Genetic and environmental perturbations drive modular transcriptome responses

Stephanie Mark[1], Joerg Weiss[1], Eesha Sharma[2], Ting Liu[1], Wei Wang[1], Julie M. Claycomb[2], Asher D. Cutter[1#]

[1]Department of Ecology & Evolutionary Biology, University of Toronto, Toronto, ON Canada
[2]Department of Molecular Genetics, University of Toronto, Toronto, ON Canada
[#]Corresponding author: asher.cutter@utoronto.ca

1

Abstract

Background: Transcriptome profiling for distinct environmental and genetic perturbations permits genomic characterization of gene expression modularity, and the robustness of that modularity in the face of extrinsic and genetically intrinsic changes. We quantified the transcriptome responses for distinct temperature-adapted genotypes of the nematode *Caenorhabditis briggsae* when exposed to chronic temperature stress.

Results: We found that 56% of the 8795 differentially-expressed genes show genotype-specific changes in expression in response to temperature (genotype-by-environment interactions, GxE). Most genotype-specific responses occur under heat stress, indicating that cold versus heat stress responses involve distinct genomic architectures. The 22 co-expression modules that we identified differ in their enrichment of genes with genetic versus environmental versus interaction effects, as well as their genomic spatial distributions, functional attributes, and rates of molecular evolution at the sequence level. Modules enriched for genes with simple effects of either genotype or temperature alone on profiles of differential expression tend to evolve especially rapidly.

Conclusions: These results illustrate how the form of transcriptome modularity, in terms of response to environmental and genetic perturbation, ramifies to distinctive profiles of evolutionary change. Chromosome-scale patterning of nucleotide differences predominates as the source of genetic differences among expression profiles rather than *cis*-acting regulatory alleles, and natural selection regimes are largely decoupled between coding sequences and non-coding flanking sequences that contain regulatory elements.

Keywords: Caenorhabditis, transcriptome, regulatory evolution, robustness, temperature, genotype-by-environment interactions, molecular evolution

Background

Evolutionary adaptation to varying environmental conditions starts with genetic variability, with alternate alleles often affecting gene regulation and expression. Consequently, understanding the plasticity, robustness, and modularity of transcriptome expression to genetic and environmental perturbation is crucial to deciphering how organisms adapt in nature [1]. Gene expression represents the most basic level at which phenotypic plasticity to a perturbation can manifest, and therefore underpins the degree of robustness of higher level phenotypes in response to the same perturbation [2, 3]. Because the transcriptome changes in response both to extrinsic factors (e.g. environmental inputs) and to factors that are intrinsic to the organism itself (e.g. genetic perturbation) [4, 5], we must consider both extrinsic and intrinsic contributions in the dynamism of genetic network composition and its genomic architecture. How much of the genome is expressed differentially in a plastic manner sensitive to environmental conditions versus a genetically deterministic manner independently of environmental conditions versus a non-additive combination of both? How modular are distinct gene expression responses and what characteristics of the genome predict their composition? What is the relative importance of transcriptional versus post-transcriptional regulation in expression plasticity, and of *cis*- versus *trans*-regulatory causes of expression differences among divergent genetic backgrounds [6, 7]? These questions frame some of the key outstanding issues in connecting transcriptome activity to environmental heterogeneity and the molecular evolution of genomes.

Temperature conditions represent a pervasive extrinsic, environmental perturbation that influences gene expression and can help reveal the relative roles of plasticity versus robustness of transcriptome profiles [8, 9]. If expression plasticity is adaptive, then we expect organisms to modulate their transcriptomes with chronic exposure to heat or cold stress in a coordinated way to maintain fitness. However, homeostasis may break down at environmental extremes and lead to non-adaptive changes in gene expression that simply reflect a 'broken' biological system. Pathways associated with the heat shock response are implicated in physiological buffering to acute heat stress [10], but chronic sublethal heat stress may not activate this same stress response. By characterizing profiles of transcriptome change across ontogenetic and physiological timescales of temperature acclimatization, we can test for the similarity and uniqueness of plastic responses from physiological timescales to evolutionary timescales of genetic divergence in the control over transcriptome change.

Allelic differences provide another kind of perturbation, a genetic perturbation, that can expose the sensitivity of genetic networks to expression changes [11]. Expression modulated by *cis*-regulatory alleles may minimize adverse pleiotropic effects and, consequently, modest effects of *cis*-regulatory SNPs might only be pronounced when they accrue over long periods of time to give rise to the kinds of expression differences that accumulate between species [12-14]. By contrast, changes to *trans*-acting regulators like transcription factors may lead to many downstream pleiotropic consequences, so large *trans*-acting effects might make up a substantial fraction of the genetic variability for gene expression differences among individuals within a species and yet rarely contribute to expression differences between species [12, 15-18]; because most changes that affect fitness are deleterious and eventually get eliminated by

natural selection [19]. The intermediate timescale of adaptive divergence between populations of the same species thus has the potential to expose whether distinct regulatory architecture must be invoked to describe transcriptome changes across the extremes of timescales from polymorphism within a population to divergence between species.

In this context, the nematode *C. elegans* has been subject to extensive transcriptome analysis in response to heat shock and knock-out mutation starting with microarrays [20], with more recent studies using recombinant inbred lines of wild strains to map polymorphic loci that contribute genotype-dependent responses to temperature [21, 22]. For example, Li et al. [22] found that among 496 detectable expression quantitative trait loci (eQTL), *trans*-eQTL were nearly 8-times as likely as *cis*-eQTL to show genotype-by-temperature responses. Moreover, eQTL are found disproportionately on SNP-dense chromosome arms in *C. elegans* [23]. Grishkevich et al. [21] reported that constitutively-expressed genes in *C. elegans* tend to have short intergenic regions, consistent with simple regulatory controls, and that genes with genotype-dependent expression or genotype-by-environment interactions have longer intergenic regions, consistent with complex regulation and a larger mutational target. How general these properties are across species remains unknown.

Here we quantified transcriptome expression for *C. briggsae* nematodes from populations with distinct genetic backgrounds thought to have adapted to temperature differences between their origins in Tropical versus Temperate latitudes [24-26]. By rearing animals at hot and cold sublethal temperatures near their fertile limits, as well as under benign thermal conditions, we characterize genotypic and environmentally-induced differential gene expression across the genome. We then describe transcriptome complexity in terms of co-expression modularity to reflect transcriptome plasticity and robustness to environmental and genetic perturbations, demonstrating distinctive genomic spatial distributions, functional attributes, and rates of molecular evolution at the sequence level.

Results

Widespread genotype- and temperature-dependent differential gene expression

Over half (54%, n=8795) of all 16,199 genes with detectable expression after quality filtering differed significantly across genotypes, temperatures, or both (of 21,827 annotated coding genes). The majority of differentially expressed genes had a significant genotype-specific response to temperature, with these genotype-by-temperature interactions ("GxT genes") comprising 56% of the 8795 differentially expressed genes (n=4919; 30.4% of all genes with detectable expression; Supplementary File S1) (Figure 1A). Another 1119 genes (13%) showed significant differential expression from the independent, additive effects of both genotype and temperature ("G & T genes"). The remaining 2757 genes exhibited a response to either just temperature or just genotype, with 2.5-fold more genes showing differential expression due to

4

temperature alone (23% "T genes", n=1987) than to genotype alone (8.8% "G genes", n=770). As a result, 64% more genes overall exhibited a simple "plastic" response to temperature than a "deterministic" response to genotype (1987 + 1119 = 3106 vs 770 + 1119 = 1889). In addition, the 4919 genes with a more complex influence of both genotype and temperature (GxT genes) further accentuates the important roles of both environmental plasticity and genetic determinism in transcriptome profiles (Figure 1A).

Genes with expression influenced by chronic cold stress (14°C) responded differently than genes affected by chronic heat stress (30°C) in terms of the number of genes involved, whether genes increased or decreased expression, and the magnitude of expression change. In particular, expression of 74% of genes with simple effects of temperature (T and G & T genes) differed significantly from benign conditions at 20°C in response to rearing under cold stress (2308 of 3106 genes) whereas it was heat stress that altered expression of the plurality of GxT genes (2393 of 4919 genes, 49%) (Figure 1B). However, among all the genes that responded to temperature in some way, more genes reduced their expression at cool temperatures and elevated their expression at warm temperatures, compared to benign conditions (Figure 1C). This contrasting bias in the direction of difference in expression level between cool and warm temperatures held true both for genes with a simple or complex influence of temperature (Figure 1C; 1.05-fold reduction for T plus G&T and 1.3-fold reduction for GxT at 14°C, 6.8-fold elevated for T plus G&T and 1.2-fold elevated for GxT at 30°C). By contrast, genes with simple expression dynamics showed similar magnitudes of expression changes for both chronic cold and heat stress (T and G&T genes; Figure 1D; 6.1 to 6.5-fold increase for heat and cold; 3.0 to 3.7-fold decrease for heat and cold). Those GxT genes that increased expression under chronic heat stress, however, showed much larger magnitudes of change than for chronic cold stress (hot 8.57-fold vs. cold 3.73-fold increase). Reciprocally, GxT genes that decreased expression under chronic cold stress had a larger magnitude change than with chronic heat stress (hot 6.50-fold vs. cold 9.19-fold decrease). These observations support the idea that distinct genetic networks mediate response to cold versus heat stress, rather than control by a single shared temperature stress response.

Co-expression modules differ in functionality

We defined 22 co-expression clusters in the *C. briggsae* transcriptome with WGCNA to capture modules of broad patterns of differential gene expression in response to distinct temperatures and genotypes (Figure 2, Figure 3, Supplementary File S1). Six modules each containing >1000 genes comprise 75% of all genes (M1-M6), whereas just two co-expression modules have <100 genes each (Modules M21 and M22, plus the 37-gene 'remainder' pseudo-module M0).

Each co-expression module displays a stereotypical expression profile for genes contained within it, represented by the "module eigengene" that is defined by the first principal component in expression space (Figure 3). These eigengene profiles illustrate how a given module reflects a dominant trend of genotype-dependence (e.g. M10), temperature-dependence (e.g. M12), additive effects of genotype and temperature (G&T, e.g. M4), or genotype-specific sensitivity to

temperature (GxT, e.g. M22) (Figure 3). When we quantified the incidence of genes with individually-significant differential expression profiles within modules, we found modules to range from a low of just 6% (M13) to a high of 84% (M15) with an average of 46% of genes in a module having individually-significant differential expression (Figure 3, Supplementary Figure S3). Moreover, we observed genes with temperature- and genotype-specific differential expression to be concentrated within particular subsets of modules (Figure 3) and that the genes within distinct co-expression modules also differed in sequence characteristics and in their enrichment with sex-related differential gene expression, as described below.

Genotype-dependent expression profiles predominate in just two modules (M7 and M10), which together include 44% (n=342) of all 770 genes with G-only differential expression. Their eigengene expression profiles show limited dynamics across temperatures, with expression for the Temperate HK104 genotype consistently higher than Tropical AF16 in M7 and consistently lower in M10 (Figure 3). Gene ontology (GO) term enrichment in M7 indicates disproportionate representation of genes with nervous system function, including 11 GABA and 11 acetylcholine receptor activity genes, including the ortholog of *C. elegans* nicotinic acetylcholine receptor *acr-9*. M10 is enriched for GO terms related to extracellular constituents (Supplementary File S2). Notably, the HK104 and AF16 strains of *C. briggsae* differ in rearing-dependent thermal taxis [26], a suite of behaviors under neural control. Genes in module M10 have several other special features compared to other modules: rapidly-evolving genes (high dN/dS'), high density of SNPs in replacement sites despite lowest SNP density in introns, the highest enrichment in arm regions of autosomes, enrichment on the X-chromosome, and exceptional rarity in operons (Figure 4, Figure 5A). We observed that genes in M10 also have the least consistent expression among replicates, with very few gene members having orthologs with "oogenic" expression according to Ortiz et al. [27] (Figure 4A; Figure 5A). These features imply weaker canalization of expression of genes in M10, reflecting either weaker purifying selection or perhaps recent adaptive divergence in average expression levels that has not yet fine-tuned expression variability.

Two other modules were especially enriched in genes with additive effects of both genotype and temperature (M4, M5; G&T genes), accounting for over half (56%) of all such genes genome-wide (Figure 3). Eigengene profiles for M4 and M5 show high expression at low temperatures, with the Temperate HK104 genotype having consistently higher expression than Tropical AF16 in M4 and vice versa for M5 (Figure 3). M4 is enriched in lipid-related GO terms, as well as for extra-cellular matrix, cell adhesion, and serine peptidase inhibitors; including for example the *C. briggsae* ortholog of *C. elegans ttm-5* and human DEGS2, with sphingolipid desaturase activity. By contrast, GO term enrichment in M5 indicates a prominent role of genes with phosphatase/kinase activity and glycogen metabolism (Supplementary File S2), including the orthologs of *C. elegans gsp-3/4* and *aagr-1*. Previous expression studies have reported male-biased and sperm-related genes to be enriched for genes with phosphatase/kinase GO terms [28, 29], and that at least some glycoproteins play crucial roles in sperm competitiveness [30]. Interestingly, we found that module M5 is extremely enriched (3.3-fold) in orthologs of "spermatogenic" genes from Ortiz et al. [27], a level unlike any other module (Figure 4A). Both sperm and oocyte fertility show temperature sensitivity differently between these genotypes of *C. briggsae* [24, 25]. Genes in M5 also are rare on the X-chromosome and nearly absent from

operons, as expected for sperm-related genes [31-33], with fewer transcription factors (TFs) than most modules (Figure 4B; Figure 5A).

By contrast to these co-expression modules with strong representation of genes with independent allelic genetic effects, two distinct modules each were comprised of >50% 'temperature only genes' (M12, M15), although they accounted for just 12% (n=247) of the 1987 total temperature genes (Figure 3). Modules M6 and M4 also contained a large fraction of temperature-only genes, and as large modules they also contain a large count of such genes (Figure 3). For both M12 and M15, eigengene expression is highest at high rearing temperatures for both the Temperate and Tropical genotypes (Figure 3). Module M12 is highly enriched (3.7-fold) for orthologs with an oogenic gene classification in *C. elegans* [27], whereas M15 is depleted of such genes by having 2.6-fold fewer than expected (Figure 4A). Module M12 GO terms show enrichment for genes associated with chromatin, like the ortholog of *C. elegans cec-7*, but with just 8 such genes of the 245 in M12, it is unclear how distinctive a property this is. Even more enigmatically, M15 has no GO term enrichment, providing little clue as to whether these heat-sensitive genes act in related functional pathways (Supplementary File S2). These modules have genes with the highest average rates of evolution (dN/dS') and that occur only rarely in operons (Figure 5). The genes in modules M12 and M15 also have among the lowest average levels of expression and codon usage bias (Figure 5A).

Eight modules contained a disproportionately large set of GxT genes (M1, M2, M3, M9, M14, M16, M18, M22), indicating a prominent influence of genotype-specific responses to temperature (Figure 3). These eight modules accounted for 71% (n=3477) of all GxT genes genome-wide. The eigengene profiles for M9, M14, M18 and M22 show dramatic crossing 'reaction norms' such that the Temperate HK104 and Tropical AF16 genotypes exhibit opposing expression responses to rearing temperature (Figure 3). By contrast, eigengene profiles for M1, M2, M3 and M16 show how one of the genotypes has a much more exaggerated expression response specifically at 30°C, the highest temperature (Figure 3). GO terms related to chromatin and transcription were enriched in M1, M2 and M9, whereas M14, M18 and M22 showed enrichment for mitochondria-related, ribosome-related, and/or translation-related GO terms; M3 was enriched for genes with nervous system-related terms (Supplementary File S2). No GO terms were enriched in M16, despite its 8 of its 18 gene orthologs assessed for gonad expression showing spermatogenic roles (1.7-fold enrichment) [27] (Figure 4A). By contrast, M9 was 2.7-fold enriched for orthologs with oogenic roles [27] (Figure 4A). Among these modules enriched with GxT genes, we observed that M18 and M22 were especially unusual in having genes that tend to be short, with the strongest codon usage bias and highest average expression, while also having the lowest incidence of replacement-site SNPs and lowest average rates of protein evolution (dN/dS') (Figure 5). Genes in modules M18 and M22 also are enriched for orthologs of "sex neutral" genes from [27], are enriched in operons, and include few TFs (Figure 4A; Figure 5A). M22 also is distinctive among modules in its genes having among the lowest density of SNPs in upstream flanking sequence (Supplementary Figure S4). These diverse observations for M18 and M22 are highly compatible with their especially strong enrichment of ribosome-related GO terms.

The seven remaining co-expression modules consisted primarily of genes that lacked individually significant differential expression, though their eigengene profiles nevertheless suggest important effects of genetic background and temperature on the stereotypical expression profile for genes in those modules (M8, M11, M13, M17, M19, M20, M21). Several of these modules showed GO term enrichment for various metabolic processes (M8, M11, M17, M21) and transcriptional or translational functions (M8, M11, M13, M20). Among these modules, M8, M19 and M20 are extremely enriched for orthologs of "oogenic" genes from Ortiz et al. [27], but include very few operonic genes (Figure 4A; Figure 5A). Oogenesis is especially sensitive to high temperatures in HK104, yielding especially strong reductions in mitotic and meiotic cell counts in the gonad [25]. M20 also has the highest incidence of TFs (29%) among all co-expression modules, and is enriched for genes on autosomal arms and on the X-chromosome (Figure 4, Figure 5A). Genome-wide, TFs are more likely to show no differential expression than other kinds of genes (no DE for 54.5% of TFs vs. 45.2% of other genes; G-test χ^2=29.2, P<0.0001). In contrast to M20, M21 is distinctive in having the highest incidence of genes in operons (60.4%), which are extremely rare on autosomal arms and the X-chromosome (Figure 4, Figure 5A). The 96 genes in M21 have extremely consistent expression across replicates, with most showing no individually-significant differential expression due to either temperature or genotype (Figure 3; Figure 5A).

Muted differential expression role among heat shock proteins

We hypothesized that if heat shock proteins (hsp) modulate transcriptomic responses to chronic temperature stress then we would detect disproportionate differential expression for hsp genes. Of the 24 hsp genes in our expression dataset, only 8 showed significant differential expression, which represents a smaller proportion of differentially-expressed genes (33%) than in the genome overall (54%) (G-test χ^2 = 4.267, df = 1, P = 0.039). The limited contribution of hsp differential expression suggests that hsp genes may play a less crucial role in chronic temperature stress, despite their profound importance to maintaining homeostasis in the face of acute heat stress [10]; even with acute heat shock, few genes show consistent upregulation in *C. elegans* [34]. These 8 hsps include 1 G-only gene, 2 T-only genes, and 5 GxT genes, indistinguishable from the representation of these categories throughout the genome (Fisher Exact Test, P = 0.35). We found 8 of the hsp genes to occur in modules with predominantly GxT eigengene profiles (M2, M3), 6 were in predominantly G&T modules (M4, M5), and 1 was in a predominantly G-only module (M7). The eigengene profiles for modules containing most of the hsps had higher expression in the Temperate genotype than the Tropical and none had highest expression at 20°C, suggesting that hsp genes that respond to temperature stress (either cooling or heating) do so by increasing expression.

Genomic position and differential gene expression

We hypothesized that genomic architectural and molecular evolutionary features might lead to local enrichment of genes with genotype-dependent differential-expression. For example, SNP variation is greater in the high recombination arm domains of autosomes in *C. briggsae* [35], and the X-chromosome exhibits a variety of distinctive features compared to autosomes [36, 37]. Therefore, we tested for non-random distributions of differentially-expressed genes along chromosomes and between chromosomes. We found that autosome arm domains contained 22% more genes with genotype-dependent expression than expected by chance, and also were slightly enriched for GxT genes (1.04-fold; Figure 1B). Chromosome arms of *C. elegans* also have been reported to contain a disproportionate representation of genes with genotype-dependent differential expression [21, 23, 38]. By contrast, it was center domains that contained 15% more G&T genes than expected (Figure 1B). Temperature genes and genes with no differential expression were randomly distributed between arm and center domains (Figure 1B). Among the 22 co-expression modules, we observed 9 modules to have significant enrichment in arms and 5 enriched in center domains of autosomes (Figure 4B).

We also found the X-chromosome to be enriched for genes with significant differential expression due to genetic background (G-only genes) as well as for genes with no individually-significant differential expression (Figure 1B). This X-linked bias held true for some co-expression modules with an overabundance of these gene classes (M10, M20; Figure 4B). Autosomes, on the other hand, contained disproportionate representation of G&T and GxT genes, also reflected among some co-expression modules (M5, M18) (Figure 1B, Figure 4B). Overall, the X-chromosome was enriched for 8 modules, autosomes were enriched for 8 modules, and genes from 6 modules were randomly distributed between autosomes and the X-chromosome (Figure 4B). Genes from module M21, in particular, are virtually absent from the X-chromosome (Figure 4B), likely associated with the prevalence of operonic genes in this co-expression module that also tend to be exceptionally rare on the X-chromosome [32, 39].

In *C. elegans*, loci with genotype-dependent expression tend to have longer upstream intergenic regions, interpreted as being consistent with more complex regulation of these genes [21]. We observed a similar pattern in *C. briggsae* (ANOVA $F_{4,15414}$=5.84, P<0.0001, Tukey post-hoc tests on log-transformed upstream intergenic length show G-only > T-only), with median upstream length of 1367bp for G-only genes versus 1074bp for T-only genes. After partitioning the genomic locations of differentially-expressed genes to account for their non-random distributions in the genome, we found that only those G-only genes in autosomal centers have significantly longer upstream intergenic regions compared to T-only genes (arms ANOVA $F_{4,5653}$=0.10, P=0.98; centers $F_{4,6410}$=5.50, P=0.0002, Tukey post-hoc tests on log-transformed upstream intergenic length show G-only > T-only). However, genes in autosomal centers with no differential expression also had longer upstream intergenic lengths than T-only genes and were not significantly different in length to GxT genes or G&T genes. We also find significant variation among co-expression modules in upstream intergenic length (arm ANOVA $F_{22,5635}$=13.59, P<0.0001; center ANOVA $F_{22,6392}$=15.48, P<0.0001), but observe no clear trend between length and the relative composition of genotype- or temperature-dependent genes.

Genomic SNP associations with differential expression of the transcriptome

Genotype-dependent differences in expression could result from allelic differences in the local vicinity of genes (*cis*-acting effects; e.g. variants in promoter or nearby enhancer elements) or in distant regulators (*trans*-acting effects; e.g. variants in the regulation or functional sequence of transcription factors or miRNAs) [40]. The allelic differences contributing to local *cis*-acting regulation are likely to occur in the upstream promoter regions for those genes showing genotype-dependent expression [21], though there are additional important roles of downstream and intronic regulatory elements in gene expression [41]. Therefore, we quantified the incidence of single-nucleotide polymorphisms (SNP) between the AF16 and HK104 genomic backgrounds in 500bp upstream and downstream flanking regions of coding sequences, as promoter regions tend to be in close proximity to coding sequences in *Caenorhabditis* [42].

We first quantified how often genes lacked SNPs altogether (for the 16,167 genes with expression and genomic coverage in both AF16 and HK104), which excludes a role for *cis*-acting SNPs for such loci. Specifically, we found that 26.2% of such genes have zero upstream SNPs (23.0% G-only, 30.9% G+T, 25.6% GxT), suggesting this value as a lower-bound estimate for the incidence of entirely *trans*-acting regulatory differences altering expression of loci with genotype-dependent expression. Moreover, of differentially-expressed genes affected by genotype, 32.6% have zero downstream SNPs, 20.3% have zero intronic SNPs, and 18.6% have zero SNPs in the coding sequence, also consistent with a major role of *trans*-regulatory control being responsible for the genotype-dependence. Genes with differential expression due to non-genetic factors might also show unusual incidence of SNPs, as *cis*-acting regulation would show no detectable allelic effects for such loci. While T-only genes had a nominally higher incidence of genes that lacked SNPs altogether, it was not significantly different from the frequency of genes with zero SNPs observed for other expression classes (G-test χ^2=8.58, P=0.073).

We further predicted that an important role of *cis*-acting SNPs would be most evident by their enrichment in association with G-only genes (as well as G&T genes and GxT genes), whereas SNPs would be underrepresented in genes with no differential expression or T-only profiles. Genome-wide, we did observe significant differences among differential expression categories in the incidence of SNPs in upstream (ANOVA $F_{4,16141}$=7.63, P<0.0001), downstream ($F_{4,16141}$=4.75, P=0.0008), and intronic portions of coding genes ($F_{4,16141}$=7.82, P<0.0001). Overall, G-only genes have significantly higher SNP densities than other expression classes at replacement sites, synonymous sites, introns and flanking sequences and genes with a GxT pattern of differential expression had a greater density of SNPs than T-only genes only in introns. These results are consistent with the report by Grishkevich et al. [21] for *C. elegans* that SNPs are enriched in promoters of genes with genotype-dependent differential expression.

Our findings therefore superficially support the idea of a key role for *cis*-acting SNPs controlling genotype-dependent differential expression. However, we observed that this trend is driven

primarily by the enrichment of G-only genes in chromosome arms (Figure 1B), where SNPs are disproportionately abundant for both functionally-constrained and unconstrained sites [35]. When we account for genomic location, SNP density remains elevated for G-only genes among genes in autosomal centers but not in arms (ANOVA $F_{4,6729}$=3.60, P=0.0062, G-only > other gene classes with Tukey HSD post-hoc test; Figure 5C).

One possible interpretation of the excess of SNPs among genes with genotype-dependent differential expression is that these genes are subject to weaker selective constraint that allows SNPs to accumulate within them. Replacement-site SNPs are rarest in non-differentially expressed genes (in both chromosome arms and centers), which would be consistent with non-DE genes having strongest selective constraint that most effectively eliminates new mutations (Figure 5C). However, replacement-site divergence that reflects a longer timescale of evolution is no different between G-only genes and non-differentially expressed genes (median G-only dN/dS' = 0.0580, no DE dN/dS' = 0.0511; no significant difference from Tukey's post-hoc test on log-transformed values). These contrasting patterns suggest that relaxed selection on G-only genes might be evolutionarily recent or that adaptive divergence between Temperate and Tropical phylogeographic groups of *C. briggsae* might contribute disproportionately to this class of loci.

Among co-expression modules, average coding SNP density correlated strongly and negatively with the average expression level and codon usage bias among genes (module mean $\pi_{nonsynonymous}$ × average expression Spearman ρ = -0.92, P<0.0001; $\pi_{nonsynonymous}$ × ENC ρ = 0.418, P=0.048; Supplementary Figure S4). Module SNP density for non-coding flanking regions, however, showed weaker negative correlations with expression level (module mean π × average expression Spearman ρ, $\rho_{upstream}$ = -0.72, P<0.0001; $\rho_{downstream}$ = -0.77, P<0.0001; $\rho_{intronic}$ = -0.54, P<0.0082; Supplementary Figure S4). In particular, coding SNPs were most enriched in modules M10, M16, M12 and M15 (Figure 5B); genes in these modules also exhibit among the lowest average levels of expression and codon bias. Coding SNPs are most under-represented in modules M22, M18 and M14, which have especially high average expression levels and strong codon usage bias (Figure 5A; Figure 5B). Coding SNP density also correlates with long-term molecular evolutionary divergence between species (dN/dS'; Figure 5B). These observations support the idea that genes and modules with many SNPs are subject to weaker selective constraints.

Finally, we tested for correspondence between coding and non-coding sequence molecular evolution. While SNP density at replacement sites correlated strongly with the incidence of SNPs in flanking or intronic non-coding sequences in a gene-wise analysis, interspecies divergence in coding sequence did not correlate with SNP density in non-coding sequences (Supplementary Figure S5). SNPs and interspecies divergence correlate positively across genes for replacement sites, consistent with similar pressures of purifying selection at both short and long evolutionary timescales (log-transformed $\pi_{nonsynonymous}$ and dN/dS', $F_{1,5741}$ = 646.7, P<0.0001). When we instead analyzed average values for co-expression modules, however, we observe positive correlations of non-coding SNP density with both coding SNPs and interspecies divergence (Supplementary Figure S5), suggesting that the distinct gene contents

11

and genomic locations of genes among modules partly contributes to the coding-noncoding correspondence at the module level. Overall, these observations support the idea that selection pressures are largely decoupled between coding sequences and non-coding flanking sequences that contain regulatory elements.

Discussion

Temperature-dependent plasticity in gene expression represents the predominant mode of differential expression in our analysis of the *C. briggsae* transcriptome, relative to genotype-dependent expression differences. Specifically, 91% (n=8025) of the 8795 genes we found to be differentially expressed showed temperature-dependence, with most of the purely plastic response being to cool relative to benign conditions. In *C. remanei*, plasticity also dominates the transcriptome response to temperature stress, at least in terms of acute heat shock [43]. Despite this profound environmental plasticity, we also found genetic effects for 77% (n=6808) of differentially-expressed genes. Moreover, more than half (56%) of differentially expressed genes show complex non-additive effects of both genotype and temperature (GxT genes), with genotype-dependence most often due to warm relative to benign conditions. If environment-dependent expression responses reflect adaptive plasticity, then our observations suggest canalization of stereotyped cool-rearing expression responses whereas responses to warm conditions might reflect adaptive evolution by the distinct genetic backgrounds of *C. briggsae* from Tropical and Temperate regions.

We identified 22 co-expression modules in the *C. briggsae* transcriptome that define collections of genes with distinctive functional and evolutionary properties, in addition to their unique profiles of expression as a function of genotype and temperature. Genes in modules with the most pronounced dependence on genotype alone (M10) or temperature alone (M12 and M15) exhibited the lowest overall expression levels and the highest average rates of sequence evolution across modules. The amount of expression appears to be a key determinant of rates of coding sequence evolution, with slower molecular evolution of highly expressed genes. Moreover, *C. elegans* genes showing non-interaction differential expression similarly tend to have low expression levels [21]. Whether these patterns primarily reflect adaptive sequence divergence or, instead, weaker purifying selection on low-expression genes is difficult to determine from the present data alone. More common among co-expression modules, gene profiles exhibited a complex genotype × temperature (GxT) interaction, either showing 'crossing reaction norms' such that the genotypes responded to temperature in opposite ways or with one genotype showing a more extreme response than the other at some temperatures. The three modules with slowest average rates of coding sequence evolution also have the highest make-up of GxT genes, which exhibit crossing reaction norms and very high average expression levels (M14, M18, M22). This finding of especially strong purifying selection implicates either adaptive plasticity in the expression control of these modules or unusually low robustness of expression levels to perturbation from both genotypic and environmental sources.

The incidence of coding SNP differences between Tropical AF16 and Temperate HK104 tracks closely the inter-species divergence across modules, implicating consistent differences in selective constraint on coding sequence among modules at both short (SNP) and long (dN/dS') evolutionary timescales. These molecular evolutionary properties of genes in co-expression modules contrasts with the less clear patterns among genes categorized by mode of differential expression (i.e. G-only, T-only etc.), suggesting that the finer-grained grouping of genes according to co-expression profiles yields more cohesive biological units for analysis.

When we cross-correlated gene identity within co-expression modules with functional attributes of their orthologs in *C. elegans*, we found that module M5 was unique in having a large representation of sperm-related genes among its orthologs. It includes overrepresentation of genes with phosophatase/kinase activity and associated with glycogen metabolism, which previous studies show to be especially important in sperm function [28-30, 44]. Both genotype and temperature were important determinants of expression profiles in M5 (Figure 3), implicating the potential for both adaptive divergence and phenotypic plasticity to influence gene responses. Sperm-dependent fertility appears to be especially sensitive to high temperature, with Tropical and Temperate genotypes of *C. briggsae* differing in sensitivity [24, 25]. As expected for sperm genes [31-33], genes from M5 are especially rare on the X-chromosome and virtually absent from operons.

Across all genes, those in operons were much less likely to show significant differential expression than non-operonic genes (43% of operonic vs. 56% of non-operonic genes; Fisher exact text P<0.0001). The trend is even more extreme among the genes in co-expression module M21, for which fully 60% of the genes occurred within operons: less than 17% of its genes had individually significant differential expression. This observation is consistent with operonic genes being disproportionately robust to both environmental and genetic perturbation. *C. elegans* operon genes, most of which are conserved in *C. briggsae* [45], are known to show high expression during growth, as for gonad tissue [32] and following growth-arrested states [46]. At the other end of the operonic spectrum, only 2% of genes in M20 are operonic, and instead 29% of this module of low-expression genes correspond to transcription factors; just 27% of genes in M20 show individually significant differential expression.

Genes with genotype-dependent differential expression located in chromosome centers are enriched for SNPs in upstream non-coding regions, consistent with local *cis*-acting alleles affecting their expression. However, long-term coding sequence divergence correlates poorly with non-coding SNP density across genes, implying that the strength of selection on coding sequence variation may be decoupled from *cis*-regulatory genetic variation [47-51] or that regulatory elements are too sparse within flanking DNA to leave a clear selective signature with our approach. Nevertheless, the abundance of loci with zero upstream SNPs suggests that distant *trans*-regulatory control is a profound source of genetic variation in the differential expression patterns that we quantified, consistent with studies of short evolutionary timescales [12, 14]. At least in yeast, eQTL analysis implicates a stronger role for *trans*- relative to *cis*-regulation of genotype-environment interactions [16].

Genes showing genotype-dependent expression were enriched on *C. briggsae* chromosomal arms, genomic regions that also are rich in SNPs and with high rates of recombination [35, 36]. This pattern is reminiscent of the excess of eQTL and loci with genotype-dependent expression on *C. elegans* chromosome arms [21, 23, 38]. Does this pattern reflect direct selection on the affected loci, as could occur from either adaptive divergence being more prevalent or from purifying selection being weaker for genes on high recombination arms? Or instead might it be a byproduct of linked selection, whereby elimination of SNPs in low recombination centers has simply led to few loci with the potential to show genotype-dependent differential expression?

The higher recombination rate of arm regions means that natural selection favoring a given allele at one locus will be subject to less interference from selection at other loci in the genome [52-54]. Experiments implicate temperature-related adaptive divergence between *C. briggsae* genotypes from Tropical and Temperate latitudes [24]. Consequently, gene-specific adaptation to distinct ecological conditions should operate more efficiently for genes on arms, which might yield the enrichment of genotype-dependent expression on arms as well as the more rapid sequence evolution of genes on arms. However, it is difficult to exclude the role of linked selection, as the high self-fertilization in *C. briggsae* leaves a substantial imprint on genomic patterns of variation for both synonymous and non-synonymous polymorphisms [35, 55]. Moreover, if our observations of genotype-dependent differential expression depend primarily on a small number distant *trans*-acting upstream regulators that influence many target loci (rather than local *cis*-acting allelic variants for many genes), then the bias toward chromosome arms of differentially-expressed genes might simply be a byproduct of non-random distributions of gene functions encoded across the genome.

Regardless of the potential role of chromosome structure in adaptation, we can largely rule out the nearly 9500 genes with T-only effects or no differential expression as encoding key determinants of temperature-dependent adaptive divergence between AF16 and HK104. In experimentally evolved populations of *C. remanei* in response to acute heat stress, Sikkink et al. (2018) report genetic responses that yielded altered baseline expression as well as GxE, reminiscent of some of the differences we observe for the Temperate and Tropical genotypes of *C. briggsae*. Our analysis of *C. briggsae* finds a stronger signal of genotype-dependent differential gene expression than the *C. remanei* study, perhaps reflecting the longer period of divergence between AF16 and HK104 than between the experimental evolution lines for *C. remanei*, in addition to technical differences between the studies [43]. Phylogenetic comparative analysis of differential expression among genotypes and environments could prove fruitful in deciphering whether shared gene networks across species provide common substrate for adaptive divergence and adaptive plasticity in organismal responses to chronic and acute temperature stress.

Conclusions

Genome-wide differential gene expression is sensitive to both extrinsic temperature conditions and to intrinsic genomic background in the nematode *C. briggsae*, with 56% of the 8795 differentially expressed genes in our study exhibiting complex non-additive effects of both factors. Most genotype-specific responses occur under heat stress, indicating that cold versus heat stress responses involve distinct genomic architectures. When we cluster genes together that have similar expression profiles, we find that the resulting 22 co-expression modules define distinctive functional features, genomic distributions and molecular evolutionary patterns of their constituent genes. The modules with the fastest-evolving protein coding sequences correspond to a predominant role of temperature (but not genotype) or genotype (but not temperature) inducing differential expression, and overall low levels of expression across conditions. One co-expression module is exceptionally enriched for gene orthologs associated with sperm function, and which exhibit strong sensitivity to both temperature and genotype. Our analysis of SNP differences in putative regulatory regions primarily reflects chromosome-scale patterning of nucleotide differences rather than evidence for a predominant role of *cis*-regulatory alleles controlling genotype-dependent expression differentiation, and that selection regimes are largely decoupled between coding sequences and non-coding flanking sequences that contain regulatory elements.

Methods

Experimental design and sequencing

To quantify the genome-wide effects of rearing temperature and genetic background on gene expression, we isolated and sequenced mRNA transcriptomes from *C. briggsae* young adult hermaphrodites of two isogenic strains (AF16 = "Tropical" strain, HK104 = "Temperate" strain) that were reared at 14°C, 20°C, and 30°C from egg to adult. Previous generations of both genotypes had been raised at 20°C prior to establishment of eggs for rearing at the treatment temperatures following stage synchronization via standard *Caenorhabditis* sodium hypochlorite ("bleaching") protocol [56]. After reaching young adulthood, total RNA was isolated with Trizol extraction and isopropanol precipitation [57] from mass isogenic cultures of each strain at each rearing temperature with three biological replicates (2 genotypes x 3 rearing temperatures x 3 replications = 18 samples). The mRNA was then separated from small RNA fractions of less than approximately 200 nucleotides using the mirVana kit from Ambion as per the manufacturer's instructions, and prepared for single-end 100bp sequencing of TruSeq libraries via Illumina HiSeq 2000 (Genome Quebec, Canada) with each of the 18 barcoded samples sequenced across 2 lanes to control for lane effects [58].

We obtained an average of 51.4 million reads per sample (range: 34.5 - 73.4 million) for 925.3 million total reads. Sequences are available in NCBI in project accession PRJNA509247. Over 96% of reads were retained after cleaning and trimming of raw FASTQ files with Trimmomatic 0.36 (894.4 million reads retained), using a seed-mismatch rate of 2, a simple clip threshold of

10, discarding reads <60bp long, and trimming bases from 5' and 3' ends if they had phred33 scores lower than 3 [59].

Read mapping and expression counts

For each sample, we mapped reads to the *C. briggsae* genome (WS253) using STAR [60], setting the maximum intron size to 5000 bp which includes 99.6% of all intron annotations in the *C. briggsae* reference genome. We applied a liberal mismatch rate of 10 to accommodate potential mapping efficacy differences between the AF16 and HK104 strains due to their genetic differences; the reference genome is based on the AF16 strain, so this liberal parameter choice minimizes the potential for mapping to bias towards the Tropical genotype that could inflate inference of differential expression due to genotype. Over 90% of the 894.4 million total reads mapped to unique locations, in all samples (except one replicate of HK104 at 30°C with 73.86% of 48.4 million reads mapping uniquely), with an average of 45.9 million reads mapping per sample to unique locations (Supplementary Table S1).

We then counted the number of reads that mapped to each exon annotated in the WS253 reference genome with htseq-count [61] and summed over all exons in a gene to give a raw measure of expression for each gene in each sample. For our analysis, we neglected alternative splicing isoforms, treating them as contributing to expression levels for the same gene, and set the "mode" parameter in htseq-count to "intersection-nonempty" to resolve ambiguity for overlapping genes [61]. Among mapped reads, 82-85% were assigned successfully to a particular gene among the 23,267 genes annotated in the WS253 *C. briggsae* genome in all samples (again excepting one replicate of HK104 at 30°C, with 24.0 million = 58% of reads assigned to genes). Among the reads that were not assigned to genes, most (9% on average) could not be associated with any exon or were counted in multiple locations (8% on average) and less than 0.1% were ambiguous.

Differential expression analysis

We first visualized gene expression counts in a Multi-Dimensional Scaling (MDS) plot [62], which showed strong clustering of most biological replicates within a treatment and differentiation among treatments (Supplementary Figure S1). We then retained only the subset of genes with at least 1 cpm (gene read count per million; using the "cpm" function in edgeR [63]) in 3 or more libraries (i.e. in one biological replicate) to exclude 7068 genes with extremely low expression that could bias downstream analysis. It is possible that the genes filtered out at this step might exhibit higher expression at different developmental stages, males, or alternative environmental conditions than those assessed here. To test for statistical evidence of differential expression, we next transformed the expression counts using limma and voom, which performs well in controlling Type I error and in detecting true positives [64, 65]; preliminary analysis (not shown) found limma to be more conservative than edgeR for our dataset [63]. Upon applying the voom transformation from the limma package to the remaining set of 16,199 genes, a Q-Q plot showed that the data closely approximated a normal distribution (Supplementary Figure S1).

We then tested these 16,199 genes for differential expression using limma by fitting a linear model to the expression profile for each gene as: expression ~ strain + temperature + strain*temperature interaction. We first tested for significance of the interaction term, and then tested for significance of the main effect terms only if the interaction was non-significant. The model intercept was set as expression for the Tropical strain at 20°C and P-values were adjusted for multiple testing using the Benjamini-Hochberg correction with significance inferred for a false discovery rate (FDR) of 0.05 [66]. To distinguish which genes responded to hot versus cold rearing conditions for genes with a significant effect of temperature (either main effect or interaction effect), we performed post-hoc tests on the individual temperature coefficients (FDR = 0.05). We then classified genes into five mutually exclusive categories based on whether they showed significant differential expression due to genotype only ("G only" genes), temperature only ("T only" genes), both genotype and temperature as independent main effects (i.e. additive effects; "G&T" genes), a non-additive interaction between genotype and temperature ("GxT" genes), or no differential expression ("no DE" genes).

Co-expression clustering of gene expression profiles

To capture distinct stereotypical profiles of gene expression differences in response to our temperature and genotype treatments, we performed a co-expression clustering analysis using the Weighted Gene Correlation Network Analysis (WGCNA) package [67]. Because WGCNA works best with normally distributed expression values, we again used the voom-transformed expression values for the 16,199 filtered genes. A preliminary hierarchical clustering analysis of the samples rejected batch effects as a source of heterogeneity among samples, instead identifying both genotype and temperature as likely and biologically interesting sources of variation in the data (Supplementary Figure S1). We determined the best soft-thresholding power parameter for our data to be 30 (R^2 correlation with a scale-free network topology = 0.75) based on fits across a range of values from 1 to 42 (Supplementary Figure S2), which also yielded an acceptable level of mean connectivity (k = 115), which is central to the assumptions of the WGCNA model [68].

Running WGCNA yielded 124 initial clusters of genes with similar patterns of expression, which we consolidated further by merging similar modules, defined as those with a correlation of 0.75 or higher with each other (Supplementary Figure S2). This procedure produced 22 co-expression modules plus one pseudo-module (M0) containing the 37 genes that could not be grouped based on expression pattern. The characteristic expression profile of genes in a module is represented by WGCNA as the first principal component in expression space, termed the "module eigengene" [69], which we plotted for each genotype separately as the module eigengene expression values averaged across the three biological replicates as a function of rearing temperature.

We performed statistical overrepresentation tests of Gene Ontology (GO) terms associated with gene lists of each co-expression module using PANTHER [70], using all four PANTHER lists available for *C. briggsae*: Pathways, GO-slim Molecular Function, GO-slim Biological Process,

and GO-slim Cellular Components. P-values were adjusted for multiple testing with the Bonferroni correction.

Genomic enrichment analysis

C. briggsae chromosomes are defined by distinct recombinational domains (high recombination arms, low recombination centres, and small tip regions with little detectable recombination), which also correlate with the density of coding genes, repetitive elements and single nucleotide polymorphism (SNPs) [35, 36, 71]. We therefore tested whether gene profiles of differential expression or module affiliation were enriched in particular chromosomal regions using Bonferroni-adjusted G-tests, defining arm-center boundaries as in Ross et al. [36]. Analyses of upstream intergenic lengths were log-transformed prior to analysis with ANOVA; genes with overlapping positions in the genome annotation were excluded from this analysis. We used the transcription factor gene designations from [72]. We also cross-referenced differential expression categories and co-expression module membership with Wormbase-defined *C. elegans* orthologs found have sex-biased differential expression by Ortiz et al. [27], which we used to test for enrichment with G-tests.

SNP and molecular evolution analysis

We called single nucleotide variants between AF16 and HK104 based on Illumina paired-end sequencing of HK104 to ~33x coverage using identical methods of Thomas et al. [35], yielding 761,531 SNPs and 173,341 indels. Sequences are available in NCBI in project accession PRJNA509247. We calculated the per-bp density of SNPs (π) in the pairwise comparison of AF16 and HK104 in a 500bp window upstream (and downstream) of coding sequences, excluding genes internal to operons (and using just the 5'-most or 3'-most operon gene for upstream or downstream sequence, respectively). 1070 operons comprising 2573 genes were identified based on orthology and synteny with annotated *C. elegans* operons, as in Tu et al. [57]. We also calculated the per-bp incidence of SNPs for different genomic features on a per-gene basis, including non-synonymous sites, synonymous sites, and introns, in addition to the 500bp flanking regions, after masking non-covered and low-quality sites. The effective number of codons (ENC) metric of biased codon usage was calculated for each gene in the *C. briggsae* reference genome WS253 with codonw (J. Peden, http://codonw.sourceforge.net). We used 6911 coding sequence divergence values (dN/dS') for 1-1 orthologs between *C. briggsae* and *C. nigoni* from Thomas et al. [35].

Declarations

Ethics approval and consent to participate: not applicable.

Consent to publish: not applicable.

Acknowledgements

References

1. Ungerer MC, Johnson LC, Herman MA: **Ecological genomics: understanding gene and genome function in the natural environment.** *Heredity* 2007, **100:**178.
2. de Visser JAGM, Hermisson J, Wagner GP, Meyers LA, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al: **Evolution and detection of genetic robustness.** *Evolution* 2003, **57:**1959-1972.
3. Flatt T: **The evolutionary genetics of canalization.** *The Quarterly Review of Biology* 2005, **80:**287-316.
4. Grishkevich V, Yanai I: **The genomic determinants of genotype × environment interactions in gene expression.** *Trends in Genetics* 2013, **29:**479-487.
5. Gagneur J, Stegle O, Zhu C, Jakob P, Tekkedil MM, Aiyar RS, Schuon A-K, Pe'er D, Steinmetz LM: **Genotype-Environment Interactions Reveal Causal Pathways That Mediate Genetic Effects on Phenotype.** *PLOS Genetics* 2013, **9:**e1003803.
6. Wray GA: **The evolutionary significance of cis-regulatory mutations.** *Nature Reviews Genetics* 2007, **8:**206-216.

7. Massey JH, Wittkopp PJ: **The genetic basis of pigmentation differences within and between Drosophila species.** In *Current Topics in Developmental Biology. Volume* 119. Edited by Orgogozo V: Academic Press; 2016: 27-61

8. Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, Jennings EG, Lee TI, True HL, Lander ES, Young RA, Silver PA: **Remodeling of yeast genome expression in response to environmental changes.** *Molecular Biology of the Cell* 2001, **12:**323-337.

9. Smith S, Bernatchez L, Beheregaray LB: **RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species.** *BMC Genomics* 2013, **14:**375.

10. Lindquist S, Craig EA: **The Heat-Shock Proteins.** *Annual Review of Genetics* 1988, **22:**631-677.

11. Hecker M, Lambeck S, Toepfer S, van Someren E, Guthke R: **Gene regulatory network inference: Data integration in dynamic models—A review.** *Biosystems* 2009, **96:**86-103.

12. Stern DL, Orgogozo V: **The loci of evolution: how predictable is genetic evolution?** *Evolution* 2008, **62:**2155-2177.

13. Carroll SB: **Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution.** *Cell* 2008, **134:**25-36.

14. Wittkopp PJ, Kalay G: **Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence.** *Nature Reviews Genetics* 2012, **13:**59-69.

15. Tirosh I, Reikhav S, Levy AA, Barkai N: **A yeast hybrid provides insight into the evolution of gene expression regulation.** *Science* 2009, **324:**659.

16. Smith EN, Kruglyak L: **Gene–environment interaction in yeast gene expression.** *PLoS Biol* 2008, **6:**e83.

17. Wittkopp PJ, Haerum BK, Clark AG: **Regulatory changes underlying expression differences within and between Drosophila species.** *Nature Genetics* 2008, **40:**346-350.

18. Wittkopp PJ, Haerum BK, Clark AG: **Evolutionary changes in cis and trans gene regulation.** *Nature* 2004, **430:**85.

19. Keightley PD, Lynch M: **Toward a realistic model of mutations affecting fitness.** *Evolution* 2003, **57:**683-685.

20. Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson GS: **A gene expression map for *Caenorhabditis elegans*.** *Science* 2001, **293:**2087-2092.

21. Grishkevich V, Ben-Elazar S, Hashimshony T, Schott DH, Hunter CP, Yanai I: **A genomic bias for genotype–environment interactions in C. elegans.** *Molecular Systems Biology* 2012, **8**.

22. Li Y, Alvarez OA, Gutteling EW, Tijsterman M, Fu J, Riksen JA, Hazendonk E, Prins P, Plasterk RH, Jansen RC, et al: **Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*.** *PLoS Genetics* 2006, **2:**e222.

23. Rockman MV, Skrovanek SS, Kruglyak L: **Selection at linked sites shapes heritable phenotypic variation in *C. elegans*.** *Science* 2010, **330:**372-376.

24. Prasad A, Croydon-Sugarman M, Murray RL, Cutter AD: **Temperature-dependent fecundity associates with latitude in *Caenorhabditis briggsae*.** *Evolution* 2011, **65:**52-63.

25. Poullet N, Vielle A, Gimond C, Ferrari C, Braendle C: **Evolutionarily divergent thermal sensitivity of germline development and fertility in hermaphroditic Caenorhabditis nematodes.** *Evolution and Development* 2015, **17:**380-397.

26. Stegeman GW, Bueno de Mesquita M, Ryu WS, Cutter AD: **Temperature-dependent behaviours are genetically variable in the nematode *Caenorhabditis briggsae*.** *Journal of Experimental Biology* 2013, **216:**850-858.
27. Ortiz MA, Noble D, Sorokin EP, Kimble J: **A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode *Caenorhabditis elegans*.** *G3: Genes|Genomes|Genetics* 2014, **4:**1765-1772.
28. Thomas CG, Li RH, Smith HE, Woodruff GC, Oliver B, Haag ES: **Simplification and desexualization of gene expression in self-fertile nematodes.** *Current Biology* 2012, **22:**2167-2172.
29. Reinke V, Gil IS, Ward S, Kazmer K: **Genome-wide germline-enriched and sex-biased expression profiles in *Caenorhabditis elegans*.** *Development* 2004, **131:**311-323.
30. Yin D, Schwarz EM, Thomas CG, Felde RL, Korf IA, Cutter AD, Schartner CM, Ralston EJ, Meyer BJ, Haag ES: **Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins.** *Science* 2018, **359:**55-61.
31. Reinke V, Smith HE, Nance J, Wang J, Van Doren C, Begley R, Jones SJM, Davis EB, Scherer S, Ward S, Kim SK: **A global profile of germline gene expression in *C. elegans*.** *Molecular Cell* 2000, **6:**605-616.
32. Reinke V, Cutter AD: **Germline expression influences operon organization in the *Caenorhabditis elegans* genome.** *Genetics* 2009, **181:**1219-1228.
33. Albritton SE, Kranz A-L, Rao P, Kramer M, Dieterich C, Ercan S: **Sex-biased gene expression and evolution of the X chromosome in nematodes.** *Genetics* 2014, **197:**865-883.
34. GuhaThakurta D, Palomar L, Stormo GD, Tedesco P, Johnson TE, Walker DW, Lithgow G, Kim S, Link CD: **Identification of a novel cis-regulatory element involved in the heat shock response in *Caenorhabditis elegans* using microarray gene expression and computational methods.** *Genome Research* 2002, **12:**701-712.
35. Thomas CG, Wang W, Jovelin R, Ghosh R, Lomasko T, Trinh Q, Kruglyak L, Stein LD, Cutter AD: **Full-genome evolutionary histories of selfing, splitting and selection in *Caenorhabditis*.** *Genome Research* 2015, **25:**667-678.
36. Ross JA, Koboldt DC, Staisch JE, Chamberlin HM, Gupta BP, Miller RD, Baird SE, Haag ES: ***Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination.** *PLoS Genetics* 2011, **7:**e1002174.
37. Cutter AD: **X exceptionalism in Caenorhabditis speciation.** *Molecular Ecology* 2018, **0**.
38. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK: **The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*.** *Nature Genetics* 2005, **37:**544-548.
39. Blumenthal T, Evans D, Link CD, Guffanti A, Lawson D, Thierry-Mieg J, Thierry-Mieg D, Chiu WL, Duke K, Kiraly M, Kim SK: **A global analysis of *Caenorhabditis elegans* operons.** *Nature* 2002, **417:**851-854.
40. Rockman MV, Kruglyak L: **Genetics of global gene expression.** *Nature Reviews Genetics* 2006, **7:**862.
41. Merritt C, Rasoloson D, Ko D, Seydoux G: **3' UTRs are the primary regulators of gene expression in the *C. elegans* germline.** *Current Biology* 2008, **18:**1476-1482.
42. Saito TL, Hashimoto S-i, Gu SG, Morton JJ, Stadler M, Blumenthal T, Fire A, Morishita S: **The transcription start site landscape of C. elegans.** *Genome Research* 2013, **23:**1348-1361.
43. Sikkink KL, Reynolds RM, Ituarte CM, Cresko WA, Phillips PC: **Environmental and evolutionary drivers of the modular gene regulatory network underlying**

phenotypic plasticity for stress resistance in the nematode Caenorhabditis remanei. *bioRxiv* 2018.

44. Wu J-c, Go AC, Samson M, Cintra T, Mirsoian S, Wu TF, Jow MM, Routman EJ, Chu DS: **Sperm development and motility are regulated by PP1 phosphatases in *Caenorhabditis elegans*.** *Genetics* 2012, **190:**143.

45. Qian W, Zhang J: **Evolutionary dynamics of nematode operons: Easy come, slow go.** *Genome Research* 2008, **18:**412-421.

46. Zaslaver A, Baugh LR, Sternberg Paul W: **Metazoan operons accelerate recovery from growth-arrested states.** *Cell* 2011, **145:**981-992.

47. Castillo-Davis CI, Hartl DL, Achaz G: ***cis*-Regulatory and protein evolution in orthologous and duplicate genes.** *Genome Research* 2004, **14:**1530-1536.

48. Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL: **Evolution of proteins and gene expression levels are coupled in *Drosophila* and are independently associated with mRNA abundance, protein length, and number of protein-protein interactions.** *Molecular Biology and Evolution* 2005, **22:**1345-1354.

49. Jordan IK, Mariño-Ramírez L, Koonin EV: **Evolutionary significance of gene expression divergence.** *Gene* 2005, **345:**119-126.

50. Liao B-Y, Zhang J: **Evolutionary conservation of expression profiles between human and mouse orthologous genes.** *Molecular Biology and Evolution* 2006, **23:**530-540.

51. Tirosh I, Barkai N: **Evolution of gene sequence and gene expression are not correlated in yeast.** *Trends in Genetics* 2008, **24:**109-113.

52. Cutter AD, Payseur BA: **Genomic signatures of selection at linked sites: unifying the disparity among species.** *Nature Reviews Genetics* 2013, **14:**262-274.

53. Comeron JM, Williford A, Kliman RM: **The Hill-Robertson effect: evolutionary consequences of weak selection and linkage in finite populations.** *Heredity* 2008, **100:**19-31.

54. Hill WG, Robertson A: **Effect of linkage on limits to artificial selection.** *Genetical Research* 1966, **8:**269-294.

55. Cutter AD, Choi JY: **Natural selection shapes nucleotide polymorphism across the genome of the nematode *Caenorhabditis briggsae*.** *Genome Research* 2010, **20:**1103-1111.

56. Stiernagle TL: **Maintenance of *C. elegans*.** In *C elegans: A Practical Approach.* Edited by Hope IA. New York: Oxford University Press; 1999

57. Tu S, Wu MZ, Wang J, Cutter AD, Weng Z, Claycomb JM: **Comparative functional characterization of the CSR-1 22G-RNA pathway in Caenorhabditis nematodes.** *Nucleic Acids Research* 2015, **43:**208-224.

58. Fang Z, Cui X: **Design and validation issues in RNA-seq experiments.** *Briefings in Bioinformatics* 2011, **12:**280-287.

59. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data.** *Bioinformatics* 2014, **30:**2114-2120.

60. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29**.

61. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, Robinson MD: **Count-based differential expression analysis of RNA sequencing data using R and Bioconductor.** *Nat Protocols* 2013, **8:**1765-1786.

62. Nikolayeva O, Robinson MD: **edgeR for differential RNA-seq and ChIP-seq analysis: an application to stem cell biology.** In *Stem Cell Transcriptional Networks: Methods and Protocols.* Edited by Kidder BL. New York, NY: Springer New York; 2014: 45-79

63. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26:**139-140.

64. Law CW, Chen Y, Shi W, Smyth GK: **voom: precision weights unlock linear model analysis tools for RNA-seq read counts.** *Genome Biology* 2014, **15:**R29.

65. Smyth GK: **limma: Linear Models for Microarray Data.** In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor.* Edited by Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S. New York, NY: Springer New York; 2005: 397-420

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

67. Langfelder P, Horvath S: **WGCNA: an R package for weighted correlation network analysis.** *BMC Bioinformatics* 2008, **9:**559.

68. Zhang B, Horvath S: **A general framework for weighted gene co-expression network analysis.** *Statistical Applications in Genetics and Molecular Biology* 2005, **4:**1-45.

69. Langfelder P, Horvath S: **Eigengene networks for studying the relationships between co-expression modules.** *BMC Systems Biology* 2007, **1:**54.

70. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD: **PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium.** *Nucleic Acids Research* 2010, **38:**D204-D210.

71. Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, Waterston RH: **Comparison of** *C. elegans* **and** *C. briggsae* **genome sequences reveals extensive conservation of chromosome organization and synteny.** *PLoS Biol* 2007, **5:**e167.

72. Haerty W, Artieri C, Khezri N, Singh RS, Gupta BP: **Comparative analysis of function and interaction of transcription factors in nematodes: extensive conservation of orthology coupled to rapid sequence evolution.** *BMC Genomics* 2008, **9:**399.
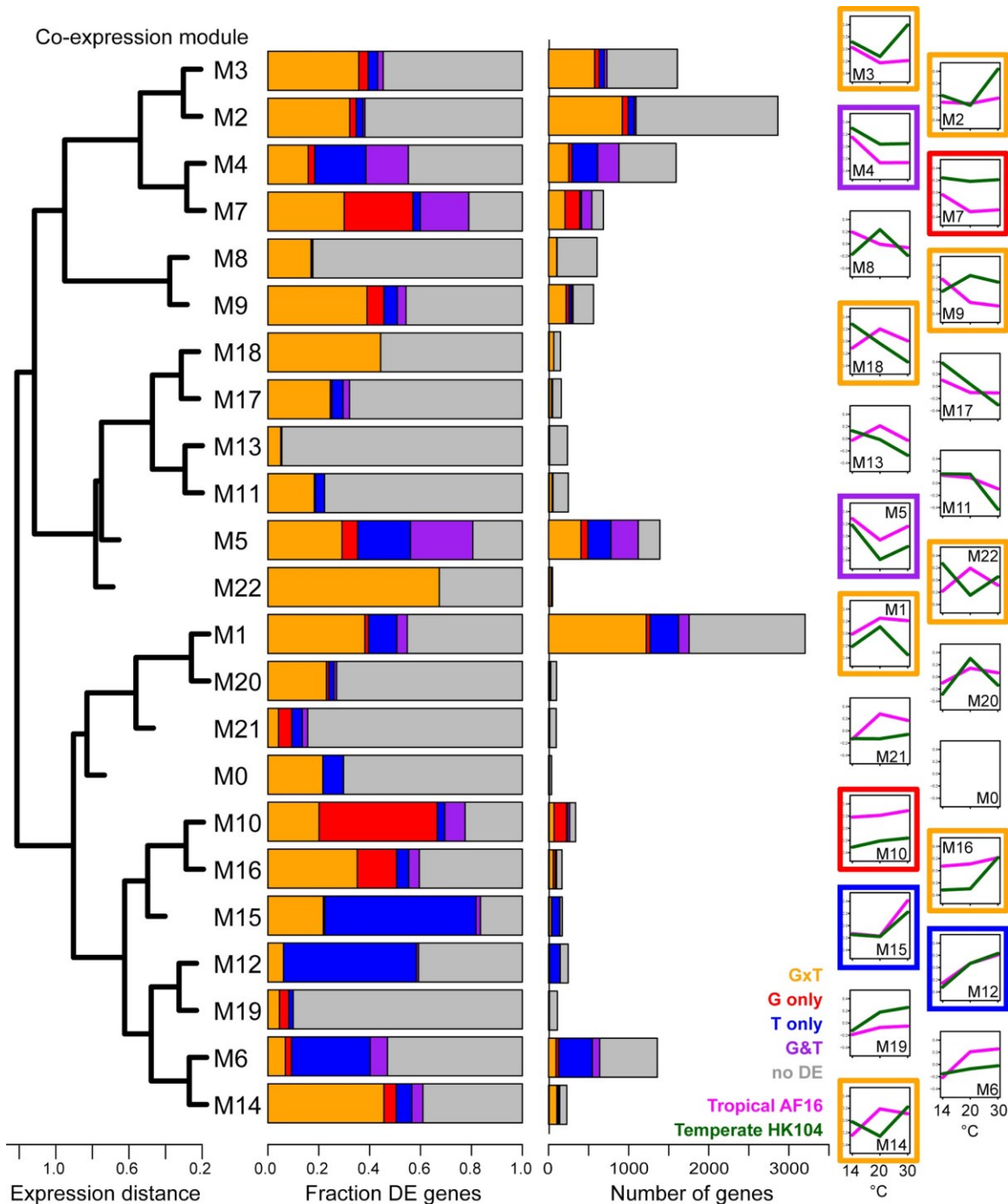
Figures
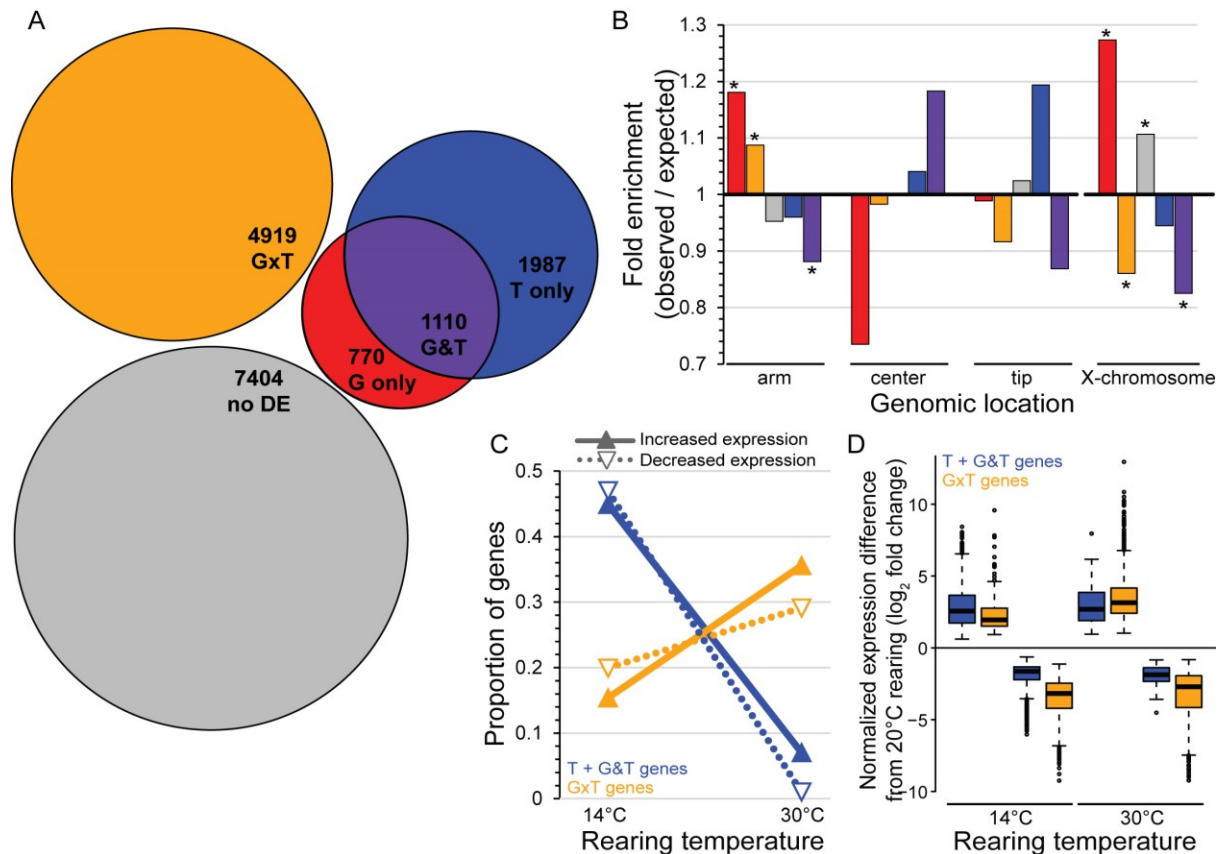
Figure 1. (**A**) Differential expression analysis identified 54% of 16,199 genes to have significant differential expression (5% FDR in limma). Of these genes, most had significant interaction effects (GxT), with significant independent effects of genotype and rearing temperature affecting 12.6% of genes with differential expression ("G&T") and the remainder of genes having altered expression due to genotype or rearing temperature treatments alone ("G only"; "T only"). (**B**) G-only and GxT genes are significantly enriched on autosomal arms, whereas G&T genes are enriched in autosomal centers (colors as in A; * indicates G-test Bonferroni corrected P<0.05). Genes with G-only or no differential expression ("no DE") are enriched on the X-chromosome, whereas genes with GxT and G&T patterns of differential expression are underrepresented on the X-chromosome (* indicates G-test Bonferroni corrected P<0.05). (**C**) In response to high 30°C rearing, similar proportions of GxE genes increase vs decrease expression relative to 20°C rearing expression profiles; similar proportions of GxE genes increase vs decrease expression when reared under cool 14°C conditions, but expression is modulated for fewer GxE genes under cool rearing than for hot rearing. By contrast, genes with a non-interacting effect of temperature on expression (T only and G&T genes) show disproportionate response to cool rearing. (**D**) The magnitude of expression change is similar for genes with a significant independent effect of temperature (T only and G&T genes) under chronic cold stress and chronic heat stress (median with interquartile range, whiskers show 1.5x interquartile range). For GxT genes, however, the magnitude of expression increase is greater under heat stress than cold stress.
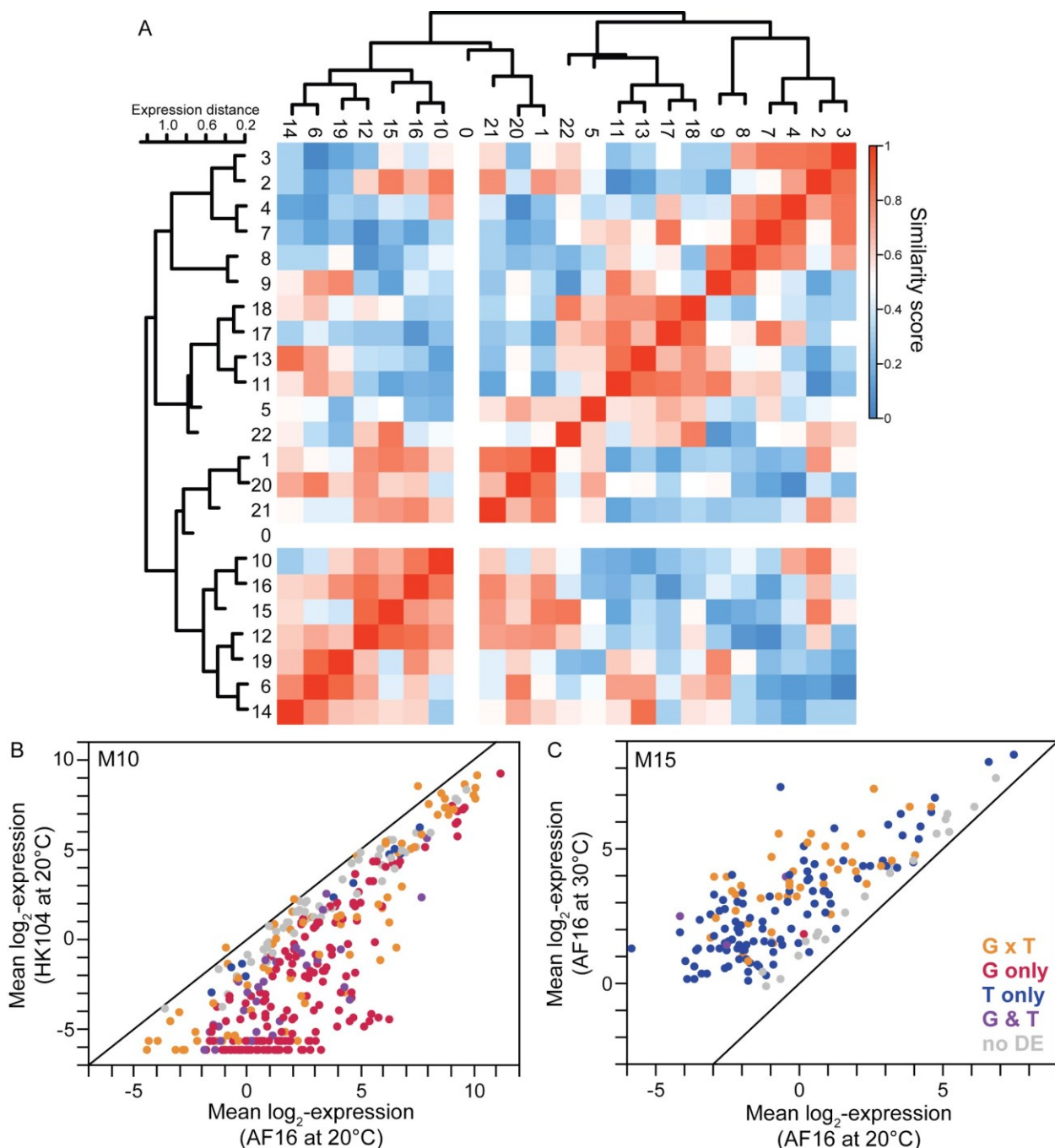
Figure 2. (**A**) WGCNA analysis yielded 23 co-expression modules for 16,199 genes (including module M0 with genes that did not cluster), after merging modules with expression similarity distance <0.25 from an initial set of 124 co-expressed gene sets. The dendrogram and heatmap summarize module expression profile similarity. For example, (**B**) module M10 is comprised of 338 genes with disproportionate representation of genotypic differences in expression, reflected as higher expression by the AF16 (Tropical) genotype. (**C**) Module M15, by contrast, is enriched in genes with individually significant differential expression due to rearing temperature, as reflected in higher expression when reared at 30°C. See also Figure 3.

25

Figure 3. Co-expression clustering of 16,199 genes into 23 modules, reflected in eigengene plots of normalized log$_2$-transformed expression profiles across rearing temperature treatments (14°C, 20°C, 30°C) for each genotype (Tropical AF16, Temperate HK104). Modules range in size from 3203 genes (M1) to 49 genes (M22). Module compositions contained distinctive representation of genes with individually-significant patterns of differential expression (cf. Figure 1; modules with highest incidence of T-only genes indicated with blue (M12, M15), G-only genes in red (M7, M10), G&T genes (M4, M5), and GxT genes with crossing or non-crossing reaction norms (M9, M14, M18, M22; M1, M2, M3, M16) (Supplementary Figure S3).
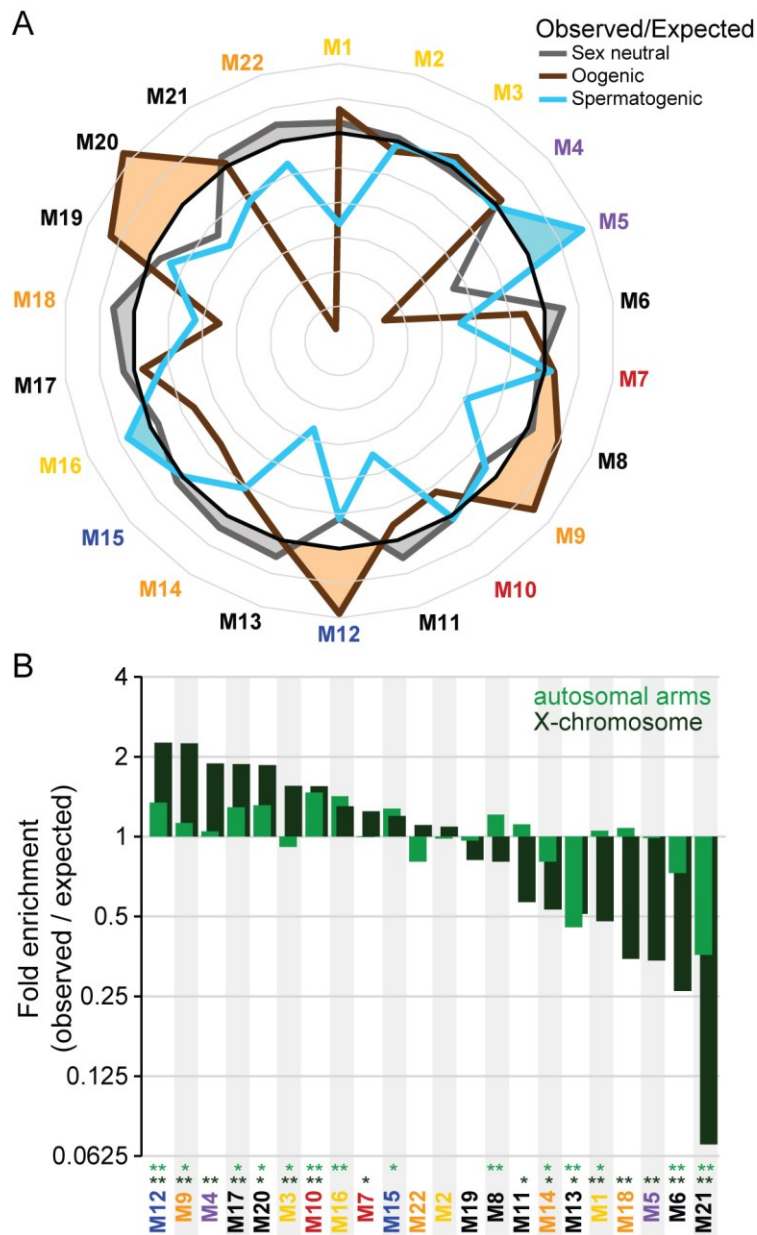
Figure 4. (**A**) Module gene over- and under-enrichment for *C. elegans* orthologs with spermatogenic, oogenic, or gender neutral expression profiles from Ortiz et al. (2014). Log-2 interval scale for observed/expected number of genes per module in radial plot, with black line indicating a value of 1 (outer curve indicates 4-fold enrichment, innermost curve indicates 2^{-5} under-enrichment). (**B**) Module gene over- and under-enrichment across the genome shows biases toward autosomal arms (values < 1 indicate enrichment in autosomal centers) and for X-linkage (values < 1 indicate enrichment on autosomes). Significant enrichment indicated by * (P<0.05 after Benjamini-Hochberg adjustment; FDR = 0.05) and ** (P<0.001 after Benjamini-Hochberg adjustment; FDR = 0.05). Coloring of module names in A and B corresponds to differential enrichment patterns indicated in Figure 3 (blue, T only; red, G only; purple, G&T; orange = GxT with crossing reaction norm profile; yellow, GxT with non-crossing reaction norm; gray, black, differential expression).

27
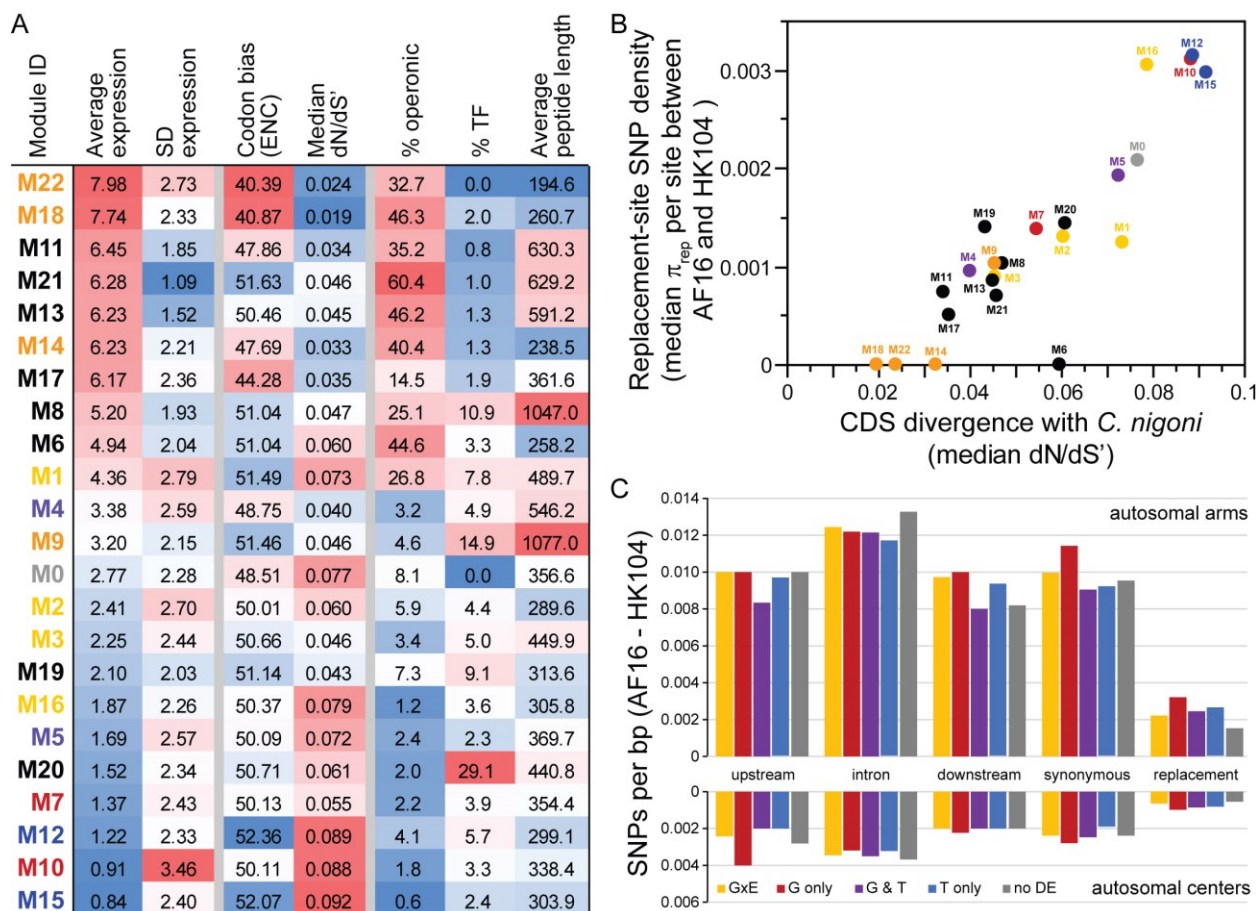
Figure 5. (**A**) Heatmap of module features, sorted by average normalized expression. Modules with high expression profiles tend to contain genes with stronger codon usage bias, greater sequence constraint (low dN/dS'), and more operons. (**B**) Modules with gene orthologs having little coding sequence divergence between *C. briggsae* and *C. nigoni* also have low densities of replacement-site SNPs in coding sequences. Coloring of module names in A and B corresponds to differential enrichment patterns indicated in Figure 3 (blue, T only; red, G only; purple, G&T; orange = GxT with crossing reaction norm profile; yellow, GxT with non-crossing reaction norm; gray, black, differential expression). (**C**) Among genes with distinctive profiles of individually-significant differential expression, linkage to autosomal arms versus centers represents the primary driver of SNP variation with little difference among DE categories for a given genomic site type (1kb upstream of CDS, intronic, 1kb downstream of CDS, synonymous sites).

Supplementary Information

Supplementary File S1. "SuppFile_voomNormFiltLog2Expr_DE_modules.csv" contains log-2 normalized expression values for each of the 16,199 genes analyzed in each replicate sample, as well as the category of differential expression (G-only, T-only, G&T, GxT, noDE) and name of the co-expression module (M0 through M22). Columns labeled with sample name (treatment-replicate), where treatment is a combination of genotype and rearing temperature for each of three biological replicates ("AF" = Tropical AF16 genotype, "HK" = Temperate HK104 genotype; 14=14°C rearing, 20=20°C rearing, 30=30°C rearing).

Supplementary File S2. "SuppTable_GO.xlsx" contains lists of gene ontology term enrichment with summary statistics for different PANTHER GO-slim categories for each co-expression module.

Supplementary Table S1. Number and percentage of reads that mapped to unique genomic locations with STAR.

Sample (treatment-rep)*	Number of uniquely mapped reads	% of reads that uniquely mapped
AF14-1	32007586	93.35%
AF14-2	47012080	93.84%
AF14-3	54523984	94.15%
AF20-1	57787850	94.17%
AF20-2	62122726	93.43%
AF20-3	43596994	93.75%
AF30-1	66295835	93.74%
AF30-2	35618842	93.20%
AF30-3	33087508	90.10%
HK14-1	51264144	93.39%
HK14-2	43616826	93.65%
HK14-3	48035023	92.74%
HK20-1	30262908	90.76%
HK20-2	49528582	93.95%
HK20-3	41332043	94.26%
HK30-1	51537189	91.44%
HK30-2	40748065	92.22%
HK30-3	35632338	73.62%

*AF=AF16, HK=HK104 genotypes; 14=14°C, 20=20°C, 30=30°C rearing
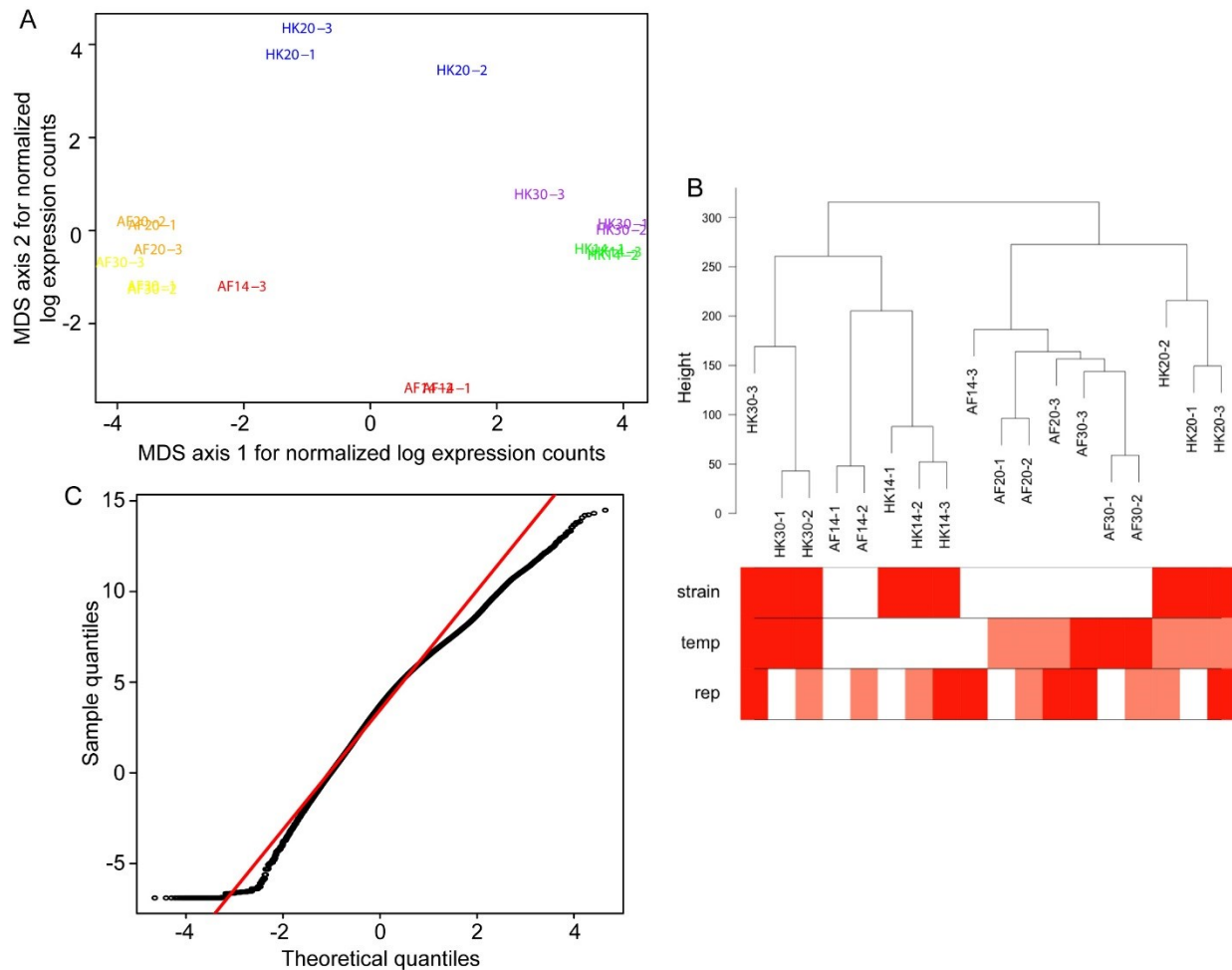
Supplementary Figure S1. (A) Multi-dimensional scaling plot (MDS) of filtered, normalized, and log-transformed count data for each sample. Sample labels (treatment-replicate) are colored by experimental treatment for a given genotype and rearing temperature combination ("AF" = Tropical AF16 genotype, "HK" = Temperate HK104 genotype; 14=14°C rearing, 20=20°C rearing, 30=30°C rearing). The x- and y-axes show the two principal components that explain most variation. Biological replicates that cluster together in the plot are more similar to each other, indicating consistency across replicates. (B) After filtering out genes with very low to no counts, TMM normalization for different library sizes, and log-transforming count data, a dendrogram reveals similarity of samples within strain and within temperature ("temp"), but not replicate ("rep"); red shading along a row indicates shared strain, temperature or replicate value. This suggests data heterogeneity among samples due to experimental treatments, not batch effects. (C) Quantile-quantile plot for normalized, voom-transformed count data shows good approximation to a normal distribution (red line).
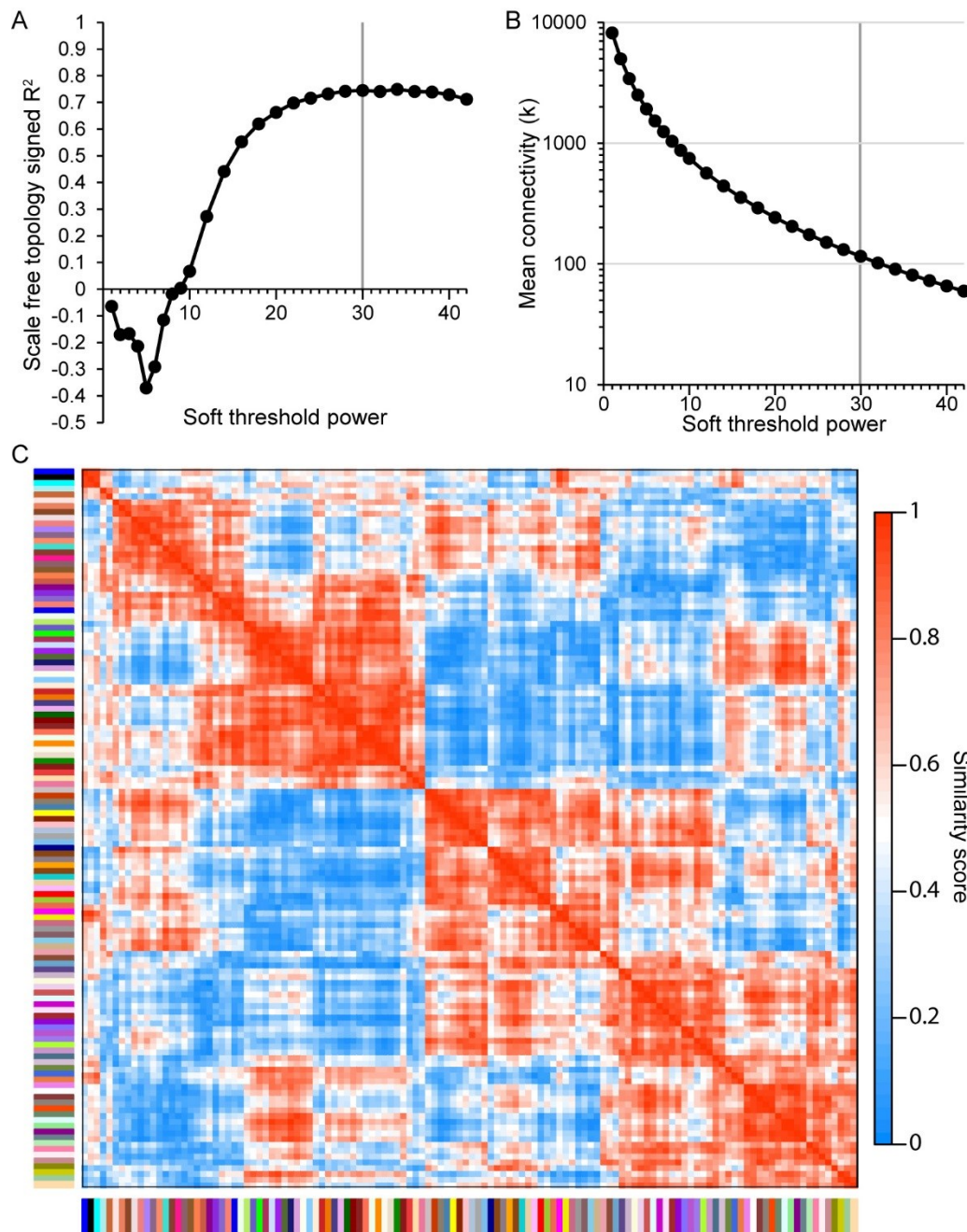
Supplementary Figure S2. (A) Analysis of soft-thresholding powers revealed 30 to be the power at which the scale-free fit is maximized (R^2 = 0.75) and most closely approximates a scale-free network. (B) Analysis of a range of soft-thresholding powers revealed a value of 30 to have a mean connectivity of at least 100 (k = 115), while also maximizing fit to a scale-free network. (C) Clustering of 16,199 genes with WGCNA into a preliminary set of 124 co-expression clusters (red in heatmap indicates maximum similarity and blue no similarity; color bars along the x- and y-axes correspond to the 124 clusters). Clusters with similarity distance <0.25 were merged, leading to the 22 co-expression modules (plus pseudo-module M0) analyzed in the main text.
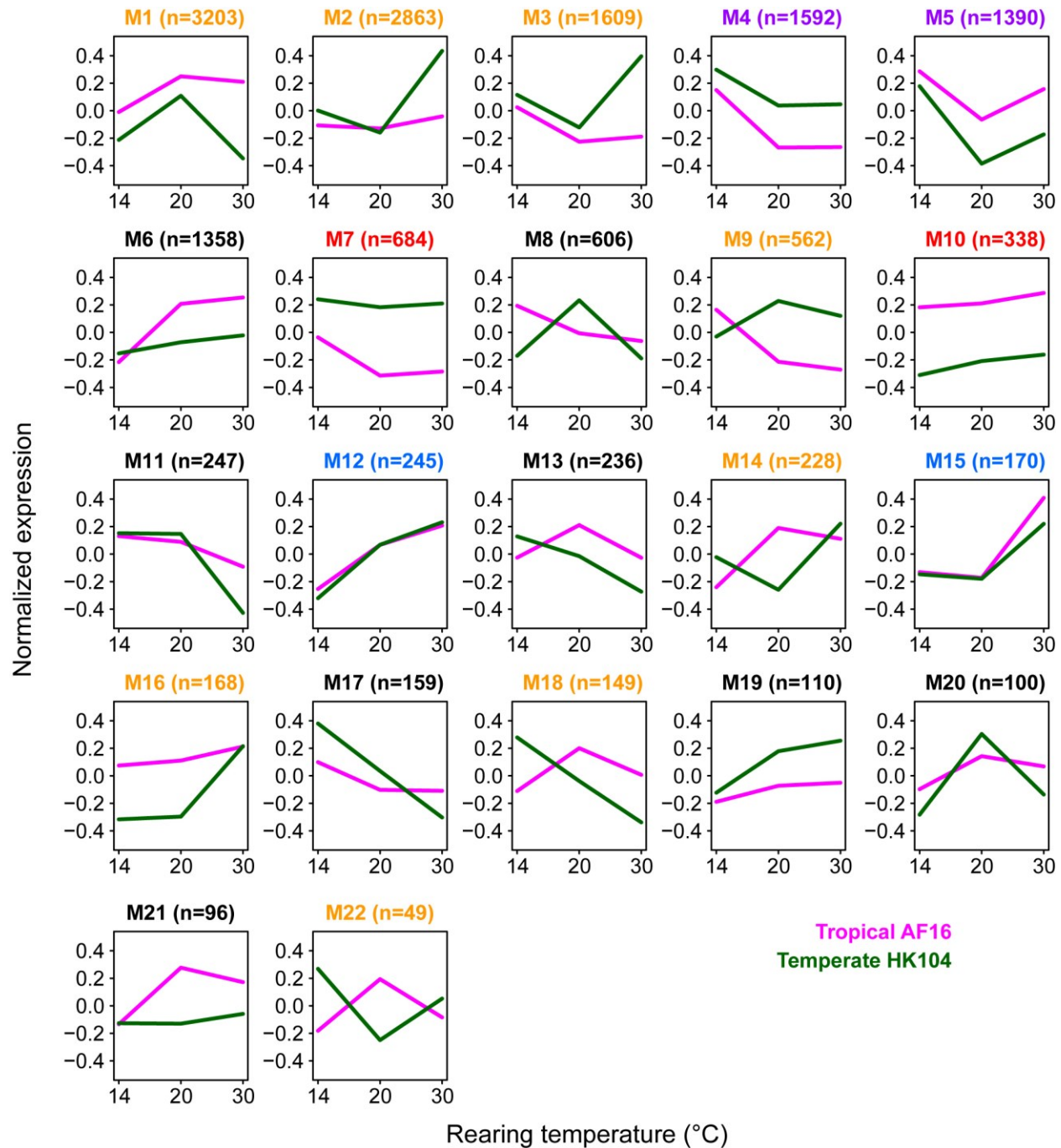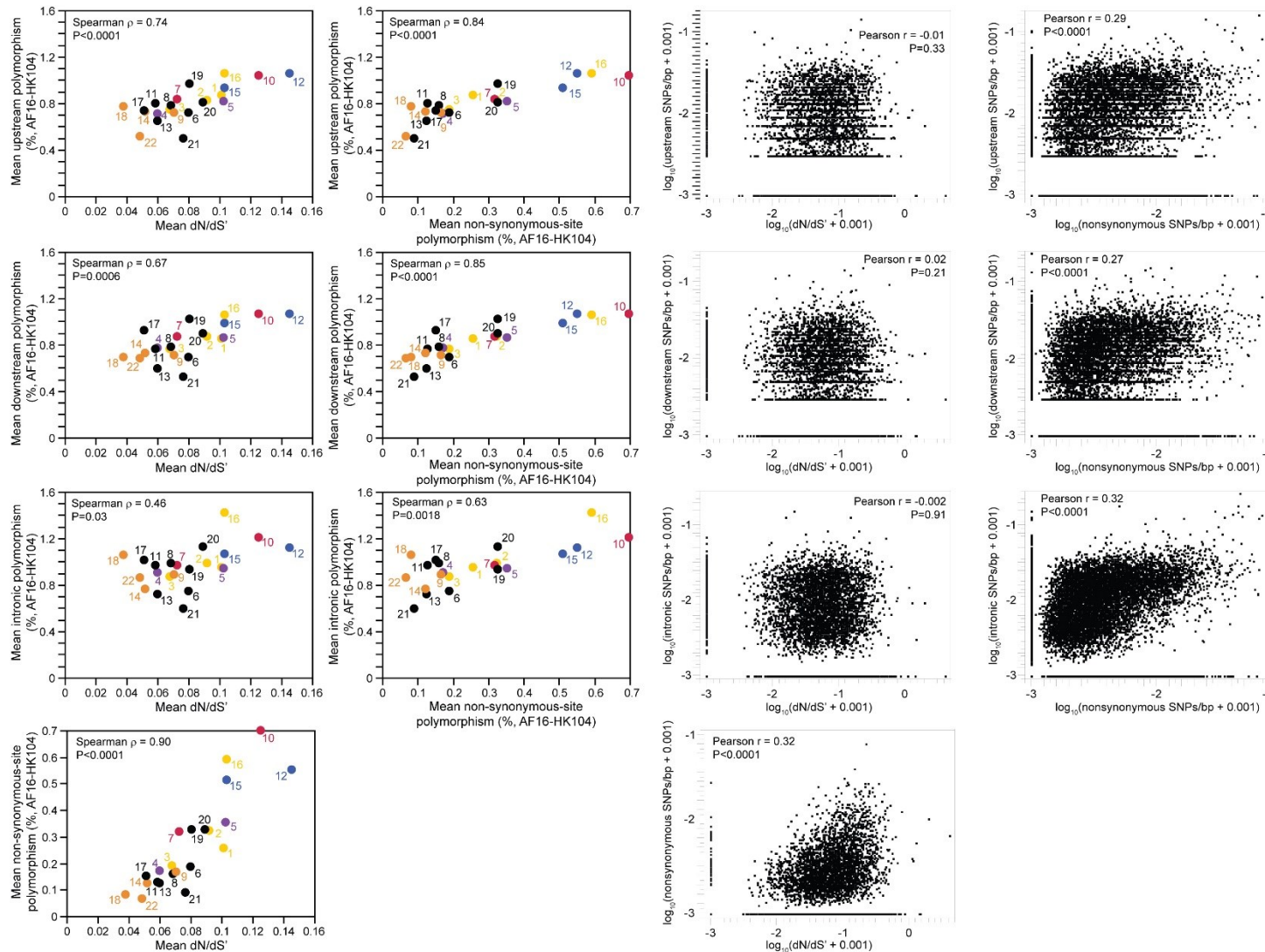
Supplementary Figure S3. Co-expression module eigengene plots of normalized, log$_2$-transformed expression across temperature treatments for each genotype (pink AF16, green HK104). Module names colored as in Figure 3 of the main text according to representation of individually significant differentially expressed genes within the module.
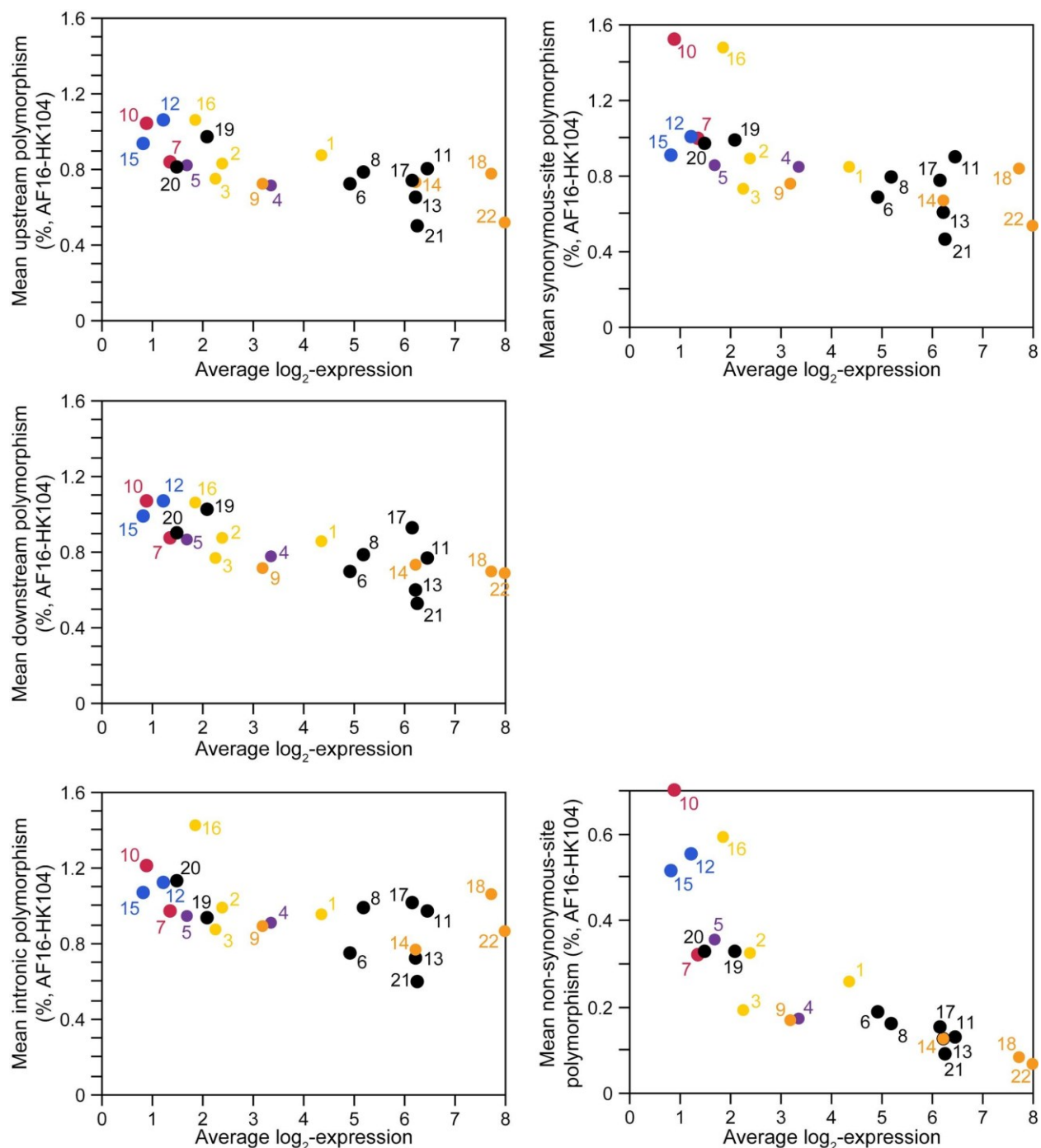
Supplementary Figure S4. SNP density in upstream, downstream, and intronic non-coding locations of genes and at synonymous and non-synonymous sites within coding sequences, averaged for genes within each co-expression module as a function of average module expression. Points are labeled and colored with module name as in Figure 3 of the main text.

Supplementary Figure S5. SNP density in upstream, downstream, and intronic non-coding locations of genes and at non-synonymous sites within coding sequences, averaged for genes within each co-expression module (left panels) or per gene (right panels) as a function of average module interspecies divergence (dN/dS') or non-synonymous site substitution. Modules are labeled and colored as in Figure 3 of the main text.