Q:1, 2, 3, 4, 5, 6, 7

The gene regulatory program of *Acrobeloides nanus* reveals conservation of phylum-specific expression

Philipp H. Schiffer[a,1], Avital L. Polsky[b,1], Alison G. Cole[c], Julia I. R. Camps[d], Michael Kroiher[e], David H. Silver[b], Vladislav Grishkevich[b], Leon Anavy[b], Georgios Koutsovoulos[f,2], Tamar Hashimshony[b], and Itai Yanai[g,3]

Q:8 [a]Department of Genetics, Evolution and Environment, University College London, London, United Kingdom; [b]Department of Biology, Technion–Israel Institute of Technology, 32000 Haifa, Israel; [c]Department of Molecular Evolution and Development, University of Vienna, Vienna, Austria; [d]Molecular Cell Biology, Institute I for Anatomy University Clinic Cologne, University of Cologne, Cologne, Germany; [e]Zoological Institute, Cologne Biocenter, University of Cologne, Cologne, Germany; [f]School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom; and [g]Institute for Computational Medicine, NYU School of Medicine, New York, NY

Edited by Paul W Sternberg, California Institute of Technology, Pasadena, CA, and approved March 14, 2018 (received for review December 14, 2017)

The evolution of development has been studied through the lens of gene regulation by examining either closely related species or extremely distant animals of different phyla. In nematodes, detailed cell- and stage-specific expression analyses are focused on the model *Caenorhabditis elegans*, in part leading to the view that the developmental expression of gene cascades in this species is archetypic for the phylum. Here, we compared two species of an intermediate evolutionary distance: the nematodes *C. elegans* (clade V) and *Acrobeloides nanus* (clade IV). To examine *A. nanus* molecularly, we sequenced its genome and identified the expression profiles of all genes throughout embryogenesis. In comparison with *C. elegans*, *A. nanus* exhibits a much slower embryonic development and has a capacity for regulative compensation of missing early cells. We detected conserved stages between these species at the transcriptome level, as well as a prominent middevelopmental transition, at which point the two species converge in terms of their gene expression. Interestingly, we found that genes originating at the dawn of the Ecdysozoa supergroup show the least expression divergence between these two species. This led us to detect a correlation between the time of expression of a gene and its phylogenetic age: evolutionarily ancient and young genes are enriched for expression in early and late embryogenesis, respectively, whereas Ecdysozoa-specific genes are enriched for expression during the middevelopmental transition. Our results characterize the developmental constraints operating on each individual embryo in terms of developmental stages and genetic evolutionary history.

evolution | development | gene expression

An insight regarding the embryo that continues to provide understanding is the notion that evolutionary constraints have shaped development (1, 2). Indeed, the field of evolutionary developmental biology posits that these two concepts are intertwined and mutually illuminating (3). The comparative approach of analyzing distant species has shed light on many processes, including the evolution and development of the bilaterian body plan by HOX genes (4, 5). Although it might be naively expected Q:11 that comparing two closely related species would result in only a few genomic and transcriptomic changes, the last two decades have provided plenty of evidence that the genome and its phenotypes are extremely plastic (6, 7). These changes are manifest, but they are not random, and we require an understanding of how constraints act on possible genomic changes.

Transcriptomics methods, beginning with DNA microarrays, later followed by RNA-SEq (8, 9), have been transformative for biological research, as they afford a comprehensive view of gene expression. Whereas previous methods examined individual genes, with the simultaneous knowledge of the expression of all the genes in a given sample, a highly resolved state of system emerged, enabling the study of cellular, developmental, and evolutionary biology. Using transcriptomics, sharp changes in gene expression were detected throughout embryogenesis, suggesting the existence of developmental milestones (10). These were observed by gene expression changes that are not gradual but, rather, punctuate the embryo. Moreover, it was shown that different stages show different levels of expression conservation, suggesting different levels of expression constraints. The different stages also showed different compositions of genes in terms of their ages (11), which supported the notion that the stages of embryogenesis have unique evolutionary histories.

One particular stage during embryogenesis stood out in comparative transcriptomics studies. Studying a collection of *Caenorhabditis* species, the ventral enclosure stage was found to correspond to a period of intense changes in gene expression (10). Studies in arthropods and chordates revealed a similar middevelopmental stage. Interestingly, the stage in each of these phyla corresponded to the phylotypic stage: a period in which the species appear the most similar, morphologically. This middevelopmental transition between an early gastrulation module and a late morphogenesis module was observed in seven additional phyla in a recent study (12). Also, when studying this middevelopmental transition using mutation accumulation lines, it was observed that genes expressed during this stage are less likely to be different within a population of *C. elegans* species (2), suggesting that the middevelopmental transition is under severe developmental constraints.

Significance

Comparing gene regulatory programs throughout developmental time and across species allows us to reveal their constraints and flexibilities. Here we study the organism *Acrobeloides nanus*, a clade IV nematode, by sequencing its genome, identifying its developmental transcriptome, and studying the patterns of embryonic conservation and divergence through a comparison with *Caenorhabditis elegans*. The gene regulatory programs of these two species show many differences early in development, but significantly converge at the middevelopmental transition. Moreover, the genes most conserved in their expression during development arose at the dawn of the superphylum Ecdysozoa. Our work shows that variation is not evenly distributed but, rather, that developmental and evolutionary constraints act to shape gene regulatory programs.

Author contributions: I.Y. designed research; P.H.S., A.L.P., A.G.C., and T.H. performed research; A.G.C., J.I.R.C., M.K., and T.H. contributed new reagents/analytic tools; P.H.S., A.L.P., D.H.S., V.G., L.A., G.K., and I.Y. analyzed data; and P.H.S. and I.Y. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: All raw data are deposited in the NCBI Sequence Read Archive (Bioproject PRJNA354072). The genome, transcriptome, and annotations are available at genomehubs.org. Q:9

[1]P.H.S. and A.L.P. contributed equally to this work.

[2]Present address: INRA, Institut Sophia Agrobiotech, France. Q:10

[3]To whom correspondence should be addressed. Email: Itai.Yanai@nyumc.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1720817115/-/DCSupplemental.

The rate of development varies drastically in nematodes, even between those that are closely related (13–18). Although *C. elegans* has a generation time of 3–7 d, other nematode species can take anywhere from days to a year (19, 20). The clade IV species *Acrobeloides nanus* has a rate of embryogenesis that is four times longer than that of *C. elegans* (at 20 °C) and differs substantially from *C. elegans* in many aspects of life cycle, mode of living, and phenotype. Although it was initially assumed that *C. elegans* development is archetypic for nematodes, it has now been shown that early development in *A. nanus* is far more regulative (21) and that, for example, gastrulation in the enoplean species *Tobrilus stefanskii* is much more similar to nonnematode Bilateria (13). It has also become apparent that the molecular toolkit of development varies across the phylum, and even between closely related taxa (22, 23).

In particular, *A. nanus* blastomeres remain multipotent until at least the five-cell stage, able to reassign their cell fates to compensate for the death of a neighboring blastomere (21). *A. nanus* also differs from *C. elegans* in its ability to tolerate a wider range of environmental stresses: it develops optimally at 25 °C, whereas *C. elegans*, typically cultured at 20 °C, is negatively affected by such a high temperature (17). Moreover, *A. nanus* has an increased tolerance to desiccation and toxins (24, 25). Finally, *A. nanus* is one of many obligate parthenogens in the nematode phylum, and as such, its development is, unlike that of *C. elegans*, initiated without sperm input (26).

Here we compare the embryogenesis of *A. nanus* and *C. elegans* at the gene expression level. We describe the genome and transcriptome of *A. nanus* and show how they allow for the study of transcriptional differences of cells and developmental stages in this species. We compare at the single-cell level the two-cell stage and find a tremendous amount of variation. Comparing the temporal developmental transcriptomes of these two species, we find that there are similar sharp changes at developmental milestones. In particular, we find that the middevelopmental transition is the stage at which gene expression differences between the pair of species begin to significantly decrease. In general, the genes that are more conserved are those that arose at the origin of the Nematode phylum and the superphylum Ecdysozoa. Further examining this observation, we found a relationship between the timing of expression of a gene and its phylogenetic origin. Genes arising during the superphylum Ecdysozoa are expressed during the middevelopmental transition, which can explain their increased conservation over evolutionary time. Our analysis illustrates how species with key phylogenetic distances may be leveraged to address evolutionary developmental biology, using molecular tools.

Results

Genome Analysis of *Acrobeloides nanus*. To study the evolution of embryogenesis, we sought to compare at the molecular and developmental level between *C. elegans* and the clade IV nematode *Acrobeloides nanus* (Fig. 1*A*). We assembled the *A. nanus* genome on the basis of Illumina sequencing of DNA and RNA (*Supporting Information*). Our genome assembly encompassed 248 Mbp comprising 30,759 contigs with an N50 of 19,614 bp. As Fig. 1*A* shows, *A. nanus* has a fairly large genome relative to the other species. To account for this difference, we investigated repetitive DNA and estimated that it constitutes ~50% of the genome, with 43% of these repeats being unclassified (Table S1). A driver for this might be parthenogenetic reproduction in *A. nanus*, as parthenogenetic species are not able to efficiently remove repeats from the population (27). Recent studies, however, did not find an inflation of transposable elements in several parthenogenetic arthropod species (28), nor in another parthenogenetic nematode (29). Thus, we propose that the accumulation of repeats in *A. nanus* is random, as observed in other species with small effective population sizes (30, 31).

Running the BUSCO3 pipeline (32) on our *A. nanus* assembly revealed that it is 89% complete and 95% partial complete for the Eukaryote gene set. We obtained 35,692 gene predictions Augustus (33), trained on the RNA-Seq data (*Supporting Information*). We annotated 20,281 of the *A. nanus* proteins with

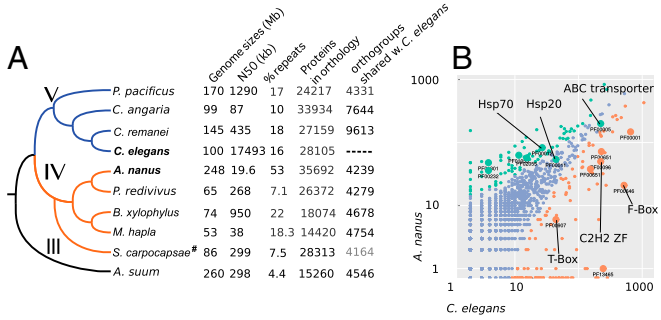

Fig. 1. The genome of the nematode *A. nanus* in comparison with that of other nematodes. (*A*) Phylogenetic tree of the indicated species. Roman numerals indicate clades according to ref. 20. Genome sizes, N50 of the assembly, repeats (23, 50), protein count, and number of orthologs with *A. nanus* are indicated in the table (see *SI Experimental Procedures*, #except for *S. carpocapsae* data, where 1–1 orthologs from ref. 43 are given). (*B*) Scatter plot of gene family sizes between *A. nanus* and *C. elegans*. Differentially enriched families are indicated by color. Larger circles indicate specific families: PF00001, Rhodopsin-like receptors; PF00001, ABC transporters; PF00011, Hsp20/alpha crystallin family; PF00012, Hsp70 protein; PF00096, zinc finger; C2H2 type; PF000232, glycosyl hydrolase family 1, transcription factors; PF00651, overrepresented Pfam domains between *A. nanus* and *C. elegans*.

PFAM domains, using InterProScan and in a bispecies comparison with *C. elegans* and screened for gene family inflations (Fig. 1*B*). Finally, employing OrthoFinder (34), we identified *A. nanus* orthologs across eight species selected on the basis of their phylogenetic position, with 4,240 groups of orthologs containing *A. nanus* and *C. elegans* proteins.

The *A. nanus* genome shows dramatic variation at the level of gene families relative to *C. elegans* (Fig. 1*B*). Pfam analysis shows more Brachyury-like (T-box) genes in *C. elegans* (22 genes) relative to *A. nanus* (six genes). The *C. elegans* genome is also overrepresented in other transcription factor families; namely, Zinc fingers of the C2H2 and C4 type, F-Box domains, and BTB/POZ domains. In contrast, *A. nanus* has more glycosyl hydrolase family genes, Hsp70, and Hsp20, as well as ABC transporters ($P < 0.05$, Fisher's exact test, FDR-corrected). Interestingly, consistent with the expansion of the Hsp gene family, *A. nanus* develops into normal adults in large numbers when kept at 30 °C; a temperature at which *C. elegans* quickly becomes sterile (35).

Studying *A. nanus* Blastomeres Using Single-Cell RNA-Seq. We sought to use the genome assembly to study the early stages of embryogenesis. We collected individual AB and P_1 blastomeres (Fig. 2*A*) and sequenced their transcriptomes using single-cell RNA-Seq (*SI Experimental Procedures*). The identity of the blastomeres could be clearly distinguished morphologically, as well as from their transcriptomes (Fig. 2*B*). To study the transcriptomes at the gene level, we identified the differentially segregated genes between the AB and P_1 blastomeres. We found that transcripts of heat shock genes are found in greater numbers in AB, whereas ribosomal genes are higher in P_1 (Fig. 2*C*). Interestingly, this was not observed in *C. elegans* (36).

We next compared the overall pattern of gene expression at the two-cell stage between *C. elegans* and *A. nanus*. For this, we compared with previously published *C. elegans* single-cell RNA-Seq data (36) and found genes with conserved and divergent AB-P_1 segregations (Fig. 2*D*). P-granule-associated genes are expressed in the same direction (36). *skn-1* is evenly expressed between AB and P_1 in *C. elegans*; however, our previous analysis using in situ staining of *skn-1* mRNA (22) showed a higher expression of this gene in the AB cell in the *A. nanus* two-cell stage embryo. Our single-cell transcriptomics data are in accordance with this previous finding, supporting the validity of the approach.

We found a small number of genes to be highly expressed in either the *A. nanus* AB or P_1 blastomere that had no expression in the *C. elegans* two-cell stage. Screening these genes for enriched

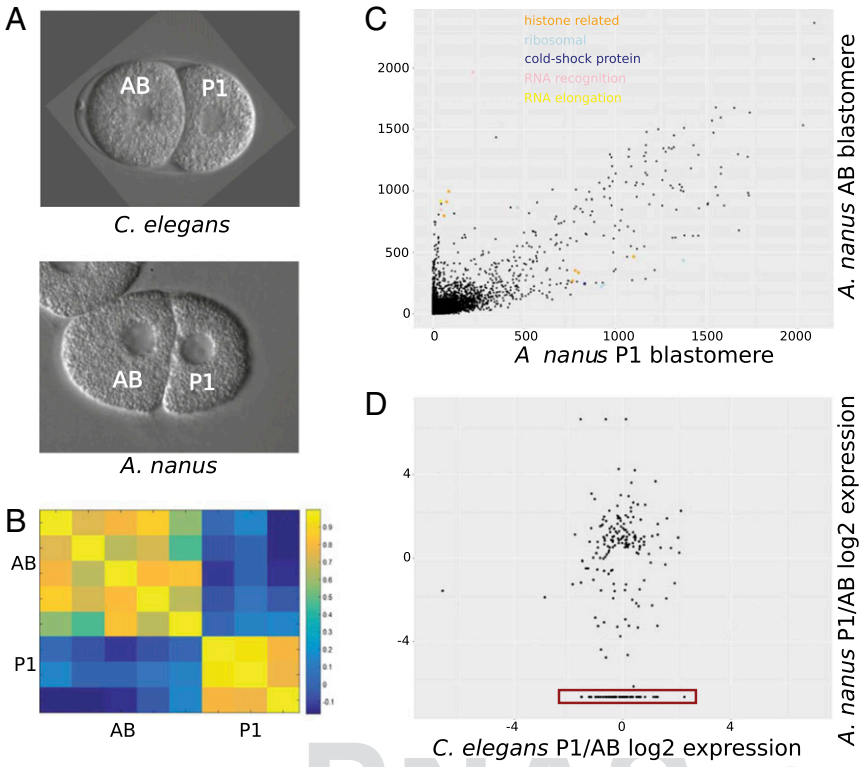

Fig. 2. Single-cell *A. nanus* blastomere analysis. (*A*) The two-cell stage in *A. nanus* and *C. elegans*, indicating also the AB and P_1 blastomeres. Embryos are 50 μm in length. (*B*) Heat map showing correlation coefficients among the *A. nanus* transcriptomes of five AB blastomeres and three P_1 blastomeres. (*C*) Comparison of the *A. nanus* gene expression levels between the AB and P_1 blastomeres. Expression levels are computed as transcripts per million (tpm; *SI Experimental Procedures*). Genes of the indicated functional groups are highlighted. (*D*) Ratios of expression between AB and P_1 in *C. elegans* and *A. nanus*, respectively. The red box indicates genes with high P_1 expression only in *A. nanus*.

functional groups according to their gene ontology terms, we found terms relating to reproduction, body morphogenesis, molting, regulation of growth, and transcription initiation ($P < 0.001$, hypergeometric distribution). This last functional description is particularly of interest because the slow and regulative development of *A. nanus* might not rely on many maternally deposited transcripts and proteins, similar to *C. elegans*, but, rather, on primarily zygotic expression. This is in accordance with the prediction that the fast development seen in *C. elegans* requires the deposition of a higher amount of maternal factors in general (37). Because comparison of the two-cell stage showed differences between the clade IV species and the model organism from clade V, we wanted to quantify the divergence in embryonic development between *A. nanus* and *C. elegans* on a global level.

Developmental Dynamics in *A. nanus* Reveal Distinct Stages. To identify the gene expression of all genes throughout embryogenesis, we assayed expression in individual embryos throughout *A. nanus* development. In contrast to the two-cell stage analysis, in this analysis, we focused exclusively on temporal resolution for the entire developmental process (Fig. 3*A*). Morphologically, *A. nanus* stages differ from those of *C. elegans*; however, at the 102-cell stage, the two species appear to have converged in their cell locations (38).

We produced a gene expression time-course dataset according to our previously described BLIND method, in which embryos are randomly collected and sorted by their transcriptomes (39). We collected 81 *A. nanus* embryos and processed each individually, using CEL-SEq (36), to obtain an expression matrix (Fig. 3*B*). For each embryo, we also noted the apparent morphological stage of development: one to eight cell-stages, ~30-cell stage, >30-cell stage, ventral enclosure, comma, or morphogenesis. Examining the transcriptomes using principal components analysis, we found that the overall ordering of the embryonic transcriptomes corresponded to the morphological stages (Fig. 3*B*). This principal components analysis on 1,314 dynamically expressed genes (*SI Experimental Procedures*) accounted for 49.8% (PC1) and 13.6% (PC2) of the gene expression variation. PC1 clearly captures developmental time, and PC2 distinguishes between the stages of the middevelopmental transition and the ends of embryogenesis. Thus, from randomly collected worm embryos, we obtained a time-course of expression throughout embryogenesis.

Studying the correlation among the transcriptomes, we found sharp developmental transitions (Fig. 3*C*). To annotate the stage of each transition, we compared with our morphological annotations and found that each transition corresponded to a shift between developmental stages (Fig. 3*D*). The first transition occurs after the likely degradation of the maternal transcriptome at the end of the 8-cell stage, and the next transition between early gastrulation (~30-cell stage) and midgastrulation (>30-cell stage). Another transition occurs at the end of the ventral enclosure stage. Finally, the comma stage was found to express a major transcriptomic transition after ventral enclosure and before morphogenesis. Thus, despite differences in the timing of embryonic development, we find a conservation in the pattern of gene expression transitions in *A. nanus* and *C. elegans* (10).

To validate the RNA-Seq data, we further examined the expression of homeodomain genes, known to play important developmental roles, between *A. nanus* and *C. elegans* (Fig. 4*A*). We found that although many genes are expressed at similar developmental stages between the two species, there were also some interesting divergences. One example is the *ceh-20* gene, which encodes one of the three *C. elegans* homeodomain proteins (CEH-20, CEH-40, and CEH-60) homologous to *Drosophila* Extradenticle (Exd/Pbx). In *C. elegans*, this gene is expressed during the ventral enclosure stage (40), whereas in *A. nanus*, the ortholog is expressed earlier, during the one to eight cell stage. To validate this difference, we performed an in situ for the *ceh-20* ortholog in *A. nanus* (Fig. 4*B*). The in situ confirmed the early *A. nanus* expression. Moreover, an additional in situ of the *ceh-34* gene, which is homologous to human SIX2, revealed expression consistent with our RNA-Seq analysis (Fig. 4*B*). This analysis further supports the credibility of the gene expression time-course.

Comparison of the *A. nanus* and *C. elegans* Developmental Transcriptomes. Seeking to compare the developmental transcriptomes of *A. nanus* and *C. elegans* in their entireties, we applied our previous approach in which dynamically expressed genes are first sorted according to their temporal expression (Fig. 5*A*) (2). Examining expression profiles of

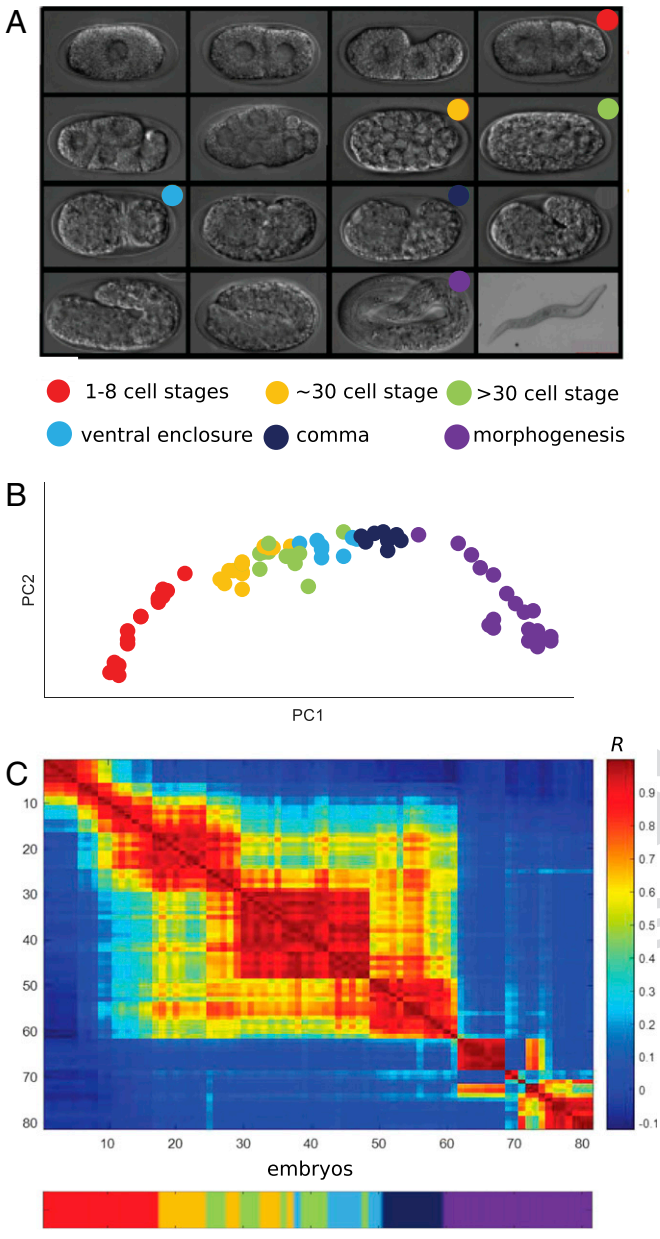

Fig. 3. A gene expression developmental time-course for *A. nanus* embryogenesis. (*A*) Micrographs of *A. nanus* embryos at the indicated stages. (*B*) RNA-Seq of 81 randomly collected *A. nanus* embryos. The embryos were sorted according to BLIND. (*C*) A correlation matrix of the BLIND-sorted *A. nanus* transcriptomes. Note the sharp transitions after the one to eight cell stages and then again at morphogenesis.

orthologous *C. elegans* genes, sorted according to expression of the corresponding *A. nanus* orthologs (Fig. 5*A*), we found an immediately apparent correspondence (Fig. 5*B*), suggesting general conservation of gene expression programs.

We asked whether gene expression at particular developmental stages is more evolvable than at other stages. To address this, we also sorted the *C. elegans* genes according to their temporal expression (Fig. 5*C*). For each pair of orthologs, we computed the difference between the relative order in which each gene appears in its respective time-course, which we refer to as the expression divergence index. We then examined whether at different stages of development, genes show different overall expression patterns between species. Proceeding from the earliest to the latest expression, we examined the distributions of expression divergence scores for *A. nanus* genes within the nonoverlapping windows shown in Fig. 5*D*.

As the distributions show, expression divergence is not uniform for genes expressed at different times. Genes expressed at the earliest stage may be considered maternal transcripts, and these appear to be highly divergent (Fig. 5*D*). The earliest zygotically expressed genes appear to be significantly more conserved in their expression (Fig. 5*D*, early) than the gastrula expressed genes, whereas genes expressed during the middevelopmental transition show significantly less divergence than those expressed at the gastrula stage ($P < 10^{-8}$, Wilcoxon test). Interestingly, this level of conservation continues throughout morphogenesis and does not increase, as would be expected from the hourglass model. This suggests a more complicated, funnel-like pattern of developmental constraints than previously recognized, although the reduction in divergence during the middevelopmental transition does mark a period of increased conservation, as expected.

Phylostratigraphic Analysis of Expression Divergence. Previous studies across animals separated by hundreds of millions of years of independent evolution has revealed that temporal expression of genes during animal development is correlated to the evolutionary age of genes (41, 42). We sought to investigate whether a similar pattern is observable between the closer-related clade IV and clade V nematode species examined here. For each pair of orthologs, we inferred the phylostratigraphic age (11), ranging from cellular life (common to all studied organisms) to Rhabditida, and restricted to this class of roundworms. To study whether genes differed in their evolvability throughout development, we studied the distributions of expression divergence for each class of gene ages (Fig. 6).

We observed a restriction of expression divergence for genes originating at superphylum, phylum, and class levels within Nematoda. The sample sizes did not allow for direct statistical comparisons of phylostratigraphic nodes. However, a Wilcoxon ranks-sum test confirmed that the distributions were significantly different between neighboring phylostrata for genes that evolved

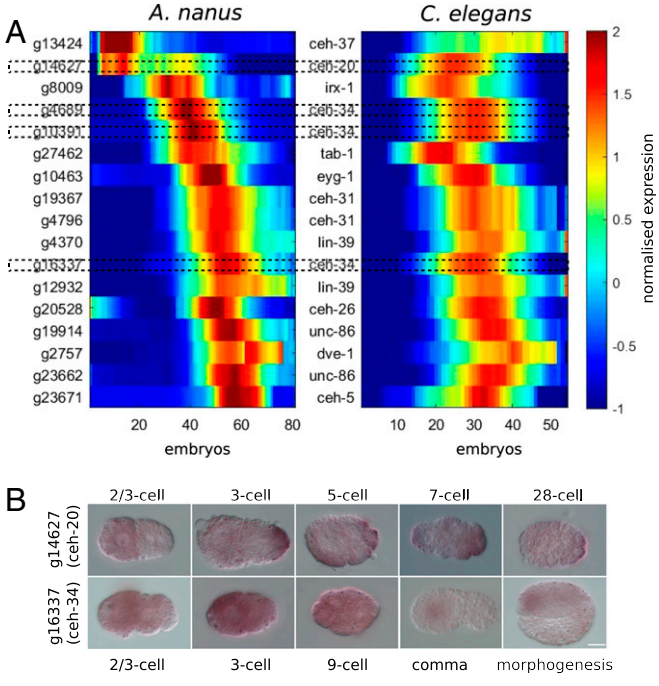

Fig. 4. Expression of homeodomain genes between *A. nanus* and *C. elegans*. (*A*) Comparison of temporal expression of selected orthologous genes in *A. nanus* and *C. elegans*. Specific homeodomain genes that were further analyzed by in situ (*B*) are emphasized with dotted outlines. (*B*) In situ hybridizations for *ceh-20* and *ceh-34* orthologs in *A. nanus*.

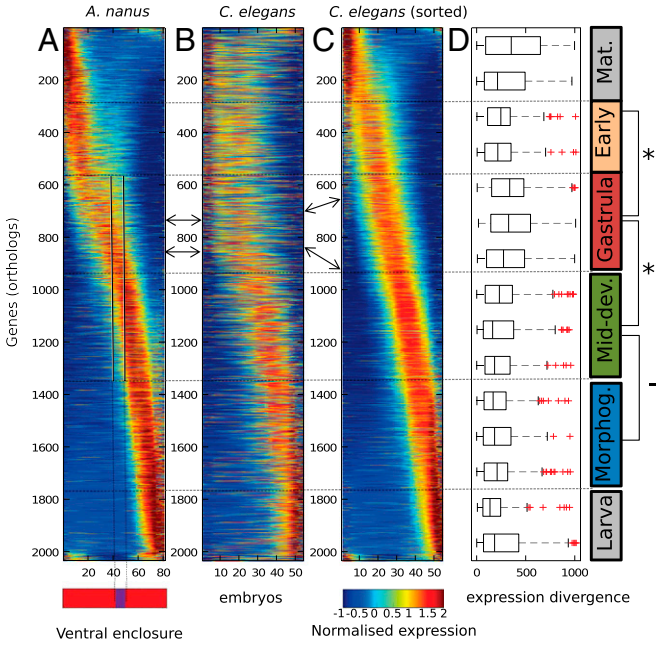

Fig. 5. Expression divergence between the developmental transcriptomes of *C. elegans* and *A. nanus*. (*A*) Developmental transcriptome of *A. nanus*. Genes are sorted by the Zavit method (2). (*B*) Developmental transcriptome of the *C. elegans* orthologs of *A. nanus*, sorted as in *A. nanus*. Arrows indicate orthologs. (*C*) Developmental transcriptome of the *C. elegans* orthologs sorted independent of *A. nanus*. Arrows indicate corresponding genes, sorted in *C* according to *C. elegans* time. (*D*) Box plots indicating the expression divergences between genes in *A* and *C* for stages along development. Developmental stages are indicated on the right (Mat., maternal; early; gastrula; Mid-dev., middevelopmental transition; Morphog., morphogenesis; and larva). Note the increased relative conservation of genes expressed early and at middevelopmental transition.

at the base of the superphylum Ecdysozoa and the phylum Nematoda (Fig. 6). Thus, in addition to genes expressed at or after the middevelopmental transition, genes originating at the dawn of the Nematode phylum are also more conserved in their expression across species than expected.

We hypothesized that the reason that genes of distinct phylostratigraphic ages are conserved in their gene expression between species at different levels follows from their expression at distinct periods during embryogenesis. In other words, if genes of different ages are expressed at different developmental stages, then their expression would evolve at different rates following our results shown in Fig. 5. Interestingly, we found that deeply conserved genes are expressed early in both *C. elegans* and *A. nanus*. Meanwhile, genes specific to the *Chromadorea* class or more specific taxa (*SI Experimental Procedures*) are expressed later in development, during differentiation (Fig. 6*B*). However, genes that originated in the metazoan and the superphylum ecdysozoan are expressed during the middevelopmental transition. We further tested this result by examining the expression of genes of different ages in the recently published developmental transcriptome of the parasitic clade IV species *Steinernema carpocapsae* (43). Again, we found the same pattern (Fig. 6*B*), suggesting that a relationship between phylogenetic age and developmental expression may be general to the Nematode phylum.

Discussion

In this work, we compared the developmental transcriptomes of two distantly related nematodes. *C. elegans* is a clade V nematode of the Rhabditoidea superfamily, whereas *A. nanus* belongs to the Cephaloboidea superfamily within clade IV. The lineages of both species most likely diverged not more than 200 million ago (44). Although the embryogenesis of *A. nanus* has been analyzed in classical cell biological studies, here we report for the first time its genome, transcriptome, and developmental gene regulation. Compared with *C. elegans*, we found important differences at the two-cell stage, in terms of transcription factor expression during the course of development and the overall pattern of development. We also compared the divergence in gene expression in terms of the phylostratigraphy and found that genes specific to Nematodes and the Ecdysozoa superphylum are more conserved. In this section, we discuss our results in light of the methodologies for evaluating developmental transcriptomics, the middevelopmental transition, developmental constraints, and phylostratigraphy.

As in other species examined by transcriptomics, we identified a clear middevelopmental transition in *A. nanus*, depicted as a sharp transition in the heat map of correlations between transcriptomes. We also observed that at this stage in development, the transcriptomes of *C. elegans* and *A. nanus* begin to converge. Interestingly, the transcriptomes do not diverge in an hourglass shape after the middevelopmental transition, as was initially suggested for vertebrates (45), and later for a variety of invertebrates (46) and plants (47). This is similar to a previous observation of two frog species (48) that converged at the tailbud stage (the phylotypic stage of chordates) and then did not diverge again. This may be a result of the large number of cell types expressed at this stage. These results also somewhat mirror what was seen when examining mutation accumulation strains of *C. elegans* (2), as well as the results of a recent study examining the developmental transcriptomes of the parasitic clade IV species *S. carpocapsae* (43).

Our phylostratigraphical analysis shows that genes that emerged during the origin of the superphylum Ecdysozoa and Nematoda are more conserved in their developmental expression. We found

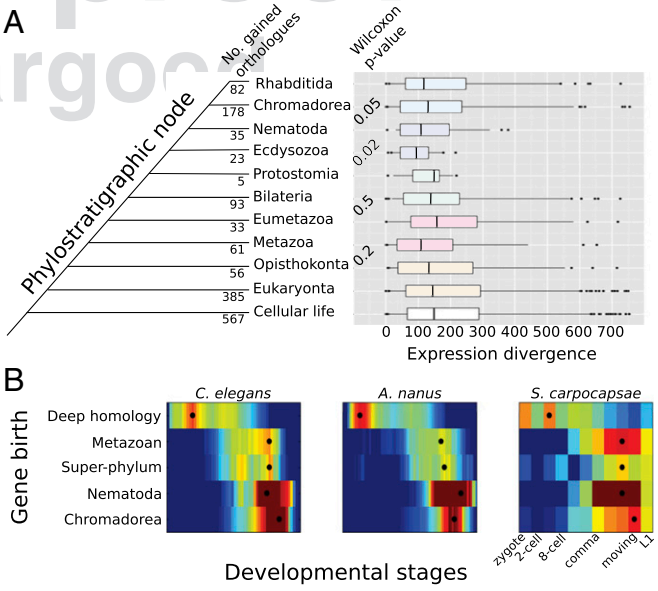

Fig. 6. Ecdysozoan- and Nematode-specific genes are more conserved in their expression between *C. elegans* and *A. nanus*. (*A*) Genes were grouped according to their phylostratigraphic age (*Left*, see *SI Experimental Procedures*). Expression divergence index (ED) of *C. elegans* and *A. nanus* orthologs in comparison with their phylostratigraphic age. Phylostratigraphic age was calculated by blasting against a previously reported database (47) using the Phylostratigraphy software (https://github.com/AlexGa/Phylostratigraphy.git). A statistical test of difference in ED distributions for phylostratigraphic nodes revealed significance of divergence for comparisons in Nematoda, but not for genes that evolved before the phylum. The ED appears to follow an hourglass shape through evolutionary time, with evolutionary very old and young genes showing less constrained ED than those acquired on intermediate nodes in Nematoda. (*B*) Average expression profiles of genes of a common phylostratigraphic age for the three indicated species. Black dots indicate the stage for each category at which average expression is at its maximum.

that this may follow from a relationship between the age of a gene and its expression during development. Although Domazet-Lošo and Tautz also found that a middevelopmental stage has an overall older transcriptome when computed by the transcriptomic age index (42), we found that genes of older origin tend to be expressed early in development. We attribute this difference to us having studied separately groups of genes of distinct ages, rather than combining ages for an age of the transcriptome. In our analysis, genes of the superphylum and phylum age category are enriched for expression during the middevelopmental transition. This suggests that genetic pathways originating at the dawn of the Ecdysozoa superphylum are more conserved in their expression program during embryogenesis because they have been integrated into the more conserved middevelopmental transition stage.

Importantly, our finding that genes of intermediate evolutionary age show a restriction in their developmental expression divergence is in line with the inference that these genes are definitive of superphyla and phyla within the sphere of animal diversity. Moreover, it has been argued that taxon specific ("orphan") genes contribute most to the differentiation of developmental between taxa (49). Thus, our evidence that evolutionarily young genes are more variable in their developmental expression and expressed at later stages, indeed suggests that these genes drive the differentiation of developmental gene expression programs. Our detailed study of the developmental gene expression and genome of *A. nanus* will allow for detailed comparative studies into these patterns, and enable deeper insights into the evolvability and constraint of molecular pathways in animal development.

Experimental Procedures

We used Illumina technology to sequence the *A. nanus* genome and transcriptome, and followed the CEL-seq protocol to establish a developmental time course and single blastomere transcriptomes. Details of the procedure and the short-read cleaning and assembly pipelines can be found in the *Supporting Information*. We annotated the genome with Augustus, inferred orthology with Orthofinder, and analyzed expression data using Matlab, R, and Python as described in the *Supporting Information*.

ACKNOWLEDGMENTS. We gratefully acknowledge the support and inspiration of Einhard Schierenberg throughout this project. We acknowledge the Technion Genome Center for assistance with the sequencing. We thank Maayan Baron for help in the processing of the data. We thank Ali Mortazavi and Marissa Macchietto for kindly providing supplementary data on orthologous genes between *C. elegans* and *S. carpocapsae*. P.H.S. was funded by the VolkswagenStiftung in their initiative for evolutionary biology and was supported by a grant from the European Research Council (ERC-2012-AdG 322790 to Max Telford).

1. Gould SJ, Lewontin RC (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* 205:581–598.
2. Zalts H, Yanai I (2017) Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nat Ecol Evol* 1:113.
3. Arthur W (1988) *A Theory of the Evolution of Development* (John Wiley & Sons Incorporated, Chichester).
4. Valentine JW, Jablonski D, Erwin DH (1999) Fossils, molecules and embryos: New perspectives on the Cambrian explosion. *Development* 126:851–859.
5. Angelini DR, Kaufman TC (2005) Comparative developmental genetics and the evolution of arthropod body plans. *Annu Rev Genet* 39:95–119.
6. Lynch M (2007) *The Origins of Genome Architecture* (Sinauer Associates Incorporated, Sunderland, MA).
7. Yanai I, Lercher M (2016) *The Society of Genes* (Harvard Univ Press, Cambridge, MA).
8. Slonim DK, Yanai I (2009) Getting started in gene expression microarray analysis. *PLoS Comput Biol* 5:e1000543.
9. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5:621–628.
10. Levin M, Hashimshony T, Wagner F, Yanai I (2012) Developmental milestones punctuate gene expression in the Caenorhabditis embryo. *Dev Cell* 22:1101–1108.
11. Domazet-Lošo T, Brajković J, Tautz D (2007) A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet* 23:533–539.
12. Levin M, et al. (2016) The mid-developmental transition and the evolution of animal body plans. *Nature* 531:637–641.
13. Schulze J, Schierenberg E (2011) Evolution of embryonic development in nematodes. *Evodevo* 2:18.
14. Goldstein B, Frisse LM, Thomas WK (1998) Embryonic axis specification in nematodes: Evolution of the first step in development. *Curr Biol* 8:157–160.
15. Laugsch M, Schierenberg E (2004) Differences in maternal supply and early development of closely related nematode species. *Int J Dev Biol* 48:655–662.
16. Skiba F, Schierenberg E (1992) Cell lineages, developmental timing, and spatial pattern formation in embryos of free-living soil nematodes. *Dev Biol* 151:597–610.
17. Schierenberg E (2001) Three sons of fortune: Early embryogenesis, evolution and ecology of nematodes. *BioEssays* 23:841–847.
18. Houthoofd W, et al. (2006) Different roads to form the same gut in nematodes. *Evol Dev* 8:362–369.
19. Houthoofd W, et al. (2003) Embryonic cell lineage of the marine nematode Pellioditis marina. *Dev Biol* 258:57–69.
20. Blaxter ML, et al. (1998) A molecular evolutionary framework for the phylum Nematoda. *Nature* 392:71–75.
21. Wiegner O, Schierenberg E (1999) Regulative development in a nematode embryo: A hierarchy of cell fate transformations. *Dev Biol* 215:1–12.
22. Schiffer PH, et al. (2014) Developmental variations among Panagrolaimid nematodes indicate developmental system drift within a small taxonomic unit. *Dev Genes Evol* 224:183–188.
23. Schiffer PH, et al. (2013) The genome of Romanomermis culicivorax: Revealing fundamental changes in the core developmental genetic toolkit in Nematoda. *BMC Genomics* 14:923.
24. Doroszuk A, Wojewodzic M, Kammenga J (2006) Rapid adaptive divergence of life-history traits in response to abiotic stress within a natural population of a parthenogenetic nematode. *Proc Biol Sci* 273:2611–2618.
25. Bird AF, Ryder MH (1993) Feeding of the nematode Acrobeloides nanus on bacteria. *J Nematol* 25:493–499.
26. Heger P, Kroiher M, Ndifon N, Schierenberg E (2010) Conservation of MAP kinase activity and MSP genes in parthenogenetic nematodes. *BMC Dev Biol* 10:51.
27. Arkhipova I, Meselson M (2005) Deleterious transposable elements and the extinction of asexuals. *BioEssays* 27:76–85.
28. Bast J, et al. (2016) No accumulation of transposable elements in asexual arthropods. *Mol Biol Evol* 33:697–706.
29. Schiffer PH, et al. (July 3, 2017) Signatures of the evolution of parthenogenesis and cryptobiosis in the genomes of panagrolaimid nematodes. *bioRXiv*, 10.1101/159152.
30. Lynch M, Conery JS (2003) The origins of genome complexity. *Science* 302:1401–1404.
31. Szitenberg A, et al. (2016) Genetic drift, not life history or RNAi, determine long term evolution of transposable elements. *Genome Biol Evol* 8:2964–2978.
32. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
33. Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19:ii215–ii225.
34. Emms DM, Kelly S (2015) OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157.
35. Wittenburg N, Baumeister R (1999) Thermal avoidance in Caenorhabditis elegans: An approach to the study of nociception. *Proc Natl Acad Sci USA* 96:10477–10482.
36. Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2:666–673.
37. Schierenberg E (2006) Embryological variation during nematode development. *WormBook* 1–13.
38. Schulze J (2008) Vergleichende Untersuchungen zur Embryonalentwicklung basaler und abgeleiteter Nematoden. Ein Beitrag zum Verständnis der Evolution von Entwicklungsprozessen. PhD thesis (University of Cologne, Cologne, Germany).
39. Anavy L, et al. (2014) BLIND ordering of large-scale transcriptomic developmental timecourses. *Development* 141:1161–1166.
40. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I (2015) Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* 519:219–222.
41. Moyers BA, Zhang J (2015) Phylostratigraphic bias creates spurious patterns of genome evolution. *Mol Biol Evol* 32:258–267.
42. Domazet-Lošo T, Tautz D (2010) A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* 468:815–818.
43. Macchietto M, et al. (2017) Comparative transcriptomics of Steinernema and Caenorhabditis single embryos reveals orthologous gene expression convergence during late embryogenesis. *Genome Biol Evol* 9:2681–2696.
44. Rota-Stabelli O, Daley AC, Pisani D (2013) Molecular timetrees reveal a Cambrian colonization of land and a new scenario for ecdysozoan evolution. *Curr Biol* 23:392–398.
45. 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. *Dev Suppl* 135–142.
46. Kalinka AT, Tomancak P (2012) The evolution of early animal embryos: Conservation or divergence? *Trends Ecol Evol* 27:385–393.
47. Drost HG, Gabel A, Grosse I, Quint M (2015) Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol* 32:1221–1231.
48. Yanai I, Peshkin L, Jorgensen P, Kirschner MW (2011) Mapping gene expression in two Xenopus species: Evolutionary constraints and developmental flexibility. *Dev Cell* 20:483–496.
49. Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. *Nat Rev Genet* 12:692–702.
50. Kikuchi T, et al. (2011) Genomic insights into the origin of parasitism in the emerging plant pathogen Bursaphelenchus xylophilus. *PLoS Pathog* 7:e1002219.

Supporting Information

Schiffer et al. 10.1073/pnas.1720817115

SI Experimental Procedures

Genome and Transcriptome Sequencing and Assembly. *A. nanus* (strain ES501) was kindly provided by Einhard Schierenberg, and then cultured at 25 °C on minimal agar plates, as described in ref. 1. We used the Illumina GaIIx and HiSeq platforms to generate paired end and mate pair reads with differing insert size from extracted DNA of many individuals. We analyzed the obtained read sets with FastQC (v0.10.1) and removed residual adapters and low-quality bases with Trimmomatic (v0.33) (2). We explored differing assembly pipelines and found SPAdes (v 3.9) (3) to give the best initial assembly results. To scaffold we choose the redundans pipeline (4), which incorporates GapFiller (5) and SSPACE (6) in an iterative way. Finally, we used a Trinity (7) assembly of RNA-Seq data to extend our scaffolding with SCUBAT2 (https://github.com/GDKO/SCUBAT2.git). Because nematode genomes are very often contaminated with sequences stemming from bacteria the animals feed on, we used Blobtools (8) to screen for contamination. We then removed the most abundant (measured in megabases) contigs with best blast hits to Proteobacteria, Actinobacteria, Cyanobacteria, Streptophyta, Ascomycota, Bacteroidetes, and Spirochetes. We sequenced mRNA across all life cycle stages using Illumina GaIIx and HiSeq machines after the general Illumina RNA-Seq protocol. We then used the Trinity pipeline to assemble the reads into a set of transcriptomic contigs.

Genome Annotation. We used BUSCO3 through the gVolante web service (https://gvolante.riken.jp) to check genome completeness. We relied on Augustus (v. 3.2.2) to annotate the *A. nanus* genome. To improve the Augustus predictions, we used our RNA-Seq data and incorporated repeats found with RepeatModeller (www.repeatmasker.org/RepeatModeler/) and masked with RepeatMasker (9). For RNA-Seq guided annotation, we followed the respective protocols on the Augustus wiki by using gmap/gsnap (v.2016–06-09) (10) to map RNA-Seq reads, incorporating SAMtools (11) and BAMtools (12) when Augustus hints are created. We set *C. elegans* as the species profile for Augustus.

Orthology Inference. We used Orthofinder (v.1.0.8) (13) to screen for orthologous proteins between *A. nanus* and *C. elegans*. To allow for links to be established along the phylogeny, we further included the second nematode model *Pristionchus pacificus* (clade V), as well as *Bursaphelenchus xylophylus*, *Meloidogyne hapla*, *Panagrellus redivivus* (all clade IV), and *Ascaris suum* from clade III as a remote outgroup. Instead of NCBI BLAST+, we used the DIAMOND blast approach (14) in the initial any versus any blast step of Orthofinder. The phylogeny among these species is well resolved, and we thus relied on the simple gene trees to species tree algorithm implemented in Orthofinder instead of implementing more sophisticated phylogenetic programs.

Protein Domain Annotation. We employed InterProScan (v.5.19–58.0) (15) in a local standalone version to screen the *A. nanus* and *C. elegans* (Wormbase version PRJNA13758) proteomes for Pfam (16) and PANTHER (17) annotations. GO terms (18) were retrieved as part of the PANTHER families.

Phylostratigraphy. To retrieve a phylostratigraphic annotation of the Augustus-predicted *A. nanus* proteins set and the *C. elegans* protein set downloaded from Wormbase, we used the Phylostratigraphy pipeline from https://github.com/AlexGa/Phylostratigraphy.git. The algorithm natively implements BLAST (19) searches against the Phylostratigraphy database from ref. 20 and subsequently orders the proteins according to the phylostratigraphic nodes based on best hits. In our assay, we replaced the BLAST+ searches by the faster, but highly sensitive, DIAMOND software.

RNA-Seq Developmental Time-Course. Individual *A. nanus* nematodes were placed on 60 mm minimal agar plates seeded with OP50 until a few embryos were observed to have hatched, at which point all embryos on a plate were collected. One hundred twenty-four embryos were collected, which spanned the course of development beginning at the single-cell stage through just before hatching. For *C. elegans*, we used a previous dataset (21). Each individual embryo was placed in 1 uL water on the cap of a microcentrifuge tube and then frozen in liquid nitrogen. Samples were stored at −80 °C until all samples were collected. Total RNA was extracted from individual embryos (samples were not pooled) at 1/5 the recommended volume using TRIzol (Invitrogen). LPA and tRNA were added to help precipitate and visualize pellets, as well as 1 uL of the ERCC spike-in kit (22) at a 1:500,000 dilution to help in quantification of amplified RNA. The TRIzol mix was added to each sample, and then frozen in liquid nitrogen and thawed in a 42 °C water bath five times immediately after adding TRIzol to ensure disruption of the chitinous egg shell. RNA isolation then proceeded according to ref. 23. Isolated RNA was eluted in ultrapure water and a uniquely barcoded primer for reverse transcription, and then half of the elution was amplified according to the CEL-seq protocol (24) and then sequenced on the Illumina HISeq2000 at the Technion Genome Center. To analyze only the high-quality embryo RNA-Seq samples, we filtered out those samples with less than 600,000 transcripts, leading to an 81-embryo sample (analyzed first in Fig. 3).

Single Cell RNA-Seq of Blastomeres. *A. nanus* blastomeres were isolated according to the methods of Edgar and Goldstien (2012), with the following modifications. All solutions were prepared with 2× salt concentrations with respect to the original recipes for *C. elegans*. After collection of fertilized eggs from gravid adult worms, the external chorion was removed by incubation in 2× bleach for 5 min, followed by an 8–12-min treatment in chitanase. As *A. nanus* blastomeres are connected by cytoplasmic bridges, individual cells from the two- and three-cell stages were separated from one another mechanically, using a fine pulled-glass needle. Both dechorionated embryos and isolated blastomeres that were cultured overnight in 2×-salt EGM developed into small juvenile worms. On dissociation, relative cell sizes were noted for identification purposes, and all cells from a single embryo were flash frozen individually in liquid nitrogen. Blastomeres were collected only from embryos where all cells survived the isolation procedure. The blastomere collection was processed for single-cell RNA-sequencing according to the CEL-Seq protocol (24), with the addition of unique molecular identifiers within the CEL-Seq2 primers (25).

In Situ Hybridization. In situ hybridization was performed according to the freeze crack procedure described for *C. elegans* (26) and modifications given by refs. 27 and 28. Before freeze cracking, the egg shell of *A. nanus* has been partly removed by incubation in alkaline-bleach solution (4.5% NaOCl and 0.75 M KOH) for about 90 s.

Digoxigenin-labeled sense and antisense RNA probes were prepared from linearized pBluescript vectors (Stratagene) containing a fragment of the *A. nanus* homologs of *C. elegans ceh-20* (g14627.t1) and *ceh-34* (g16337.t1) genes via run off in vitro transcription with T7 or T3 RNA-polymerase (Roche). *A. nanus ceh-20* and *ceh-34* fragments were amplified by PCR from *A. nanus* cDNA, cloned into pBs vector, and verified by Sanger Sequencing.

Steinernema Gene Expression Analysis. Expression data and orthologous mappings were retrieved from a recent publication (29). The phylostratigraphic groups of Steinernema genes were transferred from their *C. elegans* orthologs. Expression of transcriptomes triplicates were averaged by computing the median value of the log transformed data. Of the 2,464 one to one *C. elegans* and *S. carpocapsae* orthologs, we selected those 1,143 orthologs with overall expression higher than 6 average log10 units. We then normalized the expression using transcripts per million, as in *C. elegans* and *A. nanus* analyses. We collapsed the phylostratigraphic categories into five broader categories: deep homology, which includes cellular organisms, eukaryota, and opisthokonta; metazoan, which includes metazoa, eumetazoa, and bilateria; superphylum, which includes protostomia and ecdysozoa; Nematoda and Chromadorea are simply Nematoda and Chromadorea, respectively. To estimate the expression profile of the set of genes of each phylostratigraphic group, we computed the mean of the Z-score-normalized gene expression profiles of genes with that phylostratigraphic age.

1. Lahl V, Halama C, Schierenberg E (2003) Comparative and experimental embryogenesis of Plectidae (Nematoda). *Dev Genes Evol* 213:18–27.
2. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120.
3. Bankevich A, et al. (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477.
4. Pryszcz LP, Gabaldón T (2016) Redundans: An assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* 44:e113.
5. Boetzer M, Pirovano W (2012) Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56.
6. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W (2011) Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27:578–579.
7. Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512.
8. Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M (2013) Blobology: Exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. *Front Genet* 4:237.
9. Smit A, Hubley R, Green P RepeatMasker Open-3.0.1996-2010.
10. Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873–881.
11. Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
12. Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT (2011) BamTools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692.
13. Emms DM, Kelly S (2015) OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157.
14. Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60.
15. Jones P, et al. (2014) InterProScan 5: Genome-scale protein function classification. *Bioinformatics* 30:1236–1240.
16. Sonnhammer EL, Eddy SR, Birney E, Bateman A, Durbin R (1998) Pfam: Multiple sequence alignments and HMM-profiles of protein domains. *Nucleic Acids Res* 26:320–322.
17. Mi H, Muruganujan A, Thomas PD (2013) PANTHER in 2013: Modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 41:D377–D386.
18. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25:25–29.
19. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410.
20. Drost HG, Gabel A, Grosse I, Quint M (2015) Evidence for active maintenance of phylotranscriptomic hourglass patterns in animal and plant embryogenesis. *Mol Biol Evol* 32:1221–1231.
21. Hashimshony T, Feder M, Levin M, Hall BK, Yanai I (2015) Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer. *Nature* 519:219–222.
22. Baker SC, et al.; External RNA Controls Consortium (2005) The External RNA Controls Consortium: A progress report. *Nat Methods* 2:731–734.
23. Baugh LR, Hill AA, Slonim DK, Brown EL, Hunter CP (2003) Composition and dynamics of the Caenorhabditis elegans early embryonic transcriptome. *Development* 130:889–900.
24. Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: Single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2:666–673.
25. Wiegner O, Schierenberg E (1999) Regulative development in a nematode embryo: A hierarchy of cell fate transformations. *Dev Biol* 215:1–12.
26. Schierenberg E (2001) Three sons of fortune: Early embryogenesis, evolution and ecology of nematodes. *BioEssays* 23:841–847.
27. Doroszuk A, Wojewodzic M, Kammenga J (2006) Rapid adaptive divergence of life-history traits in response to abiotic stress within a natural. *Proc Biol Sci* 273:2611–2618.
28. Bird AF, Ryder MH (1993) Feeding of the nematode Acrobeloides nanus on bacteria. *J Nematol* 25:493–499.
29. Heger P, Kroiher M, Ndifon N, Schierenberg E (2010) Conservation of MAP kinase activity and MSP genes in parthenogenetic nematodes. *BMC Dev Biol* 10:51.

Table S1. Repetitive elements of the *A. nanus* genome

	Number of elements*	Length occupied, bp	Percentage of sequence
SINEs	26,302	4,059,848	1.51
ALUs	0	0	0
MIRs	933	116,205	0.04
LINEs	8,105	943,138	0.35
LINE1	3,264	322,726	0.12
LINE2	0	0	0
L3/CR1	456	138,845	0.05
LTR elements	22,913	2,761,436	1.02
ERVL	0	0	0
ERVL-MaLRs	0	0	0
ERV_classI	3,521	578,653	0.21
ERV_classII	323	30,967	0.01
DNA elements	127,682	16,061,090	5.96
hAT-Charlie	0	0	0
TcMar-Tigger	0	0	0
Unclassified	796,066	113,351,783	42.05
Total interspersed repeats		137,177,295	50.89
Small RNA	1,848	231,160	0.09
Satellites	997	132,877	0.05
Simple repeats	46,280	4,210,287	1.56
Low complexity	4,387	237,029	0.09

In this study, 139,424,278 bp were masked (51.72%). The table provides their composition.

*Most repeats fragmented by insertions or deletions have been counted as one element.