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. Gene ontology 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). Linear polyacrylamide and tRNA were added to help precipitate and visualize pellets, as well as 1 uL of the External RNA Control Consortium 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 Goldstein (25), 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 embryonic growth medium 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 (26).

In Situ Hybridization. In situ hybridization was performed according to the freeze crack procedure described for *C. elegans* (27) and modifications given by (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 AFA, Hubley R, Green P (2015) RepeatMasker Open-4.0.2013-2015. Available at www.repeatmasker.org. Accessed November 1, 2016.
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. Edgar LG, Goldstein B (2012) Culture and manipulation of embryonic cells. *Methods Cell Biol* 107:157–175.
26. Hashimshony T, et al. (2016) CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* 17:77.
27. Seydoux G, Fire A (1995) Whole-mount in situ hybridization for the detection of RNA in Caenorhabditis elegans embryos. *Methods Cell Biol* 48:323–337.
28. Broitman-Maduro G, Maduro MF (2011) In situ hybridization of embryos with antisense RNA probes. *Methods Cell Biol* 106:253–270.
29. 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.

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

Type of element	No. 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.