# Molecular evolution across developmental time reveals rapid divergence in early embryogenesis

Asher D. Cutter[1]*, Rose H. Garrett[1,2,3], Stephanie Mark[1], Wei Wang[1], Lei Sun[2,3]

[1]Department of Ecology & Evolutionary Biology, [2]Division of Biostatistics, Dalla Lana School of Public Health, [3]Department of Statistical Sciences
University of Toronto, Toronto, ON M6G1W3

*corresponding author: Asher Cutter, asher.cutter@utoronto.ca

# Abstract

Ontogenetic development hinges on the changes in gene expression in time and space within an organism, suggesting that the demands of ontogenetic growth can impose or reveal predictable pattern in the molecular evolution of genes expressed dynamically across development. Here we characterize co-expression modules of the *C. elegans* transcriptome, using a time series of 30 points from early-embryo to adult. By capturing the functional form of expression profiles with quantitative metrics, we find fastest evolution in the distinctive set of genes with transcript abundance that declines through development from young embryos. These genes are highly enriched for oogenic function (maternal provisioning), are non-randomly distributed in the genome, and correspond to a life stage especially prone to inviability in inter-species hybrids. These observations conflict with the "early conservation model" for the evolution of development, and provide only qualified support for the "hourglass model." Genes in co-expression modules that peak toward adulthood also evolve fast, being hyper-enriched for roles in spermatogenesis, implicating a history of sexual selection and relaxation of selection on sperm as key factors driving rapid change to ontogenetically distinguishable co-expression modules of genes. These predictable trends of molecular evolution for dynamically-expressed genes across ontogeny might predispose particular life stages, early embryogenesis in particular, to hybrid dysfunction in the speciation process.

# Introduction

Ontogenetic development hinges on the changes in gene expression in time and space within an organism. The dynamic molecular networks that specify cell proliferation and differentiation together produce morphogenesis, going from a single-celled zygote to a reproductively mature adult. Evolution favors maximal reproductive success to shape those gene expression dynamics and the functional properties of the proteins they encode, with the strength of selection pressures recorded in their sequences. Do the demands of ontogenetic growth impose or reveal predictable pattern in the molecular evolution of genes expressed dynamically across development? How do the cellular constraints to forming a whole organism in embryogenesis, and the life history constraints on a whole organism to reproduce successfully, translate into underlying gene sequence conservation? What rules are there, if any, to govern the molecular evolution of development? We can address these questions from the perspective of genetic controls or from spatio-temporal dynamics in the formation of the structures of a complete organism.

One important axis of evolutionary predictability in development emphasizes non-coding regulatory controls: the relative roles of *cis*- and *trans*-regulators versus protein functionality in developmental variation and divergence (Wray 2007; Carroll 2008; Stern and Orgogozo 2008). Adaptive changes to protein function also are important in evolution, as evinced by phenotypic examples and estimates of a large fraction of coding substitutions fixed between species by positive selection (Hoekstra and Coyne 2007; Galtier 2016). Empirical data, however, support an outsized role of *cis*-regulatory change in morphological divergence between species (Stern and Orgogozo 2008; Wittkopp and Kalay 2012), with one of the arguments explaining this pattern holding that lower pleiotropy of *cis*-regulatory changes facilitates adaptive divergence. A parallel notion about pleiotropic effects based on genetic network structure predicts that ‘hub’ genes with many interaction partners should evolve slowly, with faster evolution in genes with fewer interactions; however, data do not strongly favor this ‘hub’ effect for coding sequence evolution (Jordan, Wolf, and Koonin 2003; Batada, Hurst, and Tyers 2006). Both of these perspectives emphasize the genetic architecture, however, rather than the spatio-temporal architecture in the evolution of development.

Taking a more physical perspective on molecular evolutionary predictability in development, one ontogenetic means of limiting pleiotropy of a gene is tissue- or cell-specificity of gene expression: narrow breadth of expression in space must narrow the potentially negative pleiotropic effects of changes to gene expression or protein function (Stern 2000; Carroll 2005; Haygood et al. 2010; He et al. 2012). For example, mammalian genes with greater tissue-specificity of expression evolve faster in coding sequence but slower in terms of expression change (Liao and Zhang 2006a). Temporal specificity of gene expression provides another dimension that can restrict or exacerbate the potential for pleiotropic effects of change to gene regulation or structure. Similar to the argument for spatial extent of gene activity, narrower duration of expression in ontogeny ought to narrow the potential for negative pleiotropic effects of changes to a given gene. A counter-argument, however, points out the unidirectional nature of time: changes to early points in development can cascade through ontogeny with disproportionate force (Poe and Wake 2004; Irie and Kuratani 2014; Arthur 2015). Because most new mutations with fitness effects are deleterious (Keightley and Lynch 2003), this "early conservation" or "generative entrenchment" view predicts slower evolution of genes expressed earlier in embryonic development, as has been reported for mouse and zebrafish (Roux and Robinson-Rechavi 2008; Irie and Kuratani 2014). By contrast, the most famous temporal paradigm derives from embryological observations of a 'phylotypic stage' with greatest phenotypic constraint relative to earlier and later timepoints in development, the 'hourglass model' (Raff 1996; Kalinka and Tomancak 2012). Applications of this idea to molecular data have renewed interest in it beyond morphology (Castillo-Davis and Hartl 2002; Cutter and Ward 2005; Davis, Brandman, and Petrov 2005; Hazkani-Covo, Wool, and Graur 2005; Cruickshank and Wade 2008; Domazet-Loso and Tautz 2010; Kalinka et al. 2010; Irie and Kuratani 2011; Levin et al. 2012; Gerstein et al. 2014). Different still, population genetics arguments about weaker purifying selection on genes expressed by just one sex, like maternal-effect gene products deposited in eggs, predict disproportionately rapid evolution of such maternally-deposited genes involved in early-embryogenesis of zygotes (Cruickshank and Wade 2008).

These 'evo-devo' ideas, however, are largely focused on embryogenesis, and do not explicitly incorporate the entirety of ontogeny over an organism's life cycle (Kalinka and Tomancak 2012). Ideas from the evolution of aging and senescence, by contrast, consider late life (Flatt and Schmidt 2009). In particular, the mutation-accumulation theory of aging predicts more rapid

evolution of genes expressed following the onset of reproductive maturity than for those expressed earlier because diminishing reproductive value following maturity weakens the ability of selection to eliminate mutations (Medawar 1952; Charlesworth 1993; Promislow and Tatar 1998; Partridge 2001). Genes with expression in just one sex also ought to experience weaker purifying selection than other genes, leading to faster protein evolution, because mutations would be exposed to selection in just half of the population (Cruickshank and Wade 2008). Sexual contests and mate choice drive rapid divergence in morphological ornaments and their genetic underpinnings (Swanson and Vacquier 2002; Ellegren and Parsch 2007), so sexual selection also predicts faster evolution of sex-biased genes and of genes expressed late in life, to the extent that their development gets specified toward adulthood. The coding sequences of adult-expressed genes do tend to evolve faster than embryonic genes in a number of taxa (Cutter and Ward 2005; Davis, Brandman, and Petrov 2005; Artieri, Haerty, and Singh 2009).

*C. elegans* and related nematodes are well-known for their similarity in form (Haag et al. 2007), despite the long times since species separated from one another (Cutter 2008). Indeed, the embryonic cell lineage of different *Caenorhabditis* species is outwardly preserved to an astonishing degree (Zhao et al. 2008; Memar et al. 2018), albeit with some key differences in timing of developmental milestones (Levin et al. 2012). This similarity of form, however, masks substantial evolution of genetic interactions as revealed by pronounced embryonic mortality in interspecies hybrids (Baird and Seibert 2013; Bundus, Alaei, and Cutter 2015). Developmental system drift is thought to underlie evolutionary change to spindle movement in the first cell division of embryos (Riche et al. 2013; Farhadifar et al. 2015; Valfort et al. 2018). Experiments also demonstrate that morphological stasis and even conserved expression patterns mask profound *cis*-regulatory divergence of conserved coding genes (Barriere, Gordon, and Ruvinsky 2012; Barrière and Ruvinsky 2014; Verster et al. 2014; Barkoulas et al. 2016). Molecular evolution analysis of genes expressed differentially across post-embryonic development from microarray data reported faster evolution of coding sequences associated with the onset of reproductive maturity, but little directional effect of timing in embryogenesis (Cutter and Ward 2005). These collective observations motivate characterization of molecular evolution for gene expression dynamics across the entirety of ontogeny.

Here we characterize co-expression modules of the *C. elegans* transcriptome over the full course of development, using a time series of 30 points from early embryo to adults (Gerstein et al. 2010; Gerstein et al. 2014). By coarse graining the functional form of these ontogenetic trajectories of gene expression, we capture quantitative metrics that reveal how developmental dynamics relate to rates of molecular evolution. Surprisingly, we find the fastest evolution, both in terms of coding sequence divergence and ortholog turnover, in the unique set of genes with a high abundance of transcripts in young embryos that then decline in expression level over development. These genes are highly enriched for oogenic function, implying that they are comprised disproportionately of maternally-deposited transcripts. Genes in co-expression modules that show rising post-embryonic expression that peaks toward adulthood also exhibit especially fast evolutionary change. These latter gene modules are hyper-enriched for roles in spermatogenesis, implicating a history of gametic co-evolutionary interactions, sexual selection, and relaxation of selection on sperm in *C. elegans* history as key factors driving rapid change to ontogenetically distinguishable co-expression modules of genes. By contrast, genes in modules of co-expression with high transcript abundance and constitutive expression across all or most of ontogeny have the strongest sequence conservation. These results are consistent with temporal breadth of gene expression in development as delimiting bounds for molecular evolution, mediated by ongoing genetic interactions rather than pleiotropic cascades.

# Results

We defined 14 co-expression modules to describe clusters of the 19,711 genes that get expressed across 30 timepoints from early embryo through young adult stages of hermaphrodite *C. elegans* (Figure 1), based on ModENCODE transcriptome profiling data (Supplementary Table S1) (Gerstein et al. 2010; Gerstein et al. 2014). To obtain quantitative metrics describing the shape of each co-expression module, we then fit a cubic function to the gene expression profiles of each of the 14 developmental time series (Figure 1). The parameter values extracted from the cubic fits capture the overall expression level ($\alpha$), increasing or decreasing trends in expression across development ($\beta_1$), the degree of concave versus convex expression dynamics over ontogeny ($\beta_2$), and how S-shaped are the expression dynamics ($\beta_3$). Four modules show consistent expression

with little change across development (M3, M6, M12, M13). These 'constitutive' gene expression modules differ from one another primarily in the overall magnitude of expression (highest $\alpha$=8.91 for M6, lowest $\alpha$=0.66 for M13) and include the three largest modules by gene membership (M3, M12, M13) (Figure 1). By contrast, five modules exhibited hump-shaped expression dynamics with low expression in early embryos coupled to peak expression in late embryogenesis ($\beta_1>>0$, $\beta_2<<0$, $\beta_3<<0$; M1, M2, M5, M7, M8). Module M4 was unique among all modules in showing peak expression in early embryogenesis, which then declined across developmental time ($\beta_1<<0$). The four remaining modules displayed peak expression in post-embryonic stages (M9, M10, M11, M14), with especially strong up-regulation toward adulthood in M10 and M11 (Figure 1).

Given the utility of coarse graining the module expression patterns with a cubic function fit, we next fit the cubic functional form to each gene and extracted the gene-wise parameters for $\alpha$, $\beta_1$, $\beta_2$, and $\beta_3$ (Supplementary Figure S1, Supplementary Figure S2). Discriminant analysis demonstrated that values for these four parameters could correctly determine the co-expression module identity for 92.9% of genes, indicating that gene-wise cubic function fitting captures well the key distinguishing features of ontogenetic expression dynamics. The quadratic and cubic terms ($\beta_2$, $\beta_3$) were generally symmetric in distribution around zero (Supplementary Figure S1), whereas the $\alpha$ parameter suggests a bimodal distribution of expression-levels across genes and the linear change parameter illustrates how most genes increase in expression across ontogeny ($\beta_1>0$) but that a subset primarily composed of genes in M4 decline strongly in expression ($\beta_1<0$).

Upon defining these ontogenetically dynamic gene expression modules, we investigated their distinguishing features in terms of genomic organization, function and molecular evolution. Interestingly, genes from related expression profiles showed distinctive chromosome biases. Five modules were enriched on the X-chromosome, all of which corresponded to those with peak expression in late embryogenesis (M1, M2, M5, M7, M8; Figure 2). The early-embryogenesis module M4 showed the greatest chromosomal bias of any module, being >2-fold enriched on Chromosome II and tended to be underrepresented on all other chromosomes (Figure 2). Genes from those modules with peak post-embryonic expression, by contrast, showed enrichment on

chromosomes IV and V (M9, M10, M11, M14), and highly-expressed 'constitutive' modules showed enrichment on chromosomes I and III (M3, M6, M12; Figure 2). When we looked within chromosomes at their recombination domain structure of arms versus centers, we found genes for most modules to be present in their expected proportions given chromosomal gene densities (Figure 2). However, genes in M4 were significantly enriched in arms on Chromosome II, the chromosome where M4 genes are exceptionally abundant, and were elevated on arms relative to centers of other chromosomes (Figure 2). Post-embryonic modules M9 and M10, as well as the low-expression 'constitutive' module M13, also showed significant enrichment on arms of several chromosomes (Figure 2). By contrast, the highly-expressed 'constitutive' module M12 was under-enriched on the arms of Chromosomes II and V (Figure 2). At a more local scale of genome organization, we found that three modules were hyper-enriched for membership in operons, with each of the highly-expressed 'constitutive' modules M3, M6 and M12 containing >40% of their genes in operons (Figure 3); just 20.5% of coding genes overall occur in operons. Of the remaining modules, only M13 (the fourth 'constitutive' module) and M8 had >10% operonic genes, and <4% of genes occurred in operons for all four modules with post-embryonic peak expression (M9, M10, M11, M14) (Figure 3).

We cross-referenced the gene composition of co-expression modules with those gene sets identified by Ortiz et al. (Ortiz et al. 2014) to have sex-neutral, oogenic or spermatogenic enrichment of expression. These three expression categories had been inferred from differential expression of dissected gonads that had either active oocyte-only or sperm-only development (Ortiz et al. 2014). The early-embryogenesis module M4 showed extreme enrichment for oogenic genes (57%), with the next most enriched modules for oogenic genes being 'constitutive' modules M3 (24%) and M12 (23%) (Figure 3, yellow portion of bar plots). By contrast, the four modules with peak expression in post-embryonic stages contained almost no oogenic genes, instead being exceptionally enriched for spermatogenic genes (75% to 92%) (Figure 3; M9, M10, M11, M14). As expected of genes with sperm-related function (Reinke and Cutter 2009), operons were rarest in these modules (M9-M11, M14) (Figure 3). Eight of the 14 modules overall were comprised of >50% sex-neutral genes, including all five of those with peak expression late in embryogenesis, though three of the 'constitutive' modules contained the highest abundance of them (71% to 82%; M3, M6, M12) (Figure 3, gray portion of bar plots). Gene ontology (GO) and phenotype enrichment analysis (PEA) further showed these highly

expressed 'constitutive' modules to be enriched for basic cellular processes, like ribosomal and mitochondrial activity, embryonic defects, and chromosome segregation (M3, M6, M12; Figure 2B; Supplementary Table S2; Supplementary Table S3). By contrast, the modules showing increasing expression across embryogenesis and later stages tended to have significant enrichment of developmental GO and behavioral PEA terms, such as regulation of cell shape, neural activity, linker cell migration, and animal motility (Figure 2B, purple and green shaded modules; Supplementary Table S2; Supplementary Table S3). The most overrepresented terms across all co-expression modules were found in early-embryogenesis module M4, involving 21-fold enrichment of genes associated with protein heterodimerization activity (GO) and 19-fold enrichment of early embryonic chromatid segregation (PEA) (Figure 2B; Supplementary Table S2; Supplementary Table S3). Among the 105 genes in the *C. elegans* genome annotated with the protein heterodimerization activity GO term (GO:0046982), 69% correspond to histones, with most of the others comprised of TATA-box binding proteins, transcription factors, and CENP centromere-related proteins; M4 alone has 31 histones.

The co-expression modules also differ significantly in the rate at which their gene members evolve (Figure 3). Module M6 contains genes with strongest protein sequence conservation between species: median $K_A = 0.033$ implies that only about 3% of non-synonymous sites in codons have changed between *C. elegans* and *C. briggsae* since their common ancestor, estimated at 113 million generations ago (Cutter 2008). At the other extreme, curiously, those genes in M4 with peak expression in early embryogenesis comprise the most rapidly-evolving set of genes (median $K_A = 0.43$; Figure 3). As another sign of rapid evolution of genes in M4, this module contained the lowest percentage of genes with identifiable orthologs between *C. elegans* and *C. briggsae* (28% vs. 64% genome-wide and 92% ortholog pairs identified for M6; Figure 3). Curiously, however, module M4 has the highest fraction of genes (9.3% vs. 0.5% of genes overall) with near-zero values of $K_A$, implying exceptionally strong selective constraint on this subset of genes within M4: this subset is comprised entirely of histones, which are overrepresented in M4. These 14 histone genes, plus another subgroup of 15 genes with $K_A < 0.02$ (14 of which also are histones), imply that about 20% of M4's "early embryogenesis" genes encode histones that evolve extraordinarily slowly in stark contrast to the remaining 80% that evolve remarkably fast (Supplementary Figure S3). The only other module with substantial abundance of a group of exceptionally conserved coding sequences is 'constitutive' module M6

(4.9% of genes with near-zero $K_A$), which also shows the strongest sequence conservation on average irrespective of this exceptional subset of genes.

The four modules with peak post-embryonic expression and enrichment with spermatogenic function also evolve up to twice as rapidly as the genome-wide median $K_A = 0.121$ (median $K_A$ for "post-embryonic" modules M9, M10, M11, M14 from 0.185 to 0.261; Figure 3). Overall, co-expression modules with lower incidence of sex-neutral genes exhibit more rapid sequence divergence (Figure 4). As expected from previous analyses of *C. elegans* molecular evolution (Cutter, Dey, and Murray 2009), genes in those modules with higher average expression tend to evolve more slowly and show more sequence conservation (Figure 4); this manifests as unusually low divergence at synonymous sites only for M6 (median $K_S = 1.1$ vs. genome-wide median $K_S = 2.33$). An outlier to the $K_A$–expression relationship, however, is module M4: these early-embryogenesis genes show fast molecular evolution despite relatively high transcript levels (Figure 4). Our gene-wise analysis of coarse-grained cubic function parameters corroborate these findings (Supplementary Figure S4), with the four $\alpha$ and $\beta$ parameters alone being capable of explaining 11.5% of the variability in $K_A$ across genes (ANOVA $F_{4,12623} = 408.5$, $P<0.0001$; log-transformed $K_A$).

As a complement to the developmental time series analysis, we quantified rates of molecular evolution for a simpler partitioning of genes, by grouping genes according to the timepoint with highest observed expression. Median rates of protein sequence evolution were fastest for those genes with peak expression in the final L4 larval stage, young adults and in early embryos (Figure 5), corroborating the findings from the ontogenetic co-expression modules. Among those genes with peak expression in embryogenesis, genes with later peak expression tended to evolve more slowly (Figure 5), recapitulating the contrast of $K_A$ for "early embryogenesis" module M4 versus "late embryogenesis" co-expression modules (M1, M2, M5, M7, M8). However, genes with peak expression at timepoints 7-9 (180-240 minutes) exhibit a dip in median divergence (Figure 5), suggesting a trend of greater sequence conservation near ventral enclosure in embryogenesis reminiscent of patterns of expression divergence between species (Levin et al. 2012).

# Discussion

An understanding of the interplay between genes and phenotypes in the evolution of development must accommodate both phenotypic divergence and phenotypic conservation. The conservation of phenotype, however, including developmentally static phenotypes like *Caenorhabditis* embryogenesis (Zhao et al. 2008), need not imply conservation of the genetic pathways that produce them (Kalinka and Tomancak 2012). This idea is the essence of developmental system drift (DSD) (True and Haag 2001), and a key question is to what extent are different stages of development more or less susceptible to molecular divergence and DSD in a predictable way. Temporal trajectories of gene co-expression provide a means of interrogating these questions to determine what are the rules in the molecular evolution of development.

We observe the fastest coding sequence evolution for genes with peak expression early in embryogenesis (co-expression module M4), suggesting that this developmental stage near gastrulation may be especially prone to DSD. Why do genes with peak expression in early embryogenesis evolve so fast? This rapid evolution occurs despite an over-representation of histone proteins within this co-expression module that have exceptionally slow sequence evolution. Among the genes with rapid evolution, weaker purifying selection on maternally-provisioned transcripts provides one plausible basis for faster evolution of early-embryogenesis genes (Cruickshank and Wade 2008). A greater incidence of positive selection also could contribute to the rapid evolution of genes in M4, perhaps resulting from parent-offspring conflict or protein-protein co-evolution yielding DSD between gene partners (True and Haag 2001; Clark et al. 2009; de Juan, Pazos, and Valencia 2013). Moreover, genes in M4 are over-represented on autosomal arms (64% of M4 genes on arms vs. 37% genome average), genomic regions known to harbor genes with greater divergence (Cutter, Dey, and Murray 2009). Despite the extreme consistency of cell lineage in early embryos of different *Caenorhabditis* species (Zhao et al. 2008), the underlying molecular controls of early embryogenesis have diverged radically so that embryonic arrest near this stage represents the usual fate of inter-species hybrids (Baird, Sutherlin, and Emmons 1992; Baird and Seibert 2013; Dey et al. 2014; Bundus, Alaei, and Cutter 2015), consistent with divergence of genetic interactions with important biological consequences. Thus, the molecular evolutionary consequences of the biased composition of

genes with peak expression early in embryogenesis might be predisposed to DSD and to contribute to hybrid inviability in the speciation process.

Our observation of more rapid coding sequence evolution for genes with peak expression early in embryogenesis clearly conflicts with the "early conservation" model for the evolution of development (Kalinka and Tomancak 2012). Moreover, it has been argued that "conservation at the end of embryogenesis is not endorsed by any model" (Kalinka and Tomancak 2012), and yet the trend we observe shows just that, based on analyses of both co-expression modules and peak gene expression patterns. Our analysis of peak expression timing (Figure 5B), however, does show a hint of a phase of mid-embryonic development with strongest constraint, suggestive of the "hourglass model" that has been endorsed in *Caenorhabditis* from analysis of expression divergence (Levin et al. 2012). In our view, however, the unusually rapid evolution of early embryogenesis represents the more compelling pattern of molecular evolution requiring attention. Developmental stages associated with genes having faster rates of molecular evolution ought to be predisposed to more extensive developmental system drift, which can reveal itself by manifesting as the stage of development most sensitive to hybrid dysfunction in crosses between diverged species (Bundus, Alaei, and Cutter 2015).

Tissue-specific genes have faster coding sequence evolution in mammals (Liao and Zhang 2006a), and temporal specificity might lead to similar consequences. In our analysis, we can think of genes with extreme values of $\beta 1$, $\beta 2$, and $\beta 3$ as having greater temporal specificity of expression and therefore mutations to them having lower potential for pleiotropic effects; however, we observe no strong associations of these metrics with $K_A$ (Supplementary Figure S2). Alternately, we can think of mutations to genes with lower $\alpha$ (i.e. a profile of lower overall expression across ontogeny) as having lower potential for pleiotropic effects due to the rarity of gene products, and indeed genes with lower $\alpha$ evolve faster. Genes in module M4, with peak expression during early embryogenesis represent an important outlier to this trend, as they tend to have both fast sequence evolution and moderately high values of $\alpha$ (Supplementary Figure S2). In yeast, however, factors like translational robustness appear to be especially important in mediating the correspondence between expression level and rate of coding sequence evolution (Drummond et al. 2005), though it remains unclear how general this explanation holds across eukaryotes.

Our analysis puts to the side the question of the relative importance of regulatory versus coding changes in adaptation and morphological divergence (Wray 2007; Carroll 2008; Stern and Orgogozo 2008). Instead, we focus on coding sequence evolution to ask what features of ontogeny predict differences in the rates of evolution across genes. The differences in rates of coding sequence evolution among distinct co-expression modules implies a mapping between the nature of regulatory control and protein evolution. Previous studies of diverse animals show a weakly positive correlation between molecular evolutionary rates of coding sequences and regulatory regions (Jordan, Mariño-Ramírez, and Koonin 2005; Lemos et al. 2005; Liao and Zhang 2006b), including for *Caenorhabditis* (Castillo-Davis, Hartl, and Achaz 2004; Mark et al. 2019). Both coding sequences and gene expression are subject to purifying selection in *C. elegans* (Denver et al. 2005; Cutter, Dey, and Murray 2009), but future genome-scale analyses that couple ontogenetic transcriptome profiles with coding and regulatory sequence evolution are required to more fully determine the magnitude of inter-dependence of these modes of molecular evolution across development.

Verbal models suggest that genes will be more likely to diverge in coding sequence when they have fewer *cis*-regulatory elements, and when coding mutations are less likely to exert pleiotropic effects (Carroll 2008). The simple expression profiles of genes in the constitutive modules that we identified (M6, M12, M3, M13) versus the more complex dynamics of the modules that get up- or down-regulated across ontogeny implies the potential for simple versus complex *cis*-regulatory control. Indeed, for a given magnitude of expression ($\alpha$), genes in modules with simple profiles tend to show faster coding sequence evolution (e.g. M3 vs M7, M13 vs M1; Figure 4B), though genes with simple profiles often have high expression and slow evolution (e.g. M6). These observations set up a substrate for future analyses to test the hypothesis of a direct connection between *cis*-regulatory complexity and coding sequence evolution.

Evo-devo generally focuses on how the relative strength of constraint, which manifests as purifying selection and sequence conservation, could shape temporal ontogenetic patterns of evolution (Kalinka and Tomancak 2012). And yet, micro-evolutionary studies demonstrate that a majority of amino acid substitutions in protein coding sequence evolution may accumulate as a result of adaptive evolution in many animals, especially those with large effective population

sizes (Galtier 2016). Our analysis of divergence, however, cannot distinguish whether differences in coding sequence evolution among gene classes reflects differences in the strength of purifying selection versus the incidence of adaptive divergence. In *Drosophila*, rapidly-evolving proteins involved in chromatin regulation and genomic conflict are known to play important roles in creating post-zygotic reproductive barriers between species during early development (Presgraves 2010; Maheshwari and Barbash 2011; Cooper et al. 2018). Evolutionary conflict over allelic expression in early embryos can drive rapid sequence evolution (Haig 1997), and empirical observations demonstrate that positive selection is responsible for many coding sequence differences between species (Galtier 2016). Therefore, our findings raise the possibility that adaptive protein sequence divergence rather than weaker constraint might contribute importantly to ontogenetic patterns in the molecular evolution of development that lead to predictable developmental manifestations of hybrid dysfunction in the speciation process.

# Materials and Methods

### Expression data source and primary processing

We obtained RNAseq transcriptome sequences as sam-format files (mapped to *C. elegans* reference genome version WS248) from the public modENCODE data repository (http://data.modencode.org) for the *C. elegans* developmental time series for early embryos, each larval stage and young adult hermaphrodites (Supplementary Table S1) (Gerstein et al. 2010; Gerstein et al. 2014). We quantified expression for each gene using featureCounts (Liao, Smyth, and Shi 2014), based on exon annotations of WS248 (transposable element and pseudogene annotations were excluded; exons corresponding to all alternative splice forms of a given gene contributed to expression quantification for that gene). We then normalized expression counts following the log-counts per million method of (Law et al. 2014). Embryonic transcriptomes included a single biological replicate per timepoint whereas larval and young adult transcriptomes included duplicates; given the high correlation between duplicates ($r > 0.95$), we used the average log-normalized expression for each larval and adult timepoint for subsequent analyses. We restricted our analyses to those 19,711 genes with an expression level $\geq 1$ read count per million (cpm) in at least one timepoint (Robinson, McCarthy, and Smyth 2010). We

recalculated the log-cpm values for this set of 19,711 genes to account for the slight change in library sizes after the filtering step.

**Co-expression clustering and expression quantification of modules**

To uncover and identify distinct sets of gene expression patterns over time across the 19,711 genes in the *C. elegans* transcriptome (co-expression "modules"), we performed a functional principal components analysis (FPCA) (Madrigal, Dai, and Hadjipantelis 2018). FPCA is appropriate for longitudinal datasets (Yao, Müller, and Wang 2005), as for this transcriptome time series with just a single replicate per timepoint. First, we applied FPCA to the log-normalized gene expression data, using the "FPCA" function in the R package fdapace (Madrigal, Dai, and Hadjipantelis 2018), observing the first two components to cumulatively explain ~92% of the total variation. We then used each gene's FPC scores of the first two components as input for the clustering algorithm, implemented through the "FClust" function in R that uses a Gaussian Mixture Model approach based on EMCluster (W.C. Chen and R. Maitra, 2015, http://cran.r-project.org/package=EMCluster). We determined the optimal number of co-expression clusters or modules in our analysis to be $k = 14$, based on minimizing the Bayesian information criterion (BIC) value. We varied $k$ between 2 and 20 and observed minimum $\Delta$BIC = 11.4 occurring between $k = 12$ and $k = 14$. Visual inspection of expression trends affirmed the biological relevance of choosing $k = 14$ co-expression modules to represent the variation in expression profiles in the *C. elegans* transcriptome time series. Based on the outputs of the clustering algorithm, we assigned each gene to the module for which the gene has the highest membership probability.

To summarize quantitatively the dominant trends in expression over time for each co-expression module, we fit orthogonal cubic polynomial functions with time to log-normalized expression values, rescaled using the "poly_rescale" function in the polypoly R package (T. Mahr, 2017, https://cran.r-project.org/package=polypoly). To relate the co-expression modules to each other, we then performed hierarchical clustering on the module-wise cubic polynomial regression coefficients. The parameters extracted from the cubic fits summarize the overall expression level ($\alpha$), increasing or decreasing trends in expression across development ($\beta_1$), the degree of concave

versus convex expression dynamics over ontogeny ($\beta_2$), and how S-shaped are the expression dynamics ($\beta_3$). In order to obtain a finer-grained view of the temporal trends, we also performed a gene-level analysis, in which we fit an orthogonal cubic polynomial to each individual gene expression profile and extracted the corresponding parameters for analysis.

Finally, we classified genes according to expression pattern in the simplest of ways, by grouping genes according to which timepoint they showed peak expression across the time series.

## Enrichment analysis

To investigate trends of genomic organization for each co-expression module, we used contingency tables and $\chi^2$-test statistics to test for non-random distributions of genes for each of the 14 modules across each of the 6 chromosomes in the genome. To achieve this, we arranged the data in 84 individual two-way contingency tables, so that we could obtain $\chi^2$-test statistics on 1 degree of freedom to test for an association within each module-chromosome combination. We further investigated trends of genomic organization by looking within chromosomes, at enrichment within the arm and center regions of each chromosome, with arm vs. centre domains defined by recombination rate breakpoint positions given by (Rockman and Kruglyak 2009). MtDNA genes were excluded for these analyses, and p-values were adjusted for multiple testing using the Holm–Bonferroni method.

We conducted gene ontology (GO) and phenotype enrichment analysis (PEA) tests using the list of genes in each co-expression module as input into the WormBase Enrichment Analysis Suite (Angeles-Albores et al. 2016; Angeles-Albores et al. 2018), obtaining Benjamini-Hochberg false discovery rate corrected p-values (q-values) for statistical significance. By also cross-referencing genes with the analysis of (Tu et al. 2015), we used their determination of operon identity and calculations of coding sequence divergence between orthologs of *C. elegans* and *C. briggsae* to quantify molecular evolution of protein sequence as $K_A$, the rate of non-synonymous site substitution per non-synonymous site. Finally, we cross-referenced the genes in the transcriptome time series with those identified by (Ortiz et al. 2014) to have sex-neutral, oogenic

or spermatogenic enrichment of expression in their analysis of *C. elegans* transcriptomes from dissected gonads.

## Acknowledgements

We are grateful to the ModENCODE consortium for providing publicly available expression data for *C. elegans*. ADC is supported with funds from the Natural Sciences and Engineering Research Council (NSERC) of Canada; LS is supported with funds from NSERC and the Canadian Institutes of Health Research (CIHR). RG was supported by an NSERC Undergraduate Summer Research Award.

## Literature Cited

**Angeles-Albores D, Lee RYN, Chan J, Sternberg PW. 2018. Two new functions in the WormBase Enrichment Suite. *microPublication Biology*.https://doi.org/10.17912/W17925Q17912N.**

**Angeles-Albores D, Lee RYN, Chan J, Sternberg PW. 2016. Tissue enrichment analysis for *C. elegans* genomics. *BMC Bioinformatics.* 17:366.**

**Arthur W. 2015. Internal Factors in Evolution: The Morphogenetic Tree, Developmental Bias, and Some Thoughts on the Conceptual Structure of Evo-devo. Pp. 343-363 *in* Love CA, ed. Conceptual Change in Biology: Scientific and Philosophical Perspectives on Evolution and Development. Springer Netherlands, Dordrecht.**

**Artieri CG, Haerty W, Singh RS. 2009. Ontogeny and phylogeny: molecular signatures of selection, constraint, and temporal pleiotropy in the development of Drosophila. *BMC Biol.* 7.**

**Baird SE, Seibert SR. 2013. Reproductive isolation in the Elegans-Group of *Caenorhabditis*. *Nat Sci.* 5:18-25.**

**Baird SE, Sutherlin ME, Emmons SW. 1992. Reproductive isolation in Rhabditidae (Nematoda, Secernentea): mechanisms that isolate 6 species of 3 genera. *Evolution.* 46:585-594.**

**Barkoulas M, Vargas Velazquez AM, Peluffo AE, Felix MA. 2016. Evolution of new cis-regulatory motifs required for cell-specific gene expression in Caenorhabditis. *PLoS Genet.* 12:e1006278.**

**Barriere A, Gordon KL, Ruvinsky I. 2012. Coevolution within and between regulatory loci can preserve promoter function despite evolutionary rate acceleration. *PLoS Genet.* 8:e1002961.**

**Barrière A, Ruvinsky I. 2014. Pervasive divergence of transcriptional gene regulation in *Caenorhabditis* nematodes. *PLoS Genet.* 10:e1004435.**

**Batada NN, Hurst LD, Tyers M. 2006. Evolutionary and physiological importance of hub proteins. *PLoS Comp Biol.* 2:e88.**

**Bundus JD, Alaei R, Cutter AD. 2015. Gametic selection, developmental trajectories and extrinsic heterogeneity in Haldane's rule. *Evolution.* 69:2005-2017.**

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

**Carroll SB. 2005. Evolution at two levels: on genes and form. *Public Library of Science Biology.* 3:e245.**

**Castillo-Davis CI, Hartl DL. 2002. Genome evolution and developmental constraint in *Caenorhabditis elegans*. *Mol Biol Evol.* 19:728-735.**

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

**Charlesworth B. 1993. Evolutionary mechanisms of senescence. *Genetica.* 91:11-19.**

**Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. 2009. Coevolution of interacting fertilization proteins. *PLoS Genet.* 5:e1000570.**

**Cooper JC, Lukacs A, Reich S, Schauer T, Imhof A, Phadnis N. 2018. Altered chromatin localization of hybrid lethality proteins in Drosophila. *bioRxiv.*438432.**

**Cruickshank T, Wade MJ. 2008. Microevolutionary support for a developmental hourglass: gene expression patterns shape sequence variation and divergence in Drosophila. *Evolution & Development.* 10:583-590.**

**Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol.* 25:778-786.**

**Cutter AD, Dey A, Murray RL. 2009. Evolution of the *Caenorhabditis elegans* genome. *Mol Biol Evol.* 26:1199-1234.**

**Cutter AD, Ward S. 2005. Sexual and temporal dynamics of molecular evolution in *C. elegans* development. *Mol Biol Evol.* 22:178-188.**

**Davis JC, Brandman O, Petrov DA. 2005. Protein evolution in the context of Drosophila development. *J Mol Evol.* 60:774-785.**

**de Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. *Nat Rev Genet.* 14:249.**

**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.* 37:544-548.**

**Dey A, Jin Q, Chen Y-C, Cutter AD. 2014. Gonad morphogenesis defects drive hybrid male sterility in asymmetric hybrid breakdown of *Caenorhabditis* nematodes. *Evol Dev.* 16:362-372.**

**Domazet-Loso T, Tautz D. 2010. A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature.* 468:815-U107.**

**Drummond DA, Bloom JD, Adami C, Wilke CO, Arnold FH. 2005. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci USA.* 102:14338-14343.**

**Ellegren H, Parsch J. 2007. The evolution of sex-biased genes and sex-biased gene expression.** ***Nat Rev Genet.*** **8:689-698.**

**Farhadifar R, Baer Charles F, Valfort A-C, Andersen Erik C, Müller-Reichert T, Delattre M, Needleman Daniel J. 2015. Scaling, selection, and evolutionary dynamics of the mitotic spindle.** ***Curr Biol.*** **25:732-740.**

**Flatt T, Schmidt PS. 2009. Integrating evolutionary and molecular genetics of aging.** ***Biochimica et Biophysica Acta (BBA) - General Subjects.*** **1790:951-962.**

**Galtier N. 2016. Adaptive protein evolution in animals and the effective population size hypothesis.** ***PLoS Genet.*** **12:e1005774.**

**Gerstein MB, Lu ZJ, Van Nostrand EL et al. 2010. Integrative analysis of the** ***Caenorhabditis elegans*** **genome by the modENCODE project.** ***Science.*** **330:1775-1787.**

**Gerstein MB, Rozowsky J, Yan K-K et al. 2014. Comparative analysis of the transcriptome across distant species.** ***Nature.*** **512:445-448.**

**Haag ES, Chamberlin H, Coghlan A, Fitch DH, Peters AD, Schulenburg H. 2007. Caenorhabditis evolution: if they all look alike, you aren't looking hard enough.** ***Trends Genet.*** **23:101-104.**

**Haig D. 1997. Parental antagonism, relatedness asymmetries, and genomic imprinting.** ***Proc R Soc Lond B-Biol Sci.*** **264:1657-1662.**

**Haygood R, Babbitt CC, Fedrigo O, Wray GA. 2010. Contrasts between adaptive coding and noncoding changes during human evolution.** ***Proc Natl Acad Sci USA.*** **107:7853-7857.**

**Hazkani-Covo E, Wool D, Graur D. 2005. In search of the vertebrate phylotypic stage: A molecular examination of the developmental hourglass model and von Baer's third law.** ***Journal of Experimental Zoology Part B-Molecular and Developmental Evolution.*** **304B:150-158.**

**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.*** **29:3385-3395.**

**Hoekstra HE, Coyne JA. 2007. The locus of evolution: evo devo and the genetics of adaptation.** ***Evolution.*** **61:995-1016.**

**Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis.** ***Nat Comm.*** **2:248.**

**Irie N, Kuratani S. 2014. The developmental hourglass model: a predictor of the basic body plan?** ***Development.*** **141:4649-4655.**

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

**Jordan IK, Wolf YI, Koonin EV. 2003. No simple dependence between protein evolution rate and the number of protein-protein interactions: only the most prolific interactors tend to evolve slowly.** ***BMC Evol Biol.*** **3:1.**

**Kalinka AT, Tomancak P. 2012. The evolution of early animal embryos: conservation or divergence?** ***Trends Ecol Evol.*** **27:385-393.**

**Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, Jarrells J, Ohler U, Bergman CM, Tomancak P. 2010. Gene expression divergence recapitulates the developmental hourglass model.** ***Nature.*** **468:811-U102.**

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

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

**Lemos B, Bettencourt BR, Meiklejohn CD, Hartl DL. 2005. 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.** ***Mol Biol Evol.*** **22:1345-1354.**

**Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental milestones punctuate gene expression in the** ***Caenorhabditis*** **embryo.** ***Dev Cell.*** **22:1101-1108.**

**Liao B-Y, Zhang J. 2006a. Low rates of expression profile divergence in highly expressed genes and tissue-specific genes during mammalian evolution.** ***Mol Biol Evol.*** **23:1119-1128.**

**Liao B-Y, Zhang J. 2006b. Evolutionary conservation of expression profiles between human and mouse orthologous genes.** ***Mol Biol Evol.*** **23:530-540.**

**Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features.** ***Bioinformatics.*** **30:923-930.**

**Madrigal P, Dai X, Hadjipantelis PZ. 2018. Sparse functional data analysis accounts for missing information in single-cell epigenomics.** ***bioRxiv.*****504365.**

**Maheshwari S, Barbash DA. 2011. The genetics of hybrid incompatibilities.** ***Annu Rev Genet.*** **45:331-355.**

**Mark S, Weiss J, Sharma E, Liu T, Wang W, Claycomb JM, Cutter AD. 2019. Genetic and environmental perturbations drive modular transcriptome responses. Unpublished.**

**Medawar PB. 1952. An Unsolved Problem of Biology. H.K. Lewis & Co., London.**

**Memar N, Schiemann S, Hennig C, Findeis D, Conradt B, Schnabel R. 2018. Twenty million years of evolution: The embryogenesis of four Caenorhabditis species are indistinguishable despite extensive genome divergence.** ***Dev Biol.***

**Ortiz MA, Noble D, Sorokin EP, Kimble J. 2014. A new dataset of spermatogenic vs. oogenic transcriptomes in the nematode** ***Caenorhabditis elegans*****.** ***G3: Genes|Genomes|Genetics.*** **4:1765-1772.**

**Partridge L. 2001. Evolutionary theories of ageing applied to long-lived organisms.** ***Experimental Gerontology.*** **36:641-650.**

**Poe S, Wake Marvalee H. 2004. Quantitative tests of general models for the evolution of development.** ***Am Nat.*** **164:415-422.**

**Presgraves DC. 2010. The molecular evolutionary basis of species formation.** ***Nat Rev Genet.*** **11:175-180.**

**Promislow DEL, Tatar M. 1998. Mutation and senescence: where genetics and demography meet.** ***Genetica.*** **102/103:299-314.**

**Raff RA. 1996. The Shape of Life: Genes, Development, and the Evolution of Animal Form. University of Chicago Press, Chicago.**

**Reinke V, Cutter AD. 2009. Germline expression influences operon organization in the *Caenorhabditis elegans* genome. *Genetics.* 181:1219-1228.**

**Riche S, Zouak M, Argoul F, Arneodo A, Pecreaux J, Delattre M. 2013. Evolutionary comparisons reveal a positional switch for spindle pole oscillations in Caenorhabditis embryos. *J Cell Biol.* 201:653-662.**

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

**Rockman MV, Kruglyak L. 2009. Recombinational landscape and population genomics of *C. elegans*. *PLoS Genet.* 5:e1000419.**

**Roux J, Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. *PLoS Genet.* 4:e1000311.**

**Stern DL. 2000. Evolutionary developmental biology and the problem of variation. *Evolution.* 54:1079-1091.**

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

**Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nat Rev Genet.* 3:137-144.**

**True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evol Dev.* 3:109-119.**

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

**Valfort A-C, Launay C, Sémon M, Delattre M. 2018. Evolution of mitotic spindle behavior during the first asymmetric embryonic division of nematodes. *PLoS Biol.* 16:e2005099.**

**Verster AJ, Ramani AK, McKay SJ, Fraser AG. 2014. Comparative RNAi screens in *C. elegans* and *C. briggsae* reveal the impact of developmental system drift on gene function. *PLoS Genet.* 10:e1004077.**

**Wittkopp PJ, Kalay G. 2012. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet.* 13:59-69.**

**Wray GA. 2007. The evolutionary significance of cis-regulatory mutations. *Nat Rev Genet.* 8:206-216.**

**Yao F, Müller H-G, Wang J-L. 2005. Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association.* 100:577-590.**

**Zhao Z, Boyle TJ, Bao Z, Murray JI, Mericle B, Waterston RH. 2008. Comparative analysis of embryonic cell lineage between *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Dev Biol.* 314:93-99.**

# Figures

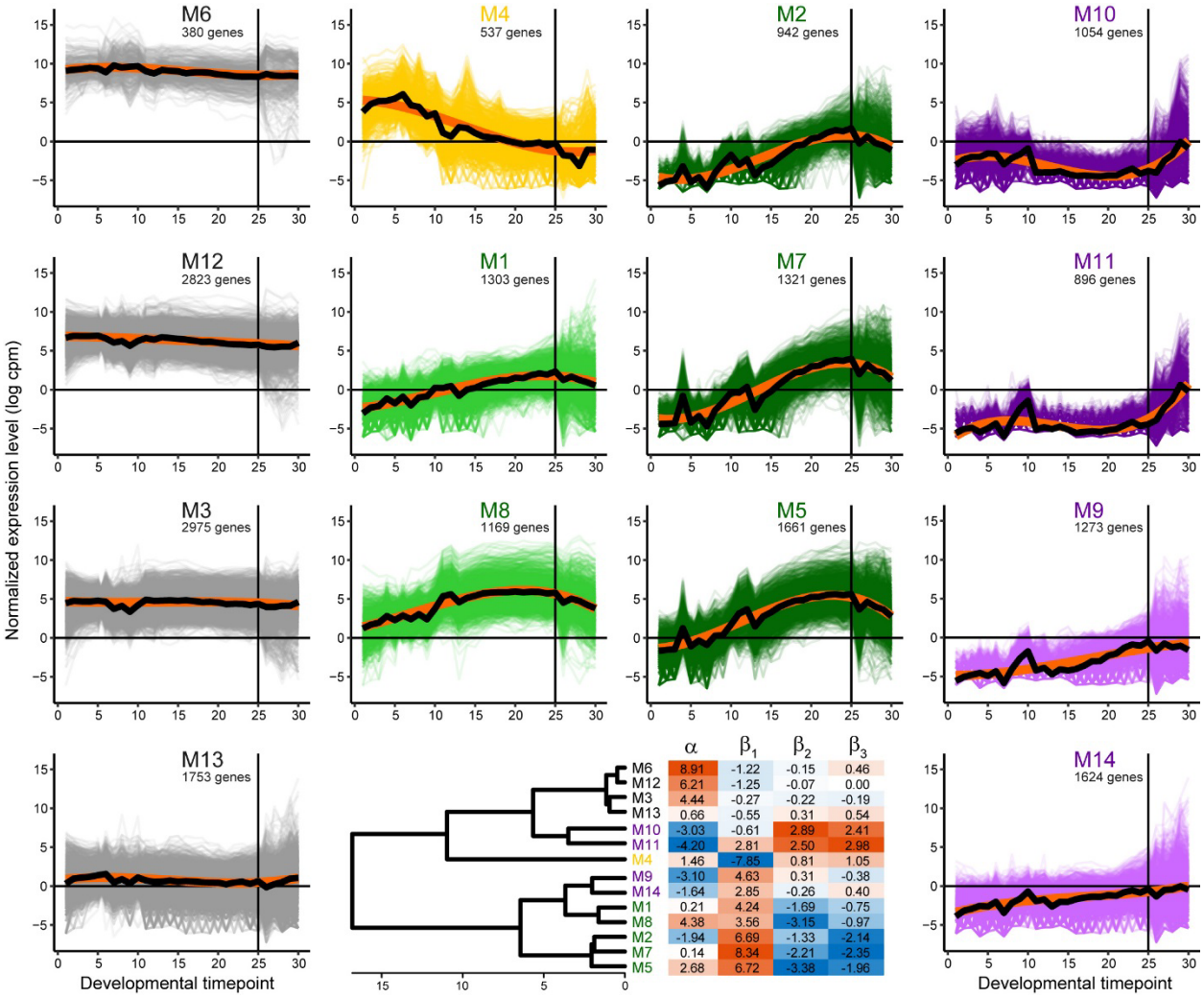


Figure 1. Ontogenetic time series of 19,711 *C. elegans* gene expression profiles clustered into 14 co-expression modules. Modules colored according to a trend of decreasing expression across development (yellow M4), peak expression in late embryogenesis (green M1, M8; M2, M7, M5), peak expression post-embryogenesis (purple M10, M11; M9, M14), or non-dynamic 'constitutive' expression across all 30 developmental timepoints (gray M6, M12, M3, M13). Thick black curves indicate expression trend across all genes in a module; thick orange curves indicate cubic polynomial fit to the expression trend. Similarity of module profiles indicated in dendrogram, with heatmap of parameter values from polynomial fit to each module expression trend ($\alpha$ = overall expression level, $\beta_1$ = linear change over time, $\beta_2$ = quadratic curvature, $\beta_3$ = cubic S-shape to expression profile over development). Vertical line at developmental timepoint

25 indicates the end of embryonic development, followed by 5 post-embryonic timepoints; embryonic timepoints taken at 30 minute intervals, with 1 timepoint for each larval stage L1-L4 and young adult (Supplementary Figure S1) (Gerstein et al. 2010; Gerstein et al. 2014).

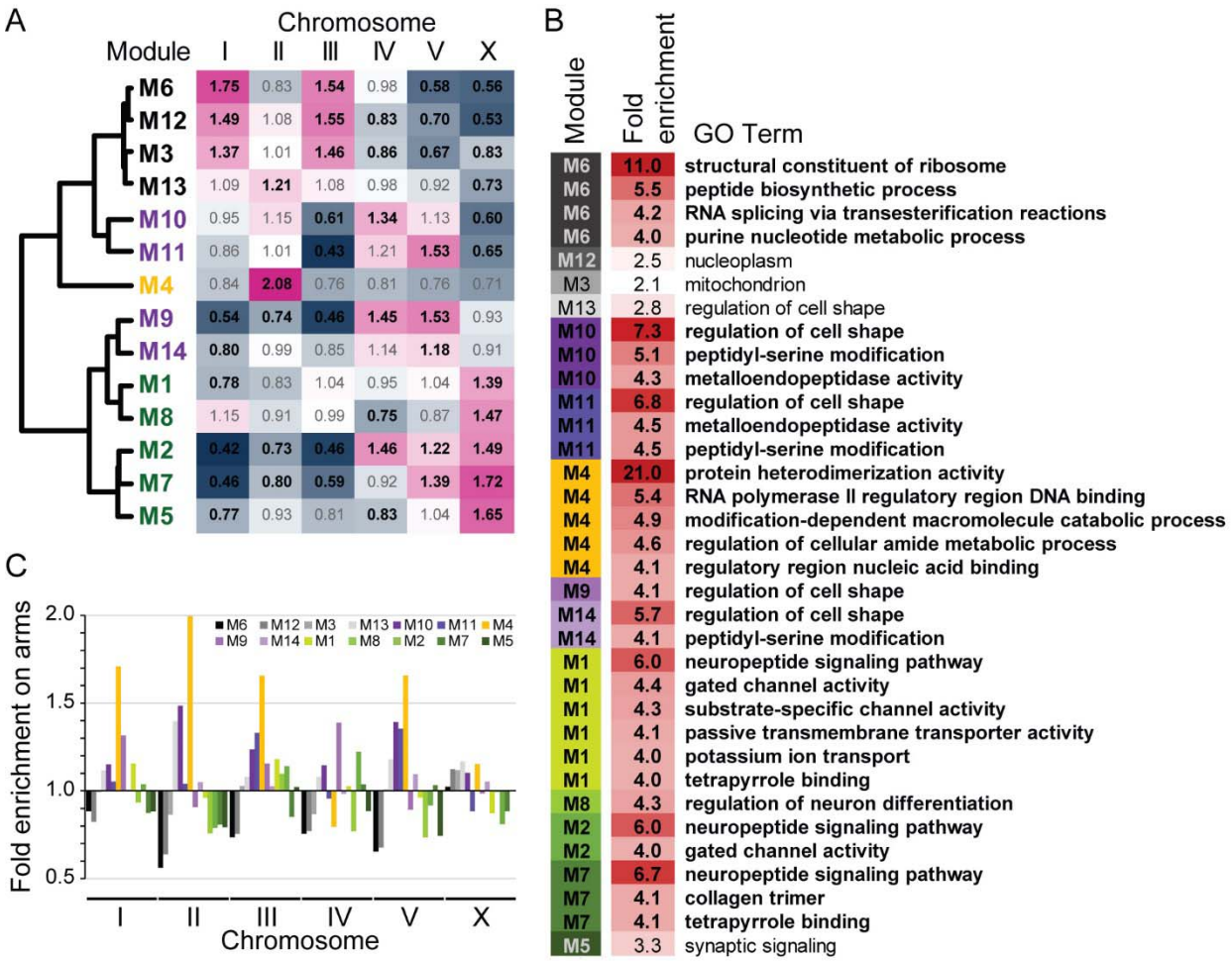


Figure 2. (A) Enrichment of gene membership among chromosomes for each co-expression module. Bold black text for observed/expected values in the heatmap indicates significant over- or under-enrichment (Holm-Bonferroni corrected p-values < 0.05,). (B) List of the 30 most enriched (>4-fold) gene ontology (GO) terms for each module, plus the single most enriched GO term observed for M3, M5, M12, M13 (all q-values < 0.005; 346 significantly enriched GO terms total across the 14 modules; Supplementary Table S1, Supplementary Table S2). (C) Enrichment of module gene membership on chromosome arms (values <1 imply enrichment in chromosome centers), where arm regions have higher recombination, higher density of repetitive elements, and lower gene density. Genome-wide significant enrichment on autosomal arms for M4, M10, and M13 and in centers for M5 and M12 (all Holm-Bonferroni corrected p-values < 0.003). Module identities colored and sorted by expression profile similarity as in Figure 1.

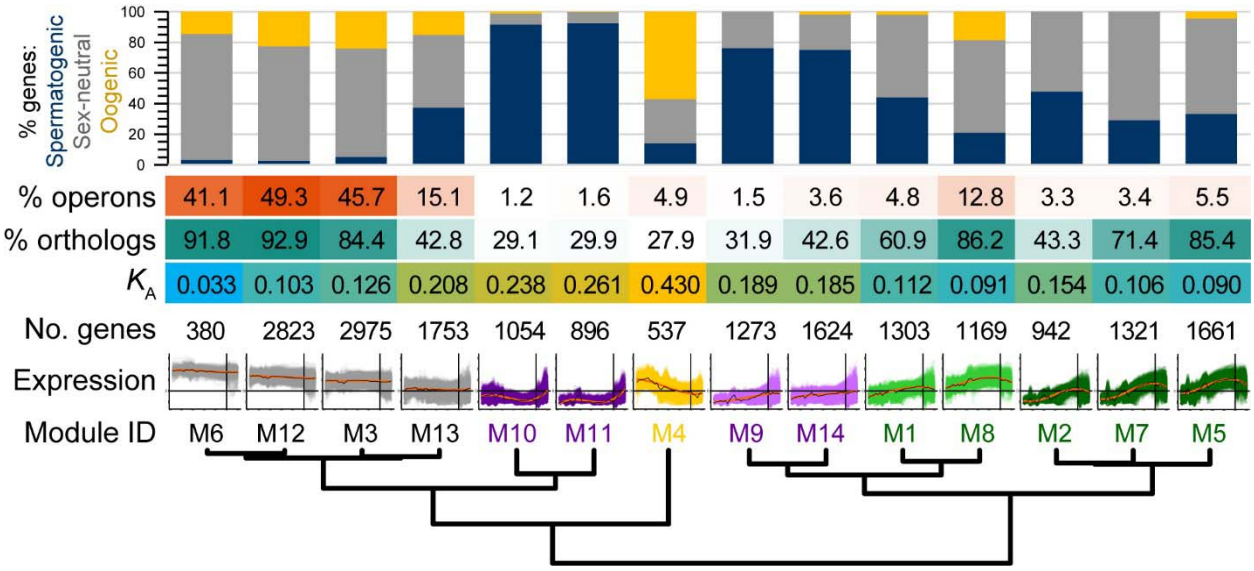


Figure 3. Functional and evolutionary properties of genes within each co-expression module. The proportion of genes with enrichment of spermatogenic, oogenic or sex-neutral expression categories defined by Ortiz et al. (2014), shown in the cumulative bar graph. Heat map shows the incidence of module genes in operons, the fraction of module members having orthologs in *C. briggsae*, and the median rate of non-synonymous site substitution ($K_A$) as a measure of protein sequence divergence. Module order sorted by expression profile similarity as in Figure 1.

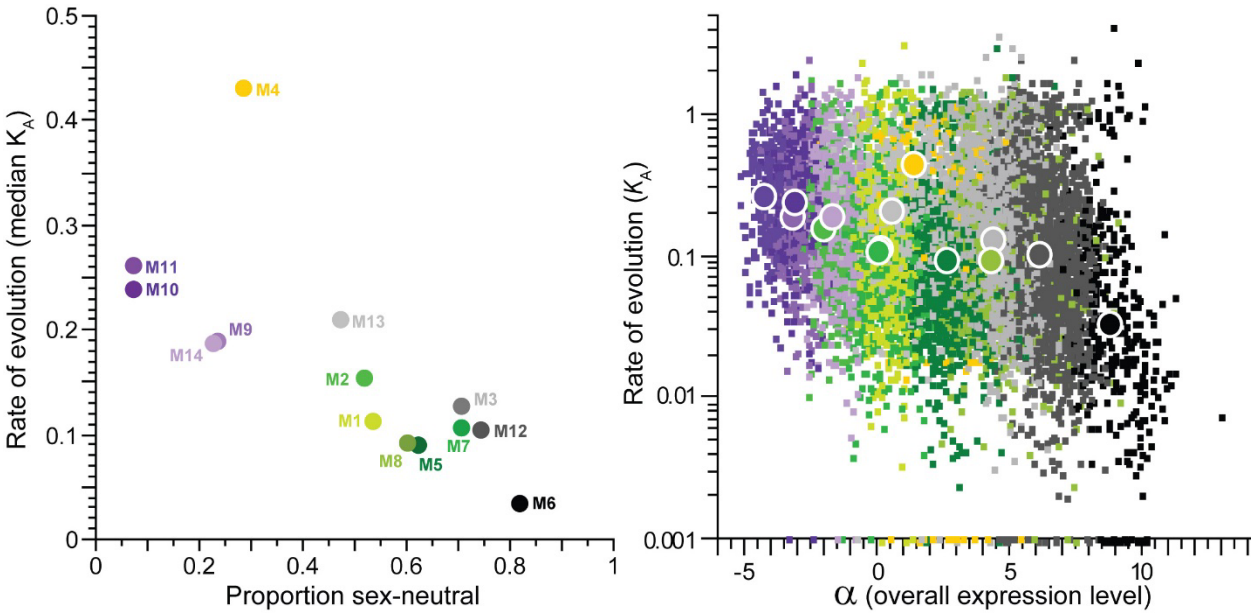


Figure 4. (A) Median rate of protein evolution (non-synonymous site substitution, $K_A$, between orthologs of *C. elegans* and *C. briggsae*) for genes within each co-expression module as a function of the proportion of module genes with the sex-neutral expression category, as defined by Ortiz et al. (2014). (B) Rates of protein evolution ($K_A$, log-scale) plotted as a function of the α parameter (overall expression level) from the polynomial fit to the expression time series. Per-gene values shown as small squares, module median values shown as large circles. Module membership color is the same in A and B.

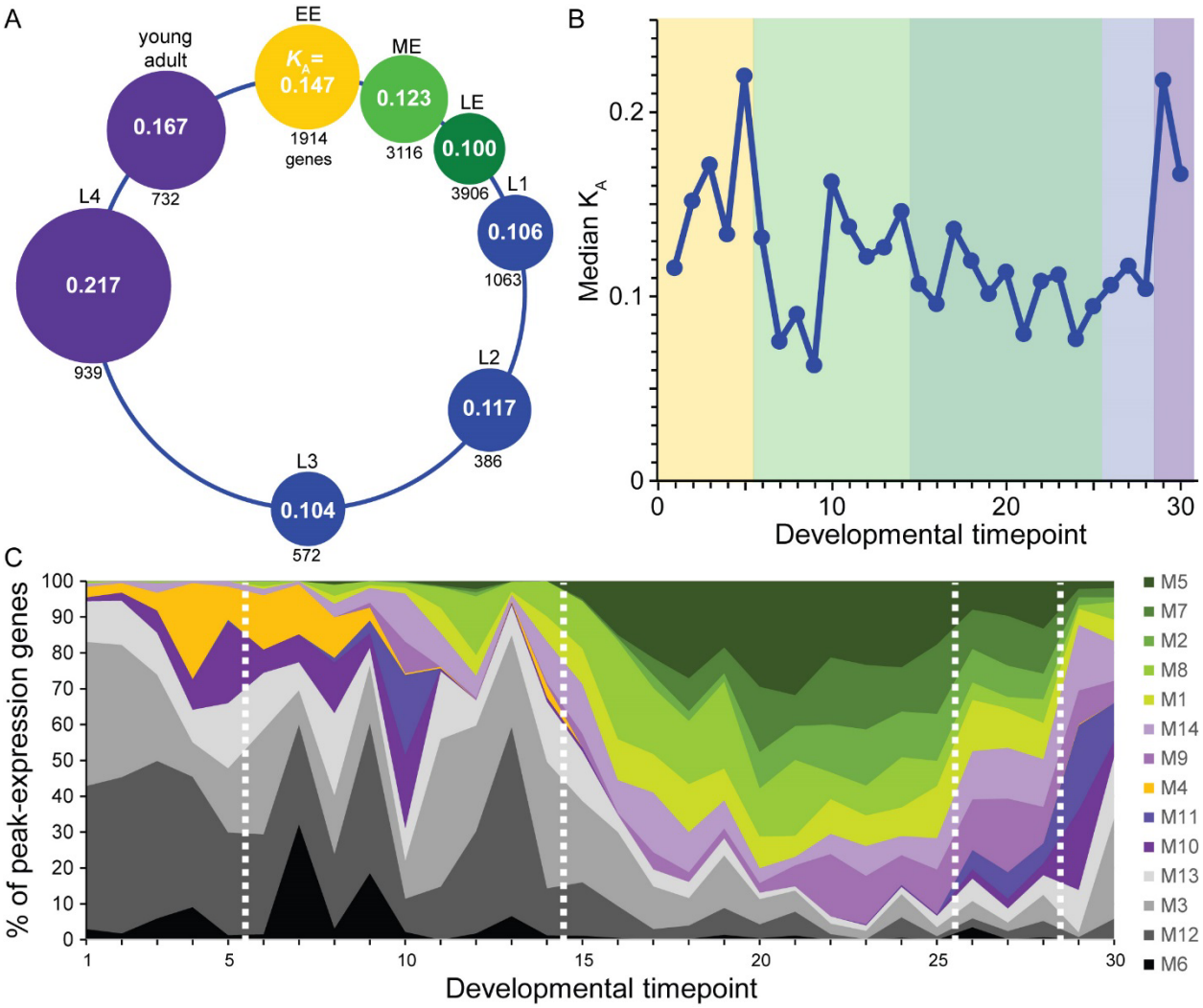


Figure 5. (A) Median rates of protein evolution ($K_A$) for genes with peak expression in different stages of the *C. elegans* life cycle (EE = early embryo, ME = mid-embryo, LE = late embryo, larval stages L1-L4, young adult); circle diameters proportional to value of $K_A$. (B) Median $K_A$ for genes with peak expression at each timepoint in the ontogenetic time series (shading: yellow = EE, light green = ME, dark green = LE, blue = larval L1-L3, purple = larval L4 to young adult). Timepoints in embryogenesis are spaced at 30 minute intervals (Gerstein et al. 2010; Gerstein et al. 2014). (C) Cumulative fraction of genes having peak expression at each timepoint that are members of the 14 co-expression modules. Module identities sorted by expression profile similarity as in Figure 1 and colored as in Figure 4; dashed vertical white lines demarcate the boundaries between EE, ME, LE, larval L1-L3, larval L4 to adult as in B.