Evolution, 2025, **79**(**10**), 1977–1995
https://doi.org/10.1093/evolut/qpaf158
Advance access publication 30 July 2025
Original Article

The genetic basis of natural variation in sociability

Arteen Torabi-Marashi [1], Dania Daanish[2], Andrew M. Scott[2], Reuven Dukas [2], Ian Dworkin [1]

[1]Department of Biology, McMaster University, Hamilton, Ontario, Canada
[2]Cognitive Ecology Group, Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada

Corresponding authors: Ian Dworkin, Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. Email: dworkin@mcmaster.ca; Reuven Dukas, Cognitive Ecology Group, Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, Ontario, Canada. Email: dukas@mcmaster.ca
A.T-M. and D.D. contributed equally.
R.D. and I.D. shared senior authorship.

Abstract

Sociability, defined as individuals' tendencies to affiliate with conspecifics, is widespread among animals, including species not traditionally categorized as social. A few studies have documented a positive association between sociability and fitness, and sociability has positive effects on health, well-being, and longevity in humans. Despite the importance of sociability, we still have limited information about its genetic basis. To address this knowledge gap, we compared gene expression in the heads of fruit flies (*Drosophila melanogaster*) from lineages artificially selected for, and divergent in, degree of sociability. We identified 174 genes that showed differential expression among low and high-sociability lineages, of which 33 genes have known effects on neuroanatomy, neurophysiology, or behavior. Experiments using knockdowns of 16 of the top candidate sociability genes revealed 6 of them significantly affecting sociability in the predicted direction. Relying on our large genomic database, we and others can further elucidate the genetic architecture and evolutionary biology of sociability.

Keywords: artificial selection, *Drosophila melanogaster*, gene expression, sociability, social behavior

Introduction

Sociability, defined as individuals' tendencies to affiliate with conspecifics, is prevalent among animal species. Sociability means that individuals seek each other while engaged in activities, including feeding, traveling, resting, and sleeping (Allee, 1938; Clutton-Brock, 2016; Scott et al., 2022a; Tinbergen, 1953; Ward & Webster, 2016; Wilson, 1975). While we often associate sociability with group-living species such as many mammals, flocking birds, schooling fish, and social insects, many more species, including a large variety of insects traditionally classified as non-social, are sociable (Costa, 2006; Prokopy & Roitberg, 2001). For example, although fruit flies (*Drosophila melanogaster*) had been historically perceived as solitary animals, they form social groups in both laboratory (Bentzur et al., 2021; Billeter et al., 2024; Chen & Sokolowski, 2022; Durisko & Dukas, 2013; Saltz, 2011; Schneider et al., 2012) and field settings (Dukas, 2020), and show heritable variation in sociability (Scott et al., 2018, 2022a). Overall, fruit flies possess a variety of social behaviors, including social synchronization of the circadian clock (Levine et al., 2002), social learning (Battesti et al., 2012; Sarin & Dukas, 2009), and collective response to danger (Ferreira & Moita, 2020; Ramdya et al., 2015).

Several studies have documented positive associations between sociability and fitness (Bond et al., 2021; Dal Pesco et al., 2022; Gerber et al., 2022; Kajokaite et al., 2022; Snyder-Mackler et al., 2020). Most notably, long-term studies have revealed positive correlations between measures of social integration and components of fitness in baboons (*Papio* spp.). Female baboons that have stronger social bonds with other females live longer (Silk et al., 2010), have higher infant survival (Silk et al., 2003), and their adult offspring have longer lifespan (Silk et al., 2009). Male baboons that are more strongly bonded to females also have longer lifespans (Campos et al., 2020). In humans, the number and quality of social contacts are positively correlated with a variety of health measures as well as with longevity (Elovainio et al., 2017; Holt-Lunstad et al., 2015; House et al., 1988; Steptoe et al., 2013; Yang et al., 2016), but this positive correlation may not indicate causation (Liang et al., 2024).

Despite the clear importance of sociability for many animals, including humans, we still have limited knowledge of its genetic architecture. Within the broader area of social behavior, defined as interactions among conspecifics, perhaps the most relevant bodies of literature include genomic studies of social hymenopterans (Kocher et al., 2018; Shpigler et al., 2017; Smith et al., 2008; Toth & Rehan, 2017), genetic studies of animal models, primarily mice, employed for investigating the mechanistic basics of social deficiencies such as autism spectrum disorder in humans (de la Torre-Ubieta et al., 2016; Moy & Nadler, 2008; Silverman et al., 2010), and genome-wide association studies focusing on socializing and loneliness in humans (Bralten et al., 2021; Clyde, 2018; Day et al., 2018). It has been suggested that there are highly conserved mechanisms for social behaviors, which are shared by a wide range of animals from insects to mammals, including humans. Evidence in support of this proposition, however,

Received August 8, 2024; revisions received January 22, 2025; accepted July 29, 2025
Associate Editor: Amanda J. Moehring; Handling Editor: Jason Wolf

is currently limited (Rittschof et al., 2014; Robinson et al., 2005; Shpigler et al., 2017).

To further advance our knowledge of the genetic architecture of sociability, we artificially selected replicated lineages of fruit flies that depict either low or high sociability (Scott et al., 2022a). Our measure of sociability was the strength of flies' tendency to join others when presented with eight food patches of identical quality. Flies could fall anywhere between showing a strong tendency to feed and rest together on a single patch, indicating high sociability, to random assortment among the eight food patches, reflecting no significant sociability. We then utilized genomics approaches to identify genes that may mediate variation in sociability. First, we compared gene expression in the heads of adult flies from the divergent lineages using RNAseq. Our major goal was to identify the set of genes that showed differential expression between the low- and high-sociability lineages. Then we examined sociability by sex interactions and differential gene transcript usage. Once we identified candidate sociability genes, we asked what biological processes they are involved in, whether they were disproportionately expressed in head tissue, and whether they were homologous to genes linked to social behavior in other species. Finally, we conducted experiments to verify the effects of top candidate genes on sociability.

Materials and methods

Artificial selection

We previously applied artificial selection on sociability (Scott et al., 2022a). For each selection treatment, we had four independently evolving lineages (four low-sociability lineages, four high-sociability lineages, and four control lineages). For each generation, we quantified sociability in 12 groups of 16 females and 12 groups of 16 males from each of the 4 low- and 4 high-sociability lineages. The 16 flies of each single-sex group were held together from sexing within 8 hr post eclosion until testing, when they were 4 days old. To quantify sociability, we placed each group of 16 flies inside a sociability arena, which had eight equal-sized compartments, each containing a food disc (Figure 1 in Scott et al., 2022a). Flies could move freely among compartments and either form one or a few groups, or disperse among all eight food patches to feed and rest. After 90 min, we blocked the passage, and recorded the number of flies in each compartment. From this record, we calculated the sociability score as the variance scaled by the mean number of flies in each compartment. With this widely used ecological measure of dispersion (Krebs, 1999), a value of 0 indicates social avoidance, 1 means random distribution, and values significantly above 1 denote significant sociability (Durisko et al., 2014; Scott et al., 2018). We then selected four flies from each arena (96 individuals total per lineage). For the low-sociability lineages, we selected flies from compartments with the lowest numbers of individuals, while for the high-sociability lineages, we selected flies from compartment(s) with the highest number of individuals. For the 4 control lineages, we randomly selected 4 flies from each of the 12 groups of 16 same-sex flies per lineage. Owing to time constraints, we quantified sociability in the control lineages only every five generations. We used the extreme 48 males and 48 females from each lineage for breeding to generate the next generation of individuals. After 25 generations of selection, the high-sociability lineages showed sociability scores about 50% greater compared with low-sociability lineages. (Scott et al., 2022a; Figure 1).

In generation 26, we collected adult fly heads for gene expression analysis. We used identical methods as in previous generations except that we had two experience treatments. In the sociability arena treatment, we placed groups of 16 same-sex flies in the sociability arenas prior to their collection for gene expression. This provided the flies with the social dynamics experienced during their evolutionary history under artificial selection. When placed in the sociability arenas, flies initially engage in exploration and frequent contact with other flies (Scott et al., 2022a). We presumed that such social interactions would affect the expression of pertinent genes. In the vial treatment, we just moved groups of 16 same-sex flies into fresh vials, so these flies did not experience the sociability arenas. After 20 min, we rapidly transferred each group of 16 individuals into a 1.5 ml tube and submerged it in liquid nitrogen. We had 3 replicates per lineage × 12 lineages × 2 sexes × 2 experience conditions for a total of 144 samples. We later separated the flies' heads and extracted RNA.

RNA extraction and sequencing

Adult head tissue was homogenized in 1.5 ml tubes using a NextAdvance Bullet Blender (NextAdvance) using metal beads. We extracted RNA with the MagMAX-96 Microarrays Total RNA Isolation Kit (Thermo Fisher). We checked RNA purity with a Nanodrop spectrophotometer and quantified the amount of RNA using a Denovix Fluorometer (Denovix) with the Qubit RNA high-sensitivity assay kit (Thermo Fisher). We sent samples to Génome Québec (Centre d'expertise et de services, Génome Québec) for library preparation and sequencing. Library preparation used NEBNext dual multiplex oligos. Sequencing was done on an Illumina NovaSeq 6000 S4, generating 100 bp paired-end sequences. One sample was rejected in the quality control check for poor quality, and another sample was rejected for low quantity of RNA, so 142 samples were sequenced. A total of approximately 6.2 billion read clusters were generated with an average of 44 million read clusters per sample (total reads per sample in Supplementary File 1).

Read processing and mapping

All computational analysis was performed using the Graham cluster from the Digital Research Alliance of Canada (formerly Compute Canada; www.alliancecan.ca). We checked the sequencing quality of reads using FASTQC (v0.11.9, Andrews, 2010) and MultiQC (v1.12, Ewels et al., 2016), which assessed adapter content, per sequence quality scores, and GC (guanine-cytosine) content. All samples had a mean Phred score value of > 35. We assessed transcript integrity number (TIN) using RSeQC (v4.0.0, Wang et al., 2012), and all but two samples had median TIN scores >60, with those two having a median TIN score of 49 and 59. Using clustering and PCA, we evaluated whether these two samples appeared as outliers, and neither were. As such, we included all samples in the analysis. We trimmed adapters using trimmomatic (v0.36, Bolger et al., 2014), with both leading and trailing set to "3" and run parameters set to "MAXINFO:20:0.2." We removed reads shorter than 36 bp from the sample. Following trimming, we again used FASTQC

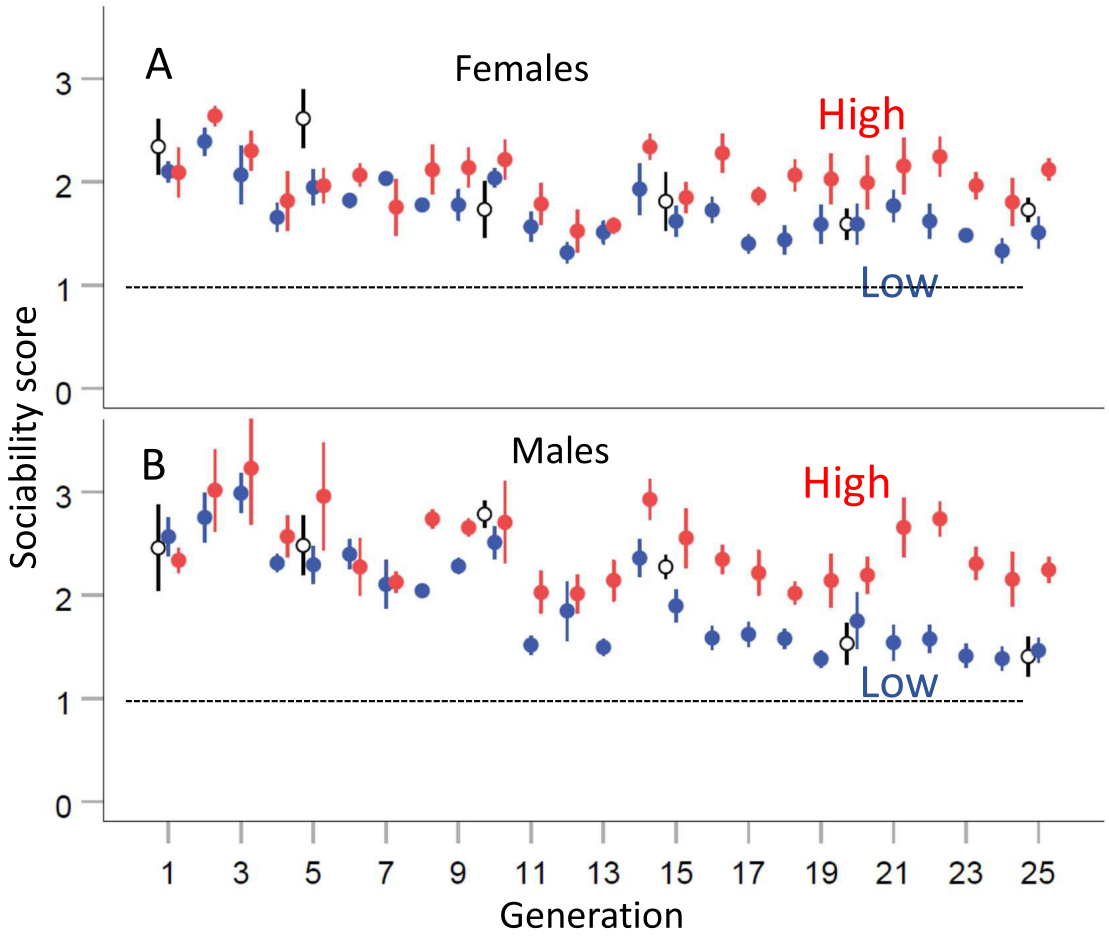

Figure 1. Mean ± SEM sociability scores over 25 generations in females (A) and males (B) of the low (blue), high (red), and control (black, unfilled) lineages (n = four lineages for each of the three treatments). Note that we quantified sociability in the control lineages only every five generations. Values significantly above 1 (dashed lines) indicate significant sociability. Data from (Scott et al. 2022a)).

and MultiQC to confirm adapters were trimmed while maintaining high-quality sequence. We mapped reads to a reference transcriptome of *D. melanogaster* from FlyBase (version r6.38, Gramates et al., 2022) using Salmon (v1.4.0, Patro et al., 2017) with decoys, which produced counts of transcripts per sample. To use Salmon, we first generated an index file from a list of decoys, a reference transcriptome, and a reference genome (version r6.38, Gramates et al., 2022), which we then used for mapping. We also separately used the splice-aware aligner STAR (v2.7.9a, Dobin et al., 2013) to map reads to a reference genome, which produced gene-level counts (Figure S1, Table S1). Counts were imported into R (v4.2.0, R Core Team, 2022) using tximport (v1.24.0, Soneson et al., 2016).

Principal component analysis

Count data were normalized using the "vst()" function from DESeq2 (v1.36.0, Love et al., 2014), which performs a variance stabilizing transformation. In the "vst()" call, we set "blind = FALSE," providing a design matrix consisting of sex, experience, selection, and the interaction between selection and sex. We also used "nsub = 5000" to filter for only the top 5,000 most variable genes. To visualize the principal component analysis (PCA) results, we used the function "plot_pca()" from RNAseqQC (v0.1.4, DeLuca et al., 2012) with "nfeats = 500" to plot the top 500 most variable genes. The robustness of the qualitative findings from the PCA were confirmed by trying different numbers of genes to include (from 500 to 10,000).

Differential gene expression analysis

To filter out low-expressed genes, we used the "filterByExpr()" function from edgeR (v3.38.4, Robinson et al., 2010). We filtered genes that had lower than 0.3 counts per million (CPM) in at least eight samples. We used CPM instead of raw counts for filtering step to avoid overrepresentation of genes expressed in larger libraries (Chen et al. 2016). From 13,701 genes, we removed 2,176 genes, leaving 11,525 genes for gene-wise analysis. For gene-level modeling, we utilized two distributional approaches, extending common approaches widely used for gene expression analysis. To model variation from counts directly, we used a negative-binomial generalized linear mixed model, while for Gaussian mixed models, counts were normalized and variance stabilized in the form of $\log_2$(CPM) using the "voom()" function in the limma package (v3.52.4, Ritchie et al., 2015). Details for each are given below. As we needed to incorporate random effects into our generalized linear models, we used glmmTMB (v 1.1.4, Brooks et al., 2017). Given that most studies of differential gene expression (DGE) use mod-

eling tools like limma-voom, EdgeR, and DESeq2, we confirmed that model estimates for fixed effects were similar when compared to estimates from the models lacking random effects fit in limma-voom, as discussed below.

The full model used lane, sex, experience (sociability arena or vial), and selection as main, fixed-effect terms. We included all second-order interactions between selection, sex, and experience. Lineage was modeled as a random effect nested within selection treatments. Variation for sex and experience was allowed to vary by lineage (i.e., random "slopes" for sex, by lineage). For each gene evaluated, the full model in glmmTMB syntax is

$$\begin{aligned}\log_2\text{cpm} \sim\ & \text{lane} + \text{sex} + \text{expMatched} + \text{selection} \\ & + \text{sex} : \text{expMatched} + \text{selection} : \text{expMatched} \\ & + \text{selection} : \text{sex} + \text{diag}(0 + \text{sex} \\ & + \text{expMatched} \mid \text{selection} : \text{lineage}).\end{aligned}$$

In standard notation:

$$\begin{aligned}y_i \sim N\big(\mu = {} & \beta_{0[j][i]} + \beta_1 x_{1[i]} + \beta_{2[j]} x_{2[i]} + \beta_{3[j]} x_{3[i]} \\ & + \beta_4 x_{4[i]} + \beta_5 x_{2[i]} x_{3[i]} + \beta_6 x_{4[i]} x_{3[i]} + \beta_7 x_{4[i]} x_{2[i]}, \sigma_y^2\big).\end{aligned}$$

With variances across lineages (for $j = 1, ..., 12$) within selection treatment:

$$\begin{pmatrix}\beta_{0[j]} \\ \beta_{2[j]} \\ \beta_{3[j]}\end{pmatrix} \sim MVN \begin{pmatrix}\mu_{\beta_0} & \sigma^2_{\beta_0} & 0 & 0 \\ \mu_{\beta_2}, & 0 & \sigma^2_{\beta_2} & 0 \\ \mu_{\beta_3} & 0 & 0 & \sigma^2_{\beta_3}\end{pmatrix},$$

where $y_i = \log_2(\text{CPM})$ (or counts for negative binomial, discussed below) for the ith sample (for $i = 1, ...,142$), x_1 = lane, x_2 = sex, x_3 = expMatched, and x_4 = artificial selection treatment.

Similarly to limma-voom, for each gene, we modeled $\log_2$ of CPM (computed from tximport), as our response, with residual variation assumed to be distributed normally. If a model failed to converge for a given gene, we adjusted the model to fit a slightly less complex random effect while keeping all other terms identical. The adjustment to the random effect was to drop experience, such that the random effect was now diag(0 + sex | selection : lineage). Importantly, examination of model fits where covariance between the sexes across lineages was set to 0, had minimal impacts on model estimates (and associated uncertainties) for coefficients of interest (selection treatment and sex) for this study. For our specific contrasts and downstream analyses, we utilized estimated marginal means (emmeans) and associated contrasts from model fits using the emmeans package (v1.8.1, Lenth, 2022).

For direct use of gene expression count data, for each gene, we fit models in glmmTMB, using gene expression counts (from tximport) as the response, utilizing a natural log link and a negative binomial distribution. We obtained and extracted normalization factors from DESeq2 using "estimateSizeFactors()" and "normalizationFactors()." Normalization factors were included as offsets (log transformed) in the models. To account for over-dispersion, we used the quadratic parameterization for the variance, "family = nbinom2()," specifying the variance as $V = \mu(1 + \mu/\phi)$, with predicted mean (μ), and dispersion parameter (ϕ). Given the large sample size of our experiment, the dispersion parameter was estimated uniquely for each gene. In contrast, in the default approach in DESeq2, information is shared (and dispersion estimates for individual genes most often shrunk toward) values for genes that share similar mean expression levels. The method used in DESeq2 is appropriate for analysis of datasets with limited sample sizes, albeit with strong assumptions. In particular, genes with similar expression levels have similar variances. As our sample sizes were relatively large, these could be estimated well for each gene in our models, and it was not necessary to make these assumptions. For both approaches, the mixed model fits in comparison to either limma-voom or DESeq2 had very similar fixed effect estimates, but with more appropriate measures of uncertainty (Figure S2, Table S2, supplementary results).

Filtering results, custom contrasts, and gene curation

We used a multi-stage approach to help identify genes where differential expression across selective treatments was relevant. In our first stage, we used the "Anova()" function from the R package car (Fox & Weisberg, 2019) to perform a Wald test on fitted models (per gene). For each gene, we extracted p-values for relevant model terms (relating to differences across sociability treatments). We adjusted p-values, using the R function "p.adjust()" with "method = "BY" (Benjamini & Yekutieli, 2001) for controlling false discovery rate. After adjustment, we filtered out genes with false discovery rate (FDR) < 0.05, using custom contrasts, via emmeans, to include genes whose expression differences across sociability treatments fit this criterion. This way, we only examined custom contrasts from genes passing initial filtering. For the next stage, we split contrast lists into three lists, which corresponded to our three contrasts: low versus high, low versus control, and control versus high. In each of these three lists, we pulled out genes with a p-value < .05 to obtain a list of genes in each of the three contrasts that potentially mediate sociability. We investigated each of these genes using the *Drosophila* database, FlyBase (vFB2023_01), focusing on whether previous work indicated expression in head tissue, links to social behavior, and orthologous human genes. We then focused on the gene list from the contrast between low versus high selection treatments and investigated each of the genes using the *Drosophila* database, FlyBase (vFB2023_01). We looked for evidence of expression in the adult head and relevant phenotypes, including neuroanatomy, neurophysiology, locomotor behavior, or circadian rhythm.

Differential transcript usage analysis

We also examined how differential transcript usage (DTU; sometimes called isoform switching) evolved amongst the evolutionary treatments. DTU evaluates the relative contribution of different transcripts within a gene to its overall abundance, and broadly speaking, can be thought of expression ratios of different transcripts. We followed recommendations for DTU analysis as outlined in Love et al. (2018). We generated transcript-level counts from Salmon and imported them into R (tximport). We normalized counts to scale to library size during import. We filtered transcripts using the "dmFilter()" function from DRIMseq (v1.24.0, Nowicka & Robinson, 2016). For a gene to be retained through filtering, the gene had to be expressed in a minimum of 28 samples (out of 142 total), with a minimum expres-

sion of 10 counts per sample, in those samples. For a given transcript to be retained, it had to be expressed in a minimum of 20 samples, with the transcript representing at least 5% of the gene's total expression in those samples. This removes rare transcripts (within sample) or genes with limited expression (across samples). A priori, we would not expect to be able to estimate coefficients for these transcripts with sufficient precision for meaningful comparisons. Prior to filtering, there were 6,559 genes with at least two transcripts representing a total of 21,143 transcripts from our samples. Post filtering, we had 4,761 genes representing 12,335 transcripts (Supplementary File 2).

For model fitting, we used an approach analogous to that used in DEXseq (Anders et al., 2012). However, as we did for total gene expression, we modified this approach to allow for the inclusion of random effects in the model in the framework of a generalized linear mixed model. This approach focuses on examining transcript-treatment interactions to assess DTUs (Love et al., 2018). For computational efficiency, DEXseq (v1.5.3) implemented a change (relative to earlier versions of the software) in how the design matrix is coded (based on changelogs, we think the library authors made this change in 2013), and thus how contrasts between transcripts are estimated (Reyes et al., 2013). The authors of DEXseq implemented this change to deal with high computational overhead, for situations where number of exons (or transcripts) per gene was very high (Anders et al., 2012). However, for our data and modeling strategy, this was not a constraint. As such, we retained treatment contrast coding for our design matrix during estimation, and, as discussed below, used emmeans to extract estimates and contrasts.

We used glmmTMB to fit a model that predicted counts for each individual transcript of a gene as described below. For a few genes, we observed complete separation (a transcript was completely absent in one treatment but varying in others). To account for this, we added a count of one to all transcripts for each sample. Thus, changes in transcript usage will be slightly underestimated. Of note is the inclusion of a random effect per sample to account for variation in transcript abundances within each biological sample. We fit both full and "null" models using the negative binomial distribution with glmmTMB. The full model was

$$\begin{aligned}\text{counts} \sim\ & 1 + \text{transcript} + \text{transcript} : \text{sex} + \text{transcript} \\ & : \text{selection} + \text{transcript} : \text{sex} : \text{selection} \\ & + \text{diag}(\text{sex} + \text{transcript} \mid \text{selection} : \text{lineage}) \\ & + (1 \mid \text{sample\_id}).\end{aligned}$$

The null model was

$$\begin{aligned}\text{counts} \sim\ & 1 + \text{transcript} + \text{transcript} : \text{sex} \\ & + \text{diag}(\text{sex} \mid \text{selection} : \text{lineage}) + (1 \mid \text{sample\_id}).\end{aligned}$$

For genes where the full model failed to converge, we re-fit with the following model:

$$\begin{aligned}\text{counts} \sim\ & 1 + \text{transcript} + \text{transcript} : \text{sex} \\ & + \text{transcript} : \text{selection} + \text{transcript} \\ & : \text{sex} : \text{selection} + (1 \mid \text{sample\_id}).\end{aligned}$$

With the corresponding null model:

$$\begin{aligned}\text{counts} \sim\ & 1 + \text{transcript} + \text{transcript} : \text{sex} \\ & + (1 \mid \text{sample\_id}).\end{aligned}$$

We confirmed model comparisons by using the "Anova()" function from the car package to perform a type II ANOVA. Concurrently, we obtained emmeans contrasts for low versus high sociability. We adjusted *p*-values using the "BY" method, filtering results to include genes with adjusted *p*-value < .05 for the transcript: selection interaction term (or transcript: sex for our checks for sex-specific DTUs). From here, we subsetted our emmeans list to only include genes that passed the cut-off for the ANOVA, and used custom contrasts to identify changes in transcript usage within genes across selective treatments, sex, and their interactions.

Gene ontology analysis

We performed gene ontology (GO) analysis using topGO (v2.48.0, Alexa & Rahnenfuhrer, 2022) on the subset of genes identified using our analysis pipelines as described above. We separately performed GO analysis on the subsets of filtered genes from DGE and DTU analyses. We set the minimum number of genes per GO term to 5 and used Fisher's exact test. We adjusted resulting *p*-values for multiple comparisons in the exact same way we did in gene curation.

Comparison to other social behavior studies

We compared differentially expressed (DE) genes from our study with those reported in four of the most relevant published studies that assessed the genetic basis of social behavior. For each of the four comparisons, we ran simulations to predict the chance occurrence of overlapping genes. First, Bralten et al. (2021) performed a genome-wide association study (GWAS) with 342,461 people from the UK Biobank, and identified 56 genes associated with sociability. We took their list of 56 genes and identified orthologous genes in *Drosophila*. We also took the orthologs of the 18 independent loci and identified the corresponding 8 orthologs (as some were single-nucleotide polymorphism (SNP) locations with no corresponding *Drosophila* orthologs).

Second, Wang et al. (2022) examined early life social experience in the bumblebee, *Bombus impatiens*. They performed RNA sequencing to look for genes DE between three separate early life conditions: colony-housed, group-housed (with others but outside of the colony), and isolation (Wang et al., 2022). They ended up with a list of 94 DE genes between isolated and colony-reared bees and 27 DE genes between isolated and group-housed bees, with 6 genes overlapping between the two contrasts (Wang et al., 2022).

Third, Woodard et al. (2011) examined the convergent evolution of eusociality across bee species. They looked across nine socially diverse bee species, which included eusocial and non-eusocial bees, and identified 212 genes that evolved more rapidly in eusocial lineages compared to non-eusocial lineages (Woodard et al., 2011).

Finally, Shpigler et al. (2017) performed DGE analysis on RNA obtained from the mushroom body of the brain of bees that only responded to a territorial threat, bees that only showed nursing behavior toward a queen larva, and bees that responded to neither. They identified 1,057 DE genes

between all three groups of social responsiveness (Shpigler et al., 2017).

To evaluate whether there was more overlap between genes associated with sociability that we identified in *Drosophila* and those in either *Bombus/Apis* or humans (Bralten et al., 2021; Shpigler et al., 2017; Wang et al., 2022; Woodard et al., 2011), we computed the total number of orthologous genes that could be compared across species (i.e., orthologs in *Drosophila* that could be unambiguously identified) used in each study. From this, we generated a random set of genes with the number of significant hits being the same observed in each study. From this, we identified the number of overlapping genes in this random set. We evaluated our observed number of overlapping genes, relative to the expectation based on overlaps for random sets of genes (based on 10,000 simulations for each study).

Candidate gene validation

We chose 20 candidate genes that had among the highest differential expression between the low- and high-sociability lineages, subject to the availability of RNAi strains that are part of either the Transgenic RNAi Project (TRiP) or Vienna Drosophila Resource Center (VDRC) collections (Dietzl et al., 2007; Zirin et al., 2020). Additionally, we used the TRiP-control strains that are co-isogenic with the TRiP RNAi knockdown strains for control crosses. Detailed genotypic information for each strain is in Table S3. We crossed males from each RNAi strain with females from a general nervous system Gal4 strain, to specifically knockdown gene expression of each chosen candidate gene. Our default Gal4 line was *elav*-Gal4, UAS-*Dicer2* (BDSC 25750). We used a weaker *elav*-Gal4 (BDSC 8765) with the *thoc5*-RNAi strain owing to high offspring mortality with the default Gal4 strain. We verified expression of *elav*-Gal4 in both Gal4 strains by crossing them to a UAS-GFP strain, which allowed us to visualize and confirm pattern of Gal4 expression in the *Drosophila* brain. Each experimental block had an equal number of assays for the control genotypes, in which we crossed males from the co-isogenic TRiP-control strains (BDSC stock 36303 or 36304 depending on the location of the transgene insertion) to *elav*-Gal4 females. While all gene knockdown experiments conducted with the TRiP strains had co-isogenic controls, we did not do this with the VDRC control strains as those are in a *white*$^{-}$ background, and the single copy of the "mini"-*white*$^{+}$ rescue in the *elav*-Gal4 insertion only partially rescued eye pigmentation in females as a heterozygote (close to wild-type in hemizygous males). As loss of function of *white* is associated with numerous behavioral defects, we wanted to avoid this situation. As such, we used the same genotypes to set up our controls as we did for the TRiP strains. On average, we did not observe substantial or consistent differences in sociability between the TRiP and VDRC collection of lines, but we cannot rule knockdown-specific genetic background effects for the latter (Chandler et al., 2013) contributing to our effects.

We maintained all strains at 25 °C and 60% RH on fly medium in which each 1 L contained 90 g sucrose, 75 g cornmeal, 10 g agar, 32 g yeast, 2 g methyl paraben dissolved in 20 ml ethanol, and water. We failed to produce normal offspring owing to either high mortality or unexpanded wings with three strains (BDSC# 50556, VDRC# 101616, and VDRC# 100094), while one RNAi strain was incorrectly identified and removed from further consideration. Hence, we ended up testing 16 candidate genes (Table S3).

We quantified sociability using a protocol modified from Scott et al. (2022a). We sexed groups of eight newly eclosed, unmated offspring from the knockdown group and control group and placed each group into same-sex and same-treatment food vials with 5 ml of standard food. We sexed 12 sets of focal flies per sex, treatment, and day, for a total of 384 flies per day, and quantified sociability once individuals were 3 days old. We used 3D-printed circular arenas 50 mm wide and 6 mm high divided into eight compartments by thin walls with openings 6 mm wide and 3.5 mm high (Figure S3). Flies readily traveled among the eight sections. The top of the arena consisted of a plexiglass sheet with a 3D-printed circular edge and a 3 mm hole. We placed a 7.5 mm wide and 2 mm high circular patch made of regular medium in each compartment, and covered it with 50 µl of juice solution made of 2 g live yeast dissolved in 10 ml orange juice. Every morning (8 a.m.), we placed fresh food patches in each arena, attached the covers, and gently aspirated groups of eight same-sex flies into each arena through the hole in the top. We then covered the hole with a small piece of transparent, sticky tape. We prepared 12 arenas per sex, per treatment, and 48 arenas in total per day, and placed the arenas inside a humidified container maintained at 25 °C and 50% RH.

We allowed flies to settle until 2 p.m. Then, an observer blind to treatment counted the number of flies in each compartment within each arena every 15 min for 1 hr. We calculated the five sociability scores for each arena as the variance over the average number of flies in each arena (Durisko et al., 2014; Scott et al., 2018). The minimum sociability score of 0 represents one individual within each of the eight sections of the arena, and the maximum sociability score of 8 is attained if all eight flies from a single group in one section of the arena. With this sociability measure, scores significantly greater than 1 indicate more social aggregation than expected at random. At the end of scoring arenas, we discarded flies, washed the arenas with detergent and water, and let them dry overnight. We conducted 3 test days for each candidate gene for a total of 144 arenas per experimental block (16 experimental blocks total).

The data used as our response variable from the sociability scores for each arena are semi-continuous, positively valued, with rare 0s. As such, we analyzed the sociability data for each gene by fitting a generalized linear mixed effects model with the glmmTMB package (v1.1.8, Brooks et al., 2017) using a Tweedie distribution with a log link function in R v4.3.3 (R-Core-Team, 2023). In this implementation, the Tweedie power parameter is constrained to lie between 1 (pure Poisson) and 2 (pure Gamma). We modeled treatment, sex, and their interaction as fixed effect, and included time from onset of scoring as a continuous predictor. We fit a random effect, allowing the intercept to vary for day of experiment, an independent random effect, allowing intercept, and slope for time within experiment to vary according to individual arena (unit of sampling). We also included a final random effect for experimental block. For the model fit, we had a singular convergence warning. As such, we confirmed the stability of fixed effect estimates with a model where we removed the random slope associated with time for the random effect of individual arenas, but otherwise identical to

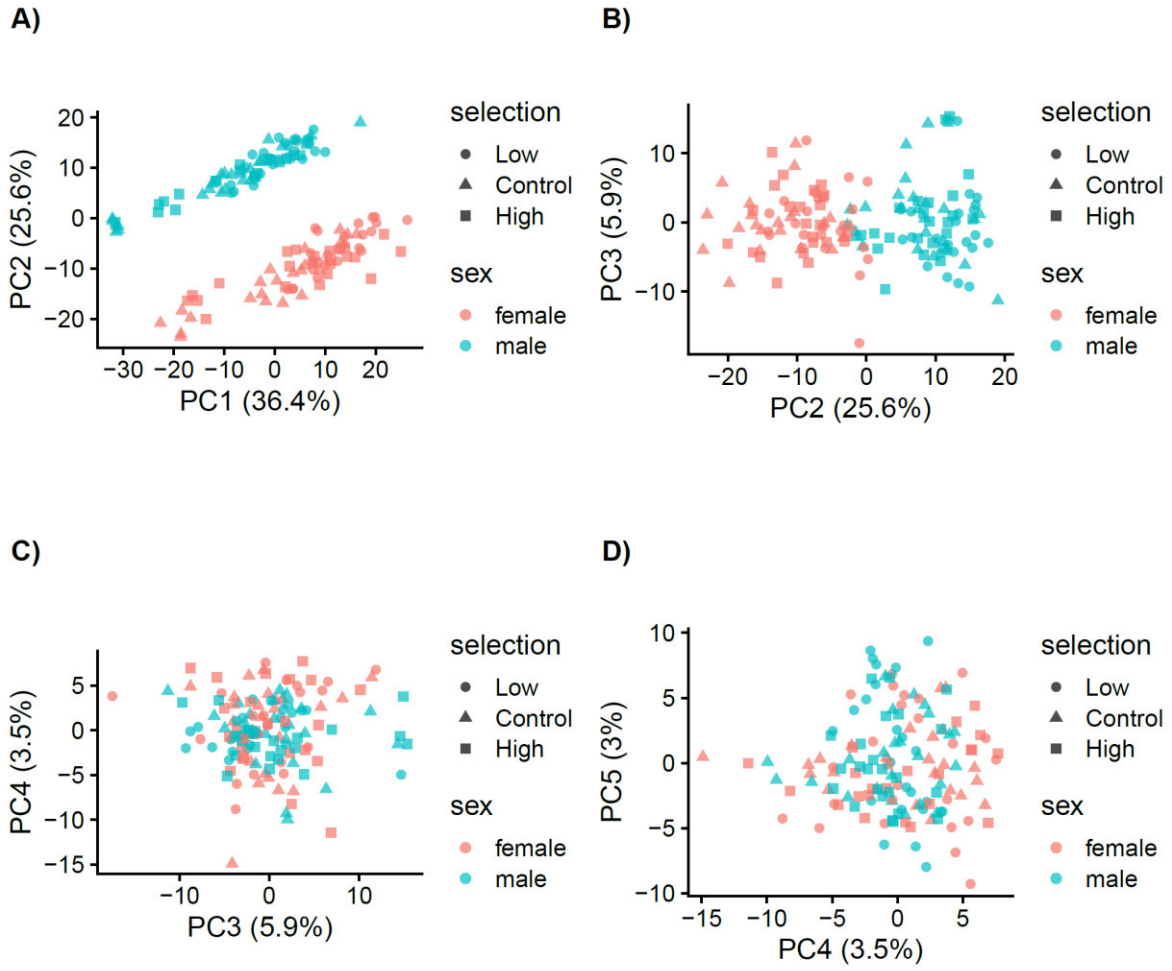

Figure 2. Principal component analysis (PCA) plots showing the variance associated with samples. Points on each plot are colored by sex, with females in red and males in blue. Different-shaped points represent different selection treatments. (A) PC1 (36.5% of variance) on the *x*-axis and PC2 (25.2% of variance) on the *y*-axis. (B) PC2 on the *x*-axis and PC3 (5.9% of variance) on the *y*-axis. (C) PC3 on the *x*-axis and PC4 (3.5% of variance) on the *y*-axis. (D) PC4 on the *x*-axis and PC5 (3% of variance) on the *y*-axis.

the model described above. Both models produced virtually identical estimates and confidence intervals for fixed effects, which are the focus of this study. We used emmeans and contrast functions from the emmeans package (v1.10.0, Lenth, 2022) to estimate custom contrasts of treatment effects, averaged over sex, as well as the interaction contrasts for treatment and sex effects (to assess sex-specific effects of RNAi-mediated gene knockdowns). These were adjusted for multiple comparisons using the DunnettX approach in emmeans, adjusting for 16 comparisons.

Results

Broad-scale co-variation in gene expression suggests changes associated with artificial selection for sociability

To examine broad-scale, qualitative patterns of variation in gene expression, we used PCA on the samples (Figure 2). Sexually dimorphic gene expression accounts for much of the (co)variation that loads on the second principal component, accounting for ~25% of the variation in gene expression, consistent with large-scale sex-biased gene expression in the adult head (Arbeitman et al., 2016; Khodursky et al., 2020; Nanni et al., 2023). Interestingly, PC1 (accounting for ~36% of the variation) shows that the lineages artificially selected for low sociability ("down") tend to have positively valued scores on PC1, while the samples from the control and "up" lineages are variable along PC1. This variation for high and control treatments in gene expression is a result of lineage-specific effects, i.e., replicate lineages within each selection treatment (Figure S4).

DGE analysis indicates that the expression of hundreds of genes may mediate natural variation in sociability

Following DGE analysis, we had results between all four possible combinations of using Salmon or STAR for mapping, and either a Gaussian ($\log_2$ CPM) or negative binomial distribution (counts with offsets). Results presented will be in reference to the Salmon-mapped counts fit with the Gaussian distribution (Figure S5). Contrasts and gene lists from the comparable analysis (STAR mapped, fit with a negative binomial with a log link) are provided in the repository (https://doi.org/10.6084/m9.figshare.29657138.

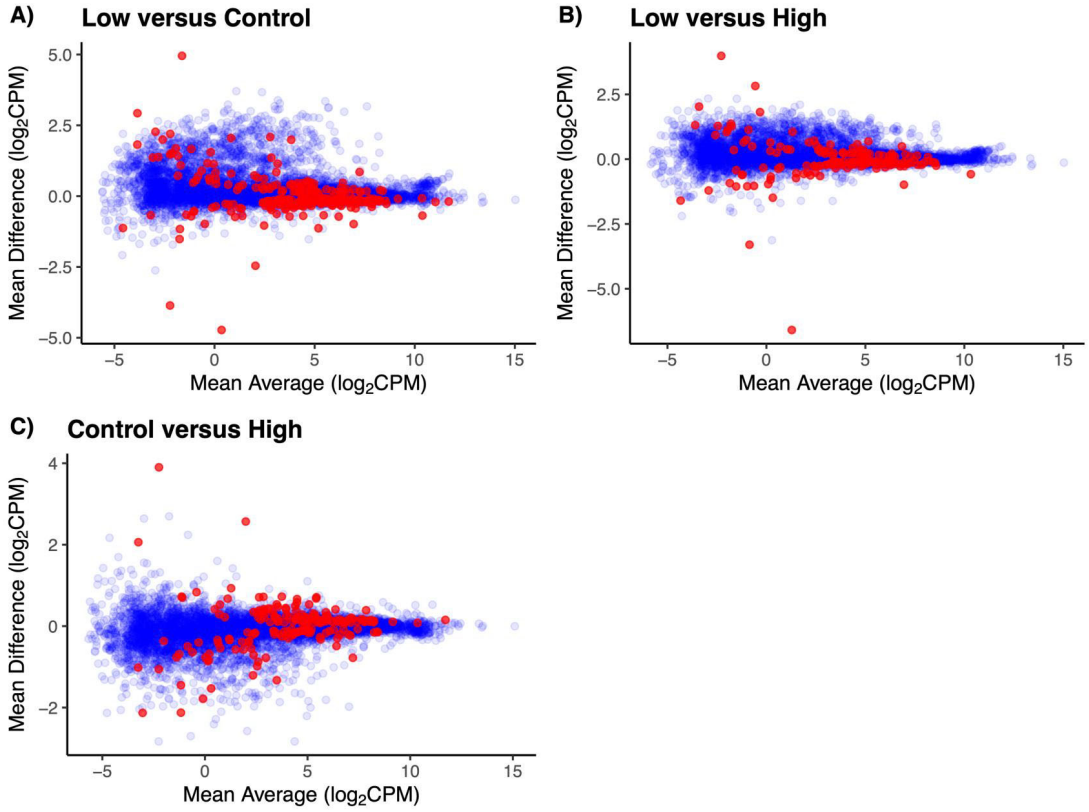

Figure 3. Mean-average (MA) plot for a given selection contrast. *X*-axis is the mean average in $\log_2$(CPM) for each gene obtained from emmeans. *Y*-axis is the mean difference between low and high sociability expression in $\log_2$(CPM), also obtained from emmeans. Red points are differentially expressed genes that have a *p*-value $< .05$ when looking at the given contrast, and blue points are genes that have a *p*-value $> .05$. (A) Low versus control sociability contrast. (B) Low versus high sociability contrast. (C) Control versus high sociability contrast.

v1). As a check, we first extracted DE genes between females and males. Previous studies have shown that within *D. melanogaster*, there is a large number of genes that show sex-biased gene expression differences (Parisi et al., 2004; Ranz et al., 2003), and relevant to our study, within the head (Arbeitman et al., 2016; Khodursky et al., 2020; Nanni et al., 2023). We found 5,331 genes that are DE between females and males (Figure S6) based on our filtering criteria.

Within the contrasts among the three artificially selected treatments, we examined the distribution of effects, and observed that a majority of the differences between treatments fell between a $\log_2$(CPM) of -1 and 1. Thus, the majority of evolved differences in gene expression had modest individual magnitudes. Based upon our filtering criteria, we observed 271 DE genes in the low sociability vs control artificial selection treatments (Figure 3A). In the low versus high selection contrast, we saw 174 DE genes (Figure 3B). In the control versus high selection contrast, we saw 194 DE genes (Figure 3C). We saw a total of 327 unique DE genes across the three selection contrasts. Figure 4 depicts the subset of 12 genes with the largest effect size in the low versus high selection contrast. Additionally, we found 213 genes DE between the vial and social arena experience contrast. We also examined if either sex or experience had an interacting effect with selection and found minimal evidence of genes altering gene expression in either the sex-by-selection interaction or the experience-by-selection interaction. Visualizations for each of these candidate genes are in Supplementary File 3.

Gene curation

We found 33 genes that are associated with relevant phenotypes, including neuroanatomy, neurophysiology, locomotor behavior, or circadian rhythm (Table 1). A subset of the 12 genes with the largest effect size is depicted in Figure 5.

Go analysis

Following GO analysis, we identified GO terms that are deemed as significantly overrepresented in our gene list of DE and differentially transcribed genes. When looking at all DE genes in our gene list, we found 60 GO terms overrepresented (Table S4). These terms included sensory perception of mechanical stimulus and synaptic assembly at neuromuscular junction. When looking at the DTU gene set, we found 43 GO terms overrepresented (Table S5), including photoreceptor cell axon guidance, regulation of neuron synaptic plasticity, and regulation of compound eye photoreceptor.

Comparison to other social behavior studies

First, of the 56 *Drosophila* genes orthologous to the human candidate sociability genes identified by Bralten et al. (2021), no specific *Drosophila* orthologs appeared in our list of "differentially expressed" genes. However, the family of solute carrier genes did appear in both lists. Our further analysis focusing on the 8 independent loci with *Drosophila* orthologs also revealed no ortholog that changed consistently across

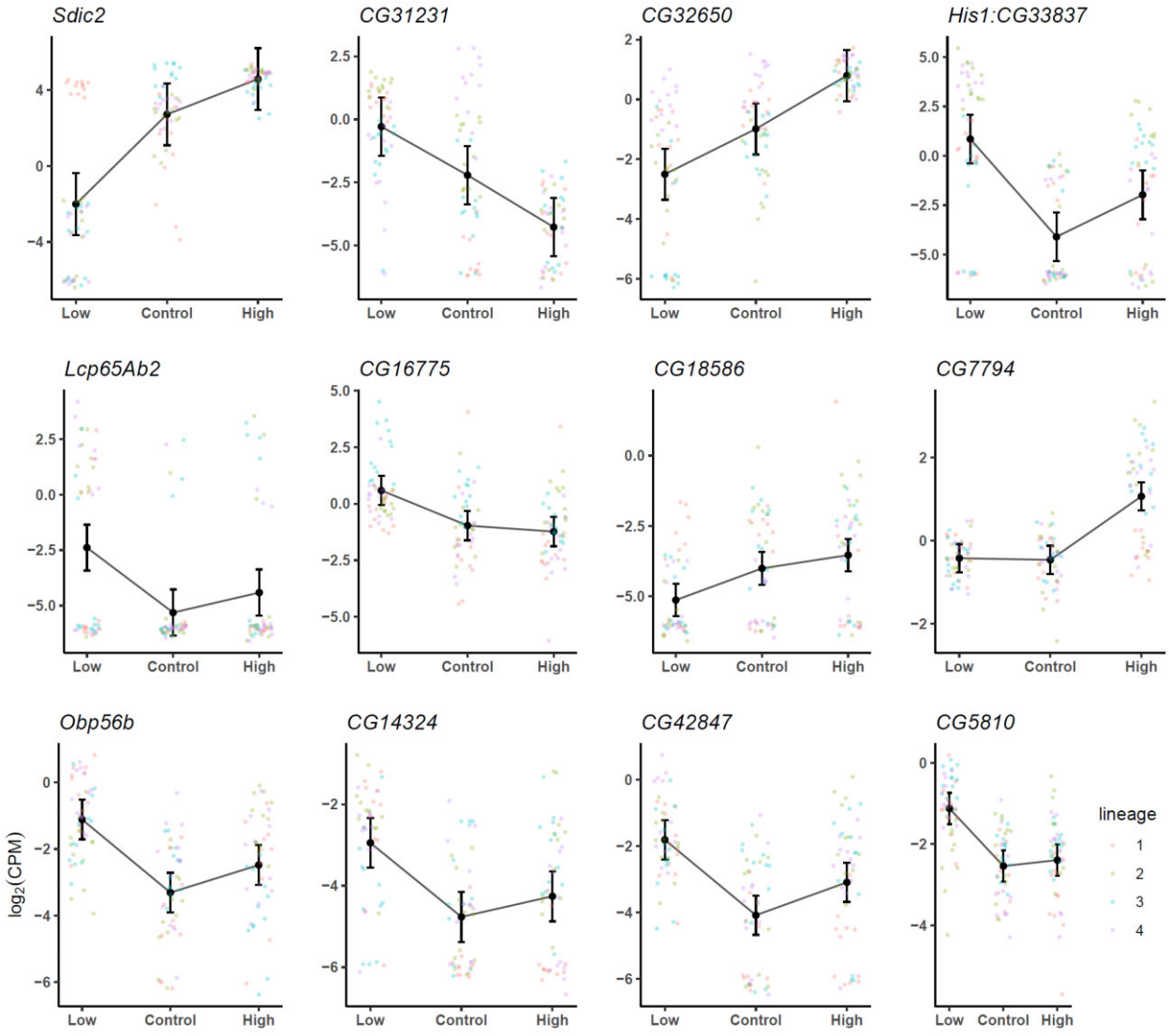

Figure 4. Reaction norms of the top 12 genes by $\log_2$(CPM) difference in the low versus high sociability contrast. Each plot shows fitted gene expression in $\log_2$(CPM) as obtained by emmeans with their 95% confidence interval. The individual points indicate the $\log_2$(CPM) of each sample, where the four colors are the four lineages of each treatment.

selection conditions in our data (Figure S7). Based on our simulations, we expected to observe a median of three overlapping genes, with the 95th percentile being an overlap of six genes, by chance alone.

Second, we identified *Drosophila* homologs of 68 out of the 115 genes reported by Wang et al. (2022) and found two genes that appeared in our DE gene list, *yellow-c* and *CG43066* (Figure S8). Based on our simulations, we expected to observe a median of five genes with the 95th percentile being an overlap of nine genes by chance alone.

Third, of the 212 genes identified by Woodard et al. (2011), we found four orthologs in *Drosophila* that were DE among our contrasts (Figure S9). Based on our simulations, we expected to observe a median of nine genes with the 95th percentile being an overlap of 14 genes by chance alone.

Finally, from the list of 1,057 genes identified by Shpigler et al. (2017), we found 14 orthologs in *Drosophila* that appeared in any of our low versus high, low versus control, or high versus control DE gene lists (Figure S10). Based on our simulations, we expected to observe a median of 45 genes with the 95th percentile being an overlap of 55 genes by chance alone.

DTU analysis

Following DTU analysis, we obtained a list of genes (and their corresponding transcripts) that were differentially transcribed. As a confirmation that our approach was detecting known sex-specific patterns, we looked at DTU between females and males and saw 2,631 genes with DTU, including the known sex determination genes in *D. melanogaster* (i.e., *doublesex* and *transformer*). In the low versus high selection contrast, we saw 191 genes with DTU (Figure 6; Figure S11, Supplementary Files 4, 5). In the low versus control selection contrast, we saw 384 genes with DTU, and in the control versus high selection contrast, we saw 252 genes with DTU. In total, we saw 619 genes overlap between all three selection contrasts. When looking at a sex-by-selection interaction, we found 14 genes with DTU. When comparing our

Table 1. Manually curated list of genes with relevant phenotypes.

FBgnID	Gene	Estimate	*p*-value	Phenotype
FBgn0010222	*Nmdmc*	0.686	1.45×10^{-5}	Abnormal locomotor behavior and stress response
FBgn0015773	*NetA*	0.654	.0011	Abnormal neuroanatomy and involved in axon guidance
FBgn0033885	*DJ-1α*	–0.624	1.47×10^{-4}	Abnormal locomotor behavior, neuroanatomy, and dopaminergic neuron
FBgn0036150	*Ir68a*	–0.518	.0021	Abnormal behavior and involved in sensory neurons
FBgn0030795	*ppk28*	0.380	5.3×10^{-4}	Abnormal memory, neurophysiology, and taste perception
FBgn0037217	*CG14636*	0.349	.0016	Abnormal auditory perception
FBgn0031435	*Elba2*	0.311	.0488	Abnormal locomotor behavior
FBgn0027783	*SMC2*	–0.285	.0015	Abnormal neuroanatomy
FBgn0016672	*Ipp*	–0.254	8.82×10^{-6}	Abnormal learning in males and abnormal neurophysiology
FBgn0261563	*wb*	0.241	.0093	Abnormal neuroanatomy
FBgn0003174	*pwn*	–0.223	1.4×10^{-4}	Abnormal neurophysiology
FBgn0266670	*Sec5*	–0.220	1.5×10^{-4}	Abnormal developmental rate, neuroanatomy, and size
FBgn0000565	*MsrA*	0.215	.0170	Involved in neuron projection
FBgn0032701	*CG10341*	–0.190	.0120	Abnormal neuroanatomy
FBgn0003654	*sw*	0.189	.0007	Abnormal neuroanatomy, paralytic, and dendritic arborizing neuron
FBgn0035464	*PIG-B*	–0.188	1.22×10^{-5}	Abnormal locomotor behavior
FBgn0030932	*Ggt-1*	–0.187	.0018	Abnormal behavior
FBgn0030969	*Usp39*	–0.176	2.95×10^{-5}	Abnormal locomotor behavior
FBgn0266418	*wake*	0.167	.0017	Abnormal locomotor, courtship behavior, and abnormal sleep
FBgn0003301	*rut*	–0.144	.0208	Abnormal behavior, locomotor behavior, neurophysiology, and neuroanatomy
FBgn0034585	*Rbpn-5*	–0.141	2.36×10^{-5}	Abnormal developmental rate and locomotor behavior
FBgn0052982	*CG32982*	0.139	9.9×10^{-4}	Abnormal locomotor behavior
FBgn0029992	*Upf2*	0.138	.0011	Abnormal neurophysiology
FBgn0260635	*Diap1*	–0.131	5.37×10^{-7}	Abnormal neuroanatomy, oxidative stress response, larval neurons, peptidergic neurons, abnormal size, and cell death
FBgn0030352	*sicily*	–0.130	1.27×10^{-4}	Abnormal neuroanatomy and neurophysiology
FBgn0026083	*tyf*	–0.114	1.28×10^{-5}	Abnormal circadian behavior and rhythm and abnormal locomotor rhythm
FBgn0001316	klar	0.110	.0153	Abnormal locomotor
FBgn0023095	*caps*	0.108	.0201	Abnormal neuroanatomy and axon guidance
FBgn0039861	*pasha*	–0.0988	.0088	Abnormal neuroanatomy and neurophysiology
FBgn0037574	*Coq2*	–0.0957	.0016	Abnormal locomotor rhythm
FBgn0024179	*wit*	–0.0858	.0303	Abnormal neurophysiology and neuroanatomy
FBgn0032222	*Cox10*	–0.0762	.0183	Abnormal locomotor behavior
FBgn0039635	*Pdhb*	–0.0728	.0232	Abnormal locomotor behavior

Note. The list contains FlyBase ID (FBgnID), gene name, low versus high sociability contrast estimate, *p*-value, and a brief description of phenotypes reported on FlyBase for different alleles of the gene.

DTU results back to the DGE results, we saw 39 genes that appear in both the overall DGE list and the overall DTU list (Figure S12, Supplementary File 6). For none of these comparisons did we observe biological ontology processes showing significant over-enrichment. Nonetheless, we observed that a substantial number of genes with known neuronal or behavioral functions show evidence of DTU (e.g., *lilli*, *Pyrokinin 1 receptor*, *Focal adhesion kinase*, *Optix*, *Rat1*, *Goosecoid*, *ben*, *Endophilin A*, *wacky*, *Proctolin*, *Lar*, *hattifattener*, *boule*, and *Bacchus*). A number of genes that influence circadian rhythm showed DTU as well. One transcript (*tim-RR*/FBtr0333258) of the *timeless* (*tim*) gene showed an average ~4 fold greater abundance of transcript in the low relative to high sociability treatment. We observed a similar pattern (albeit with a smaller, 1.25 fold increase in transcript usage in low versus high sociability) for the *Shaker cognate w* transcript, *Shaw-RA*. *Shaw* encodes a voltage-gated potassium channel, and its misregulation results in altered locomotor rhythms (Buhl et al., 2016). The *GABA-B-R3-RG* transcript of the *GABA-B-R3* gene showed similar differences in expression (magnitude and direction) to *Shaw*, and knockdown of its function leads to misregulation of sleep and the circadian clock (Dahdal et al., 2010; Haynes et al., 2015). In the high sociability treatment, there was modest upregulation (1.2 fold) of the *qvr-RC* transcript of *quiver*, known to interact with *shaker*, influencing sleep (Koh et al., 2008). Finally, the low sociability treatment showed a modest increase (~1.2 fold) in the amount of the *CkIIbeta-RI* transcript of the *CkIIbeta* gene (Konopka et al., 1991).

Candidate gene validation

Ten of the sixteen candidate genes that we functionally tested via RNAi-mediated gene knockdown showed significant (accounting for multiple comparison) differences in sociability (Figure 7, Figure S13). This included several with substantial effects, most notably *Sec5*, which showed ~50% reduction in the sociability score when knocked down. Averaged over sex, only 6 of these 10 gene knockdowns showed differences in sociability in the direction predicted by gene expression changes from the artificial selection experiment (Figure 7). This is in part due to gene knockdowns for several candidate genes having sex-specific effects on sociability (Figures S14 and S15). For example, *Est-P* knockdown females showed much higher sociability than control females, whereas the male treatments showed modest differences (Figures S14 and S15). Notably, CG31231 knockdowns showed an increase

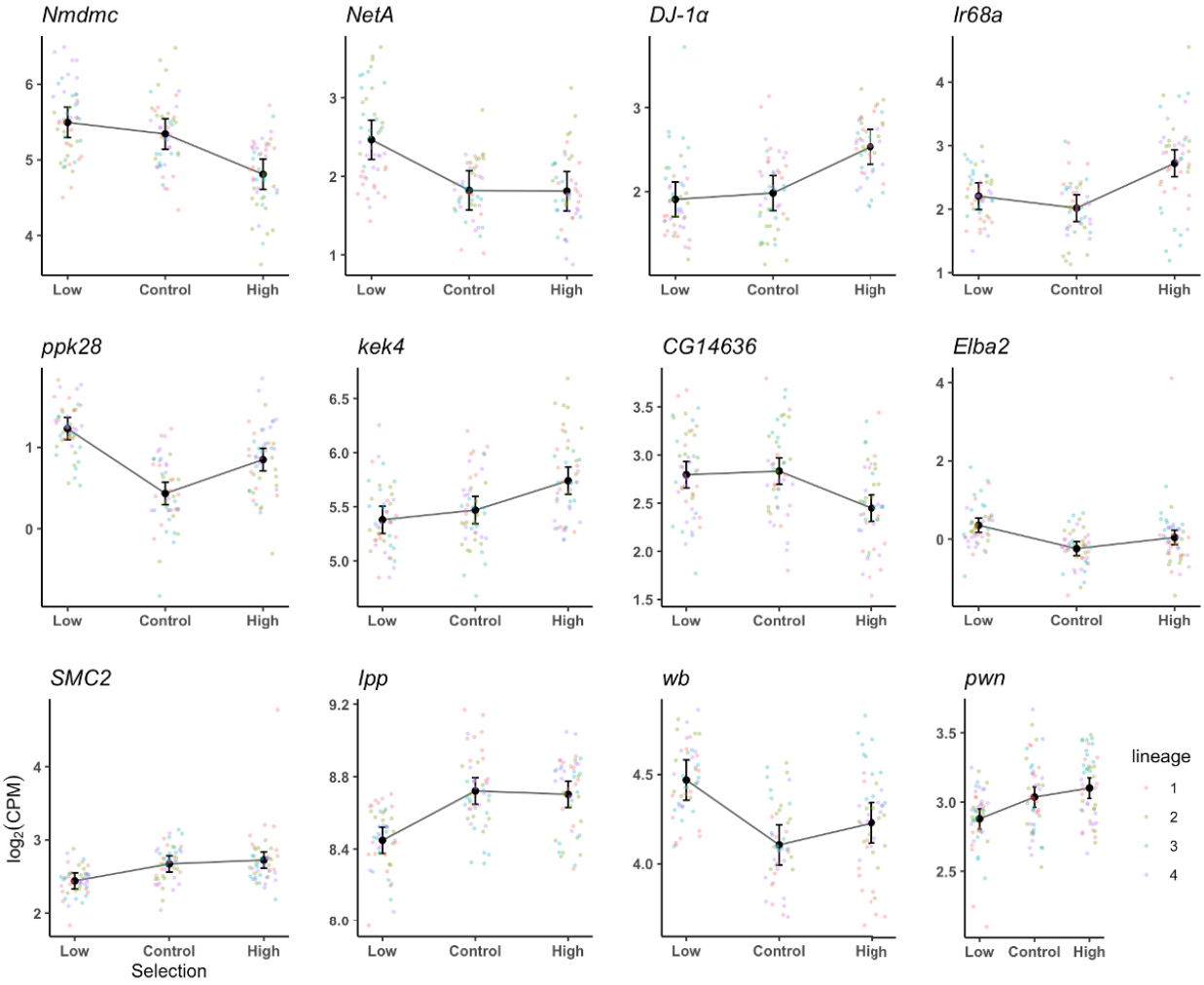

Figure 5. Reaction norms of the top 12 genes with relevant phenotypes, including neuroanatomy, neurophysiology, locomotor behavior, or circadian rhythm (Table 1), by $\log_2$(CPM) difference in the low versus high sociability contrast. Each plot shows fitted gene expression in $\log_2$(CPM) as obtained by emmeans with their 95% confidence interval. The individual points indicate the $\log_2$(CPM) of each sample, where the four colors are the four lineages of each treatment.

in sociability in males, and a decrease in females, relative to their controls.

Discussion

Our overarching goal was to identify genes underlying natural variation in sociability that ultimately may contribute to evolutionary changes in this behavior. To this end, we relied on gene expression data from our artificial selection on sociability (Scott et al., 2022a), followed up by experiments on knockdowns of top candidate genes to validate their direct effect on sociability.

Gene expression

Utilizing a linear mixed model framework, followed by planned contrasts between artificial selection treatments, we identified 327 genes showing differential expression across all three selection contrasts, with 174 attributed to the low versus high selection contrast. While the maximum (minimum) change in gene expression was 3.98 (−6.59) in $\log_2$(CPM), the majority (159/174) of DE genes showed more modest changes between −1 and 1 $\log_2$(CPM) (Figure S16). Fold changes in expression, however, are rarely proportional to either phenotypic changes (Dworkin et al., 2009) or causality. Exclusively between the low versus control and high versus control DE gene lists, we see 133 DE genes that overlap between the two contrasts (Figure S17). While it cannot be stated unequivocally, the fact that many genes, associated with a broad array of biological functions, showed changes in expression is broadly consistent with the underlying genetic response being polygenic. That is, gene expression variation that may modulate sociability in *Drosophila* is a result of modest changes in function of a large number of genes across a diverse set of biological processes.

Among the DE genes, we identified 33 genes with known effects on neuroanatomy, neurophysiology, or behavior (Table 1). For example, the Sec5 protein and its human ortholog, EXOC2, are part of the exocytosis complex, which is involved in membrane traffic within neurons and has critical roles in neuronal function (Evers et al., 2014; Halim et al., 2023; Martin-Urdiroz et al., 2016; Murthy et al.,

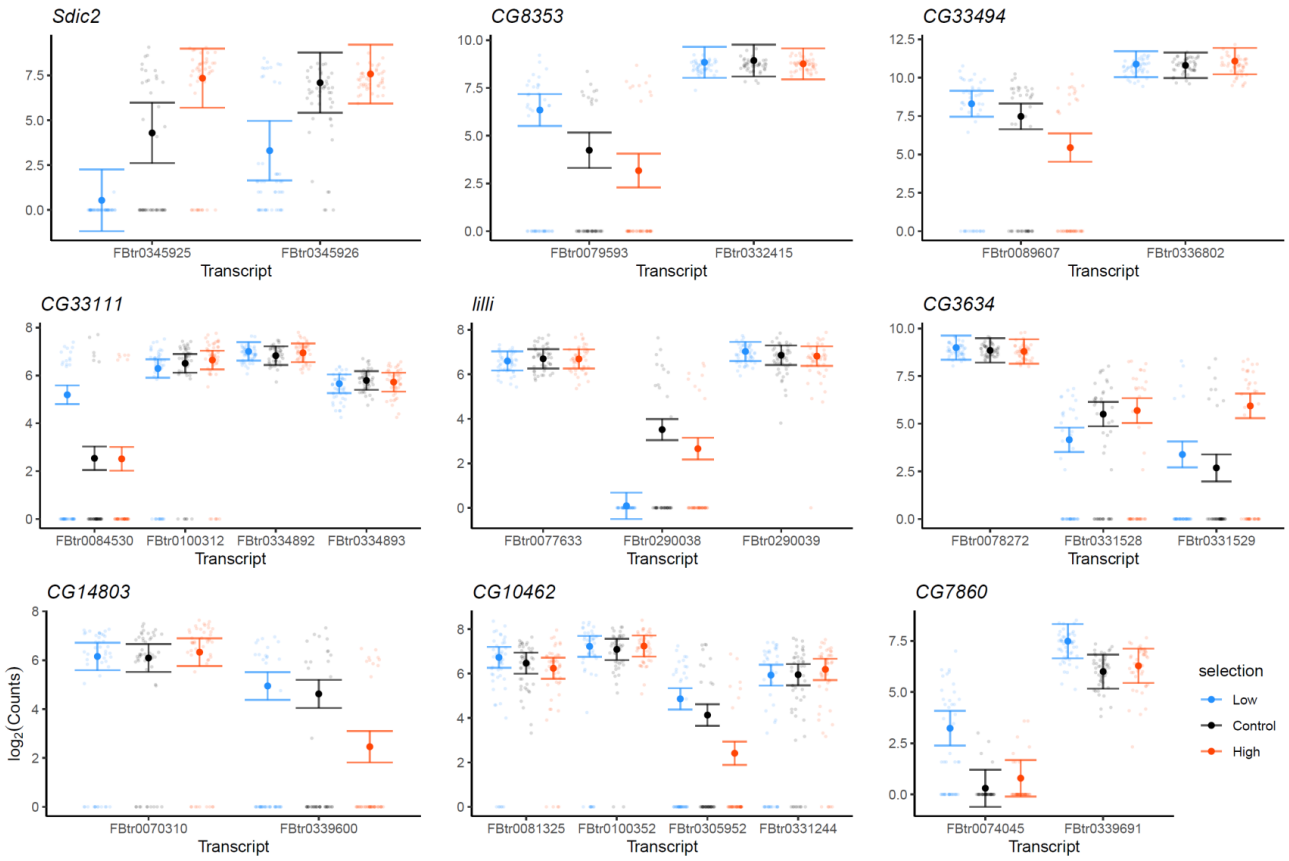

Figure 6. Reaction norms of the top nine genes by $\log_2$(CPM) difference in the low versus high sociability contrast within the differential transcript usage gene list. Three genes with a large number of transcripts are depicted in Figure S13. Each plot shows expression of the given gene and its associated transcripts. Along the *x*-axis is the transcript of a given gene, and the *y*-axis is $\log_2$(counts). Each transcript is colored by selection, with blue representing low sociability, black representing control, and red representing high sociability. The large points are the fitted expression values of a transcript as obtained by emmeans with their 95% confidence intervals. The small points are the $\log_2$(counts) of each sample.

2003; Swope et al., 2022; Van Bergen et al., 2020). Similarly, *MsrA* encodes a methionine-S-sulfoxide reductase, which is involved in responses to oxidative stress (Moskovitz et al., 2001; Roesijadi et al., 2007), and has been linked to neurological deficits, including autism spectrum disorder (Grove et al., 2019). There is currently, however, limited information about the majority of the most DE genes (Figure 4). Relevant exceptions include *Sdic2* and *Obp56b*. Although *Sdic2* (Sperm dynein intermediate chain 2) is part of a gene family assumed to be involved in sperm motility (Yeh et al., 2012), its human ortholog, *DYNC1I2* (Dynein Cytoplasmic 1 Intermediate Chain 2), is critical for neurodevelopment (Ansar et al., 2019). *Obp56b* codes for an odorant binding protein, which might be pertinent for social behavior.

The number of unique replicate lineages (4) for each of the three artificial selection treatments (low, high, and control), as well as the high sampling per unique replicate allowed us to gain some important insights on the degree of shared versus unique transcriptional responses to selection. For instance, the gene *Sdic2* demonstrated a substantial change in magnitude between treatments, where three of four low-sociability lineages showed a consistent reduction in gene expression relative to the control and high-sociability lineages. One low sociability replicate lineage, however, showed no change (Figure 4). Do these results suggest that *Sdic2* has a causal role to play, but that the sampling of allelic variation influencing gene expression was not captured in this replicate lineage (i.e., due to the interaction of selection-drift)? Or was it related via linkage disequilibrium with an unknown causal change, where recombination (along with selection-drift) resulted in the observed pattern? Further studies linking lineage-specific changes in allele frequencies with gene expression changes will provide important insights into the potential for heterogeneous response to selection in terms of what alleles are captured by selection.

Global variation in gene expression from adult heads among lineages (Figure 2) shows the expected pattern of sex-biased gene expression, and relatively modest aggregation of the low-sociability lineages. One possible explanation for the modest changes in the overall expression profiles, and a limitation of our experiment, is that relevant changes in gene expression associated with evolutionary changes in sociability can occur during brain development, but (at least at the transcript level) may not vary substantially in adults. While this would reduce the number of candidate genes identified, it is valuable, as even identifying genes that show differential expression due to changes during brain development across the treatments provides insight into the biological processes of interest.

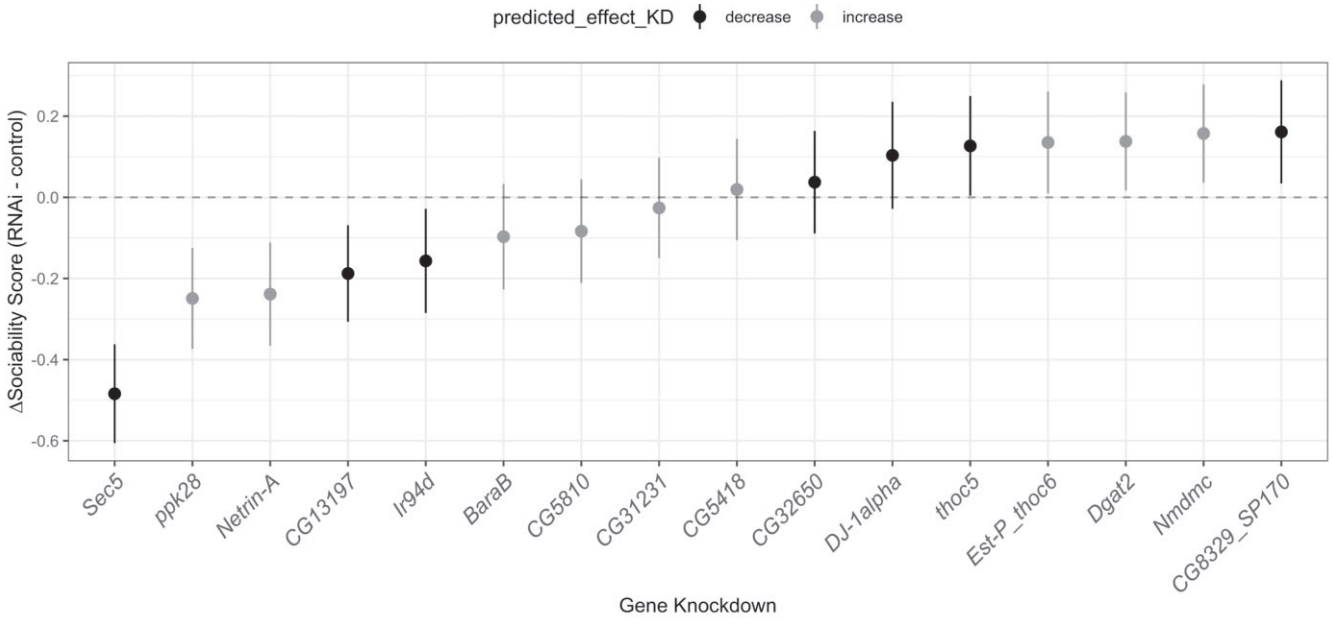

Figure 7. Treatment contrasts between RNAi knockdown (KD) and their respective controls. Contrasts are on a natural log scale, with corresponding 95% CI, adjusted for 16 comparisons to the control treatment (Dunnett adjustment). Values below zero indicate lower sociability in knockdowns. Black and grey colors indicate predicted effects of gene knockdowns on sociability, based on the gene expression data. $N = 48$ arenas for each of the four treatments (males, females, knockdown, and control), for a total of 144 arenas for each gene.

The role that alternative splicing plays in mediating the generation of phenotypic variation and contributing to adaptive divergence is becoming increasingly clear (Singh & Ahi, 2022; Verta & Jacobs, 2022; Wright et al., 2022). From our analysis, we identified hundreds of genes demonstrating DTU across the sociability treatments (Figure 6, Figure S11, Supplementary Files 4 and 5). In some cases, like *Sdic2*, it may be that DTU drives overall changes in gene expression, as in this gene, one of the transcripts (FBtr0345925) is not expressed in any of the low sociability treatment samples (Figure 6). Many of the genes that showed evidence of DTU have known neuronal and behavioral functions, including at least five genes (*tim*, *Shaw*, *GABA-B-R3*, *qvr*, and *CkIIbeta*), with known circadian or sleep-related phenotypes when misregulated. While the mutational target size for variation in circadian rhythm in *Drosophila* may be greater than previously assumed based on the "core" clock genes (Harbison et al., 2019; Kumar et al., 2021), this set of genes may suggest variation in an, as of yet unknown, regulator of splicing influencing variation in sociability, acting pleiotropically with circadian rhythm. In humans, there is a known association between disruption in circadian rhythms/sleep and dysregulation of various social behaviors, although without a clear understanding of causality (Grandin et al., 2006; Kohyama, 2014; Walker et al., 2021). Despite the known functions of many of the genes that show DTU, we did not observe significant enrichment of biological processes for the sets of genes for which we identified DTUs during GO analysis.

Candidate gene validation

Ten of the 16 genes that we tested via RNA interference showed significant effects on sociability (Figures 7 and S13). This proportion is similar to ones in comparable studies. For example, in our related work on another social behavior, sexual aggression, four of seven candidate genes tested showed significant effects on male sexual aggression toward females (Scott et al., 2022b). Intriguingly, however, 4 of the 10 genes with significant effects on sociability did so in the opposite direction than we predicted based on the gene expression data (Figure 7). Unlike the gene expression data, however, which we acquired from adult heads, the candidate gene tests, we performed involved knocking down genes broadly throughout the nervous system both in adults and during development. With respect to our experimental design for testing candidate genes, in all crosses using RNAi strains from the TRiP collection, the genetic background of the control genotypes was the same as the experimental RNAi genotypes (with the exception of the RNAi construct itself). However, RNAi strains from the VDRC collection had a different genetic background from the TRiP collection. As such, in addition to the RNAi construct itself, there were also overall genetic background differences from the controls. As discussed in the methods, this was done on purpose to avoid having individuals who were homozygous or hemizygous for a mutant allele in the *white* gene, which influences numerous behaviors. The downside of this approach was that there was potential for background-dependent effects that could confound interpretation with these strains (de Belle & Heisenberg, 1996; Mullis et al., 2018; Taylor & Ehrenreich, 2015). However, we did not see any overall differences in patterns between our results from the TRiP or VDRC collections.

There are limits to what global gene expression differences as a "snapshot"—in our case in adult heads—inform us about causal influences of individual genes on trait variation. Correlation between mRNA and protein expression is generally strong and positive (~0.6 in *Drosophila*), but less

than 1. These represent readouts of changes associated with variation in sociability, but do not capture all relevant variation in expression occurring in the adult brain (both temporally and anatomically), let alone all evolved changes associated with the developing brain and nervous system. The response to artificial selection starting with a genetically variable population is likely polygenic. Individual allelic variants likely have modest impacts, and it is aggregate effects of alleles that result in substantial phenotypic changes. As such, it is likely that some differences in gene expression associated with artificial selection are correlated changes, and the magnitude of expression differences does not necessarily reflect phenotypic impact (Dworkin et al., 2011, 2009). The consequence of this is that an unknown proportion of DE genes do not directly mediate sociability. Despite such limitations, these approaches provide insight and identify candidate genes for further research on the genetic basis of natural variation in sociability, even when, as in our case, they can provide some confusing results when validating genetic effects.

Natural genetic variation in sociabiliity

Despite over 500 million years of evolution, there remains considerable shared gene function involved with specific biological processes throughout animals. This includes core aspects of development, such as the shared roles of the Hox genes in anterior-posterior patterning and positional identity, co-option of *Pax6* (*eyeless*) and *Distal-less* orthologs during the repeated evolution of complex eyes and limbs across taxa (Kozmik, 2005; Panganiban et al., 1997; Quiring et al., 1994), shared role for orthologs of *tinman* in heart development (Bodmer, 1995), and numerous genes involved with nervous system development that appear to be shared across phyla (Freeman & Doherty, 2006; Holland et al., 2013; Lichtneckert & Reichert, 2005). Despite rapid evolution of primary sex-determination signals, many species across multiple phyla utilize *doublesex/mab3* genes (albeit a diverse gene family) as part of the sex-determination cascade (Bachtrog et al., 2014; Haag & Doty, 2005; Kopp, 2012). Taken together, these findings suggest re-utilization or homologous functioning of gene pathways in these processes, despite extensive evolutionary diversification being common for many phenotypes. Despite some distinct features of behaviors, there is some tantalizing evidence for shared function of genes related to feeding/foraging (Fitzpatrick & Sokolowski, 2004), circadian rhythm (Chong et al., 2012), and more recently, for aspects of social behavior shared between hymenopterans (arthropods) and humans (chordates) (Liu et al., 2016; Shpigler et al., 2017; Wang et al., 2022). As discussed in the sections above, orthologs of several of our candidate sociability genes have been linked to social behavior in humans. However, our formal comparisons between our study and a few other relevant studies on the genetics of social behavior (Shpigler et al., 2017; Woodard et al., 2011) did not reveal more shared genes than expected by chance, based on our simulations (Figures S6–S9). The most likely explanation for these results is that social behavior is a broad category, which includes all types of interactions among conspecifics. Examples include parameters of social networks (Bentzur et al., 2021; Schneider et al., 2012; Wice & Saltz, 2021), inter-individual distance (Simon et al., 2012), social effects on oviposition (Bailly et al., 2023; Fowler et al., 2022), nursing behavior in honey bees (Shpigler et al., 2017), and effects of social isolation (Wang et al., 2022). Effectively, there are additional challenges in establishing comparisons among seemingly similar behaviors across taxa, as compared to anatomical traits like hearts and eyes.

We have deliberately chosen to focus our research on a core aspect of social behavior, sociability, defined as the tendency to affiliate with conspecifics. Specifically, we designed an ecologically realistic apparatus (Dukas, 2020) in which flies could decide to join, stay with, or evict others from food patches (Scott et al., 2022a). Using this measure allows us to quantify how strongly flies prefer to affiliate with others while feeding or resting. Such preference to engage in friendly activities with conspecifics is prevalent among animals, including humans (Allee, 1938; Ward & Webster, 2016; Wilson, 1975). It is thus possible that there are shared genetic networks underlying the narrowly defined trait of sociability among animals. We would need many more studies on the genetics of sociability as well as of distinct social-behavior traits in order to possess a deeper understanding of the genetic architecture of natural variation in social behavior in general and sociability in particular. In ongoing work in our labs, we investigate the changes in allele frequencies that accompanied the evolution of sociability, validate the effects on sociability of candidate genes identified in our population genomics analyses, and quantify the dynamics of group formation in normal and socially deficient flies. We also hope that our publicly available data sets will facilitate research in other labs on the evolutionary biology of sociability.

Supplementary material

Supplementary material is available online at *Evolution*.

Data availability

Intermediate data, scripts, and outputs are available at GitHub: https://github.com/DworkinLab/DrosophilaSociabilityTranscriptomics, and a static version is available at FigShare: https://doi.org/10.6084/m9.figshare.29657138.v1. Raw sequence data will be available on NCBI SRA (BioProject PRJNA1311514, sample accessions SAMN50809994 to SAMN50810135).

Author contributions

A.M.S., I.D., and R.D. designed, and A.M.S. ran the artificial selection experiment. A.M.S. extracted RNA, and A.T.M., and I.D. conducted genomic analyses. I.D., D.D., and R.D. designed, and D.D. conducted gene validation tests. A.T.M., D.D., R.D., and I.D. wrote the manuscript. R.D. and I.D. revised the manuscript.

Funding

This work was supported by the Natural Sciences and Engineering Research Council of Canada, RGPIN-2019-05420, Canada Foundation for Innovation, and Ontario Ministry of Research and Innovation.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgments

We thank V. Ricciardi, A. Roth, S. Abdullah, B. Luo, V. Sugunanavalan, and C. Kloppenburg for assistance in the gene validation experiments, A. Moehring and three anonymous referees for perceptive comments on the manuscript.

References

Alexa, A., & Rahnenfuhrer, J. (2022). topGO: Enrichment analysis for gene ontology: R package version 2.48.0. https://bioconductor.org/packages/release/bioc/html/topGO.htmlDate accessed January 10, 2024.

Allee, W. C. (1938). *The social life of animals*. Norton. https://doi.org/10.5962/bhl.title.7226

Anders, S., Reyes, A., & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research*, *22*(10), 2008–2017. https://doi.org/10.1101/gr.133744.111

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/Date accessed December 14, 2023.

Ansar, M., Ullah, F., Paracha, S. A., Adams, D. J., Lai, A., Pais, L., Iwaszkiewicz, J., Millan, F., Sarwar, M. T., Agha, Z., Shah, S. F., Qaisar, A. A., Falconnet, E., Zoete, V., Ranza, E., Makrythanasis, P., Santoni, F. A., Ahmed, J., Katsanis, N., … Antonarakis, S. E. (2019). Bi-allelic variants in DYNC1I2 cause syndromic microcephaly with intellectual disability, cerebral malformations, and dysmorphic facial features. *The American Journal of Human Genetics*, *104*(6), 1073–1087. https://doi.org/10.1016/j.ajhg.2019.04.002

Arbeitman, M. N., New, F. N., Fear, J. M., Howard, T. S., Dalton, J. E., & Graze, R. M. (2016). Sex differences in *Drosophila* somatic gene expression: Variation and regulation by doublesex. *G3 Genes|Genomes|Genetics*, *6*(7), 1799–1808. https://doi.org/10.1534/g3.116.027961

Bachtrog, D., Mank, J. E., Peichel, C. L., Kirkpatrick, M., Otto, S. P., Ashman, T. L., Hahn, M. W., Kitano, J., Mayrose, I., Ming, R., Perrin, N., Ross, L., Valenzuela, N., & Vamosi, J. C., The Tree of Sex, C. (2014). Sex determination: Why so many ways of doing it? *PLoS Biology*, *12*(7), e1001899. https://doi.org/10.1371/journal.pbio.1001899

Bailly, T. P. M., Kohlmeier, P., Etienne, R. S., Wertheim, B., & Billeter, J. C. (2023). Social modulation of oogenesis and egg laying in *Drosophila melanogaster*. *Current Biology*, *33*, 2865–2877.e4. https://doi.org/10.1016/j.cub.2023.05.074

Battesti, M., Moreno, C., Joly, D., & Mery, F. (2012). Spread of social information and dynamics of social transmission within *Drosophila* groups. *Current Biology*, *22*, 309–313. https://doi.org/10.1016/j.cub.2011.12.050

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, *29*(4), 1165–1188. https://doi.org/10.1214/aos/1013699998

Bentzur, A., Ben-Shaanan, S., Benichou, J. I. C., Costi, E., Levi, M., Ilany, A., & Shohat-Ophir, G. (2021). Early life experience shapes male behavior and social networks in *Drosophila*. *Current Biology*, *31*, 486–501.e3. https://doi.org/10.1016/j.cub.2020.10.060

Billeter, J. C., Bailly, T. P. M., & Kohlmeier, P. (2025). The social life of *Drosophila melanogaster*. *Insectes Sociaux*, *72*, 127–140, . https://doi.org/10.1007/s00040-024-00990-3

Bodmer, R. (1995). Heart development in *Drosophila* and its relationship to vertebrates. *Trends in Cardiovascular Medicine*, *5*(1), 21–28. https://doi.org/10.1016/1050-1738(94)00032-Q

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bond, M. L., Lee, D. E., Farine, D. R., Ozgul, A., & König, B. (2021). Sociability increases survival of adult female giraffes. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1944), 20202770. https://doi.org/10.1098/rspb.2020.2770

Bralten, J., Mota, N. R., Klemann, C. J. H. M., De Witte, W., Laing, E., Collier, D. A., de Kluiver, H., Bauduin, S. E. E. C., Arango, C., Ayuso-Mateos, J. L., Fabbri, C., Kas, M. J., van der Wee, N., Penninx, B. W. J. H., Serretti, A., Franke, B., & Poelmans, G. (2021). Genetic underpinnings of sociability in the general population. *Neuropsychopharmacology*, *46*(9), 1627–1634. https://doi.org/10.1038/s41386-021-01044-z

Brooks, M. E., Kristensen, K., Benthem, K. J. V., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*, 378–400. https://doi.org/10.32614/RJ-2017-066

Buhl, E., Bradlaugh, A., Ogueta, M., Chen, K. F., Stanewsky, R., & Hodge, J. J. L. (2016). Quasimodo mediates daily and acute light effects on *Drosophila* clock neuron excitability. *Proceedings of the National Academy of Sciences*, *113*(47), 13486–13491. https://doi.org/10.1073/pnas.1606547113

Campos, F. A., Villavicencio, F., Archie, E. A., Colchero, F., & Alberts, S. C. (2020). Social bonds, social status and survival in wild baboons: A tale of two sexes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1811), 20190621. https://doi.org/10.1098/rstb.2019.0621

Chandler, C. H., Chari, S., & Dworkin, I. (2013). Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. *Trends in Genetics*, *29*(6), 358–366. https://doi.org/10.1016/j.tig.2013.01.009

Chen, M., & Sokolowski, M. B. (2022). How social experience and environment impacts behavioural plasticity in *Drosophila*. *Flyer*, *16*(1), 68–84. https://doi.org/10.1080/19336934.2021.1989248

Chen, Y., Lun, A., & Smyth, G. (2016). From reads to genes to pathways: Differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline [version 2; peer review: 5 approved]. *F1000Research*, *5*(1438), 1438. https://doi.org/10.12688/f1000research.8987.2

Chong, S. Y. C., Ptáček, L. J., & Fu, Y. H. (2012). Genetic insights on sleep schedules: This time, it's PERsonal. *Trends in Genetics*, *28*(12), 598–605. https://doi.org/10.1016/j.tig.2012.08.002

Clutton-Brock, T. (2016). *Mammal societies*. Wiley.

Clyde, D. (2018). The genetics of loneliness. *Nature Reviews Genetics*, *19*(9), 532–533. https://doi.org/10.1038/s41576-018-0036-8

Costa, J. T. (2006). *The other insect societies*. Harvard University Press.

Dahdal, D., Reeves, D. C., Ruben, M., Akabas, M. H., & Blau, J. (2010). *Drosophila* pacemaker neurons require g protein signaling and GABAergic inputs to generate twenty-four hour behavioral rhythms. *Neuron*, *68*(5), 964–977. https://doi.org/10.1016/j.neuron.2010.11.017

Dal Pesco, F., Trede, F., Zinner, D., & Fischer, J. (2022). Male–male social bonding, coalitionary support and reproductive success in wild Guinea baboons. *Proceedings of the Royal Society B: Biological Sciences*, *289*(1975), 20220347. https://doi.org/10.1098/rspb.2022.0347

Day, F. R., Ong, K. K., & Perry, J. R. B. (2018). Elucidating the genetic basis of social interaction and isolation. *Nature Communications*, *9*(1), 2457. https://doi.org/10.1038/s41467-018-04930-1

de Belle, J. S., & Heisenberg, M. (1996). Expression of *Drosophila* mushroom body mutations in alternative genetic backgrounds: A case study of the mushroom body miniature gene (mbm). *Proceedings of the National Academy of Sciences*, *93*(18), 9875–9880. https://doi.org/10.1073/pnas.93.18.9875

de la Torre-Ubieta, L., Won, H., Stein, J. L., & Geschwind, D. H. (2016). Advancing the understanding of autism disease mechanisms through

genetics. *Nature Medicine*, *22*(4), 345–361. https://doi.org/10.1038/nm.4071

DeLuca, D. S., Levin, J. Z., Sivachenko, A., Fennell, T., Nazaire, M. D., Williams, C., Reich, M., Winckler, W., & Getz, G. (2012). RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics*, *28*(11), 1530–1532. https://doi.org/10.1093/bioinformatics/bts196

Dietzl, G., Chen, D., Schnorrer, F., Su, K. C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., & Dickson, B. J. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, *448*(7150), 151–156. https://doi.org/10.1038/nature05954

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dukas, R. (2020). Natural history of social and sexual behavior in fruit flies. *Scientific Reports*, *10*, 21932. https://doi.org/10.1038/s41598-020-79075-7

Durisko, Z., & Dukas, R. (2013). Attraction to and learning from social cues in fruit fly larvae. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 20131398. https://doi.org/10.1098/rspb.2013.1398

Durisko, Z., Kemp, B., Mubasher, A., & Dukas, R. (2014). Dynamics of social interactions in fruit fly larvae. *PLoS One*, *9*(4), e95495. https://doi.org/10.1371/journal.pone.009549

Dworkin, I., Anderson, J. A., Idaghdour, Y., Parker, E. K., Stone, E. A., & Gibson, G. (2011). The effects of weak genetic perturbations on the transcriptome of the wing imaginal disc and its association with wing shape in *Drosophila melanogaster*. *Genetics*, *187*(4), 1171–1184. https://doi.org/10.1534/genetics.110.125922

Dworkin, I., Kennerly, E., Tack, D., Hutchinson, J., Brown, J., Mahaffey, J., & Gibson, G. (2009). Genomic consequences of background effects on scalloped mutant expressivity in the wing of *Drosophila melanogaster*. *Genetics*, *181*(3), 1065–1076. https://doi.org/10.1534/genetics.108.096453

Elovainio, M., Hakulinen, C., Pulkki-Råback, L., Virtanen, M., Josefsson, K., Jokela, M., Vahtera, J., & Kivimäki, M. (2017). Contribution of risk factors to excess mortality in isolated and lonely individuals: An analysis of data from the UK Biobank cohort study. *The Lancet Public Health*, *2*(6), e260–e266. https://doi.org/10.1016/S2468-2667(17)30075-0

Evers, C., Maas, B., Koch, K. A., Jauch, A., Janssen, J. W. G., Sutter, C., Parker, M. J., Hinderhofer, K., & Moog, U. (2014). Mosaic deletion of EXOC6B: Further evidence for an important role of the exocyst complex in the pathogenesis of intellectual disability. *American Journal of Medical Genetics Part A*, *164*(12), 3088–3094. https://doi.org/10.1002/ajmg.a.36770

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Ferreira, C. H., & Moita, M. A. (2020). Behavioral and neuronal underpinnings of safety in numbers in fruit flies. *Nature Communications*, *11*(1), 4182. https://doi.org/10.1038/s41467-020-17856-4

Fitzpatrick, M. J. (2004). In search of food: Exploring the evolutionary link between cGMP-dependent protein kinase (pkg) and behaviour. *Integrative and Comparative Biology*, *44*(1), 28–36. https://doi.org/10.1093/icb/44.1.28

Fowler, E. K., Leigh, S., Rostant, W. G., Thomas, A., Bretman, A., & Chapman, T. (2022). Memory of social experience affects female fecundity via perception of fly deposits. *BMC Biology*, *20*(1), 244. https://doi.org/10.1186/s12915-022-01438-5

Fox, J., & Weisberg, S. (2019). *An r companion to applied regression*. 3rd ed. Sage Publications.

Freeman, M. R., & Doherty, J. (2006). Glial cell biology in *Drosophila* and vertebrates. *Trends in Neurosciences*, *29*(2), 82–90. https://doi.org/10.1016/j.tins.2005.12.002

Gerber, L., Connor, R. C., Allen, S. J., Horlacher, K., King, S. L., Sherwin, W. B., Willems, E. P., Wittwer, S., & Krützen, M. (2022). Social integration influences fitness in allied male dolphins. *Current Biology*, *32*(7), 1664–1669.e3. https://doi.org/10.1016/j.cub.2022.03.027

Gramates, L. S., Agapite, J., Attrill, H., Calvi, B. R., Crosby, M. A., dos Santos, G., Goodman, J. L., Goutte-Gattat, D., Jenkins, V. K., Kaufman, T., Larkin, A., Matthews, B. B., Millburn, G., Strelets, V. B., Perrimon, N., Gelbart, S. R., Agapite, J., Broll, K., Crosby, L., … Lovato, T. (2022). FlyBase: A guided tour of highlighted features. *Genetics*, *220*(4), iyac035. https://doi.org/10.1093/genetics/iyac035

Grandin, L. D., Alloy, L. B., & Abramson, L. Y. (2006). The social zeitgeber theory, circadian rhythms, and mood disorders: Review and evaluation. *Clinical Psychology Review*, *26*(6), 679–694. https://doi.org/10.1016/j.cpr.2006.07.001

Grove, J., Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J. H., … Børglum, A. D. (2019). Identification of common genetic risk variants for autism spectrum disorder. *Nature Genetics*, *51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Haag, E. S., & Doty, A. V. (2005). Sex determination across evolution: Connecting the dots. *PLoS Biology*, *3*(1), e21. https://doi.org/10.1371/journal.pbio.0030021

Halim, D. O., Munson, M., & Gao, F. -B. (2023). The exocyst complex in neurological disorders. *Human Genetics*, *142*, 1263–1270. https://doi.org/10.1007/s00439-023-02558-w

Harbison, S. T., Kumar, S., Huang, W., McCoy, L. J., Smith, K. R., & Mackay, T. F. C. (2019). Genome-wide association study of circadian behavior in *Drosophila melanogaster*. *Behavior Genetics*, *49*(1), 60–82. https://doi.org/10.1007/s10519-018-9932-0

Haynes, P. R., Christmann, B. L., & Griffith, L. C. (2015). A single pair of neurons links sleep to memory consolidation in *Drosophila melanogaster*. *eLife*, *4*, e03868. https://doi.org/10.7554/eLife.03868

Holland, L. Z., Carvalho, J. E., Escriva, H., Laudet, V., Schubert, M., Shimeld, S. M., & Yu, J. K. (2013). Evolution of bilaterian central nervous systems: A single origin? *EvoDevo*, *4*(1), 27. https://doi.org/10.1186/2041-9139-4-27

Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, *10*(2), 227–237. https://doi.org/10.1177/1745691614568352

House, J. S., Landis, K. R., & Umberson, D. (1988). Social relationships and health. *Science*, *241*(4865), 540–545. https://doi.org/10.1126/science.3399889

Kajokaite, K., Whalen, A., Koster, J., & Perry, S. (2022). Social integration predicts survival in female white-faced capuchin monkeys. *Behavioral Ecology*, *33*(4), 807–815. https://doi.org/10.1093/beheco/arac043

Khodursky, S., Svetec, N., Durkin, S. M., & Zhao, L. (2020). The evolution of sex-biased gene expression in the *Drosophila* brain. *Genome Research*, *30*(6), 874–884. https://doi.org/10.1101/gr.259069.119

Kocher, S. D., Mallarino, R., Rubin, B. E. R., Yu, D. W., Hoekstra, H. E., & Pierce, N. E. (2018). The genetic basis of a social polymorphism in halictid bees. *Nature Communications*, *9*(1), 1–7. https://doi.org/10.1038/s41467-018-06824-8

Koh, K., Joiner, W. J., Wu, M. N., Yue, Z., Smith, C. J., & Sehgal, A. (2008). Identification of SLEEPLESS, a sleep-promoting factor. *Science*, *321*(5887), 372–376. https://doi.org/10.1126/science.1155942

Kohyama, J. (2014). The possible long-term effects of early-life circadian rhythm disturbance on social behavior. *Expert Review of Neu-*

rotherapeutics, *14*(7), 745–755. https://doi.org/10.1586/14737175.2014.927735

Konopka, R. J., Smith, R. F., & Orr, D. (1991). Characterization of andante, a new *Drosophila* clock mutant, and its interactions with other clock mutants. *Journal of Neurogenetics*, *7*(2-3), 103–114. https://doi.org/10.3109/01677069109066214

Kopp, A. (2012). Dmrt genes in the development and evolution of sexual dimorphism. *Trends in Genetics*, *28*(4), 175–184. https://doi.org/10.1016/j.tig.2012.02.002

Kozmik, Z. (2005). Pax genes in eye development and evolution. *Current Opinion in Genetics & Development*, *15*(4), 430–438. https://doi.org/10.1016/j.gde.2005.05.001

Krebs, C. J. (1999). *Ecological methodology*. (2nd ed.). Addison-Wesley.

Kumar, S., Tunc, I., Tansey, T. R., Pirooznia, M., & Harbison, S. T. (2021). Identification of genes contributing to a long circadian period in *Drosophila melanogaster*. *Journal of Biological Rhythms*, *36*(3), 239–253. https://doi.org/10.1177/0748730420975946

Lenth, R. V. (2022). emmeans: Estimated marginal means, aka least-squares means: R package version 1.8.1-1. https://CRAN.R-project.org/package=emmeansDate accessed January 10, 2024.

Levine, J. D., Funes, P., Dowse, H. B., & Hall, J. C. (2002). Resetting the circadian clock by social experience in *Drosophila melanogaster*. *Science*, *298*(5600), 2010–2012. https://doi.org/10.1126/science.1076008

Liang, Y. Y., Zhou, M., He, Y. u, Zhang, W., Wu, Q., Luo, T., Zhang, J., Jia, F., Qi, L. u, Ai, S., & Zhang, J. (2024). Observational and genetic evidence disagree on the association between loneliness and risk of multiple diseases. *Nature Human Behaviour*, *8*(11), 2209–2221. https://doi.org/10.1038/s41562-024-01970-0

Lichtneckert, R., & Reichert, H. (2005). Insights into the urbilaterian brain: Conserved genetic patterning mechanisms in insect and vertebrate brain development. *Heredity*, *94*(5), 465–477. https://doi.org/10.1038/sj.hdy.6800664

Liu, H., Robinson, G. E., & Jakobsson, E. (2016). Conservation in mammals of genes associated with aggression-related behavioral phenotypes in honey bees. *PLOS Computational Biology*, *12*(6), e1004921. https://doi.org/10.1371/journal.pcbi.1004921

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Love, M. I., Soneson, C., & Patro, R. (2018). Swimming downstream: Statistical analysis of differential transcript usage following Salmon quantification. *F1000Research*, *7*, 952. https://doi.org/10.12688/f1000research.15398.3

Martin-Urdiroz, M., Deeks, M. J., Horton, C. G., Dawe, H. R., & Jourdain, I. (2016). The exocyst complex in health and disease. *Frontiers in Cell and Developmental Biology*, *4*: 24. https://doi.org/10.3389/fcell.2016.00024

Moskovitz, J., Bar-Noy, S., Williams, W. M., Requena, J., Berlett, B. S., & Stadtman, E. R. (2001). Methionine sulfoxide reductase (MsrA) is a regulator of antioxidant defense and lifespan in mammals. *Proceedings of the National Academy of Sciences*, *98*(23), 12920–12925. https://doi.org/10.1073/pnas.231472998

Moy, S. S., & Nadler, J. (2008). Advances in behavioral genetics: Mouse models of autism. *Molecular Psychiatry*, *13*, 4–26. https://doi.org/10.1038/sj.mp.4002082

Mullis, M. N., Matsui, T., Schell, R., Foree, R., & Ehrenreich, I. M. (2018). The complex underpinnings of genetic background effects. *Nature Communications*, *9*(1), 3548. https://doi.org/10.1038/s41467-018-06023-5

Murthy, M., Garza, D., Scheller, R. H., & Schwarz, T. L. (2003). Mutations in the exocyst component Sec5 disrupt neuronal membrane traffic, but neurotransmitter release persists. *Neuron*, *37*(3), 433–447. https://doi.org/10.1016/S0896-6273(03)00031-X

Nanni, A. V., Martinez, N., Graze, R., Morse, A., Newman, J. R. B., Jain, V., Vlaho, S., Signor, S., Nuzhdin, S. V., Renne, R., & McIntyre, L. M. (2023). Sex-biased expression is associated with chromatin state in *Drosophila melanogaster* and *Drosophila simulans*. *Molecular Biology and Evolution*, *40*(5), msad078. https://doi.org/10.1093/molbev/msad078

Nowicka, M., & Robinson, M. D. (2016). DRIMSeq: A Dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000Research*, *5*, 1356. https://doi.org/10.12688/f1000research.8900.2

Panganiban, G., Irvine, S. M., Lowe, C., Roehl, H., Corley, L. S., Sherbon, B., Grenier, J. K., Fallon, J. F., Kimble, J., Walker, M., Wray, G. A., Swalla, B. J., Martindale, M. Q., & Carroll, S. B. (1997). The origin and evolution of animal appendages. *Proceedings of the National Academy of Sciences*, *94*(10), 5162–5166. https://doi.org/10.1073/pnas.94.10.5162

Parisi, M., Nuttall, R., Edwards, P., Minor, J., Naiman, D., Lü, J., Doctolero, M., Vainer, M., Chan, C., Malley, J., Eastman, S., & Oliver, B. (2004). A survey of ovary-, testis-, and soma-biased gene expression in *Drosophila melanogaster* adults. *Genome Biology*, *5*(6), R40. https://doi.org/10.1186/gb-2004-5-6-r40

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4), 417–419. https://doi.org/10.1038/nmeth.4197

Prokopy, R. J., & Roitberg, B. D. (2001). Joining and avoidance behavior in nonsocial insects. *Annual Review of Entomology*, *46*(1), 631–665. https://doi.org/10.1146/annurev.ento.46.1.631

Quiring, R., Walldorf, U., Kloter, U., & Gehring, W. J. (1994). Homology of the eyeless gene of *Drosophila* to the small eye gene in mice and aniridia in humans. *Science*, *265*(5173), 785–789. https://doi.org/10.1126/science.7914031

R Core Team. (2022). *R: A language and environment for statistical computing. (Version 4.2.0)*. R Foundation for Statistical Computing. https://www.R-project.org/Date accessed June 14, 2022.

R-Core-Team. (2023). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Ramdya, P., Lichocki, P., Cruchet, S., Frisch, L., Tse, W., Floreano, D., & Benton, R. (2015). Mechanosensory interactions drive collective behaviour in *Drosophila*. *Nature*, *519*, 233–236. https://doi.org/10.1038/nature14024

Ranz, J. M., Castillo-Davis, C. I., Meiklejohn, C. D., & Hartl, D. L. (2003). Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science*, *300*(5626), 1742–1745. https://doi.org/10.1126/science.1085881

Reyes, A., Anders, S., & Huber, W. (2013). Inferring differential exon usage in RNA-Seq data with the DEXSeq package. https://bioconductor.org/packages/release/bioc/vignettes/DEXSeq/inst/doc/DEXSeq.html Date accessed May 4, 2024.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47. https://doi.org/10.1093/nar/gkv007

Rittschof, C. C., Bukhari, S. A., Sloofman, L. G., Troy, J. M., Caetano-Anollés, D., Cash-Ahmed, A., Kent, M., Lu, X., Sanogo, Y. O., Weisner, P. A., Zhang, H., Bell, A. M., Ma, J., Sinha, S., Robinson, G. E., & Stubbs, L. (2014). Neuromolecular responses to social challenge: Common mechanisms across mouse, stickleback fish, and honey bee. *Proceedings of the National Academy of Sciences*, *111*(50), 17929–17934. https://doi.org/10.1073/pnas.1420369111

Robinson, G. E., Grozinger, C. M., & Whitfield, C. W. (2005). Sociogenomics: Social life in molecular terms. *Nature Reviews Genetics*, *6*(4), 257–270. https://doi.org/10.1038/nrg1575

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital

gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Roesijadi, G., Rezvankhah, S., Binninger, D. M., & Weissbach, H. (2007). Ecdysone induction of MsrA protects against oxidative stress in *Drosophila*. *Biochemical and Biophysical Research Communications*, *354*(2), 511–516. https://doi.org/10.1016/j.bbrc.2007.01.005

Saltz, J. B. (2011). Natural genetic variation in social environment choice: Context-dependent gene–environment correlation in *Drosophila melanogaster*. *Evolution*, *65*(8), 2325–2334. https://doi.org/10.1111/j.1558-5646.2011.01295.x

Sarin, S., & Dukas, R. (2009). Social learning about egg laying substrates in fruit flies. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 4323–4328. https://doi.org/10.1098/rspb.2009.1294

Schneider, J., Dickinson, M. H., & Levine, J. D. (2012). Social structures depend on innate determinants and chemosensory processing in *Drosophila*. *Proceedings of the National Academy of Sciences*, *109*(Supplement 2), 17174–17179. https://doi.org/10.1073/pnas.1121252109

Scott, A. M., Dworkin, I., & Dukas, R. (2022). Evolution of sociability by artificial selection. *Evolution*, *76*(3), 541–553. https://doi.org/10.1111/evo.14370

Scott, A. M., Dworkin, I., & Dukas, R. (2018). Sociability in fruit flies: Genetic variation, heritability and plasticity. *Behavior Genetics*, *48*, 247–258. https://doi.org/10.1007/s10519-018-9901-7

Scott, A. M., Yan, J. L., Baxter, C. M., Dworkin, I., & Dukas, R. (2022). The genetic basis of variation in sexual aggression: Evolution versus social plasticity. *Molecular Ecology*, *31*(10), 2865–2881. https://doi.org/10.1111/mec.16437

Shpigler, H. Y., Saul, M. C., Corona, F., Block, L., Cash Ahmed, A., Zhao, S. D., & Robinson, G. E. (2017). Deep evolutionary conservation of autism-related genes. *Proceedings of the National Academy of Sciences*, *114*, 9653–9658. https://doi.org/10.1073/pnas.1708127114

Silk, J. B., Alberts, S. C., & Altmann, J. (2003). Social bonds of female baboons enhance infant survival. *Science*, *302*(5648), 1231–1234. https://doi.org/10.1126/science.1088580

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., & Cheney, D. L. (2010). Strong and consistent social bonds enhance the longevity of female baboons. *Current Biology*, *20*(15), 1359–1361.

Silk, J. B., Beehner, J. C., Bergman, T. J., Crockford, C., Engh, A. L., Moscovice, L. R., Wittig, R. M., Seyfarth, R. M., & Cheney, D. L. (2009). The benefits of social capital: Close social bonds among female baboons enhance offspring survival. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1670), 3099. https://doi.org/10.1098/rspb.2009.0681

Silverman, J. L., Yang, M., Lord, C., & Crawley, J. N. (2010). Behavioural phenotyping assays for mouse models of autism. *Nature Reviews Neuroscience*, *11*(7), 490–502. https://www.nature.com/articles/nrn2851

Simon, A. F., Chou, M. T., Salazar, E. D., Nicholson, T., Saini, N., Metchev, S., & Krantz, D. E. (2012). A simple assay to study social behavior in *Drosophila*: Measurement of social space within a group. *Genes, Brain and Behavior*, *11*(2), 243–252. https://doi.org/10.1111/j.1601-183X.2011.00740.x

Singh, P., & Ahi, E. P. (2022). The importance of alternative splicing in adaptive evolution. *Molecular Ecology*, *31*(7), 1928–1938. https://doi.org/10.1111/mec.16377

Smith, C. R., Toth, A. L., Suarez, A. V., & Robinson, G. E. (2008). Genetic and genomic analyses of the division of labour in insect societies. *Nature Reviews Genetics*, *9*(10), 735–748. https://doi.org/10.1038/nrg2429

Snyder-Mackler, N., Burger, J. R., Gaydosh, L., Belsky, D. W., Noppert, G. A., Campos, F. A., Bartolomucci, A., Yang, Y. C., Aiello, A. E., O'Rand, A., Harris, K. M., Shively, C. A., Alberts, S. C., & Tung, J. (2020). Social determinants of health and survival in humans and other animals. *Science*, *368*(6493), eaax9553. https://doi.org/10.1126/science.aax9553

Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences. *F1000Research*, *4*, 1521. https://doi.org/10.12688/f1000research.7563.2

Steptoe, A., Shankar, A., Demakakos, P., & Wardle, J. (2013). Social isolation, loneliness, and all-cause mortality in older men and women. *Proceedings of the National Academy of Sciences*, *110*(15), 5797–5801. https://doi.org/10.1073/pnas.1219686110

Swope, R. D., Hertzler, J. I., Stone, M. C., Kothe, G. O., & Rolls, M. M. (2022). The exocyst complex is required for developmental and regenerative neurite growth in vivo. *Developmental Biology*, *492*, 1–13. https://doi.org/10.1016/j.ydbio.2022.09.005

Taylor, M. B., & Ehrenreich, I. M. (2015). Higher-order genetic interactions and their contribution to complex traits. *Trends in Genetics*, *31*(1), 34–40. https://doi.org/10.1016/j.tig.2014.09.001

Tinbergen, N. (1953). *Social behaviour in animals: With special reference to vertebrates London*. Methuen & Co.

Toth, A. L., & Rehan, S. M. (2017). Molecular evolution of insect sociality: An eco-evo-devo perspective. *Annual Review of Entomology*, *62*(1), 419–442. https://doi.org/10.1146/annurev-ento-031616-035601

Van Bergen, N. J., Ahmed, S. M., Collins, F., Cowley, M., Vetro, A., Dale, R. C., Hock, D. H., de Caestecker, C., Menezes, M., Massey, S., Ho, G., Pisano, T., Glover, S., Gusman, J., Stroud, D. A., Dinger, M., Guerrini, R., Macara, I. G., & Christodoulou, J. (2020). Mutations in the exocyst component EXOC2 cause severe defects in human brain development. *Journal of Experimental Medicine*, *217*(10), e20192040. https://doi.org/10.1084/jem.20192040

Verta, J. —P., & Jacobs, A. (2022). The role of alternative splicing in adaptation and evolution. *Trends in Ecology & Evolution*, *37*(4), 299–308. https://doi.org/10.1016/j.tree.2021.11.010

Walker, W. H., Walton, J. C., & Nelson, R. J. (2021). Disrupted circadian rhythms and mental health. In D. F. Swaab, F. Kreier, P. J. Lucassen, A. Salehi, & R. M. Buijs (Eds.), *Handbook of clinical neurology*. (Vol. *179*, pp. 259–270): Elsevier.

Wang, L., Wang, S., & Li, W. (2012). RSeQC: Quality control of RNA-seq experiments. *Bioinformatics*, *28*(16), 2184–2185. https://doi.org/10.1093/bioinformatics/bts356

Wang, Z. Y., McKenzie-Smith, G. C., Liu, W., Cho, H. J., Pereira, T., Dhanerawala, Z., Shaevitz, J. W., & Kocher, S. D. (2022). Isolation disrupts social interactions and destabilizes brain development in bumblebees. *Current Biology*, *32*(12), 2754–2764.e5. https://doi.org/10.1016/j.cub.2022.04.066

Ward, A., & Webster, M. (2016). *Sociality: The behaviour of group living animals*. Springer. https://doi.org/10.1007/978-3-319-28585-6

Wice, E. W., & Saltz, J. B. (2021). Selection on heritable social network positions is context-dependent in *Drosophila melanogaster*. *Nature Communications*, *12*(1), 3357. https://doi.org/10.1038/s41467-021-23672-1

Wilson, E. O. (1975). *Sociobiology: The new synthesis*. Harvard University Press.

Woodard, S. H., Fischman, B. J., Venkat, A., Hudson, M. E., Varala, K., Cameron, S. A., Clark, A. G., & Robinson, G. E. (2011). Genes involved in convergent evolution of eusociality in bees. *Proceedings of the National Academy of Sciences*, *108*(18), 7472–7477. https://doi.org/10.1073/pnas.1103457108

Wright, C. J., Smith, C. W. J., & Jiggins, C. D. (2022). Alternative splicing as a source of phenotypic diversity. *Nature Reviews Genetics*, *23*(11), 697–710. https://doi.org/10.1038/s41576-022-00514-4

Yang, Y. C., Boen, C., Gerken, K., Li, T., Schorpp, K., & Harris, K. M. (2016). Social relationships and physiological determinants of longevity across the human life span. *Proceedings of the National Academy of Sciences*, *113*(3), 578–583. https://doi.org/10.1073/pnas.1511085112

Yeh, S. —D., Do, T., Chan, C., Cordova, A., Carranza, F., Yamamoto, E. A., Abbassi, M., Gandasetiawan, K. A., Librado, P., Damia, E., Dimitri, P., Rozas, J., Hartl, D. L., Roote, J., & Ranz, J. M. (2012). Functional evidence that a recently evolved *Drosophila* sperm-specific gene boosts sperm competition. *Proceedings of the National Academy of Sciences*, *109*(6), 2043–2048. https://doi.org/10.1073/pnas.1121327109

Zirin, J., Hu, Y., Liu, L., Yang-Zhou, D., Colbeth, R., Yan, D., Ewen-Campen, B., Tao, R., Vogt, E., VanNest, S., Cavers, C., Villalta, C., Comjean, A., Sun, J., Wang, X., Jia, Y. u, Zhu, R., Peng, P., Yu, J., … Perrimon, N. (2020). Large-scale transgenic *Drosophila* resource collections for loss- and gain-of-function studies. *Genetics*, *214*(4), 755–767. https://doi.org/10.1534/genetics.119.302964

Received August 8, 2024; revisions received January 22, 2025; accepted July 29, 2025
Associate Editor: Amanda J. Moehring; Handling Editor: Jason Wolf