communications biology

Article

https://doi.org/10.1038/s42003-024-07255-6

cis- and *trans*-regulatory contributions to a hierarchy of factors influencing gene expression variation

Check for updates

Siddhant Kalra, Stephen Lanno, Guadalupe Sanchez & Joseph D. Coolon

Gene expression variation results from numerous sources including genetic, environmental, life stage, and even the environment experienced by previous generations. While the importance of each has been demonstrated in diverse organisms, their relative contributions remain understudied because few investigations have simultaneously determined each within a single experiment. Here we quantified genome-wide gene expression traits in *Drosophila*, quantified the contribution of multiple different sources of trait variation and determined the molecular mechanisms underlying observed variation. Our results show that there is a clear hierarchy in our data with genome and developmental stage contributing on average considerably more than current and finally previous generation environmental effects. We also determined the role of *cis* and *trans*-regulatory changes across different sources of trait variation, highlighting their importance in adaptation and environmental responses and showing unexpectedly that transgenerational effects herein were predominantly associated with changes in *trans*-regulation.

Gene expression is the process by which information stored in DNA is first transcribed into an mRNA that can then be translated into proteins with diverse functions that generate all the traits and trait variations we see across the entire tree of life. Genome-wide gene expression levels are molecular traits that, like organismal traits, vary as a consequence of numerous intrinsic and extrinsic factors. Regulation of gene expression is controlled by an intricate network of biochemical interactions between *cis*-acting[1–4] DNA sequences near a regulated gene (e.g., transcription factor binding sites) and *trans*-acting[5,6] regulatory factors encoded elsewhere in the genome (e.g., transcription factor proteins) that bind *cis*-acting sites and either activate or repress gene expression. These molecular traits govern the larger-order organism-level traits[7,8] we typically associate with a determination of form, function, and ultimately organismal fitness[4,9,10].

Intrinsic and extrinsic factors that are commonly associated with substantial contributions to variation in traits like gene expression levels are an organism's genome sequence[11–14], the cell type/developmental stage/sex considered[15–18], and the organism's environment including both current conditions[19–21] as well as transgenerational effects of environmental conditions experienced by previous generations[14,22–27] and increasingly complex interactions among all these different sources[28–33]. While a multitude of studies have demonstrated diverse and abundant effects of each in a wide variety of organisms with focus on a multitude of different traits, the relative contribution of these different sources of variation remains understudied. This has largely resulted from a paucity of studies that simultaneously determine the contributions of multiple sources of variation within a single fully factorial experiment such that their relative contributions can be properly compared. Most studies in this area have focused on a single or few such sources of variation with fewer studies published the more treatments that are simultaneously examined within a given study[8,11,34–38]. This is especially true for transgenerational effects, which are generally understudied, remain poorly understood, and are not frequently included in larger multi-treatment experiments.

The evolution of sequencing technologies provides the ability to perform large-scale multi-omics studies that incorporate a wide range of biological and technical conditions influencing gene expression[39]. However, it simultaneously presents the challenge of understanding the intricate relationships and interactions among different sources of variation and their influence on molecular architecture. This is crucial for quantifying their contributions to influence gene expression. To address these complexities, a thorough understanding of the patterns of transcriptional variance across genes, their context-specific regulation, and their functional implications is necessary along with the mechanisms responsible for driving these patterns and their respective magnitudes. The interaction of these processes in shaping gene expression variance remains an open question.

To begin to fill these knowledge gaps we first determined the relative hierarchy of contributions to genome-wide gene expression traits in *Drosophila* with RNA-seq focused on investigating the roles of species differences (*D. simulans* vs. *D. sechellia*, which diverged approximately 0.2

Wesleyan University, Middletown, CT, USA. e-mail: jcoolon@wesleyan.edu

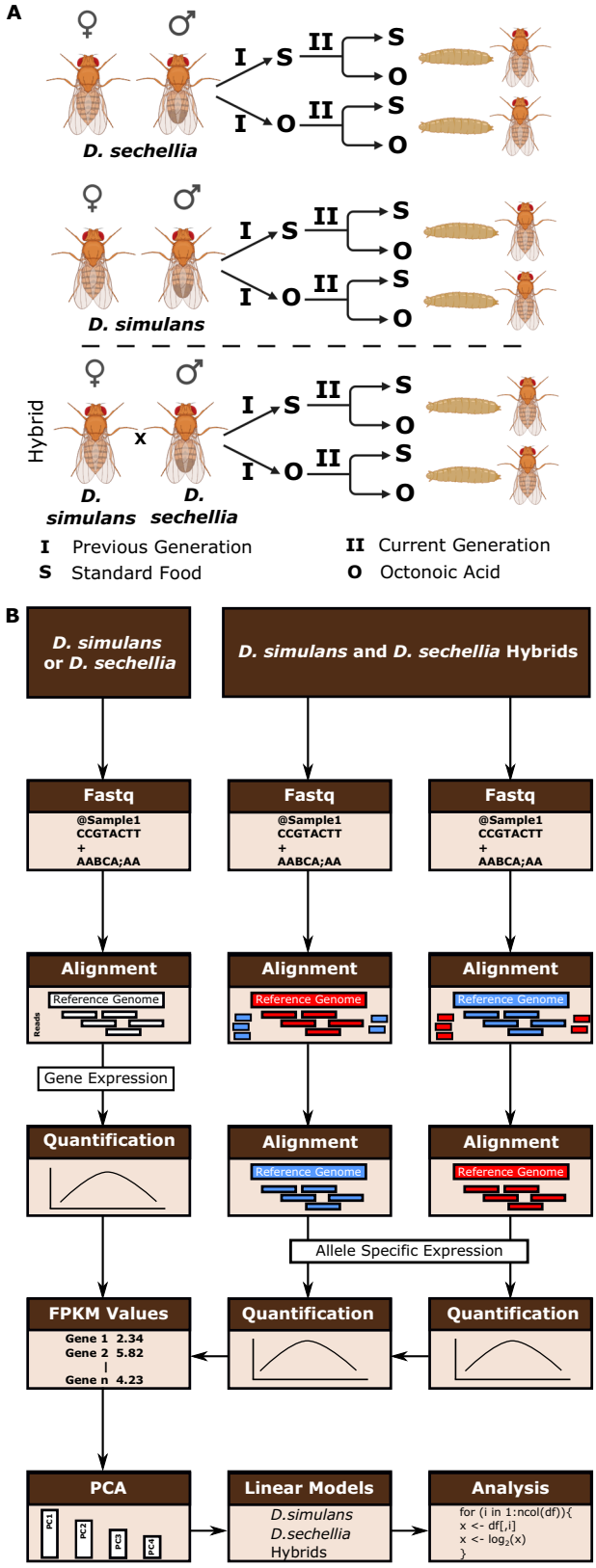


Fig. 1 | Experimental design. A Experimental setup to produce the samples ($n = 3$ biologically independent samples of 10 pooled individuals) for a fully factorial design testing effects of stage (larvae vs. adult), species (*D. sechellia* vs. *D. simulans*), previous generation environment (S standard food vs O octanoic acid food), and current generation environment (S vs. O) on genome-wide gene expression. F_1 hybrids were also included in sample collection for identification of regulatory mechanisms underlying observed differences between species (see Methods). **B** Bioinformatics workflow used is shown from the generation of FASTQ files, sequence alignment, gene and allele-specific expression (ASE) quantification and downstream analysis. The larvae and adult fly icons shown in this figure were obtained from Biorender (https://www.biorender.com/).

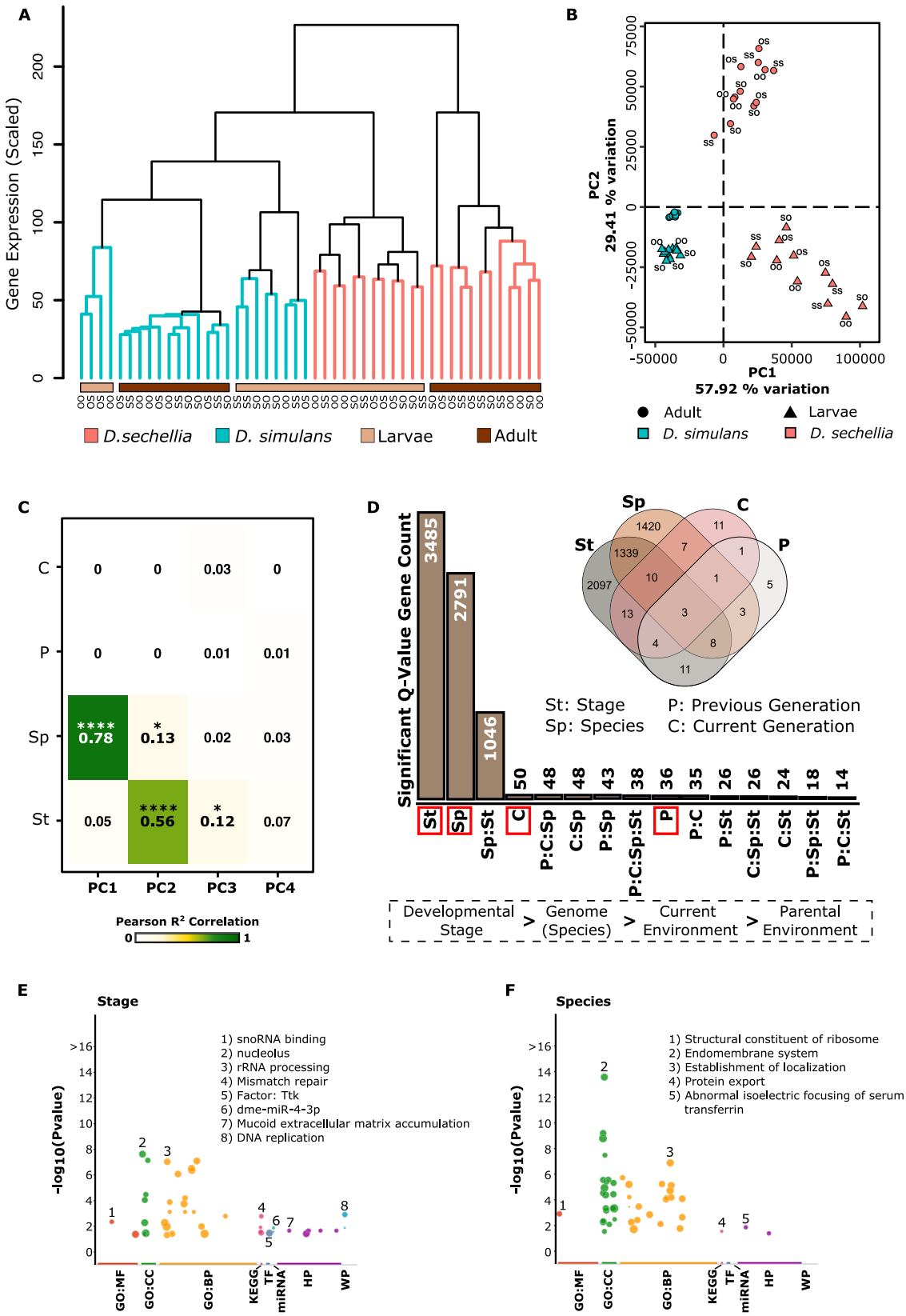
MYA[40,41] and differences observed are interpreted herein as genomic differences), age/developmental stage (L3 larvae vs adult), environment (standard food vs octanoic acid (OA) containing food, a relevant ecological difference between these species[42–46]), and the previous generation's environment (standard food vs OA food) in gene expression variation. Second, we followed this with analysis of genome-wide allele-specific expression analysis in F_1 hybrids between *D. simulans* and *D. sechellia* to disentangle the molecular mechanisms (*cis* and *trans*-regulatory contributions) underlying gene expression trait variation[13,14,23,47,48] (Fig. 1A, B) (Supplementary Table 1). Species differences identified herein could include epigenetic effects but they likely play a minimal role due to long-term lab culture in a common environment. Finally, we chose the environments included in this study (OA vs control media) because *D. sechellia* and *D. simulans* differ greatly in their ecology and we hoped to identify candidate genes whose expression level(s) are critical for *D. sechellia* specific adaptations to life feeding on a novel host-plant. *D. sechellia* is an island endemic species restricted to the Seychelles Islands in the Indian Ocean that has evolved to be a host-plant feeding specialist on the toxic fruit of *Morinda citrifolia* that is abundant on this archipelago[42,43,49,50]. The toxicity of *M. citrifolia* is primarily attributed to its production of high levels of the volatile medium-chain fatty acid OA[51]. *D. sechellia* has evolved resistance to the toxic effects of OA while sister species *D. simulans* and other *Drosophila* species are sensitive to OA toxicity[45,46]. Despite this sensitivity, both *D. sechellia* and *D. simulans* are found on *M. citrifolia* fruit on the Seychelles Islands making this an ecologically relevant environmental condition for these species[40].

Understanding the different impacts of tested factors is important because it allows understanding at numerous levels of how gene expression works and relates to organismal traits and trait variance. We set out to identify patterns in the data (e.g., a hierarchy) that will help in making predictions about other organism's responses to similar treatments in the future, allow for contrasting different factors based on their impacts, identify any genes that are hubs in response pathways or common to all treatment responses, and begin to understand the dynamic regulatory network underlying organismal traits big and small. Additionally, information from main effect responses and their interactions gives clues for how nature fine-tunes gene expression in a context-specific (e.g., the environment) manner to increase fitness, and understanding gene expression responses genome-wide may help in making predictions about the consequences of environmental change among other ecological scenarios. The interactions specifically, which are typically absent from lower complexity experimental designs, aid in understanding how systems work in the wild where there are numerous factors interacting that an organism experiences in combination. Finally, using a hierarchical inference combined with quantification of *cis*- and *trans*-regulatory changes allows for testing if factors with differing degrees of impact on genome-wide gene expression are regulated in different ways.

Results and discussion

Determining the relative impact of different factors on genome-wide gene expression

To begin to uncover the hierarchy of effects for the different factors in this experiment, we used hierarchical clustering analysis of genome-wide gene expression. This analysis of structural information revealed important aspects of the pattern of variation in the data. We found distinct grouping of treatments first based on species, followed by developmental stage, and little grouping was observed associated with current or previous environmental effects (Fig. 2A). Next, to identify the amount of variation explained by each



factor in our analysis, we used principal component analysis (PCA)[38] and with the elbow method we found that 4 PCs were sufficient to understand the majority of the variation in our dataset (PC1 explained 57.92%, PC2 29.41%, PC3 4.49%, and PC4 2.4% of the variance, Fig. 2B and Supplementary Fig. 1A). We then correlated these PCs with our tested factors: species, developmental stage, current generation environment, and previous generation environment (Figs. 1A, 2C). We found that PC1 primarily reflects species/genome differences ($R^2 = 0.78$, $P \leq 0.0001$), PC2 primarily reflects differences between developmental stages ($R^2 = 0.56$, $P \leq 0.0001$) but also to a lesser degree by species/ genomic differences ($R^2 = 0.13$, $P \leq 0.05$),

Fig. 2 | Bioinformatics workflow and hierarchy of factor effects. A The hierarchical clustering plot for the parental FPKM values. **B** PCA of variation in the gene expression data generated for this study. Species are color-coded (*D. simulans* blue, *D. sechellia* red), stage indicated by point shape (triangle = larvae, circle = adult), and the two letter label on each point is: first letter for the previous environment (S or O) and second letter for current environment (S or O). **C** Plot illustrating Pearson correlation coefficients (R^2) for well-supported principal components (PCs) and tested factors: stage, species, previous generation environment, and current generation environment. $^{**}p < 0.0001$ $^{*}p < 0.05$). **D** Bar graph showing number of DEGs identified for each main effect (indicated in red) as well as interactions among main effects with inset Venn diagram showing overlap among DEGs for each main effect (St stage, Sp species, P previous generation environment, C current generation environment). Below is the hierarchy observed for the number of DEGs for the main effects. **E** Gene ontology plot for 2097 genes exclusive to the stage effect. The top entry for each category is labeled in the plot. **F** Gene ontology plot for 1420 genes exclusive to the species effect. The top entry for each category is labeled in the plot.

PC3 primarily reflects differences between developmental stages ($R^2 = 0.12$, $P \leq 0.05$), and PC4 did not have significant correlation with any of our tested factors (Fig. 2B, C). These data strongly suggest that there is a hierarchy of importance among the factors tested with species/genomic differences and differences between developmental stages having considerably greater effects on gene expression variation than either the current or previous generation environment experienced in this experiment. The results of the PCA are, in general, in agreement with the results of hierarchical clustering (Fig. 2A) strengthening this result.

Gene-specific analysis of transcriptomic variation

The PCA provided an unbiased whole-transcriptome measure of variation in our data and allowed the identification of treatments correlated with observed variation but could potentially be influenced by a small number of genes with large effects. Therefore, as an additional complementary approach, we next performed gene-specific multivariate linear regression to identify genes that varied significantly by treatment(s) or treatment interactions (Fig. 2D). When investigating the main effects in our model, we found that the largest number of differentially expressed genes (DEGs) were caused by differences in life stage (3485) followed by species (2791) then current environment (50) and finally previous environment (36) (Supplementary Table 2). Because we had a full factorial design, we could also identify genes with significant interactions among main effects. We observed that the interactions followed the same general hierarchy as the main effects, further strengthening confidence in our results (Fig. 2D). These data indicated that, in this system and with this experimental design, substantially more DEGs were caused by genomic differences between species and regulatory differences between life stages than either current or previous generation environmental plasticity. While the specific genes altered and the quantity of DEGs identified were certainly the consequence of the specific genotypes, life stages, and environments chosen; the data suggest this general pattern in the observed hierarchy of effects would remain regardless of which environments were chosen based on the abundance of the differences observed. Furthermore, we chose these environments as they represent ecologically relevant conditions that these species would encounter in their shared geographic range and, as such, should provide an estimate of environmental effects that are actually experienced by these species in the wild[52,53].

Functional enrichment of identified differentially expressed genes

To determine the functional significance of DEGs for the different main effects, we used Gene Ontology (GO) term enrichment analysis. For the 2097 genes exclusive to developmental stage effects (Fig. 2D) we found several prominent biological processes, including ribosome biogenesis (GO:0042254), rRNA processing (GO:0006364), ncRNA processing (GO:0034470), and eggshell formation (GO:0030703) (Fig. 2E and Supplementary Table 3) were significantly enriched. Additionally, snoRNA binding (GO:0030515) and catalytic activity (GO:0140640) emerge as key molecular functions (Fig. 2E and Supplementary Table 3) that were enriched in the developmental stage DEGs. GO term enrichment analysis also identified cellular components, including the nucleolus (GO:0005730), egg chorion (GO:0042600), and small subunit processome (GO:0032040) (Fig. 2E and Supplementary Table 3) as significantly enriched among developmental stage DEGs. The complete list of significantly enriched GO terms and pathways for genes with developmental stage-specific effects is provided in Supplementary Table 3.

Similarly, among the 1420 DEGs exclusive to species effects (Fig. 2D), our analysis identified biological processes GO terms including the establishment of localization (GO:0051234), cytoplasmic translation (GO:0002181), and endoplasmic reticulum to Golgi vesicle-mediated transport (GO:0006888), among others (Fig. 2F and Supplementary Table 4) as significantly enriched in this set of genes. Notably, the only significantly enriched molecular function GO term identified was the structural constituent of the ribosome (GO:0003735) (Fig. 2F and Supplementary Table 4) while endomembrane system (GO:0012505), cytoplasm (GO:0005737), Golgi apparatus (GO:0005794), and more emerge as the key cellular components (Fig. 2F and Supplementary Table 4) that were enriched among the DEGs identified between species. Identification of endomembrane system as enriched in DEGs between species is consistent with our prior work identifying genes in the Osiris family with known involvement in endomembrane system function as important contributors to differences between *D. sechellia* and *D. simulans* resistance to OA toxicity[46]. The complete list of significantly enriched GO terms and pathways for DEGs with species effects is provided in Supplementary Table 4. For the 11 DEGs genes specific to current environment plasticity, we only found the structural constituent of chitin-based cuticle (GO:0005214) as significantly enriched in molecular function GO terms while no significant enrichment was found for biological processes and cellular components for this effect. This enrichment is consistent with proper work exposing *D. sechellia* to much higher levels of OA where similar DEGs involved in chitin-based cuticle formation were also identified[45]. We did not find any significant enrichment for the main effects of trans-generational plasticity, likely due to the low (only 5) number of DEGs identified specific to this category. Additionally, we also performed the enrichment analysis not only on the exclusive gene list but also on the overall categories (stage, species, current, and previous) and found similar results which are provided in Supplementary Table 5.

cis and *trans*-regulatory changes associated differently with different sources of variation

To disentangle the molecular mechanisms underlying the observed regulatory changes we created F_1 hybrids between female *D. simulans* and male *D. sechellia*, performed allele-specific RNA-seq with our custom bioinformatics approach[13,14,27,54] and classified changes into those caused by *cis*-acting (having allele-specific effects), *trans*-acting, or a combination of both types of change including (1) compensating: *cis*- and *trans*-acting changes opposing each other or (2) enhancing: *cis*- and *trans*-acting changes acting to increase differences (Supplementary Table 6). Overall, we found that *cis*-acting changes were an order of magnitude more abundant than *trans*-acting changes, consistent with prior studies documenting relationships between regulatory change and evolutionary divergence between species[14,23,47,48,54,55] (Fig. 3A–E and Supplementary Fig. 1B, C–F, K–N, O–R, W–Z).

Furthermore, compensating changes, those with multiple underlying molecular changes acting in opposition to each other to reduce differences between species, were the most abundant type of regulatory change, suggesting that stabilizing selection is the primary driver of regulatory evolution in this system. Interestingly, this pattern was more pronounced in adults (and enhancing regulatory changes were more abundant in larvae) (Fig. 3B, C), inconsistent with prior studies showing that buffering is more pronounced early in development but reduced in later life stages[56]. However,

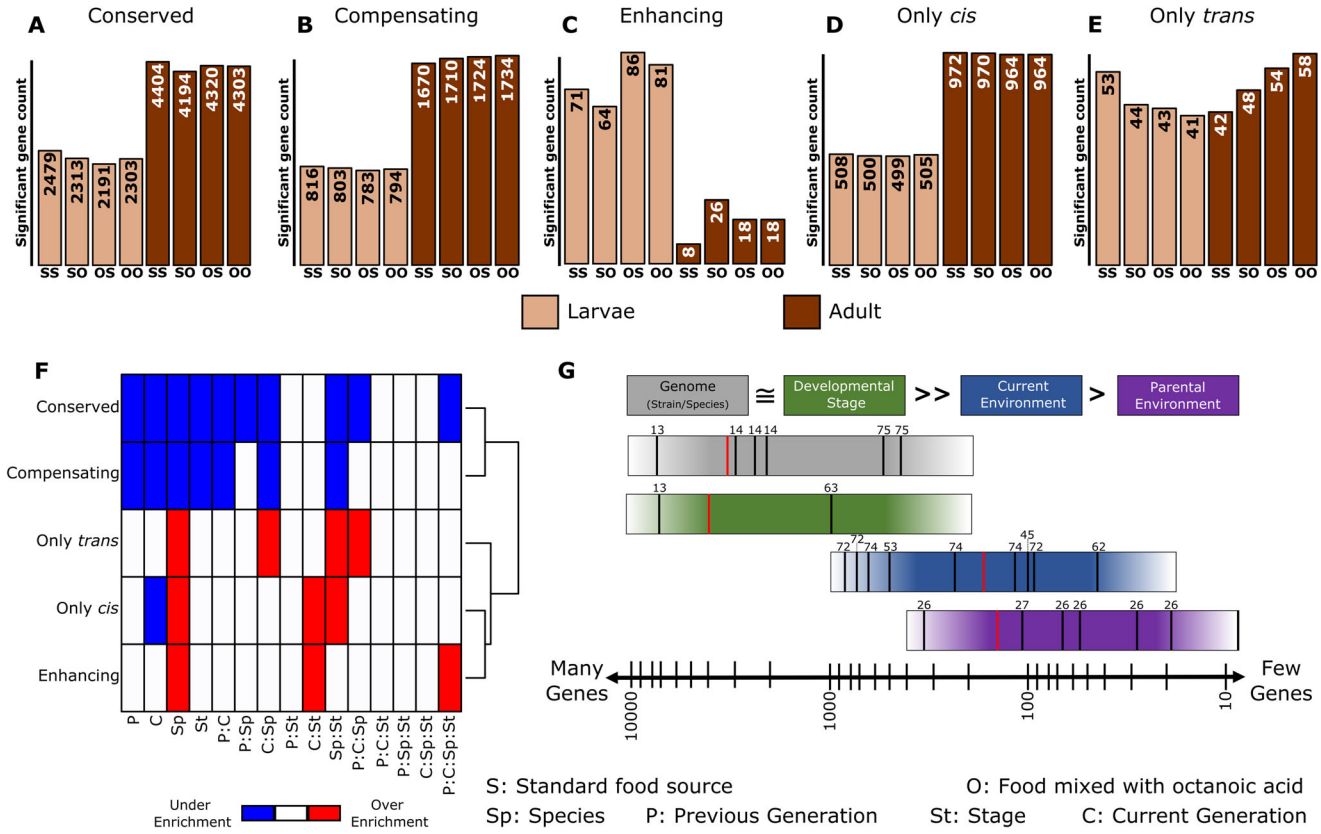


Fig. 3 | Regulatory mechanisms underlying observed differences in gene expression. A–E Barplots depicting the number of DEGs with **A** conserved, **B** Compensating, **C** enhancing, **D** only *cis*, and **E** only *trans*-regulatory mechanism classification (see methods). The two letter x-axis labels are: first letter for previous environment (S or O) and second letter for current environment (S or O). **F** Heatmap for the degree of overlap of genes between regulatory mechanism classifications (as in **A**–**E**) and DEGs for main effects and all possible interactions among them (as in Fig. 1E bar plot). Red colors indicate statistically significant overenrichment (more than expected), while blue colors indicate significant underenrichment (less than expected), and white indicates non-significant relationships. **G** Diagram depicting a generalized hierarchy of the magnitude of different sources of variation in gene expression and proposed variability of each from this work (red lines) and examples from published studies performed by our group using the same methodology such that studies are maximally comparable (black lines)[13,14,26,27,45,53,62,72–75].

we did find that *cis*-regulatory changes were more abundant in adults than larvae (Fig. 3D and Supplementary Fig. 1K–N, W–Z) which is consistent with such predictions, and we found no relationship between genes with *trans*-only regulatory change and developmental stage suggesting such changes might be less context-specific or tunable and therefore more visible phenotypically and therefore vulnerable to loss via natural selection when deleterious.

Stabilizing selection can maintain phenotypes over evolutionary time in one of two general ways: (i) mutations altering a phenotype can be removed from the population by purifying selection resulting in conserved DNA sequence and molecular mechanisms over evolutionary time, or (ii) compensating mutations can fix within a population, resulting in conserved phenotypes despite changes in the underlying DNA sequence and/or molecular mechanisms. Our findings (Fig. 3B) and those of other recent studies that compared gene expression among strains and/or species of diverse organisms suggested that compensating changes underlie the widespread pattern of stabilizing selection observed for mRNA abundance traits[47,48,57–61]. This phenotypic stasis may, paradoxically, enable the evolution of novelty by maintaining function while allowing cryptic genetic variation to accumulate with the potential to contribute to future adaptation as well as interact with other sources of variation in gene expression in unexpected ways (e.g., the environment).

To test this prediction and more generally identify the molecular mechanisms contributing to gene expression variation in our experiment, we identified the types of regulatory changes enriched among DEGs from the different effects tested (e.g., species, stage, current, previous environment, Fig. 3F and Supplementary Fig. 1G–J, S–V). Genes with species differences were enriched for *cis*-regulatory changes, *trans*-regulatory changes, and enhancing changes while depleted for conserved and compensating changes as expected (this is a condition for the categorization performed, see Supplementary Table 7 and Supplementary Fig. 1G–J, S–V), serving as an important quality control check. We further found that genes with compensating evolutionary changes in gene regulation between species were significantly underrepresented among genes responsive to either current or previous generation environments, differences in life stage, the interaction between previous and current environmental effects, the interaction between species and current environment and the interaction between species and life stage (Fig. 3F). This strongly suggests that the cryptic genetic variation underlying the observed compensating regulatory changes is not contributing to excess variation that is unlocked by exposure to different environments or the other sources of variation tested herein.

Interestingly, we found that genes with plastic responses to the current environment were depleted for *cis*-regulatory changes but genes with stage-specific current-environment plasticity were enriched with *cis*-regulatory changes (Fig. 3F and Supplementary Fig. 1B). Genes with enhancing regulatory changes were also enriched for stage-specific current-environment plasticity suggesting directional selection may be acting to promote stage-specific current-environment responses and these may represent interesting candidates for roles in environment-specific adaptation(s). We also found that genes with species-specific current-environment plasticity and species-specific environmental transgenerational plasticity were enriched for *trans*-regulatory changes (Fig. 3F). Finally, we found that genes with species-specific life stage effects were enriched for both *cis*-regulatory and *trans*-regulatory changes (Fig. 3F). These data suggest that the transgenerational

plasticity found in our experiment was caused by *trans*-acting changes like inheritance of mRNA or proteins in the egg rather than methylation at the gene itself (which would act allele-specifically and be classified as a *cis*-regulatory change). One caveat of the experimental design used herein is that we only made one cross to make F_1 hybrids, this was between female *D. simulans* and male *D. sechellia*. Without reciprocal crosses, we cannot be sure about the presence or scope of possible imprinting or other parent-of-origin epigenetic effects, but our prior work showed that there was no evidence supporting imprinting in female *D. melanogaster* with normal karyotypes[27] suggesting that this should not be a major concern given we worked with females only and would likely only affect few genes if present.

Functional enrichment differed among different regulatory change categories

To further investigate the functional enrichment of genes identified with different classifications of regulatory change (conserved, compensating, enhancing, only *cis*, and only *trans*), we performed a GO term enrichment analysis (Supplementary Table 8). Specifically, for the genes categorized as conserved (Fig. 3A), our analysis revealed significant enrichment in several biological processes, including developmental process (GO:0032502), anatomical structure development (GO:0048856), and cell differentiation (GO:0030154). Regarding molecular functions, these conserved genes exhibited significant enrichment in transcription regulator activity (GO:0140110), DNA-binding transcription factor activity (GO:0003700), and transcription regulatory region nucleic acid binding (GO:0001067). Additionally, analysis of cellular component GO terms indicated significant enrichment for the nucleus (GO:0005634) and the transcription regulator complex (GO:0005667) (Supplementary Table 8).

For the genes classified as compensating (Fig. 3B), gene ontology term analysis revealed significant enrichment in several biological processes, including cellular processes (GO:0009987), cellular metabolic processes (GO:0044237), and organic substance biosynthetic processes (GO:1901576). For GO molecular function, these genes were significantly enriched for binding (GO:0005488), structural constituent of ribosome (GO:0003735), and RNA binding (GO:0003723). Regarding cellular components, GO term analysis showed significant enrichment for intracellular anatomical structure (GO:0005622), intracellular organelle (GO:0043229), and organelle (GO:0043226) (Supplementary Table 8). For the genes classified as enhancing (Fig. 3C) we observed electron transport chain (GO:0022900), mitochondrial ATP synthesis coupled electron transport (GO:0042775), and ATP synthesis coupled electron transport (GO:0042773) significantly enriched as biological processes, while structural constituent of ribosome (GO:0003735) as enriched in molecular function and cytoplasm (GO:0005737), mitochondrion (GO:0005739), and mitochondrial envelope (GO:0005740) as enriched in cellular component GO terms (Supplementary Table 8).

For the genes categorized as only *cis* (Fig. 3D), our analysis revealed significant enrichment in several biological processes, including the small molecule metabolic process (GO:0044281), carboxylic acid metabolic process (GO:0019752), and oxoacid metabolic process (GO:0043436). This aligns with our expectations as these flies were reared in an octanoic acid environment suggesting this enrichment is functionally important for fitness in an OA environment. Regarding molecular functions, the *cis*-only genes exhibited significant enrichment in catalytic activity (GO:0003824), carboxylic ester hydrolase activity (GO:0052689), and carboxylesterase activity (GO:0106435), again aligning well with the the treatments used and the ecological differences between the species used in this work. Additionally, the cellular component analysis indicated significant enrichment for the cytoplasm (GO:0005737), mitochondrion (GO:0005739), and organelle membrane (GO:0031090) (Supplementary Table 8). We found no significant enrichment for genes classified as *trans*-only (Fig. 3E).

A hierarchical model of gene expression variation

By using a fully factorial design we have identified a hierarchy of the impact from different sources of variation in gene expression traits (Fig. 3G). We found species/ genomic differences and differences between developmental stages had the largest effects and were each supported by different metrics as having the largest effect herein. Both had a substantially greater impact on gene expression than the current environment, which in itself had a greater effect than transgenerational plasticity. We supplemented the data generated for this project and overlaid them with published data from our lab resulting from experiments with incredibly similar experiment designs and bioinformatic analyses. From this, we produced hypothetical ranges for the proposed hierarchy (Fig. 3G), generating an ordering of the influences that different sources have on variation in gene expression traits that we hope will serve as a model for how to think about the hierarchy of importance of different effects.

Materials and methods

Drosophila strains and maintenance

Strains of two *Drosophila* species were used in this study: *D. simulans* (14021-0251.195) and *D. sechellia* (14021-0428.25). All flies were reared on standard food using a 16:8 light:dark cycle at 20 °C. Virgin females were collected over four days. Crosses were set at a ratio of 10 virgin females to 10 males per bottle in all cross designs (pure species crosses of *D. simulans* x *D. simulans* and *D. sechellia* x *D. sechellia*, and a hybrid cross of female *D. simulans* x *D. sechellia* males). This cross was only made in one direction because the reproductive output for this direction of cross is much higher than the reciprocal cross. This is largely the result of behavioral differences where male *D. sechellia* are found to be more persistent in courtship than *D. simulans* males and *D. sechellia* females more selective than *D. simulans* females, making the two directions of cross differentially successful at producing offspring in numbers that are required for this experiment. Prior to crossing, all flies were 0–3 days old when exposed to either standard food (0.75 g *Drosophila* medium and 2.5 mL H_2O per vial) or standard food supplemented with 0.2% octanoic acid (0.75 g *Drosophila* medium, 2.5 mL H_2O, and 6.5 µl octanoic acid per vial[46,62–64], for a period of 4 days. Flies were then moved to a new vial of standard food for a 24-h period to purge any eggs experiencing octanoic acid from the females. All flies were crossed on standard medium and were 5–8 days old at the time of crossing. Flies were crossed in replicates across both standard food and standard food containing 0.2% octanoic acid representing 12 different sample types (Fig. 1A). Crossed flies were transferred to new bottles after larval pupation was observed. Adult F_1 flies emerging from each cross were then exposed to either standard food or standard food containing 0.2% OA for a period of 24 h. All exposed adult F_1 flies were 0–4 days old at the time of exposure and 1–5 days old at the time of freezing and eventual sequencing. Wandering F_1 larvae (stage L3) were collected from each cross and exposed to either standard food or standard food containing 0.2% OA for an exposure period of one hour. Following exposure periods, adult and larval flies were snap-frozen in liquid nitrogen and stored at −80 °C until RNA extraction.

RNA extraction, library preparation, and RNA sequencing

RNA was extracted from a single tissue homogenate of either ten whole adult females or ten whole female L3 larvae using the Promega SV total RNA extraction system (catalog number: Z3105) with modified protocol (Promega)[65] in triplicate for each sample type. About 5 µL of RNA from each extraction was checked via gel electrophoresis to confirm successful RNA extraction. All further RNA quality control (BioAnalyzer and NanoDrop), concentration adjustment, library preparation, and RNA sequencing (Illumina HiSeq-4000, H4K Single End 52 Cycle) were performed by the University of Michigan Medical School DNA Sequencing Core generating 1,440,852,310 total sequence reads with an average of 20,011,837 and a range of 11,639,269–32,199,639 sequence reads per sample (Supplementary Table 9).

Bioinformatics pipeline

To identify the reads that map allele-specifically (i.e., they uniquely align to only one of the two species genomes present in the hybrids), we used

Bowtie2[66] and aligned hybrid sample sequencing reads to the *D. sechellia* genome and reads that did not align were then aligned to the *D. simulans* genome (Fig. 1B) and any that did align allele-specifically to the *D. simulans* genome must contain sequence differences that are differentiated by this two-step alignment process (and not mapped to *D. sechellia* genome). In parallel, we also took all reads and used Bowtie2 to align reads to the *D. simulans* genome and the reads that did not map were aligned to the *D. sechellia* genome to identify reads that aligned allele-specifically to the *D. sechellia* genome. Like that described above. Cufflinks[67] and Cuffdiff[68] were used for expression quantification, data normalization, and false discovery rate correction. The downstream analysis was performed by using in-house R scripts executed in RStudio with R version 4.0.1 (Fig. 1B) and the allele-specific read mapping percentage is provided in Supplementary Table 10.

Statistics and reproducibility

The statistical methods are detailed in their respective sections (see below). The statistical criteria used to categorize allele-specific interactions are provided in Supplementary Table 7. To ensure reproducibility, the raw data files have been deposited in GEO, and the processed data are available in the Supplementary Data files. Furthermore, the analysis script, written in R (v 4.3.1), is accessible on GitHub (see below).

Hierarchical clustering

Only the parental values from the entire dataset containing FPKM expression values (Supplementary Table 1) was used for generating the hierarchical clustering plot in R using the hclust()[69] function. The gene expression values were first scaled and euclidean distances were computed which were then passed to hclust() function to compute the clustering and generate a plot.

Principal component analysis and regression models

The entire dataset containing FPKM expression values (Supplementary Table 1) was subjected to the PCA using the R package PCAtools (v2.4.0)[70]. The number of principal components was determined by the elbow method. The eigenvector of the covariance matrix generated by PCAtools was used to represent the correlations between various factors in our study (species, stage, current and previous generation environment) and the principal components. Using the Pearson R^2 correlation method, the correlation among the PCs with the different variation factors was calculated. With the help of the inbuilt lm() function in R, we generated the multivariate linear regression models and computed all the estimates and parameters for our data. The *p*-values obtained were subjected to Bonferroni false discovery rate correction.

Gene ontology term enrichment analysis

The gene ontology analysis was performed in R using the gprofiler2 package (version 0.2.3)[71]. The list of genes was passed into the gost() function, and the graphs were plotted using the gostplot() function.

Allele-specific quantification and analysis of *cis*- and *trans*-regulatory change

After aligning the hybrid reads first to the *D. sechellia* genome and the non-aligned reads to the *D. simulans* genome and vice-versa (Fig. 1B) we obtained the allele-specific expression for each gene. These genes were further classified into various categories such as conserved, compensating, enhancing, only *cis*, and only *trans* using the criteria based on Supplementary Table 7. Briefly, we used the results from statistical analysis testing for differences between species, differences in allele-specific expression in hybrids (differences in *cis*), and differences between species ratio and allelic ratio (differences in *trans*). The results were then represented in the form of bar graphs (Fig. 3A–E), allele-specific graphs (Supplementary Fig. 1C–F, O–R), and pie charts (Supplementary Fig. 1K–N, W–Z). The enrichment of these categories (conserved, compensating, enhancing, only *cis*, and only *trans*) in various factors (species, stage, current, and parental environment) and their interactions (Supplementary Fig. 1G, H, S–V) was computed using the Fisher test in R. The significant odd ratios generated were $\log_2$-transformed and the positive values were assigned as 1 (over enriched) while the negative values were assigned as −1 (under enriched). The remaining values were assigned 0. For the representative heatmap in Fig. 3F, if any of the bins in the individual conditions (SS, SO, OS, and OO) were enriched, it was included.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The RNA-seq data generated for this manuscript is available in the Gene Expression Omnibus of NCBI under accession number GSE280534. Additional processed data files are available as Supplemental Data Files 1–3: Supplemental Data File 1 contains Supplemental Tables 1–7, Supplemental Data File 2 contains Supplemental Table 8, Supplemental Data File 3 contains Supplemental Tables 9, 10, see Supplementary Information for table legends with detailed description of each Supplemental Table.

Code availability

The entire R code written to compute the analysis and produce the necessary graphs and table is available on GitHub (https://github.com/sidkalra512/Trans_generational_Plasticity.git).

Received: 20 November 2023; Accepted: 12 November 2024;
Published online: 25 November 2024

References

1. Howard, M. L. & Davidson, E. H. *cis*-Regulatory control circuits in development. *Dev. Biol.* **271**, 109–118 (2004).
2. Avalos, D. et al. Genetic variation in cis-regulatory domains suggests cell type-specific regulatory mechanisms in immunity. *Commun. Biol.* **6**, 1–11 (2023).
3. Marand, A. P., Eveland, A. L., Kaufmann, K. & Springer, N. M. *cis*-Regulatory elements in plant development, adaptation, and evolution. *Annu. Rev. Plant Biol.* **74**, 111–137 (2023).
4. Farfán-Pira, K. J., Martínez-Cuevas, T. I., Evans, T. A. & Nahmad, M. A *cis*-regulatory sequence of the selector gene vestigial drives the evolution of wing scaling in *Drosophila* species. *J. Exp. Biol.* **226**, jeb244692 (2023).
5. Delaneau, O. et al. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**, eaat8266 (2019).
6. Hummel, N. F. C. et al. The *trans*-regulatory landscape of gene networks in plants. *Cell Syst.* **14**, 501–511.e4 (2023).
7. Charlesworth, B. Causes of natural variation in fitness: evidence from studies of *Drosophila* populations. *Proc. Natl Acad. Sci. USA* **112**, 1662–1669 (2015).
8. Warringer, J. et al. Trait variation in yeast is defined by population history. *PLoS Genet.* **7**, e1002111 (2011).
9. Hughes, J. T., Williams, M. E., Rebeiz, M. & Williams, T. M. Widespread *cis*- and *trans*-regulatory evolution underlies the origin, diversification, and loss of a sexually dimorphic fruit fly pigmentation trait. *J. Exp. Zool. B Mol. Dev. Evol.* **340**, 143–161 (2023).
10. Vande Zande, P., Hill, M. S. & Wittkopp, P. J. Pleiotropic effects of *trans*-regulatory mutations on fitness and gene expression. *Science* **377**, 105–109 (2022).
11. Vesth, T. C. et al. Investigation of inter- and intraspecies variation through genome sequencing of *Aspergillus* section *Nigri*. *Nat. Genet.* **50**, 1688–1695 (2018).
12. Ioerger, T. R. et al. Variation among genome sequences of H37Rv strains of *Mycobacterium tuberculosis* from multiple laboratories. *J. Bacteriol.* **192**, 3645–3653 (2010).
13. McManus, C. J. et al. Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Res.* **20**, 816–825 (2010).

14. Coolon, J. D., McManus, C. J., Stevenson, K. R., Graveley, B. R. & Wittkopp, P. J. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* **24**, 797–808 (2014).
15. Wang, X. et al. Stage-specific and location-specific cartilage calcification in osteoarthritis development. *Ann. Rheum. Dis.* **82**, 393–402 (2023).
16. Alsubaie, B. et al. Highly sex specific gene expression in Jojoba. *BMC Plant Biol.* **23**, 440 (2023).
17. Pal, S., Oliver, B. & Przytycka, T. M. Stochastic modeling of gene expression evolution uncovers tissue- and sex-specific properties of expression evolution in the *Drosophila* genus. *J. Comput. Biol.* **30**, 21–40 (2023).
18. Shen-Orr, S. S. et al. Cell type–specific gene expression differences in complex tissues. *Nat. Methods* **7**, 287–289 (2010).
19. Smith, E. N. & Kruglyak, L. Gene–environment interaction in yeast gene expression. *PLoS Biol.* **6**, e83 (2008).
20. Choi, J. K. & Kim, S. C. Environmental effects on gene expression phenotype have regional biases in the human genome. *Genetics* **175**, 1607–1613 (2007).
21. Migliore, L. & Coppedè, F. Gene–environment interactions in Alzheimer disease: the emerging role of epigenetics. *Nat. Rev. Neurol.* **18**, 643–660 (2022).
22. Brawand, D. et al. The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
23. Metzger, B. P. H. et al. Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol. Evol.* **9**, 843–854 (2017).
24. Hales, N. R. et al. Contrasting gene expression programs correspond with predator-induced phenotypic plasticity within and across generations in *Daphnia*. *Mol. Ecol.* **26**, 5003–5015 (2017).
25. Colicchio, J. M., Monnahan, P. J., Kelly, J. K. & Hileman, L. C. Gene expression plasticity resulting from parental leaf damage in *Mimulus guttatus*. *N. Phytol.* **205**, 894–906 (2015).
26. Earley, T. S., Feiner, N., Alvarez, M. F., Coolon, J. D. & Sultan, S. E. The relative impact of parental and current environment on plant transcriptomes depends on type of stress and genotype. *Proc. R. Soc. B Biol. Sci.* **290**, 20230824 (2023).
27. Coolon, J. D., Stevenson, K. R., McManus, C. J., Graveley, B. R. & Wittkopp, P. J. Genomic imprinting absent in *Drosophila melanogaster* adult females. *Cell Rep.* **2**, 69–75 (2012).
28. Sultan, S. E., Moczek, A. P. & Walsh, D. Bridging the explanatory gaps: what can we learn from a biological agency perspective? *BioEssays* **44**, 2100185 (2022).
29. Wilkins, J. F. & Haig, D. What good is genomic imprinting: the function of parent-specific gene expression. *Nat. Rev. Genet.* **4**, 359–368 (2003).
30. Fox, N. A., Hane, A. A. & Pine, D. S. Plasticity for affective neurocircuitry: how the environment affects gene expression. *Curr. Dir. Psychol. Sci.* **16**, 1–5 (2007).
31. Meaney, M. J. Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annu. Rev. Neurosci.* **24**, 1161–1192 (2001).
32. Carone, B. R. et al. Paternally induced transgenerational environmental reprogramming of metabolic gene expression in mammals. *Cell* **143**, 1084–1096 (2010).
33. Combes, M.-C., Joët, T., Stavrinides, A. K. & Lashermes, P. New cup out of old coffee: contribution of parental gene expression legacy to phenotypic novelty in coffee beans of the allopolyploid *Coffea arabica* L. *Ann. Bot.* **131**, 157–170 (2023).
34. Cuomo, A. S. E. et al. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nat. Commun.* **11**, 810 (2020).
35. Oliva, M. et al. The impact of sex on gene expression across human tissues. *Science* **369**, eaba3066 (2020).
36. Raja, V., Qadir, S. U., Alyemeni, M. N. & Ahmad, P. Impact of drought and heat stress individually and in combination on physio-biochemical parameters, antioxidant responses, and gene expression in *Solanum lycopersicum*. *3 Biotech* **10**, 208 (2020).
37. Lee, H.-O. et al. Lineage-dependent gene expression programs influence the immune landscape of colorectal cancer. *Nat. Genet.* **52**, 594–603 (2020).
38. Raychaudhuri, S., Stuart, J. M. & Altman, R. B. Principal components analysis to summarize microarray experiments: application to sporulation time series. *Pac Symp Biocomput*. 455–466 https://doi.org/10.1142/9789814447331_0043 (2000).
39. Satam, H. et al. Next-generation sequencing technology: current trends and advancements. *Biology* **12**, 997 (2023).
40. Matute, D. R. & Ayroles, J. F. Hybridization occurs between *Drosophila simulans* and *D. sechellia* in the Seychelles archipelago. *J. Evol. Biol.* **27**, 1057–1068 (2014).
41. Garrigan, D. et al. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. *Genome Res.* **22**, 1499–1511 (2012).
42. Legal, L., Chappe, B. & Jallon, J. M. Molecular basis of *Morinda citrifolia* (L.): toxicity on *Drosophila*. *J. Chem. Ecol.* **20**, 1931–1943 (1994).
43. Jones, C. D. The genetics of adaptation in *Drosophila sechellia*. *Genetica* **123**, 137–145 (2005).
44. Salazar-Jaramillo, L. & Wertheim, B. Does *Drosophila sechellia* escape parasitoid attack by feeding on a toxic resource? *PeerJ* **9**, e10528 (2021).
45. Lanno, S. M. et al. Transcriptomic analysis of octanoic acid response in *Drosophila sechellia* using RNA-sequencing. *G3* **7**, 3867–3873 (2017).
46. Andrade López, J. M. et al. Genetic basis of octanoic acid resistance in *Drosophila sechellia*: functional analysis of a fine-mapped region. *Mol. Ecol.* **26**, 1148–1160 (2017).
47. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene regulation. *Nature* **430**, 85–88 (2004).
48. Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nat. Genet.* **40**, 346–350 (2008).
49. R'Kha, S., Capy, P. & David, J. R. Host-plant specialization in the *Drosophila melanogaster* species complex: a physiological, behavioral, and genetical analysis. *Proc. Natl Acad. Sci. USA* **88**, 1835–1839 (1991).
50. Amlou, M., Moreteau, B. & David, J. R. Genetic analysis of *Drosophila sechellia* specialization: oviposition behavior toward the major aliphatic acids of its host plant. *Behav. Genet.* **28**, 455–464 (1998).
51. Farine, J.-P., Legal, L., Moreteau, B. & Le Quere, J.-L. Volatile components of ripe fruits of *Morinda citrifolia* and their effects on *Drosophilia*. *Phytochemistry* **41**, 433–438 (1996).
52. Yassin, A. et al. Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl Acad. Sci. USA* **113**, 4771–4776 (2016).
53. Drum, Z. et al. Genomics analysis of *Drosophila sechellia* response to Morinda citrifolia fruit diet. *G3* **12**, jkac153 (2022).
54. Coolon, J. D. et al. Molecular mechanisms and evolutionary processes contributing to accelerated divergence of gene expression on the *Drosophila* X chromosome. *Mol. Biol. Evol.* **32**, 2605–2615 (2015).
55. Verta, J.-P. & Jones, F. C. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *eLife* **8**, e43785 (2019).
56. Garfield, D. A. et al. The impact of gene expression variation on the robustness and evolvability of a developmental gene regulatory network. *PLoS Biol.* **11**, e1001696 (2013).
57. Rifkin, S. A., Kim, J. & White, K. P. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* **33**, 138–144 (2003).
58. Gilad, Y., Oshlack, A. & Rifkin, S. A. Natural selection on gene expression. *Trends Genet.* **22**, 456–461 (2006).

59. Lemos, B., Meiklejohn, C. D., Cáceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evol. Int. J. Org. Evol.* **59**, 126–137 (2005).
60. Kalinka, A. T. et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010).
61. Hsieh, W.-P., Chu, T.-M., Wolfinger, R. D. & Gibson, G. Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* **165**, 747–757 (2003).
62. Lanno, S. M., Shimshak, S. J., Peyser, R. D., Linde, S. C. & Coolon, J. D. Investigating the role of Osiris genes in *Drosophila sechellia* larval resistance to a host plant toxin. *Ecol. Evol.* **9**, 1922–1933 (2019).
63. Lanno, S. M. & Coolon, J. D. Derived esterase activity in *Drosophila sechellia* contributes to evolved octanoic acid resistance. *Insect Mol. Biol.* **28**, 798–806 (2019).
64. Peyser, R. D., Lanno, S. M., Shimshak, S. J. & Coolon, J. D. Analysis of cytochrome P450 contribution to evolved plant toxin resistance in *Drosophila sechellia*. *Insect Mol. Biol.* **26**, 715–720 (2017).
65. Coolon, J. D., Webb, W. & Wittkopp, P. J. Sex-specific effects of *cis*-regulatory variants in *Drosophila melanogaster*. *Genetics* **195**, 1419–1422 (2013).
66. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
67. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).
68. Trapnell, C. et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* **31**, 46–53 (2013).
69. Nowakowski, S. hclust1d: hierarchical clustering of univariate (1d) data. 0.1.1. https://doi.org/10.32614/CRAN.package.hclust1d (2023).
70. Blighe, K. kevinblighe/PCAtools. (2024).
71. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2 - an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res.* **9**, ELIXIR–709 (2020).
72. Drum, Z. A. et al. Genomics analysis of hexanoic acid exposure in *Drosophila* species. *G3* **12**, jkab354 (2022).
73. DeAngelis, M. W., Coolon, J. D. & Johnson, R. I. Comparative transcriptome analyses of the *Drosophila* pupal eye. *G3* **11**, jkaa003 (2021).
74. Lanno, S. M. et al. Genomics analysis of L-DOPA exposure in *Drosophila sechellia*. *G3* **9**, 3973–3980 (2019).
75. Meiklejohn, C. D., Coolon, J. D., Hartl, D. L. & Wittkopp, P. J. The roles of *cis*- and *trans*-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. *Genome Res.* **24**, 84–95 (2014).

Acknowledgements

We thank Dr. Ni Ye Feng (Assistant Professor of Biology at Wesleyan University) for providing access to Biorender (https://www.biorender.com/). Research reported in this publication was supported by Wesleyan University (start-up funds to J.D.C. and Department of Biology funds to J.D.C.), the Wesleyan Ronald E. McNair Post Baccalaureate Program (G.S.), and the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R15GM135901 (awarded to J.D.C.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions

S.K.: Formal analysis, writing—original draft, writing—review & editing, and visualization. S.L.: Experimental design, formal analysis, and data curation. G.S.: Formal analysis. J.D.C.: Conceptualization, methodology, data curation, writing—original draft, writing—review & editing, supervision, and funding acquisition.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-07255-6.

Correspondence and requests for materials should be addressed to Joseph D. Coolon.

Peer review information *Communications Biology* thanks Siyuan Feng and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editors: Luciano Matzkin, George Inglis and Johannes Stortz. A peer review file is available.