COMPARATIVE HEAD TRANSCRIPTOME ANALYSIS IN TWO SPECIES OF THE *DROSOPHILA ELEGANS* SPECIES SUBGROUP

Luo, Fangning (Sophia)
Department of Ecology and Evolutionary Biology

Laboratory of José Ranz

The molecular basis of interspecific trait evolution remains poorly understood. The sister species *Drosophila elegans* and *D. gunungcola* have diverged in one component of their courtship behavior, the frontal wing display in males. This study investigates the genes and pathways that have diverged in expression between the two species in relation to frontal wing display, and more generally how their head transcriptome has become differentiated between sexes and developmental stages. The analysis of RNAseq data from males and females of both species at two different stages (0 days and 5 days) revealed that roughly 30% of the expressed genes were differentially expressed between the sexes, developmental stages, or both. The number of differentially expressed genes between stages within the same sex was substantially larger than the number of differences between the sexes for the same stage. Focusing on differentially expressed genes at Day 5 in males, the sex that features the interspecific phenotypic difference, the search for overrepresented functional terms uncovered functional signatures potentially related to wing display behavior and mating duration. Likewise, in male-biased genes in expression in Day 5, functional terms associated with synaptic transmission and signaling were significantly overrepresented. These findings provide new insights into the identity of genes and pathways potentially relevant for neural circuit functioning during courtship rituals while informing about candidate genes to be targeted in future functional studies.

In natural populations, every organism is exposed to selection pressures at two levels, reproduction and viability. Some traits are functionally related as they are influenced, at least partially, by the same genes, some of them contributing to the developmental program of the organism. As a result, the co-evolution of functionally related traits is shaped by genetic and developmental constraints (Jacob, 1977). While the theory of co-evolution of functionally related

traits is not recent (e.g. Merilä & Sheldon, 2001), the understanding of its molecular mechanisms remains insufficient, postponing the study of its empirical evidence at the mechanistic level.

With the advancement of high-throughput genome technologies, we can gain a better understanding of the molecular mechanisms of biological processes behind phenotypic evolution. Species in the *Drosophila melanogaster* species group are suitable for such purposes. This species group includes several species subgroups. Two sister species, *Drosophila gunungcola* and *D. elegans* belong to the *Drosophila elegans* species subgroup that is part of the mentioned species subgroup. The indicated species are found in the Oriental part of Asia, featuring differences in functionally related traits (Koshikawa, 2020). *D. elegans* shows male-specific wing pigmentation and frontal wing display while *D. gunungcola* does not show either trait (Massey et al., 2020). This species pair serves as an excellent system to study the potential coevolution of new traits at the molecular level.

Courtship is a complex behavior. Frontal wing display dance during courtship involves males extending their wings outward, positioning their dorsal wing surfaces towards the female, and moving them in an up-and-down motion (Massey et al., 2020). In *D. melanogaster*, this behavior is specified by the gene *fruitless (fru)* (Baker et al., 2001). In *D. elegans*, however, the genetic basis for courtship remains unknown. Gain- or loss-of-function mutations affecting behavioral genes in *D. melanogaster* may also impact the courtship behavior in *D. elegans*. The pleiotropic effect of pigmentation genes could also play a role as melanin biosynthesis-related genes are known to contribute to the pigmentation variation among *Drosophila* species, being involved in dopamine metabolism (Takahashi, 2013), and therefore potentially impacting behavior (Kaźmierczak & Nicola, 2022). Further, it has been suggested that courtship can be triggered by genes that function in neural circuits acting as a decision switch between aggression

and courtship (Koganezawa et al., 2016). Evolutionary changes in decision switch genes' expression of *D. elegans* and *D. gunungcola* may result in disparities in their courtship behavior. Additionally, since frontal display is a male-specific trait, it is also possible that the decision switch genes are expressed in a sexually-biased manner. In summary, the species pair *D. elegans-D. gunungcola* is also an excellent system for studying the molecular evolution of trait divergence in the context of sexual selection.

The goal of this project is to identify candidate genes and pathways that might underlie the divergence between *D. elegans* and *D. gunungcola* in mating display traits. In order to do so, I aim to find the genes responsible for the frontal wing mating display which is exclusively present in *D. elegans* and absent in its sister species *D. gunungcola*. I hypothesized that *D. elegans* and *D. gunungcola* have diverged in gene expression in the brain area, with genes responsible for wing display being differentially expressed in relation to *D. elegans*. These differences should be sexually dependent. Here, I compare the head transcriptomes of the mentioned species in order to find species-specific, stage-specific, and sexually-dependent differential gene expression signatures at the developmental stage and sex levels.

MATERIALS AND METHODS

A. *Drosophila* RNA-seq Data

Thirty-two RNA-seq libraries were downloaded from NCBI SRA archive under BioProject PRJNA837195. Paired-end 150 bp RNA-seq reads were obtained from the head of *D. elegans* (native to Hong Kong, China) and *D. gunungcola* (native to Sumatra Island, Indonesia) from the two sexes. The collection of head tissues by other authors was done from two life stages, one from young adult phase (0-3 hr after pupal stage, Day 0) and the other from sexually matured adults (5 days after pupal stage, Day 5).

B. RNA-seq data preprocessing, mapping, and quantification

Raw RNA-seq reads were stored in the High Performance Community Computing Cluster (HPC3). During the preprocessing stage, PCR duplicates, polyAT tails, and remaining uncertain nucleotide calls were removed, adapter sequences were trimmed, and low-quality reads shorter than 50 bps were filtered out using htstream (Settles et al. 2023). Next, the aligner STAR was employed for genome indexing and read mapping (Dobin et al., 2013). Filtered reads were aligned to the reference genomes of *D. elegans* and *D. gunungcola* with NCBI accession numbers GCF_018152505.1 and GCF_025200985.1, respectively. Processed sequencing reads were downloaded to a local workspace for subsequent analysis. Read counts were normalized across sexes and developmental stages for each species. Only genes with counts per million (CPM) values greater than 1 in at least 2 RNA-seq libraries in each sex and stage combination were considered to be expressed and used for subsequent downstream analyses.

C. Differential expression analysis

Sequencing reliability and reproducibility were examined using principal component analysis (PCA) and Pearson's correlation heatmaps. For the identification of differentially expressed (DE) genes, separate analyses were carried out for each species, considering Sex and Developmental Stage as categorical factors, and taking into account as well their interaction in a two-way ANOVA (Table 1). This analysis utilized the edgeR package version 3.42.4 (Robinson et al., 2010) and the limma package version 3.56.2 (Ritchie et al., 2015). To increase the rate of true positives while controlling for false negatives, a log2 fold-change threshold of 0.5 and a 5% false-discovery rate (FDR) were applied, respectively.

Table 1. Analytical design for the DE analysis performed on the head transcriptome data of each species.

	Sex

		Male	Female
Developmental Stage	Day 0	Dele_M_0_Head	Dele_F_0_Head
	Day 5	Dele_M_5_Head	Dele_F_5_Head

D. Functional enrichment analysis

Resulting lists of differentially expressed genes by sex and developmental stage in both *D. elegans* and *D. gunungcola* were examined for overrepresentation of functionally coherent patterns involving gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. This was done in clusterProfiler version 4.8.3 (Wu et al., 2021) with *D. melanogaster* GO terms and with the background gene list set corresponding to all unique genes with one-to-one ortholog between *D. elegans* and *D. gunungcola* (kindly provided by the authors that generated the sequencing data). To correct for multiple tests, a 5% FDR threshold and a 5% q-value threshold were applied (Benjamini & Hochberg, 1995), where the q-value controls the FDR tailored at individual tests in multiple test settings (Storey, 2002).

RESULTS

A. Transcriptome quality assessment

PCA and Pearson's correlation matrices were generated for each species, finding that the grouping of the data per sex and developmental stage was correct. Nevertheless, biological replicates formed distinct clusters per stage, and per sex for Day 5 but not for Day 0, this being the case both in *D. elegans* and *D. gunungcola* (Fig. 1). Lack of substantial separation at Day 0 is expected as sexual dimorphism is comparatively minimal compared to Day 5. Further, based on the criteria used, 11,224 and 11,307 genes were found to be expressed in *D. elegans* and *D. gunungcola,* respectively, and were subject to further downstream analysis.

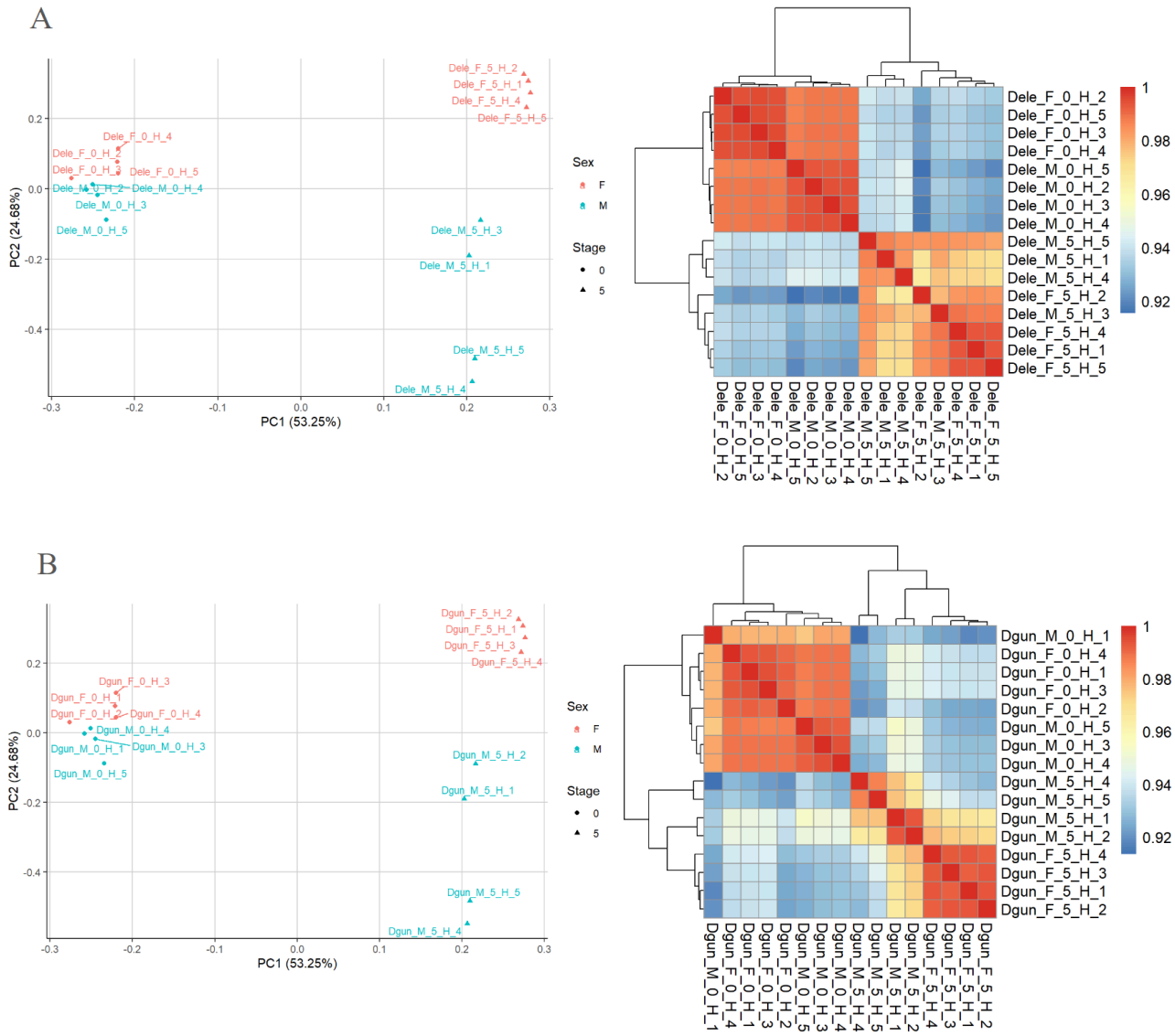

Figure 1. PCA and Pearson's correlation heatmap of RNA-seq head samples. A *D. elegans* B *D. gunungcola*. Left: PCA of 16 head RNA-seq outputs corresponding to biological replicates of different types of samples. Male, blue; female, red; day 0, circles; day 5, triangles. Right: Pearson correlation heatmap. Dele, *D. elegans*; Dgun, *D. gunungcola*; F: Female; M: Male; 0: day 0; 5: day 5; H: Head; 1-5, replicate id number.

B. Stage- and sex-pairwise expression differences in the head transcriptome

A two-way ANOVA was applied to the data of *D. elegans* and *D. gunungcola*, revealing notable intraspecific expression differences between the sexes and between developmental stages in the head transcriptome (Fig. 2). The number of DE genes in at least one of the contrasts performed, having 1-to-1 orthologous relationship between the species and an ortholog in *D.*

melanogaster, was 3115 and 3699 for *D. elegans* and *D. gunungcola,* respectively. Consistent with the expectation, sexual dimorphism in expression was lower at Day 0 compared to Day 5, whereas most differential expressions were found between stages within the same sex, involving in more cases downregulation in Day 5. A very limited number of genes showed significant developmental-stage sex-dependent differences; 0 in *D. elegans* and 164 genes in *D. gunungcola*.

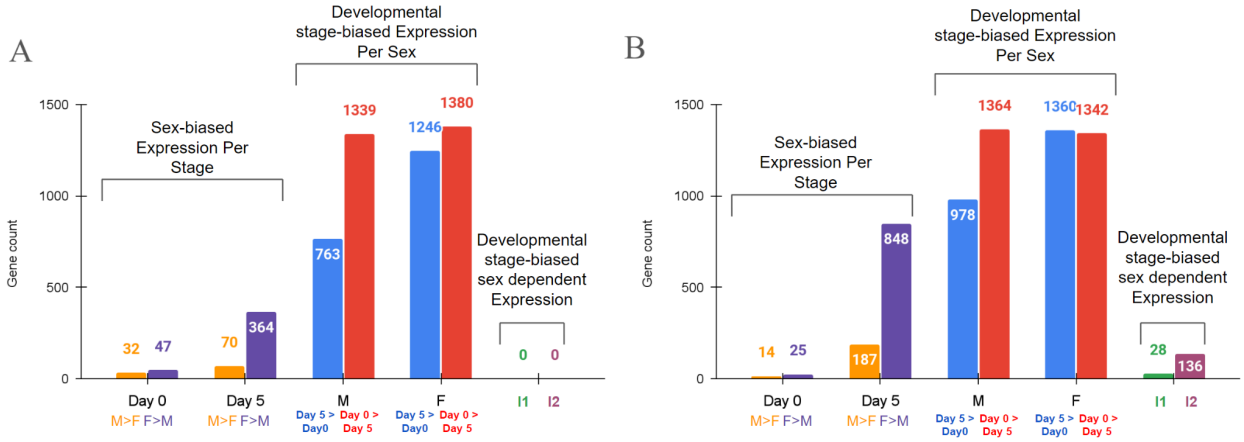

Figure 2. Differentially expressed genes by developmental stage and sex. A. *D. elegans*. B. *D. gunungcola*. The bar plot displays the results of two-way ANOVA with an interaction term to identify DE genes at a 5% FDR and FC >0.5 while controlling for stage differences or sex differences. M, male; F, female; I1, (M Day 5 > M Day 0, F Day 5 < F Day 0); I2, (M Day 5 < M Day 5, F Day 5 > F Day 0).

Between Day 5 and Day 0, 41.3% (1230/2979) and 54.7% (1900/3472) of the stage DE genes were found to be sex-biased in *D. elegans* and *D. gunungcola,* respectively (Fig. 3A), and 66.6% (2546/3823) and 67.5% (2174/3219) of the stage DE genes were unique to one of the two species in females and males, respectively (Fig. 3B). When examined from the sex differences in expression perspective, 78.9% (382/484) and 96.6% (944/977) of the sex-biased genes were found to be unique to Day 5, among which 86.4% (330/382) and 82.2% (776/944) were predominantly female-biased, a pattern that was consistent across species (Fig. 4A). In the context of the two different stages, 92.4% (85/92) and 85.0% (957/1126) of the DE genes were unique to one of the species in Day 0 and Day 5, respectively (Fig. 4B).

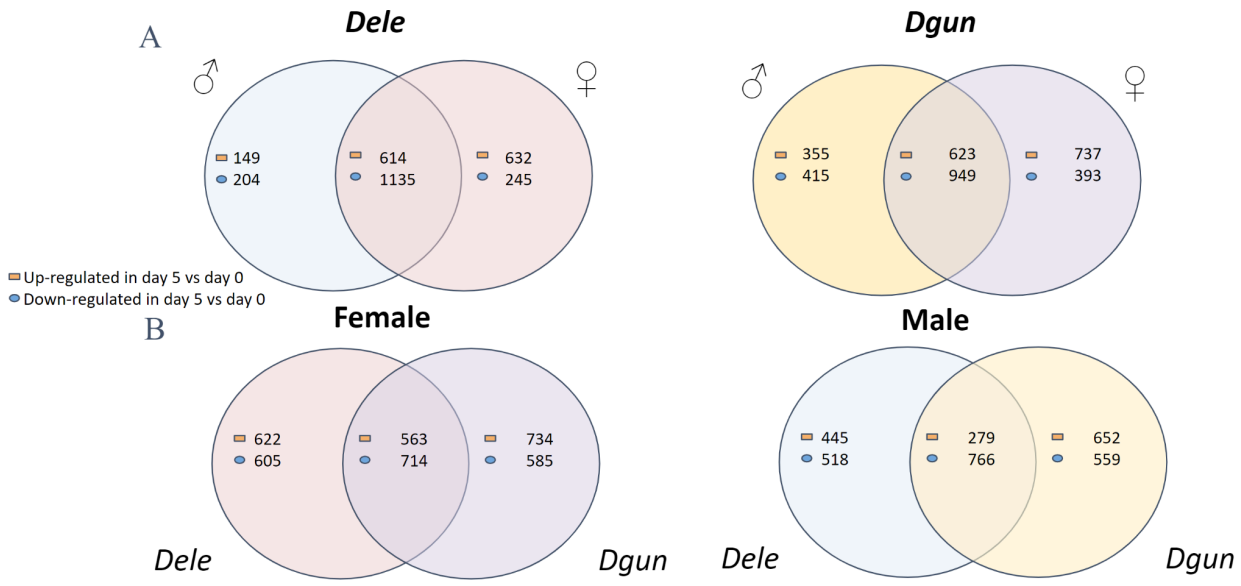

Figure 3. Venn diagrams for developmental-stage genes. A. Developmental stage-biased genes in males and females of *D. elegans* (left) and *D. gunungcola* (right). Intersection, stage DE genes in both sexes. B. Developmental stage-biased genes in females (left) and male (right) in the two species. Intersection, non-species specific genes. Dele, *D. elegans*; Dgun, *D. gunungcola*.

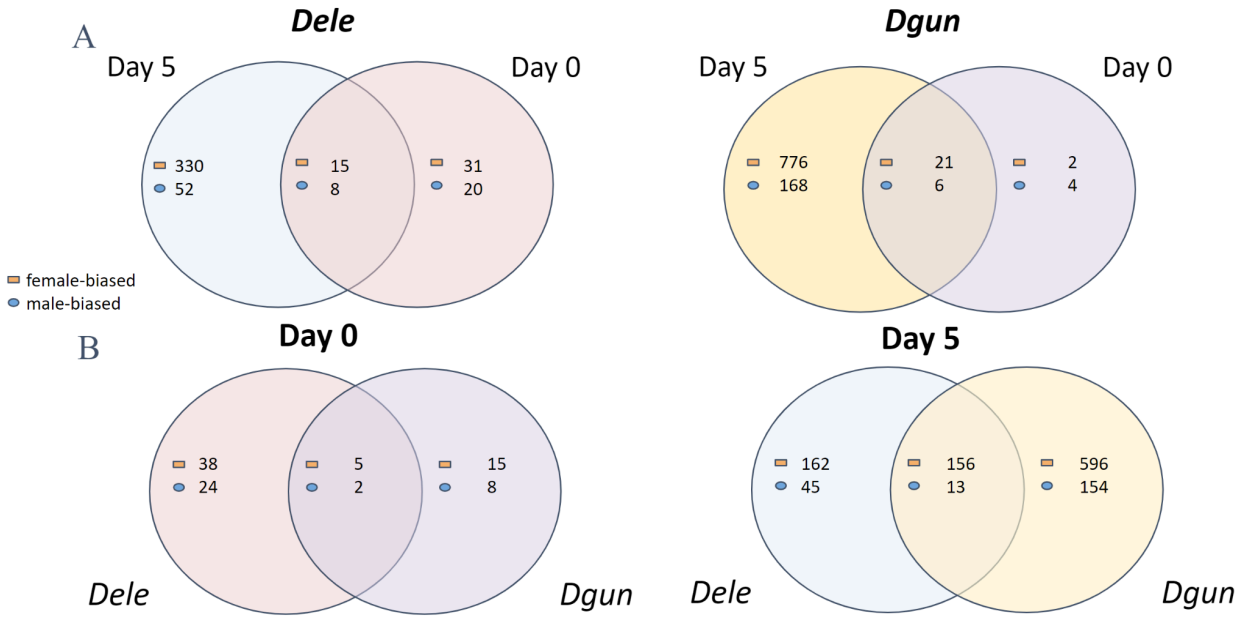

Figure 4. Venn diagrams for sex-biased genes in expression. A Sex-biased genes in particular developmental stages of *D. elegans* (left) and *D. gunungcola* (right). Intersection, sex-biased genes in both developmental stages. B Sex-biased genes in the two species on Day 0 (left) and Day 5 (right). Intersection, DE genes not unique to any of the species. Dele, *D. elegans*; Dgun, *D. gunungcola*.

C. Enriched functional patterns

I sought for biologically coherent patterns of functional enrichment among DE genes. I focused on DE genes between Day 5 and Day 0 in males as the phenotypic trait difference between species affects this sex at an age at which the males have reached sexual maturity. Up-regulated genes in Day 5 relative to Day 0 in males were found to be enriched for six biological processes (GO:BP) in *D. elegans* and *D. gunungcola* separately, with some being unique to one of the species (e.g. response to biotic stimulus in *D. elegans,* and G protein-coupled receptor signaling pathway in *D. gunungcola*). Other enriched functional terms were common to both species, e.g. neuropeptide signaling pathway (Figs. 5A & 5B). A similar picture was found for KEGG pathway terms (Fig. 5C & 5D).

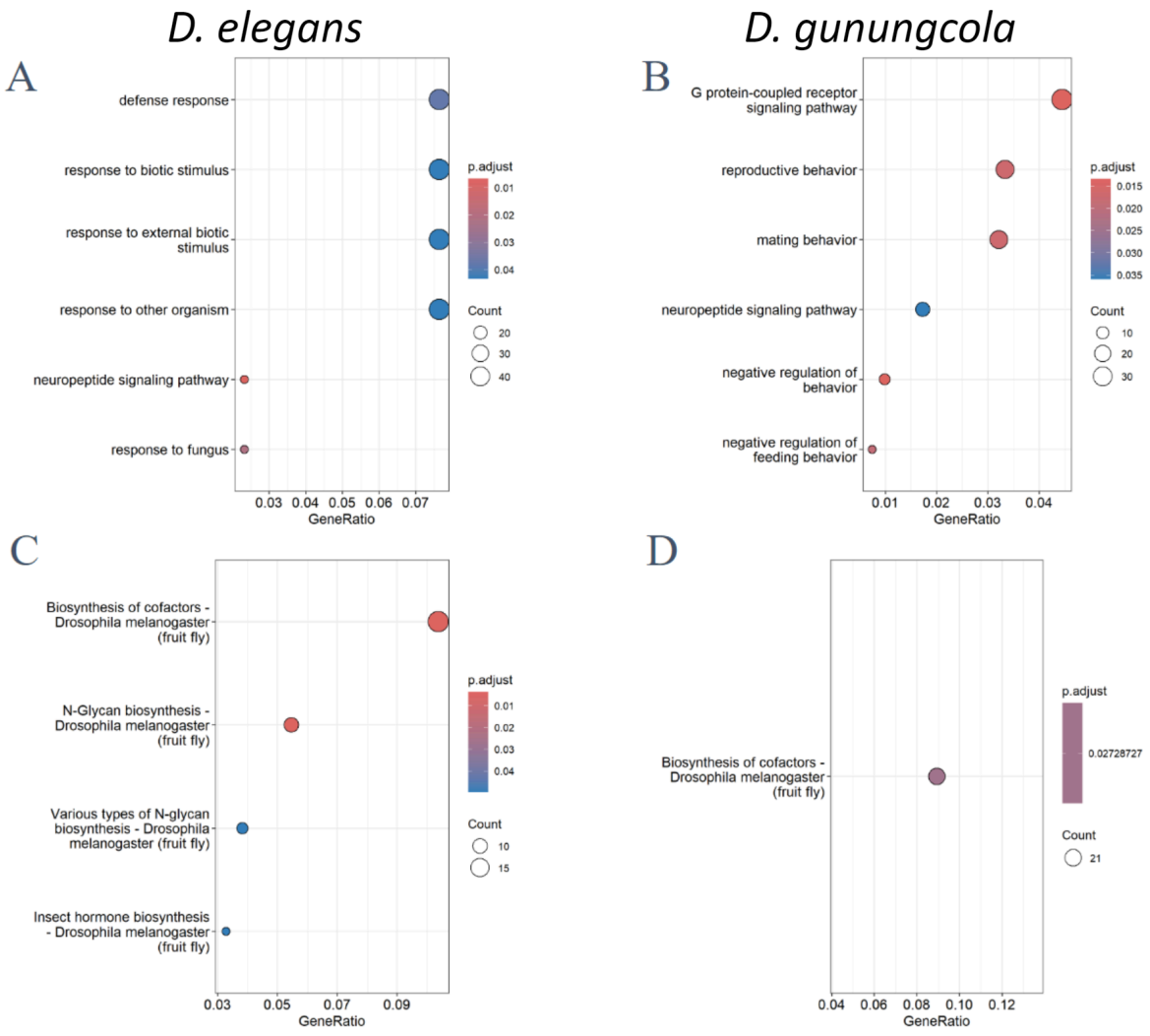

Figure 5. GO and KEGG analysis of up-regulated genes in males in Day 5 relative to Day 0. Bubble charts show significantly overrepresented GO:BP terms (A and B) and KEGG pathways (C and D) for genes up-regulated in Day 5 compared to Day 0 in males. A and C, *D. elegans*; B and D, *D. gunungcola*. Gene Ratio, proportion of DE genes relative to all genes known to be associated with a particular GO term or KEGG pathway. The level of statistical significance is indicated with a color-coded scale on the right.

Subsequently, in the context of DE genes between the sexes, I focused on male-biased genes in expression in Day 5 as the frontal wing display interspecific difference is restricted to fully sexually mature males. These genes were found to be uniquely enriched for eight and thirty-nine GO:BP terms in *D. elegans* and *D. gunungcola,* respectively (Figs. 6A & 6B). No KEGG pathway term was found in *D. elegans* whereas two terms were found significant in *D. gunungcola* (Fig. 6C).

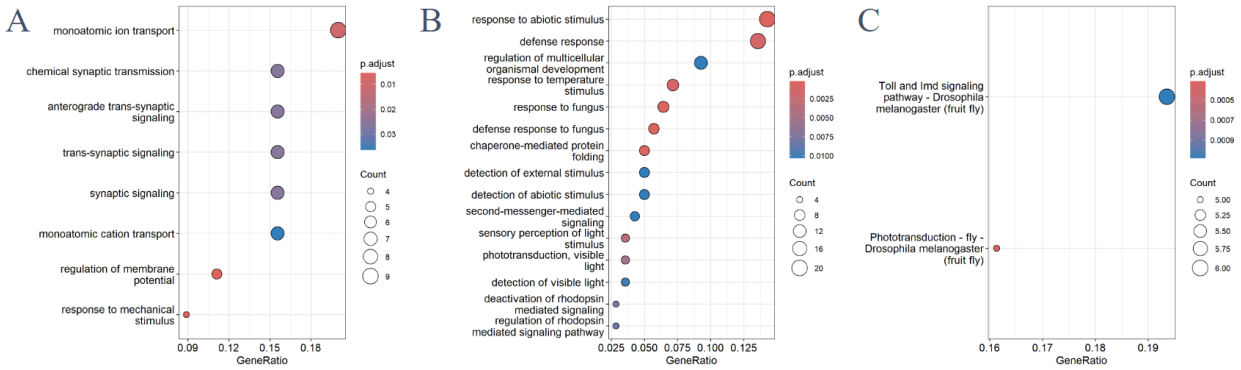

Figure 6. Gene Ontology (GO) analysis of male-biased genes on Day 5. Bubble charts show significantly overrepresented GO:BP terms (A and B) and KEGG pathways (C) for Day 5 male-biased genes. A, *D. elegans*; B and C, *D. gunungcola*. Gene Ratio, proportion of DE genes relative to all genes known to be associated with a particular GO term or KEGG pathway. The level of statistical significance is indicated with a color-coded scale on the right.

DISCUSSION

The differential gene expression analysis performed on RNAseq data from heads revealed the dynamic nature of head transcriptome between the sister species *D. elegans* and *D. gunungcola*, finding differences associated with developmental stage and sex. Developmental-stage expression differences within the same sex appear to be more prevalent than expression differences between the sexes within the same developmental stage (95.6% or

2979/3115, and 93.9% or 3472/3699, in *D. elegans* and *D. gunungcola,* respectively). Further, two general patterns of differential expression become apparent. First, as the individuals of both species grow older, downregulation prevails over upregulation. Second, there is an increase in the number of sex-biased expressed genes on Day 5, with a larger proportion of them being female-biased, both in *D. elegans* and *D. gunungcola*.

Some notable functional patterns of enrichment among DE genes have been documented. In the context of developmental stage differences, the overrepresentation in Day 5 upregulated genes in *D. elegans* males for the GO:BP term response to biotic stimulus agrees well with our expectation as the *D. elegans* males' frontal wing display behavior is one of the responses to the presence of females during mating. The neuropeptide signaling pathway cluster has been indicated to be associated with the regulation of genes for prolonging mating duration in *D. melanogaster* (Kim et al., 2013), therefore the enrichment for this GO term in *D. elegans* is particularly interesting given its function. Nevertheless, this GO:BP term, in combination with those relative to reproductive and mating behavior, is also found in Day 5 males of *D. gunungcola*, which do not feature frontal wing display behavior. One potential explanation is that though upregulated in *D. elegans* for mating and reproduction, their orthologs in *D. gunungcola* have experienced divergent evolution and serve modified or different functions.

Interestingly, when considering sex-biased gene expression, I found a significantly overrepresented GO:BP term in Day 5 male-biased expressed genes in *D. elegans* related to synaptic transmission and signaling. This enrichment pattern could be potentially associated with the unique frontal wing display behavior in the males of this species. Synaptic transmission and signaling are crucial for the proper functioning of neural circuits involved in courtship behavior,

including mate recognition, decision-making processes, and interestingly courtship rituals. Frontal wing display is in fact considered ritual specific to *D. elegans* (Shen et al., 2023).

This work has enabled the generation of a first portrait of the expression differentiation between two sister species that differ in their courtship behavior, including lists of candidate genes to be targeted in future silencing studies. This will clarify how those candidate genes contribute to the sex-dependent frontal wing display behavior in *D. elegans*, which is absent in *D. gunungcola*. Further, considering that *D. elegans* males are also characterized by wing pigmentation, comparative expression analyses for this body part between *D. elegans* and *D. gunungcola* should facilitate the study of the genetic basis of possible coevolution between both traits.

AUTHOR'S CONTRIBUTIONS

Under the supervision of Prof. José Ranz (lab PI) and Ashlyn Kimura (graduate student), I performed the data preprocessing, quality control, differential gene expression analysis, and functional enrichment analysis of *D. elegans and D. gunungcola* RNA-seq data.

ACKNOWLEDGMENTS

I would like to express my appreciation to Dr. Ranz and Ashlyn Kimura for their guidance and support in the completion of this study. I am also grateful to Hailin Liang and Tina Soroudi for their feedback on the writing of this paper. Lastly, I acknowledge the contributions of my family and partner Junping Luo for their unwavering encouragement and emotional support.

LITERATURE CITED

Baker BS, Taylor BJ, Hall JC. 2001. Are complex behaviors specified by dedicated regulatory genes? reasoning from drosophila. Cell 105:13–24.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) 57:289–300.

Jacob F. 1977. Evolution and tinkering. Science 196:1161–1166.

Kaźmierczak M, Nicola SM. 2022. The arousal-motor hypothesis of dopamine function: Evidence that dopamine facilitates reward seeking in part by maintaining arousal. Neuroscience 499:64–103.

Kim WJ, Jan LY, Jan YN. 2013. A PDF/NPF neuropeptide signaling circuitry of male drosophila

melanogaster controls rival-induced prolonged mating. Neuron 80:1190–1205.
Koganezawa M, Kimura K, Yamamoto D. 2016. The neural circuitry that functions as a switch for courtship versus aggression in drosophila males. Current Biology 26:1395–1403.
Koshikawa S. 2020. Evolution of wing pigmentation in *drosophila*: Diversity, physiological regulation, and *cis*‐regulatory evolution. Development, Growth & Differentiation 62:269–278.
Massey JH, Li J, Stern DL, Wittkopp PJ. 2021. Distinct genetic architectures underlie divergent thorax, leg, and wing pigmentation between drosophila elegans and D. Gunungcola. Heredity 127:467–474.
Massey JH, Rice GR, Firdaus AS, Chen C, Yeh S, Stern DL, Wittkopp PJ. 2020. Co‐evolving wing spots and mating displays are genetically separable traits in *drosophila*. Evolution 74:1098–1111.
Merilä J, Sheldon BC. 2001. Avian quantitative genetics. Current Ornithology 16:179–255.
Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, Vilo J. 2019. G:profiler: A web server for functional enrichment analysis and conversions of Gene Lists (2019 update). Nucleic Acids Research 47.
Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. Limma powers differential expression analyses for RNA-sequencing and Microarray Studies. Nucleic Acids Research 43.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edger: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.
Settles, M., Streett, D., Angell, J., Hunter, S., & Hoogstrate, Y. (2023). HTStream. GitHub repository. https://github.com/s4hts/HTStream
Shen P, Wan X, Wu F, Shi K, Li J, Gao H, Zhao L, Zhou C. 2023. Neural circuit mechanisms linking courtship and reward in drosophila males. Current Biology 33.
Storey JD. 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society Series B: Statistical Methodology 64:479–498.
Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. 2019. String V11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Research 47:D607–D613.
Takahashi A. 2013. Pigmentation and behavior: Potential association through pleiotropic genes in drosophila. Genes & Genetic Systems 88:165–174.
Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. 2021. Clusterprofiler 4.0: A universal enrichment tool for interpreting OMICS data. The Innovation 2:100141.