

# The regulatory architecture of gene expression variation in *Caenorhabditis elegans* revealed by multi-strain allele-specific analysis

Avery Davis Bell <sup>1,\*</sup> Francisco Valencia,<sup>1</sup> Annalise B. Paaby <sup>1,\*</sup>

<sup>1</sup>School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive NW, Atlanta, GA 30332-2000, USA

\*Corresponding author: School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive NW, Atlanta, GA 30332-2000, USA. Email: [averydavisbell@gmail.com](mailto:averydavisbell@gmail.com); \*Corresponding author: School of Biological Sciences, Georgia Institute of Technology, 950 Atlantic Drive NW, Atlanta, GA 30332-2000, USA. Email: [paaby@gatech.edu](mailto:paaby@gatech.edu)

An outstanding question in the evolution of gene expression is the composition of the underlying regulatory architecture and the processes that shape it. Mutations affecting a gene's expression may reside locally in cis or distally in trans; the accumulation of these changes, their interactions, and their modes of inheritance influence how traits are expressed and how they evolve. Here, we interrogated gene expression variation in *Caenorhabditis elegans*, including the first allele-specific expression analysis in this system, capturing effects in cis and in trans that govern gene expression differences between the reference strain N2 and 7 wild strains. We observed extensive compensatory regulation, in which opposite effects in cis and trans at individual genes mitigate expression differences among strains, and that genes with expression differences exhibit strain specificity. As the genomic distance increased between N2 and each wild strain, the number of genes with expression differences also increased. We also report for the first time that expression-variable genes are lower expressed on average than genes without expression differences, a trend that may extend to humans and *Drosophila melanogaster* and may reflect the selection constraints that govern the universal anticorrelation between gene expression and rate of protein evolution. Together, these and other observed trends support the conclusion that many *C. elegans* genes are under stabilizing selection for expression level, but we also highlight outliers that may be biologically significant. To provide community access to our data, we introduce an easily accessible, interactive web application for gene-based queries: <https://wildworm.biosci.gatech.edu/ase/>.

**Keywords:** *C. elegans*; gene expression; allele-specific expression; regulatory architecture; natural genetic variation; regulatory evolution; RNA sequencing

## Introduction

Gene expression is an essential step in the translation of genotype to phenotype. Thus, to understand the evolution of traits and genomes, we need to elucidate the regulatory mechanisms, modes of inheritance, and biological processes shaping gene expression variation within a system. Mutational variants that affect gene expression may act in cis, locally within the focal gene haplotype, such as promoter variants, and/or in trans, on a separate molecule and potentially affecting all targets, such as mutations in transcription factors (Signor and Nuzhdin 2018). Regulatory variants that mediate gene expression may represent adaptive change, neutral differences, or relaxed selection (reviewed in, e.g. Landry, Lemos, et al. 2007; Fay and Wittkopp 2008; Romero et al. 2012; Signor and Nuzhdin 2018, 2019; Hill et al. 2020; Price et al. 2022a). They may also act to stabilize expression by dampening changes to expression induced by other variants.

An incisive way to study gene expression regulation is to examine variation by simultaneously capturing expression among wild strains and their F1 hybrid offspring (Wittkopp et al. 2004; Landry, Hartl, et al. 2007). Within the F1, expression differences observed between the parental alleles (allele-specific expression [ASE])

may be assigned to mutations in cis, on the same molecule, because the diffusible trans environment is shared within cells (Cowles et al. 2002; Yan et al. 2002). Thus, comparisons of expression between alleles, between parents, and between F1s and parents enable inference of the regulatory architecture and inheritance mode of gene expression (Wittkopp et al. 2004; McManus et al. 2010). This approach has been employed in a number of systems to interrogate various phenomena, including domestication, adaptation, and speciation in wild and crop plants (Zhang and Borevitz 2009; He et al. 2012; Lemmon et al. 2014; Steige et al. 2015; He et al. 2016; Verta et al. 2016; Rhone et al. 2017; Steige et al. 2017; Bao et al. 2019); adaptation and the evolution of embryogenesis in *Drosophila* (McManus et al. 2010; Coolon et al. 2014; Juneja et al. 2016; Cartwright and Lott 2020); speciation and cis regulatory variation in mice (Crowley et al. 2015; Mack et al. 2016); human-specific regulatory evolution in chimpanzee-human hybrid cell lines (Gokhman et al. 2021; Starr et al. 2023; Wang et al. 2024); RNA and protein regulation in yeast (Artieri and Fraser 2014; Muzzey et al. 2014; Wang et al. 2015); and speciation and evolution of reproductive mode in nematodes (Sanchez-Ramirez et al. 2021; Xie et al. 2022; Viswanath and Cutter 2023).

*Caenorhabditis elegans* has long been a leading developmental and genetic model organism (Sternberg et al. 2024), and the recent establishment of a global collection of wild strains has pushed *C. elegans* to the forefront of quantitative genetics and evolutionary genomics research (Frézal and Félix 2015; Cook et al. 2017; Crombie et al. 2019; Andersen and Rockman 2022; Crombie et al. 2024). Yet, while the genetic basis of expression variation has been interrogated via well-powered eQTL studies (Rockman et al. 2010; Vinuela et al. 2010; Francesconi and Lehner 2014; Kamkina et al. 2016; Evans and Andersen 2020; Zhang et al. 2022), the regulatory architecture and inheritance mode of gene expression variation in *C. elegans* has not been assessed by allele-specific analyses. However, the biology of *C. elegans* offers rich opportunity for investigating gene expression variation and its evolution, beyond its well-established resources. *C. elegans* strains persist as predominantly selfing lineages in diverse ecological habitats across the globe, with the population structure exhibiting a spectrum of genetic divergence between wild strains, from closely related to highly diverged (Barriere and Felix 2005a; Barriere and Felix 2005b; Crombie et al. 2019; Lee et al. 2021; Crombie et al. 2024). The genomes harbor extensive linkage disequilibrium, including long haplotypes arising from historical adaptive sweeps, and interstrain crosses often exhibit fitness deficits, suggesting disruption of the selfed, coadapted genotype combinations (Barriere and Felix 2005a; Dolgin et al. 2007; Rockman and Kruglyak 2009; Andersen et al. 2012). Thus, *C. elegans* is optimally suited to facilitate investigations into whether and how genetic divergence translates to differences in expression, into the scope and correlates of compensatory interactions in the evolution of gene expression regulation, and into the broader evolutionary pressures shaping these trends.

The role of compensatory interactions in the evolution of gene expression is incompletely understood, but a growing body of literature suggests that such dynamics are influential and pervasive. Gene expression changes often fail to result in protein-level changes (Schrimpf et al. 2009; Khan et al. 2013; Brion et al. 2020; Buccitelli and Selbach 2020) and regulatory changes to expression arising in cis often fail to produce overall differences in gene expression, implying that they are compensated by regulation in trans (Landry et al. 2005; Signor and Nuzhdin 2018, 2019). Studies have reported such compensation of cis-regulated differences in hybrids of different species, subspecies, and occasionally strains of fruit flies, sticklebacks, cotton, mice, yeast, spruce, and more (Landry et al. 2005; McManus et al. 2010; Goncalves et al. 2012; Coolon et al. 2014; Verta et al. 2016; Metzger et al. 2017; Signor and Nuzhdin 2018, 2019; Bao et al. 2019; Verta and Jones 2019). In yeast, dissection of the trans regulatory architecture of a single gene estimated hundreds of variants mediating its expression, with extensive compensatory interactions and evidence of stabilizing selection on overall expression level (Metzger and Wittkopp 2019). However, methodological constraints and analytical artifacts can limit confidence at both protein and RNA levels, hindering widespread inferences (Fraser 2019; Buccitelli and Selbach 2020; Fraser 2022; Price et al. 2022a; Price et al. 2022b). In *C. elegans*, fitness-related traits exhibit compensatory-like architecture, with epistasis and tightly linked opposite-direction effects shaping fertility and fecundity (Noble et al. 2017; Bernstein et al. 2019); these traits may be governed by similar dynamics at the level of gene expression. Overall, the extent to which *C. elegans* gene expression has evolved compensatory interactions remains an open question.

Here, to elucidate the regulatory architecture and evolutionary dynamics shaping gene expression in *C. elegans*, we examine

expression variation in 7 intraspecific crosses. In each, the reference strain N2 was crossed to 1 of 7 wild strains representing a spectrum of genomic differentiation from the reference. We define the regulatory patterns and inheritance modes of expression variation in this system and then assess how regulatory effects are influenced by factors such as nucleotide diversity, genome evolutionary history, gene essentiality and biological role, and expression level.

## Materials and methods

### Experimental methods

A detailed protocol describing the experimental methods is available at protocols.io ([dx.doi.org/10.17504/protocols.io.5jyl8p15rg2w/v1](https://doi.org/10.17504/protocols.io.5jyl8p15rg2w/v1), Bell et al. 2024).

### Worm strains

Supplementary Table 1 provides the complete list of strains used in this study. In selecting parental strains to cross with the N2 laboratory reference strain to generate F1s in which to investigate ASE, we aimed to represent the range of nucleotide diversity present in the species as well as capture outlier strains. All chosen strains differed at more than 127,000 nucleotides from N2 (>1.27 variants per kilobase average) (per CaenDR, Crombie et al. 2024) to ensure that the F1s harbored many genes with differences from the reference in coding regions. To ensure that we generated F1s with 1 copy of the genome from each parent, rather than N2 self-progeny, we used the N2 strain feminized via a deletion of *fog-2* as the N2 “female” parent (referred to in the text as N2<sup>fog-2</sup>, strain CB4108): *fog-2*-deficient hermaphrodites are incapable of producing sperm and therefore function as females (Schedl and Kimble 1988; Hodgkin 2002).

### Worm husbandry

We thawed fresh aliquots of each wild strain and grew them without starving for at least 3 generations, but for no more than 1 month, prior to starting the experiment. We followed standard protocol (Stiernagle 2006) for worm culture, using 1.25% agarose plates to prevent wild strains’ burrowing. Prior to the start of the experiment, all strains were maintained at 18°C to allow slower growth of large quantities of worms and to avoid inducing QX1211s mortal germline phenotype, which is more penetrant at higher temperatures (Frézal et al. 2018).

### Generating parallel F1 crosses and self-progeny

As detailed in our protocol (Bell et al. 2024), we first bleach synchronized all parental strains to ensure that the parents that would be mated were of similar developmental stage, as parental age can impact offspring development and transcriptional program (Perez et al. 2017; Webster et al. 2025). To ensure that we would have many L4 parent worms to move to mating plates, we grew several plates of all bleached strains at 18, 19, and 20°C and additionally grew the N2<sup>fog-2</sup> parent (from whom we needed the highest number of worms) at room temperature.

After allowing these worms to grow for 2 days, we generated mating plates by placing 60–80 N2<sup>fog-2</sup> L4 pseudohermaphrodites onto each of 5 6-cm plates with small bacteria spots and added 40 L4 males of the appropriate strain to each plate. We concurrently moved 80 individual L4 hermaphrodites to each of 3 6-cm plates for each parental strain (N2 and 7 wild strains) to simultaneously generate the parental strains used for sequencing from self-matings, while the F1 crosses were generated from cross-matings.

After allowing mating for 48 h, we collected and synchronized the offspring for the crosses and self-matings by collecting all parental worms and embryos from the bacterial lawn, treating with bleach, and allowing embryos to develop into L1 larvae and arrest over 30 h in liquid buffer. After 30 h, L1s were transferred directly to the bacterial lawn of 6-cm plates at a density of ~400 L1s per plate.

After allowing the worms to develop for ~36 h, we removed males from the F1 plates as soon as they were detectable and screened the parental plates for any spontaneously generated males, which were also removed. Plates used for RNA sequencing (at least 3 per strain) had all males removed as L4s or young adults.

### Worm harvesting

Worms were harvested as day 1 reproductively mature young adults: when most worms were gravid with embryos and laid embryos were visible on the plates. Because developmental timing differs across wild strains (Hodgkin and Doniach 1997; Gems and Riddle 2000; Harvey and Viney 2007; Pouillet et al. 2015; Stastna et al. 2015; Zhang et al. 2021), we chose to match developmental stage rather than hours of development; even so, all worms reached reproductive maturity and were harvested within 3 h of each other. Worms were rinsed off plates, washed with M9 buffer, and resuspended in TRIzol (Invitrogen #15596026) in 3 tubes (replicates) per strain before immediate flash freezing in liquid nitrogen and storage at  $-80^{\circ}\text{C}$  until RNA extraction.

### RNA library preparation and sequencing

RNA was extracted from worms stored in TRIzol (Invitrogen #15596026) following standard procedure (following He 2011, see Bell et al. 2024) using a TRIzol (Invitrogen #15596026) chloroform (Fisher #C298-500) extraction and RNeasy columns (Qiagen #74104). This extraction was performed in 3 batches of 15 over 2 consecutive days, with 1 replicate from each strain included in each batch. RNA was stored at  $-80^{\circ}\text{C}$  for ~1 wk prior to library generation. Library preparation and sequencing for all samples were performed by the Molecular Evolution Core Laboratory at the Georgia Institute of Technology. Following RNA quality checks (all RINs 9.8 or greater), mRNA was enriched from 1  $\mu\text{g}$  RNA with the NEBNext Poly(A) mRNA magnetic isolation module (NEB #E7490) and sequencing libraries generated using the NEBNext Ultra II directional RNA library preparation kit (NEB #E7760) with 8 cycles of PCR. Libraries were quality checked and fluorometrically quantified prior to pooling and sequencing. Libraries were sequenced on an Illumina NovaSeq X using a 300 cycle 10B flowcell. A median of 65 million  $150 \times 150$  bp sequencing read pairs were generated per library (range 25–93 million, Supplementary Table 1).

### Analytical methods

The code written for this study is available at <https://github.com/paabylab/wormase>.

### Expression quantification

Before expression quantification, we generated strain-specific transcriptomes as described previously (Bell et al. 2023) by inserting known SNV and INDEL polymorphisms (from the CeNDR (Cook et al. 2017; Crombie et al. 2024) 2021021 release hard-filter VCF) into the *C. elegans* reference genome (ws276 from WormBase, Sternberg et al. 2024) and extracting transcripts. We created pseudodiploid strain transcriptomes by combining these strain-specific transcriptomes for the 2 parent strains. Tools

used in generating these transcriptomes included g2gtools (v0.1.31) (<https://github.com/churchill-lab/g2gtools>), gffread (v0.12.7) (Pertea and Pertea 2020), seqkit (v0.16.1) (Shen et al. 2016), and bioawk (v1.0) (<https://github.com/lh3/bioawk>). For comparison purposes, we also created pseudodiploid and strain-specific transcriptomes using script *create\_personalized\_transcriptome.py* from the Ornaments code suite (initial version) (Adduri and Kim 2024) tool, with the ws286 genome build and 20220216 CeNDR VCF.

For quantification used in ASE and differential expression (DE) analyses, we estimated allele-specific and total RNA counts using EMASE (emase-zero v0.3.1) (Raghupathy et al. 2018) with input quantifications generated by running Salmon (v1.4) (Patro et al. 2017) against the pseudodiploid transcriptomes. Specifically, we generated a salmon index for the diploid transcriptome using *salmon index* with options *-k 31 --keepDuplicates* (no decoy, all other parameters default). To prepare RNA-seq data for quantification, we trimmed Illumina adapters using trimmomatic (v0.39) (Bolger et al. 2014) with parameters *ILLUMINACLIP: TruSeq3-PE-2.fa:1:30:12:2:True*. Salmon quantification with equivalence class outputs saved was performed against the pseudodiploid transcript's index with *salmon quant -l ISR -dumpeq -fldMean < sample-specific mean > -fldSD < sample-specific SD > -rangeFactorizationBins 4 -seqBias -gcBias*. Salmon outputs were converted to .bin inputs for emase-zero using *alntools salmon2ec* (v0.1.1) (<https://churchill-lab.github.io/alntools/>). Finally, emase-zero was run on this input using parameters *-model 4 -t 0.0001 -i 999*. For comparison, we separately generated quantification estimates using kallisto (v0.50.1) (Bray et al. 2016) against strain-specific transcriptomes generated by Ornaments and estimated allele-specific RNA counts using *ornaments quant* (initial version), which implements WASP (van de Geijn et al. 2015)-style allele-specific quantification on top of kallisto quantification and includes INDELs in its analysis. Workflows to perform these steps are in our code repository in internal directories: *data\_generation\_scripts/getdiploidtranscriptomes*; *data\_generation\_scripts/emase*; *data\_generation\_scripts/ornaments*.

We used DESeq2 (v1.42.0) (Love et al. 2014) to obtain final RNA quantifications for downstream modeling. For DE analyses, we used the “total” column of the “gene.counts” output from emase-zero. For allele-specific analyses, we used the allelic counts columns of the “gene.counts” output from emase-zero. For kallisto quantifications, transcript TPMs were combined to gene-level, normalized quantifications for DESeq2 using *tximport* (v1.30.0) (Soneson et al. 2015). In all cases, genes with at least 10 total reads summed across samples were retained for downstream analysis. For obtaining general best expression quantification estimates (rather than for DE modeling), we used DESeq2's variance-stabilizing transformation (*vst* function) to get log-scale, variance-normalized, and length- and library size-normalized gene expression estimates.

### Age estimation

We estimated each sample's age in hours against a developmental timing “ruler” from the N2 strain via RAPToR (v1.2.0) (Bulteau and Francesconi 2022) using DESeq2's *vst*-corrected gene counts from total emase-zero outputs. The age reference used (provided with RAPToR) was *Cel\_YA\_2*. The script used to perform this analysis is in our code repository: *data\_classification\_scripts/RAPToR.R*.

### DE and ASE calling

Each sample was assigned to its generation-strain group (e.g. CB4856 F1). Total gene counts from emase-zero “total”

gene.counts output were binomially negatively modeled by DESeq2 as

$$\log_2(q_{ij}) = \beta_{1i}x_j + \beta_{2i}y_j$$

where, for gene  $i$ , sample  $j$ ,  $q$  is proportional to RNA concentration/counts (Love et al. 2014) and  $\beta$ s give the effects for gene  $i$  for RNA extraction replicate ( $x$ ) and each generation-strain pair ( $y$ ). The Wald test was used for significance testing. The results were pulled out for each pairwise comparison of interest using DESeq2's contrasts. All  $\log_2$  fold changes (LFCs) were adjusted using *ashr* (v2.2-63) (Stephens 2016). For DE to be called, both a fold change (FC) of  $>1.5$  after *ashr* adjustment (for significance testing and calling) and a genome-wide adjusted  $P$ -value of  $<0.05$  were required.

For genes to be included in ASE analyses, we required them to have 5 gene and allele-specific alignments. The total counts of alignments per gene and those that were gene and allele-specific were derived by analyzing of salmon's equivalence class output file, which assigns equivalence classes of kmers to transcripts from which they derive and gives the counts of reads aligning to each equivalence class. We investigated several thresholds of gene- and allele-specific alignments for considering a gene ASE-informative; we found that our RNA sequencing was deep enough that once genes in each F1 genotype had more than 3 allele- and gene-specific alignments in each sample from that genotype, they usually had many allele- and gene-specific alignments. Therefore, we required genes to have a slightly conservative 5 allele- and gene-specific alignments to be considered ASE-informative.

To model ASE in the F1s, each allele's count was represented in its own column in the model matrix. Within each strain, each sample was assigned its sample blocking factor, controlling for sample during modeling. We used DESeq2's negative binomial modeling to model allele counts:

$$\log_2(q_{ij}) = \beta_{1i}x_j + \beta_{2i}x_jy_j + \beta_{3i}z_j$$

where, for gene  $i$ , allele (rather than sample)  $j$ ,  $q$  is proportional to allelic RNA concentration/counts (Love et al. 2014),  $\beta_1$  gives the effect of RNA extraction replicate ( $x$ ),  $\beta_2$  gives the effect of the interaction between RNA extraction replicate and specific sample ( $xy$ ), and  $\beta_3$  gives the effect of the allele/genotype ( $z$ ). Library size correction was not used in this modeling because all comparisons were being made within sample, where library size was identical, and counts were of alleles rather than total. Library size correction was excluded by setting all DESeq2 size factors to 1 prior to DE testing. The results were extracted for each allelic pairwise comparison of interest and were used in downstream analysis for ASE-informative genes. ASE-informative genes were considered to have ASE if their *ashr*-adjusted FC was greater in magnitude than 1.5 (equivalent to having 60% of alleles come from 1 haplotype) and their genome-wide-adjusted  $P$ -value was  $<0.05$  (the same thresholds required for DE calls; FC threshold used in both significance testing and calling). Both LFCs and the proportion of alleles deriving from the reference and alternate genomes were used for downstream analytical interpretation; alternate allele proportion was calculated from the *ashr*-adjusted LFC as

$$\frac{2^{\text{LFC}}}{(1 + 2^{\text{LFC}})}$$

The scripts used for these analyses are in our code repository: equivalence class processing for ASE-informative decisions in

`data_generation_scripts/salmonalleleeqclasses.py` and ASE and DE modeling in `data_classification_scripts/ase_de_annotategenes_deseq2_frommemaseout.R`.

### Inheritance mode classifications

Inheritance mode categories were called from DE testing results (from global RNA counts); categories and definitions followed McManus et al. (2010) and others, with thresholds tuned for our specific statistical testing framework. All  $P$ -values were genome-wide adjusted and FCs/LFCs were *ashr* adjusted. Genes were called "no\_change" if there was no DE between the parents, between the F1 and the N2 parent, or between the F1 and the other parent (all  $P > 0.05$  or  $|FC| < 1.5$ ). Genes were called "overdominant" if the F1 had higher expression than both parents ( $FC > 1.5$  and  $P < 0.05$ ). Genes were called "underdominant" if the F1 had lower expression than both parents ( $FC < -1.5$  and  $P < 0.05$ ). Genes were called "N2\_dominant" if the parents were differentially expressed and the F1 was potentially differentially expressed from the wild parent in the same direction as N2 was (N2 vs wild strain  $|FC| > 1.5$  and  $P < 0.05$ , F1 vs wild strain  $P < 0.05$  and FC in the same direction as N2s), or if the parents were potentially differentially expressed and the F1 was differentially expressed in the same direction from the wild parent as N2 was (N2 vs wild strain  $P < 0.05$  and FC in the same direction as F1s; F1 vs wild strain  $|FC| > 1.5$  and  $P < 0.05$ ). Genes were called "alt\_dominant" the same way as N2\_dominant but requiring the F1 to be differentially expressed from the N2 parent in the same way as its wild parent. Genes were called "additive" if the parent strains were differentially expressed ( $P < 0.05$  and  $|FC| > 1.5$ ) and the F1 had nominally called DE with expression amount falling between the 2 parents ( $P < 0.05$ ,  $FC > 0$  if parental  $FC > 0$  and  $FC < 0$  if parental  $FC < 0$ ). Therefore, genes with midparental gene expression closer to the expression level of 1 parent or another, which might more formally be termed "nearly additive," are included in the "additive" category. Genes whose DE results did not meet any of the above requirements were called "ambiguous," for example, when parental DE was not called but the F1 had DE called from 1 parent (these genes might be either additively inherited or dominantly inherited, but the statistical evidence was not strong enough to discern which). The inheritance mode classification script is available in our code repository: `data_classification_scripts/f1_parental_inhmode_withinstrain.R`.

### Regulatory pattern and related classifications

Regulatory pattern categories were called from comparisons of ASE (N2 vs wild strain allele) calls and DE (N2 vs wild strain total RNA counts) calls; categories and definitions followed McManus et al. (2010) and others, with the specific thresholds tuned for our specific statistical testing framework. All  $P$ -values were genome-wide adjusted and FCs/LFCs were *ashr* adjusted, and categorizations were only considered if genes were ASE-informative. Genes were called "conserved" if they had neither ASE nor DE (both allelic and strain-wise  $P > 0.05$  and  $|FC| < 1.5$ ). Genes were called "cis" (i.e. cis-only or cis-dominant regulatory divergence) if ASE and DE were both present and in the same direction and if their 99.9% C.I. on effect size overlapped (allelic  $P < 0.05$  and  $|FC| > 1.5$ , strain-wise  $P < 0.05$  without FC threshold,  $\log_2FC(ASE)/\log_2FC(ASE) > 0$ ). Genes were called "trans" (i.e. trans-only or trans-dominant regulatory divergence) if they did not have ASE but did have DE (allelic  $P > 0.05$ , strain-wise  $P < 0.05$  and  $|FC| > 1.5$ ). Genes were called "enhancing" (i.e. cis-trans enhancing or cis + trans) if they had both ASE and DE in the same direction and DE was of greater magnitude than ASE with nonoverlapping



99.9% C.I. of the ASE and DE estimates ( $ASE\ P < 0.05$  and  $|FC| > 1.5$  and  $DE\ P < 1$ , or  $ASE\ P < 0.05$  and  $DE\ P < 0.05$  and  $|FC| > 1.5$ ; and  $\log_2FC(DE)/\log_2FC(ASE) > 1$ ). Genes were called “compensating” (i.e. cis and trans regulatory changes in opposite directions, with the cis effect larger than the trans effect) if they had ASE and DE in the same direction with larger ASE than DE and nonoverlapping 99.9% C.I. on the ASE and DE estimates ( $0 > \log_2FC(DE)/\log_2FC(ASE) > 1$ , allelic  $P < 0.05$  and  $|FC| > 1.5$  and strain-wise  $P < 0.05$  or allelic  $P < 0.05$  and strain-wise  $P < 0.05$  and  $|FC| > 1.5$ ). Genes were called “compensatory” (i.e. cis and trans regulatory changes in opposite directions, with trans changes fully offsetting the cis changes) if there was ASE but not DE (allelic  $P < 0.05$  and  $|FC| > 1.5$ , strain-wise  $P > 0.05$ ). Genes were called “overcompensating” (i.e. cis and trans regulatory changes in opposite directions, with the trans change more than offsetting the cis effect) if they had ASE and DE in different directions with nonoverlapping 99.9% C.I. on the ASE and DE estimates ( $\log_2FC(DE)/\log_2FC(ASE) < 0$ ; allelic  $P < 0.05$  and  $|FC| > 1.5$  and strain-wise  $P < 0.05$  or allelic  $P < 0.05$  and strain-wise  $P < 0.05$  and  $|FC| > 1.5$ ). Genes were called “ambiguous” if they did not meet the above criteria, e.g. when ASE and DE were called with overlapping widened C.I. on opposite-direction effects. The regulatory pattern classification script is in our code repository: `data_analysis_scripts/ase_de_cistransclassifications.R`.

We simplified these regulatory patterns for ease of understanding and visualization in a couple of ways. First, genes were classified as “cis–trans opposing” anytime they had opposite-direction cis and trans effects, i.e. when their regulatory pattern was “compensating,” “compensatory,” or “overcompensating.” Second, we used the regulatory patterns to investigate compensation in a more targeted way, classifying genes as compensated if their simplified regulatory pattern was “cis–trans opposing” and as not compensated if their regulatory pattern was “cis” or “enhancing.” Genes without cis regulatory changes therefore are neither compensated or not compensated and were not included in compensation-specific analyses.

To evaluate the extent of compensatory regulatory effects, i.e. to assess whether individual genes showed an enrichment for effects in both cis and trans, and specifically for opposing effects in cis vs trans, we first considered the numbers of genes with only effects in cis (“cis” class), with only effects in trans (“trans”), with both cis and trans (“compensating,” “compensatory,” “overcompensating,” and “enhancing”), and with no change (“conserved”). The distribution of these numbers across the  $2 \times 2$  table was assessed for significance using a Fisher’s exact test (FET)/hypergeometric test. We also repeated this test excluding enhancing genes from the category of genes with both cis and trans effects (both were highly significant, all  $P < 2 \times 10^{-112}$  and all  $P < 4 \times 10^{-102}$ , respectively). To assess the enrichment of opposing effects within the set of genes that showed both cis and trans regulation, we used a binomial test (always 87–92% of genes with cis and trans effects had those effects oppose each other across strains). To assess what proportion of genes called cis–trans opposing would have to be wrongly classified to result in the FET being not significant (after Bonferroni-correction), we sequentially removed 1–100% of these genes from this category in 2 ways and then recomputed the FET. First, to simulate DE being missed at a gene, we took genes away from the cis–trans opposing group and added them to the cis-only group. Second, to simulate ASE being spuriously called at a gene, we took genes away from the cis–trans opposing group and added them to the conserved group. In each such simulation, we computed the proportion of all called DE genes simulated as being missed and all called ASE genes

simulated as being spurious and report these numbers in the main text where the FET was no longer significant. This analysis script is in our code repository: `data_analysis_scripts/compensationamounttesting.R`.

### Gene filtering

We performed all analyses including all nominally expressed genes, excluding genes overlapping hyperdivergent haplotypes or with aberrantly low or high DNA sequence coverage in the focal strain, and excluding all genes called hyperdivergent in any of 328 strains analyzed by CeNDR (Lee et al. 2021). Focal strain gene haplotype hyperdivergence was inferred when the gene region overlapped any hyperdivergent haplotype in this strain in the hyperdivergent haplotype BED file from the CeNDR 20210121 release (Lee et al. 2021). Genes were flagged as having aberrantly low or high DNA sequence coverage if they had  $<0.3$ - or  $>2.5$ -fold the median gene’s coverage in that strain, with coverage calculated across all exonic bases from CeNDR DNA sequence BAMs (20210121 release), as described previously (Bell et al. 2023). The list of genes hyperdivergent in any strain population wide was from Lee et al. (2021).

### Gene set enrichment analyses

We used WormCat (Holdorf et al. 2020) to perform gene set enrichment analyses by writing a script extension to the WormCat R package (v2.0.1) that allowed us to provide a custom background gene set for enrichment tests. We performed the following tests with genes from each strain separately (formatted here as test gene set vs background gene set): DE genes vs all analyzed genes, ASE genes vs ASE-informative genes, compensatory genes vs ASE-informative genes, compensatory genes vs ASE genes, transgressive (overdominant + underdominant) genes vs all analyzed genes, overdominant genes vs all analyzed genes, underdominant genes vs all analyzed genes, DE genes that are ASE-informative vs ASE-informative genes, ASE-informative genes vs all analyzed genes, N2-dominant genes vs all analyzed genes, wild dominant genes vs all analyzed genes, cis genes that were not called additive inheritance mode vs ASE-informative genes, and cis genes that were not called additive inheritance mode vs ASE genes. The WormCat extension and analysis scripts are in our code repository: `data_analysis_scripts/wormcat_givebackgroundset.R` and `data_analysis_scripts/combinewormcatout_aseetc.R`.

### Meta-strain results: combined comparisons across strains

We performed all analyses within each strain/strain pair, and we also combined strains’ results into 1 “meta-strain” to be able to display and report 1 set of results when results across strains were largely consistent. For this meta-strain, genes were considered ASE-informative if they were ASE-informative in all 7 strains and not ASE-informative if they were not informative for ASE in any strain; genes had to be informative in all strains or not informative in any strain to be compared in informative-vs-not analyses. To compare ASE vs not, DE vs not, and regulatory pattern, genes informative in all strains were included for each strain: each gene is present on each plot 7 times, in the category of its classification for each strain. For example, 1 gene might be called ASE in 3 strains and not ASE in 4 strains and would be represented by 3 points in the ASE group and 4 points in the non-ASE group. In some cases, other characteristics of the gene (such as essentiality, see below) were the same across strains and therefore represented identically 7 times, while in others (such as expression level, see below), both the ASE characterization and the other characteristic are different in each strain.

## Comparing to regulatory divergence between *Caenorhabditis nigoni* and *Caenorhabditis briggsae*

To determine how our estimates of regulatory divergence within *C. elegans* compare to the interspecies regulatory differences between nematode species *C. nigoni* and *C. briggsae*, we obtained the number of regulatory-diverged genes from the ASE analyses in Sanchez-Ramirez et al. (2021). In their `df.cis_trans.inherit.merge.csv` ([https://github.com/santiagosnchez/competitive\\_mapping\\_workflow/raw/refs/heads/master/analyses/tables/cis\\_trans\\_and\\_expression\\_inheritance/df.cis\\_trans.inherit.merge.csv](https://github.com/santiagosnchez/competitive_mapping_workflow/raw/refs/heads/master/analyses/tables/cis_trans_and_expression_inheritance/df.cis_trans.inherit.merge.csv)) dataset, 10,502 genes had assigned regulatory patterns, and 8,653 (82.4%) of these had diverged regulation (cis, trans, or compensatory; ambiguous excluded). These species differ at 20.7% of synonymous sites (Thomas et al. 2015), equivalent to a strain differing from N2 at 20,756,530 bp. We used these numbers allowed and our within-species *C. elegans* linear model to predict the proportion of regulatory-diverged genes between *C. nigoni* and *C. briggsae* and compare that to the observed proportion of regulatory-diverged genes and to genomic divergence.

## Genome, population genetic, and gene essentiality metrics

Genes were assigned to chromosome region bins (centers, arms, tips) based on which region from Rockman and Kruglyak (2009) contained the gene's midpoint. Nucleotide diversity statistics population-wide pairwise segregating sites ( $\pi$ ) and among-parental-pair proportion segregating sites  $p$  were calculated from the 20210121 hard-filter CeNDR VCF from biallelic SNVs only using PopGenome (v2.7.5) (Pfeifer et al. 2014). Nucleotide diversity ( $\pi$ ) and Tajima's  $D$  were also obtained from Lee et al. (2021), with their per-kb  $\pi$  per site converted to per-gene  $\pi$  per site by taking the median (missing data excluded) of all 1-kb windows overlapping the gene  $\pm 500$  bp. Tajima's  $D$ , Fay and Wu's  $H$ , and  $F_{ST}$  in non-Hawaiian and Hawaiian subpopulations were obtained from Ma et al. (2021). When we had multiple sources for the same statistic, we tested all of them, finding that results were generally consistent across statistic source when they were internally consistent across strains and gene sets; we use  $\pi$  from Lee et al. (2021) in the figures in this study. Whether the gene fell in a haplotype with a selective sweep in N2 was inferred from the swept haplotype data from Lee et al. (2021). To assign genes as essential or not, we downloaded gene annotations including "RNAi Phenotype Observed" and "Allele Phenotype Observed" for all genes in the *C. elegans* genome from WormBase using SimpleMine (Sternberg et al. 2024). Genes with lethality or sterility phenotypes from RNAi or alleles were considered essential (specifically, we searched for "lethal" and "steril" in the "RNAi Phenotype Observed" and "Allele Phenotype Observed" columns). Scripts used for these analyses are in our code repository: `data_generation_scripts/nucdivcendr_geneswindows_allandasestrains.R`, `data_analysis_scripts/chroclenrichment_asederpim.R`, and `data_analysis_scripts/aseetc_vs_general.R`.

## Expression-level analyses

For comparing gene categories to the expression level of each gene, we used the average normalized expression level from the 6 relevant parents in each cross. Specifically, kallisto quantification estimates to strain-specific transcriptomes were length and library size normalized followed by variance-stabilizing transformation (all via DESeq2) and then averaged across the appropriate samples.

For analyses of human gene expression variability vs human gene expression level, we used the S4 dataset from Wolf et al.

(2023), which comprises ranks of gene variation and expression level derived from principal components analysis of across-57-study correlation in gene expression variation and (separately) mean gene expression. Prior to this cross-study variance and level ranking, the authors corrected for the mean-variance relationship of gene expression within each study. We performed correlation tests on the input data as well as assigned genes to deciles of gene expression variability (1,313 or 1,314 genes per decile, 13,139 genes in dataset), testing these deciles for differences in central tendency of gene expression level via ANOVA.

For analyses of *Drosophila melanogaster* gene regulatory pattern and variation vs expression level, we used data from Glaser-Schmitt et al. (2024). We obtained ASE, DE, tests of their difference, and regulatory pattern assignments for midgut and hindgut from Data S3 and Data S4 and raw gene counts from the related GEO repository (GSE263264). To generate normalized expression values to use in our analyses, we determined the length and GC content of each transcript in the *D. melanogaster* genome from the transcript FASTA from FlyBase (release FB2024\_06) (Ozturk-Colak et al. 2024) and used the median across transcripts within a gene for the gene's value. We corrected count data from all samples from the parent strains for length bias (highly relevant) and GC bias (not very relevant in *D. melanogaster*) using R package `cqn` (v1.48) (Hansen et al. 2012) and then normalized to correct for library size and variance using DESeq2's `vst` and `normTransform` functions. The point estimate for each gene's expression was the mean across all parents' normalized values in the specific cross; we did all comparisons with both `vst` and `normTransform` normalized data, which yielded qualitatively similar results. Downstream, we used `normTransform` data as it removed more of these data's mean-variance correlation than did `vst`. To assign genes to regulatory patterns, we used Glaser-Schmitt et al. (2024)'s ASE and DE LFCs and their Cochran-Mantel-Haenszel test P-values comparing the parental and F1 allelic ratios and then followed our regulatory pattern assignment criteria without requiring a LFC threshold, as the original authors did not. Our regulatory pattern assignments were concordant with Glaser-Schmitt et al. (2024)'s except in cases where they had not taken directionality of effects into account and where we called genes as conserved if they did not have ASE or DE as called by DESeq2, even if they had a significant CMH based on underlying allelic ratios. Because patterns were consistent across strains, we plotted all strains' results together. We focused our interpretations on genes with changed expression regulation as determined by their ASE and DE estimates and comparisons between them, as the threshold for ASE-informative genes was different enough between the studies to preclude comparing genes without called expression changes.

Scripts used in these analyses are in our code repository: `data_analysis_scripts/aseetc_vs_general.R`, `data_analysis_scripts/wolf2023_humexpanalyses.R`, and `data_analysis_scripts/glaserschmitt_drosase_meanexprvsaseetc.R`.

## General software tools used for analyses and figures

Tools used for specific analytical purposes are described in the relevant sections; here, we share tools used for general data processing and figure creation.

Analysis scripts were largely written in R (v4.3.2) (R Core Team 2023), with a few written in Python (v3.7) ([www.python.org](http://www.python.org)). Workflow scripts were written and run using Nextflow (v22.10.7) ([www.nextflow.io](http://www.nextflow.io)). Compute-intensive analyses and workflows were run via the Partnership for an Advanced Computing

Environment (PACE), the high-performance computing environment at the Georgia Institute of Technology.

General data wrangling R packages used included *data.table* (v1.14.99) (Dowle and Srinivasan 2022), *argparser* (v0.7.1) (Shih 2021), and *formattable* (v0.2.1) (Ren and Russell 2021). R packages used for data display and figure creation included *ggplot2* (v3.5.1) (Wickham 2016), *cowplot* (v1.1.2) (Wilke 2020), *eulerr* (v7.0.2) (Larsson and Gustafsson 2018), *ggforce* (v0.4.1) (Pedersen 2022), *ggVennDiagram* (v1.2.3) (Gao 2021), *ggsignif* (v0.6.4) (Ahlmann-Eltze and Patil 2021), and *ggpmisc* (v0.5.6) (Aphalo 2024). Color schemes were developed using *RColorBrewer* (v1.1-3) (Neuwirth 2022) and Paul Tol's color palettes (<https://personal.sron.nl/~pault/>).

## Results

### An experiment to reveal extent and mode of gene expression variation in *C. elegans*

To interrogate intraspecific gene expression variation in *C. elegans*, we captured expression differences among the reference strain N2 and 7 wild strains. We estimated pairwise differential expression (DE) between each wild strain and N2, as well as allele-specific expression (ASE) in the F1 offspring of each strain crossed to N2 (Fig. 1a; Supplementary Table 1). ASE analyses are uniquely sensitive at identifying cis regulatory changes (Cowles et al. 2002; Yan et al. 2002; Wittkopp et al. 2004), and analyzed in conjunction with DE of parental strains, they can reveal the regulatory pattern and inheritance mode of gene expression across the genome (Fig. 1b). The 7 wild strains were chosen to represent a range of nucleotide divergence from N2 and spanned the species' genetic diversity: EG4348; DL238; CB4856 ("Hawaii"); ECA722; QX1211; and ECA701 and XZ1516, 2 extremely diverged strains (Fig. 1c).

To maximize power and limit confounding effects, we conducted the experiment in 1 batch, generating young adult selfed offspring of the parental strains simultaneously with their cross offspring with N2 (Fig. 1a, see 'Materials and Methods'). Replicate RNA-seq samples clustered in gene expression space, indicating true differences between strains and generations (principal components analysis, Supplementary Fig. 1). To analyze these gene expression data for signatures of DE and ASE, we developed a framework that (1) minimized reference bias, wherein sequence reads from the reference genome have higher rates of alignment than reads from the nonreference genome (Degner et al. 2009), (2) equivalently handled strains and genomes with varying levels of difference from each other while minimizing differences in power among strains as much as possible (Supplementary Note 1), and (3) generated comparable estimates of expression differences among strains (DE) and between alleles within the F1s (ASE), enabling direct comparison (see 'Materials and Methods'). Although the wild strains exhibit a substantial span in their genetic differentiation from the reference, we observed no reference bias; the proportion of reference alleles called per gene was tightly centered around 50% for all strains (Fig. 1d). To estimate DE among strains, we analyzed 18,647 genes with nominal expression (those with 10 or more total reads observed across samples). To detect ASE within the F1 hybrids, transcripts must carry genomic variant(s) that discriminate between the parental genotypes, so not all expressed genes permit ASE analysis. The genes informative for ASE comprised 22–53% of all nominally expressed genes; the proportion scales with genetic difference from N2 (Fig. 1e). In this manuscript, we refer to these as "ASE-informative" genes.

Here, we present the insights derived from these gene expression data for all *C. elegans* genes, including those in hyperdivergent

haplotypes (Lee et al. 2021; Moya et al. 2025), as global trends persisted across different gene inclusion criteria (see 'Discussion').

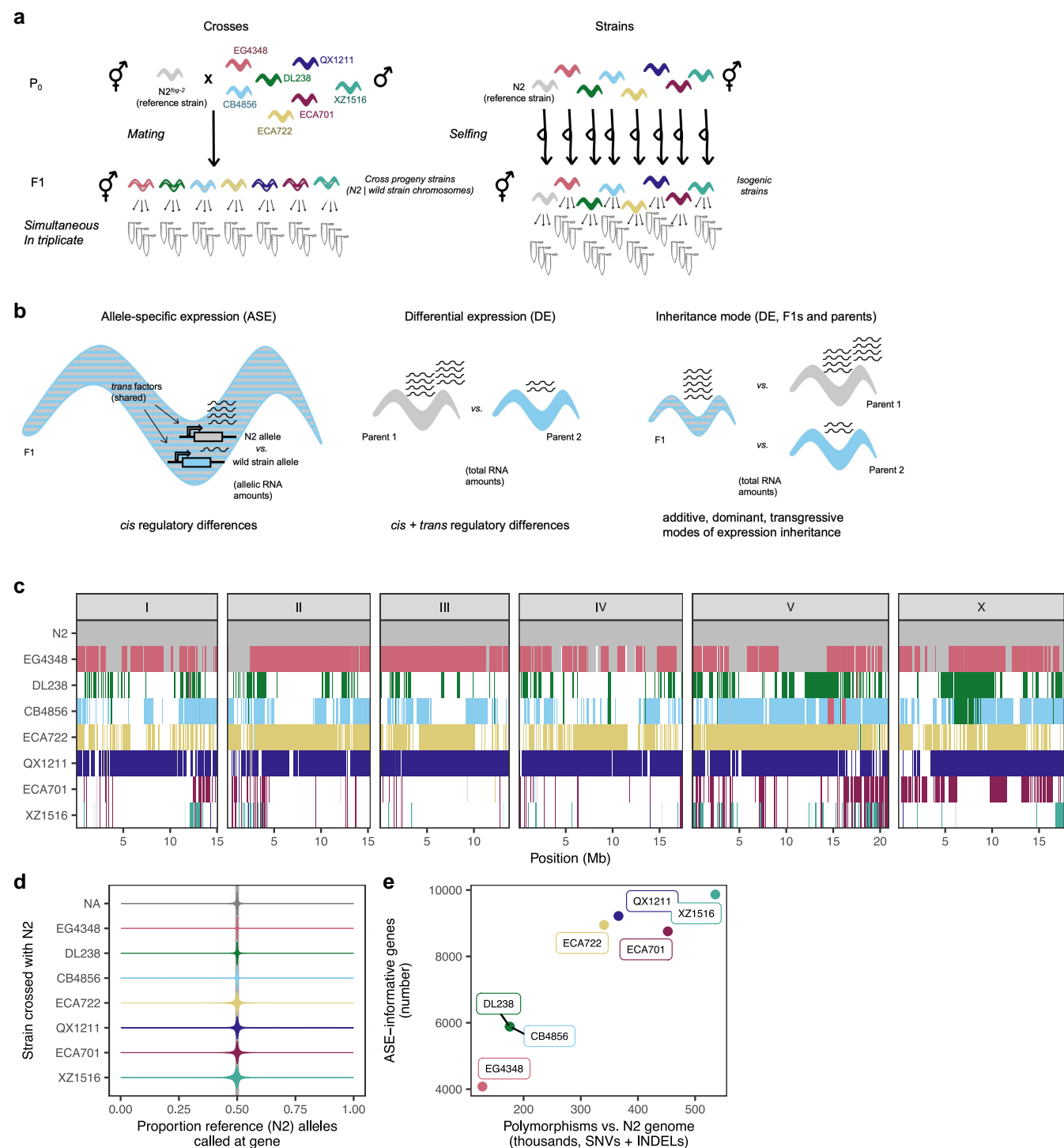
### Modes of gene expression inheritance include transgressive variation

Of all analyzed genes, 26% exhibited DE within at least 1 cross, either between the parental strains or between generations (Fig. 2). To evaluate how these expression changes were inherited, we compared, for each gene, the expression of the F1 offspring to each of its parents (McManus et al. 2010): we classified genes for which the F1 exhibits the same expression as one parent but different expression from the other as "N2 dominant" or "alt dominant" (wild strain dominant); genes with expression intermediate to the parents as "additive"; and genes with expression significantly higher or lower in the F1 than in both parents as transgressive, either "overdominant" or "underdominant" (Fig. 2a; Supplementary Fig. 2; see 'Materials and Methods'). Each cross exhibited genes with each inheritance classification (Fig. 2b), including overdominant and underdominant transgressive expression. Across strains, transgressive genes comprised 0.2–0.6% of all genes and 0.9–7.2% of genes with unambiguous expression changes. These transgressive genes are of outsized interest: expression values beyond the range of the parents may contribute to hybrid dysfunction and eventual speciation, and may help to explain outbreeding depression in *C. elegans* (Renaut et al. 2009; Gomes and Civetta 2015; Sanchez-Ramirez et al. 2021).

To determine whether expression of functionally related genes tends to be inherited the same way within and across strains, we performed gene set enrichment analyses (Supplementary Fig. 3) (Holdorf et al. 2020). Genes with transgressive expression were heavily and consistently enriched for collagen genes relative to all other categories (Fig. 2c). Yet, the pattern of expression varied by gene and by strain. Some collagen genes were lower expressed in the F1 than in either parent in all strains, with either idiosyncratic intermediate expression in the wild parent (e.g. *col-81*) or consistently high expression in both parents (e.g. *dpy-5*) (Fig. 2d). Strain XZ1516 often showed unique patterns, suggesting its collagen network may be strain-specifically regulated. At least some of the expression variation in collagen genes likely originates with the N2 genotype, which participated in each cross; N2 carries a derived mutation that modifies the phenotypic penetrance of cuticle mutations commonly used as markers in lab work (Noble et al. 2020). However, the differences by gene and expression patterns across strains suggest that collagen genes may be especially evolutionarily labile within this species, consistent with earlier work suggesting their rapid evolution and expression differentiation (Cutter and Ward 2005; Denver et al. 2005). Collagen genes interact in complex networks to form the worm cuticle (Higgins and Hirsh 1977; Cox et al. 1980; Kramer 1994; McMahon et al. 2003), and pathway architecture, including redundancies, may facilitate functional diversification across strains.

### Regulatory patterns reveal extensive cis-trans compensation within gene expression variation

We next sought to elucidate how expression differences are regulated in cis vs trans. For each ASE-informative gene, we compared the allele-specific difference in expression, which must occur in cis, to the expression difference between the parents, which may arise from regulation in cis, trans, or both (McManus et al. 2010). We classified genes with similar magnitude ASE and DE as regulated largely or solely in cis ("cis" category); genes with DE but no ASE as "trans"; and genes with DE exceeding ASE as

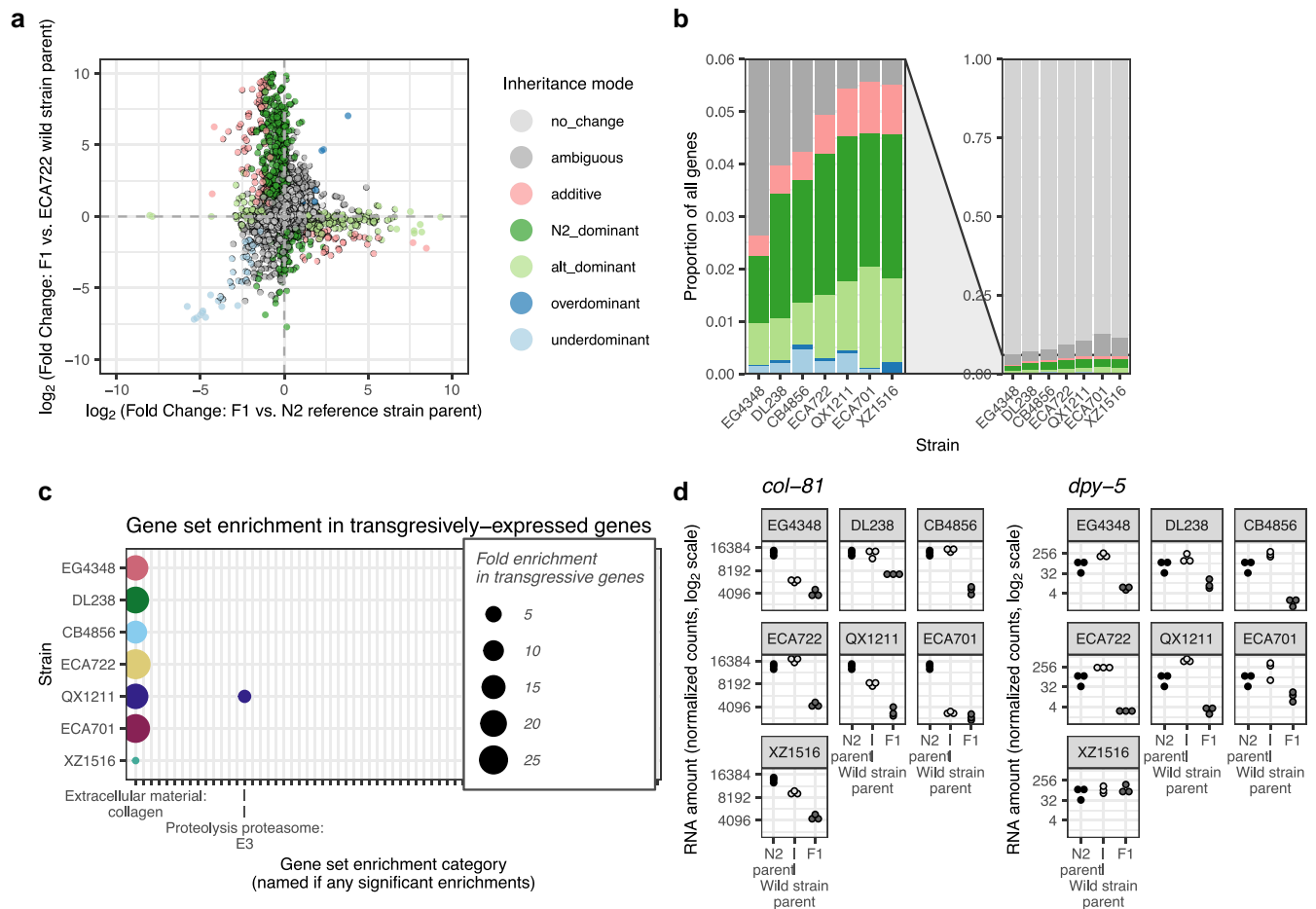


**Fig. 1.** Interrogating gene expression variation in wild *C. elegans*. a) Experimental regime. b) The 3 expression-level comparisons from this experiment. Left: ASE is estimated from per-allele, allele-specific read quantification within each set of F<sub>1</sub>s. Center: comparison of total RNA amounts between parental strains yields DE estimates. Comparisons of ASE and DE enable determination of regulatory pattern of expression differences. Right: comparison of total RNA amounts between the F<sub>1</sub> and its parents enables inference of inheritance mode of each gene's expression. c) Genetic similarity of the strains in this study. White indicates haplotypes containing variants private to this strain in the entire population, whereas colors show haplotypes shared with at least one other strain in the population. Specific color denotes the first strain in this study in which the given haplotype was observed; the same color shows that haplotype as identical-by-descent with at least one other strain in the entire population (haplotype identity-by-descent data from Lee et al. 2021). d) Proportion reference alleles in each ASE-informative gene's RNA-seq (see Supplementary Table 2 for all gene ns). e) The relationship between number of ASE-informative genes (see main text) to the genome divergence between the wild parental strain and reference genome N2.

“enhancing.” Genes with ASE but no or lower-magnitude DE reflect cis differences that are compensated by opposite-effect regulation in trans, producing either an incomplete offset of the cis

effect (“compensating”), a complete offset (“compensatory”), or an opposite effect exceeding the change in cis (“overcompensating”) (Fig. 3a; Supplementary Fig. 4; see ‘Materials and Methods’).





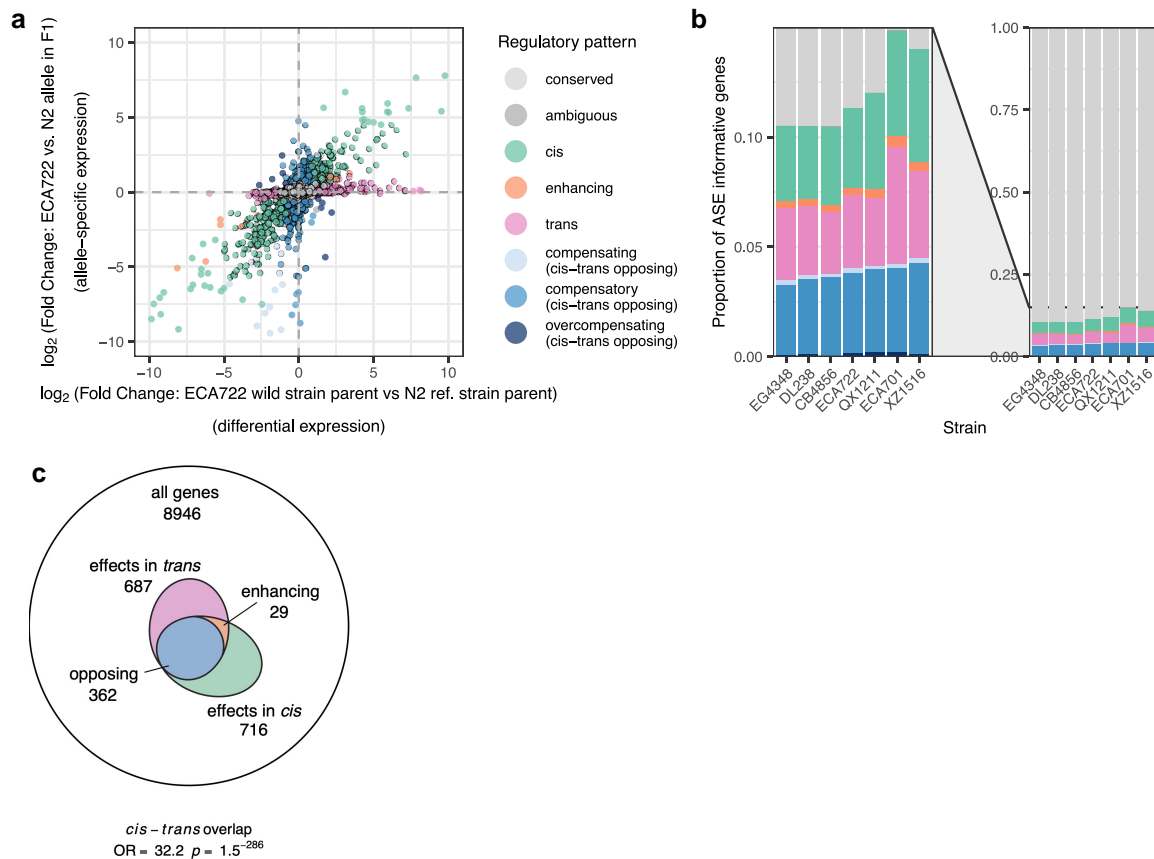
**Fig. 2.** Gene expression inheritance modes. a) For each cross, the inheritance mode of each gene was inferred by comparing DE between the F1 and their N2 parent (x axis) and DE between the F1 and their wild strain parent (y axis) (McManus et al. 2010). One point per analyzed gene, excluding 20 exceeding the axis limits. Supplementary Figure 2 shows this classification for all strains. b) Colors show the proportion of genes with each inheritance mode across strains (colors map to inheritance mode patterns as in a). Right: all genes; left: zoomed on genes with expression differences. Supplementary Table 2 has numbers (rather than proportions) of genes in each category in each strain cross. c) Gene set enrichment analysis results (Holdorf et al. 2020) for transgressively inherited underdominant genes, in which the F1 expression is lower than either parent, vs all analyzed genes. X axis ticks indicate all gene categories analyzed in this comparison; only significant enrichments are labeled (Bonferroni-adjusted  $P < 0.05$ ). Supplementary Figure 3 shows gene set enrichment analysis results for all analyzed inheritance mode gene sets. d) An example of 2 collagen genes with underdominant expression in multiple strains. N2 parental gene expression is the same in each subplot (the same 3 N2 samples serve as the N2 parent for all strains).  $n = 45$ . Web app [wildworm.biosci.gatech.edu/ase](http://wildworm.biosci.gatech.edu/ase) shows these plots and further information for any queried gene.

This classification regime operates equivalently across strains, enabling interstrain comparisons. Importantly, this method avoids a common pitfall wherein the influences of cis and trans effects are artifactually negatively correlated (Supplementary Note 2; Fraser 2019; Zhang and Emerson 2019). Such correlation can inflate inferences of compensation because any false positive estimates of cis effects, leading to spurious calls of ASE, result in equal-but-opposite trans false positives to explain the absence of observed DE (Supplementary Note 2; Fraser 2019; Zhang and Emerson 2019).

Across strain pairs, 9–15% of the ASE-informative genes exhibited expression differences in cis, trans, or a combination; roughly similar numbers of genes were regulated primarily in cis, primarily in trans, or both in opposite directions indicating compensatory regulation (Fig. 3b). As expected, the genes with no observed expression differences in the inheritance mode analysis were likewise classified as unchanged (“conserved”) or as compensatory in this analysis (Supplementary Fig. 5). Genes with expression differences driven solely in trans were much more likely to be inherited as dominant than as additive, and likewise genes with

expression differences driven solely in cis were more likely than trans genes to be inherited additively. This trend makes sense, as a trans factor inherited from 1 parent can target alleles from both while a change in cis affects only the local allele (Lemos et al. 2008). However, many cis-regulated genes were also inherited dominantly, as has been observed in other nematode species (Sanchez-Ramirez et al. 2021). While this result may reflect real biological mechanisms (Sanchez-Ramirez et al. 2021), we note that estimates of dominance may be inflated since the statistical threshold for additivity, which requires the intermediate F1 expression level to be distinct from both parents, is harder to achieve than that for dominance, which requires distinction from only 1 parent.

The substantial fraction of genes exhibiting compensatory regulation (Fig. 3b) reflects the ubiquity of opposing cis and trans effects at the same individual genes. For example, of genes with detected ASE, i.e. evidence of gene regulation in cis, 44–51% were inferred to be partially or completely attenuated in trans, resulting in no or reduced DE at the organismal level (Fig. 3c; Supplementary Fig. 6). This is a substantially higher overlap



**Fig. 3.** Gene expression regulatory patterns. a) The regulatory mode for each gene was inferred by comparing DE between the 2 parental strains (x axis) with ASE between the 2 alleles in the F1 (y axis) (McManus et al. 2010). One point per ASE-informative gene, excluding 10 exceeding the axis limits. [Supplementary Figure 4](#) shows this classification for all strains. b) Colors show the proportion of genes with each regulatory pattern in each strain cross (colors map to regulatory patterns as in a). Right: all genes; left: zoomed on genes with expression differences. [Supplementary Table 2](#) shows numbers (rather than proportions) of genes in each category in each strain. c) Relative to all genes, the subset of genes with regulatory differences in cis and in trans (in isolation or in combination), shown for strain *ECA722*; many more genes have both cis and trans effects than would be expected by chance if the effects are independent (hypergeometric/FET results in panel). Numbers show total number of genes in category inclusive of category-specific and overlapping genes. [Supplementary Figure 6](#) shows this result for all strains.

than would be expected by chance if the probabilities of cis and trans effects were independent (FET, all  $P < 2 \times 10^{-112}$ , [Supplementary Fig. 6](#)). Moreover, of genes regulated in both cis and trans, opposing effects (the compensatory classes) dominated, comprising 88–94% of these genes. This significant excess of genes with cis-trans overlapping effects is unlikely to be artifactual: for this enrichment of compensatory effects not to be significant, 90% of these calls would need to be in error, meaning at least 40% of our ASE calls would have to be spurious (false positives) or 35% of DE calls would have had to have been missed (false negatives) (see also [Supplementary Note 2](#)). Such pervasive opposition between regulatory effects may reflect a history of stabilizing selection on gene expression levels (see ‘Discussion’).

We next considered whether the expression differences we observed, and their modes of regulation, were common or idiosyncratic across strains. As all crosses shared *N2* as a parent, expression differences arising from derived changes in *N2* are likely to be shared; this is evidenced by the shared expression pattern at *fog-2*, which exhibited ASE in each cross (we deleted this spermatogenesis gene from the *N2* worms used to create the F1s, to facilitate obligate outcrossing, though we used wild-type *N2* for sequencing this “parental” strain). Strain pairs showed consistency in expression variation with respect to which genes exhibited

any differences at all: 1,798 (85%) of shared informative genes had the same regulatory pattern at 6 or more strains and 1,480 (70%) at all 7 strains ([Supplementary Table 3](#)); most of these genes shared conserved expression. However, of genes with expression differences, 82 exhibited cis-dominant regulation in at least 1 strain and trans-dominant regulation in at least another strain, and 109 genes exhibited uncompensated cis-driven expression changes in at least 1 strain that were compensated in at least 1 other strain ([Supplementary Table 3](#)). While apparent differences among strains may be inflated by the statistical challenge of observing a given gene as significant in multiple crosses, the overall trend is that genes with expression differences were strain specific: of genes that were ASE-informative in all strains, a preponderance (51.9%, 275 of 530) of those exhibiting ASE did so in only a single strain ([Supplementary Fig. 7](#) and [Supplementary Table 3](#)). Genes with specific regulatory patterns tended not to be obviously associated with common biological categories across strains ([Supplementary Fig. 8](#)).

### Expression variation increases with genomic divergence

*C. elegans* strains persist predominantly as selfing lineages, resulting in the accumulation of genetic changes and a spectrum of genomic differentiation between more closely or more

distantly related strains (Barriere and Felix 2005a; Barriere and Felix 2005b). We leveraged this aspect of *C. elegans* biology to assess the relationship between genomic differentiation and gene expression variation, asking whether the proportion of genes with expression differences changes with genomic differentiation. Overall, yes: for each strain, the proportion of genes with differences in expression scaled positively with genetic distance from N2 (Fig. 4).

This pattern persisted across all measures of expression variation: for genes exhibiting DE between strains (Fig. 4a), for genes by regulatory class (Fig. 4b–d; Supplementary Fig. 9), and for genes by inheritance mode (Supplementary Fig. 10), with weaker trends for the infrequently observed gene categories with less estimate precision. Considering all genes with evidence of differential regulation (all those not called conserved), we estimate that increasing the number of genetic variants by 100 thousand increases the proportion of variable expression genes by 1 percentage point (1%) (linear regression per 1,000 variants:  $\beta = 1.2 \times 10^{-4}$ ,  $P = 0.004$ ). This trend is not explained by the increased number of ASE-informative genes in more highly differentiated strains, as the estimates are specific to the ASE-informative genes for each strain (though see Supplementary Note 1).

Given that *C. elegans* live as independent lineages and show significant outbreeding depression (Dolgin et al. 2007), we had hypothesized that gene expression variation might plateau between the most diverged strains, as observed between species. For example, the nematode species *C. briggsae* and *C. nigoni* exhibit regulatory divergence at 82% of ASE-analyzed genes (Sanchez-Ramirez et al. 2021), which is 13.5-fold more expression differences than our most diverged *C. elegans* strain pair even as these sister species are 39-fold more diverged at the genome level (Thomas et al. 2015). However, here, gene expression differences scale with genomic differentiation and do not show any softening of the trend among the most diverged strains, as we might expect between incipient or distinct species.

## Location, nucleotide diversity, and essentiality define genes with expression differences

To consider processes that may have shaped gene expression variation, we interrogated gene sets with different regulatory patterns for association with genomic location, nucleotide diversity metrics, and gene essentiality.

The *C. elegans* genome harbors extensive evidence of its recombination history, with more recombination in the chromosome arms and less in chromosome centers (Rockman and Kruglyak 2009): gene density tends to be higher in the centers while nucleotide diversity is higher on arms (Rockman and Kruglyak 2009; Andersen et al. 2012). ASE-informative genes must have coding sequence polymorphisms; commensurately, they are enriched in chromosome arms and exhibit higher nucleotide diversity across all strains (Fig. 5a and b; Supplementary Figs. 11–13). However, even accounting for this background enrichment, genes with expression differences (in cis or trans) were more likely to reside on chromosome arms than on centers (Fig. 5a; Supplementary Fig. 11). These trends held for all 7 strains and reinforce prior observations that genes variably expressed across wild *C. elegans* strains are more likely to reside in arms, as shown for DE (Denver et al. 2005) and as mapped as eQTLs by linkage (Rockman et al. 2010) or by association (Zhang et al. 2022). Relatedly, genes with expression differences showed an excess of polymorphism—beyond that which makes them ASE-informative—and were associated with both elevated

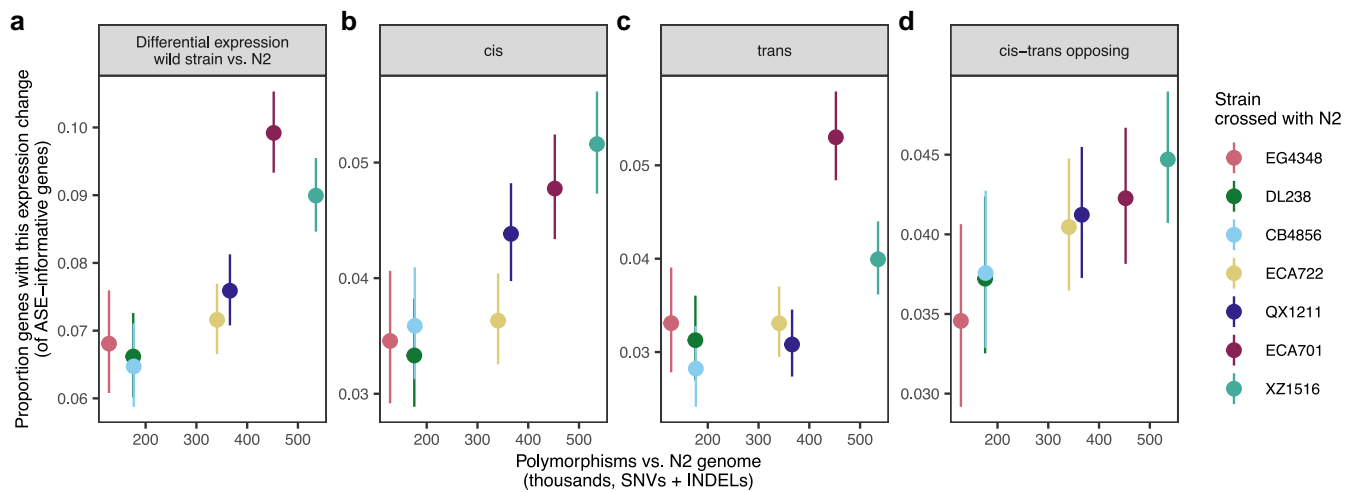
genetic differences between the 2 parents (Supplementary Fig. 12) and elevated nucleotide diversity across the species (Fig. 5c; Supplementary Fig. 13). This trend parallels recent findings in humans that genes with higher variation in expression harbor more genetic polymorphism (Wolf et al. 2023). This trend likely reflects differences in historical selection, including the possibility of selection relaxation (see ‘Discussion’); consistent with this, expression-stabilized genes exhibiting compensatory regulation with opposing cis and trans effects had lower nucleotide diversity and tended to be less enriched in chromosome arms, than non-compensated genes, but not to the extent shown by genes with completely conserved expression (Fig. 5a–c).

The *C. elegans* genome shows evidence of selective sweeps, in which haplotypes comprising large portions of individual chromosomes have risen in frequency across the population (Andersen et al. 2012; Lee et al. 2021). A footprint of strong historical selection, these sweeps dominate the genomes of non-Hawaiian isolates and may underlie adaptation associated with the colonization of new habitats (Zhang et al. 2021). We hypothesized that swept haplotypes are also associated with changes to gene expression. In our study, the non-Hawaiian strains N2 and EG4348 carry swept haplotypes over 65 and 37% of their genomes, respectively; the other Hawaiian strains harbor no swept haplotypes (Lee et al. 2021). Therefore, all our F1s share swept haplotypes inherited from N2, and only F1s derived from EG4348 carry additional swept haplotypes. Across strains, ASE-informative genes were less likely to reside in locations associated with N2 swept haplotypes (Fig. 5d; Supplementary Fig. 14). However, genes with cis regulatory differences (ASE) and genes with expression differences (DE) were both more likely to reside in locations associated with sweeps in N2 (Fig. 5d; Supplementary Fig. 14). These expression differences may have helped drive shifts in allele frequency and facilitated adaptation as *C. elegans* lineages colonized new habitats (Zhang et al. 2021), though the actual targets of selection within the swept haplotypes are unknown. Genes with cis regulatory differences compensated in trans tended to be less likely to be associated with swept haplotypes, but these trends were not always statistically significant across strains and gene sets (Fig. 5d; Supplementary Fig. 14).

Next, we asked whether gene essentiality was associated with differences in expression. Essential genes, defined as those with an RNAi or allele phenotype leading to lethality or sterility (Sternberg et al. 2024), were significantly depleted among genes with any observed expression differences (ASE or DE), even as informative genes were enriched for essentiality (Fig. 5e; Supplementary Fig. 15). These results reinforce earlier findings that essential genes are depleted among eQTL genes (Rockman et al. 2010; Zhang et al. 2022) and parallel observations from humans that genes with less expression variability tend to be less tolerant of loss of heterozygosity (Wolf et al. 2023). However, genes with opposing effects in cis and trans tended not to be depleted for essential genes (Fig. 5e; Supplementary Fig. 15). Essential genes are therefore likelier to have expression differences compensated by another mechanism, stabilizing their expression.

## Genes with expression differences are less highly expressed

Finally, we examined the relationship between overall expression level and the tendency of genes to show expression differences. As higher expression enables the detection of significant differences, a positive association might arise as an artifact of the method; in fact, genes informative for ASE were higher expressed



**Fig. 4.** Increasing expression variation with genomic divergence. For each wild strain, the proportion of genes exhibiting a given regulatory difference is plotted against the genetic distance from reference strain *N2*. a) DE between the wild strain and *N2* at ASE-informative genes; b)–d) proportion of genes with specific regulatory patterns in each strain cross. Cis–trans opposing genes comprise compensating, compensatory, and overcompensating gene categories from Fig. 3a and b. All nonconserved expression regulatory patterns combined: correlation  $\rho = 0.89$ ,  $P = 0.01$ . In all panels, error bars denote 95% binomial C.I. of proportion and y-axis bounds are specific to the data shown. [Supplementary Figure 9](#) shows proportion of each individual regulatory pattern category vs genetic distance from *N2*. [Supplementary Figure 10](#) shows proportion of each individual inheritance mode vs genetic distance from *N2*. See [Supplementary Table 2](#) for all gene ns.

than those not ASE-informative (Fig. 6a; [Supplementary Fig. 16](#)). Alternatively, a negative association between expression level and expression variation might reflect a phenomenon arising from biological function or evolutionary history.

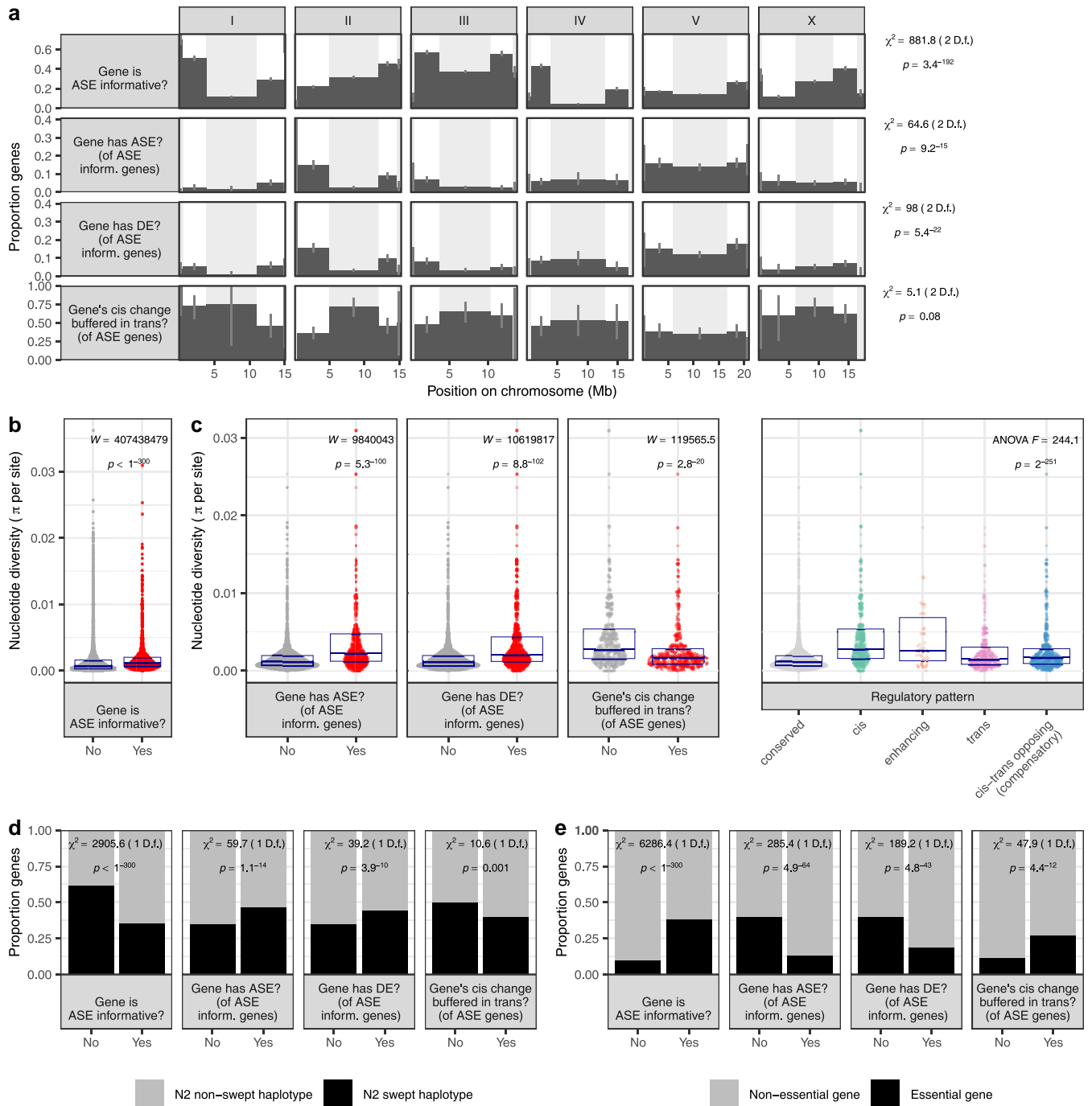
Here, genes with expression differences tended to exhibit lower average expression: of ASE-informative genes, those with detected ASE or DE were on average less expressed than those without (Fig. 6b; [Supplementary Fig. 16](#)). Moreover, genes with no or reduced DE between strains because of cis–trans opposing effects exhibited intermediate expression levels, higher than uncompensated genes with expression differences but lower than conserved expression genes (Fig. 6b; [Supplementary Fig. 16](#)). Because these genes with compensatory effects have higher expression than those with DE or uncompensated ASE, these calls are unlikely to be explained by missed DE calls or spurious ASE calls. To our knowledge, these are the first observations to explicitly demonstrate a relationship between gene expression level and gene expression variation, though genes with higher expression tend to evolve more slowly over interspecific timescales (Pal et al. 2001; Krylov et al. 2003; Liao and Zhang 2006; Zhang and Yang 2015). Thus, our results may reflect, at the intraspecific level within *C. elegans*, evidence of the constraint that has been hypothesized to govern the anticorrelation between expression and evolutionary rate (Zhang and Yang 2015) (Discussion). Because this pattern was clear in each strain, it is likely a general feature of *C. elegans* gene expression rather than a strain-specific idiosyncrasy ([Supplementary Fig. 16](#)).

To evaluate whether this relationship between gene expression level and variability extended beyond *C. elegans*, we examined expression data from humans and *D. melanogaster* ([Supplementary Fig. 17](#)). For humans, we reanalyzed data from a meta-analysis of gene expression studies, comprising 57 studies with a median of 251 individuals included per study, which ultimately generating a robust across-study rank of mean expression and expression variance for each gene that encompassed variation driven by genotype and other sources (Wolf et al. 2023). In our analysis, more variably expressed human genes tended to be less

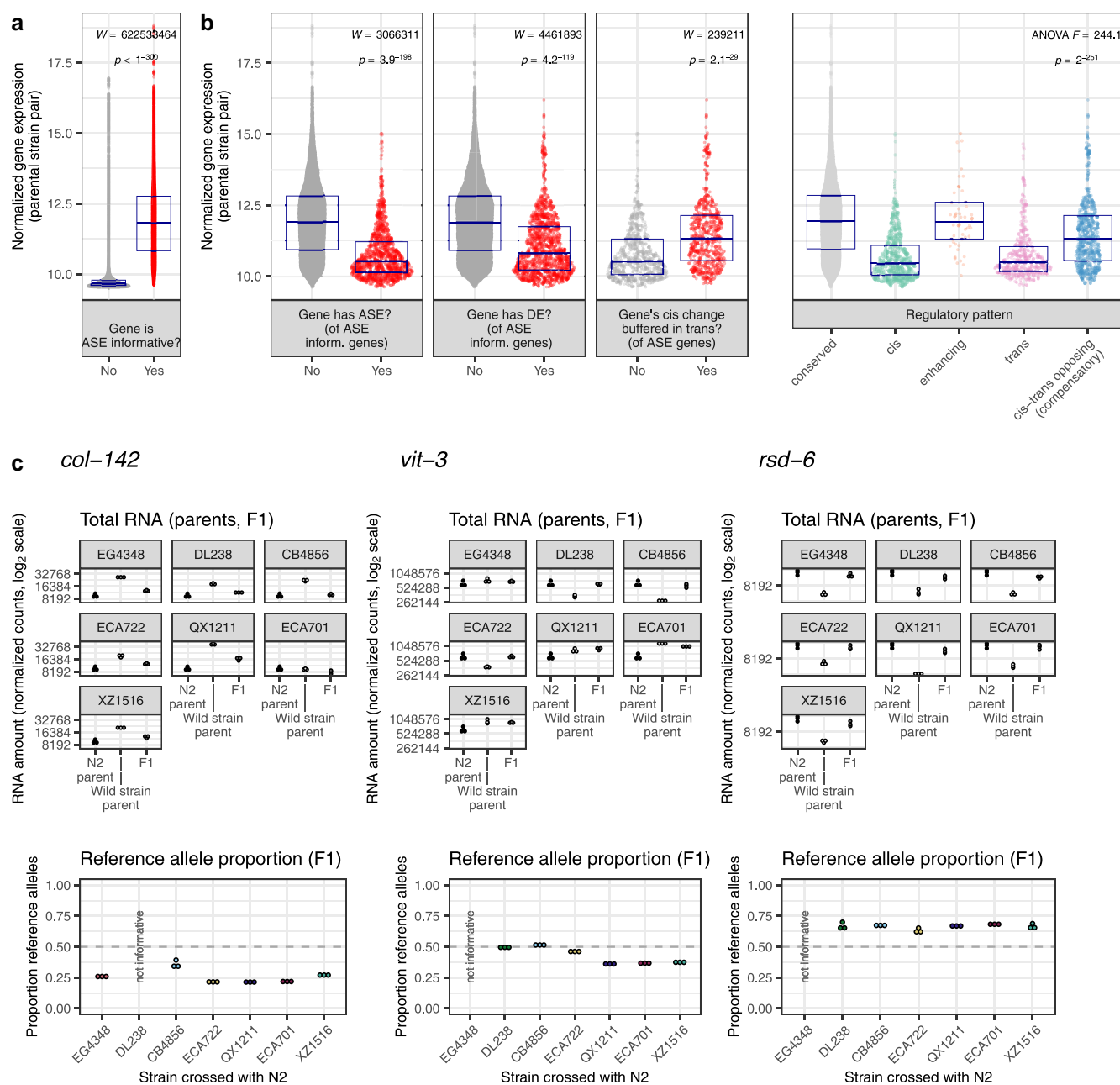
expressed; the relationship is modest in magnitude but statistically significant ([Supplementary Fig. 17a](#)). For *D. melanogaster*, we re-analyzed and carefully normalized data from a recent ASE study of intestinal tissues from 4 strain crosses (Glaser-Schmitt et al. 2024). As we saw in *C. elegans*, *D. melanogaster* genes with compensatory cis–trans opposing effects as well as enhancing cis–trans effects had higher average expression than genes regulated only in cis, but trans effects were also elevated ([Supplementary Fig. 17b](#)). How much of these similarities and differences derive from study-specific analysis pipelines vs species-specific evolution remain open questions.

The observation that differentially expressed genes have lower expression on average provides a platform for identifying potentially important outliers: genes with very high expression that nonetheless have expression differences might be targets of strain-specific adaptive evolution. Of genes in the top 10% of gene expression, 9 had cis-regulated DE in 1 or more strains ([Supplementary Table 4](#)). Anecdotally, these genes reflect dominant aspects of *C. elegans* biology: first, collagen genes *col-8* and *col-142* (Fig. 6c) are part of the extensive, epistatic network of genes coding for the collagen cuticle matrix. Second, vitellogenin genes *vit-3* (Fig. 6c) and *vit-5* code for extremely highly expressed yolk proteins that dominate young adult *C. elegans*’ mRNA and protein generation (Perez and Lehner 2019) and whose gene products are even hypothesized to be used for offspring provisioning as a sort of “milk” (Kern et al. 2021). Third, *rsd-6* (Fig. 6c) and *deps-1* are involved in the P granule and piRNA processing (Grishok 2013; Sternberg et al. 2024). Such small RNA pathways predominate worm biology and exhibit remarkable diversity in function and gene makeup across strains (Felix 2008; Youngman and Claycomb 2014; Chou et al. 2024). Although these genes exhibit similar high expression level and expression regulation, they are likely shaped by different evolutionary histories. For example, *rsd-6* is expressed at a lower level in all wild strains than in *N2*, suggesting an *N2*-specific mutation or function at this gene, while *vit-3* exhibits a diversification of expression levels across strains.





**Fig. 5.** Location, nucleotide diversity, haplotype, and essentiality differentiate expression diverged genes. The results shown here are for all strains combined (see 'Materials and Methods'). See [Supplementary Table 2](#) for all gene ns. a) Proportion of genes in each region of the chromosome (tip, arm, and center, denoted by alternating white and gray background) that have the described attribute. [Supplementary Figure 11](#) shows similar data for all strains individually. b) and c) Average per-site nucleotide diversity, estimated species-wide from 300+ wild *C. elegans* strains, is shown for genes with various expression patterns. Each point represents 1 gene and points fill a violin plot; boxes denote median  $\pm$  interquartile range. In c) (right), Tukey's HSD on annotated ANOVA cis > conserved ( $P = 9.8 \times 10^{-3}$ ); enhancing > conserved ( $P = 9.8 \times 10^{-3}$ ); trans > conserved ( $P = 9.8 \times 10^{-3}$ ); cis-trans opposing > conserved ( $P = 9.8 \times 10^{-9}$ ); cis > trans ( $P = 9.8 \times 10^{-9}$ ); cis > cis-trans opposing ( $P = 9.8 \times 10^{-9}$ ), enhancing > trans ( $P = 4.5 \times 10^{-5}$ ), enhancing > cis-trans opposing ( $P = 0.0003$ ) (all  $P$ -values Bonferroni corrected; other comparisons nonsignificant). [Supplementary Figure 12](#) shows pairwise, rather than population-wide, nucleotide diversity for all strains individually. [Supplementary Figure 13](#) shows same population-wide nucleotide diversity data for all strains individually. d) For various expression characteristics of interest, bars show the proportion of genes in a region in which there is evidence of historical positive selection (selective sweep) in the N2 parent. [Supplementary Figure 14](#) shows this breakdown for each strain individually. e) As in d), but each bar shows the proportion of genes in that category that are predicted to be essential in *C. elegans*. [Supplementary Figure 15](#) shows this breakdown for each strain individually.



**Fig. 6.** The relationship between expression level and expression variation. a), b) Results shown are for all strains combined (see 'Materials and Methods'). Y axis denotes gene expression amount (length and library size normalized and variance stabilized, averaged across the 2 parental strains). Each point represents 1 gene and points inhabit a violin plot; boxes denote median  $\pm$  interquartile range. See [Supplementary Table 2](#) for all gene ns. In b) (right), ANOVA Tukey's HSD conserved > cis ( $P = 9.6 \times 10^{-9}$ ); conserved > trans ( $P = 9.6 \times 10^{-9}$ ); conserved > cis-trans opposing ( $P = 9.6 \times 10^{-9}$ ); cis > enhancing ( $P = 9.6 \times 10^{-9}$ ), cis-trans opposing > cis ( $P = 9.6 \times 10^{-9}$ ), trans > enhancing ( $P = 9.6 \times 10^{-9}$ ), enhancing > cis-trans opposing ( $P = 0.014$ ); ( $P = 9.6 \times 10^{-9}$ ), cis-trans opposing > trans ( $P = 9.6 \times 10^{-9}$ ), (all  $P$ -values Bonferroni corrected; other comparisons nonsignificant). [Supplementary Figure 16](#) shows expression vs these various gene categories for all strains individually. c) Three example *C. elegans* genes with top 10% expression levels that nonetheless exhibit DE caused by cis regulatory divergence. Top: total gene expression for each sample. N2 samples are the same across plots/crosses. Bottom: within-sample allelic proportion from allelic counts.  $n = 3$  per strain per generation (45 total). Web app [wildworm.biosci.gatech.edu/ase](http://wildworm.biosci.gatech.edu/ase) shows these plots and further information for any queried gene.

## Discussion

### Main findings

Our study of intraspecific variation in gene expression included the first allele-specific analysis in *C. elegans* and elucidates the gene regulatory architecture of this system. We observed substantial compensatory regulation, in which opposite effects in cis and trans at individual genes mitigate expression differences among

strains. The most common gene expression pattern was one of no change and, while this majority gene set was well conserved among strains, genes with expression differences exhibited strain specificity and diversity in regulatory classification. We also report for the first time that expression-variable genes are lower expressed on average than genes without expression differences. Additionally, the number of expression-variable genes between strains increased with genomic distance. In addition to these

broad trends, we also highlighted genes with outlier expression patterns, such as collagen genes heavily enriched for transgressive variation (Fig. 2d) and expression-variable genes with extremely high average expression (Fig. 6c).

The compensation of gene expression changes by cis-trans opposing interactions (Fig. 3) has been observed across the tree of life, though with varying degrees of quantitative characterization and control for methodological artifacts (Supplementary Note 2) (Goncalves et al. 2012; Coolon et al. 2014; Fear et al. 2016; Mack et al. 2016; Verta et al. 2016; Metzger et al. 2017; Fraser 2019; Zhang and Emerson 2019). While these interactions are often interpreted as evidence for stabilizing selection on expression level, recent work in yeast has demonstrated how an initial mutation, by skewing the mutational target space, may predispose later mutations to induce opposing, compensatory effects on gene expression (McQueen et al. 2025). Under this regime, compensatory changes may accumulate neutrally, providing an alternative explanation for the significant excess of opposite effects relative to enhancing effects we observe in genes with changes in both cis and trans. Nevertheless, we found that genes exhibiting compensatory regulation were more likely to be essential than genes with uncompensated expression changes, suggesting stabilizing selection on expression. *C. elegans* may be especially prone to fitness-associated compensatory evolution due to extensive linkage across the genome arising from its predominantly selfing mode of reproduction (Barriere and Felix 2005a; Barriere and Felix 2005b; Rockman and Kruglyak 2009): fitness in *C. elegans* is mediated by opposite-effect, closely linked regions of the genome (Bernstein et al. 2019), and compensatory regulatory elements are closely linked in self-fertilizing spruce trees (cis-trans) (Verta et al. 2016) and yeast (trans-trans) (Metzger and Wittkopp 2019).

Our discovery that genes with expression differences tended to be expressed at lower levels (Fig. 6) may have been overlooked previously given that most studies control for the positive correlation between mean and variance in RNA quantification, discouraging investigation into the larger phenomenon. Indeed, breaking this relationship is difficult, which limited our ability to assess this phenomenon in published data from other systems. However, we did also see evidence of this trend in humans and *D. melanogaster* (Supplementary Fig. 17). These sparse findings may relate to the more widely observed, and more thoroughly considered, phenomenon in which highly expressed genes evolve at a slower evolutionary rate (Pal et al. 2001; Krylov et al. 2003; Liao and Zhang 2006; Zhang and Yang 2015). This expression-rate anticorrelation is universal across the domains of life and has been explained by a number of mechanisms independent of selection on biological essentiality: avoidance of protein misfolding or mis-interaction, selection on mRNA folding, and cost of protein expression (Zhang and Yang 2015). Very few studies have explicitly investigated this phenomenon over microevolutionary timescales, but venom protein expression in pit vipers evolves more slowly, i.e. is less variable, at higher levels (Margres et al. 2016), and higher expressed genes in yeast exhibit lower nucleotide polymorphism due to purifying selection directly targeting amino acid identity (Marek and Tomala 2018). We posit that our observed anticorrelation of gene expression level and variability is likely a function of stabilizing selection on higher expressed genes in *C. elegans*, but the fitness-associated traits under selection may relate to the efficiency or efficacy of gene products rather than, or in addition to, ultimate biological function. We also note that the mutations that define the expression-rate anticorrelation must arise within populations, and exploration of the question within a tractable experimental system enables tests of longstanding hypotheses

about the driving mechanisms (Biesiadecka et al. 2020). Overall, the expression level-variability anticorrelation invites many questions, including whether it occurs as a common feature of heritable expression variation in most systems; whether it scales to interspecific differences in expression; whether and how it translates to other molecular phenotypes, such as the expression of proteins; and the proximate and evolutionary forces that govern it.

We found that as genomic differentiation between the wild strains and N2 increased, the proportion of genes with expression differences also increased (Fig. 4; see also Supplementary Note 1 for potential caveats). Such a trend is not necessarily to be expected: regulatory divergence has been observed to scale with genetic divergence among marine-freshwater ecotypes in sticklebacks (Verta and Jones 2019) but to plateau at high genetic divergence between yeast species (Metzger et al. 2017) and to not necessarily increase with divergence within and among *Drosophila* species, but to accelerate in particular crosses (Coolon et al. 2014). Though analyses of this relationship between genomic and regulatory divergence can shed light on the evolution of the genotype-phenotype map and the interplay between genetic variation, gene expression, and speciation (Orr 1995; Mack and Nachman 2017), it remains incompletely understood. Ultimately, this correlation must plateau at sufficient evolutionary distance, as genomic differences accumulate but not all genes show expression divergence; while 82% of genes exhibit regulatory divergence between sister species *C. briggsae* and *C. nigoni* (Sanchez-Ramirez et al. 2021), their genomic distance would need to be 2.9 times higher to scale with the trend we see within *C. elegans*. While it is difficult to directly compare ASE studies due to differences in experimental and analytical frameworks, we note similarities in the inheritance modes and regulatory patterns observed within *C. elegans* and between *C. briggsae* and *C. nigoni*, including the extent of compensatory regulation (Sanchez-Ramirez et al. 2021). The relationship between gene expression divergence and genomic divergence within *C. elegans* may offer an access point for deeper investigation within a highly tractable genetic system.

Consistent with earlier work testing the neutrality of *C. elegans* gene expression variation using mutation accumulation lines (Denver et al. 2005), our observations suggest that the majority of *C. elegans* genes may be under stabilizing selection for expression level. Genes with expression changes exhibit higher nucleotide variation, lower average expression, and a depletion of essentiality, potentially reflecting a history of relaxed selection relative to genes with stabilized expression. This result also comports with the observation that population structure estimated from gene expression data of 207 *C. elegans* wild strains is less differentiated than that estimated from nucleotide data, suggesting possible stabilizing selection on expression relative to nucleotide differentiation (Zhang et al. 2022), and with a recent analysis that inferred stabilizing selection at most *C. elegans* transcripts (Inskeep and Groen 2025). However, in our results, genes with expression changes were also more likely to reside on chromosome arms than centers; this pattern may be explained in part by stronger stabilizing selection on expression in center-located genes but is also likely shaped by background selection, which acts independently of the expression phenotype of any individual gene (Rockman et al. 2010). Consequently, the role of stabilizing selection on expression phenotypes vs the indirect influence of background selection on reducing standing variation in chromosomal centers is not clear. Our observation that genes in chromosome centers show elevated compensation of expression

differences by cis-trans opposing effects (Fig. 5a) suggests that these may be direct targets of stabilizing selection.

We also observed evidence for adaptive evolution of gene expression variation. Namely, genes with expression differences were more likely to reside in locations at which the N2 haplotype experienced a selective sweep, which may include genes that facilitated adaptation during colonization of new habitats (Zhang et al. 2021). However, as in the case of background selection, such associations with genomic location do not distinguish between direct targets of selection and those indirectly targeted through linkage. Moreover, strain-specific changes in gene expression, even as they are associated with higher nucleotide diversity and other markers of relaxed selection, may instead reflect adaptive diversification; we cannot distinguish between these alternatives. For genes involved in environmental sensitivity or immune response, for example, the lower expression of expression-variable genes may be mediated by the absence of pathogens or other inducible factors in the lab environment. Thus, genes with expression differences yet high average expression may be useful candidates for identifying fitness-associated traits governed by expression variation.

In our study, each wild strain was crossed to the common reference strain N2, so N2-specific differences such as laboratory-derived adaptations would likely show up as common differences across the strain set. We observed only a small number of genes with common differences across all wild strains; instead, many genes had expression differences only in a single wild strain (Supplementary Fig. 7 and Supplementary Table 3). Genes in the worm cuticle network exhibited both shared and strain-specific trends. For example, most wild strains showed transgressive expression at the same collagen genes (Fig. 2c–d), suggesting N2-specific differentiation. This result may relate to the derived mutation in *col-182* in N2, which increases the phenotypic penetrance of classical lab mutations affecting cuticle phenotype (such as *rol-1*) that are suppressed in the ancestral background (Noble et al. 2020). However, strain XZ1516 and its F1s exhibited distinct collagen gene expression phenotypes, suggesting divergent evolution in collagen or cuticle pathways along the XZ1516 lineage. The collagen gene network is especially large and complex with evidence of accelerated adaptation (Cox et al. 1980; Kramer 1994; McMahon et al. 2003; Cutter and Ward 2005), features that might facilitate lineage-specific changes arising from directional selection on function or from diversification under either stabilizing or relaxed selection. Anecdotally, in our hands, XZ1516 was difficult to manipulate on the plate, which we hypothesize may be due to a sensitive cuticle. Moreover, another wild strain, XZ1514, was so fragile that we refrained from using it in this study, suggesting potential further genetic differentiation in collagen function across *C. elegans*.

## Comments about experimental system and design

Controlling for confounding variation poses a particular challenge in gene expression studies. For example, wild strains mature at different rates (Hodgkin and Doniach 1997; Gems and Riddle 2000; Harvey and Viney 2007; Pouillet et al. 2015; Stastna et al. 2015; Zhang et al. 2021). We observed differences in developmental rate among our experimental strains, including that parental strain QX1211, and to a lesser extent XZ1516, its F1 with N2, and the N2 parent, developed more slowly than other strains (Supplementary Table 1). While most F1 offspring developed at a rate similar to 1 parent or intermediate between the parents, the F1 offspring of QX1211 and N2 reached young adulthood

over an hour faster than either parent (Supplementary Table 1). To reduce the influence of developmental variation on gene expression differences, we harvested worms at a consistent developmental stage rather than a consistent chronological age, nevertheless all within 3 h of one another (see ‘Materials and Methods’). Further, we estimated the transcriptional age of each sample using an N2 gene expression time course as a “ruler” (Bulteau and Francesconi 2022); all estimates fell within a 5.5-h time range (Supplementary Table 1). These computational estimates differed across samples within strains even though such samples appeared identical and were harvested at the same time, suggesting further work is needed to understand discordance between experimental observations and computational predictions as well as interindividual timing variation. While these measures control for developmental differences as much as is experimentally tractable, they do not guarantee the elimination of confounding developmental variation at the level of the whole organism or in tissue- or process-specific events. Thus, conclusions around specific genes should be considered carefully; for example, small differences in fertilization timing might contribute to substantial differences in the highly expressed vitellogenin genes (Fig. 6c).

In this study, we chose the number of strains to maximize the range of genetic diversity while performing crosses simultaneously in triplicate. Maximizing the number of strains did restrict the direction of the crosses, however: we always used N2 as the maternal parent and did not include reciprocal crosses to reduce the burden of effort. The nonreciprocal cross design thus limits inferences about parent-specific transgenerational effects or sex-specific gene regulation, which have biological significance in other *Caenorhabditis* nematodes (Sanchez-Ramirez et al. 2021; Viswanath and Cutter 2023). For the same reason, we generated the parental strains via selfing rather than by crossing, while the F1s were obligately generated by crossing; these differences in reproductive mode may conflate with differences between F1s and parents, though we are unaware of any such broad effects on gene expression. An analytical concern with using multiple, variably diverged strains is whether ASE inferences could be standardized across strains. In theory, different strains could be differently vulnerable to reference genome bias or could exhibit increasing power to detect ASE with increasing variants within a given gene. However, the methods used here largely avoid these pitfalls (Fig. 1d; Supplementary Note 1).

Our inferences in this study, including expression classifications and trends between differently regulated genes, were robust to the inclusion or exclusion of genes in hyperdivergent haplotypes (Lee et al. 2021). Hyperdivergent regions differ substantially from the N2 reference sequence, making alignment and variant calling from short-read data unreliable (Lee et al. 2021; Moya et al. 2025); recent RNA-seq studies in *C. elegans* sensibly and conservatively exclude genes in these regions (Lee et al. 2021; Zhang et al. 2022). However, our recent gene expression analyses showed that genome-wide trends appear robust to including or excluding genes in hyperdivergent haplotypes (Bell et al. 2023). Therefore, here we performed each of our genome-wide analyses both including all genes and excluding genes classified as hyperdivergent as well as genes with evidence of other possible analytical hurdles (see ‘Materials and Methods’). The vast majority of trends detected when all genes were included were recapitulated when excluding hyperdivergent genes. We note, though, that results at individual genes are still likely to be influenced by genomic context, so these features should be considered when assessing small numbers of genes or conducting gene-specific queries. For example, our



gene set enrichment analysis results (Fig. 2d; Supplementary Figs. 3 and 8) were similar when including or excluding hyperdivergent genes, and whenever specific genes were used as exemplars of trends these genes were not hyperdivergent or otherwise concerning (e.g. Figs. 2d and 6c).

In this study, we focused on global, large-scale patterns in gene-level expression and did not quantify specific isoforms. However, recent evidence and common sense suggest that wild strains differ in expression of specific transcripts (Zhang et al. 2022). The extent to which nonreference strains express novel isoforms and how F1 cross progeny mediate the expression of parent-specific isoforms remain unexplored questions. A particularly intriguing possibility is that transgressive isoforms could be expressed in F1 heterozygous backgrounds but not in their native background, akin to cis regulatory changes that are revealed in hybrids but compensated among the parents.

### Interactive web application

Our experimental approach had many advantages, among them our study organism: the wealth of experimental data in *C. elegans* and its curation and accessibility via WormBase (Sternberg et al. 2024) makes this system especially amenable to analyses that add new molecular detail to existing experimental phenotypes. In turn, our characterization of gene expression variation improves our understanding of *C. elegans* and the processes that shape gene expression in this system. To aid in future genetics, trait mapping, and other *C. elegans* research, we have made the data from this study accessible via an interactive web application, where users can query their favorite gene to view its expression, regulatory pattern, inheritance mode, and other information: <https://wildworm.biosci.gatech.edu/ase/>.

### Data availability

Raw and processed gene expression data are available at GEO with accession number GSE272616. Per-gene per-strain data (used to perform all analyses and generate all figures), including regulatory pattern and inheritance mode classifications and underlying statistical DE results, are available via the Zenodo repository at <https://doi.org/10.5281/zenodo.13270636>. Per-gene information is interactively available via user query at web app <https://wildworm.biosci.gatech.edu/ase/>. Code used in this study's data processing and analysis is available at <https://github.com/paabylab/wormase>. Materials and methods fully describe all existing and new software and analyses used in this study.

Supplemental material available at GENETICS online.

### Acknowledgments

The authors thank Samiksha Kaul and Ling Wang for help with mating plate set up and male removal during the experiment and Han Ting Chou for undertaking a similar pilot experiment and sharing her expertise. Matt Rockman, members of the Rockman lab, Steve Burger, and 3 anonymous reviewers contributed helpful discussions about and suggestions for improving this work. The authors thank Shweta Biliya, Naima Djeddar, and Anton Bryskin at the Molecular Evolution Core Laboratory at Georgia Tech for their expert library preparation and sequencing guidance. Troy Hilley of Academic & Research Computing Services at Georgia Tech's College of Sciences provided expert web server configuration support for the interactive web app. The authors gratefully acknowledge individuals in the worm community who have collected and disseminated wild *C. elegans*

isolates. This research was supported in part through research cyberinfrastructure resources and services provided by the PACE at Georgia Tech.

### Funding

This work was funded by an National Science Foundation Postdoctoral Research Fellowship in Biology 2109666 to A.D.B., a National Institutes of Health grant R35 GM119744 to A.B.P., and support from the Georgia Institute of Technology.

### Conflicts of interest

The author(s) declare no conflict of interest.

### Literature cited

- Adduri A, Kim S. 2024. Ornaments for efficient allele-specific expression estimation with bias correction. *Am J Hum Genet.* 111(8): 1770–1781. doi:10.1016/j.ajhg.2024.06.014.
- Ahlmann-Eltze C, Patil I. 2021. ggsignif: R Package for Displaying Significance Brackets for 'ggplot2'. *PsyArxiv*. <https://doi.org/10.31234/osf.io/7awm6>.
- Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, Felix MA, Kruglyak L. 2012. Chromosome-scale selective sweeps shape *Caenorhabditis elegans* genomic diversity. *Nat Genet.* 44(3):285–290. doi:10.1038/ng.1050.
- Andersen EC, Rockman MV. 2022. Natural genetic variation as a tool for discovery in *Caenorhabditis* nematodes. *Genetics.* 220(1): iyab156 doi:10.1093/genetics/iyab156.
- Aphalo PJ. 2024. ggpmisc: Miscellaneous Extensions to 'ggplot2'. Accessed 1 February 2025. <https://CRAN.Rproject.org/package=ggpmisc>.
- Artieri CG, Fraser HB. 2014. Evolution at two levels of gene expression in yeast. *Genome Res.* 24(3):411–421. doi:10.1101/gr.165522.113.
- Bao Y, Hu G, Grover CE, Conover J, Yuan D, Wendel JF. 2019. Unraveling cis and trans regulatory evolution during cotton domestication. *Nat Commun.* 10(1):5399. doi:10.1038/s41467-019-13386-w.
- Barriere A, Felix MA. 2005a. High local genetic diversity and low outcrossing rate in *Caenorhabditis elegans* natural populations. *Curr Biol.* 15(13):1176–1184. doi:10.1016/j.cub.2005.06.022.
- Barriere A, Felix MA. 2005b. Natural variation and population genetics of *Caenorhabditis elegans*. *WormBook.* 1–19. doi:10.1895/wormbook.1.43.1.
- Bell AD, Chou HT, Valencia F, Paaby AB. 2023. Beyond the reference: gene expression variation and transcriptional response to RNAi in *C. elegans*. *G3 (Bethesda).* 13(8):jkad112. doi:10.1093/g3journal/jkad112.
- Bell AD, Valencia F, Paaby AB. 2024. Generating multiple stage-matched *C. elegans* hybrids and parental strains simultaneously. *protocols.io.* doi:10.17504/protocols.io.5jyl8p15rg2w/v1.
- Bernstein MR, Zdrachek S, Andersen EC, Rockman MV. 2019. Tightly linked antagonistic-effect loci underlie polygenic phenotypic variation in *C. elegans*. *Evol Lett.* 3(5):462–473. doi:10.1002/evl3.139.
- Biesiadecka MK, Sliwa P, Tomala K, Korona R. 2020. An overexpression experiment does not support the hypothesis that avoidance of toxicity determines the rate of protein evolution. *Genome Biol Evol.* 12(5):589–596. doi:10.1093/gbe/evaa067.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

- Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol.* 34(5):525–527. doi:[10.1038/nbt.3519](https://doi.org/10.1038/nbt.3519).
- Brion C, Lutz SM, Albert FW. 2020. Simultaneous quantification of mRNA and protein in single cells reveals post-transcriptional effects of genetic variation. *Elife.* 9:e60645 doi:[10.7554/eLife.60645](https://doi.org/10.7554/eLife.60645).
- Buccitelli C, Selbach M. 2020. mRNAs, proteins and the emerging principles of gene expression control. *Nat Rev Genet.* 21(10):630–644. doi:[10.1038/s41576-020-0258-4](https://doi.org/10.1038/s41576-020-0258-4).
- Bulteau R, Francesconi M. 2022. Real age prediction from the transcriptome with RAPToR. *Nat Methods.* 19(8):969–975. doi:[10.1038/s41592-022-01540-0](https://doi.org/10.1038/s41592-022-01540-0).
- Cartwright EL, Lott SE. 2020. Evolved differences in cis and trans regulation between the maternal and zygotic mRNA complements in the *Drosophila* embryo. *Genetics.* 216(3):805–821. doi:[10.1534/genetics.120.303626](https://doi.org/10.1534/genetics.120.303626).
- Chou HT, Valencia F, Alexander JC, Bell AD, Deb D, Pollard DA, Paaby AB. 2024. Diversification of small RNA pathways underlies germline RNA interference incompetence in wild *Caenorhabditis elegans* strains. *Genetics.* 226(1):iyad191. doi:[10.1093/genetics/iyad191](https://doi.org/10.1093/genetics/iyad191).
- Cook DE, Zdravljec S, Roberts JP, Andersen EC. 2017. CeNDR, the *Caenorhabditis elegans* natural diversity resource. *Nucleic Acids Res.* 45(D1):D650–D657. doi:[10.1093/nar/gkw893](https://doi.org/10.1093/nar/gkw893).
- Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ. 2014. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Res.* 24(5):797–808. doi:[10.1101/gr.163014.113](https://doi.org/10.1101/gr.163014.113).
- Cowles CR, Hirschhorn JN, Altshuler D, Lander ES. 2002. Detection of regulatory variation in mouse genes. *Nat Genet.* 32(3):432–437. doi:[10.1038/ng992](https://doi.org/10.1038/ng992).
- Cox GN, Laufer JS, Kusch M, Edgar RS. 1980. Genetic and phenotypic characterization of roller mutants of CAENORHABDITIS ELEGANS. *Genetics.* 95(2):317–339. doi:[10.1093/genetics/95.2.317](https://doi.org/10.1093/genetics/95.2.317).
- Crombie TA, McKeown R, Moya ND, Evans KS, Widmayer SJ, LaGrassa V, Roman N, Tursunova O, Zhang G, Gibson SB, et al. 2024. CaenDR, the *Caenorhabditis* natural diversity resource. *Nucleic Acids Res.* 52(D1):D850–D858. doi:[10.1093/nar/gkad887](https://doi.org/10.1093/nar/gkad887).
- Crombie TA, Zdravljec S, Cook DE, Tanny RE, Brady SC, Wang Y, Evans KS, Hahnel S, Lee D, Rodriguez BC, et al. 2019. Deep sampling of Hawaiian *Caenorhabditis elegans* reveals high genetic diversity and admixture with global populations. *Elife.* 8:e50465. doi:[10.7554/eLife.50465](https://doi.org/10.7554/eLife.50465).
- Crowley JJ, Zhabotynsky V, Sun W, Huang S, Pakatci IK, Kim Y, Wang JR, Morgan AP, Calaway JD, Aylor DL, et al. 2015. Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nat Genet.* 47(4):353–360. doi:[10.1038/ng.3222](https://doi.org/10.1038/ng.3222).
- Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol Biol Evol.* 22(1):178–188. doi:[10.1093/molbev/msh267](https://doi.org/10.1093/molbev/msh267).
- Degner JF, Marioni JC, Pai AA, Pickrell JK, Nkadori E, Gilad Y, Pritchard JK. 2009. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics.* 25(24):3207–3212. doi:[10.1093/bioinformatics/btp579](https://doi.org/10.1093/bioinformatics/btp579).
- Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat Genet.* 37(5):544–548. doi:[10.1038/ng1554](https://doi.org/10.1038/ng1554).
- Dolgin ES, Charlesworth B, Baird SE, Cutter AD. 2007. Inbreeding and outbreeding depression in *Caenorhabditis* nematodes. *Evolution.* 61(6):1339–1352. doi:[10.1111/j.1558-5646.2007.00118.x](https://doi.org/10.1111/j.1558-5646.2007.00118.x).
- Dowle M, Srinivasan A. 2022. data.table: Extension of 'data.frame'. Accessed 1 February 2025. <https://Rdatatable.gitlab.io/data.table>.
- Evans KS, Andersen EC. 2020. The gene *scb-1* underlies variation in *Caenorhabditis elegans* chemotherapeutic responses. *G3 (Bethesda).* 10(7):2353–2364. doi:[10.1534/g3.120.401310](https://doi.org/10.1534/g3.120.401310).
- Fay JC, Wittkopp PJ. 2008. Evaluating the role of natural selection in the evolution of gene regulation. *Heredity (Edinb).* 100(2):191–199. doi:[10.1038/sj.hdy.6801000](https://doi.org/10.1038/sj.hdy.6801000).
- Fear JM, Leon-Novelo LG, Morse AM, Gerken AR, Van Lehmann K, Tower J, Nuzhdin SV, McIntyre LM. 2016. Buffering of genetic regulatory networks in *Drosophila melanogaster*. *Genetics.* 203(3):1177–1190. doi:[10.1534/genetics.116.188797](https://doi.org/10.1534/genetics.116.188797).
- Felix M-A. 2008. RNA interference in nematodes and the chance that favored Sydney Brenner. *J Biol.* 7(9):34. doi:[10.1186/jbiol97](https://doi.org/10.1186/jbiol97).
- Francesconi M, Lehner B. 2014. The effects of genetic variation on gene expression dynamics during development. *Nature.* 505(7482):208–211. doi:[10.1038/nature12772](https://doi.org/10.1038/nature12772).
- Fraser HB. 2019. Improving estimates of compensatory cis-trans regulatory divergence. *Trends Genet.* 35(1):88. doi:[10.1016/j.tig.2018.10.003](https://doi.org/10.1016/j.tig.2018.10.003).
- Fraser HB. 2022. Existing methods are effective at measuring natural selection on gene expression. *Nat Ecol Evol.* 6(12):1836–1837. doi:[10.1038/s41559-022-01889-7](https://doi.org/10.1038/s41559-022-01889-7).
- Frezal L, Demoinet E, Braendle C, Miska E, Felix MA. 2018. Natural genetic variation in a multigenerational phenotype in *C. elegans*. *Curr Biol.* 28(16):2588–2596.e8. doi:[10.1016/j.cub.2018.05.091](https://doi.org/10.1016/j.cub.2018.05.091).
- Frézal L, Félix M-A. 2015. *C. elegans* outside the Petri dish. *eLife.* 4:e05849. doi:[10.7554/eLife.05849](https://doi.org/10.7554/eLife.05849).
- Gao C-H. 2021. ggVennDiagram: A 'ggplot2' Implement of Venn Diagram. Accessed 1 February 2025. <https://CRAN.R-project.org/package=ggVennDiagram>.
- Gems D, Riddle DL. 2000. Defining wild-type life span in *Caenorhabditis elegans*. *J Gerontol A Biol Sci Med Sci.* 55(5):B215–B219. doi:[10.1093/gerona/55.5.b215](https://doi.org/10.1093/gerona/55.5.b215).
- Glaser-Schmitt A, Lemoine M, Kaltenpoth M, Parsch J. 2024. Pervasive tissue-, genetic background-, and allele-specific gene expression effects in *Drosophila melanogaster*. *PLoS Genet.* 20(8):e1011257. doi:[10.1371/journal.pgen.1011257](https://doi.org/10.1371/journal.pgen.1011257).
- Gokhman D, Agolia RM, Kinnebrew M, Gordon W, Sun D, Bajpai VK, Naqvi S, Chen C, Chan A, Chen C, et al. 2021. Human-chimpanzee fused cells reveal cis-regulatory divergence underlying skeletal evolution. *Nat Genet.* 53(4):467–476. doi:[10.1038/s41588-021-00804-3](https://doi.org/10.1038/s41588-021-00804-3).
- Gomes S, Civetta A. 2015. Hybrid male sterility and genome-wide misexpression of male reproductive proteases. *Sci Rep.* 5(1):11976. doi:[10.1038/srep11976](https://doi.org/10.1038/srep11976).
- Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC. 2012. Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. *Genome Res.* 22(12):2376–2384. doi:[10.1101/gr.142281.112](https://doi.org/10.1101/gr.142281.112).
- Grishok A. 2013. Chapter one—Biology and mechanisms of short RNAs in *Caenorhabditis elegans*. In: Friedmann T, Dunlap JC, Goodwin SF, editors. *Advances in genetics*. Academic Press. p. 1–69.
- Hansen KD, Irizarry RA, Wu Z. 2012. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics.* 13(2):204–216. doi:[10.1093/biostatistics/kxr054](https://doi.org/10.1093/biostatistics/kxr054).
- Harvey SC, Viney ME. 2007. Thermal variation reveals natural variation between isolates of *Caenorhabditis elegans*. *J Exp Zool B Mol Dev Evol.* 308(4):409–416. doi:[10.1002/jez.b.21161](https://doi.org/10.1002/jez.b.21161).
- He F. 2011. Total RNA extraction from *C. elegans*. *Bio Protoc.* 1(6):e47. doi:[10.21769/BioProtoc.47](https://doi.org/10.21769/BioProtoc.47).
- He F, Arce AL, Schmitz G, Koornneef M, Novikova P, Beyer A, de Meaux J. 2016. The footprint of polygenic adaptation on stress-responsive cis-regulatory divergence in the *Arabidopsis* genus. *Mol Biol Evol.* 33(8):2088–2101. doi:[10.1093/molbev/msw096](https://doi.org/10.1093/molbev/msw096).

- He F, Zhang X, Hu J, Turck F, Dong X, Goebel U, Borevitz J, de Meaux J. 2012. Genome-wide analysis of cis-regulatory divergence between Species in the Arabidopsis genus. *Mol Biol Evol.* 29(11):3385–3395. doi:[10.1093/molbev/mss146](https://doi.org/10.1093/molbev/mss146).
- Higgins BJ, Hirsh D. 1977. Roller mutants of the nematode *Caenorhabditis elegans*. *Mol Gen Genet.* 150(1):63–72. doi:[10.1007/BF02425326](https://doi.org/10.1007/BF02425326).
- Hill MS, Vande Zande P, Wittkopp PJ. 2020. Molecular and evolutionary processes generating variation in gene expression. *Nat Rev Genet.* 22(4):203–215. doi:[10.1038/s41576-020-00304-w](https://doi.org/10.1038/s41576-020-00304-w).
- Hodgkin J. 2002. Exploring the envelope. Systematic alteration in the sex-determination system of the nematode *Caenorhabditis elegans*. *Genetics.* 162(2):767–780. doi:[10.1093/genetics/162.2.767](https://doi.org/10.1093/genetics/162.2.767).
- Hodgkin J, Doniach T. 1997. Natural variation and copulatory plug formation in *Caenorhabditis elegans*. *Genetics.* 146(1):149–164. doi:[10.1093/genetics/146.1.149](https://doi.org/10.1093/genetics/146.1.149).
- Holdorf AD, Higgins DP, Hart AC, Boag PR, Pazour GJ, Walhout AJM, Walker AK. 2020. WormCat: an online tool for annotation and visualization of *Caenorhabditis elegans* genome-scale data. *Genetics.* 214(2):279–294. doi:[10.1534/genetics.119.302919](https://doi.org/10.1534/genetics.119.302919).
- Inskeep TR, Groen SC. 2025. Network properties constrain natural selection on gene expression in *Caenorhabditis elegans* [Preprint]. *bioRxiv* 2025.02.2019.639144. <https://doi.org/10.1101/2025.02.19.639144>.
- Juneja P, Quinn A, Jiggins FM. 2016. Latitudinal clines in gene expression and cis-regulatory element variation in *Drosophila melanogaster*. *BMC Genom.* 17(1):981. doi:[10.1186/s12864-016-3333-7](https://doi.org/10.1186/s12864-016-3333-7).
- Kamkina P, Snoek LB, Grossmann J, Volkers RJM, Sterken MG, Daube M, Roschitzki B, Fortes C, Schlapbach R, Roth A, et al. 2016. Natural genetic variation differentially affects the proteome and transcriptome in *Caenorhabditis elegans*. *Mol Cell Proteomics.* 15(5):1670–1680. doi:[10.1074/mcp.M115.052548](https://doi.org/10.1074/mcp.M115.052548).
- Kern CC, Townsend S, Salzmann A, Rendell NB, Taylor GW, Comisel RM, Foukas LC, Bahler J, Gems D. 2021. *C. elegans* feed yolk to their young in a form of primitive lactation. *Nat Commun.* 12(1):5801. doi:[10.1038/s41467-021-25821-y](https://doi.org/10.1038/s41467-021-25821-y).
- Khan Z, Ford MJ, Cusanovich DA, Mitrano A, Pritchard JK, Gilad Y. 2013. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science.* 342(6162):1100–1104. doi:[10.1126/science.1242379](https://doi.org/10.1126/science.1242379).
- Kramer JM. 1994. Genetic analysis of extracellular matrix in *C. elegans*. *Annu Rev Genet.* 28:95–116. doi:[10.1146/annurev.ge.28.120194.000523](https://doi.org/10.1146/annurev.ge.28.120194.000523).
- Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 13(10):2229–2235. doi:[10.1101/gr.1589103](https://doi.org/10.1101/gr.1589103).
- Landry CR, Hartl DL, Ranz JM. 2007a. Genome clashes in hybrids: insights from gene expression. *Heredity (Edinb).* 99(5):483–493. doi:[10.1038/sj.hdy.6801045](https://doi.org/10.1038/sj.hdy.6801045).
- Landry CR, Lemos B, Rifkin SA, Dickinson WJ, Hartl DL. 2007b. Genetic properties influencing the evolvability of gene expression. *Science.* 317(5834):118–121. doi:[10.1126/science.1140247](https://doi.org/10.1126/science.1140247).
- Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL. 2005. Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of *Drosophila*. *Genetics.* 171(4):1813–1822. doi:[10.1534/genetics.105.047449](https://doi.org/10.1534/genetics.105.047449).
- Larsson J, Gustafsson P. 2018. A case study in fitting area-proportional Euler diagrams with ellipses using eulerr. *Proceedings of International Workshop on Set Visualization and Reasoning*; Edinburgh, UK. p. 84–91. Accessed 11 June 2025. <https://ceur-ws.org/Vol-2116/paper7.pdf>.
- Lee D, Zdravljek S, Stevens L, Wang Y, Tanny RE, Crombie TA, Cook DE, Webster AK, Chirakar R, Baugh LR, et al. 2021. Balancing selection maintains hyper-divergent haplotypes in *Caenorhabditis elegans*. *Nat Ecol Evol.* 5(6):794–807. doi:[10.1038/s41559-021-01435-x](https://doi.org/10.1038/s41559-021-01435-x).
- Lemmon ZH, Bukowski R, Sun Q, Doebley JF. 2014. The role of cis regulatory evolution in maize domestication. *PLoS Genet.* 10(11):e1004745. doi:[10.1371/journal.pgen.1004745](https://doi.org/10.1371/journal.pgen.1004745).
- Lemos B, Araripe LO, Fontanillas P, Hartl DL. 2008. Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proc Natl Acad Sci U S A.* 105(38):14471–14476. doi:[10.1073/pnas.0805160105](https://doi.org/10.1073/pnas.0805160105).
- Liao BY, Zhang J. 2006. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution. *Mol Biol Evol.* 23(6):1119–1128. doi:[10.1093/molbev/msj119](https://doi.org/10.1093/molbev/msj119).
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biol.* 15(12):550. doi:[10.1186/s13059-014-0550-8](https://doi.org/10.1186/s13059-014-0550-8).
- Ma F, Lau CY, Zheng C. 2021. Large genetic diversity and strong positive selection in F-box and GPCR genes among the wild isolates of *Caenorhabditis elegans*. *Genome Biol Evol.* 13(5):evab048. doi:[10.1093/gbe/evab048](https://doi.org/10.1093/gbe/evab048).
- Mack KL, Campbell P, Nachman MW. 2016. Gene regulation and speciation in house mice. *Genome Res.* 26(4):451–461. doi:[10.1101/gr.195743.115](https://doi.org/10.1101/gr.195743.115).
- Mack KL, Nachman MW. 2017. Gene regulation and speciation. *Trends Genet.* 33(1):68–80. doi:[10.1016/j.tig.2016.11.003](https://doi.org/10.1016/j.tig.2016.11.003).
- Marek A, Tomala K. 2018. The contribution of purifying selection, linkage, and mutation bias to the negative correlation between gene expression and polymorphism density in yeast populations. *Genome Biol Evol.* 10(11):2986–2996. doi:[10.1093/gbe/evy225](https://doi.org/10.1093/gbe/evy225).
- Margres MJ, Wray KP, Seavy M, McGivern JJ, Herrera ND, Rokyta DR. 2016. Expression differentiation is constrained to low-expression proteins over ecological timescales. *Genetics.* 202(1):273–283. doi:[10.1534/genetics.115.180547](https://doi.org/10.1534/genetics.115.180547).
- McMahon L, Muriel JM, Roberts B, Quinn M, Johnstone IL. 2003. Two sets of interacting collagens form functionally distinct substructures within a *Caenorhabditis elegans* extracellular matrix. *Mol Biol Cell.* 14(4):1366–1378. doi:[10.1091/mbc.e02-08-0479](https://doi.org/10.1091/mbc.e02-08-0479).
- McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. 2010. Regulatory divergence in *Drosophila* revealed by mRNA-Seq. *Genome Res.* 20(6):816–825. doi:[10.1101/gr.102491.109](https://doi.org/10.1101/gr.102491.109).
- McQueen E, Yang B, Wittkopp PJ. 2025. Epistatic impacts of cis- and trans-regulatory mutations on the distribution of mutational effects for gene expression in *Saccharomyces cerevisiae* [Preprint]. *bioRxiv* 2025.03.26.645465. <https://doi.org/10.1101/2025.03.26.645465>.
- Metzger BPH, Wittkopp PJ. 2019. Compensatory trans-regulatory alleles minimizing variation in TDH3 expression are common within *Saccharomyces cerevisiae*. *Evol Lett.* 3(5):448–461. doi:[10.1002/evl3.137](https://doi.org/10.1002/evl3.137).
- Metzger BPH, Wittkopp PJ, Coolon JD. 2017. Evolutionary dynamics of regulatory changes underlying gene expression divergence among *Saccharomyces* species. *Genome Biol Evol.* 9(4):843–854. doi:[10.1093/gbe/evx035](https://doi.org/10.1093/gbe/evx035).
- Moya ND, Yan SM, McCoy RC, Andersen EC. 2025. The long and short of hyperdivergent regions. *Trends Genet.* 41(4):303–314. doi:[10.1016/j.tig.2024.11.005](https://doi.org/10.1016/j.tig.2024.11.005).
- Muzzey D, Sherlock G, Weissman JS. 2014. Extensive and coordinated control of allele-specific expression by both transcription and translation in *Candida albicans*. *Genome Res.* 24(6):963–973. doi:[10.1101/gr.166322.113](https://doi.org/10.1101/gr.166322.113).



- Neuwirth E. 2022. RColorBrewer: ColorBrewer Palettes. Accessed 1 February 2025. <https://CRAN.R-project.org/package=RColorBrewer>.
- Noble LM, Chelo I, Guzella T, Afonso B, Riccardi DD, Ammerman P, Dayarian A, Carvalho S, Crist A, Pino-Querido A, et al. 2017. Polygenicity and epistasis underlie fitness-proximal traits in the *Caenorhabditis elegans* multiparental experimental evolution (CeMEE) panel. *Genetics*. 207(4):1663–1685. doi:10.1534/genetics.117.300406.
- Noble LM, Miah A, Kaur T, Rockman MV. 2020. The ancestral *Caenorhabditis elegans* cuticle suppresses rol-1. *G3 (Bethesda)*. 10(7):2385–2395. doi:10.1534/g3.120.401336.
- Orr HA. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. *Genetics*. 139(4):1805–1813. doi:10.1093/genetics/139.4.1805.
- Ozturk-Colak A, Marygold SJ, Antonazzo G, Attrill H, Goutte-Gattat D, Jenkins VK, Matthews BB, Millburn G, Dos Santos G, Tabone CJ, et al. 2024. FlyBase: updates to the *Drosophila* genes and genomes database. *Genetics*. 227(1):iyad211. doi:10.1093/genetics/iyad211.
- Pal C, Papp B, Hurst LD. 2001. Highly expressed genes in yeast evolve slowly. *Genetics*. 158(2):927–931. doi:10.1093/genetics/158.2.927.
- Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 14(4):417–419. doi:10.1038/nmeth.4197.
- Pedersen TL. 2022. ggforce: Accelerating 'ggplot2'. Accessed 1 February 2025. <https://CRAN.R-project.org/package=ggforce>.
- Perez MF, Francesconi M, Hidalgo-Carcedo C, Lehner B. 2017. Maternal age generates phenotypic variation in *Caenorhabditis elegans*. *Nature*. 552(7683):106–109. doi:10.1038/nature25012.
- Perez MF, Lehner B. 2019. Vitellogenins—yolk gene function and regulation in *Caenorhabditis elegans*. *Front Physiol*. 10:1067. doi:10.3389/fphys.2019.01067.
- Pertea G, Pertea M. 2020. GFF utilities: gffRead and GffCompare. *F1000Res*. 9:ISCB Comm J-304. doi:10.12688/f1000research.23297.2.
- Pfeifer B, Wittelsburger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. *Mol Biol Evol*. 31(7):1929–1936. doi:10.1093/molbev/msu136.
- Pouillet N, Vielle A, Gimond C, Ferrari C, Braendle C. 2015. Evolutionarily divergent thermal sensitivity of germline development and fertility in hermaphroditic *Caenorhabditis* nematodes. *Evol Dev*. 17(6):380–397. doi:10.1111/ede.12170.
- Price PD, Palmer Drogue DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE. 2022a. Detecting signatures of selection on gene expression. *Nat Ecol Evol*. 6(7):1035–1045. doi:10.1038/s41559-022-01761-8.
- Price PD, Palmer Drogue DH, Taylor JA, Kim DW, Place ES, Rogers TF, Mank JE, Cooney CR, Wright AE. 2022b. Reply to: existing methods are effective at measuring natural selection on gene expression. *Nat Ecol Evol*. 6(12):1838–1839. doi:10.1038/s41559-022-01916-7.
- Raghupathy N, Choi K, Vincent MJ, Beane GL, Sheppard KS, Munger SC, Korstanje R, de Villena F P-M, Churchill GA. 2018. Hierarchical analysis of RNA-Seq reads improves the accuracy of allele-specific expression. *Bioinformatics*. 34(13):2177–2184. doi:10.1093/bioinformatics/bty078.
- R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Accessed 1 December 2023. <https://www.R-project.org/>.
- Ren K, Russell K. 2021. formattable: Create 'Formattable' Data Structures. Accessed 1 February 2025. <https://CRAN.R-project.org/package=formattable>.
- Renaut S, Nolte AW, Bernatchez L. 2009. Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. Salmonidae). *Mol Biol Evol*. 26(4):925–936. doi:10.1093/molbev/msp017.
- Rhone B, Mariac C, Couderc M, Berthouly-Salazar C, Ousseini IS, Vigouroux Y. 2017. No excess of cis-regulatory variation associated with intraspecific selection in wild pearl millet (*Cenchrus americanus*). *Genome Biol Evol*. 9(2):388–397. doi:10.1093/gbe/evv004.
- Rockman MV, Kruglyak L. 2009. Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet*. 5(3):e1000419. doi:10.1371/journal.pgen.1000419.
- Rockman MV, Skrovanek SS, Kruglyak L. 2010. Selection at linked sites shapes heritable phenotypic variation in *C. elegans*. *Science*. 330(6002):372–376. doi:10.1126/science.1194208.
- Romero IG, Ruvinsky I, Gilad Y. 2012. Comparative studies of gene expression and the evolution of gene regulation. *Nat Rev Genet*. 13(7):505–516. doi:10.1038/nrg3229.
- Sanchez-Ramirez S, Weiss JG, Thomas CG, Cutter AD. 2021. Widespread misregulation of inter-species hybrid transcriptomes due to sex-specific and sex-chromosome regulatory evolution. *PLoS Genet*. 17(3):e1009409. doi:10.1371/journal.pgen.1009409.
- Schedl T, Kimble J. 1988. fog-2, a germ-line-specific sex determination gene required for hermaphrodite spermatogenesis in *Caenorhabditis elegans*. *Genetics*. 119(1):43–61. doi:10.1093/genetics/119.1.43.
- Schrimpf SP, Weiss M, Reiter L, Ahrens CH, Jovanovic M, Malmstrom J, Brunner E, Mohanty S, Lercher MJ, Hunziker PE, et al. 2009. Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol*. 7(3):e48. doi:10.1371/journal.pbio.1000048.
- Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One*. 11(10):e0163962. doi:10.1371/journal.pone.0163962.
- Shih DJH. 2021. argparser: Command-Line Argument Parser. Accessed 1 February 2025. <https://CRAN.R-project.org/package=argparser>.
- Signor SA, Nuzhdin SV. 2018. The evolution of gene expression in cis and trans. *Trends Genet*. 34(7):532–544. doi:10.1016/j.tig.2018.03.007.
- Signor SA, Nuzhdin SV. 2019. Compensatory evolution of gene expression. *Trends Genet*. 35(12):890–891. doi:10.1016/j.tig.2019.09.008.
- Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-Seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 4:1521. doi:10.12688/f1000research.7563.2.
- Starr AL, Gokhman D, Fraser HB. 2023. Accounting for cis-regulatory constraint prioritizes genes likely to affect species-specific traits. *Genome Biol*. 24(1):11. doi:10.1186/s13059-023-02846-8.
- Stastna JJ, Snoek LB, Kammenga JE, Harvey SC. 2015. Genotype-dependent lifespan effects in peptone deprived *Caenorhabditis elegans*. *Sci Rep*. 5:16259. doi:10.1038/srep16259.
- Steige KA, Laenen B, Reimegard J, Scofield DG, Slotte T. 2017. Genomic analysis reveals major determinants of cis-regulatory variation in *Capsella grandiflora*. *Proc Natl Acad Sci U S A*. 114(5):1087–1092. doi:10.1073/pnas.1612561114.
- Steige KA, Reimegard J, Koenig D, Scofield DG, Slotte T. 2015. Cis-regulatory changes associated with a recent mating system shift and floral adaptation in *Capsella*. *Mol Biol Evol*. 32(10):2501–2514. doi:10.1093/molbev/msv169.
- Stephens M. 2016. False discovery rates: a new deal. *Biostatistics*. 18(2):275–294. doi:10.1093/biostatistics/kxw041.
- Sternberg PW, Van Auken K, Wang Q, Wright A, Yook K, Zarowiecki M, Arnaboldi V, Becerra A, Brown S, Cain S, et al. 2024. WormBase 2024: status and transitioning to alliance infrastructure. *Genetics*. 227(1):iyae050. doi:10.1093/genetics/iyae050.



- Stiernagle T. 2006. Maintenance of *C. elegans*. WormBook. 1–11. doi:[10.1895/wormbook.1.101.1](https://doi.org/10.1895/wormbook.1.101.1).
- Thomas CG, Wang W, Jovelín R, Ghosh R, Lomasko T, Trinh Q, Kruglyak L, Stein LD, Cutter AD. 2015. Full-genome evolutionary histories of selfing, splitting, and selection in *Caenorhabditis*. *Genome Res.* 25(5):667–678. doi:[10.1101/gr.187237.114](https://doi.org/10.1101/gr.187237.114).
- van de Geijn B, McVicker G, Gilad Y, Pritchard JK. 2015. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods.* 12(11):1061–1063. doi:[10.1038/nmeth.3582](https://doi.org/10.1038/nmeth.3582).
- Verta JP, Jones FC. 2019. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *Elife.* 8:e43785. doi:[10.7554/eLife.43785](https://doi.org/10.7554/eLife.43785).
- Verta JP, Landry CR, MacKay J. 2016. Dissection of expression-quantitative trait locus and allele specificity using a haploid/diploid plant system—insights into compensatory evolution of transcriptional regulation within populations. *New Phytol.* 211(1):159–171. doi:[10.1111/nph.13888](https://doi.org/10.1111/nph.13888).
- Vinuela A, Snoek LB, Riksen JA, Kammenga JE. 2010. Genome-wide gene expression regulation as a function of genotype and age in *C. elegans*. *Genome Res.* 20(7):929–937. doi:[10.1101/gr.102160.109](https://doi.org/10.1101/gr.102160.109).
- Viswanath A, Cutter AD. 2023. Regulatory divergence as a mechanism for X-autosome incompatibilities in *Caenorhabditis* Nematodes. *Genome Biol Evol.* 15(4):evad055. doi:[10.1093/gbe/evad055](https://doi.org/10.1093/gbe/evad055).
- Wang B, Starr AL, Fraser HB. 2024. Cell-type-specific cis-regulatory divergence in gene expression and chromatin accessibility revealed by human-chimpanzee hybrid cells. *Elife.* 12:e89594. doi:[10.7554/eLife.89594](https://doi.org/10.7554/eLife.89594).
- Wang Z, Sun X, Zhao Y, Guo X, Jiang H, Li H, Gu Z. 2015. Evolution of gene regulation during transcription and translation. *Genome Biol Evol.* 7(4):1155–1167. doi:[10.1093/gbe/evv059](https://doi.org/10.1093/gbe/evv059).
- Webster AK, Willis JH, Johnson E, Sarkies P, Phillips PC. 2025. Gene expression variation across genetically identical individuals predicts reproductive traits. *eLife.* 14:RP106525. <https://doi.org/10.7554/eLife.106525.1>.
- Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag. Accessed 1 February 2025. <https://ggplot2.tidyverse.org>.
- Wilke CO. 2020. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. Accessed 1 February 2025. <https://CRAN.R-project.org/package=cowplot>.
- Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. *Nature.* 430(6995):85–88. doi:[10.1038/nature02698](https://doi.org/10.1038/nature02698).
- Wolf S, Melo D, Garske KM, Pallares LF, Lea AJ, Ayroles JF. 2023. Characterizing the landscape of gene expression variance in humans. *PLoS Genet.* 19(7):e1010833. doi:[10.1371/journal.pgen.1010833](https://doi.org/10.1371/journal.pgen.1010833).
- Xie D, Ye P, Ma Y, Li Y, Liu X, Sarkies P, Zhao Z. 2022. Genetic exchange with an outcrossing sister species causes severe genome-wide dysregulation in a selfing *Caenorhabditis* nematode. *Genome Res.* 32(11-12):2015–2027. doi:[10.1101/gr.277205.122](https://doi.org/10.1101/gr.277205.122).
- Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW. 2002. Allelic variation in human gene expression. *Science.* 297(5584):1143. doi:[10.1126/science.1072545](https://doi.org/10.1126/science.1072545).
- Youngman EM, Claycomb JM. 2014. From early lessons to new frontiers: the worm as a treasure trove of small RNA biology. *Front Genet.* 5:416. doi:[10.3389/fgene.2014.00416](https://doi.org/10.3389/fgene.2014.00416).
- Zhang G, Mostad JD, Andersen EC. 2021. Natural variation in fecundity is correlated with species-wide levels of divergence in *Caenorhabditis elegans*. *G3 (Bethesda).* 11(8):jkab168. doi:[10.1093/g3journal/jkab168](https://doi.org/10.1093/g3journal/jkab168).
- Zhang G, Roberto NM, Lee D, Hahnel SR, Andersen EC. 2022. The impact of species-wide gene expression variation on *Caenorhabditis elegans* complex traits. *Nat Commun.* 13(1):3462. doi:[10.1038/s41467-022-31208-4](https://doi.org/10.1038/s41467-022-31208-4).
- Zhang J, Yang JR. 2015. Determinants of the rate of protein sequence evolution. *Nat Rev Genet.* 16(7):409–420. doi:[10.1038/nrg3950](https://doi.org/10.1038/nrg3950).
- Zhang X, Borevitz JO. 2009. Global analysis of allele-specific expression in *Arabidopsis thaliana*. *Genetics.* 182(4):943–954. doi:[10.1534/genetics.109.103499](https://doi.org/10.1534/genetics.109.103499).
- Zhang X, Emerson JJ. 2019. Inferring compensatory evolution of cis- and trans-regulatory variation. *Trends Genet.* 35(1):1–3. doi:[10.1016/j.tig.2018.11.003](https://doi.org/10.1016/j.tig.2018.11.003).

Editor: P. Wittkopp