Comparative Transcriptomics of *Steinernema* and *Caenorhabditis* Single Embryos Reveals Orthologous Gene Expression Convergence during Late Embryogenesis

Marissa Macchietto[1,2], Dristi Angdembey[2], Negar Heidarpour[2], Lorrayne Serra[2], Bryan Rodriguez[2], Nicole El-Ali[2], and Ali Mortazavi[1,2,*]

[1]Center for Complex Biological Systems, University of California Irvine

[2]Department of Developmental and Cell Biology, University of California Irvine

*Corresponding author: E-mail: ali.mortazavi@uci.edu.

Accepted: September 25, 2017

Data deposition: This project has been deposited at the Gene Expression Omnibus (GEO) under the accession GSE86381.

Abstract

Cells express distinct sets of genes in a precise spatio-temporal manner during embryonic development. There is a wealth of information on the deterministic embryonic development of *Caenorhabditis elegans*, but much less is known about embryonic development in nematodes from other taxa, especially at the molecular level. We are interested in insect pathogenic nematodes from the genus *Steinernema* as models of parasitism and symbiosis as well as a satellite model for evolution in comparison to *C. elegans*. To explore gene expression differences across taxa, we sequenced the transcriptomes of single embryos of two *Steinernema* species and two *Caenorhabditis* species at 11 stages during embryonic development and found several interesting features. Our findings show that zygotic transcription initiates at different developmental stages in each species, with the *Steinernema* species initiating transcription earlier than *Caenorhabditis*. We found that ortholog expression conservation during development is higher at the later embryonic stages than at the earlier ones. The surprisingly higher conservation of orthologous gene expression in later embryonic stages strongly suggests a funnel-shaped model of embryonic developmental gene expression divergence in nematodes. This work provides novel insight into embryonic development across distantly related nematode species and demonstrates that the mechanisms controlling early development are more diverse than previously thought at the transcriptional level.

Key words: RNA-seq, single-embryo, steinernema, caenorhabditis, comparative transcriptomics, embryonic development.

Introduction

Embryonic development in *Caenorhabditis elegans* is deterministic and is characterized by invariable cell lineages (Sulston et al. 1983). Studies have been done to perturb a large gamut of regulatory factors to uncover their roles in *C. elegans* lineage specification during embryonic development, and many factors have been well characterized and documented (Gerstein et al. 2010; Araya et al. 2014). However, far fewer molecular and genetic studies have been conducted on nematodes that are distantly related to *Caenorhabiditis*, and comparative developmental studies across nematodes have been based primarily on observations (Schierenberg 2006). These studies have noted and compared features of early divisions across nematodes, such as the synchronicity of the divisions, the sizes of cells produced from the divisions, the cell–cell interactions ("T" shape embryo vs. "I" shape embryo after removal of egg shell) after the divisions, the specification of the anterior and posterior axis, and when the timing of cell fate commitment occurs in them (Voronov and Panchin 1998; Goldstein et al. 1998; Schierenberg 2006). Many of these developmental features segregate based on their phylogeny. For example, clade 2 nematodes (Dorylaimia) in the phylogeny proposed by De Ley and Blaxter have synchronous cell divisions and produce cells of equivalent sizes that are unspecified, whereas clades 3–12 (Chromadorea) follow asynchronous divisions and produce cells of different sizes with determined cell fates (Voronov et al. 1998; De Ley and Blaxter 2002). Differences in the timing between developmental stages and the occurrence of certain developmental landmarks such as gastrulation spur questions about how

similar gene expression is at equivalent stages across diverse nematode species, such as whether different nematodes species express the same genes at the same stages of development, how conserved is the expression of orthologous genes during development, how much of the transcriptome changes from one stage to another in a species, and how much of gene expression similarity across species depends on absolute time versus morphological stage?

Molecular studies of comparative development in nematodes have focused primarily on the genus *Caenorhabditis*. A comparative study of embryonic developmental gene expression was conducted across five *Caenorhabditis* species in order to investigate the relationship between embryonic developmental morphology and gene expression in the genus (Levin et al. 2012) in order to determine whether there is a "phylotypic" stage or multiple phylotypic stages during embryonic development. The idea of a phylotypic stage dates back as early as 1828, and it is currently defined as a stage of development where morphological variation, and by extension, gene expression variation, across species is minimal (von Baer 1828; Kalinka et al. 2010). Levin and colleagues found that the time for each species to reach the same developmental stage (morphological stage) varied and found that the degree of transcriptome divergence between any two stages is dependent on time. If the timing between stages in one species took 3 h and the timing in another took 4 h, then the transcriptome should in theory be more divergent in the second species because the transcriptome has had more time to change in expression from the first state. They found that this generally occurred, except when two specific developmental stages were considered. Levin et al. found that at the 4th division of the AB lineage (~24-cell stage) and especially at the ventral enclosure stage (~421–560-cell stage), divergence in gene expression became independent of time suggesting that the evolutionary constraints at these stages are stronger than at other developmental stages. Crucial developmental regulators involved in muscle and neuron tissue differentiation, and proteins containing homeobox, immunoglobulin-like, SH3, PDZ, and PH (cell–cell signaling) domains were also enriched at the ventral enclosure stage, suggesting that this stage could be the "phylotypic" stage (Levin et al. 2012). While this study showed that time plays an important role in gene expression during development, it did not delve into the degree of ortholog expression conservation during development across the species. In addition, it also only compared closely related species that are all from the same genus. Given that nematodes are so diverse, we were interested in investigating how gene expression varies during development across species of more distant genera.

Although clade ten nematodes such as *Steinernema*, a genus of insect pathogenic nematodes, are thought to develop very similarly to clade 9 worms such as *C. elegans*, we found in a previous study that early mixed-stage (zygote to 24–44-cell) embryonic gene expression showed little conservation between *C. elegans* and *Steinernema* (De Ley and Blaxter 2002; Dillman et al. 2015). We were interested in whether these expression differences reflected variations in their modes of embryonic development. In order to answer this question, we produced a high-resolution RNA-seq time course of embryonic development in *Steinernema carpocapsae*, *Steinernema feltiae*, and *C. elegans* along with the more distantly related congener, *C. angaria* for which a genome has already been sequenced (Mortazavi et al. 2010) and which was not considered in the Levin et al. study (fig. 1*A*). In this study, we investigate 1) the degree of conservation of embryonic developmental ortholog expression between these genera and within each genus, 2) how the timing of embryogenesis varies across them, and 3) what pathways could be significantly different between them during embryogenesis.

Materials and Methods

Strains

S. carpocapsae (strain ALL) and *S. feltiae* (strain SN) were cultured and maintained according to Dillman et al. (2015). *C. elegans* (N2) were grown on Nematode Growth Media (NGM) plates seeded with OP50. *C. angaria* (PS1010) were grown on nutrient agar + 0.1% cholesterol plates seeded with OP50.

Caenorhabditis Nematode Culture and Embryo Isolation

Mixed-stage populations of *C. elegans* and *C. angaria* grown on OP50 plates were collected by adding ddH_2O to the agar plates and swirling to lift the nematodes off of the plates. The nematode suspensions were poured into 15 ml conical tubes, and repeated until plates were clean. The suspensions were spun down at 2,000 RPM for 1 min, and washed twice with ddH_2O. Nematode pellets were treated for 5 min in a 5 ml solution containing 1.25 ml fresh bleach, 2.25 ml 1 M NaOH, and 1.5 ml ddH_2O in 15 ml conicals with intermittent vortexing. After the 5-min incubation, the conical tubes were topped off with M9 buffer, spun at 2,000 RPM for 2 min, and embryo pellets were washed three times to remove traces of bleach solution.

Steinernema Nematode Culture and Embryo Isolation

Approximately 10,000 *S. carpocapsae* and *S. feltiae* IJs were seeded on lipid agar plates on top of lawns of *Xenorhabdus nematophila* and *Xenorhabdus bovienii*, respectively. Nematodes were grown at room temperature until gravid adults were present (3–4 days for *S. carpocapsae* and 2–3 days for *S. feltiae*), and adults were bleached to obtain embryos using the same protocol that was used for *C. elegans* above, except that the embryos were washed and collected in Ringer's solution instead of M9 buffer.

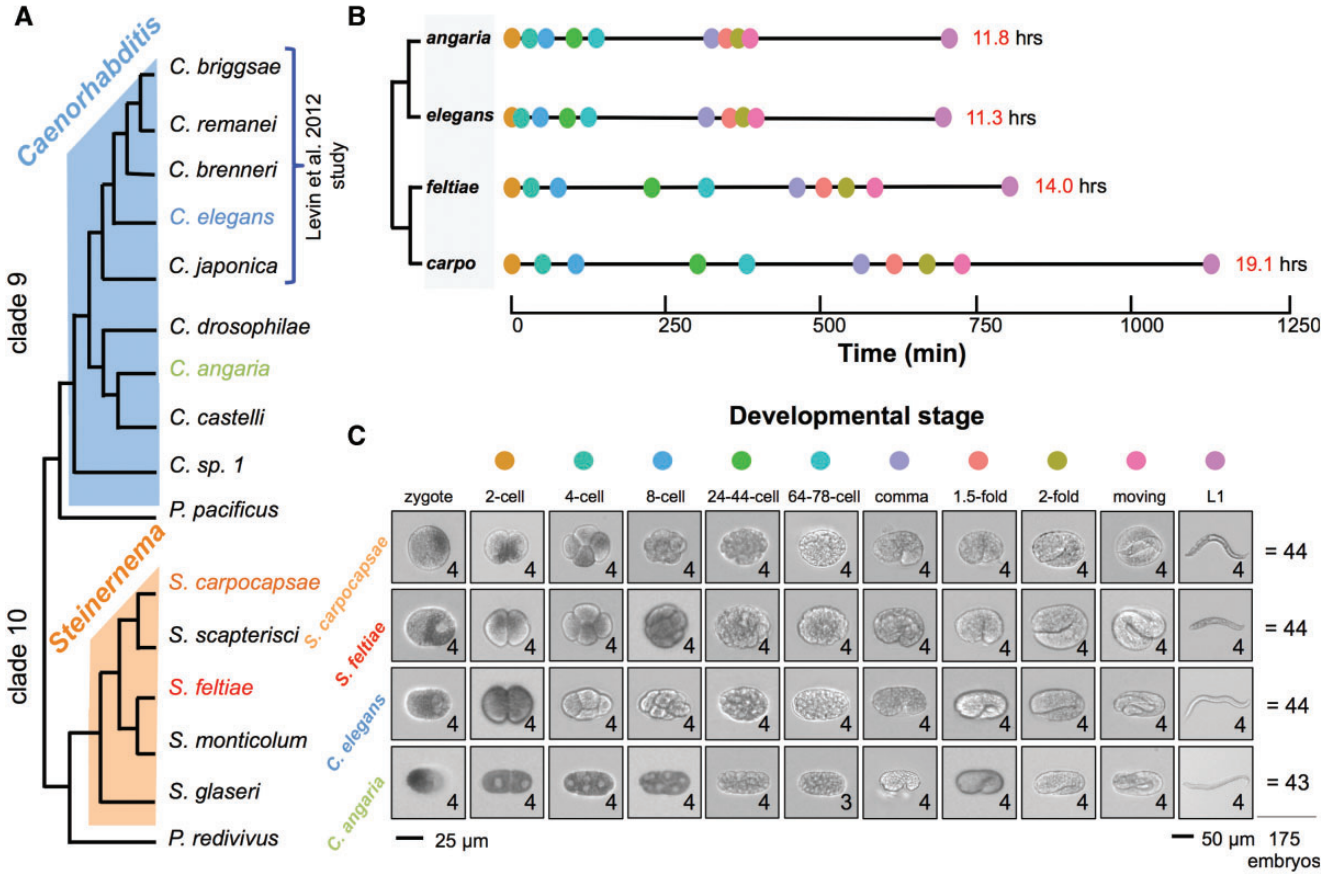

Fig. 1.—(*A*) Phylogenetic tree showing the relationships of the four nematodes in this study (*S. carpocapsae*, *S. feltiae*, *C. angaria*, and *C. elegans*). Several species from each genus and an outgroup species are included to highlight the evolutionary distances between the nematodes under investigation. Of note, the evolutionary distance between the *Caenorhabditids* in our study (*C. elegans* and *C. angaria*) is further than the distances between *C. elegans* and any of the four *Caenorhabditis* species chosen for the Levin et al. 2012 study. Branch lengths are not to scale. (*B*) Embryonic development was tracked using a time-lapse microscope for each species at 24 °C with representative images of each stage shown. The timeline shows the average timing between stages based on at least three embryos imaged for the transition between pairs of stages. Stage key is in 3C. (*C*) Images of the morphologies of 11 embryonic stages of two *Steinernema* and two *Caenorhabditis* species. Three to four embryos of each embryonic stage for each species were collected for single embryo RNA-sequencing with Smart-seq2. Embryos are on one scale (scale bar = 25 µm) and the L1s are on another (scale bar = 50 µm).

Embryonic Time Course

Embryos of *S. carpocapsae*, *S. feltiae*, *C. elegans*, and *C. angaria* were imaged every 5 or 10 min for 24 h at 24 °C on the EVOS inverted microscope (fig. 1*B*). *S. carpocapsae* and *S. feltiae* embryos were imaged in Ringer's solution, whereas *C. elegans* and *C. angaria* were imaged in M9 buffer. Time data for each stage transition was collected for at least three embryos. The average number of embryos collected per stage is ten embryos. Developmental timeline was made using the timeline library in R version 3.2.3 (Bryer 2013).

Experimental Design

We collected and sequenced single embryos at 11 embryonic stages per species (*S. carpocapsae*, *S. feltiae*, *C. elegans*, and *C. angaria*) in quadruplicates (fig. 1*C*). We amplified the very low quantities of mRNA from each of these individual embryos into cDNA by following Smart-seq2 protocol with minor modifications detailed below (Picelli et al. 2014). We sequenced a total of 175 single embryos; each was sequenced an average depth of 10 million reads.

Embryo Collection for Smart-Seq2

Pellets of embryos were resuspended in 2 ml of Ringer's solution (made with DEPC water) + 0.01% tween 20. DEPC was used in the Ringer's solution to limit RNase contamination, and tween 20 was used to prevent embryos from sticking to any surfaces. Resuspended embryos were passed through at 40 µm mesh filter into a 60 mm x 15 mm petri dish to remove debris. Enough Ringer's solution + 0.01% tween 20 was added to coat the bottom of the petri dish and reduce the density of the embryos so that they could easily be collected with a pipette. Embryos were visualized in the dish using an

EVOS inverted microscope, and single embryos were imaged and collected in 1.5 µl using a micropipette. If more than one embryo was collected, embryos were diluted further by pipetting them into 20 µl Ringer's solution + 0.01% tween 20 on a clean slide that was pretreated with RNase ZAP or 100% ethanol. Single embryos were collected in 1.5 µl into PCR tube strip, and 2 µl of lysis buffer (18 µl 0.3% Triton-X 100 + 2 µl RNase inhibitor SIGMA), 1 µl of oligo-dT primer, and 1 µl of dNTP mix were added to each embryo. Embryos were heated to reverse secondary structure of RNA, reverse transcribed and PCR amplified according to the Smart-seq2 protocol by Picelli (Picelli et al. 2014). All embryos, regardless of embryonic stage, were amplified for 18 cycles through PCR. PCR primers were cleaned up from the embryo samples by adding a 1:1 ratio of Ampure XP beads to sample, which were both equilibrated to room temperature, incubated for 8 min, placed on a magnet, and washed with 200 µl of 80% ethanol three times. Beads were dried at room temperature for approximately 5 min (until the beads cracked), after which, 17.5 µl of EB was added and incubated off the magnet for 3 min. Samples were placed back on the magnet, and 15 µl of cDNA was collected for each sample. Sample cDNA concentration was quantified using the Qubit fluorometer and bioanalyzed using the Agilent 2100 Bioanalyzer to check the cDNA quality.

Single Embryo Library Preparation and Sequencing

For library preparation, 20 ng of cDNA from each sample was prepared using the regular Nextera tagmentation protocol (Gertz et al. 2012). The protocol reagents were scaled down, so that 2 µl of transposase, 10 µl of buffer, and 8 µl of cDNA (20 ng total) were used yielding a total volume of 20 µl. Transposase was cleaned up from the tagmented DNA using the QIAGEN column following the manufacturer's instructions.

In a PCR tube, 30 µl of sample, 35 µl of Phusion high fidelity master mix, 2.5 µl and 25 µM Nextera adapter ID, and 2.5 µl and 25 µM Nextera adapter Ad_noMX were combined and mixed well with a pipette. Samples were spun down quickly, and amplified for six cycles using the PCR program with the following settings: 72 °C for 5 min, 98 °C for 30 s, (98 °C for 10 s, 63 °C for 30 s, 72 °C for 1 min) for six cycles, 72 °C for 5 min, and hold at 4 °C.

PCR amplified libraries were cleaned up using a 1:1 ratio of Ampure XP beads to sample, and prepared in the same way as the bead cleanup above, except that 30 µl of EB was added to the beads to resuspend the library sample, which was then collected in 27.5 µl after 2 min.

Sample library fragments were between 200 and 600 bps with an average size of 360 bps after the Nextera tagmentation protocol. Samples were sequenced as paired-end 43 bp on the Illumina NextSeq 500 to an average depth of ~10 million reads. This project has been deposited at the Gene Expression Omnibus (GEO) under the accession GSE86381.

Gene Expression Analyses

Unstranded, paired-end RNA-seq reads for all species were trimmed to 40 bp from their 3′ ends to remove low quality nucleotide sequences. Transcriptome indexes were prepared for *S. carpocapsae* (PRJNA202318 downloaded from WormBase ParaSite version WS254), *S. feltiae* (PRJNA204661 downloaded from WormBase ParaSite version WS254), *C. elegans* (WS220), and *C. angaria* (PRJNA51225 downloaded from WormBase ParaSite W254) using the RSEM command (version 1.2.12) rsem-prepare-reference (Li et al. 2011). Reads were mapped to each respective species' annotations using bowtie 0.12.8 with the following options: –S, –offrate 1, –v 1, –k 10, –best, –strata, –m 10 (Langmead et al. 2009). Gene expression was quantified using the RSEM command, rsem-calculate-expression, with the following options: –bam, –fragment-length-mean (Li et al. 2011). For all analyses, gene expression was reported in Transcripts Per Million (TPM).

Orthology Relationships Analysis

Orthologs and paralogs were determined across the four species by blasting their protein sequences to each other and to seven additional species using OrthoMCL 1.4 with the default settings (Li and Dewey 2003). Protein sequences were downloaded from WormBase ParaSite for *S. carpocapsae* (PRJNA202318. WBPS8), *S. feltiae* (PRJNA204661.WBPS8), *S. glaseri* (PRJNA204943.WBPS8), *S. monticolum* (PRJNA205067. WBPS8), *S. scapterisci* (PRJNA204942.WBPS8), *C. elegans* (PRJNA13758), *C. remanei* (PRJNA53967), *C. japonica* (PRJNA12591), *C. briggsae* (PRJNA10731), *C. angaria* (PRJNA51225), and *H. bacteriophora* (PRJNA13977.WBPS8), an entomopathogenic nematode from a different genus. Protein sequences were filtered to retain only the longest proteins corresponding to each gene sequence. In table 1, we report 1) the number of annotated protein coding genes in each genome, 2) the number of genes in each species that are within genus-specific orthology clusters (*N*:*N*:0:0 and 0:0:*N*:*N*), 3) the number of genes that are conserved across the genera that have multiple-to-multiple, or multiple-to-one gene relationships (*N*:*N*:*N*:*N*, excluding 1:1:1:1s) across species, where $N \geq 1$, 4) the number of genes that exist as a single copy (1:1:1:1) across the four species, 5) the number of genes that are conserved across genera that have multiple-to-one relationships (*N*:*N*:1:*N*), where there is only a single ortholog in *C. elegans* and $N \geq 1$ orthologs in the other species (excluding 1:1:1:1s), 6) the number of genes that are conserved between two or three species in a combination that is not covered by 2–5, 7) the number of genes whose proteins clustered in an orthology cluster with genes from one or more of the species in this

Table 1
OrthoMCL Gene Orthology Relationships (Further Description in Materials and Methods)

		Steinernema		*Caenorhabditis*	
		carpocapsae	*feltiae*	*elegans*	*angaria*
1	# Annotated genes in genome	28,313	33,459	20,389	27,970
2	# Conserved genes in genus (*N*:*N*:0:0, 0:0:*N*:*N*; where $N \geq 1$)	5,272	6,674	2,862	3,125
3	# Conserved genes across genera—*N*:*N*:*N*:*N* (where $N \geq 1$ and N does not have to equal *N*, excludes 1:1:1:1 orthologs)	3,271	3,334	3,017	3,593
4	# Conserved genes across all four species—1:1:1:1 orthologs	4,164	4,164	4,164	4,164
5	# conserved genes across genera—*N*:*N*:1:*N* (where $N \geq 1$ and *C. elegans* = 1, excludes 1:1:1:1 orthologs)	2,141	2,252	1,501	2,526
6	# Genes conserved in two or three of the species (all combinations that are not in lines 2–5)	1,585	1,856	1,529	1,056
7	# Clustered genes (in OrthoMCL orthology clusters - sum of lines 2, 3, 4, 5, and 6)	19,493	23,203	17,877	18,123
8	# of species-specific genes (in paralogy cluster (*N*:0:0:0) or in orthology cluster with other species (*N*:0:0:0)	5,201	7,175	6,305	6,185
9	# of species-specific genes (does not cluster with any other species or own genes)	8,820	10,256	2,512	9,847

study (sum of 2–6), 8) the number of genes that are species-specific genes in paralogy clusters, and 9) the number of species-specific genes that did not cluster at all in OrthoMCL ((1) − ((2) + (7)) = (8)). Manual annotation of orthologs and paralogs of select genes for analyses (oma gene) was done using WormBase ParaSite. In order to determine the robustness of the PCA results to ortholog detection method, we also obtained 1:1:1:1 orthologs from WormBase ParaSite and determined orthology relationships with OMA using the default settings and only the protein sequences of the four species in the manuscript (Altenhoff et al. 2015).

Differential Expression Analyses

Differential gene expression was determined using the Bioconductor package, edgeR v.3.2.4 (Robinson et al. 2010). The RSEM count data was used for calculating differential expression, and genes were called as differentially expressed if they had an FDR < 0.05 and a fold change $> 2\times$. Four replicates were used per stage for the analysis, except for the 64–78-cell stage RNA-seq data for *C. angaria*, which had three replicates. Early adjacent stages were pair-wise compared with detect the onset of the maternal to zygotic transcription (fig. 4 and supplementary fig. 12, Supplementary Material online).

Correlation Matrices

A pseudocount of 1 TPM was added to the gene expression of each gene for all the single embryos of each species and $\log_2$ scaled. Pearson's correlation coefficient (ρ) was determined from the data using the corr() function in R version 3.2.3 (R Development Core Team 2008).

Heat Maps

Heat maps of gene expression were mean-centered, normalized, and hierarchically clustered with Cluster 3.0 and visualized using Java Treeview (de Hoon et al. 2004, Saldanha 2004).

Differential Temporal Dynamics during Development with maSigPro

9,844 1:1 orthologs shared between *S. carpocapsae* and *S. feltiae* and 6,840 1:1 orthologs shared between *C. elegans* and *C. angaria* were run through maSigPro (Nueda et al. 2014) as multiple time series using *S. carpocapsae*'s and *C. elegans*' time course data, respectively (supplementary fig. 13, Supplementary Material online). A pseudo count of 1 was added to each gene for each sample, and the gene counts were normalized in edgeR using calcNormFactors() and cpm(). maSigPro was run with counts = TRUE setting for count-based expression. Significance threshold (*P* value) was adjusted to 0.01. Significant genes were clustered into nine expression profiles for each species.

Determining the Degree of Haplotype Contamination in the *S. carpocapsae* and *S. feltiae* Genomes Using HaploMerger2

The soft-masked *S. carpocapsae* and *S. feltiae* genomes (downloaded from WormBase ParaSite) were cleaned using the faDnaPolishing.pl script provided by HaploMerger2 (Huang et al. 2017). To generate score matrices, 5% of the genomic sequence (longest three scaffolds) was used as a target and the remaining 95% was used as a query for alignment with Lastz. Two alignment identity thresholds were tested for each species: 90% identity (excluding Ns, indels,

and gaps), which is equivalent to 4–5% allelic difference (heterozygosity), and 95% identity, which is equivalent to 2% allelic difference. We found that very little sequence (~200 kb, 0.2% of genome) was lost through collapsing the genomic sequence on the haplotypes for *S. carpocapsae*, indicating that the inflated gene annotations are not due to additional haplotypes. There was more sequence loss in the *S. feltiae* genomes, which lost 2.6 Mb.

Results

Embryonic Developmental Timing Varies across Nematodes

The Levin et al. study showed that within *Caenorhabditis*, the time for a 4-cell embryo to reach the first larval stage could take between 800 min (*C. elegans* or *C. briggsae*) and 1,300 min (*C. brenneri*) (Levin et al. 2012). We compared the timing of development between *Steinernema* and *Caenorhabditis*. We imaged the embryonic development of *S. carpocapsae*, *S. feltiae*, *C. elegans*, and *C. angaria* at 24 °C, and found that *Steinernema* species take much longer than the *Caenorhabditis* species to develop from the 2-cell stage to the L1 stage (fig. 1*B*, supplementary fig. 1, Supplementary Material online). This increase in developmental time is largely due to delayed early cleavage divisions in *Steinernema*, as the time windows are much larger between these stages in *Steinernema*. Specifically, the timing between the 4-cell to 8-cell and 8-cell to 24–44-cell stage is approximately 50% longer in *S. carpocapsae* and *S. feltiae* than it is in *Caenorhabditis*. Later embryonic developmental stages show overall less variation than between zygote and 24–44 cells.

Caenorhabditis Embryonic Transcriptomes Are More Highly Correlated with Each Other during Early Development (Zygote to 8-Cell) than *Steinernema* Embryonic Transcriptomes

In order to explore the transcriptomic changes that are occurring during embryogenesis across species, we sequenced the mRNA from single embryos spanning 11 developmental stages (zygote, 2-cell, 4-cell, 8-cell, 24–44-cell, 64–78-cell, comma, 1.5-fold, 2-fold, moving, and L1) for *S. carpocapsae*, *S. feltiae*, *C. elegans*, and *C. angaria* in quadruplicates using Smart-seq2 (fig. 1*C*). Our ultimate goal is to determine whether ortholog expression patterns are conserved over the course of embryogenesis in order to gain insight into the degree of conservation during development between *Steinernema* and *Caenorhabditis* at the transcriptional level.

Since the time between early embryonic stages is longer in *Steinernema* species, we postulated that the gene expression between pairs of early embryonic stages is potentially more divergent (less correlated) in *Steinernema* when compared with *Caenorhabditis* (Levin et al. 2012). To verify this, we calculated the Pearson's correlation coefficient between all pairs of single embryo transcriptomes for each species (fig. 2). We confirmed that 1) replicate embryo transcriptomes were highly correlated with each other, and 2) there was no contamination from embryos of other stages due to sample swaps. We found, as expected, that embryos that are more distant in time showed lower correlations than embryos that are closer temporally in all four species (fig. 2*A–D*). However, the degree of correlation between corresponding adjacent embryonic stages showed marked differences between the two genera with *Caenorhabditis* species showing higher correlation between adjacent early embryonic stages than *Steinernema* species (fig. 2). Early stage (zygote to 4-cell) *Steinernema* embryo replicates (fig. 2*A* and *B*) have lower correlations with each other (between 0.8 and 0.9) compared with early embryos in *Caenorhabditis* (0.9–1) (fig. 2*C* and *D*). In terms of the overall structure of the correlation matrices, we found similar structures between species of each genus, in contrast to the different structures observed across genera. Interestingly, the *Steinernema* correlation matrices show a decrease in correlation between 2-cell and 4-cell, and then a drastic decrease in transcriptome correlation (from >0.9 to <0.6) between 4-cell and 8-cell embryos (fig. 2*A* and *B*). Maternal transcript degradation in *C. elegans* occurs between the 4-cell and 8-cell stage (Edgar et al. 1994; Baugh et al. 2003). Thus, this substantial change in transcriptomes could reflect the earlier onset of maternal transcript degradation in *Steinernema*. These stage-to-stage transcriptome changes were less pronounced in *Caenorhabditis* because most of the early embryonic stages (from the zygote to the 4-cell) correlated so highly with each other that the stages could not be differentiated from each other globally (fig. 2*C* and *D*). Because the global gene expression of the zygote and 2-cell are representative of the maternal transcriptome, comparing the correlations between these stages with more distant neighboring stages could aid in determining when zygotic transcriptional changes commence in *Caenorhabditis*. With this strategy, a slight drop in correlation can be detected at the 8-cell stage in *C. elegans* (fig. 2*C*) and at the 24–44-cell in *C. angaria* (fig. 2*D*). This suggests that the transcriptional landscape of *Steinernema* is changing faster than *Caenorhabditis* in the early embryo and that the onset of maternal transcript degradation is occurring at a later stage in *Caenorhabditis angaria* compared with *C. elegans*. Thus, we observe both a set of within-genus differences as well as more dramatic differences between genera at the earliest embryonic stages.

Maternal *oma-1/2* Transcript Degradation Occurs at Different Stages across Species

In *C. elegans* embryos, the degradation of the maternally deposited proteins and transcripts *oma-1* and *oma-2* are crucial for the activation of zygotic gene expression

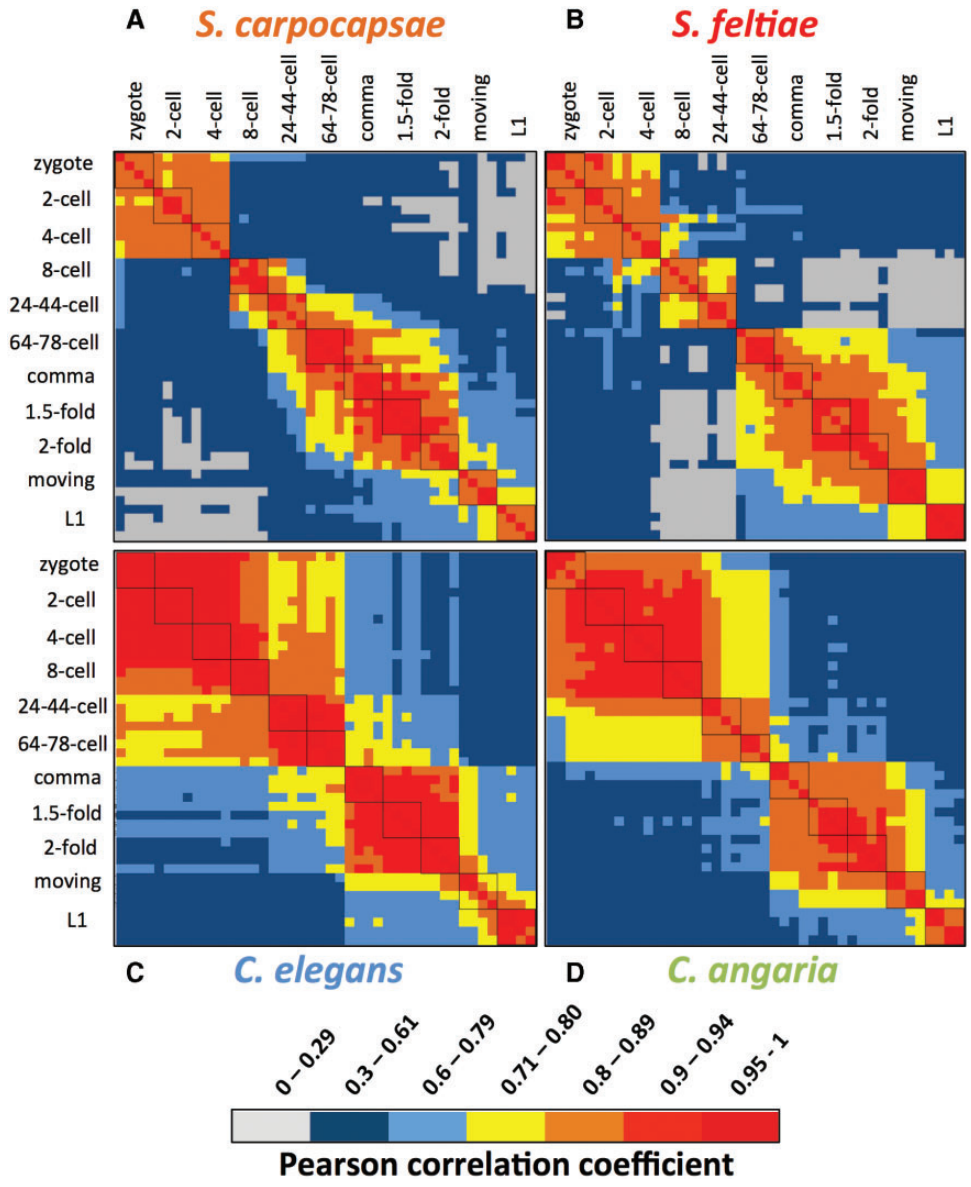

Fig. 2.—Transcriptome correlations (Pearson's correlation coefficients) across single embryos for each species for all annotated genes. Heatmaps showing the correlation coefficients of all single embryo comparisons for (*A*) *S. carpocapsae*, (*B*) *S. feltiae*, (*C*) *C. elegans*, and (*D*) *C. angaria*. Four replicate embryos are shown per developmental stage, except for the 64–78-cell stage in *C. angaria*, which has three replicate embryos. Red indicates almost perfect correlations (0.9–1), whereas grey indicates little to no correlation (0–0.29).

(Stitzel et al. 2006; Tadros and Lipshitz 2009). We explored whether the embryonic stages at which we detect the first upregulation of zygotic gene expression across all four species coincide with downregulation/degradation of maternal *oma-1/2* transcripts (supplementary fig. 2, Supplementary Material online). We investigated the orthology and expression of the *oma-1/2* gene across the four species and found that *C. elegans* underwent a triplication of an ancestral *oma* gene to produce *oma-1*, *oma-2*, and *moe-3*. Both *oma-1* and *oma-2* transcripts are highly expressed in the *C. elegans* zygote, but we found that *oma-1* is downregulated one stage earlier (8-cell vs. 24–44-cell) than *oma-2* (supplementary fig. 2, Supplementary Material online). Although *C. elegans* has three *oma* genes involved in oocyte maturation, we found that all of the other species have only a single copy of the *oma* gene that shares homology with these *C. elegans* genes, which suggests that the molecular effectors of zygotic transcription initiation are varied across species and has implications for how the early transcriptional and posttranscriptional programs are carried out.

Focusing on the gene expression dynamics of these closely related *oma* genes, we find that the *oma-1/2* transcripts in *S. carpocapsae*, *S. feltiae*, and *C. angaria* are downregulated by the 2-cell, 8-cell, and 24–44-cell stage, respectively (supplementary fig. 2, Supplementary Material online). We further

found that more distant paralogs of the *oma* genes in all four species (*pos-1, mex-3, mex-5, mex-6, ccch-1, ccch-2, ccch-5, Y11A8C.20, dcf-13, C35D6.4, F38C2.7, Y60A9.3, Y116A8C.19*) are also strictly maternally expressed (supplementary fig. *3A* and *B*, Supplementary Material online). There are fewer *oma* paralogs in the *Steinernema* species and *C. angaria* (8 in *S. carpocapsae*, 7 in *S. feltiae*, and 5 in *C. angaria*) than in *C. elegans* (16 in *C. elegans*), indicating that these paralogs in other species may combine the roles of more than one paralog in *C. elegans*. Although we find evidence of degradation of the *oma-1/2* transcripts earlier in *Steinernema*, we lack data on when the OMA-1/2 proteins are degraded to establish whether *oma-1/2* transcript degradation is responsible for the earlier upregulation of genes that we observe.

Genus-Specific Trajectories during Embryonic Development

To assess how gene expression of single embryos varies across species during embryonic development, we performed Principal Component Analysis (PCA) on all of the single embryos (175) from all four species for a set of 4,164 1:1:1:1 orthologous genes (fig. 3*A* and table 1). We found that Principle Component 1 (PC1), which accounts for 21.9% of the variance across the single embryos, separated the embryos based on developmental time (early embryos vs. intermediate embryos vs. late embryos). We found that PC3 (8.9%) separated embryos by genus, and PC4 (6.9%) separated *C. elegans* and *C. angaria* embryos, but not *Steinernema* embryos (fig. 3*B* and supplementary fig. *4A*, Supplementary Material online). We tracked the developmental trajectories of each species on a plot of PC1 versus PC2, and found a clear distinction between the early embryos (from the zygote to 24–44-cell stage) of *Steinernema* and *Caenorhabditis* along PC2, but observed a convergence in later embryos from the 64–78-cell stage to the L1 stage (fig. 3*A*). PC1 corresponds to developmental time and PC2 corresponds to early differences in development between genera. We took the top 476 genes with positive PC2 loadings (loadings >0.025) and the top 207 genes with negative PC2 loadings (loadings <-0.01) and ran them through a gene ontology analysis separately. We found that PC2 positive loading genes were enriched for GO terms related to nematode larval development (P value = 2.2e-13, count = 94), rRNA processing (P value = 3.7e-6, count = 10), and Wnt signaling (P value = 1.0e-3, count = 5), whereas PC2 negative loading were enriched for neuropeptide hormone signaling and activity (P value = 1.3e-5, count = 8) (supplementary figs. *5B* and 6, Supplementary Material online). In particular, the Wnt genes *mom-2*, *mom-5*, and *pop-1* have altered their expression profiles between the genera, where they are expressed from the 8-cell stage to the L1 stage in the *Steinernema* species while they are expressed from the zygote to comma/1.5-fold/2-fold stage in *Caenorhabditis* (supplementary fig. *5C*, Supplementary Material online). PC3 also shows large differences between genera. Inspection of the top and bottom 100 PC3 gene loadings shows orthologs that have taken on diverse expression profiles during development between the two genera (supplementary fig. *4B*, Supplementary Material online). The PCA plots clearly display divergence of ortholog expression between genera at the earliest stages of development followed by convergence in expression at later stages.

Using the classical genome assembly statistic of N50 to rank genome quality, we see that they range from the finished genome of *C. elegans* being the most complete (chromosomes = 6, N50 = 17.49 Mb, genome size = 100 Mb) to the increasingly more fragmented draft genomes of *S. carpocapsae* (contigs = 1,577, N50 = 300 kb, genome size = 86 Mb), *S. feltiae* (contigs = 5,838, N50 = 47 kb, genome size = 82 Mb), and *C. angaria* (contigs = 34,620, N50 = 79.8 kb, genome size = 105 Mb). We found that both *Steinernema* species had higher numbers of expressed genes (defined as >1 Transcript per Million [TPM]) than the *Caenorhabditis* species and that this was also true for genes that are present in a single copy across species and share ancestry (1:1:1:1 orthologs) (fig. 3*C* and *D* and supplementary fig. *7A–C*, Supplementary Material online). However, the number of expressed genes, as well as 1:1:1:1 orthologs, were more comparable between genera at the later stages of embryonic development (from the comma to L1 stage), than at the earlier stages. Over 20,000 genes are detected at the 2-cell stage in *S. feltiae*, whereas the other species express between 7,000 and 11,000 genes at this stage. As development proceeds toward L1, the numbers of genes detected in *S. feltiae* (14,000 genes) becomes more similar to the other species (~14,000 genes in *S. carpocapsae*, ~11,500 in *C. angaria*, and ~10,500 in *C. elegans*). We considered whether the larger numbers of expressed genes in *Steinernema* were due to more annotated genes in the *Steinernema* genomes. We therefore analyzed the fraction of genes in the genomes that are expressed at different thresholds (1, 5, and 10 TPM) and found roughly comparable percentages across species (supplementary fig. *7D–F*, Supplementary Material online).

However, the spurious assembly of divergent haplotypes for the same gene could potentially lead to inflated gene numbers for some of the species as well. We therefore used CEGMA and BUSCO measures of genome completeness provided by WormBase ParaSite to evaluate the completeness and degree of predicted single-copy ortholog duplications of these genomes (Parra et al. 2007; Simão et al. 2015). All of these genomes showed low gene duplication levels by BUSCO (table 2). The number of genes expressed in *S. feltiae* during development fluctuates between being higher than the other species and then lower than the other species during early development, and then decreases to about the same number of genes as the other species towards later development, which suggests that this is not the result of assembled

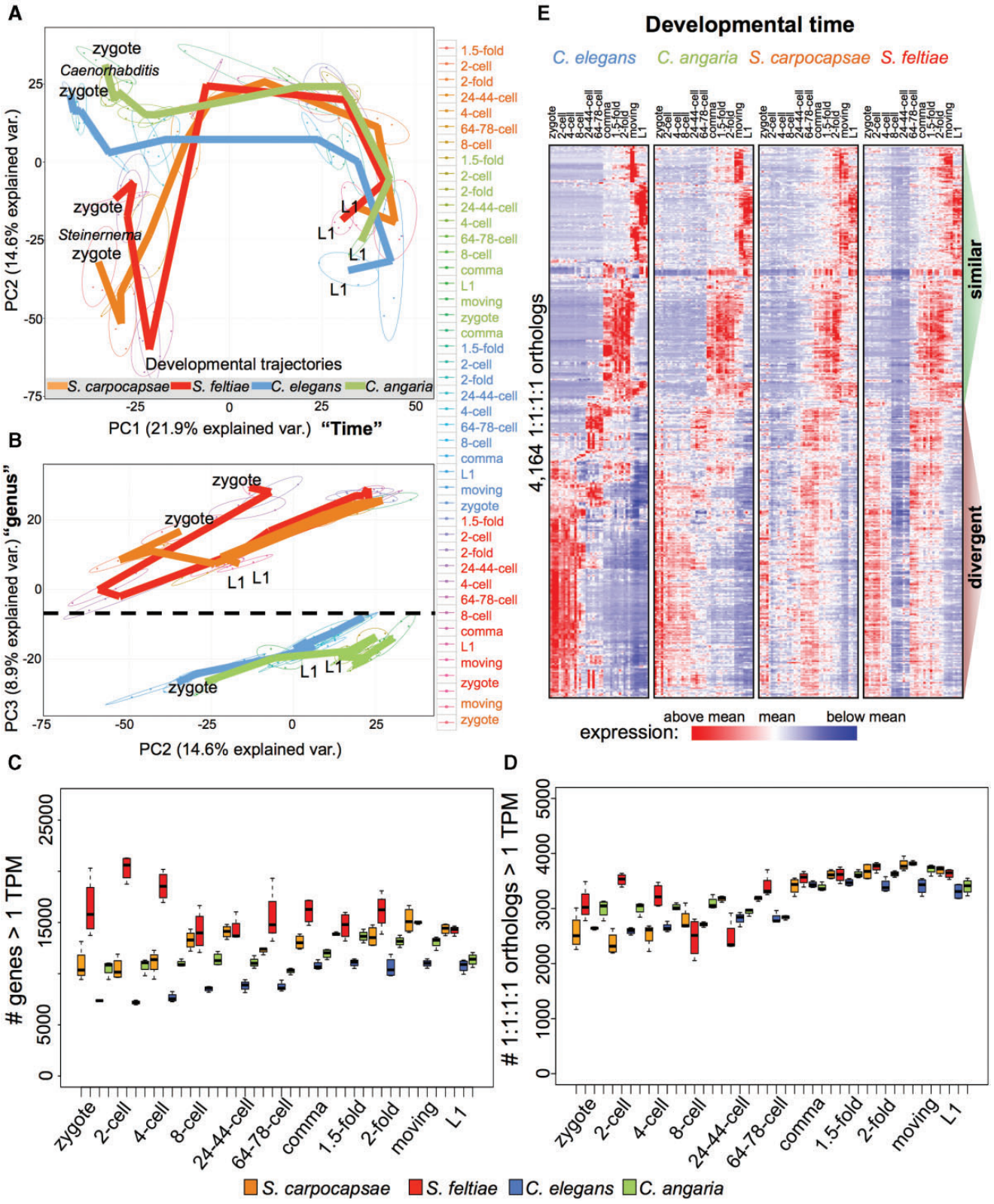

FIG. 3.—(*A*) Numbers of genes expressed during development across four species. Box plot showing the variance in the number of genes expressed >1 transcript per million (TPM) across embryo quadruplicates for each developmental stage. Number of genes displayed is out of 28,313 annotated genes in *S. carpocapsae*, 33,459 annotated genes in *S. feltiae*, 20,389 annotated genes in *C. elegans*, 27,970 annotated genes in *C. angaria*. (*B*) Box plot showing

Table 2
Genome Completeness Statistics

	CEGMA[a]		BUSCO[b]	
	Complete (%)	Partial (%)	Complete (%) [duplicated (%), single (%)]	Fragmented (%)
S. carpocapsae	100	100	89.4 [4.6, 84.8]	6.30
S. feltiae	97.58	98.39	86.7 [3.6, 83.1]	6.20
C. elegans	100	100	98.6 [0.6, 98]	0.80
C. angaria	83.47	91.53	77.7 [1.5, 76.2]	10.30

[a]Set of highly conserved eukaryotic genes.
[b]Set of single-copy orthologs that are present in > 90% of animals.

haplotypes. If this were the result of haplotype duplications, we would expect to see higher numbers of genes in *S. feltiae* across all developmental stages, not just a subset of them. To exclude whether haplotype contamination within the *Steinernema* genomes is affecting our analyses in figure 3, we used HaploMerger2 to collapse potential haplotype sequences in the *S. carpocapsae* and *S. feltiae* genomes (Huang et al. 2017). We collapsed haplotypes using two allelic difference (heterozygosity) thresholds, 2% and 4–5%, and found similar results at both thresholds for both genomes. Approximately 300 kb or 0.2% of the *S. carpocapsae* genome encompassing 181 genes was eliminated upon haplotype collapse, whereas 2.6 Mb or 3.1% of the *S. feltiae* genome encompassing 1,121 genes was also eliminated (table 3). These analyses show that the number of expressed genes and orthologs (> 1 TPM) is highly variable across the species during early embryonic development and less variable during later stages and that haplotype contamination is a negligible issue in the *S. carpocapsae* assembly and a minor issue in the *S. feltiae* assembly.

We repeated the global orthology analysis using a different orthology package (OMA) and WormBase ParaSite orthologs to make sure that our results were not sensitive to the specific ortholog detection method (Altenhoff et al. 2015). We found the same profiles as we saw with OrthoMCL indicating that the results are robust to changes in ortholog detection method (supplementary fig. 8*A* and *C*, Supplementary Material online). In addition, we repeated the analysis for 1,502 *N*:*N*:1:*N* orthologs (not including 1:1:1:1s), where 1 is the number of orthologs in *C. elegans* and $N \geq 1$ in the other species, with paralog expression averaged (supplementary fig. 8*B*, Supplementary Material online). We performed this analysis to control for the effects of potential divergent haplotypes on the results. If haplotypes are present in these genomes, then they should be represented in the orthology clustering as paralogs. We found that the *N*:*N*:1:*N* ortholog relationships (supplementary fig. 8*B*, Supplementary Material online) show the same general profile as the 1:1:1:1 orthologs plots (fig. 3*C* and supplementay fig. 8*A* and *C*, Supplementary Material online), where again ortholog expression is more variable at earlier time points than at later ones. In addition, we checked whether our restriction to 1:1:1:1 orthologs was biased by the presence of two haplotypes in the assembly, which affected 3.3% of the predicted genes in *S. feltiae* based on running HaploMerger2 (Huang et al. 2017) (table 3), but found that the results stayed essentially the same when incorporating the five 1:1:1:1 genes that had been duplicated by haplotype contamination (supplementary fig. 9, Supplementary Material online). We conclude that our result showing global convergence of gene expression during later development is robust to orthology methods and potential haplotype duplications.

Orthologs Expressed Specifically during Early Embryonic Development in *Caenorhabditis* Are Expressed at Both Early and Mid-Embryonic Stages in *Steinernema*

A heatmap of 1:1:1:1 ortholog expression used for PCA confirms that a set of orthologs which are expressed primarily during later embryonic development (comma to L1) shows conserved expression over the embryonic stages across all four of the species. However, we can also see that another set of orthologs, which appear to be strictly maternal in *C. elegans* and *C. angaria*, that is, are expressed only from the zygote stage up until the early or intermediate stages (8-cell to 24–44-cell), show downregulation at earlier stages (4-cell to 8-cell) in *Steinernema*, and interestingly, are then reexpressed in later stages of development (fig. 3*E*). This suggests that maternal-specific and other early embryonic orthologs in *Caenorhabditis* have new, additional roles in later embryonic

Fig. 3. Continued
the variability in the number of 1:1:1:1 orthologous expressed >1 TPM out of 4,164 orthologs shared between the four species at each developmental stage. (*C*) Principal Component Analysis of 4,164 1:1:1:1 shared orthologs between four species. Plot shows Principal Component (PC) 1 versus PC2. (*D*) Plot of PC2 versus PC3. Plots A and B show the developmental trajectories of each species and the clear genus-specific clustering in PC2 and PC3. (*E*) Heat maps of 1:1:1:1 ortholog expression during embryonic development across species. Gene expression (TPM—transcripts per million) during embryonic development of 4,164 1:1:1:1 orthologs was mean-centered and hierarchically clustered based on expression in *C. elegans*.

Table 3

S. carpocapsae and *S. feltiae* Genome Statistics After Haplotype Collapsing with HaploMerger2

	Original Genome	Allelic Difference = 2%		Allelic Difference = 4–5%	
		Reference (haplotype 1)	Alternate (haplotype 2)	Reference (haplotype 1)	Alternate (haplotype 2)
Haplomerger2—*S. carpocapsae*					
# scaffolds	1,577	958	958	961	961
GC content	45.53	45.5	45.5	45.5	45.5
N10	979,322	979,322	979,322	979,322	979,322
N50	299,566	303,283	303,283	300,834	300,834
N90	54,505	57,634	57,634	57,634	57,634
Largest scaffold size	1,694,367	1,694,367	1,694,314	1,694,367	1,694,314
Genome size	85,643,095	85,346,383	85,323,956	85,346,806	85,327,544
Haplomerger2—*S. feltiae*					
# Scaffolds	5,838	4,257	4,256	4,248	4,247
GC content	46.99	47	47	47	47
N10	303,346	308,236	308,236	308,236	312,167
N50	47,851	51,278	50,899	51,278	50,899
N90	7,114	8,593	8,564	8,593	8,564
Largest scaffold size	1,446,875	1,446,875	1,446,875	1,446,875	1,446,875
Genome size	82,626,797	79,950,914	79,719,430	79,929,219	79,693,975

development in *Steinernema*. Alternatively, these 1:1:1:1 orthologs may have been expressed in these later stages in ancestral species and have been lost expression at these time points in *Caenorhabditis*.

Another noticeable feature of the ortholog heat maps is a lack of highly expressed 1:1:1:1 orthologs at the 8-cell and the 24–44-cell stages in *S. feltiae*, and to a lesser extent the 2-cell through 8-cell stages in *S. carpocapsae* when the heat maps are clustered based on expression pattern in *C. elegans*. We hierarchically clustered the 1:1:1:1 orthologs based on expression in other species and found 305 orthologs in *S. feltiae* and 403 orthologs in *S. carpocapsae* that are expressed most highly in the 8-cell and 24–44-cell stages, showing that there are in fact 1:1:1:1 orthologs expressed at these stages in the *Steinernema* species (supplementary fig. 10*A–C*, Supplementary Material online).

Since transcription factors (TFs) are responsible for regulating the expression of genes during development, we suspected that the expression of transcription factors would mirror the profiles observed for the 1:1:1:1 orthologs and would also show major differences in the early and intermediate embryonic stages between the genera. We plotted the expression of TFs that are orthologous across all four species (1:1:1:1), three out of the four species (1:1:1:0, 1:1:0:1, 1:0:1:1, 0:1:1:1), two out of the four species (1:1:0:0, 0:0:1:1, 1:0:1:0,1:0:01, 0:1:1:0, 0:1:0:1), and one out of four species to assess their expression profiles (supplementary fig. 11, Supplementary Material online). The 253 1:1:1:1 orthologous TFs showed identical expression dynamics as the set of all 4,164 1:1:1:1 orthologs (fig. 3*E*). The subset of TFs that are expressed primarily during early embryogenesis in *Caenorhabditis* show less early embryo-specificity in *Steinernema*, with these TFs most highly expressed at the 24–44-cell and 64–78-cell stages in *Steinernema*. Focusing on TFs across all species combinations, we find that the maternal and early transcription factors are species- and genus-specific. We found many TFs that were specific to *C. elegans* (159) or *C. elegans* and *C. angaria* (99) that have diverse expression profiles during the time course. The group of 159 *C. elegans*-specific TFs includes 66 nuclear hormone receptors and the GATA TFs *end-3* and *end-1* that specify the endoderm at the 8-cell and 24–44-cell stage, respectively.

Focusing on TFs that are expressed in *S. carpocapsae* and one or more species but not in *C. elegans* (189 TFs), we find 26 TFs (14%) that have early embryo-specific expression. The set of 189 TFs found in *S. carpocapsae*, but not *C. elegans*, had GO enrichments, such as positive mesodermal fate specification (FDR = 9.53e-9, fold enrichment = 64.2), response to retinoic acid (FDR = 1.12e-10, fold enrichment = 70.4), dorsal/ventral pattern formation (FDR = 9.99e-7, fold enrichment = 21.8), positive regulation of cell proliferation (FDR = 9.5e-4, fold enrichment = 9.2), regulation of establishment of cell polarity (FDR = 6.02e-3, fold enrichment = 59.9), and embryonic digestive tract morphogenesis (FDR = 4.14e-2, fold enrichment = 21.4) (table 4). These results suggest that the *Steinernema* specific-TFs are likely to participate in the regulation of multiple developmental processes.

Differential Gene Expression Analysis of Adjacent Stages Suggests Specific-Specific Initiation of Zygotic Transcription

In order to detect specific transcriptional changes between early embryos, we performed differential gene expression

Table 4
GO Terms for *S. carpocapsae* Transcription Factors that Are Not in *C. elegans*

GO Term	FDR	Fold Enrichment
Positive regulation of transcription from RNA pol II promoter	6.29E-40	49.4
Negative regulation of transcription from RNA pol II promoter	3.97E-26	31.3
Response to retinoic acid	4.31E-12	70.4
Retinoic acid receptor signaling	1.12E-10	112.3
Mesodermal cell fate specification	9.53E-09	64.2
Dorsal/ventral pattern formation	9.99E-07	21.8
N-terminal peptidyl-lysine acetylation	9.27E-07	99.9
Somatic stem cell population maintenance	6.59E-04	44.9
Regulation of transcription involved in primary germ layer cell commitment	8.26E-04	149.8
Positive regulation of cell proliferation	9.50E-04	9.2
Regulation of mesoderm development	2.06E-03	99.9
Regulation of establishment of cell polarity	6.02E-03	59.9
Embryonic digestive tract morphogenesis	4.14E-02	21.4

(DE) analyses between pairs of adjacent early developmental stages using either all of the genes within each species (fig. 4) or the 4,164 1:1:1:1 orthologs shared between them (supplementary fig. 12, Supplementary Material online). In *Caenorhabditis*, very few genes or orthologs were differentially expressed (FDR < 0.05 and fold change $> 2\times$) between stages before 4-cell. Once the embryos reached the 8-cell stage in *C. elegans*, 972 genes became differentially upregulated relative to the 4-cell stage, consistent with our previous correlation matrix results (fig. 2) and with previous published results showing that zygotic expression begins at the 4-cell stage in *C. elegans* (Edgar et al. 1994, Baugh et al. 2003). In contrast, *C. angaria* showed very little change in gene expression until the 8-cell to 24–44-cell stage transition. At that point, 1,440 genes were upregulated in the 24–44-cell stage relative to the 8-cell stage, indicating that zygotic transcriptional changes are occurring at later developmental stages in *C. angaria* than in *C. elegans*. Both *Steinernema* species showed a substantial upregulation of gene expression (4,787 genes in *S. carpocapsae* and 2,938 genes in *S. feltiae*) at the 8-cell stage similar to *C. elegans*. However, both *S. carpocapsae* and *S. feltiae* show upregulation in a subset of genes prior to the 8-cell stage, in the 2-cell (541 genes) and 4-cell stages (251 genes), respectively. A Gene Ontology (GO) analysis of these early upregulated genes at 2-cell in *S. carpocapsae* found that they are enriched for terms involved in yolk granules (FDR = 3.91e-3, fold enrichment = 23.7) and ubiquitination (FDR = 2.11e-2, fold enrichment = 3.98). We did not find any significant GO term enrichments for the early upregulated *S. feltiae* genes. It is unclear why there is an upregulation from zygote to 2-cell and then a plateau in gene expression from 2-cell to 4-cell in *S. carpocapsae*. It may be that zygotic transcription starts for a small subset of genes at an earlier stage in *Steinernema* than in *Caenorhabditis*.

maSigPro Ortholog Expression Clustering within Each Genus

To discover orthologs that show significant temporal expression dynamics between the species in *Steinernema* and *Caenorhabditis*, we used maSigPro (Conesa et al. 2006; Nueda et al. 2014) on 9,844 1:1 orthologs shared between *S. carpocapsae* and *S. feltiae* and 6,840 1:1 orthologs that are shared between *C. elegans* and *C. angaria* (supplementary fig. 13, Supplementary Material online). We found that 4,819 (48.9%) of the *Steinernema* orthologs and 4,462 (65.2%) of the *Caenorhabditis* orthologs were dynamically expressed during embryonic development (Benjamini Hochberg FDR < 0.01). These dynamically expressed genes partitioned into nine different clusters (supplementary fig. 13*B* and *C*, Supplementary Material online) based on their expression profile. Clusters 1 and 2 show the dynamics of the early orthologous embryonic or "maternal" transcripts, clusters 3 and 4 show the dynamics of early to intermediate embryonic development (8-cell to 24–44-cell or 74–78-cell), clusters 5 and 6 show the dynamics of intermediate to late genes, and clusters 7–9 show the dynamics of very late development until hatching. The clusters also represent orthologs that are higher on average in one species than another at around the same time points during development. For example, *Steinernema* cluster 1 and 2 show genes that are "high" in both *Steinernema* species very early on during embryogenesis, but it is clear that cluster 1 genes are much higher in *S. carpocapsae* than *S. feltiae*, whereas cluster 2 genes are higher in *S. feltiae* than *S. carpocapsae* (supplementary fig. 13*B*, Supplementary Material online). We found GO terms associated with nuclear lumen (FDR = 2.96e-2, fold enrichment = 5.4), mitotic spindle checkpoint (FDR = 4.11e-2, fold enrichment = 36.6), and AMP-activated protein kinase (FDR = 4.78e-2, fold enrichment = 54.9) for cluster 1. In clusters 3 and 4, we found proteasomal catabolic process (FDR = 2.7e-2, fold

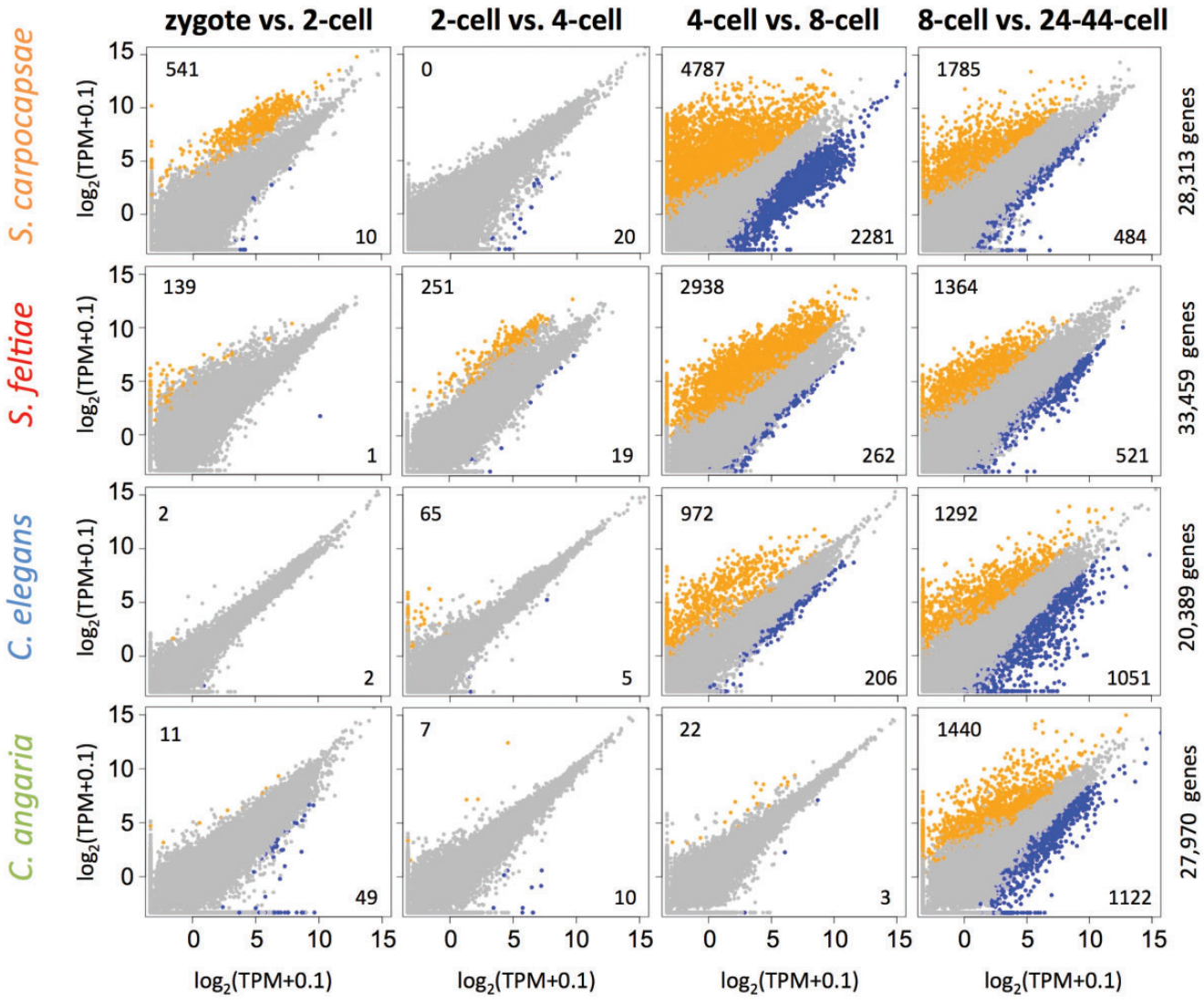

Fig. 4.—Differential gene expression of all genes during early embryonic development across species. Gene expression $\log_2(\text{TPM} + 0.1)$ of all genes was plotted for adjacent early developmental stages for all four species. The earlier stages are displayed on the x-axis and the later stages are displayed on the y-axis. Genes that are differentially expressed (FDR < 0.05 and fold change $> 2\times$) between the stages, and are more highly expressed in the earlier stage or later stage are shown in blue and yellow, respectively. Genes in gray are not differentially expressed.

enrichment = 13.3) and many terms related to neddylation and ubiquitination. It is interesting that for all of these clusters, except for cluster 9 in *Steinernema*, there is a fairly large difference in the magnitude of expression between species of the same genus.

Contributions of nonorthologous genes to embryonic development in *S. carpocapsae*

Given our finding that our 1:1:1:1 orthologs show more conserved expression during later development, we asked what the contribution of the other 75% of genes are to development. We focused on *S. carpocapsae* genes that share homology with at least one other *Steinernema* species (*S. feltiae*, *S. glaseri*, *S. monticolum*, and *S. scapterisci*) but no homology to any of the *Caenorhabditis* species (*C. elegans, C. angaria, C. briggsae, C. japonica*, and *C. remanei*), and that are both expressed at an average of 10 TPM during embryonic development and have at least one replicate with expression > 50 TPM (supplementary fig. 14, Supplementary Material online). These expression thresholds were set to ensure that these genes are true expressed genes and not pseudogenes. We found 5,679 genes that fit these criteria and that 1,036 genes (18.1%) are expressed between zygote to 4-cell (clusters 1 and 2), 2,272 genes (39.8%) are expressed between 8-cell and 64–78-cell (cluster 2 and 3), and the remaining 2,389 (41.9%) genes are expressed at some point between comma to L1 (clusters 4 and 5). Approximately half of these *Steinernema*-only genes

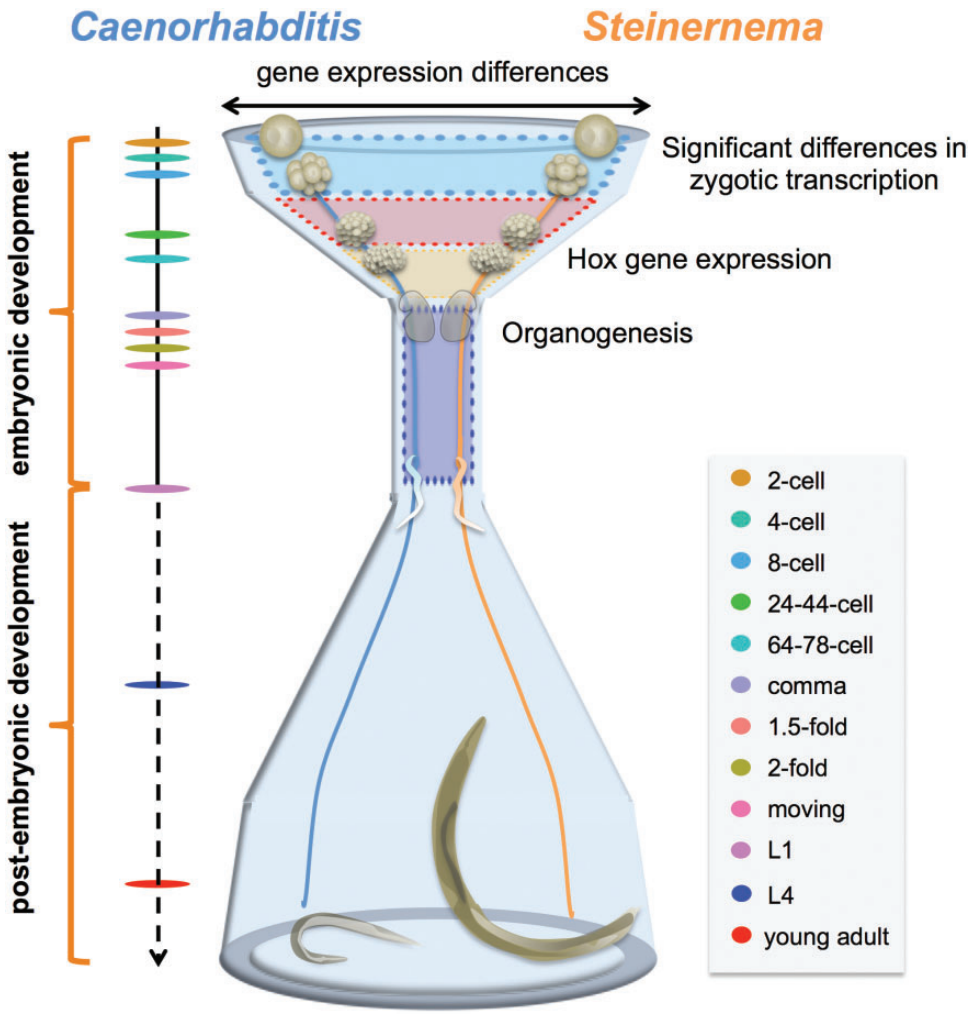

Fig. 5.—A model of gene expression divergence over embryonic (to scale) and postembryonic development (not to scale) between distant genera. A funnel model of nematode embryonic development, where 1:1:1:1 ortholog expression variation is high in early stage embryos and low in later stage embryos during embryonic development, and high during postembryonic development across genera. Embryonic and postembryonic stages are in the gray figure legend. The solid line shows the developmental time points that have been scaled to fit the funnel. The dashed line shows a few postembryonic developmental time points that are not to scale.

(2,674, 47%) have no match to proteins from any other species. Of the 3,005 genes that do have annotations, 24 are fatty-acid and retinol binding proteins, fatty-acid amide hydrolases, fatty-acid desaturases or fatty-acid elongation protein annotations, and 53 are ubiquitin E3 ligases or ubiquitin-related proteins. Another 26 are homeobox-domain containing proteins. This could suggest alternative gene expression cascades or programs governing *Steinernema* development. Together, *Steinernema*-conserved genes that have no *Caenorhabditis* orthologs are expressed throughout embryogenesis and are likely to affect several processes during their development.

Discussion

We generated high-resolution single embryo RNA-seq time courses spanning 11 embryonic stages for two *Steinernema* and two *Caenorhabditis* species to determine the extent of ortholog expression conservation during embryogenesis across distantly related genera. During the early stages of embryogenesis, we observed large transcriptional changes, which are representative of the maternal-to-zygotic transition (MZT) of each species, and found that that the MZT occurs at earlier developmental stages in *Steinernema* (2-cell and 4-cell) in comparison to *Caenorhabditis* (8-cell and 24–44-cell). The large upregulation of gene expression at these early stages also coincides with the degradation of the *oma-1/2* transcripts in these species providing additional support that the MZT is initiating at earlier stages in *Steinernema*.

In 1828, Von Baer proposed a reverse funnel model of animal development where developmental similarities are highest in the earliest stages of embryogenesis and lowest at the end of development (von Baer 1828). In 1994,

Duboule proposed a variant of the model called the hourglass model. In his model, embryonic divergence during development follows an hourglass-like shape, where embryos of different species are most divergent at the earliest and latest stages of development, but not the middle stages of development when the body plan is being set (von Baer 1828; Duboule 1994). Duboule proposed that critical developmental processes are occurring during this middle period of development, and that these processes are governed by evolutionarily conserved genes and gene networks, which place a large constraint on animal development (Bateson 1894; Duboule 1994; Irie and Kuratani 2014). Since then, various studies across many organisms from nematodes to flies have supported this hourglass model (Domazet-Lošo et al. 2010; Kalinka et al. 2010; Levin et al. 2012). The transcriptome analysis conducted by Levin et al. during the embryonic development of five *Caenorhabditis* species found that gene expression was constrained at several points during the middle of embryogenesis within the genus (Levin et al. 2012). They termed these points of convergence developmental milestones, and their findings are reminiscent of the highly debated "phylotypic" stage of the hourglass model of animal development. But many of these studies factored in species heterochrony into their determination of the phylotypic stage.

Our embryonic analysis, which focused on equivalent morphological stages instead of time, shows a slightly different picture of embryonic development. In our nematode comparative system, expression of orthologs from mid to late developmental time points show greater conservation than earlier ones between *Caenorhabditis* and *Steinernema*, unlike what would be expected in the hourglass model. The lower degree of expression conservation between 1:1:1:1 orthologs during earlier embryonic development (zygote to 8-cell) in contrast to the later stages of development (64–78-cell to L1) leads us to propose the funnel model of embryonic development for nematodes who are more distantly related than the ones considered in the Levin et al. study (fig. 5). The large amount of variation we see in gene expression in early development is reminiscent of what nematologists have previously seen at the macroscopic level between different species, such as differences in the timing of gastrulation, AP axis specification, and when the endoderm and mesoderm cells are specified (Goldstein et al. 1998; Schierenberg 2006). Our results are similar to the results of a study on gene expression conservation between equivalent morphological stages during *Xenopus laevis* and *Xenopus tropicalis* embryogenesis (Yanai et al. 2011), where the authors found that the number of orthologous genes that are differentially expressed across species during development steadily decreased. It should be noted that these two *Xenopus* species have diverged more recently than the species in our study (about 60–90 Ma vs. 200+ Ma). The differences in the staging of the MZT between species would also contribute to the high divergence in ortholog expression we observed across species during early embryogenesis. However, our model is not in contradiction with a broader, whole-lifecycle version of the hourglass model. The last larval stage and adult stages are very diverse morphologically between species of different genera, including between *Steinernema* and *Caenorhabditis*. In effect, the gene expression funnel that we observe during embryogenesis is only the top half of the hourglass that is evident when development up to the adult stage is taken into consideration and represents an interesting variation on the standard hourglass model.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We would like to thank all of the members of the Mortazavi lab, Dr. Adler Dillman, Dr. Steve Shen, Dr. Olivier Cinquin, Dr. Brandon Gaut, Dr. Ken Cho, Dr. Anne Calof, and Dr. Shaun Hug for helpful discussions. This work was supported by an NIH New Innovator Award to A.M. (NIGMS DP2 GM111100).

Literature Cited

Altenhoff AM, et al. 2015. The OMA orthology database in 2015: function predictions, better plant support, synteny view and other improvements. Nucleic Acids Res. 43(Database issue): D240–D249.

Araya CL, et al. 2014. Regulatory analysis of the *C. elegans* genome with spatiotemporal resolution. Nature 512(7515): 400–405.

Bateson W. 1894. Materials for the study of variation. New York, NY: Macmillan.

Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP. 2003. Composition and dynamics of the *Caenorhabditis elegans* early embryonic transcriptome. Development 130(5): 889–900.

Conesa A, Nueda MJ, Ferrer A, Talon M. 2006. maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. Bioinformatics 22(9): 1096–1102.

Bryer J. 2013. timeline: Timelines for a Grammar of Graphics. R package version 0.9. Available from: http://CRAN.R-project.org/package=timeline, last accessed September 27, 2017.

de Hoon MJ, Imoto S, Nolan J, Miyano S. 2004. Open source clustering software. Bioinformatics 20(9): 1453–1454.

De Ley P, Blaxter ML. 2002. Systematic position and phylogeny. In: Lee D, editor. The biology of nematodes. London: Taylor and Francis. p. 1–30.

Dillman AR, et al. 2015. Comparative genomics of Steinernema reveals deeply conserved gene regulatory networks. Genome Biol. 16: 200.

Duboule D. 1994. Temporal colinearity and the phylotypic progression: a basis for the stability of a vertebrate Bauplan and the evolution of morphologies through heterochrony. Development (Suppl): 135–142.

Edgar LG, Wolf N, Wood WB. 1994. Early transcription in *Caenorhabditis elegans* embryos. Development 120(2): 443–451.

Gerstein MB, et al. 2010. Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. Science 330(6012): 1775–1787.

Gertz J, et al. 2012. Transposase mediated construction of RNA-seq libraries. Genome Res. 22(1): 134–141.

Goldstein B, Frisse LM, Thomas WK. 1998. Embryonic axis specification in nematodes: evolution of the first step in development. Curr Biol. 8(3): 157–160.
Huang S, Kang M, Xu A. 2017. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. Bioinformatics 33(16): 2577–2579.
Irie N, Kuratani S. 2014. The developmental hourglass model: a predictor of the basic body plan. Development 141(24): 4649–4655.
Kalinka AT, et al. 2010. Gene expression divergence recapitulates the developmental hourglass model. Nature 468(7325): 811–814.
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 10(3): R25.
Levin M, Hashimshony T, Wagner F, Yanai I. 2012. Developmental milestones punctuate gene expression in the Caenorhabditis embryo. Dev Cell 22(5): 1101–1108.
Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12: 323.
Li L, Stoeckert CJJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13(9): 2178–2189.
Mortazavi A, et al. 2010. Scaffolding a Caenorhabditis nematode genome with RNA-seq. Genome Res. 20(12): 1740–1747.
Nueda MJ, Tarazona S, Conesa A. 2014. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. Bioinformatics 30(18): 2598–2602.
Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23(9): 1061–1067.
Picelli S, et al. 2014. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 24(12): 2033–2040.
Picelli S, et al. 2014. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 9(1): 171–181.
R Development Core Team. 2008. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0, Available from: http://www.R-project.org, last accessed September 27, 2017.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1): 139–140.
Saldanha AJ. 2004. Java Treeview: extensible visualization of microarray data. Bioinformatics 20(17): 3246–3248.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210–3212.
Schierenberg E. 2006. Embryological variation during nematode development. WormBook 2006: 1–13.
Stitzel ML, Pellettieri J, Seydoux G. 2006. The *C. elegans* DYRK kinase MBK-2 marks oocyte proteins for degradation in response to meiotic maturation. Curr Biol. 16(1): 56–62.
Sulston JE, Schierenberg E, White JG, Thomson JN. 1983. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev Biol. 100(1): 64–119.
Tadros W, Lipshitz HD. 2009. The maternal-to-zygotic transition: a play in two acts. Development 136(18): 3033–3042.
von Baer KE. (1828). Uber Entwickelungsgeschichte der Thiere. Koenigsberg: Beobachtung und Reflektion.
Voronov DA, Panchin YV. 1998. Cell lineage in marine nematode *Enoplus brevis*. Development 125(1): 143–150.
Voronov DA, Panchin YV, Spiridonov SE. 1998. Nematode phylogeny and embryology. Nature 395(6697): 28.
Yanai I, Peshkin L, Jorgensen P, Kirschner MW. 2011. Mapping gene expression in two Xenopus species: evolutionary constraints and developmental flexibility. Dev Cell 20(4): 483–496.

Associate editor: Gunter Wagner