientific) following the manufacturer's protocol. For sexually produced eggs, each embryo was processed individually for RNA extraction and subsequently subjected to sex identification using a PCR-based method prior to pooling. Specifically, 1 µl of the RNA

solution prior to DNase treatment was used as a source of genomic DNA for sex-specific PCR with Phusion High-Fidelity DNA Polymerase (New England Biolabs) and the RADtag_467410 marker (Hayashi et al., 2017; Takata et al., 2020). PCR conditions were as follows: an initial denaturation at 98°C for 30 seconds; 35 cycles of 98°C for 10 seconds, 65°C for 30 seconds, and 72°C for 1 minute; followed by a final extension at 72°C for 10 minutes. The doublesex (dsx) gene, located on an autosomal chromosome, was used as a positive control under the same PCR conditions. Samples showing a strong band for the male-specific marker were classified as male, while those showing only the dsx control band and no male-specific band were classified as female. Only embryos clearly identified as male or female were selected for RNA-seq. Approximately 10 embryos of the same sex from each colony were then pooled to obtain sufficient RNA for downstream analysis. For parthenogenetic eggs, which were all genetically female, RNA was extracted directly from pooled samples of 10 embryos. The samples used for gene expression analysis were derived from colonies A to E. However, due to poor RNA yield from sexually produced male embryos in colony E, a sample from colony F was used instead.

The library preparation, RNA sequencing, and initial data processing were outsourced to Rhelixa (Tokyo, Japan). RNA samples from parthenogenetically produced female embryos (PFE), sexually produced female embryos (SFE), and sexually produced male embryos (SME) were processed using the SMART-Seq® HT Plus Kit (Takara Bio). Sequencing libraries were prepared and sequenced on the Illumina NovaSeq 6000 platform to generate paired-end reads (150 bp × 2). The data are available in DDBJ under the accession number RJDB19121. The overview of the experimental methods is shown in Fig. 2b.

Gene expression analysis

Raw reads were processed to remove adapter sequences and reads shorter than 25 bp using Fastp (version 0.23.2). The cleaned reads were then mapped to the reference genome of *R. speratus* (Accession: PRJNA779705; (Takata et al., 2024)) and its mitochondrial genome (Accession: LC850934) using STAR (version 2.7.10a). The whole mitochondrial genome assembly was conducted as part of this study and is described in detail below. The reference genome index for mapping was prepared using RSEM (version 1.3.3). Gene expression levels were quantified using RSEM (version 1.3.3), and differentially expressed genes (DEGs) were identified using edgeR (version 3.40.2) based on RSEM count data. Statistical analyses were performed to determine significant

differences in gene expression between the groups using a false discovery rate (FDR) threshold of 0.05. The differences in gene expression levels among the three groups (PFE, SFE, and SME) were statistically analyzed using ANOVA and Tukey's HSD test at a 95% confidence level. DEGs were further analyzed for sequence similarity using BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to identify homologous genes. BLAST searches were primarily conducted using blastx; however, for sequences that were too long to be searched with blastx or whose blastx results did not align with the gene expression regions, searches were performed using blastp with the amino acid sequences of *R. speratus* (Takata et al., 2024). The analysis code is provided in the Supplementary Materials.

Splicing variant analysis was performed using rMATS (version 4.1.2), a software tool designed to detect differential alternative splicing events from RNA-seq data. Pairwise comparisons were conducted among PFE, SFE, and SME groups, and splicing events were considered significant based on a Bonferroni-corrected false discovery rate (FDR < 0.05). The list of genes with significant splicing variants is provided in Table S8.

Colony foundation and sample preparation for allele-specific expression analysis

The dealates from colony A (collected in Kyoto city) and colony E (collected in Tsukuba Mountain), prepared in the "Termite collection" section, were used. Five pairs of males from the Kyoto colony and females from the Tsukuba colony, as well as five pairs of males from the Tsukuba colony and females from the Kyoto colony, were formed and placed in petri dishes containing BPC blocks (ca. 40 × 11 mm) and kept at 20°C in darkness.

The eggs laid by each pair were fostered by the pair until they reached the stage immediately after katatrepsis and were then collected for RNA extraction. RNA extraction was performed using the RNAqueous-Micro Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. DNA was extracted from the dealate males and females, which were the parents of the eggs. For the females, the spermatheca was removed prior to DNA extraction. DNA extraction was carried out following proteinase K treatment using the Maxwell RSC Plant DNA Kit protocol.

The library preparation, RNA sequencing, and initial data processing of the egg samples were outsourced to Rhelixa (Tokyo, Japan). The data are available in DDBJ under the accession number PRJDB19121. The resequencing of parental samples was supported by JSPS KAKENHI Grant Number JP22H04925 (PAGS). The data are available in DDBJ under the accession number PRJDB16588. The overview of the experimental methods is shown in Fig. 4a.

Allele-specific expression analysis

Raw reads from both parental resequencing data and egg RNA-seq data were processed using fastp (version 0.23.2) to remove adapter sequences and reads shorter than 25 bp. Cleaned reads from parental resequencing data were mapped to the reference genome (Accession: PRJNA779705; (Takata et al., 2024)) and its mitochondrial genome (<https://doi.org/10.6084/m9.figshare.27435600>) using bwa-mem2 (version 2.2.1), while reads from egg RNA-seq data were mapped using hisat2 (version 2.2.1), with the reference genome index prepared using hisat2-build. The whole mitochondrial genome assembly was conducted as part of this study and is described in detail below. SAM files were converted to BAM files using samtools (version 1.15.1), and BAM files were filtered with bamtools (version 2.5.2). Indexing of BAM files was performed using samtools. For SNP analysis, GATK (version 4.2.6.1) was used. Variant calling was performed using parental resequencing and egg RNA-seq data, and the resulting variants were used for BQSR. Allele-specific expression levels in eggs were obtained using ASEReadCounter and merged with parental genotype data. Statistical analyses for expression differences between alleles were performed using edgeR (version 3.40.2). The analysis code is provided in the Supplementary Materials.

Whole mitochondrial genome assembly

The nuclear genome of *R. speratus* is known (<https://0-www-ncbi-nlm-nih-gov.brum.beds.ac.uk/bioproject/PRJNA779705/>), but the mitochondrial genome sequence was not available. Therefore, we performed sequencing and annotation of the mitochondrial genome in this study. In May 2022, colony G was collected in Shiga Prefecture, Japan, and DNA was extracted from a final instar male nymph following the protocol of the Maxwell® RSC Plant DNA Kit (Promega).

The library construction, short-read sequencing, and initial data processing were outsourced to MacroGen (Tokyo, Japan). The library was constructed using the TruSeq DNA PCR-Free Kit (Illumina, San Diego, CA, USA) with 1 µg of DNA, and sequencing was performed on the Illumina NovaSeq 6000 platform. Raw reads were converted into FASTQ format. The mitochondrial genome assembly was performed using GetOrganelle v1.7.0. The mitochondrial genome, including protein-coding genes (PCGs), rRNAs, and tRNAs, was annotated using the MITOS2 webserver (<http://mitos2.bioinf.uni-leipzig.de/index.py>). The analysis code is provided in the Supplementary Materials.

484

485 AUTHOR CONTRIBUTIONS

486 Riho Kamiyama: Formal analysis; investigation; methodology; data curation;
487 visualization; writing – original draft.

488 Michihiko Takahashi: Formal analysis; investigation; methodology; writing – review and
489 editing.

490 Mamoru Takata: Methodology; supervision.

491 Kenji Matsuura: Conceptualization; funding acquisition; investigation; methodology;
492 supervision; writing – review and editing.

493

494

495 REFERENCE

496 Brott, B. K., Pinsky, B. A., & Erikson, R. L. (1998). Nlk is a murine protein kinase related
497 to Erk/MAP kinases and localized in the nucleus. *Proceedings of the National Academy*
498 *of Sciences of the United States of America*, 95(3), 963–968.

499 Cahan, S. H., Julian, G. E., Rissing, S. W., Schwander, T., Parker, J. D., & Keller, L.
500 (2004). Loss of phenotypic plasticity generates genotype-caste association in harvester
501 ants. *Current Biology: CB*, 14(24), 2277–2282.

502 Cornette, R., Gotoh, H., Koshikawa, S., & Miura, T. (2008). Juvenile hormone titers and
503 caste differentiation in the damp-wood termite *Hodotermopsis sjostedti* (Isoptera,
504 Termopsidae). *Journal of Insect Physiology*, 54(6), 922–930.

505 Elliott, K. L., & Stay, B. (2008). Changes in juvenile hormone synthesis in the termite
506 *Reticulitermes flavipes* during development of soldiers and neotenic reproductives
507 from groups of isolated workers. *Journal of Insect Physiology*, 54(2), 492–500.

508 Emery, C. (1896). Le polymorphisme des fourmis et la castration alimentaire. *Compte*
509 *Rendu Des Séances Du Troisième Congrès Internationale de Zoologie*, 395–410.

510 Gilbert, L. I., Granger, N. A., & Roe, R. M. (2000). The juvenile hormones: historical
511 facts and speculations on future research directions. *Insect Biochemistry and Molecular*
512 *Biology*, 30(8–9), 617–644.

513 Hayashi, Y., Lo, N., Miyata, H., & Kitade, O. (2007). Sex-linked genetic influence on
514 caste determination in a termite. *Science*, 318(5852), 985–987.

- Hayashi, Y., Oguchi, K., Yamaguchi, K., Kitade, O., Maekawa, K., Miura, T., & Shigenobu, S. (2017). Male-specific molecular genetic markers in the Japanese subterranean termite *Reticulitermes speratus*. *Insectes Sociaux*, 64(3), 357–364.
- Helms Cahan, S., Parker, J. D., Rissing, S. W., Johnson, R. A., Polony, T. S., Weiser, M. D., & Smith, D. R. (2002). Extreme genetic differences between queens and workers in hybridizing *Pogonomyrmex* harvester ants. *Proceedings. Biological Sciences / The Royal Society*, 269(1503), 1871–1877.
- Helms Cahan, S., & Vinson, S. B. (2003). Reproductive division of labor between hybrid and nonhybrid offspring in a fire ant hybrid zone. *Evolution; International Journal of Organic Evolution*, 57(7), 1562–1570.
- Hrdy, I. (1985). The role of juvenile hormones and juvenoids in soldier formation in Rhinotermitidae. *Caste Differentiation in Social Insects*.
- Jablonka, E., & Raz, G. (2009). Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *The Quarterly Review of Biology*, 84(2), 131–176.
- Julian, G. E., Fewell, J. H., Gadau, J., Johnson, R. A., & Larrabee, D. (2002). Genetic determination of the queen caste in an ant hybrid zone. *Proceedings of the National Academy of Sciences of the United States of America*, 99(12), 8157–8160.
- Katsuyama, T., Sugawara, T., Tatsumi, M., Oshima, Y., Gehring, W. J., Aigaki, T., & Kurata, S. (2005). Involvement of winged eye encoding a chromatin-associated bromo-adjacent homology domain protein in disc specification. *Proceedings of the National Academy of Sciences of the United States of America*, 102(44), 15918–15923.
- Kramer, K. J., Sanburg, L. L., Kézdy, F. J., & Law, J. H. (1974). The juvenile hormone binding protein in the hemolymph of *Manduca sexta* Johannson (Lepidoptera: Sphingidae). *Proceedings of the National Academy of Sciences of the United States of America*, 71(2), 493–497.
- Li, R., Dai, X., Zheng, J., Larsen, R. S., Qi, Y., Zhang, X., et al. (2024). Juvenile hormone as a key regulator for asymmetric caste differentiation in ants. *Proceedings of the National Academy of Sciences of the United States of America*, 121(46), e2406999121.
- Luchetti, A., Velonà, A., Mueller, M., & Mantovani, B. (2013). Breeding systems and reproductive strategies in Italian *Reticulitermes* colonies (Isoptera: Rhinotermitidae). *Insectes Sociaux*, 60(2), 203–211.
- MacKay, J. O., Soanes, K. H., Srivastava, A., Simmonds, A., Brook, W. J., & Bell, J. B. (2003). An in vivo analysis of the vestigial gene in *Drosophila melanogaster* defines the domains required for Vg function. *Genetics*, 163(4), 1365–1373.

550 Maekawa, K., Mizuno, S., Koshikawa, S., & Miura, T. (2008). Compound eye
551 development during caste differentiation in the termite *Reticulitermes speratus*
552 (Isoptera: Rhinotermitidae). *Zoological Science*, 25(7), 699–705.

553 Mao, L., Henderson, G., Liu, Y., & Laine, R. A. (2005). Formosan subterranean termite
554 (Isoptera: Rhinotermitidae) soldiers regulate juvenile hormone levels and caste
555 differentiation in workers. *Annals of the Entomological Society of America*, 98(3), 340–
556 345.

557 Matsuura, K., Fujimoto, M., & Goka, K. (2004). Sexual and asexual colony foundation
558 and the mechanism of facultative parthenogenesis in the termite *Reticulitermes*
559 *speratus* (Isoptera, Rhinotermitidae). *Insectes Sociaux*, 51(4), 325–332.

560 Matsuura, K. (2020). Genomic imprinting and evolution of insect societies. *Population*
561 *Ecology*, 62(1), 38–52.

562 Matsuura, K., & Kobayashi, N. (2007). Size, hatching rate, and hatching period of
563 sexually and asexually produced eggs in the facultatively parthenogenetic termite
564 *Reticulitermes speratus* (Isoptera: Rhinotermitidae). *Applied Entomology and Zoology*,
565 42(2), 241–246.

566 Matsuura, K., Mizumoto, N., Kobayashi, K., Nozaki, T., Fujita, T., Yashiro, T., et al.
567 (2018). A genomic imprinting model of termite caste determination: not genetic but
568 epigenetic inheritance influences offspring caste fate. *The American Naturalist*, 191(6),
569 677–690.

570 Matsuura, K., & Nishida, T. (2001). Comparison of colony foundation success between
571 sexual pairs and female asexual units in the termite *Reticulitermes speratus* (Isoptera:
572 Rhinotermitidae). *Population Ecology*, 43(2), 119–124.

573 Matsuura, K., Vargo, E. L., Kawatsu, K., Labadie, P. E., Nakano, H., Yashiro, T., & Tsuji,
574 K. (2009). Queen succession through asexual reproduction in termites. *Science*,
575 323(5922), 1687.

576 Michener, C. D. (1974). The social behavior of the bees: a comparative study. *Harvard*
577 *University Press*.

578 Mitaka, Y., Akino, T., & Matsuura, K. (2023). Development of a standard medium for
579 culturing the termite *Reticulitermes speratus*. *Insectes Sociaux*.

580 Miyata, H., Furuichi, H., & Kitade, O. (2004). Patterns of neotenic differentiation in a
581 subterranean termite, *Reticulitermes speratus* (Isoptera: Rhinotermitidae).
582 *Entomological Science*, 7(4), 309–314.

583 Miyazaki, S., Fujiwara, K., Kai, K., Masuoka, Y., Gotoh, H., Niimi, T., et al. (2021).
 584 Evolutionary transition of doublesex regulation from sex-specific splicing to male-
 585 specific transcription in termites. *Scientific Reports*, 11(1), 15992.
 586 Molaro, A., Hodges, E., Fang, F., Song, Q., McCombie, W. R., Hannon, G. J., & Smith,
 587 A. D. (2011). Sperm methylation profiles reveal features of epigenetic inheritance and
 588 evolution in primates. *Cell*, 146(6), 1029–1041.
 589 Nijhout, H. F. (2021). Insect hormones. *Princeton University Press*.
 590 Nozaki, T., Yashiro, T., & Matsuura, K. (2018). Preadaptation for asexual queen
 591 succession: queen tytoparthenogenesis produces neotenic queens in the termite
 592 *Reticulitermes okinawanus*. *Insectes Sociaux*, 65(2), 225–231.
 593 Oster, G. F., & Wilson, E. O. (1978). Caste and ecology in the social insects. *Monographs*
 594 *in Population Biology*, 12, 1–352.
 595 Peng, X., Khan, Z., Liu, X.-M., Deng, S.-L., Fang, Y.-G., Zhang, M., et al. (2023).
 596 Embryonic development of parthenogenetic and sexual eggs in lower termites. *Insects*,
 597 14(7).
 598 Nijhout, H. F., & Wheeler, D. E. (1982). Juvenile hormone and the physiological basis of
 599 insect polymorphisms. *The Quarterly Review of Biology*, 57(2), 109–133.
 600 Rinn, J. L., & Chang, H. Y. (2012). Genome regulation by long noncoding RNAs. *Annual*
 601 *Review of Biochemistry*, 81(1), 145–166.
 602 Takata, M., Inagaki, T., Ishibashi, T., Tasaki, E., & Matsuura, K. (2020). A non-invasive
 603 method for sexing first and second instar larvae of termites using external morphology.
 604 *Insectes Sociaux*, 67(4), 487–493.
 605 Takata, M., Nagai, S., Inagaki, T., Ohkubo, Y., Tasaki, E., & Matsuura, K. (2023).
 606 Heritable effects on caste determination and colony-level sex allocation in termites
 607 under field conditions. *IScience*, 26(3), 106207.
 608 Takata, M., Takahashi, M., Ishibashi, T., Tasaki, E., Rueppell, O., Vargo, E. L., &
 609 Matsuura, K. (2024). *Reticulitermes speratus* genome annotation [Data set]. figshare.
 610 Tamaki, C., Takata, M., & Matsuura, K. (2021). The lose-to-win strategy of the weak:
 611 intraspecific parasitism via egg abduction in a termite. *Biology Letters*, 17(12),
 612 20210540.
 613 Okot-Kotber, B. M., Prestwich, G. D., Strambi, A., & Strambi, C. (1993). Changes in
 614 morphogenetic hormone titers in isolated workers of the termite *Reticulitermes flavipes*
 615 (Kollar). *General and Comparative Endocrinology*, 90(3), 290–295.

- Park, Y. I., & Raina, A. K. (2003). Factors regulating caste differentiation in the Formosan subterranean termite with emphasis on soldier formation. *Sociobiology*, 41, 49–60.
- Park, Y. I., & Raina, A. K. (2004). Juvenile hormone III titers and regulation of soldier caste in *Coptotermes formosanus* (Isoptera: Rhinotermitidae). *Journal of Insect Physiology*, 50(6), 561–566.
- Park, Y. I., & Raina, A. K. (2005). Regulation of juvenile hormone titers by soldiers in the Formosan subterranean termite, *Coptotermes formosanus*. *Journal of Insect Physiology*, 51(4), 385–391.
- Quque, M., Villette, C., Criscuolo, F., Sueur, C., Bertile, F., & Heintz, D. (2021). Eusociality is linked to caste-specific differences in metabolism, immune system, and somatic maintenance-related processes in an ant species. *Cellular and Molecular Life Sciences: CMLS*, 79(1), 29.
- Scharf, M. E., Ratliff, C. R., Hoteling, J. T., Pittendrigh, B. R., & Bennett, G. W. (2003). Caste differentiation responses of two sympatric *Reticulitermes* termite species to juvenile hormone homologs and synthetic juvenoids in two laboratory assays. *Insectes Sociaux*, 50(4), 346–354.
- Scharf, Michael E., Ratliff, C. R., Wu-Scharf, D., Zhou, X., Pittendrigh, B. R., & Bennett, G. W. (2005). Effects of juvenile hormone III on *Reticulitermes flavipes*: changes in hemolymph protein composition and gene expression. *Insect Biochemistry and Molecular Biology*, 35(3), 207–215.
- Trerotola, M., Relli, V., Simeone, P., & Alberti, S. (2015). Epigenetic inheritance and the missing heritability. *Human Genomics*, 9(1), 17.
- Truman, J. W., & Riddiford, L. M. (1999). The origins of insect metamorphosis. *Nature*, 401(6752), 447–452.
- Ulitsky, I., & Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell*, 154(1), 26–46.
- Vargo, E. L., Labadie, P. E., & Matsuura, K. (2012). Asexual queen succession in the subterranean termite *Reticulitermes virginicus*. *Proceedings. Biological Sciences / The Royal Society*, 279(1729), 813–819.
- Verhulst, E. C., van de Zande, L., & Beukeboom, L. W. (2010). Insect sex determination: it all evolves around transformer. *Current Opinion in Genetics & Development*, 20(4), 376–383.
- Volny, V. P., & Gordon, D. M. (2002). Genetic basis for queen–worker dimorphism in a social insect. *Proceedings of the National Academy of Sciences*, 99(9), 6108–6111.

- Wheeler, D. E. (1986). Developmental and physiological determinants of caste in social Hymenoptera: Evolutionary implications. *The American Naturalist*, 128(1), 13–34.
- Wilson, E. O. (1971). The insect societies.
- Zafar, J., Huang, J., Xu, X., & Jin, F. (2023). Recent advances and future potential of long non-coding RNAs in insects. *International Journal of Molecular Sciences*, 24(3), 2605.
- Zeng, Y. A., Rahnema, M., Wang, S., Sosu-Sedzorme, W., & Verheyen, E. M. (2007). *Drosophila* Nemo antagonizes BMP signaling by phosphorylation of Mad and inhibition of its nuclear accumulation. *Development (Cambridge, England)*, 134(11), 2061–2071.
- Zhang, B., Zhang, C., Zhang, J., Lu, S., Zhao, H., Jiang, Y., & Ma, W. (2024). Regulatory roles of long non-coding RNAs in short-term heat stress in adult worker bees. *BMC Genomics*, 25(1), 506.
- Zhou, X., Tarver, M. R., & Scharf, M. E. (2007). Hexamerin-based regulation of juvenile hormone-dependent gene expression underlies phenotypic plasticity in a social insect. *Development (Cambridge, England)*, 134(3), 601–610.

SUPPORTING INFORMATION

Supplementary Materials. Supplementary discussion and experimental procedures.

Table S1: Embryonic developmental stages and nucleic acid concentration. To determine the developmental stage of embryos used in the experiment, the following procedure was performed. Embryonic development was classified into five stages based on external morphology: I: Stage where the embryo is not yet visible (approximately 1–11 days old), II: Stage where the germ band becomes visible (approximately 12–15 days old), III: Stage where katatrepsis is completed (approximately 16–18 days old), IV: Stage of dorsal closure (approximately 19–21 days old), V: Stage where legs and antennae can be distinguished (approximately 22–35 days old). These classifications were made with reference to Matsuura and Kobayashi (2007). For these five stages, nucleic acids were extracted, and DNA and RNA concentrations were measured using a spectrophotometer. Embryos at Stage III, where sufficient nucleic acid yield could be obtained, were used for the experiment.

Table S2: Termite colony information.

Table S3: Sex identification of eggs. The number of male and female individuals identified using the sex-specific genetic marker is shown for each colony.

Table S4: RNA-seq sample statistics. PFE: parthenogenetically produced female embryo. SFE: sexually produced female embryo, SME: sexually produced male embryo. Each group includes five colony replicates, with each sample consisting of pooled multiple eggs. The number of obtained paired reads and mapping results are shown.

Table S5: Statistical results of DEGs among sexually produced male embryo (SME), sexually produced female embryo (SFE) and parthenogenetically produced female embryo (PFE). The category "PFE-high" refers to genes that showed significantly higher expression levels in PFE compared to other embryo types. Similarly, "PFE-low" refers to genes that exhibited significantly lower expression levels in PFE than in other embryo types. "SME-high" indicates genes that were significantly upregulated in SME compared to PFE and SFE, while "SME-low" denotes genes with significantly lower expression levels in SME. Finally, "SFE-high" refers to genes with significantly higher expression levels in SFE than in PFE and SME. The columns include the false discovery rates (FDR) of the ANOVA and pairwise comparisons between different conditions: SME-PFE (sexually produced male embryo vs. parthenogenetically produced female embryo), SFE-PFE (sexually produced female embryo vs. parthenogenetically produced female embryo), and SME-SFE (sexually produced male embryo vs. sexually produced female embryo). Values where $FDR < 0.05$ are indicated in bold.

Table S6: Summary results of blast homology searches of DEGs. The category "PFE-high" refers to genes that showed significantly higher expression levels in PFE compared to other embryo types. Similarly, "PFE-low" refers to genes that exhibited significantly lower expression levels in PFE than in other embryo types. "SME-high" indicates genes that were significantly upregulated in SME compared to PFE and SFE, while "SME-low" denotes genes with significantly lower expression levels in SME. Finally, "SFE-high" refers to genes with significantly higher expression levels in SFE than in PFE and SME. The "Method" column indicates whether the search was performed using blastx or blastp, and genes with blank entries in this column did not yield any hits in the BLAST search. The columns following the "Gene Name" column show information on the top BLAST search hit for each gene.

Table S7: Gene function of DEGs. The category "PFE-high" refers to genes that showed significantly higher expression levels in PFE compared to other embryo types. Similarly, "PFE-low" refers to genes that exhibited significantly lower expression levels in PFE than in other embryo types. "SME-high" indicates genes that were significantly upregulated in

SME compared to PFE and SFE, while "SME-low" denotes genes with significantly lower expression levels in SME. Finally, "SFE-high" refers to genes with significantly higher expression levels in SFE than in PFE and SME. Rows left blank indicate genes that did not have any hits in the BLAST search.

Table S8: Genes with splicing isoforms detected among parthenogenetically and sexually produced embryos. A checkmark (y) indicates significant splicing variation in the corresponding pairwise comparison (PFE_SME, SFE_PFE, SFE_SME); "n" indicates no significant difference. Statistical significance was determined using a Bonferroni-corrected false discovery rate ($FDR < 0.05$).

Table S9: Sample statistics for WGS of parents and RNA-seq of embryos. The number of paired reads and the mapping results are shown for 20 WGS samples, consisting of five replicates each of fathers and mothers from the Kyoto colony and fathers and mothers from the Tsukuba colony, as well as for 10 RNA-seq samples, consisting of five replicates each of embryos with a Kyoto colony father and Tsukuba colony mother, and embryos with a Tsukuba colony father and Kyoto colony mother.

Table S10: Genes with maternal biased expression. Crosses were performed between a Kyoto-father and a Tsukuba-mother, and between a Tsukuba-father and a Kyoto-mother. The "mean of maternal expression" represents the average expression of maternal alleles in the offspring from both crosses. Similarly, the "mean of paternal expression" represents the average expression of paternal alleles. All genes showing maternal-biased expression were mitochondrial-related.

Table S11: Statistical results of colony biased genes. "KY-high" refers to genes that exhibited higher expression in the Kyoto allele, while "TS-high" indicates genes with higher expression in the Tsukuba allele. The FDR of paKY-maTS represents the false discovery rates (FDR) comparing the expression levels of the paternal Kyoto allele and the maternal Tsukuba allele in the egg samples obtained from the cross between Kyoto father and Tsukuba mother. Similarly, the FDR of paTS-maKY compares the expression levels of the paternal Tsukuba allele and the maternal Kyoto allele in the egg samples obtained from the cross between Tsukuba father and Kyoto mother.

Table S12: Summary results of blast homology searches of colony biased genes. "KY-high" refers to genes that exhibited higher expression in the Kyoto allele, while "TS-high" indicates genes with higher expression in the Tsukuba allele. The "Method" column indicates whether the search was performed using blastx or blastp, and genes with blank entries in this column did not yield any hits in the BLAST search. The columns following the "Gene Name" column show information on the top BLAST search hit for each gene.

Table S13: Gene function of colony biased genes. "KY-high" refers to genes that exhibited higher expression in the Kyoto allele, while "TS-high" indicates genes with higher expression in the Tsukuba allele. Rows left blank indicate genes that did not have any hits in the BLAST search.

FIGURE LEGENDS

FIGURE 1. Experimental overview and rationale for three-way comparison. (a) Developmental fate of termite embryos under identical rearing conditions. Parthenogenetically produced female embryos (PFE) are predisposed to differentiate into nymphs when reared by workers, while both sexually produced female embryos (SFE) and sexually produced male embryos (SME) typically develop into workers. This classification highlights differences in reproductive origin, sex, and caste fate tendency. (b) Schematic diagram of the experimental procedure. Parthenogenetically produced and sexually produced eggs were obtained from female–female and male–female pairs, respectively, and were reared by workers until the stage immediately after katatrepsis. Total RNA was extracted, and sexually produced embryos were sexed using a sex-specific genetic marker. RNA-seq analysis was conducted on three groups—PFE, SFE, and SME—to identify genes differentially expressed with respect to reproductive mode, sex, and caste fate.

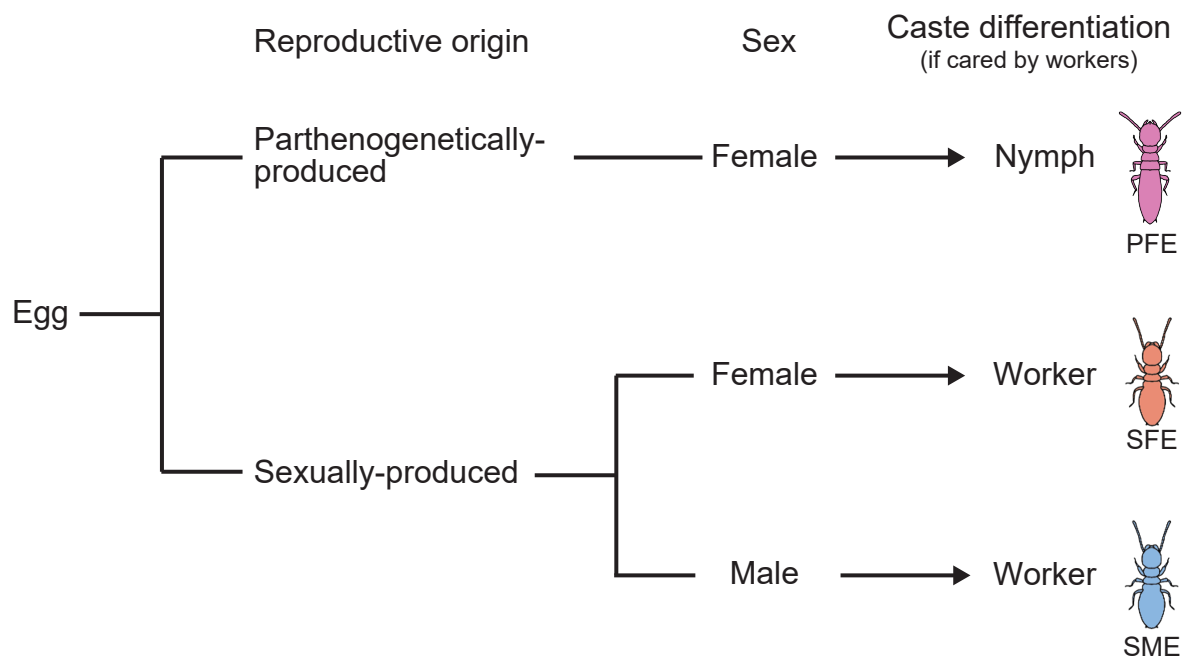
FIGURE 2. Genes showing specific expression in parthenogenetically produced female embryos. (a) Five representative genes showing high expression levels in PFE. The remaining genes are listed in Table S5. Error bars denote standard errors. Significant differences at the 5% level are marked by different lowercase letters (Tukey's HSD test, False Discovery Rate (FDR) < 0.05). The expression levels in PFE are significantly higher than in SFE and SME, with no difference in expression levels between SFE and SME. (b) Five representative genes showing high expression levels in PFE. The remaining genes are listed in Table S5. Error bars denote standard errors. Significant differences at the 5% level are marked by different lowercase letters (Tukey's HSD test, False Discovery Rate (FDR) < 0.05). The expression levels in PFE are significantly lower than in SFE and SME, with no difference in expression levels between SFE and SME.

FIGURE 3. Genes showing different patterns of expression. Error bars denote standard errors. Significant differences at the 5% level are marked by different lowercase letters (Tukey's HSD test, False Discovery Rate (FDR) < 0.05). (a) Five representative genes showing high expression levels in SME. The remaining genes are listed in Table S5. (b) Five representative genes showing high expression levels in SME. The remaining genes are listed in Table S5. (c) Five representative genes showing high expression levels in SFE. The remaining genes are listed in Table S5.

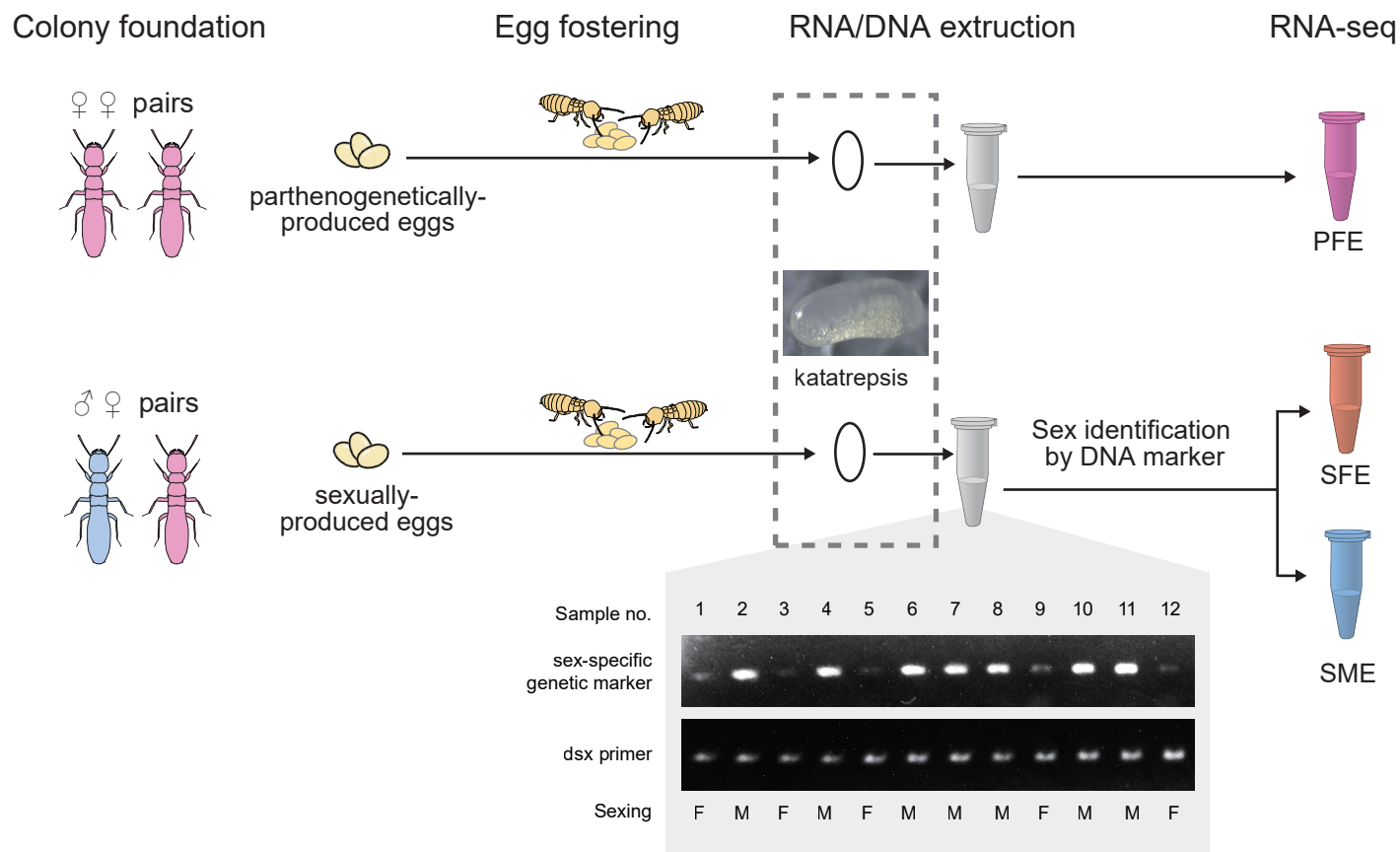
FIGURE 4. Expression of paternal and maternal alleles. (a) A schematic diagram of the experiment comparing expression of paternal and maternal alleles. (b) The expression levels of paternal and maternal alleles for each gene. For the 847 genes that could be analyzed, the average values from two types of crosses (Kyoto father × Tsukuba mother, Tsukuba father × Kyoto mother) are plotted. There are clusters of plots showing a 1:1 expression ratio and clusters with maternal-biased expression, the latter of which are all mitochondrial-related genes. (c) Allele expression of genes that showed differences in expression among PFE, SFE, and SME. The expression levels of KY-pa (allele from Kyoto father) and TS-mo (allele from Tsukuba mother), as well as TS-fa (allele from Tsukuba father) and KY-mo (allele from Kyoto mother), were compared (n.s.: not significant, ***: False Discovery Rate (FDR) < 0.001).

FIGURE 5. Genes showing colony-specific expression levels. (a) The expression levels of Kyoto alleles and Tsukuba alleles for each gene. For 847 genes, the average expression values from two types of crosses (Kyoto father × Tsukuba mother, Tsukuba father × Kyoto mother) were plotted. (b) A subset of the plots shown in (a). Genes marked with red dots had significant differences in expression levels between Kyoto-derived alleles and Tsukuba-derived alleles, indicating colony-specific expression. (c) The top five genes showing higher expression levels in Kyoto-derived alleles. The remaining genes are listed in Table S11. The expression levels of KY-pa (allele from Kyoto father) and TS-mo (allele from Tsukuba mother), as well as TS-fa (allele from Tsukuba father) and KY-mo (allele from Kyoto mother), were compared (*: False Discovery Rate (FDR) < 0.05, **: FDR < 0.01, ***: FDR < 0.001). (d) The top five genes showing higher expression levels in Tsukuba-derived alleles. The remaining genes are listed in Table S11. The expression levels of KY-pa (allele from Kyoto father) and TS-mo (allele from Tsukuba mother), as well as TS-fa (allele from Tsukuba father) and KY-mo (allele from Kyoto mother), were compared (*: False Discovery Rate (FDR) < 0.05, **: FDR < 0.01, ***: FDR < 0.001).

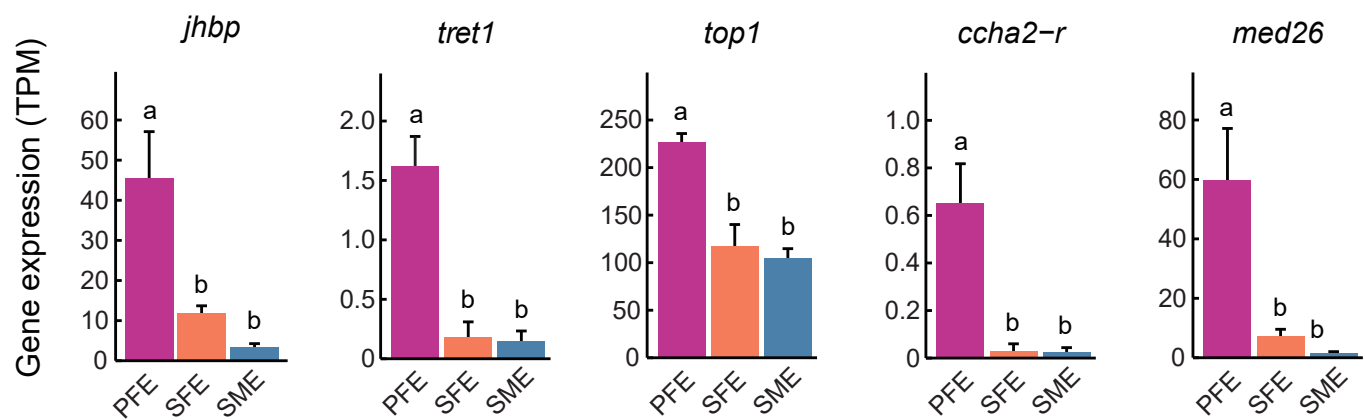
(a)



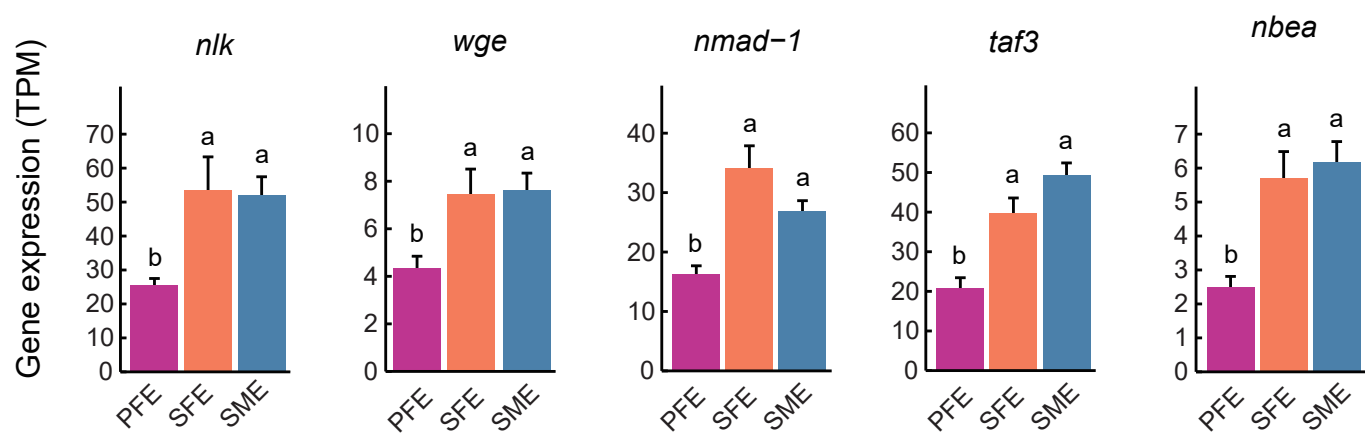
(b)



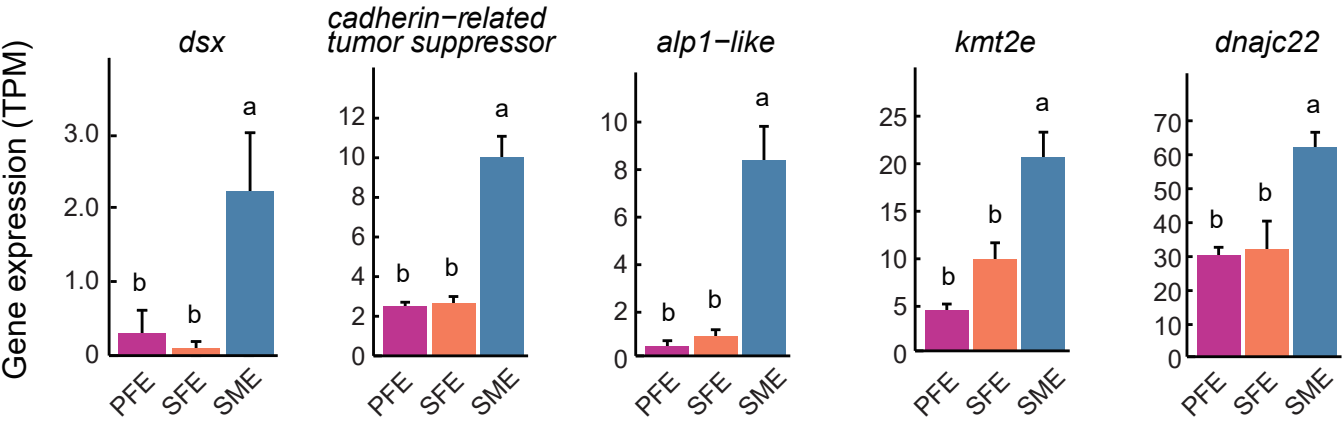
(a) PFE > SFE, SME



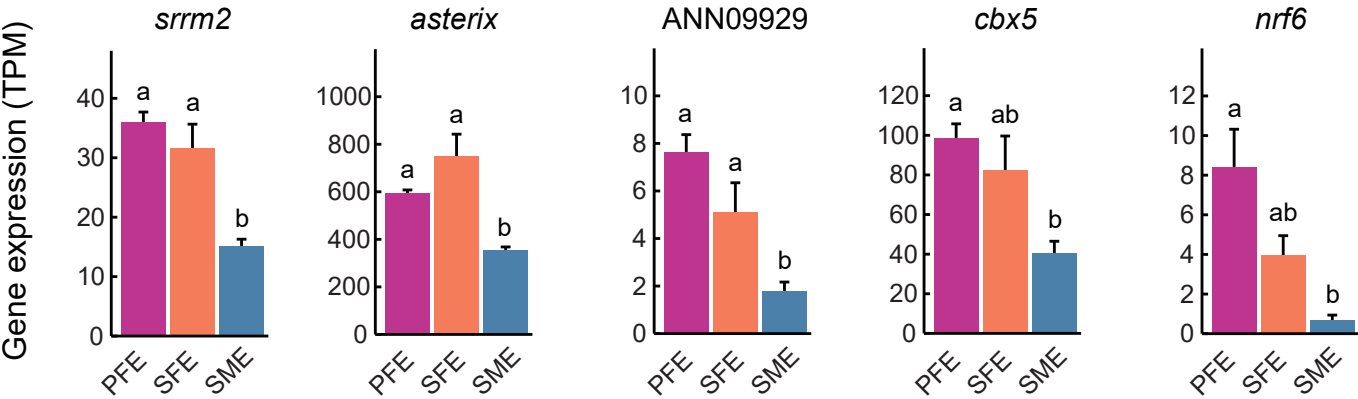
(b) PFE < SFE, SME



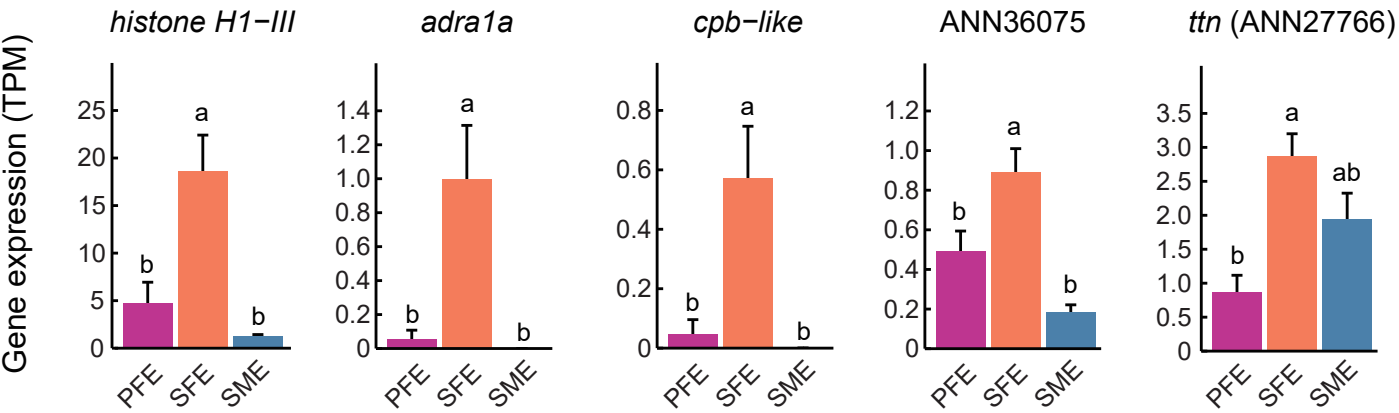
(a) SME > PFE, SFE

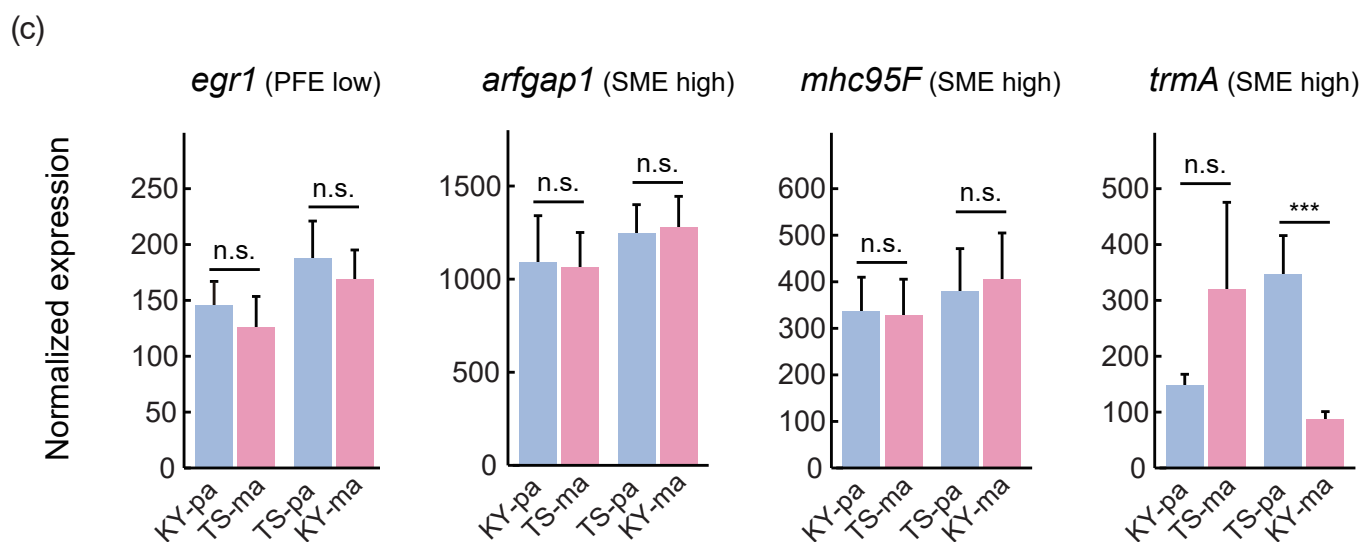
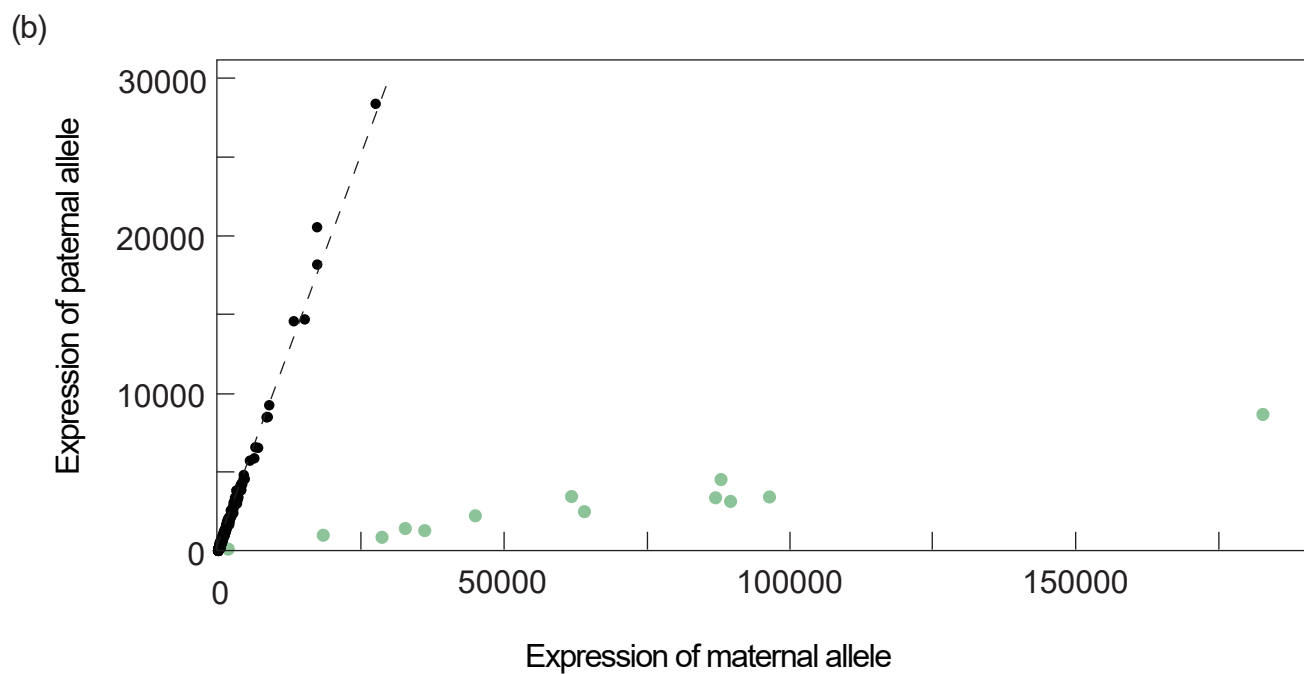
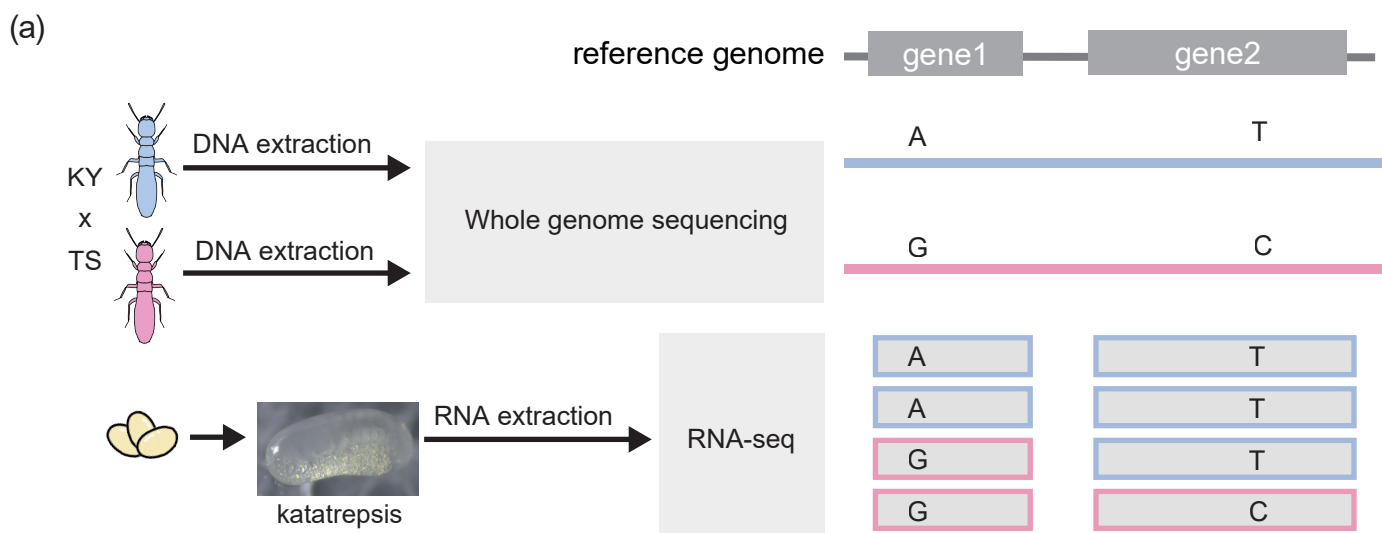


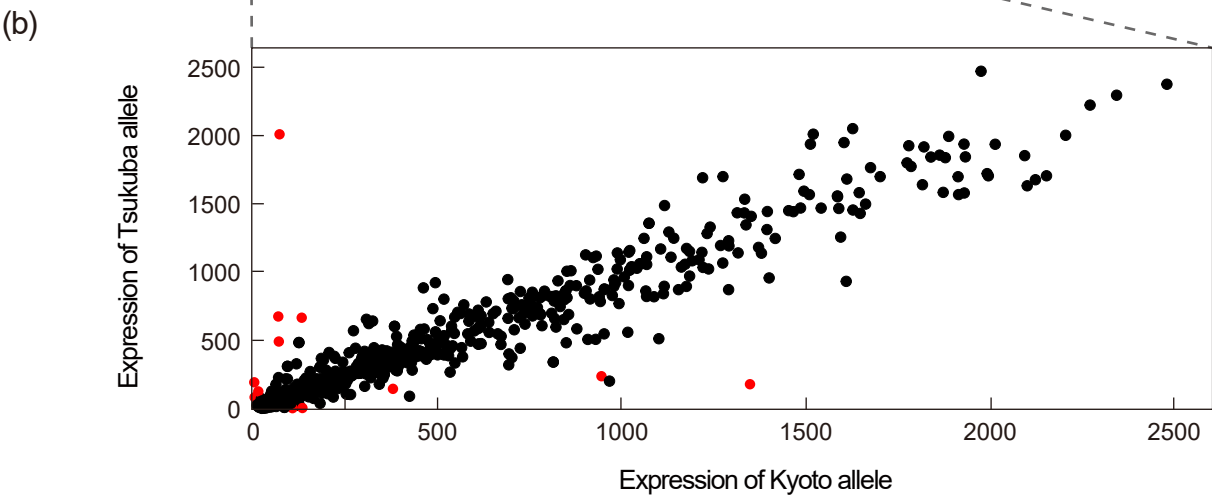
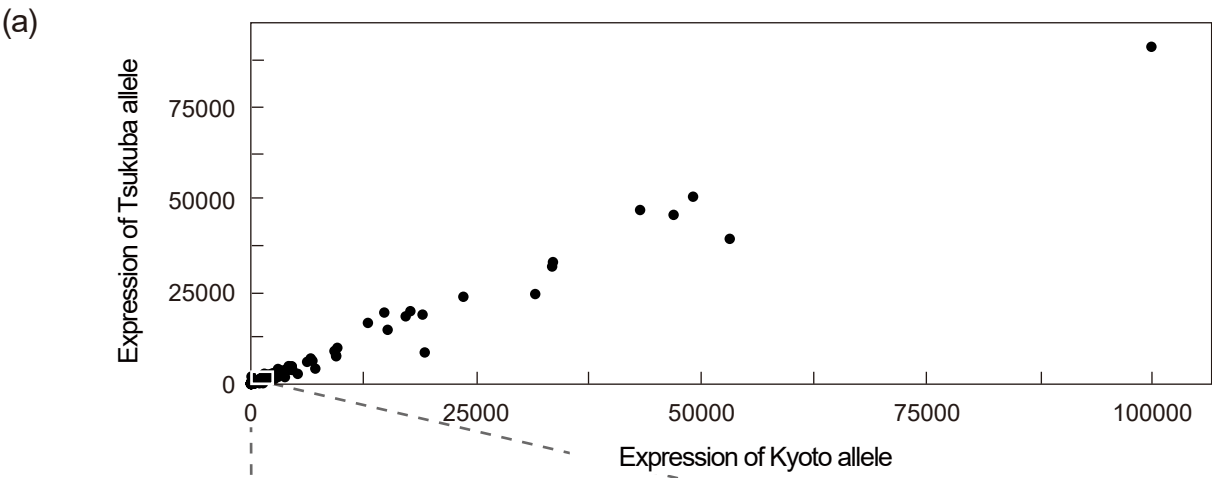
(b) SME < PFE, SFE



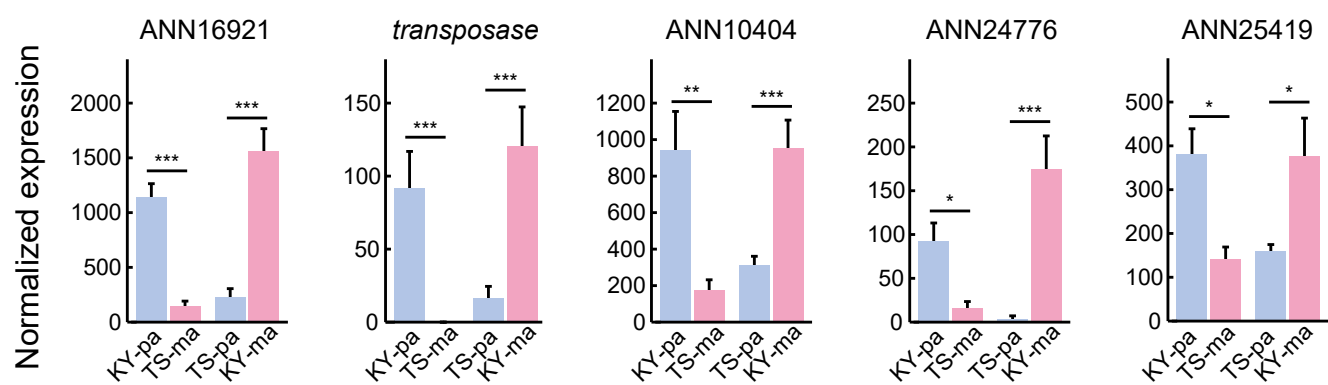
(c) SFE > PFE, SME



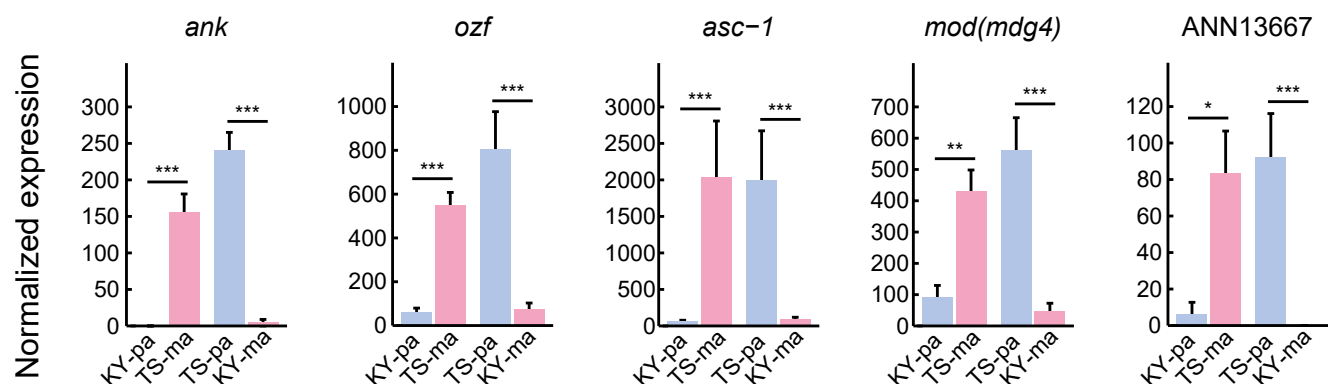




(c) KY > TS



(d) TS > KY



Differences in gene expression patterns between parthenogenetically and sexually produced offspring during early development of *Reticulitermes speratus*

EXPERIMENTAL PROCEDURES

Mitochondrial Genome Assembly

Mitochondrial Genome Assembly with GetOrganelle (v1.7.7.0)

```
mkdir Output_Dir

get_organelle_config.py -a animal_mt

get_organelle_from_reads.py \
-1 /PATH/to/fastq_File_of_Whole_Genome/Rs_male_short_1.fastq.gz \
-2 /PATH/to/fastq_File_of_Whole_Genome/Rs_male_short_2.fastq.gz \
-s /PATH/to/fasta_File_of_Related_Species/Rflavipes.fasta \
-o /PATH/to/Output_Dir \
-R 10 \
-k 21,45,65,85,105 \
-F animal_mt
```

Annotation using Mitos (<http://mitos.bioinf.uni-leipzig.de/index.py>)

Gene Expression Analysis

Quality Control and Read Filtering

FastQC (v0.12.1) for Raw Data Quality Check

```
mkdir FastQC_Output_Dir

fastqc \
-t 8 \
-o /PATH/to/FastQC_Output_Dir \
--nogroup \
/PATH/to/Raw_Data_Dir/${SAMPLE}.fq.gz
```

Filtering Reads with fastp (v0.23.2)

```
mkdir fastp_Output_Dir

fastp \
-i /PATH/to/Raw_Data_Dir/${SAMPLE}_R1.fq.gz \
-I /PATH/to/Raw_Data_Dir/${SAMPLE}_R2.fq.gz \
```

```
-o /PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R1.fq.gz \  
-O /PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R2.fq.gz \  
-w 5 \  
-l 25
```

FastQC for Post-fastp Quality Check

```
mkdir FastQC_Output_Dir2  
  
fastqc \  
-t 8 \  
-o /PATH/to/FastQC_Output_Dir2 \  
--nogroup \  
/PATH/to/fastp_Output_Dir/fastp_${SAMPLE}.fq.gz
```

Mapping and Gene Expression Quantification

Create Reference Index for Mapping (RSEM v1.3.3)

```
rsem-prepare-reference \  
--star --star-path /PATH/to/STAR \  
-p 4 \  
--gtf /PATH/to/GTF_File/Rsperatus.gtf \  
/PATH/to/FASTA_file/Rsperatus.fasta \  
RSEMindex
```

Mapping with STAR (v2.7.10a)

```
mkdir STAR_Output_Dir  
  
STAR --runThreadN 12 \  
--genomeDir /PATH/to/RSEMindex_Dir \  
--readFilesIn /PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R1.fq.gz /PATH/to/fastp_${SAMPLE}_R2.fq.gz \  
\  
--quantMode TranscriptomeSAM \  
--outSAMtype BAM SortedByCoordinate \  
--genomeLoad NoSharedMemory \  
--outFilterMultimapNmax 1 \  
--outFileNamePrefix /PATH/to/STAR_Output_Dir/${SAMPLE}_map
```

Gene Expression Quantification with RSEM (v1.3.3)

```
mkdir RSEM_Output_Dir  
  
rsem-calculate-expression \  
--paired-end --alignments \  
--estimate-rspd \  
--strandedness none \  
--no-bam-output \  
-p 6 \  
/PATH/to/STAR_Output_Dir/${SAMPLE}_mapAligned.toTranscriptome.out.bam \  
/PATH/to/RSEMindex_Dir/RSEMindex \  
/PATH/to/RSEM_Output_Dir/${SAMPLE}_rsem
```

Differential Expression Analysis

ANOVA and Differential Gene Expression Analysis with edgeR (v3.40.2)

```
# Load edgeR library
library(edgeR)

# Set working directory
setwd("/PATH/to/Your_Dir/")

# Read RSEM data
FM1 <- read.table("FM1_rsem.genes.results", header=T)
FM2 <- read.table("FM2_rsem.genes.results", header=T)
FM3 <- read.table("FM3_rsem.genes.results", header=T)
FM4 <- read.table("FM4_rsem.genes.results", header=T)
FM5 <- read.table("FM5_rsem.genes.results", header=T)
FF1 <- read.table("FF1_rsem.genes.results", header=T)
FF2 <- read.table("FF2_rsem.genes.results", header=T)
FF3 <- read.table("FF3_rsem.genes.results", header=T)
FF4 <- read.table("FF4_rsem.genes.results", header=T)
FF5 <- read.table("FF5_rsem.genes.results", header=T)
PF1 <- read.table("PF1_rsem.genes.results", header=T)
PF2 <- read.table("PF2_rsem.genes.results", header=T)
PF3 <- read.table("PF3_rsem.genes.results", header=T)
PF4 <- read.table("PF4_rsem.genes.results", header=T)
PF5 <- read.table("PF5_rsem.genes.results", header=T)

# Extract expected counts and combine into a single data frame
expression <- data.frame(
  row.names = FM1$gene_id,
  FM1 = FM1$expected_count,
  FM2 = FM2$expected_count,
  FM3 = FM3$expected_count,
  FM4 = FM4$expected_count,
  FM5 = FM5$expected_count,
  FF1 = FF1$expected_count,
  FF2 = FF2$expected_count,
  FF3 = FF3$expected_count,
  FF4 = FF4$expected_count,
  FF5 = FF5$expected_count,
  PF1 = PF1$expected_count,
  PF2 = PF2$expected_count,
  PF3 = PF3$expected_count,
  PF4 = PF4$expected_count,
  PF5 = PF5$expected_count
)

# Define groups
group <- factor(c(rep("FM", 5), rep("FF", 5), rep("PF", 5)))

# Create DGEList object
d <- DGEList(counts = expression, group = group)

# Create design matrix
design <- model.matrix(~ group)

# Calculate normalization factors
d <- calcNormFactors(d)

# Estimate dispersions
d <- estimateGLMCommonDisp(d, design)
d <- estimateGLMTrendedDisp(d, design)
d <- estimateGLMTagwiseDisp(d, design)
```

```

# Fit the GLM
fit <- glmFit(d, design)

# Perform ANOVA
lrt <- glmLRT(fit, coef = c(2, 3))
table <- as.data.frame(topTags(lrt, n = nrow(expression)))

# Extract DEGs with FDR < 0.05
table_0.05 <- table[table$FDR < 0.05,]

# Output results
write.table(table_0.05, file = "ANOVA_FDR_0.05.txt", col.names = T, row.names = T, sep = "\t")

```

Three-group Comparison with edgeR (v3.40.2)

```

# Load edgeR library
library(edgeR)

# Set working directory
setwd("/PATH/to/Your_Dir/")

# Read ANOVA results
ANOVA <- read.table("ANOVA_FDR_0.05.txt", header=T)

# Read RSEM data
FM1 <- read.table("FM1_rsem.genes.results", header=T)
FM2 <- read.table("FM2_rsem.genes.results", header=T)
FM3 <- read.table("FM3_rsem.genes.results", header=T)
FM4 <- read.table("FM4_rsem.genes.results", header=T)
FM5 <- read.table("FM5_rsem.genes.results", header=T)
FF1 <- read.table("FF1_rsem.genes.results", header=T)
FF2 <- read.table("FF2_rsem.genes.results", header=T)
FF3 <- read.table("FF3_rsem.genes.results", header=T)
FF4 <- read.table("FF4_rsem.genes.results", header=T)
FF5 <- read.table("FF5_rsem.genes.results", header=T)
PF1 <- read.table("PF1_rsem.genes.results", header=T)
PF2 <- read.table("PF2_rsem.genes.results", header=T)
PF3 <- read.table("PF3_rsem.genes.results", header=T)
PF4 <- read.table("PF4_rsem.genes.results", header=T)
PF5 <- read.table("PF5_rsem.genes.results", header=T)

# Extract TPM values and combine into a single data frame
expression <- data.frame(
  row.names = FM1$gene_id,
  FM1 = FM1$TPM,
  FM2 = FM2$TPM,
  FM3 = FM3$TPM,
  FM4 = FM4$TPM,
  FM5 = FM5$TPM,
  FF1 = FF1$TPM,
  FF2 = FF2$TPM,
  FF3 = FF3$TPM,
  FF4 = FF4$TPM,
  FF5 = FF5$TPM,
  PF1 = PF1$TPM,
  PF2 = PF2$TPM,
  PF3 = PF3$TPM,
  PF4 = PF4$TPM,
  PF5 = PF5$TPM
)

```

```

expression_ANOVA <- subset(expression, rownames(expression) %in% rownames(ANOVA))

# Define function to perform Tukey HSD and summarize results
tukey_summary <- function(expression_ANOVA) {
  results <- list()
  for (i in 1:nrow(expression_ANOVA)) {
    gene <- data.frame(
      Value = unlist(expression_ANOVA[i, ]),
      Group = factor(rep(c("FM", "FF", "PF"), each = 5), levels = c("FM", "FF", "PF"))
    )
    aov_result <- aov(Value ~ Group, data = gene)
    tukey_result <- TukeyHSD(aov_result)
    results[[i]] <- data.frame(gene = rownames(expression_ANOVA)[i], tukey_result$Group)
  }
  final_result <- do.call(rbind, results)
  return(final_result)
}

# Perform Tukey HSD and summarize results
tukey_results <- tukey_summary(expression_ANOVA)

```

Splicing variant analysis (rMATS v4.1.2)

```

rmats.py \
--b1 group1_bam.txt \
--b2 group2_bam.txt \
--gtf /PATH/to/GTF_File/Rsperatus.gtf \
--od rmats_output \
--tmp tmp_dir \
--readLength 150 \
--variable-read-length \
--nthread 4 \
--libType fr-firststrand \
--cstat 0.0001

```

Allele Expression Analysis

Mapping

Filtering Reads with fastp (v0.23.2)

```

mkdir fastp_Output_Dir

fastp \
-i /PATH/to/Raw_Data_Dir/${SAMPLE}_R1.fq.gz \
-I /PATH/to/Raw_Data_Dir/${SAMPLE}_R1.fq.gz \
-o /PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R1.fq.gz \
-O /PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R2.fq.gz \
-w 5 \
-l 25

```

Creating Index for Mapping Whole Genome Resequencing Data of Parents with bwa-mem (v2.2.1)


```
bwa-mem2 index /PATH/to/FASTA_file
```

Mapping Whole Genome Resequencing Data of Parents with bwa-mem (v2.2.1)

```
mkdir BWA_Output_Dir

bwa-mem2 \
mem -t 10 \
/PATH/to/FASTA_file \
/PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R1.fq.gz \
/PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R2.fq.gz ¥
/PATH/to/FastQC_Output_Dir2/${SAMPLE}_mapped.sam ¥
> /PATH/to/BWA_Output_Dir/${SAMPLE}.sam
```

Creating Index for Mapping Egg RNA-seq Data with histat2 (v2.2.1)

```
hisat2-build \
-p 6 \
/PATH/to/FASTA_file \
hisat2_genome_index
```

Mapping RNA-seq Data of Eggs with histat2 (v2.2.1)

```
mkdir HISAT2_Output_Dir

hisat2 \
-x /PATH/to/hisat2_genome_index \
-1 /PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R1.fq.gz \
-2 /PATH/to/fastp_Output_Dir/fastp_${SAMPLE}_R2.fq.gz \
-k 3 \
-p 6 \
-S /PATH/to/HISAT2_Output_Dir/${SAMPLE}.sam
```

Converting Mapping Results from SAM Files to BAM Files with samtools (v1.15.1)

```
mkdir BAM_Output_Dir

samtools sort \
-@ 8 \
-O bam \
-o /PATH/to/BAM_Output_Dir/${SAMPLE}.bam \
/PATH/to/BWA_Output_Dir of HISAT2_Output_Dir/${SAMPLE}.sam
```

Filtering BAM Files with bamtools (v2.5.2)

```
mkdir filteredBAM_Output_Dir

bamtools filter \
-isMapped true \
-isPaired true \
-isProperPair true \
-in /PATH/to/BAM_Output_Dir/${SAMPLE}.bam \
-out /PATH/to/filteredBAM_Output_Dir/${SAMPLE}_filtered.bam
```

Indexing BAM Files with samtools (v1.15.1)

```
samtools index \  
/PATH/to/filteredBAM_Output_Dir/${SAMPLE}_filtered.bam
```

Preprocessing for Analysis

Creating Index for SNP Analysis with GATK (v4.2.6.1)

```
samtools faidx /PATH/to/FASTA_file  
  
gatk CreateSequenceDictionary ¥  
-R /PATH/to/FASTA_file ¥  
-O FASTA_file_gatk.dict
```

Removing Duplicate Reads with GATK (v4.2.6.1)

```
mkdir MarkDuplicats_Output_Dir  
  
gatk MarkDuplicates \  
-I /PATH/to/filteredBAM_Output_Dir/${SAMPLE}_filtered.bam \  
-O /PATH/to/MarkDuplicats_Output_Dir/${SAMPLE}_no-dup.bam \  
-CREATE_INDEX true \  
-VALIDATION_STRINGENCY SILENT \  
-M /PATH/to/MarkDuplicats_Output_Dir/${SAMPLE}_out_metrics.txt
```

Indexing BAM Files with samtools (v1.15.1)

```
samtools index  
/PATH/to/MarkDuplicats_Output_Dir/${SAMPLE}_no-dup.bam
```

Grouping Reads with GATK (v4.2.6.1)

```
mkdir ReadGroups_Output_Dir  
  
gatk AddOrReplaceReadGroups \  
I=/PATH/to/MarkDuplicats_Output_Dir/${SAMPLE}_no-dup.bam \  
O=/PATH/to/ReadGroups_Output_Dir/${SAMPLE}_added.bam \  
SO=coordinate \  
RGID=1 \  
RGLB=lib \  
RGPL=ILLUMINA \  
RGPU=hogehoge \  
RGSM=${SAMPLE}
```

Removing N Cigar with GATK (v4.2.6.1)

```
mkdir SplitN_Output_Dir  
  
gatk SplitNCigarReads \  
-R /PATH/to/FASTA_file \  
-I /PATH/to/ReadGroups_Output_Dir/${SAMPLE}_added.bam \  
-O /PATH/to/SplitN_Output_Dir/${SAMPLE}_split.bam
```

Obtaining Variants Information and BQSR

Estimation of Haplotypes with GATK (v4.2.6.1)

```
mkdir GVCF_Output_Dir

gatk HaplotypeCaller \
-R /PATH/to/FASTA_file \
--emit-ref-confidence GVCF \
-I /PATH/to/SplitN_Output_Dir/${SAMPLE}_split.bam \
-O /PATH/to/GVCF_Output_Dir/${SAMPLE}.g.vcf
```

Genotyping with GATK (v4.2.6.1)

```
echo -e "223-193-659193-circular\nScaffold_1\nScaffold_2\nScaffold_3\nScaffold_4\nScaffold_5\nScaffold_6\nScaffold_7\nScaffold_8\nScaffold_9\nScaffold_10\nScaffold_11\nScaffold_12\nScaffold_13\nScaffold_14\nScaffold_15\nScaffold_16\nScaffold_17\nScaffold_18\nScaffold_19\nScaffold_20\nScaffold_21" > intervals.list

gatk GenomicsDBImport ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/GVCF_Output_Dir/SAMPLE1.g.vcf \
-V /PATH/to/GVCF_Output_Dir/SAMPLE2.g.vcf \
...
-V /PATH/to/GVCF_Output_Dir/SAMPLEn.g.vcf \
-L intervals.list \
--genomicsdb-workspace-path gvcfs_db

gatk GenotypeGVCFs ¥
-R /PATH/to/FASTA_file \
-V gendb://gvcfs_db \
-O merged.vcf
```

Extracting SNPs with GATK (v4.2.6.1)

```
gatk SelectVariants ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/merged.vcf \
--select-type-to-include SNP \
-O merged_snps.vcf

gatk VariantFiltration ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/merged_snps.vcf \
-O merged_snps_filtered.vcf \
-filter "QD < 2.0" --filter-name "QD2" \
-filter "QUAL < 30.0" --filter-name "QUAL30" \
-filter "SOR > 4.0" --filter-name "SOR4" \
-filter "FS > 60.0" --filter-name "FS60" \
-filter "MQ < 40.0" --filter-name "MQ40" \
-filter "MQRankSum < -12.5" --filter-name "MQRankSum-12.5" \
-filter "ReadPosRankSum < -8.0" --filter-name "ReadPosRankSum-8"
```

Extracting Indels with GATK (v4.2.6.1)

```
gatk SelectVariants ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/merged.vcf \
--select-type-to-include INDEL \
-O merged_indels.vcf
```

```

gatk VariantFiltration ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/merged_indels.vcf \
-O merged_indels_filtered.vcf \
-filter "QD < 2.0" --filter-name "QD2" \
-filter "QUAL < 30.0" --filter-name "QUAL30" \
-filter "FS > 200.0" --filter-name "FS200" \
-filter "SOR > 10.0" --filter-name "SOR10" \
-filter "ReadPosRankSum < -20.0" --filter-name "ReadPosRankSum-20"

```

BQSR with GATK (v4.2.6.1)

```

mkdir TABLE_Output_Dir

gatk BaseRecalibrator \
-R /PATH/to/FASTA_file \
-I /PATH/to/SplitN_Output_Dir/${SAMPLE}_split.bam \
--known-sites /PATH/to/merged_snps.vcf \
--known-sites /PATH/to/merged_indels.vcf \
-O /PATH/to/TABLE_Output_Dir/${SAMPLE}_recal_data.table

mkdir BQSR_Output_Dir

gatk ApplyBQSR \
-R /PATH/to/FASTA_file \
-I /PATH/to/SplitN_Output_Dir/${SAMPLE}_split.bam \
-bqsr /PATH/to/TABLE_Output_Dir/${SAMPLE}_recal_data.table \
-O /PATH/to/BQSR_Output_Dir/${SAMPLE}_bqsr.bam

```

SNP Calling

Estimation of Haplotypes with GATK (v4.2.6.1)

```

mkdir GVCF_Output_Dir2

gatk HaplotypeCaller \
-R /PATH/to/FASTA_file \
--emit-ref-confidence GVCF \
-I /PATH/to/BQSR_Output_Dir/${SAMPLE}_bqsr.bam \
-O /PATH/to/GVCF_Output_Dir2/${SAMPLE}.g.vcf

```

Genotyping with GATK (v4.2.6.1)

```

echo -e "223-193-659193-circular\nScaffold_1\nScaffold_2\nScaffold_3\nScaffold_4\nScaffold_5\nScaffold_6\nScaffold_7\nScaffold_8\nScaffold_9\nScaffold_10\nScaffold_11\nScaffold_12\nScaffold_13\nScaffold_14\nScaffold_15\nScaffold_16\nScaffold_17\nScaffold_18\nScaffold_19\nScaffold_20\nScaffold_21" > intervals.list

gatk GenomicsDBImport ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/GVCF_Output_Dir2/SAMPLE1.g.vcf \
-V /PATH/to/GVCF_Output_Dir2/SAMPLE2.g.vcf \
...
-V /PATH/to/GVCF_Output_Dir2/SAMPLEn.g.vcf \
-L intervals.list \
--genomicsdb-workspace-path gvcfs_db2

```

```
gatk GenotypeGVCFs ¥
-R /PATH/to/FASTA_file \
-V gendb://gvcfs_db2 \
-O merged2.vcf
```

Extracting SNP Information with GATK (v4.2.6.1)

```
gatk SelectVariants ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/merged2.vcf \
--select-type-to-include SNP \
-O merged_snps2.vcf

gatk VariantFiltration ¥
-R /PATH/to/FASTA_file \
-V /PATH/to/merged_snps2.vcf \
-O merged_snps_filtered2.vcf \
-filter "QD < 2.0" --filter-name "QD2" \
-filter "QUAL < 30.0" --filter-name "QUAL30" \
-filter "SOR > 4.0" --filter-name "SOR4" \
-filter "FS > 60.0" --filter-name "FS60" \
-filter "MQ < 40.0" --filter-name "MQ40" \
-filter "MQRankSum < -12.5" --filter-name "MQRankSum-12.5" \
-filter "ReadPosRankSum < -8.0" --filter-name "ReadPosRankSum-8"
```

Obtaining Allele Expression Levels in Eggs

Counting Expression Levels for Each Allele with GATK (v4.2.6.1)

```
gatk ASEReadCounter \
-R /PATH/to/FASTA_file \
-I /PATH/to/ReadGroups_Output_Dir/${SAMPLE}_added.bam \ #卵の bam ファイルのみ
-V merged_snps_filtered2.vcf \
-O ${SAMPLE}.table \
-DF NotDuplicateReadFilter
```

Editing the File

Merging Allele Expression Levels of Eggs for Each Sample

```
# Extract and merge parental GT and egg allele expression levels
awk -v OFS='\t' '
BEGIN {
    file_count = 10
}
FNR==1 {next} # Skip header row
{
    key = $1 OFS $2
    refCount = $6
    altCount = $7

    if (key in counts) {
        counts[key] = counts[key] OFS refCount OFS altCount
        file_keys[key]++
    } else {
        counts[key] = refCount OFS altCount
        file_keys[key] = 1
    }
}
```

```

    }
}
END {
    for (key in file_keys) {
        if (file_keys[key] == file_count) {
            split(counts[key], fields, OFS)
            if (length(fields) > 20) {
                print key, fields[1], fields[2], fields[3], fields[4], fields[5], fields[6], fields[7], fields[8], fields[9],
fields[10], fields[11], fields[12], fields[13], fields[14], fields[15], fields[16], fields[17], fields[18], fields[19],
fields[20]
            } else {
                # Fill missing fields with 0
                for (i=length(fields)+1; i<=20; i++) {
                    fields[i] = 0
                }
                print key, fields[1], fields[2], fields[3], fields[4], fields[5], fields[6], fields[7], fields[8], fields[9],
fields[10], fields[11], fields[12], fields[13], fields[14], fields[15], fields[16], fields[17], fields[18], fields[19],
fields[20]
            }
        }
    }
}
}
' sample1.table sample2.table ... | sort -k1,1 -k2,2n > merged_allele_expression.txt

```

Extracting Parental GT

```

python3 extract_genotype_fields.py merged_snps_filtered2.vcf parents_GT.txt

- - - - - extract_genotype_fields.py - - - - -
import sys
import pandas as pd

def extract_genotype_fields(vcf_file, output_file):
    # Read the VCF file and find the header line
    with open(vcf_file, 'r') as f:
        lines = f.readlines()

    # Find the header line starting with '#CHROM'
    header_line = None
    for i, line in enumerate(lines):
        if line.startswith('#CHROM'):
            header_line = i
            break

    if header_line is None:
        print("Error: Header line not found in the VCF file.")
        sys.exit(1)

    # Load the VCF file into a DataFrame, starting from the header line
    vcf_df = pd.read_csv(vcf_file, sep='\t', skiprows=header_line)

    # Specify the columns to extract
    columns_to_extract = ['#CHROM', 'POS', 'ABf1', 'ABf2', 'ABf3', 'ABf4', 'ABf5',
                          'ABm1', 'ABm2', 'ABm3', 'ABm4', 'ABm5',
                          'BAf1', 'BAf2', 'BAf3', 'BAf4', 'BAf5',
                          'BAm1', 'BAm2', 'BAm3', 'BAm4', 'BAm5']

    # Ensure that the columns exist in the VCF dataframe
    if not all(col in vcf_df.columns for col in columns_to_extract):
        print("Error: One or more columns not found in the VCF file.")
        sys.exit(1)

```

```

# Extract the specified columns
extracted_df = vcf_df[columns_to_extract].copy()

# Extract the GT (genotype) part from each of the sample columns
for col in columns_to_extract[2:]:
    extracted_df[col] = extracted_df[col].apply(lambda x: x.split(':')[0])

# Save the extracted data to a new TXT file
extracted_df.to_csv(output_file, sep='\t', index=False, header=True)

if __name__ == '__main__':
    if len(sys.argv) != 3:
        print("Usage: python3 extract_genotype_fields.py <vcf_file> <output_file>")
        sys.exit(1)

    vcf_file = sys.argv[1]
    output_file = sys.argv[2]

    extract_genotype_fields(vcf_file, output_file)
- . . . . .

```

Filtering Parental GT

```

python3 filter_homozygous_diff.py parents_GT.txt filtered_parents_GT.txt

- . . . . . filter_homozygous_diff.py - . . . . .

import sys
import pandas as pd

def filter_homozygous_diff(input_file, output_file):
    # Load the TXT file into a DataFrame
    df = pd.read_csv(input_file, sep='\t')

    # Define the valid homozygous genotype pairs
    valid_pairs = {("0/0", "1/1"), ("0/0", "1|1"), ("0|0", "1/1"), ("0|0", "1|1"),
                   ("1/1", "0/0"), ("1/1", "0|0"), ("1|1", "0/0"), ("1|1", "0|0")}

    # Filter rows based on the specified conditions
    def is_valid_row(row):
        pairs = [("ABf1", "ABm1"), ("ABf2", "ABm2"), ("ABf3", "ABm3"), ("ABf4", "ABm4"), ("ABf5", "ABm5"),
                 ("BAf1", "BAm1"), ("BAf2", "BAm2"), ("BAf3", "BAm3"), ("BAf4", "BAm4"), ("BAf5", "BAm5")]
        for f, m in pairs:
            if (row[f], row[m]) not in valid_pairs:
                return False
        return True

    filtered_df = df[df.apply(is_valid_row, axis=1)]

    # Save the filtered data to a new TXT file
    filtered_df.to_csv(output_file, sep='\t', index=False, header=True)

if __name__ == '__main__':
    if len(sys.argv) != 3:
        print("Usage: python3 filter_homozygous_diff.py <input_file> <output_file>")
        sys.exit(1)

    input_file = sys.argv[1]
    output_file = sys.argv[2]

    filter_homozygous_diff(input_file, output_file)

```

```
.....
```

Merging Parental GT and Egg Allele Expression Levels

```
python3 merge_files.py filtered_parents_GT.txt merged_allele_expression.txt GT_and_allele_expression.txt

..... merge_files.py- .....
import sys
import pandas as pd

def merge_files(filtered_gt_file, merged_output_file, output_file):
    # Load the filtered_GT.txt file
    filtered_df = pd.read_csv(filtered_gt_file, sep='\t')
    filtered_df.rename(columns={'#CHROM': 'CHROM'}, inplace=True)

    # Load the merged_output2.txt file
    merged_df = pd.read_csv(merged_output_file, sep='\t', header=None)
    merged_df.columns = ['CHROM', 'POS', 'ABe1_ref', 'ABe1_alt', 'ABe2_ref', 'ABe2_alt', 'ABe3_ref', 'ABe3_alt',
                        'ABe4_ref', 'ABe4_alt', 'ABe5_ref', 'ABe5_alt', 'BAe1_ref', 'BAe1_alt', 'BAe2_ref',
                        'BAe2_alt', 'BAe3_ref', 'BAe3_alt', 'BAe4_ref', 'BAe4_alt', 'BAe5_ref', 'BAe5_alt']

    # Merge the two dataframes on CHROM and POS
    merged = pd.merge(filtered_df, merged_df, how='inner', on=['CHROM', 'POS'])

    # Save the merged data to a new TXT file
    merged.to_csv(output_file, sep='\t', index=False, header=True)
    print("マージが完了しました。")

if __name__ == '__main__':
    if len(sys.argv) != 4:
        print("Usage: python3 merge_files.py <filtered_GT_file> <merged_output_file> <output_file>")
        sys.exit(1)

    filtered_gt_file = sys.argv[1]
    merged_output_file = sys.argv[2]
    output_file = sys.argv[3]

    merge_files(filtered_gt_file, merged_output_file, output_file)
.....
```

Reordering Columns

```
python3 process_gt_file.py GT_and_allele_expression.txt sorted_GT_and_allele_expression.txt

..... process_gt_file.py- .....
import sys
import pandas as pd

def process_gt_file(input_file, output_file):
    # Load the merged_GT.txt file
    df = pd.read_csv(input_file, sep='\t')

    # Define the valid homozygous genotype pairs
    ref_alt_pairs = {("0/0", "1/1"), ("0/0", "1|1"), ("0|0", "1/1"), ("0|0", "1|1")}
    alt_ref_pairs = {("1/1", "0/0"), ("1|1", "0|0"), ("1/1", "0|0"), ("1|1", "0/0")}

    # Initialize the list to store the processed rows
    processed_rows = []
```



```

# Iterate over each row in the dataframe
for _, row in df.iterrows():
    new_row = [f"{row['CHROM']}:{row['POS']}"]
    pairs = [("ABf1", "ABm1", "ABe1_ref", "ABe1_alt"), ("ABf2", "ABm2", "ABe2_ref", "ABe2_alt"),
             ("ABf3", "ABm3", "ABe3_ref", "ABe3_alt"), ("ABf4", "ABm4", "ABe4_ref", "ABe4_alt"),
             ("ABf5", "ABm5", "ABe5_ref", "ABe5_alt"), ("BAf1", "BAm1", "BAe1_ref", "BAe1_alt"),
             ("BAf2", "BAm2", "BAe2_ref", "BAe2_alt"), ("BAf3", "BAm3", "BAe3_ref", "BAe3_alt"),
             ("BAf4", "BAm4", "BAe4_ref", "BAe4_alt"), ("BAf5", "BAm5", "BAe5_ref", "BAe5_alt")]

    valid = True
    for f, m, ref_col, alt_col in pairs:
        if (row[f], row[m]) in ref_alt_pairs:
            new_row.extend([row[ref_col], row[alt_col]])
        elif (row[f], row[m]) in alt_ref_pairs:
            new_row.extend([row[alt_col], row[ref_col]])
        else:
            valid = False
            break

    if valid:
        processed_rows.append(new_row)

# Create a dataframe from the processed rows
columns = ['CHROM:POS'] + [f'ABe{i}_ref' for i in range(1, 6)] + [f'ABe{i}_alt' for i in range(1, 6)] + \
          [f'BAe{i}_ref' for i in range(1, 6)] + [f'BAe{i}_alt' for i in range(1, 6)]
processed_df = pd.DataFrame(processed_rows, columns=columns)

# Save the processed data to a new TXT file
processed_df.to_csv(output_file, sep='\t', index=False, header=True)

if __name__ == '__main__':
    if len(sys.argv) != 3:
        print("Usage: python3 process_gt_file.py <input_file> <output_file>")
        sys.exit(1)

    input_file = sys.argv[1]
    output_file = sys.argv[2]

    process_gt_file(input_file, output_file)
- . . . . .

```

Merging Gene Names

```

python annotate_snps.py sorted_GT_and_allele_expression.txt GTF_file annotated_allele_expression.txt

- . . . . . annotate_snps.py- . . . . .

# Import necessary libraries
import csv
import sys

# Function to parse a line from the GTF file
def parse_gtf_line(line):
    # Split the line into columns
    parts = line.strip().split('\t')
    # Ensure the line has at least 4 columns
    if len(parts) < 4:
        return None
    chrom = parts[0]
    start = int(parts[1])
    end = int(parts[2])

```

```

gene_info = parts[3]
return chrom, start, end, gene_info

# Function to annotate SNPs with gene information from a GTF file
def annotate_snps(vcf_file_path, gtf_file_path, output_file_path):
    # Read the VCF file
    with open(vcf_file_path, 'r') as vcf_file:
        data1 = csv.reader(vcf_file, delimiter="\t", skipinitialspace=True)

    # Read the GTF file
    with open(gtf_file_path, 'r') as gtf_file:
        data3_list = []
        for line in gtf_file:
            parsed_line = parse_gtf_line(line)
            if parsed_line:
                data3_list.append(parsed_line)
                # Debug output
                print(f"Parsed GTF entry: {parsed_line}")
            else:
                print(f"Skipping malformed GTF entry: {line.strip()}")

    # Prepare to write the output file
    with open(output_file_path, 'w', newline="") as out_file:
        data2 = csv.writer(out_file, delimiter="\t")

    # Flag to skip the header row
    is_header = True

    # Process each row in the VCF file
    for row in data1:
        # Write and skip the header row
        if is_header:
            data2.writerow(row + ['Gene'])
            is_header = False
            continue

        # Parse the CHROM:POS column
        chrom_pos = row[0]
        chrom, pos = chrom_pos.split(':')
        pos = int(pos)

        # Debug output
        print(f"Processing SNP: {chrom}:{pos}")

        # Find the corresponding gene in the GTF file
        merge = []
        for gtf_entry in data3_list:
            gtf_chrom, gtf_start, gtf_end, gene_info = gtf_entry
            # Debug output
            print(f"Checking against GTF entry: {gtf_chrom}:{gtf_start}-{gtf_end}")
            if gtf_chrom == chrom and gtf_start <= pos <= gtf_end:
                merge.append(gene_info)

        # If no corresponding gene is found, add "None"
        if not merge:
            merge.append("None")

        # Add the result as a new row
        for gene in merge:
            new_row = row + [gene]
            data2.writerow(new_row)

if __name__ == '__main__':

```

```

# Check if the script is called with exactly three arguments
if len(sys.argv) != 4:
    print("Usage: python annotate_snps.py <snp_file> <gtf_file> <output_file>")
    sys.exit(1)

vcf_file_path = sys.argv[1]
gtf_file_path = sys.argv[2]
output_file_path = sys.argv[3]

# Call the function to annotate SNPs
annotate_snps(vcf_file_path, gtf_file_path, output_file_path)
- . . . . .

```

Summarize expression levels by gene

```

python sum_by_gene.py annotated_allele_expression.txt allele_gene_expression.txt

- . . . . . - python sum_by_gene.py - . . . . . -

import pandas as pd
import sys

def sum_by_gene(input_file, output_file):
    # Read the input file into a DataFrame
    df = pd.read_csv(input_file, sep='\t')

    # Remove rows where the gene name is "None"
    df = df[df['Gene'] != 'None']

    # Select only the numeric columns
    numeric_cols = df.columns[1:-1]

    # Group the DataFrame by the 'Gene' column and sum the numeric columns
    summed_df = df.groupby('Gene')[numeric_cols].sum().reset_index()

    # Save the resulting DataFrame to the output file
    summed_df.to_csv(output_file, sep='\t', index=False)

if __name__ == '__main__':
    # Ensure the script is called with exactly two arguments
    if len(sys.argv) != 3:
        print("Usage: python sum_by_gene.py <input_file> <output_file>")
        sys.exit(1)

    # Get the input and output file paths from the command-line arguments
    input_file = sys.argv[1]
    output_file = sys.argv[2]

    # Call the sum_by_gene function with the provided file paths
    sum_by_gene(input_file, output_file)
- . . . . .

```

Statistical Analysis

Test for expression differences between alleles with edgeR (v3.40.2)

```

# Load necessary libraries
library(limma)

```

```

library(edgeR)

# Read the gene expression data from a file
egg_expr <- read.csv("allele_expression.txt", header = TRUE, row.names = 1, sep = "\t", quote = "")

# Split the data into two groups: AB and BA
AB = egg_expr[,1:10]
BA = egg_expr[,11:20]

# Define the group factor for the experimental conditions
famo_group <- factor(rep(c("fa", "mo"), times = 5))

# Create DGEList objects for both AB and BA datasets
d <- DGEList(counts = AB, group = famo_group)
e <- DGEList(counts = BA, group = famo_group)

# Create design matrices for the differential expression analysis
design_AB <- model.matrix(~ famo_group)
design_BA <- model.matrix(~ famo_group)

# Calculate normalization factors to account for library size differences
d <- calcNormFactors(d)
e <- calcNormFactors(e)

# Normalize the counts per million (CPM) for both datasets
AB_Norm <- cpm(d)
BA_Norm <- cpm(e)

# Estimate dispersions (common, trended, and tagwise) for both datasets
d <- estimateDisp(d, design_AB)
e <- estimateDisp(e, design_BA)

# Estimate tagwise dispersion specifically for both datasets
d <- estimateTagwiseDisp(d)
e <- estimateTagwiseDisp(e)

# Perform the exact test for differential expression for both datasets
result_AB <- exactTest(d)
result_BA <- exactTest(e)

# Convert the test results to data frames
table_AB <- as.data.frame(topTags(result_AB, n = nrow(AB)))
table_BA <- as.data.frame(topTags(result_BA, n = nrow(BA)))

# Merge the results from both tests by gene identifier
result1 <- merge(table_AB, table_BA, by = "row.names")

```

Table S1: Embryonic developmental stages and nucleic acid concentration. To determine the developmental stage of embryos used in the experiment, the following procedure was performed. Embryonic development was classified into five stages based on external morphology : I: Stage where the embryo is not yet visible (approximately 1–11 days old), II: Stage where the germ band becomes visible (approximately 12–15 days old), III: Stage where katasynthesis is completed (approximately 16–18 days old), IV: Stage of dorsal closure (approximately 19–21 days old), V: Stage where legs and antennae can be distinguished (approximately 22–35 days old). These classifications were made with reference to Matsuura and Kobayashi (2007). For these five stages, nucleic acids were extracted, and DNA and RNA concentrations were measured using a spectrophotometer. Embryos at Stage III, where sufficient nucleic acid yield could be obtained, were used for the experiment.

Group	Replication	DNA concentration (ng / ul)	RNA concentration (ng / ul)
I	1	5.928	4.871
I	2	6.206	4.937
II	1	10.787	9.130
II	2	10.985	9.017
III	1	17.939	14.406
III	2	19.948	16.035
IV	1	21.819	16.831
IV	2	18.985	15.007
V	1	41.408	33.024
V	2	44.085	34.924

Table S2: Termite colony information.

Colony	Location	Sampling date	Latitude	Longitude
A	Ichijoji Kotani, Kyoto, Kyoto	06 April 2023	35.04	135.80
B	Ichijoji Kotani, Kyoto, Kyoto	06 April 2023	35.04	135.80
C	Makabe Hadori, Sakuragawa, Ibaraki	16 April 2023	36.24	140.10
D	Iwakuramuramatsu Kyoto, Kyoto	21 April 2023	35.09	135.78
E	Makabe Yamao, Sakuragawa, Ibaraki	16 April 2023	36.25	140.13
F	Iwakuramuramatsu Kyoto, Kyoto	21 April 2023	35.09	135.78
G	Imaduchomukugawa Takashima, Shiga	21 May 2022	35.39	135.89

Table S3: Sex identification of eggs. The number of male and female individuals identified using the sex-specific genetic marker is shown for each colony.

Colony	No. eggs investigated	Male	Female	Unidentified
A	24	11	13	0
B	29	18	11	0
C	24	10	13	1
D	24	10	13	1
E	35	19	15	1
F	24	8	16	0

Table S4: RNA-seq sample statistics. PFE: parthenogenetically-produced female embryo. SFE: sexually-produced female embryo, SME: sexually-produced male embryo. Each group includes five colony replicates, with each sample consisting of pooled multiple eggs. The number of obtained paired reads and mapping results are shown. The data are available through the DRA accession.

Group	Colony	Number of pooled eggs	Number of pair-reads	Uniquely mapped reads number (mapped rate)	DRA accession
PFE	A	10	14936872	13290426 (88.98%)	DRX595295
PFE	B	10	14065182	12340964 (87.74%)	DRX595296
PFE	C	10	16641453	14733377 (88.53%)	DRX595297
PFE	D	10	17534795	15507359 (88.44%)	DRX595298
PFE	E	10	14285311	12693979 (88.86%)	DRX595299
SFE	A	13	14139201	12652715 (89.49%)	DRX595305
SFE	B	11	14310417	11557895 (80.77%)	DRX595306
SFE	C	14	13903616	12202159 (87.76%)	DRX595307
SFE	D	13	15225545	13566299 (89.10%)	DRX595308
SFE	E	15	16574761	13807804 (83.31%)	DRX595309
SME	A	10	19760820	17665324 (89.40%)	DRX595300
SME	C	9	13981840	12011124 (85.91%)	DRX595301
SME	D	10	15465962	13525802 (87.46%)	DRX595302
SME	E	15	18492549	16251227 (87.88%)	DRX595303

Table S5: Statistical results of DEGs among sexually-produced male embryo (SME), sexually-produced female embryo (SFE) and parthenogenetically-produced female embryo (PFE). The category "PFE-high" refers to genes that showed significantly higher expression levels in PFE compared to other embryo types. Similarly, "PFE-low" refers to genes that exhibited significantly lower expression levels in PFE than in other embryo types. "SME-high" indicates genes that were significantly upregulated in SME compared to PFE and SFE, while "SME-low" denotes genes with significantly lower expression levels in SME. Finally, "SFE-high" refers to genes with significantly higher expression levels in SFE than in PFE and SME. The columns include the false discovery rates (FDR) of the ANOVA and pairwise comparisons between different conditions: SME-PFE (sexually-produced male embryo vs. parthenogenetically-produced female embryo), SFE-PFE (sexually-produced female embryo vs. parthenogenetically-produced female embryo), and SME-SFE (sexually-produced male embryo vs. sexually-produced female embryo). Values where $FDR < 0.05$ are indicated in bold.

Gene ID	Category	FDR of ANOVA	FDR of SME-PFE	FDR of SFE-PFE	FDR of SME-SFE
ANN35001	PFE-high	2.542E-02	1.790E-03	1.976E-03	9.982E-01
ANN13362	PFE-high	7.186E-05	2.340E-03	8.940E-03	7.333E-01
ANN12308	PFE-high	1.540E-06	2.558E-03	1.172E-02	6.744E-01
ANN12985	PFE-high	3.473E-09	4.239E-03	8.716E-03	9.138E-01
ANN03666	PFE-high	9.973E-04	3.189E-04	8.000E-04	8.387E-01
ANN06780	PFE-high	4.800E-02	1.547E-04	1.939E-04	9.880E-01
ANN03715	PFE-low	6.631E-03	1.001E-03	2.951E-02	1.739E-01
ANN08428	PFE-low	9.199E-05	2.933E-02	8.233E-04	1.451E-01
ANN04383	PFE-low	2.199E-02	1.082E-02	1.186E-02	9.985E-01
ANN29344	PFE-low	2.932E-03	3.167E-03	8.262E-03	8.524E-01
ANN21136	PFE-low	3.377E-06	3.493E-02	2.684E-02	9.881E-01
ANN28791	PFE-low	1.936E-05	1.249E-04	3.770E-03	1.343E-01
ANN25122	PFE-low	6.631E-03	2.945E-02	3.907E-02	9.862E-01
ANN32106	PFE-low	2.987E-02	7.567E-03	3.312E-02	6.962E-01
ANN13837	SME-high	3.473E-09	7.451E-05	9.228E-01	1.311E-04
ANN10061	SME-high	3.346E-11	1.034E-05	9.885E-01	1.245E-05
ANN10097	SME-high	3.473E-02	5.166E-03	8.402E-01	1.421E-02
ANN26436	SME-high	8.991E-03	5.116E-03	9.779E-01	7.317E-03
ANN09810	SME-high	2.996E-02	4.505E-02	9.514E-01	2.636E-02

ANN07501	SME-high	1.461E-05	1.469E-04	1.463E-01	4.153E-03
ANN14927	SME-high	6.035E-03	1.365E-03	2.406E-01	2.804E-02
ANN28792	SME-high	3.323E-06	2.441E-02	6.408E-02	3.177E-04
ANN05965	SME-high	3.475E-03	8.607E-03	7.267E-01	3.434E-02
ANN27764	SME-high	2.067E-02	1.337E-03	5.652E-01	8.254E-03
ANN33511	SME-high	3.473E-02	5.117E-03	6.602E-01	2.489E-02
ANN01717	SME-high	2.726E-03	1.059E-03	8.319E-02	6.908E-02
ANN06279	SME-low	1.740E-03	8.718E-03	5.833E-01	5.325E-02
ANN06587	SME-high	4.985E-02	2.954E-02	8.665E-01	7.287E-02
ANN09929	SME-low	1.214E-03	1.033E-03	1.346E-01	4.062E-02
ANN10398	SME-high	3.637E-02	1.201E-02	5.838E-01	7.253E-02
ANN11722	SME-high	2.888E-02	2.800E-02	4.268E-01	2.436E-01
ANN14138	SME-high	6.987E-05	5.972E-06	2.979E-03	3.294E-03
ANN17384	SFE-high	1.360E-02	9.439E-01	1.050E-02	5.882E-03
ANN20636	SME-low	1.807E-03	2.991E-04	4.985E-01	2.019E-03
ANN21288	SME-high	2.970E-03	5.752E-03	6.017E-01	3.336E-02
ANN22200	SME-high	3.473E-02	2.762E-03	2.710E-01	5.029E-02
ANN22399	SME-low	6.667E-08	3.767E-02	7.597E-02	9.159E-01
ANN22638	SME-high	4.147E-02	3.495E-03	8.061E-01	1.072E-02
ANN24502	SME-low	4.763E-05	2.559E-03	6.812E-02	1.959E-01
ANN25403	SFE-high	3.473E-02	9.771E-01	9.856E-03	6.833E-03
ANN25845	SME-high	1.886E-03	4.995E-03	2.908E-01	8.288E-02
ANN26462	SME-low	2.996E-02	1.067E-02	8.558E-02	4.899E-01
ANN26796	SME-high	6.035E-03	8.079E-03	8.029E-02	4.258E-01
ANN27365	SME-high	2.996E-02	2.296E-02	2.972E-01	3.080E-01
ANN27766	SFE-high	6.631E-03	9.515E-02	2.907E-03	1.619E-01
ANN28024	SME-low	5.092E-03	2.281E-02	1.608E-01	7.261E-04
ANN28522	SME-high	2.932E-03	2.613E-02	8.704E-01	6.381E-02
ANN29977	SME-high	1.081E-03	4.426E-03	5.352E-01	3.133E-02
ANN31090	SFE-high	1.264E-03	1.431E-01	4.032E-01	1.391E-02
ANN31937	SFE-high	1.994E-04	6.194E-01	5.582E-03	1.077E-03
ANN33925	SFE-high	6.324E-03	8.452E-02	8.463E-01	3.203E-02

ANN35037	SME-high	1.562E-02	3.616E-03	7.260E-01	1.432E-02
ANN35549	SFE-high	4.209E-03	2.633E-01	9.918E-03	1.749E-01
ANN36075	SFE-high	2.067E-02	8.553E-02	2.619E-02	4.418E-04

Table S6: Summary results of blast homology searches of DEGs. The category "PFE-high" refers to genes that showed significantly higher expression levels in PFE compared to other embryo types. Similarly, "PFE-low" refers to genes that exhibited significantly lower expression levels in PFE than in other embryo types. "SME-high" indicates genes that were significantly upregulated in SME compared to PFE and SFE, while "SME-low" denotes genes with significantly lower expression levels in SME. Finally, "SFE-high" refers to genes with significantly higher expression levels in SFE than in PFE and SME. The "Method" column indicates whether the search was performed using blastx or blastp, and genes with blank entries in this column did not yield any hits in the BLAST search. The columns following the "Gene Name" column show information on the top BLAST search hit for each gene.

Gene ID	Category	Method	Gene Name	Accession No.	Organism	E-value	Identity
ANN35001	PFE-high	blastx	neuropeptide CCHamide-2 receptor-like isoform X5	XP_021937590.1	<i>Zootermopsis nevadensis</i>	4.E-50	71.03%
ANN13362	PFE-high	blastp	cold and drought-regulated protein CORA-like	XP_023714537.1	<i>Cryptotermes secundus</i>	3.E-24	55.33%
ANN12308	PFE-high	blastp	hemolymph juvenile hormone binding protein (JHBP)	AGM32526.1	<i>Coptotermes formosanus</i>	1.E-113	91.62%
ANN12985	PFE-high	blastx	putative mediator of RNA polymerase II transcription subunit 26	XP_021941098.1	<i>Zootermopsis nevadensis</i>	4.E-92	55.36%
ANN03666	PFE-high	blastx	topoisomerase 1	QPZ88800.1	<i>Noonilla sp. zhg-my02</i>	4.E-34	88.16%
ANN06780	PFE-high	blastx	facilitated trehalose transporter Tret1-like	XP_021939138.1	<i>Zootermopsis nevadensis</i>	3.E-30	89.04%
ANN03715	PFE-low	blastx	protein amalgam-like	XP_021925622.1	<i>Zootermopsis nevadensis</i>	5.E-31	90.14%
ANN08428	PFE-low	blastx	DNA N6-methyl adenine demethylase	XP_033610652.1	<i>Cryptotermes secundus</i>	0.E+00	67.16%
ANN04383	PFE-low	blastx	early growth response protein 1	XP_023714639.1	<i>Cryptotermes secundus</i>	4.E-120	79.38%
ANN29344	PFE-low	blastx	neurobeachin-like	XP_023727031.2	<i>Cryptotermes secundus</i>	4.E-43	94.19
ANN21136	PFE-low	blastx	Serine/threonine-protein kinase NLK	PNF40964.1	<i>Cryptotermes secundus</i>	9.E-35	97.26%
ANN28791	PFE-low	blastx	transcription initiation factor TFIID subunit 3 isoform X3	XP_023723453.1	<i>Cryptotermes secundus</i>	2.E-146	84.38%
ANN25122	PFE-low	blastx	protein winged eye isoform X1	XP_023707360.2	<i>Cryptotermes secundus</i>	3.E-135	80.16%
ANN32106	PFE-low	-	-	-	-	-	-
ANN13837	SME-high	blastx	protein ALP1-like	XP_026732211.1	<i>Trichoplusia ni</i>	6.E-128	60.42%
ANN10061	SME-high	blastx	cadherin-related tumor suppressor	XP_023713333.2	<i>Cryptotermes secundus</i>	0.E+00	89.57%
ANN10097	SME-high	blastx	coiled-coil domain-containing protein 191	XP_023721521.1	<i>Cryptotermes secundus</i>	3.E-78	50.57%

ANN26436	SME-high	blastx	dnaJ homolog subfamily C member 22	XP_021915260.1	<i>Zootermopsis nevadensis</i>	7.E-46	86.02%
ANN09810	SME-high	blastx	protein doublesex	BCX65401.1	<i>Reticulitermes speratus</i>	0.E+00	99.57%
ANN07501	SME-high	blastx	Histone-lysine N-methyltransferase MLL5	KDR16326.1	<i>Zootermopsis nevadensis</i>	2.E-47	60.11%
ANN14927	SME-high	blastx	Myosin heavy chain 95F	PNF18982.1	<i>Cryptotermes secundus</i>	1.E-44	98.86%
ANN28792	SME-high	blastx	Pyruvate dehydrogenase E1 component subunit beta	PSN52030.1	<i>Blattella germanica</i>	1.E-31	93.06%
ANN05965	SME-high	blastx	Pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16	PSN46456.1	<i>Blattella germanica</i>	1.E-38	98.73%
ANN27764	SME-high	blastx	titin isoform X10	XP_033607269.1	<i>Cryptotermes secundus</i>	5.E-06	63.64%
ANN33511	SME-high	blastx	tRNA (uracil(54)-C(5))-methyltransferase homolog isoform X2	XP_023720596.1	<i>Cryptotermes secundus</i>	2.E-35	83.33%
ANN22638	SME-high	blastx	XK-related protein 4-like	XP_021921864.1	<i>Zootermopsis nevadensis</i>	1.E-39	96.84%
ANN29977	SME-high	blastx	zinc finger MYND domain-containing protein 11 isoform X2	XP_023724206.1	<i>Cryptotermes secundus</i>	2.E-26	92.42%
ANN21288	SME-high	-	-	-	-	-	-
ANN35037	SME-high	-	-	-	-	-	-
ANN14138	SME-high	blastx	histone-lysine N-methyltransferase 2C-like	XP_046395799.1	<i>Ischnura elegans</i>	2.E-32	98.36%
ANN01717	SME-high	blastx	ADP-ribosylation factor GTPase-activating protein 1	KAG8334055.1	<i>Homalodisca vitripennis</i>	1.E-27	98.25%
ANN27365	SME-high	blastx	protein dachous	XP_033606430.1	<i>Cryptotermes secundus</i>	3.E-49	88.79%
ANN10398	SME-high	blastx	luc7-like protein 3 isoform X2	XP_046988031.1	<i>Schistocerca americana</i>	2.E-41	87.36%
ANN26796	SME-high	blastx	39S ribosomal protein L37, mitochondrial	KDQ71517.1	<i>Zootermopsis nevadensis</i>	6.E-41	74.31%
ANN28522	SME-high	blastx	Plectin	PSN48665.1	<i>Blattella germanica</i>	5.E-38	97.70%
ANN06587	SME-high	blastp	tRNA (uracil(54)-C(5))-methyltransferase homolog	XP_021938392.1	<i>Zootermopsis nevadensis</i>	7.E-129	41.12%
ANN25845	SME-high	blastx	ubiquitin-conjugating enzyme E2 Q2 isoform X2	XP_023721647.1	<i>Cryptotermes secundus</i>	3.E-25	98.15%
ANN11722	SME-high	blastx	ubiquitin-like protein 7 isoform X2	XP_023726645.1	<i>Cryptotermes secundus</i>	5.E-18	90.91%
ANN22200	SME-high	blastx	zinc finger protein 37 isoform X1	XP_023727804.1	<i>Cryptotermes secundus</i>	1.E-69	80.32%
ANN28024	SME-low	blastx	Protein Asterix	PSN52025.1	<i>Blattella germanica</i>	3.E-17	82.50%
ANN20636	SME-low	blastx	serine/arginine repetitive matrix protein 2 isoform X2	XP_023726304.1	<i>Cryptotermes secundus</i>	2.E-130	72.32%
ANN09929	SME-low	-	-	-	-	-	-
ANN06279	SME-low	blastx	chromobox protein homolog 5 isoform X1	XP_023705422.1	<i>Cryptotermes secundus</i>	9.E-24	55.56%
ANN22399	SME-low	blastx	L-selectin	KAK3926265.1	<i>Frankliniella fusca</i>	9.E-26	82.09%
ANN24502	SME-low	blastx	nose resistant to fluoxetine protein 6 isoform X2	XP_033607889.1	<i>Cryptotermes secundus</i>	6.E-23	73.97%

ANN26462	SME-low	-	-	-	-	-	-
ANN25403	SFE-high	blastx	alpha-1A adrenergic receptor	XP_023713291.1	<i>Cryptotermes secundus</i>	2.E-132	77.08%
ANN17384	SFE-high	blastx	carboxypeptidase B-like isoform X2	XP_021915563.1	<i>Zootermopsis nevadensis</i>	1.E-111	91.62%
ANN31937	SFE-high	blastx	histone H1-III	XP_023710263.1	<i>Cryptotermes secundus</i>	5.E-33	91.43%
ANN36075	SFE-high	-	-	-	-	-	-
ANN33925	SFE-high	blastx	MKL/myocardin-like protein 2 isoform X2	XP_021929358.1	<i>Zootermopsis nevadensis</i>	3.E-66	86.50%
ANN31090	SFE-high	blastx	5'-nucleotidase isoform X2	XP_023719629.1	<i>Cryptotermes secundus</i>	2.E-26	70.33%
ANN35549	SFE-high	blastx	ataxin-2-like	XP_021933346.1	<i>Zootermopsis nevadensis</i>	5.E-32	87.06%
ANN27766	SFE-high	blastx	titin isoform X2	XP_033606476.1	<i>Cryptotermes secundus</i>	4.E-51	92.52%

Table S7: Gene function of DEGs. The category "PFE-high" refers to genes that showed significantly higher expression levels in PFE compared to other embryo types. Similarly, "PFE-low" refers to genes that exhibited significantly lower expression levels in PFE than in other embryo types. "SME-high" indicates genes that were significantly upregulated in SME compared to PFE and SFE, while "SME-low" denotes genes with significantly lower expression levels in SME. Finally, "SFE-high" refers to genes with significantly higher expression levels in SFE than in PFE and SME. Rows left blank indicate genes that did not have any hits in the BLAST search.

Gene ID	Category	Gene name	Function
ANN35001	PFE-high	<i>ccha2-r</i>	Neurotransmission
ANN13362	PFE-high	<i>cora</i>	Stress response
ANN12308	PFE-high	<i>hjhb</i>	Transport of juvenile hormone
ANN12985	PFE-high	<i>med26</i>	Mediation of transcription
ANN03666	PFE-high	<i>top1</i>	Relaxation of DNA supercoiling
ANN06780	PFE-high	<i>tret1</i>	Trehalose transport
ANN03715	PFE-low	<i>ama</i>	Cell adhesion, Neural development
ANN08428	PFE-low	<i>nmad-1</i>	Regulation of epigenetic modifications, DNA repair
ANN04383	PFE-low	<i>egr1</i>	Regulation of gene expression, Cell differentiation
ANN29344	PFE-low	<i>nbea</i>	Neurodevelopment
ANN21136	PFE-low	<i>nlk</i>	Cell signaling
ANN28791	PFE-low	<i>taf3</i>	Regulation of transcription initiation
ANN25122	PFE-low	<i>wge</i>	Eye development
ANN32106	PFE-low	-	-
ANN13837	SME-high	<i>alp1-like</i>	Stress response

ANN10061	SME-high	<i>cadherin-related tumor suppressor</i>	Cell adhesion
ANN10097	SME-high	<i>ccdc191</i>	Cytoskeletal organization
ANN26436	SME-high	<i>dnajc22</i>	Protein folding
ANN09810	SME-high	<i>dsx</i>	Sex determination
ANN07501	SME-high	<i>kmt2e</i>	Epigenetic modification
ANN14927	SME-high	<i>mhc95F</i>	Muscle contraction
ANN28792	SME-high	<i>pdhb</i>	Energy metabolism
ANN05965	SME-high	<i>prp16</i>	Regulation of splicing
ANN27764	SME-high	<i>ttn</i>	Maintenance of muscle structure
ANN33511	SME-high	<i>trmA</i>	tRNA modification
ANN22638	SME-high	<i>xkr4</i>	Regulation of cell death
ANN29977	SME-high	<i>zmynd11</i>	Transcriptional repression
ANN21288	SME-high	-	-
ANN35037	SME-high	-	-
ANN14138	SME-high	<i>kmt2c-like</i>	Epigenetic modification
ANN01717	SME-high	<i>arfgap1</i>	Vesicle transport
ANN27365	SME-high	<i>ds</i>	Planar cell polarity
ANN10398	SME-high	<i>luc7l3</i>	RNA splicing
ANN26796	SME-high	<i>mrpl37</i>	Mitochondrial translation
ANN28522	SME-high	<i>plec</i>	Cytoskeletal stabilization, Cell adhesion
ANN06587	SME-high	<i>trmA</i>	tRNA modification
ANN25845	SME-high	<i>ube2q2</i>	Protein ubiquitination
ANN11722	SME-high	<i>ubl7</i>	Intracellular protein modification, Ubiquitin-proteasome system
ANN22200	SME-high	<i>zfp37</i>	Transcriptional repression
ANN28024	SME-low	<i>asterix</i>	Cell differentiation
ANN20636	SME-low	<i>srrm2</i>	Regulation of splicing
ANN09929	SME-low	-	-
ANN06279	SME-low	<i>cbx5</i>	Heterochromatin formation, Gene silencing
ANN22399	SME-low	<i>cd62l</i>	Cell adhesion

ANN24502	SME-low	<i>nrf6</i>	Stress response, Xenobiotic transport
ANN26462	SME-low	-	-
ANN25403	SFE-high	<i>adrala</i>	Cell signaling, Regulation of vasoconstriction
ANN17384	SFE-high	<i>cpb-like</i>	Protein degradation
ANN31937	SFE-high	<i>histone H1-III</i>	Maintenance of chromatin structure
ANN36075	SFE-high	-	-
ANN33925	SFE-high	<i>mkl2</i>	Regulation of muscle development
ANN31090	SFE-high	<i>nt5</i>	Nucleotide metabolism
ANN35549	SFE-high	<i>atxn2l</i>	Regulation of RNA metabolism, Formation of stress granules
ANN27766	SFE-high	<i>ttn</i>	Maintenance of muscle structure

Table S8: Genes with splicing isoforms detected among parthenogenetically and sexually produced embryos. A checkmark (y) indicates significant splicing variation in the corresponding pairwise comparison (PFE_SME, SFE_PFE, SFE_SME); “n” indicates no significant difference. Statistical significance was determined using a Bonferroni-corrected false discovery rate (FDR < 0.05).

Gene ID	PFE_SME	SFE_PFE	SFE_SME
ANN00585	y	n	n
ANN00923	n	y	n
ANN00973	y	n	y
ANN01092	y	n	n
ANN02011	y	n	n
ANN05093	n	n	y
ANN05544	n	y	n
ANN05785	y	n	n
ANN05981	y	n	y
ANN06898	n	n	y
ANN07611	n	n	y
ANN07637	n	n	y
ANN07641	n	y	n
ANN07656	n	y	n
ANN07884	n	n	y

ANN08052	n	n	y
ANN08280	y	n	n
ANN09059	n	n	y
ANN09534	y	n	n
ANN09659	y	n	y
ANN09673	n	y	n
ANN09927	y	n	n
ANN09932	n	n	y
ANN10542	n	y	n
ANN11027	n	n	y
ANN11189	y	y	n
ANN11359	n	y	n
ANN11494	y	n	n
ANN12100	y	y	y
ANN13060	y	n	y
ANN13286	n	y	n
ANN13320	y	y	n
ANN13960	n	n	y
ANN14265	n	y	y
ANN14686	y	n	n
ANN15535	n	y	n
ANN15714	n	y	n
ANN16629	y	y	n
ANN16730	y	y	n
ANN16822	y	n	n
ANN17085	n	y	n
ANN18304	y	n	n
ANN18326	y	n	n
ANN18360	y	y	n
ANN18525	n	n	y
ANN18931	y	y	n
ANN19005	n	n	y

ANN19827	n	n	y
ANN20222	n	y	n
ANN20240	n	y	n
ANN20384	y	n	y
ANN21980	n	y	n
ANN22012	y	n	n
ANN22457	y	n	y
ANN22881	n	y	y
ANN23580	y	n	n
ANN23813	n	y	y
ANN25029	n	y	n
ANN25237	y	n	n
ANN25347	y	n	n
ANN25442	n	y	y
ANN25731	n	y	n
ANN26463	n	y	n
ANN26616	n	y	n
ANN27551	y	n	n
ANN28514	y	n	n
ANN28563	y	n	y
ANN28578	y	y	n
ANN28610	y	n	y
ANN28673	y	n	y
ANN29321	n	n	y
ANN30158	n	y	y
ANN30290	n	y	n
ANN30355	n	y	y
ANN30387	n	n	y
ANN30749	n	y	n
ANN30766	y	n	n
ANN30842	n	y	n
ANN31181	n	y	y

ANN31803	n	y	n
ANN31808	n	y	n
ANN32230	n	y	n
ANN32643	n	n	y
ANN33079	n	y	y
ANN33271	n	y	n
ANN33580	y	y	n
ANN33924	y	n	n
ANN34433	n	y	n
ANN34527	y	n	n
ANN34613	n	y	y
ANN34769	n	y	y
ANN34846	n	y	n
ANN34875	n	y	y
ANN35137	n	y	n
ANN35518	y	n	n
ANN35662	y	n	n

Table S9: Sample statistics for WGS of parents and RNA-seq of embryos. The number of paired reads and the mapping results are shown for 20 WGS samples, consisting of five replicates each of fathers and mothers from the Kyoto colony and fathers and mothers from the Tsukuba colony, as well as for 10 RNA-seq samples, consisting of five replicates each of embryos with a Kyoto colony father and Tsukuba colony mother, and embryos with a Tsukuba colony father and Kyoto colony mother. The data are available through the DRA accession.

Sample	Number of pair-reads	Uniquely mapped reads number (mapped rate)	DRA accession
Kyoto father 1	220264393	212881193 (97.49 %)	DRX492617
Kyoto father 2	203284855	196504970 (97.51 %)	DRX492618
Kyoto father 3	199506207	192884144 (97.50 %)	DRX492619
Kyoto father 4	263301377	254522389 (97.50 %)	DRX492620
Kyoto father 5	190235705	183941092 (97.52 %)	DRX492621
Kyoto mother 1	200037528	192770004 (97.25 %)	DRX492622
Kyoto mother 2	259564596	250780293 (97.50 %)	DRX492623

Kyoto mother 3	261511207	252942548 (97.55 %)	DRX492624
Kyoto mother 4	256903720	248313476 (97.53 %)	DRX492625
Kyoto mother 5	202991263	195934776 (97.53 %)	DRX492626
Tsukuba father 1	196557202	189832920 (97.50 %)	DRX492627
Tsukuba father 2	230153834	222263606 (97.50 %)	DRX492628
Tsukuba father 3	203658638	196703751 (97.52 %)	DRX492629
Tsukuba father 4	217808683	210432981 (97.50 %)	DRX492630
Tsukuba father 5	281508710	272064991 (97.51 %)	DRX492631
Tsukuba mother 1	237775928	229770178 (97.55 %)	DRX492632
Tsukuba mother 2	224659782	217158855 (97.57 %)	DRX492633
Tsukuba mother 3	204976500	189659161 (93.47 %)	DRX492634
Tsukuba mother 4	276714640	267250122 (97.52 %)	DRX492635
Tsukuba mother 5	246850729	238453226 (97.49 %)	DRX492636
faKY-moTS embryo 1	14224186	12511596 (87.96 %)	DRX595310
faKY-moTS embryo 2	24447062	21685776 (88.71 %)	DRX595311
faKY-moTS embryo 3	16269910	14394325 (88.47 %)	DRX595312
faKY-moTS embryo 4	18164562	16051650 (88.37 %)	DRX595313
faKY-moTS embryo 5	16882849	14926228 (88.41 %)	DRX595314
faTS-moKY embryo 1	17835287	15683975 (87.94 %)	DRX595315
faTS-moKY embryo 2	20478805	17954291 (87.67 %)	DRX595316
faTS-moKY embryo 3	18306220	16178875 (88.38 %)	DRX595317
faTS-moKY embryo 4	14981861	13201057 (88.11 %)	DRX595318
faTS-moKY embryo 5	13931935	12342365 (88.59 %)	DRX595319

Table S10: Genes with maternal biased expression. Crosses were performed between a Kyoto-father and a Tsukuba-mother, and between a Tsukuba-father and a Kyoto-mother. The "mean of maternal expression" represents the average expression of maternal alleles in the offspring from both crosses. Similarly, the "mean of paternal expression" represents the average expression of paternal alleles. All genes showing maternal-biased expression were mitochondrial-related.

gene_name	mean of maternal expression	mean of paternal expression	ratio (maternal / paternal)
<i>trnA</i>	1789.29	122.54	14.60
<i>cox3</i>	61848.62	3457.94	17.89

<i>nad6</i>	18399.24	1002.37	18.36
<i>trnL1</i>	87992.16	4537.70	19.39
<i>cox2</i>	45011.89	2237.59	20.12
<i>cob</i>	182804.87	8670.38	21.08
<i>nad2</i>	32757.84	1430.56	22.90
<i>trnS2</i>	64101.78	2499.29	25.65
<i>coI</i>	87041.34	3380.82	25.75
<i>coIII</i>	36157.72	1298.00	27.86
<i>nad5</i>	96478.34	3431.31	28.12
<i>cox1</i>	89668.97	3151.37	28.45
<i>nad4</i>	28685.33	873.10	32.85

Table S11: Statistical results of colony biased genes. "KY-high" refers to genes that exhibited higher expression in the Kyoto allele, while "TS-high" indicates genes with higher expression in the Tsukuba allele. The FDR of paKY-maTS represents the false discovery rates (FDR) comparing the expression levels of the paternal Kyoto allele and the maternal Tsukuba allele in the egg samples obtained from the cross between Kyoto father and Tsukuba mother. Similarly, the FDR of paTS-maKY compares the expression levels of the paternal Tsukuba allele and the maternal Kyoto allele in the egg samples obtained from the cross between Tsukuba father and Kyoto mother.

Gene ID	Category	FDR of paKY-maTS	FDR of paTS-maKY
ANN15985	KY-high	2.003E-04	7.628E-04
ANN10404	KY-high	3.888E-03	1.765E-04
ANN16921	KY-high	3.287E-07	5.463E-09
ANN24776	KY-high	3.685E-02	2.893E-08
ANN25419	KY-high	4.444E-02	4.850E-02
ANN30814	TS-high	3.191E-08	2.935E-14
ANN08526	TS-high	6.278E-12	4.226E-12
ANN13889	TS-high	2.369E-03	7.793E-10
ANN30700	TS-high	5.580E-09	1.667E-09
ANN27892	TS-high	3.685E-02	5.694E-05
ANN03633	TS-high	1.743E-02	5.31E-06
ANN13667	TS-high	1.403E-02	3.419E-05

Table S12: Summary results of blast homology searches of colony biased genes. "KY-high" refers to genes that exhibited higher expression in the Kyoto allele, while "TS-high" indicates genes with higher expression in the Tsukuba allele. The "Method" column indicates whether the search was performed using blastx or blastp, and genes with blank entries in this column did not yield any hits in the BLAST search. The columns following the "Gene Name" column show information on the top BLAST search hit for each gene.

Gene ID	Category	Method	Gene Name	Accession No.	Organism	E-value	Identity
ANN15985	KY-high	blastx	transposase	KMQ86389.1	<i>Lasius niger</i>	2.E-45	70.29%
ANN10404	KY-high	-	-	-	-	-	-
ANN16921	KY-high	-	-	-	-	-	-
ANN24776	KY-high	-	-	-	-	-	-
ANN25419	KY-high	-	-	-	-	-	-
ANN30814	TS-high	blastx	activating signal cointegrator 1 complex subunit	KAJ4428810.1	<i>Periplaneta americana</i>	4.E-14	79.07%
ANN08526	TS-high	blastx	ankyrin repeat protein	AGM32371.1	<i>Coptotermes formosanus</i>	7.E-51	60.62%
ANN13889	TS-high	blastx	modifier of mdg4- like	XP_021928731.1	<i>Zootermopsis nevadensis</i>	3.E-30	44.68%
ANN30700	TS-high	blastx	zinc finger protein OZF	XP_023714764.1	<i>Cryptotermes secundus</i>	4.E-02	50.91%
ANN27892	TS-high	blastx	zinc finger protein OZF-like	XP_021924092.1	<i>Zootermopsis nevadensis</i>	1.E-44	61.06%
ANN03633	TS-high	-	-	-	-	-	-
ANN13667	TS-high	-	-	-	-	-	-

Table S13: Gene function of colony biased genes. "KY-high" refers to genes that exhibited higher expression in the Kyoto allele, while "TS-high" indicates genes with higher expression in the Tsukuba allele. Rows left blank indicate genes that did not have any hits in the BLAST search.

Gene ID	Category	Gene name	Function
ANN15985	KY-high	<i>transoposase</i>	DNA cutting and rejoining
ANN10404	KY-high	-	-
ANN16921	KY-high	-	-
ANN24776	KY-high	-	-
ANN25419	KY-high	-	-
ANN30814	TS-high	<i>asc-I</i>	Transcription co-activation

ANN08526	TS-high	<i>ank</i>	Protein-protein interactions
ANN13889	TS-high	<i>mod(mdg4)</i>	Regulation of transcription
ANN30700	TS-high	<i>ozf</i>	Regulation of gene expression
ANN27892	TS-high	<i>ozf-like</i>	Regulation of gene expression
ANN03633	TS-high	-	-
ANN13667	TS-high	-	-