



**HAL**  
open science

## Signatures of the evolution of Parthenogenesis and Cryptobiosis in the Genomes of Panagrolaimid Nematodes

Philipp H. Schiffer, Etienne Danchin, Ann M. Burnell, Christopher J. Creevey,  
Simon Wong, Ilona Dix, Georgina O'Mahony, Bridget A. Culleton, Corinne  
Rancurel, Gary Stier, et al.

► **To cite this version:**

Philipp H. Schiffer, Etienne Danchin, Ann M. Burnell, Christopher J. Creevey, Simon Wong, et al..  
Signatures of the evolution of Parthenogenesis and Cryptobiosis in the Genomes of Panagrolaimid  
Nematodes. *iScience*, 2019, 21, pp.587-602. 10.1016/j.isci.2019.10.039 . hal-02444450

**HAL Id: hal-02444450**

**<https://hal.science/hal-02444450v1>**

Submitted on 17 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

## Article

# Signatures of the Evolution of Parthenogenesis and Cryptobiosis in the Genomes of Panagrolaimid Nematodes

Philipp H. Schiffer,  
Etienne G.J.  
Danchin, Ann M.  
Burnell, ..., Einhard  
Schierenberg,  
Thomas Wiehe,  
Mark Blaxter

philipp.schiffer@gmail.com

## HIGHLIGHTS

Parthenogenesis in  
*Panagrolaimus*  
nematodes arose once  
through a hybridization  
event

Parthenogenetic  
*Panagrolaimus* are triploid  
but retain meiosis after  
millions of years

*Panagrolaimus* can be  
used as a model system to  
study aspects of animal  
parthenogenesis

The genomes of  
*Panagrolaimus*  
nematodes are shaped by  
their adaptation to  
cryptobiosis

Schiffer et al., iScience 21,  
587–602  
November 22, 2019 © 2019  
The Authors.  
[https://doi.org/10.1016/  
j.isci.2019.10.039](https://doi.org/10.1016/j.isci.2019.10.039)



## Article

# Signatures of the Evolution of Parthenogenesis and Cryptobiosis in the Genomes of Panagrolaimid Nematodes

Philipp H. Schiffer,<sup>1,2,12,14,\*</sup> Etienne G.J. Danchin,<sup>3</sup> Ann M. Burnell,<sup>4</sup> Christopher J. Creevey,<sup>5</sup> Simon Wong,<sup>6</sup> Llona Dix,<sup>4</sup> Georgina O'Mahony,<sup>4</sup> Bridget A. Culleton,<sup>4,7</sup> Corinne Rancurel,<sup>3</sup> Gary Stier,<sup>2</sup> Elizabeth A. Martínez-Salazar,<sup>8</sup> Aleksandra Marconi,<sup>9</sup> Urmi Trivedi,<sup>10</sup> Michael Kroiher,<sup>2</sup> Michael A.S. Thorne,<sup>11</sup> Einhard Schierenberg,<sup>2</sup> Thomas Wiehe,<sup>12</sup> and Mark Blaxter<sup>9,10,13</sup>

## SUMMARY

**Most animal species reproduce sexually and fully parthenogenetic lineages are usually short lived in evolution. Still, parthenogenesis may be advantageous as it avoids the cost of sex and permits colonization by single individuals. Panagrolaimid nematodes have colonized environments ranging from arid deserts to Arctic and Antarctic biomes. Many are obligatory meiotic parthenogens, and most have cryptobiotic abilities, being able to survive repeated cycles of complete desiccation and freezing. To identify systems that may contribute to these striking abilities, we sequenced and compared the genomes and transcriptomes of parthenogenetic and outcrossing panagrolaimid species, including cryptobionts and non-cryptobionts. The parthenogens are triploids, most likely originating through hybridization. Adaptation to cryptobiosis shaped the genomes of panagrolaimid nematodes and is associated with the expansion of gene families and signatures of selection on genes involved in cryptobiosis. All panagrolaimids have acquired genes through horizontal gene transfer, some of which are likely to contribute to cryptobiosis.**

## INTRODUCTION

Despite the general prevalence of sexual reproduction in animals, parthenogenesis has evolved in many taxa. Obligatory parthenogenesis, or more strictly the cessation of outcrossing after meiotic recombination, is thought to result in the gradual accumulation of deleterious mutations that cannot be purged independently of all linked loci (Lynch et al., 1993). This would condemn non-recombining parthenogenetic animals to evolutionary dead-ends through Muller ratchet (Felsenstein, 1974) and predicts restriction of these species to terminal phylogenetic branches. The evolutionary scandals of the phylogenetic persistence, and abundant speciation, of some asexual taxa (such as bdelloid rotifers and darwinulid ostracods) may be explained by the recruitment of recombination-like mechanisms (Signorovitch et al., 2015). Up to 38% of species in some arthropod genera may be parthenogenetic (Van Der Kooi et al. 2017), and this mode of reproduction has arisen frequently in nematodes (Denver et al., 2011). Parthenogens have been frequently observed to amplify their genomes through allo- or autopolyploidy, and this may buffer lineages against mutation accumulation in essential genes (Archetti, 2004). Allopolyploid genomes might also enable diversification of gene copies by associating several adapted and diverged genomes in a single hybrid species (Blanc-Mathieu et al., 2017). Comparisons of related sexual and asexual taxa are required to understand the origins and persistence of asexuality and the effects of asexuality on genome evolution.

Asexual reproduction has long been associated with organisms found in extreme or peripheral environments, such as arid deserts where water availability is ephemeral and unpredictable, or polar ecosystems where much of the annual cycle involves temperatures well below zero degree Celsius (Vrijenhoek and Davis Parker, 2009). Because they do not need to find a partner to mate and reproduce, parthenogens can colonize ephemeral or widely spaced niches as single reproductive propagules (Lynch, 1984). Organisms that inhabit extreme environments are often also cryptobionts. Cryptobiosis is the ability of an organism to survive with no evidence of active metabolism under conditions inimical to life. Cryptobionts survive freezing, avoiding lethal ice-crystal disruption of biological membranes, whereas anhydrobionts survive loss of effectively all their body water. Cryptobionts are frequently also anhydrobiotic, as ice

<sup>1</sup>CLOE, Department for Biosciences, University College London, London, UK

<sup>2</sup>Zoologisches Institut, Universität zu Köln, 50674 Köln, Germany

<sup>3</sup>INRA, Université Côte d'Azur, CNRS, ISA, France

<sup>4</sup>Maynooth University Department of Biology, National University of Ireland Maynooth, Maynooth, Co. Kildare, Ireland

<sup>5</sup>School of Biological Sciences, Queen's University Belfast, Belfast, UK

<sup>6</sup>Irish Centre for High-End Computing, Tower Building, Trinity Technology & Enterprise Campus, Grand Canal Quay, Dublin D02 HP83, Ireland

<sup>7</sup>Megazyme, Bray Business Park, Bray, Co. Wicklow A98 YV29, Ireland

<sup>8</sup>Unidad Académica de Ciencias Biológicas, Laboratorio de Colecciones Biológicas y Sistemática Molecular, Universidad Autónoma de Zacatecas, Zacatecas, México

<sup>9</sup>Institute of Evolutionary Biology, The University of Edinburgh, Edinburgh EH9 3FL, UK

<sup>10</sup>Edinburgh Genomics, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3FL, UK

<sup>11</sup>British Antarctic Survey, Natural Environment Research Council, High Cross, Madingley Road, Cambridge CB3 0ET, UK

<sup>12</sup>Institut für Genetik, Universität zu Köln, 50674 Köln, Germany

Continued



Species	Span (Mbp)	Ns (k) % of Total	No. Scaffolds (k)	Max. Scaffold Size (k)	N50 Scaffolds	Protein Models (k)	Repeats (%)	GC (%)	BUSCO C + P (%) <sup>a</sup>
<i>Panagrolaimus superbis</i>	76	1.095 (1,4)	53.2	29	1,894	24.2	8,4	31,8	96
<i>P. sp. ES5</i>	90	1.578 (1,7)	25.2	96.2	6,312	24.8	8	29	94
<i>P. sp. PS1159</i>	85	1.702 (2)	17.6	142.9	9,924	27.4	6,8	28,2	94
<i>P. sp. PS1579</i>	49	1.185 (2,4)	23.4	24.9	2,935	18 (47.7) <sup>b</sup>	4,2	30,5	95
<i>P. sp. DAW1</i>	118	73 (0.06)	39.6	42.2	4,454	31.6	10,3	27,8	68
<i>Propanagrolaimus sp. JU765</i>	64	807 (1,3)	13.3	907.5	10,861	24.9	9,1	31,8	94

**Table 1. Assembly and Annotation Metrics for the *Panagrolaimus* and *Propanagrolaimus* Genomes Presented in This Study**

Data are given for scaffolds >500 bp and contigs >100 bp (scaffolds split at  $\geq 10$  Ns).

<sup>a</sup>See Supplement Information for BUSCO output tables.

<sup>b</sup>The *P. sp. PS1579* genomic dataset was of low coverage (see main text) and transcriptome (RNA-seq)-based models were used in a complementary purpose.

sublimation in very low temperatures can result in freeze-drying. The mechanistic bases of cryptobiosis have been explored in several animal species including *Aphelenchus* nematodes (Reardon et al., 2010), the chironomid insect *Polypedilum vanderplanki* (Gusev et al., 2014), and the tardigrades *Ramazzottius varieornatus* and *Hybysibius exemplaris* (Hashimoto et al., 2016; Yoshida et al., 2017). Intrinsically unstructured proteins and the carbohydrate trehalose have been identified as important molecular components in anhydrobiotic mechanisms. Intrinsically unstructured proteins, such as the late embryogenesis abundant (LEA)-like proteins (Tyson et al., 2012; Gusev et al., 2011), and the tardigrade-specific SAHS, MAHS, and CAHS family proteins (Hashimoto et al., 2016) do not adopt a stable secondary structure in solution and are thought to replace structural water on dehydration or freezing. The polar sugar trehalose is likely to act in a similar way. Some species are able to undergo cryptobiosis rapidly at any stage of development (Hashimoto et al., 2016), whereas others require extensive preconditioning, presumably to activate physiological systems that produce anhydro- or cryo-protectants. Asexuality and cryptobiosis may be mechanistically unlinked, but both traits are advantageous to organisms colonizing extreme habitats.

Many asexual and/or cryptobiotic nematode taxa have been described. Panagrolaimidae are members of Tylenchina, in clade IV (Blaxter et al., 1998) of Nematoda. They include both outcrossing (with male-female) and parthenogenetic (female-only) taxa. Many panagrolaimids are cryptobionts, with both cryobiotic and anhydrobiotic abilities, and are found in extreme environments in polar and temperate regions (Shannon et al., 2005). Although parthenogenetic panagrolaimids are widely distributed around the globe, a single origin of parthenogenesis in the genus *Panagrolaimus* is likely (Lewis et al., 2009). The evolutionary age of the asexual *Panagrolaimus* species is unknown. Some *Panagrolaimus* species are very good cryptobionts, surviving rapid desiccation, and many species are able to survive desiccation at any life cycle stage without pre-conditioning (Shannon et al., 2005). Here we compared the genomes and transcriptomes of parthenogenetic and outcrossing cryptobiotic *Panagrolaimus* species and one *Propanagrolaimus* species. The *Propanagrolaimus* species is not a good cryptobiont. We identified a single, recent origin of parthenogenesis in *Panagrolaimus* species involving triploidization. In contrast, using the same methodology, we found retention of diploidy in two other nematode parthenogens. We identified candidate genes that may be important for cryptobiosis, some of which are likely to have been acquired horizontally from non-metazoan sources.

## RESULTS

### Parthenogenetic *Panagrolaimus* Have Larger Genomes Than Their Sexual Congeners

We sequenced and assembled *de novo* the genomes and transcriptomes of five panagrolaimid nematodes (Table 1, Figure 1) using contamination- and heterozygosity-aware methodologies. We reassembled the genome of *Panagrolaimus sp. DAW1*, which was highly fragmented and contaminated with a considerable amount of bacterial DNA in the original assembly (Thorne et al., 2014).

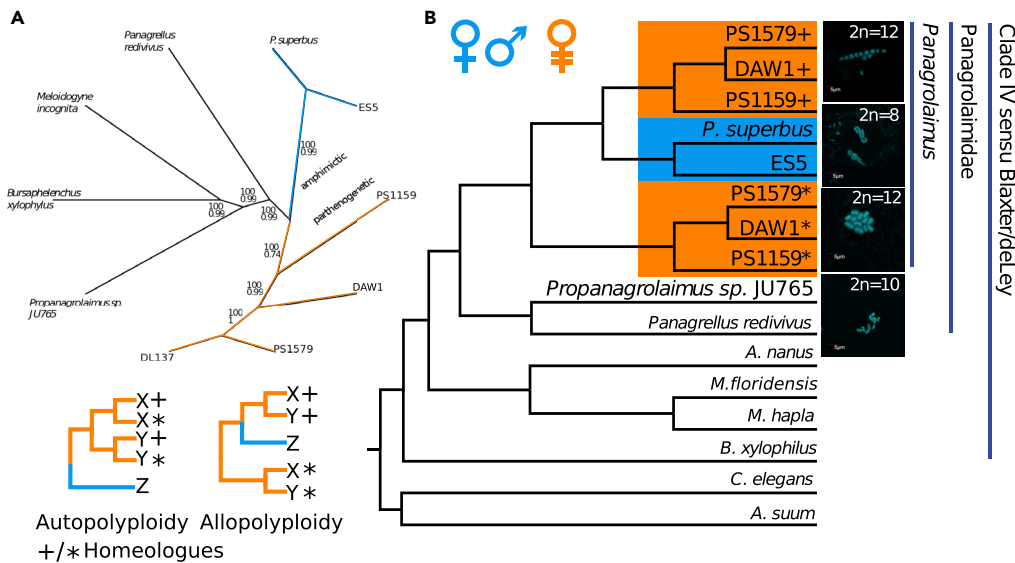
<sup>13</sup>Present address: Wellcome Sanger Institute, Cambridge CB10 1SA, UK

<sup>14</sup>Lead Contact

\*Correspondence:

philipp.schiffer@gmail.com

<https://doi.org/10.1016/j.isci.2019.10.039>



**Figure 1. Polyploidy in Parthenogenetic Panagrolaimus**

(A) Phylogenomic analysis of panagrolaimids confirms a single origin of parthenogenesis in the genus. The RAxML tree of the *Panagrolaimus* (including transcriptomic data for PS1579 and DL137), *Panagrellus*, *Propanagrolaimus* is based on CEGMA KOGs, and the topology is congruent with a Phylobayes tree of the same species set.

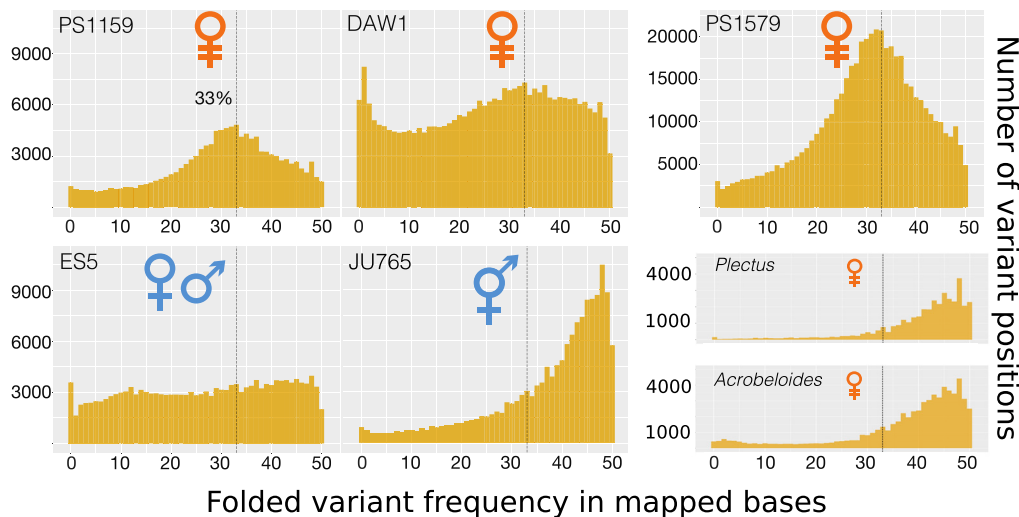
(B) Additional gene copies derive from an independent genome. MUL-trees based gene-tree reconciliation of 15,000 trees from OrthoFinder showing the most parsimonious phylogenetic position for proteins where the parthenogenetic species had multiple copies in a given orthologous group. The + and \* symbols designate the different copies per protein in the parthenogenetic species. This topology supports the assumption of polyploid hybrids (allopolyploidy), where the additional proteins in the parthenogens derive from a lineage outside the sampled outcrossing and parthenogenetic species (see tree sketches bottom left for a companions of patterns expected under allopolyploidy and autopolyploidy). All analyzed parthenogenetic species had 12 chromosomes in comparison with 8 in the diploid state of the outcrossing species. The outgroup *Propanagrolaimus* has  $2n = 10$  chromosomes.

We independently estimated the genome sizes of strains in culture using Feulgen Image Analysis Densitometry (FIAD) and quantitative PCR (RT-D-PCR). For the outcrossing *Panagrolaimus* isolates diploid genome sizes were estimated to be between 0.14 (*Panagrolaimus* sp. ES5, using FIAD) and 0.18 (*Panagrolaimus superbus*, using RT-D-PCR) pg DNA per nucleus, or 140–180 Mb. For the parthenogens *Panagrolaimus* sp. PS1159 and *Panagrolaimus* sp. PS1579 we obtained FIAD values between 0.23 and 0.27 pg DNA per nucleus (230–270 Mb), i.e., ~50% larger than the outcrossing strains. The diploid genome size of *Propanagrolaimus* sp. JU765 (previously known as *Panagrolaimus* sp. JU765, see Schiffer et al. (2014) for re-naming) was estimated at 0.13 pg DNA per nucleus (130 Mb).

The spans of the haploid genome assemblies of *P.* sp. ES5 and *P. superbus* (90 and 76 Mb, respectively, Table 1) are congruent with the sizes estimated by densitometry and PCR. Similarly, the haploid genome span of the hermaphrodite outgroup *Prop.* sp. JU765 was 64 Mb, in line with the FIAD estimate. The correspondence between assembly span and measured nuclear DNA mass was less clear-cut in the parthenogens. The assembly span of *P.* sp. PS1159 was only 85 Mb, similar to that of the haploid genomes of the sexual species but one-third of the physical measurement. The initial assembly span for *P.* sp. DAW1 was 175 Mb. After cleaning and removal of some duplicate contigs stemming from heterozygosity the final scaffolded assembly measured 118 Mb. The difference between estimated genome sizes and assembly sizes for the parthenogens would suggest the presence of genome copies not separated during the assembly.

### Parthenogenetic *Panagrolaimus* Species Are Likely Triploid

We karyotyped 15 *Panagrolaimus* species and found that all the outcrossing species had a total set of 8 chromosomes (suggesting  $2n=8$ ), whereas all the parthenogenetic species had 12 chromosomes in total (suggesting  $xn=12$ ) (Figure 1, Figure S1). This could correspond to a shift in haploid chromosome number



**Figure 2. Ploidy and Heterozygosity Estimates Based on Distribution of Variant Frequencies**

Mapping RNA-seq reads to predicted coding sequences of the respective genomes and counting the occurrence of variant frequencies, we observe a peak of variants at a frequency of  $1/3$  for the parthenogenetic species. Distribution of variant frequencies in the homozygous genome of the outcrossing diploid ES5 does not show any clear peak, whereas the hermaphrodite JU765 shows the  $1/2$  frequency peak expected under diploid heterozygosity. Patterns observed in the parthenogens *Acroboloides nanus* and *Plectus sambesii* are similar to those in JU765, thus suggesting a diploid heterozygous structure too.

from  $n=4$  to  $n=6$ , or triploidy (where  $n=4$  is maintained but somatic nuclei are  $3n$ ). Observation of polar bodies in the parthenogenetic species shows that meiosis is present (Figure S1).

In polyploid systems, extra copies of most genomic regions will exist. However, if sequence similarity between these copies is high, the genome assembly process will collapse them and only one copy will be present in the predicted gene and protein set. In evolutionarily old parthenogens that lack recombination, independent accumulation of mutation in alleles can lead to sequence divergence (Welch and Meselson, 2000) and thus representation of both alleles in the genome assembly. Similarly, in allopolyploids homeologs may be divergent enough to be independently assembled (Blanc-Mathieu et al., 2017; Glover et al., 2016; Krasileva et al., 2013).

The assembled panagrolaimid genomes showed high completeness, comparable with other *de novo* sequenced genomes for the eukaryotic gene set in the BUSCO pipeline (Simão et al., 2015) (Table 1, Data S1). We performed gene-finding on the new and reassembled genomes and predicted proteins from the transcriptomes of *Panagrolaimus* sp. DL137 and *P.* sp. PS1579. We retrieved domain annotations for 77%–88% of the predicted coding genes in each species. These were clustered with OrthoFinder and analyzed with kinfin (Laetsch and Blaxter, 2017). We included proteomes from *Propangrolaimus* sp. JU765 (this work) and other Tylenchina (Clade IV) nematodes: *Panagrellus redivivus* (Srinivasan et al., 2013), *Meloidogyne hapla* (Opperman et al., 2008), *Acroboloides nanus* (Schiffer et al., 2018), and *Bursaphelenchus xylophilus* (Kikuchi et al., 2011).

A previous phylogenetic study using neighbor joining and maximum parsimony methods based on fragments of the two rRNA genes (18S, 28S) and one mitochondrial gene (ND5) had indicated a single origin of parthenogenesis in *Panagrolaimus* (Lewis et al., 2009). Drastically extending the set of analyzed genes to all CEGMA KOGs and using the superior maximum likelihood and Bayesian methods to infer phylogenetic trees we were able to confirm a single origin of the parthenogens (Figure 1). We found an excess of protein copies in each of the parthenogenetic *Panagrolaimus* species in over 1,600 clusters of orthologs (orthogroups). Using the gene-tree reconciliation approach implemented in GRAMPA (Gregg et al., 2017) on FastME (Lefort et al., 2015) trees built from these orthogroups, we tested if auto- or allopolyploidy was the source of an extra set of proteins in the parthenogens (Figure 1). If the extra copies in the parthenogens resulted from autopolyploidy they should be more similar to each

Species A	Repro	Species B	Repro	Myr (8 days/gen)	Myr (50 days/gen)
<i>P. superbus</i>	A	ES5	A	1	3.2
PS1159	P	ES5	A	1.33	8.25
PS1159	P	<i>P. superbus</i>	A	1.37	8.55
ES5	A	<i>P. redivivus</i>	A	1.35	8.35
PS1159	P	<i>P. redivivus</i>	A	1.45	9.3
<i>P. superbus</i>	A	<i>P. redivivus</i>	A	1.88	11.6
ES5	A	JU765	H	2.4	14.95
PS1159	P	JU765	H	2.4	15.1
<i>P. superbus</i>	A	JU765	H	2.5	15.8

**Table 2. Pairwise Divergence Time Estimates in *Panagrolaimus* spp. Based on 8 Days per Generation (Measured by Us) and 50 Days per Generation as Lower Bound**

Calibrated with divergence times calculated for *Caenorhabditis* spp. by Cutter (2008). Mode of reproduction: Amphimictic, Hermaphroditic, Parthenogenetic. See Supplement Information.

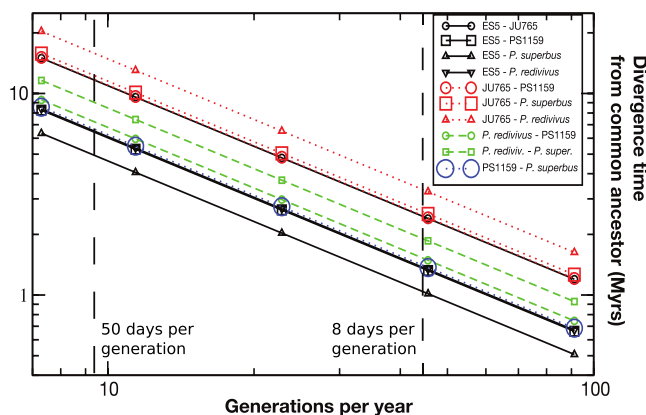
other than to orthologous copies in the outcrossing outgroup. The most parsimonious placement for these extra proteins was as basal to a clade containing both parthenogenetic and sexual species (Figure 1). It is thus unlikely that the extra copies arose by whole genome duplication within the parthenogenetic lineage (autopolyploidy). The tree topology rather supports an allopolyploid (hybrid) origin for these extra gene copies in the parthenogens.

To further assess ploidy levels in the parthenogens, we also examined read coverage of variants in *P. sp.* PS1159 and presence of homeologous loci in *P. sp.* DAW1. We mapped RNA sequencing (RNA-seq) data on the predicted coding sequences (CDS) and after variant calling plotted the distribution of allele (variants) frequencies. For the diploid outcrossing *P. sp.* ES5 sequenced from a homozygous population, we observed a flat distribution of allele frequencies, with no evident peak, which is consistent with its homozygous nature (Figure 2). A similar pattern was observed in *P. superbus* (not shown). In the hermaphrodite *P. sp.* JU765, distribution of variant frequencies showed a mode of variants present at 50%:50%, consistent with a diploid heterozygous genome. However, in the parthenogenetic taxa we observed a peak in allele frequencies at ~33%, consistent with the presence of three gene copies at these sites (Figure 2). Using a two-sample Kolmogorov-Smirnov test we confirmed that the allele frequency spectra of the parthenogenetic and outcrossing *Panagrolaimus* species, as well as the parthenogenetic *Panagrolaimus* and androdioecious (hermaphrodite) *Propanagrolaimus*, belong to different distributions (see Table S1). Mapping of genomic reads to the CDS, repeat-masked genomes, and un-masked genomes resulted in similar allele frequency profiles (data not shown). These data, together with observed chromosome numbers support triploidy in the parthenogenetic *Panagrolaimus* species.

We explored the genomes of other parthenogenetic nematodes for similar signals of polyploidy. We analyzed the genome and transcriptome of the Tylenchine (Clade IV) species *Acrobeloides nanus* (Schiffer et al., 2018) and the genome of the Chromadorid (Clade C) species *Plectus sambesii* (Rosic et al., 2018). We found that the allele frequency spectra of both species resembled that of the heterozygous (selfing) hermaphrodite *P. sp.* JU765 (Figure 2). These findings argue for non-polyploid status of these parthenogens.

### Parthenogenesis in *Panagrolaimus* Is 1.3–8.5 Million Years Old

Because they lack the advantages provided by sexual reproduction (e.g., allele shuffling, efficiency of selection), most parthenogenetic species are expected to be recent and to be evolutionary dead ends not giving rise to daughter taxa. To estimate divergence time between the parthenogenetic and outcrossing panagrolaimids, we calculated pairwise divergence of entire genome sequences with ANDI (Haubold et al., 2015). As calibration we used available divergence time estimates for *Caenorhabditis* species (Cutter, 2008), corrected by generation time measurements in *Panagrolaimus* and *Propanagrolaimus*. The



**Figure 3. Panagrolaimus Species Diversification Age**

Using our own generation time measures, genome-wide divergence calculated with Andi, and previous data for *Caenorhabditis* species as reference, we calculated possible age of divergence between the different species and the origin of parthenogenesis in *Panagrolaimus*. Eight days per generation was measured in the laboratory and would indicate an approximate age of 1.3 mya for the PS1159/*P. superbus* split (see Table 3). Fifty days per generation would be a realistic estimate for species living in harsh conditions and would suggest a more ancient origin.

outcrossing species *P. superbus* and *P. sp. ES5* were the most recently separated species (Table 2, Figure 3, see phylogeny in Figure 1). Assuming an average of 8 days per generation, the split between parthenogenetic and outcrossing *Panagrolaimus* species is estimated to have occurred 1.3–1.4 Mya. In natural conditions, it is possible that the *Panagrolaimus* species spend most of the time as non-metabolic cryptobionts. Applying a very conservative generation time of 50 days, the split would have occurred 8.0 to 8.5 Mya.

### Gene Copies of Potentially Lethal Genes Are less Likely Divergent in the Triploid Hybrids

The evolution of parthenogenesis through hybridization is rapid. Newly formed hybrids may undergo genome and transcriptome shock, as co-adapted complexes can be disrupted by the new ploidy and interacting partners that have accumulated differences through drift are brought into association. Such maladapted interactions need to be eliminated or moderated, and, in the case of parthenogens especially, the machineries of sex determination, fertilization/oocyte activation, and development are likely to undergo accelerated evolution. In parthenogens, the cost of producing males, which play no part in species fitness, is particularly high.

No males have been observed by us in any of the parthenogenetic *Panagrolaimus* species, even under stress conditions (e.g., prolonged culturing at elevated temperatures or after reviving cultures from cryptobiosis). We therefore examined panagrolaimid orthologues of genes involved in several genetic regulatory networks implicated in reproduction and sex determination that have been experimentally defined in *Caenorhabditis elegans*. Using stringent orthology analysis using OrthoMCL (Li et al., 2003) followed by validation with OrthoInspector (Linard et al., 2015) we identified changes consequent to the evolution of polyploidy and loss of sex (see Supplemental Information). Surprisingly, we found that many *C. elegans* genes acting in important developmental processes were absent not only from all the panagrolaimids tested, but also from other Tylenchina species. We were also unable to detect these genes through synteny analysis (see Supplemental Information), an approach previously successfully applied between *Caenorhabditis* species (Kuwabara and Shah, 1994; Streit et al., 1999; Haag et al., 2002). For example, in the sex determination GRN, Tylenchina nematodes lacked orthologs of the *C. elegans* master regulator *xol-1*, as well as *fem-1*, *fem-3*, *her-1*, and other genes. Similarly, in the endoderm- and mesoderm-forming GRN, tylenchines lacked *med-1*, *med-2*, *end-1*, *end-3*, and *spn-4*, among others. A similar pattern was observed for genes acting in spermatogenesis, early axis definition, DNA repair, and oogenesis. This core divergence from the systems as defined in *C. elegans* makes it difficult to identify changes associated with divergence in reproductive mode and ploidy.



In comparison with their outcrossing relatives, the parthenogenetic *Panagrolaimus* species had an excess of genes, many of which were duplicates. The most common fate of duplicated genes is loss through accumulation of deleterious mutations or other mechanisms. This fate is especially likely for loci that are sensitive to dosage or have strong phenotypes when mutated. We identified clusters of orthologs that contained exactly one protein in *C. elegans*, the two outcrossing, and all three parthenogenetic *Panagrolaimus* species (1:1: ...:1 orthologs), as well as clusters where the parthenogenetic *P. sp.* PS1159 had more proteins than the obligate outcrossing *P. sp.* ES5 (parthenogen-excess clusters). Using the database of *C. elegans* RNAi phenotypes on [wormbase.org](http://wormbase.org) (<https://wormbase.org>; Lee et al., 2018) we retrieved phenotypes for 172 of the 703 single-copy one-to-one orthologues and 695 phenotypes for the 1,618 parthenogen-excess clusters. A chi-square test confirmed ( $p$  value = 0.0015) that the proportion of lethal phenotypes (53%) in the 1:1 orthologues was higher and significantly different from the proportion (40%) in parthenogen-excess clusters. This result is in accordance with the expectation that essential genes will be less tolerant to changes in dosage.

### **Panagrolaimid Species Have Acquired Genes through Horizontal Transfer and These May Impact Cryptobiosis**

We identified potential horizontal gene transfers (HGTs) using the Alien Index (AI) approach as implemented in Alienness (Rancurel et al., 2017). We identified from 22 (*P. redivivus*) to 232 (*P. sp.* ES5) likely HGT candidates from non-metazoan donors into the genomes of panagrolaimids (Table 3, Data S4, S5, S6, S7, S8, and S9). We used a series of additional screens to confirm HGT and rule out contamination, including a check for spliceosomal introns, codon usage, presence of evolutionarily conserved metazoan genes on the same contigs, and transcriptional support from RNA-seq data (Figure 4). The inferred source taxa for the majority of high-confidence HGT candidates were Bacteria followed by Fungi in all *Panagrolaimus* species except *P. sp.* DAW1 (in which Fungi were the most frequent, followed by Bacteria).

We identified 110 OrthoMCL groups that contained candidate HGT from at least two *Panagrolaimus* species. These genes have most likely been acquired horizontally in a common ancestor for these species (multiple independent acquisitions is less parsimonious). We reconstructed the relative timing of acquisition via HGT for these 110 orthogroups using parsimony (Methods, Figure 5). Most (102; 93%) HGT events were predicted to have taken place in an ancestor of two or more *Panagrolaimus* species. Interestingly, the highest number of acquisitions (49 events) took place in the lineage leading to the last common ancestor of all the cryptobiotic *Panagrolaimus* species. Only seven HGT orthogroups were present in the last common ancestor of all eight panagrolaimid species analyzed. These seven ancestral HGT orthogroups plus the 49 novel acquisitions suggest that at least 56 independently acquired HGT loci were present in the common ancestor of the cryptobiotic *Panagrolaimus* species (Figure 5).

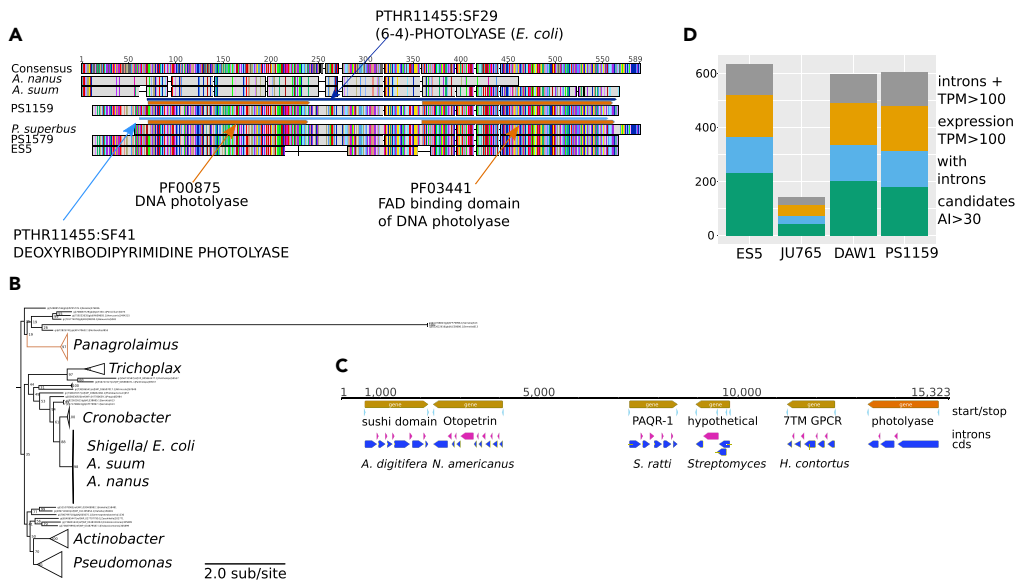
The 56 HGT orthogroups present in the last common cryptobiont ancestor contained a total of 910 proteins, annotated with 192 different Pfam domains (Table 3). Only three Pfam domains were conserved in the candidate HGT sets of all six *Panagrolaimus* species and all corresponded to enzymatic functions (alcohol dehydrogenase GroES-like domain, pyridine nucleotide-disulfide oxidoreductase, and zinc-binding dehydrogenase) (Data S3). HGTs conserved between *P. superbus*, *P. sp.* ES5, *P. sp.* DAW1, and *P. sp.* PS1159 were annotated with thirty conserved Pfam domains that included GH43, GH28, and GH32 glycoside hydrolases, previously identified as loci acquired by HGT in plant-parasitic nematodes (Rancurel et al., 2017) (Haegeman et al., 2011). We found that 47 Pfam domains functionally connected to cryptobiosis were annotated in proteins putatively acquired via HGT, 19 acquired by the last common ancestor of *Panagrolaimus* and 5 only found in the parthenogenetic species. Functions associated with these domains included peptidase, diapausin, killer toxins of the Kp4 family, LURP-one-related proteins, flavin-binding monooxygenase-like protein, caleosin, and photolyase. Proteins with these functions fulfill various roles in defense against pathogens and the removal of xenobiotics. The photolyase (Figure 4) was of particular interest, since photolyases are able to repair UV-induced damage in DNA and have been lost in several branches of Metazoa (Lucas-Lledó and Lynch, 2009). Desiccation can induce DNA damage and thus genome repair is important for desiccating organisms (Jönsson and Schill, 2007). The caleosin found in *Panagrolaimus* is the first report to date of animal caleosins. Caleosins have previously been described only in multicellular plants, green algae, and fungi (Partridge and Murphy, 2009). Plant caleosins expressed in non-seed tissues are responsive to a range of environmental stresses, particularly salinity and dehydration (Blée et al., 2014).

Species	Total # Proteins	Possible HGT Candidates (AI >0)	Candidates Identified as Contaminants <sup>a</sup>	# HGT Candidates with AI >30	High-Confidence HGT Candidates <sup>b</sup>					
					# on Same Contig with Other Gene	# with Neighbors Hitting Metazoa	# with introns	# with Expression >10/100/1000 TPM in RNA-seq	# with Pfam Domains (Total Number of Pfam Domains)	Distinct Pfam Domains
<i>Panagrolaimus superbus</i> <sup>b</sup>	24,160	528	20	88	NA	NA	NA	NA	70 (91)	56
<i>Panagrolaimus</i> sp. ES5	24,756	975	141	232	215	73	132	185/155/97	183 (429)	136
<i>Panagrolaimus</i> sp. PS1159	27,350	899	19	182	159	98	133	174/164/112	141 (268)	93
<i>Panagrolaimus</i> sp. DAW1	31,646	1,177	18	201	194	84	136	178/153/84	131(172)	62
<i>Propanagrolaimus</i> sp. JU765	24,878	530	39	44	42	25	27	44/44/36	34 (100)	23
<i>Panagrellus redivivus</i>	26,372	516	12	22	21	21	16	NA	14 (18)	12

**Table 3. Protein Numbers, HGT Candidates, and High-Confidence HGT Candidates in the *Panagrolaimus* Genomes and Outgroups**

<sup>a</sup>AI>0 and >70% id to non-metazoan protein.

<sup>b</sup>Owing to the shortness of the contigs and a low-coverage 454 transcriptome we refrained from calculating high-confidence candidates for *P. superbus*.



**Figure 4. Validating HGT Candidates**

(A) Protein alignment of putative HGT photolyases found in *Panagrolaimus*, *A. suum*, and *A. nanus* with bacterial homologues. Brown horizontal bars denote Pfam domains, whereas blue colored bars are PANTHER classification (both InterProScan derived). The alignment shows that the *A. nanus* (clade IV) and *A. suum* (clade III) sequences are similar but different from the ones in the three panagrolaimids. The PANTHER classification indicates that the former two are very similar to *E. coli* photolyases in these very divergent nematode species, making a contamination originating from bacteria used in worm culturing likely.

(B) Phylogenetic analysis of photolyases from *Panagrolaimus*, *A. suum*, and *A. nanus* and bacteria. The analysis further supports that the *A. suum* and *A. nanus* proteins are *E. coli* contamination, whereas the *Panagrolaimus* proteins are distinct acquisitions from other taxa.

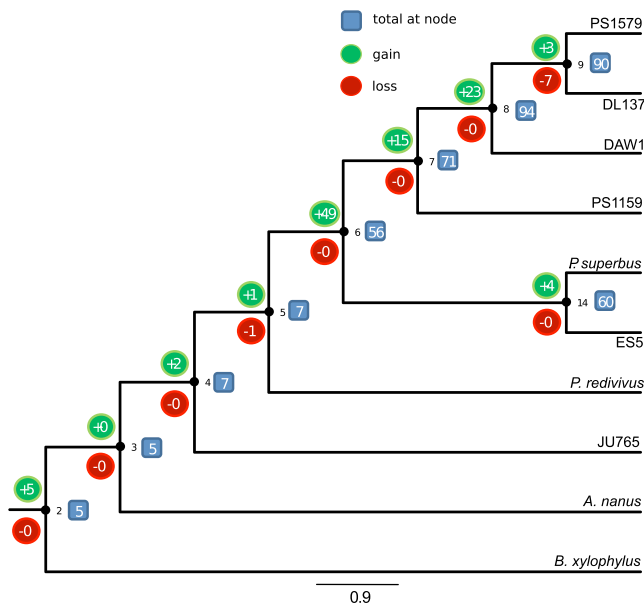
(C) Genomic integration of HGT photolyase. The photolyase locus in *P. sp.* PS1159 was found on the same contig as *bona fide* nematode proteins, further substantiating the conclusions that the gene was horizontally acquired and fully integrated into the nematode genome.

(D) Classification of HGT candidates in Panagrolaimidae. A stacked histogram of AI>30 HGT candidates for the different species classified by their features. Those with introns and expression levels of TPM>100 can be regarded as the most reliable.

### Panagrolaimus Genomes Are Shaped by Gene Families Potentially Implicated in Cryptobiosis

Adaptation to novel (extreme) habitats can also involve expansion and diversification of gene families. We analyzed gene family expansions in the cryptobiotic panagrolaimid species compared with related taxa. Forty-three Pfam domains were overrepresented in the cryptobiotic *Panagrolaimus* species compared with an outgroup containing *Prop. sp.* JU765 and *Panagrellus redivivus* (which are not cryptobionts), after Benjamini-Hochberg correction, and 123 Pfam domains were overrepresented in Panagrolaimidae (*Panagrolaimus*, *Propanagrolaimus*, and *Panagrellus*) compared with the remaining taxa (Figure S2).

As an approach complementary to overrepresentation tests, we programmed a support vector machine (SVM) classifier based on Pfam domain annotations to distinguish cryptobiotic *Panagrolaimus* from *Propanagrolaimus* plus *Panagrellus*. In the *Panagrolaimus* versus *Panagrellus/Propanagrolaimus* SVM analyses, 64 domains were frequently associated with successful classification, of which 13 were common to the SVM and the Pfam enrichment Null hypothesis significance testing (NHST) approaches (Figure S2). These 13 domain annotations included several that may play a role in cryptobiosis, including Hsp70 protein domains, serine protease inhibitors, ubiquitin family domains, lipase domains, and enoyl-reductase domains. Annotations found to be overrepresented only in the NHST analyses were proteolysis (subtilase family and matrixin domains) and maintenance of genome integrity (PIF1-like helicase, deoxyribonuclease II, linker histone H1 and H5 family).



**Figure 5. Reconstruction of the Timing of Acquisition and Loss of HGT Candidates**

The distribution of HGT candidates across these species was mapped to a phylogeny of Panagrolaimidae and two outgroups and ancestral states deduced with a parsimony approach. Internal nodes are numbered (2–9 and 14). The number of gene families acquired via HGT/lost at each node are represented in green/red bubbles, respectively. The total number of gene families acquired via HGT present at each node is indicated within blue boxes.

### Repeat and Transposon Content and Functional Analysis of the *Pangrolaimus* Gene Space

For additional information on Panagrolaimid genome assemblies as well as pertinent comparative features, please refer to the Supplementary Results and Discussion section within the [Supplemental Information](#) file.

## DISCUSSION

Understanding how parthenogenetic animals evolve and whether they persist through evolutionary time is critical to answering the fundamental biological question of why recombinational sex is near-ubiquitous in animals. How animal species adapt to environmental challenges is similarly of fundamental interest. It has been proposed that both traits are linked in geographical parthenogenesis (Vrijenhoek and Davis Parker, 2009). Many parthenogenetic nematode species have been shown to be polyploids of hybrid origin (Lunt, 2008; Castagnone-Sereno and Danchin, 2014; Lunt et al., 2014; Hiraki et al., 2017; Kraus et al., 2017; Blanc-Mathieu et al., 2017). In animals, it is assumed that polyploidy, especially when uneven, leads to non-viability and that living hybrids may suffer a reduced fitness (Burke and Arnold, 2001). Alternatively, polyploid parthenogens might have a fitness advantage in that they are buffered against Muller ratchet (Lynch, 1984). Hybridization between species frequently leads to aberrant meiosis in animals, and a possible escape from this dilemma could be parthenogenesis (Avisé, 2008). In such parthenogens, incorporation of a round of chromosomal endoreplication is a possible route to maintain meiotic triploidy through oogenesis (Lutes et al., 2010).

We have analyzed the genomes and transcriptomes of parthenogenetic and outcrossing *Panagrolaimus* nematodes, which, in contrast to closely related *Propanagrolaimus* and *Panagrellus*, are capable of surviving under prolonged drying (anhydrobiosis) and freezing (cryobiosis). Our combined karyotypic and molecular data indicate that the parthenogenetic *Panagrolaimus* are triploid hybrids originating from a single event. All analyzed parthenogens have a karyotype of  $n = 12$  (compared with diploid  $n = 8$  in amphimicts), an excess of duplicated genes that share a single phylogenetic origin, and variant frequency spectra compatible with triploidy. Triploidy could have arisen by autopolyploidy (duplication of one set of ancestral chromosomes, for example, by fertilization of a diploid oocyte by a haploid sperm from the same species) or allopolyploidy (where the additional haploid complement comes from a distinct lineage). These contrasting origins predict contrasting patterns of divergence between the three

copies of each locus. In parthenogenetic *Panagrolaimus* we find phylogenetic evidence for deep divergence of the extra copies and probable acquisition from a donor species positioned between the *Propanagrolaimus* and *Panagrolaimus* species present in our analysis. These results support allotriploidization. Analyzing two further parthenogenetic nematodes, the tylenchine cephalobe *A. nanus*, and the distantly related plectid *P. sambesii*, we found contrasting signal. Both species appear to be diploid. It is thus likely that they became parthenogenetic through a different route than the *Panagrolaimus* species. In the Rhabditine (Clade V) genus *Diploscapter* (the sister genus to *Caenorhabditis*) parthenogenetic species appear to show allelic divergence, but whether parthenogenesis involved hybridization in the genus remains unresolved (Fradin et al., 2017; Hiraki et al., 2017; Kraus et al., 2017). Thus parthenogenesis has arisen independently and repeatedly in Nematoda (Holovachov et al., 2013), offering a rich set of systems in which to investigate the mechanisms and subsequent effects of the loss of sexual reproduction.

Several kinds of genes associated with cryptobiosis have been identified in the panagrolaimid nematodes. Plant embryos express at high levels LEA proteins, natively unstructured peptides that are believed to surround proteins in a “molecular shield” (Chakrabortee et al., 2012). Natively unstructured proteins have been implicated in anhydrobiotic responses in nematodes (*Aphelenchus avenae* [Reardon et al., 2010]), arthropods (*Polypedilum vanderplanki* [Cornette et al., 2010]), and tardigrades (*R. varieornatus* and *H. dujardini* [Horikawa et al., 2013]). Other gene families associated with anhydrobiosis include protein kinases, proteasomal components, ubiquitin, protease inhibitors, proteases, DNA repair enzymes, enzymes with roles in oxidative stress protection, and heat shock proteins (Cornette et al., 2010).

In cryptobiotic tardigrades, in addition to a series of tardigrade-specific natively unstructured protein families, genomic and transcriptomic analysis highlighted the involvement of key stress response pathways in anhydrobiosis. These include horizontally acquired catalases and anhydrobiosis-upregulated superoxide dismutases and glutathione-S transferases, enzymes involved in detoxification and protection against oxyradical damage. DNA damage signaling and repair genes were also associated with anhydrobiosis. Surprisingly, the mTOR signaling pathway was absent in the tardigrades, as were the HIF-1 $\alpha$  cascade and many genes involved in peroxisomal functions including beta oxidation and other H<sub>2</sub>O<sub>2</sub> producing pathways. It has been proposed (Hashimoto et al., 2016; Yoshida et al., 2017) that their absence in tardigrades is a mechanism to avoid induction of stress-related apoptosis in cells undergoing anhydrobiosis. In the panagrolaims, however, we found no loss of these pathways. Thus the molecular machinery of desiccation does not follow convergent evolutionary paths in these taxa.

Horizontal gene transfer has emerged as an important process in animal evolution (Danchin, 2016) and has been associated (not un-controversially) with cryptobiosis. We identified about 150 gene families in *Panagrolaimus* and Panagrolaimidae that had signatures of acquisition through HGT, many of which were mapped to the root of the cryptobiotic *Panagrolaimus* species group. These candidate HGT genes have acquired spliceosomal introns, have similar GC content and codon usage to resident *Panagrolaimus* genes, are assembled in the vicinity of bona fide nematode genes, and are expressed as poly(A)<sup>+</sup> RNA. They can be regarded as domesticated and functionally integrated into the host genomes. Functional annotation of the putative HGT loci in *Panagrolaimus* suggested that they contribute to anhydrobiotic physiology and feeding. Genes acquired through HGT were also implicated in cryptobiosis in the rotifer *A. vaga* and the tardigrades *R. varieornatus* and *H. dujardini*, and a functional link between anhydrobiosis and HGT has been proposed (Hespeels et al., 2014). However, comparative genomic analyses in rotifers showed that the proportion of genes acquired by HGT was equally high in desiccation-intolerant and -tolerant rotifers (Nowell et al., 2018). The same analysis extended to other protostomes showed no clear link exists between abundance of HGT and desiccation tolerance (ibid).

The level of HGT in panagrolaimid nematodes appears similar to that found in hysibid tardigrades (Yoshida et al., 2017). Only a small fraction of the total gene number, around 1% or less, is horizontally acquired. Nevertheless, the relative proportion of strongly supported HGT candidates per genome is higher in the cryptobiotic panagrolaimid species in comparison with the non-cryptobiotic outgroups, hinting at a potential connection between the lateral acquisition of genes and cryptobiosis. Whether this is mechanistic (i.e., that taxa that experience drying or freezing are more likely to take up DNA from their environment) or causal (i.e., the HGTs underpin cryptobiosis) is presently not resolvable.

We identified karyotypic signatures of meiosis in the parthenogenetic *Panagrolaimus* species. However, the exact mode of restoration of their triploid status after meiosis remains unresolved. An aberrant tripolar spindle forming during meiosis I contributes to the maintenance of the correct (3n) ploidy in triploid carp (the goldfish *Carassius auratus langsdorffi*) (Yamashita et al., 1993). In parthenogenetic triploid *Poeciliopsis* fish endomitosis precedes meiosis, raising the number of chromosomes to 6n to ensure that a triploid egg is generated at the end of meiosis (Cimino, 1972). Identification of the parthenogenetic mechanism in the *Panagrolaimus* species will need detailed microscopic and cytogenetic analysis.

The parthenogenetic *Panagrolaimus* species appear to be 1.3–8.5 mya old. Although this is older than previously reported cases in vertebrates and most arthropods, this is much younger than the extreme estimates for the bdelloid rotifers and oribatid mites (up to 100 mya) (Neiman et al., 2009). It should be noted that estimating the age of loss of sexual reproduction is not straightforward, and the methods used are completely different for each species. Furthermore, estimates of generation times cannot be easily transferred from the laboratory to nature. For example, the Antarctic species *P. sp.* DAW1 could potentially be restricted to very few generations per year. However, other species are present in temperate regions and thrive in semi-arid (sandy) soils or leaf litter (Shannon et al., 2005) (McGill et al., 2015) where many generations per year are likely. The theory underpinning geographical parthenogenesis holds that parthenogenetic species have an advantage in adapting to new and/or extreme environments. Since we did not observe a strong signal of specific adaptation in the parthenogenetic species (including to the mode of reproduction itself) in comparison with sexual congeners it appears possible that a general-purpose genotype is indeed maintained in these evolutionarily young parthenogens, potentially by combining several differently adapted genomes during hybridization. The same hypothesis was suggested to resolve the surprising parasitic success of root-knot nematodes despite obligatory parthenogenesis (Blanc-Mathieu et al., 2017).

Only a few parthenogenetic species are suitable for genomic and molecular research. We have not yet linked expression of specific *Panagrolaimus* gene families directly to cryptobiosis or parthenogenesis. Our ability to infer function for *Panagrolaimus* genes was limited by phylogenetic distance from the well-annotated *C. elegans* orthologues. *Panagrolaimus* nematodes are easily cultured in the laboratory and accessible to molecular genetic analysis (Schiffer et al., 2014), including RNAi (Reardon et al., 2010). Given the amenability of many cultured nematodes species to genetic manipulation through CRISPR-Cas and other toolkits, we expect these *Panagrolaimus* species will be a valuable platform for further exploration of the genes important for and mechanisms of cryptobiosis.

### Limitations of the Study

Ideally, the genomes presented in this study would have been sequenced with long-read technology and single individuals as source for the input DNA. This might have allowed totally separating and phasing the individual haplotypes of the parthenogenetic and outcrossing species. Unfortunately, sequencing single individuals, although theoretically interesting to analyze the intra-species divergence in the parthenogens, remains technically extremely challenging (and very cost-intensive) with long-read technologies. Hence, one limitation is the fragmented state of the assembled genomes, which precludes most of the synteny analyses.

Although we could assert meiosis in the parthenogenetic *Panagrolaimus* species, the exact mode of restoration of the triploid status remains unresolved. It has been hypothesized that an aberrant tripolar spindle forming during meiosis I in triploid carps (the goldfish species *Carassius auratus langsdorffi*) is contributing to the maintenance of the correct (3n) ploidy (Yamashita et al., 1993). In triploid parthenogenetic *Poeciliopsis* fish, an endomitosis precedes meiosis. By first raising the number of chromosomes to 6n it is thus ensured that at the end of meiosis a triploid egg is present (Cimino, 1972). Whether similar mechanisms are at play in the *Panagrolaimus* species will need detailed microscopic analysis of the germline in the future.

A further limitation was imposed by the phylogenetic position of the species themselves. Owing to the distant relation to *C. elegans*, a substantial portion of orthologues involved in the earliest steps of development could not be found in the parthenogenetic panagrolaimid species and more generally in the clade IV nematodes. Thus, although we present a tractable model to explore the mechanisms of parthenogenesis in the future, there is still development needed to identify the genes involved.

Finally, unlike the tardigrades analysis, whereby differential expression data for species with different amenability to desiccation was conducted (Yoshida et al., 2017), we were not able to link specific gene families differentially regulated upon cryptobiosis. Our analysis remains a genomic study, yielding no expression data on cryptobiosis. To study which of the identified candidate genes might play an active role during desiccation or freezing, functional molecular studies, like RNAi or CRISPR-mediated gene knock-outs, would be needed.

## METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

## DATA AND CODE AVAILABILITY

All sequencing libraries are deposited in the SRA under the BioProject PRJNA374706. All genome assemblies are deposited in ENA under PRJEB32708 (*Propanagrolaimus* JU765: ERS3438151; *Panagrolaimus* PS1159: ERS3467865; *Panagrolaimus* ES5\_v2 ERS3467866; *Panagrolaimus superbus*: ERS3467867; *Panagrolaimus* PS1579\_v2 ERS3468321; *Panagrolaimus* sp. DAW1: ERS3468322). Transcriptome assemblies are deposited in TSA and can be found under the same BioProject as the raw read data (PRJNA374706). All genome assemblies, as well as gene predictions, and corresponding peptide sequences will be made available through WormBaseParasite shortly.

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.isci.2019.10.039>.

## ACKNOWLEDGMENTS

The authors are particularly thankful to Anne-Marike Schiffer Maraun for providing the SVM classifier code and her help in reporting on the data from this analysis. The authors thank Isabel Goebel for generation time measurements and Janine Beyert and Andrea Kraemer-Eis for help in culturing nematodes. The authors are grateful to the GenoToul bioinformatics platform Toulouse Midi-Pyrenees for providing computing resources, as well as to Nordrhein-Westfalen for the use of the CHEOPS computing cluster at the University of Cologne.

A.M.B was funded by the VolkswagenStiftung in their initiative for evolutionary biology, the Deutsche Forschungsgemeinschaft in a grant to T.W. under SPP1819, and was supported by the European Research Council (ERC-2012-AdG 322790 to Max Telford). A.M.B. was funded by Science Foundation Ireland (Project RFP/EOB2506). E.A.M.-S. thanks Dr. Ryan Gregory (University of Guelph) for providing access to his laboratory to conduct genome size estimation and CONACyT for a Postdoctoral Fellowship (2010), and UAZ-2016-37096. C.J.C. was funded under the Science Foundation Ireland (SFI) Stokes lecturer scheme (07/SK/B1236A) and the Biotechnology and Biological Sciences Research Council (BBSRC) Institute Strategic Program Grant, Rumen Systems Biology, (BB/E/W/10964A01). Edinburgh Genomics has core funding from the UK Natural Environment Research Council (UKSBS PR18037).

## AUTHOR CONTRIBUTIONS

P.H.S.: planned and conceived the study, managed the collaboration, analyzed the data, supervised G.S., wrote the paper.

E.G.J.D.: analyzed the HGT data and planned and wrote the manuscript.

A.M.B.: analyzed the anhydrobiosis and HGT data, conceived the study, supervised G.O.'M. and B.C., and wrote the paper.

C.J.C.: analyzed the selection data, drew figures, and contributed to writing.

S.W.: annotated the genomic data.

I.D.: karyotyped species, drew figures, and contributed to writing.

G.O.'M.: analyzed anhydrobiosis data for *P. superbus*.

B.A.C.: conducted molecular experiments for transcriptomics.

C.R.: generated and analyzed the HGT data.

G.S.: conducted genome size measurements.

E.A.M.S.: conducted genome size measurements and wrote and revised the manuscript.

U.T.: analyzed raw sequencing data.

M.K.: conceived parts of the study and co-supervised P.H.S.

M.A.S.T.: conceived parts of the study, co-initiated the collaboration, and analyzed the anhydrobiosis data.

E.S.: conducted microscopic analyses, supervised PHS, and revised the manuscript.

T.W.: analyzed generation time data and wrote and revised the manuscript.

M.B.: planned and conceived the study, analyzed the data, and wrote the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 21, 2018

Revised: July 17, 2019

Accepted: October 21, 2019

Published: November 22, 2019

## SUPPORTING CITATIONS

The following reference appears in the Supplemental Information: [Zerbino and Birney, 2008](#).

## REFERENCES

- Archetti, M. (2004). Loss of complementation and the logic of two-step meiosis. *J. Evol. Biol.* 17, 1098–1105.
- Avisé, J. (2008). *Clonality* (Oxford University Press).
- Blanc-Mathieu, R., Perfus-Barbeoch, L., Aury, J.-M., Da Rocha, M., Gouzy, J., Sallet, E., Martin-Jimenez, C., Bailly-Bechet, M., Castagnone-Sereno, P., Flot, J.-F., et al. (2017). Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. *PLoS Genet.* 13, e1006777.
- Blaxter, M.L., De Ley, P., Garey, J.R., Le Guédard, M., Hanano, A., Heintz, D., Ehling, J., Herrfurth, C., Feussner, I., and Bessoule, J.-J. (1998). A molecular evolutionary framework for the phylum Nematoda. *Nature* 392, 71–75.
- Blée, E., Boachon, B., Burcklen, M., Le Guédard, M., Hanano, A., Heintz, D., Ehling, J., Herrfurth, C., Feussner, I., and Bessoule, J.J. (2014). The reductase activity of the *Arabidopsis* caleosin RESPONSIVE TO DESSICATION20 mediates gibberellin-dependent flowering time, abscisic acid sensitivity, and tolerance to oxidative stress. *Plant Physiol.* 166, 109–124.
- Burke, J.M., and Arnold, M.L. (2001). Genetics and the fitness of hybrids. *Annu. Rev. Genet.* 35, 31–52.
- Castagnone-Sereno, P., and Danchin, E.G.J. (2014). Parasitic success without sex – the nematode experience. *J. Evol. Biol.* 27, 1323–1333.
- Chakrabortee, S., Tripathi, R., Watson, M., Schierle, G.S.K., Kurniawan, D.P., Kaminski, C.F., Wise, M.J., and Tunnacliffe, A. (2012). Intrinsically disordered proteins as molecular shields. *Mol. BioSystems* 8, 210–219.
- Cimino, M.C. (1972). Meiosis in triploid all-female fish (Poeciliopsis, Poeciliidae). *Science* 175, 1484–1486.
- Cornette, R., Kanamori, Y., Watanabe, M., Nakahara, Y., Gusev, O., Mitsumasu, K., Kadono-Okuda, K., Shimomura, M., Mita, K., Kikawada, T., and Okuda, T. (2010). Identification of anhydrobiosis-related genes from an expressed sequence tag database in the cryptobiotic midge *Polypedilum vanderplanki* (Diptera; Chironomidae). *J. Biol. Chem.* 285, 35889–35899.
- Cutter, A.D. (2008). Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol. Biol. Evol.* 25, 778–786.
- Danchin, E.G.J. (2016). Lateral gene transfer in eukaryotes: tip of the iceberg or of the ice cube? *BMC Biol.* 14, 101.
- Denver, D.R., Clark, K.A., and Raboin, M.J. (2011). Reproductive mode evolution in nematodes: insights from molecular phylogenies and recently discovered species. *Mol. Phylogenet. Evol.* 61, 584–592.
- Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* 78, 737–756.
- Fradin, H., Kiontke, K., Zegar, C., Gutwein, M., Lucas, J., Kovtun, M., Corcoran, D.L., Baugh, L.R., Fitch, D.H.A., Piano, F., and Gunsalus, K.C. (2017). Genome architecture and evolution of a unichromosomal asexual nematode. *Curr. Biol.* 27, 2928–2939.e6.



- Glover, N.M., Redestig, H., and Dessimoz, C. (2016). Homoeologs: what are they and how do we infer them? *Trends Plant Sci.* *21*, 609–621.
- Gregg, W.C.T., Ather, S.H., and Hahn, M.W. (2017). Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Syst. Biol.* *66*, 1007–1018.
- Gusev, O., Cornette, R., Kikawada, T., and Okuda, T. (2011). Expression of heat shock protein-coding genes associated with anhydrobiosis in an African chironomid *Polypedilum vanderplanki*. *Cell Stress Chaperones* *16*, 81–90.
- Gusev, O., Suetsugu, Y., Cornette, R., Kawashima, T., Logacheva, M.D., Kondrashov, A.S., Penin, A.A., Hatanaka, R., Kikuta, S., Shimura, S., et al. (2014). Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat. Commun.* *5*, 4784.
- Haag, E.S., Wang, S., and Kimble, J. (2002). Rapid Coevolution of the Nematode Sex-Determining Genes *fem-3* and *tra-2*. *Curr. Biol.* *12*, 2035–2041.
- Haegeman, A., Jones, J.T., and Danchin, E.G.J. (2011). Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol. Plant Microbe Interact.* *24*, 879–887.
- Hashimoto, T., Horikawa, D.D., Saito, Y., Kuwahara, H., Kozuka-Hata, H., Shin-I, T., Minakuchi, Y., Ohishi, K., Motoyama, A., Aizu, T., Enomoto, A., et al. (2016). Extremotolerant tardigrade genome and improved radiotolerance of human cultured cells by tardigrade-unique protein. *Nat. Commun.* *7*, 12808.
- Haubold, B., Klötzl, F., and Pfaffelhuber, P. (2015). andi: fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics* *31*, 1169–1175.
- Hespeels, B., Knapen, M., Hanot-Mambres, D., Heuskin, A.C., Pineux, F., Lucas, S., Koszul, R., and van Doninck, K. (2014). Gateway to genetic exchange? DNA double-strand breaks in the bdelloid rotifer *Adineta vaga* submitted to desiccation. *J. Evol. Biol.* *27*, 1334–1345.
- Hiraki, H., Kagoshima, H., Kraus, C., Schiffer, P.H., Ueta, Y., Kroiher, M., Schierenberg, E., and Kohara, Y. (2017). Genome analysis of *Diploscapter coronatus*: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. *BMC Genomics* *18*, 478.
- Holovachov, O., Boström, S., De Ley, I.T., Robinson, C., Mundo-Ocampo, M., and Nadler, S.A. (2013). Morphology, molecular characterisation and systematic position of the genus *Cynura* Cobb, 1920 (Nematoda: Plectida). *Nematology* *15*, 611–627.
- Horikawa, D.D., Cumbers, J., Sakakibara, I., Dana Rogoff, D., Leuko, S., Harnoto, R., Arakawa, K., Katayama, T., Kunieda, T., Toyoda, A., et al. (2013). Analysis of DNA repair and protection in the tardigrade *Ramazzottius varieornatus* and *Hyphibius dujardini* after exposure to UVC radiation. *PLoS One* *8*, e64793.
- Jönsson, K.I., and Schill, R.O. (2007). Induction of Hsp70 by desiccation, ionising radiation and heat-shock in the eutardigrade *Richtersius coronifer*. *Comp. Biochem. Physiol. B* *146*, 456–460.
- Kikuchi, T., Cotton, J.A., Dalzell, J.J., Hasegawa, K., Kanzaki, N., McVeigh, P., Takahashi, T., Tsai, I.J., Assefa, S.A., Cock, P.J.A., et al. (2011). Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog.* *7*, e1002219.
- Krasileva, K.V., Buffalo, V., Bailey, P., Pearce, S., Ayling, S., Tabbita, F., Soria, M., Wang, S., IWGS Consortium, Akhunov, E., et al. (2013). Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biol.* *14*, R66.
- Kraus, C., Schiffer, P.H., Kagoshima, H., Hiraki, H., Vogt, T., Kroiher, M., Kohara, Y., and Schierenberg, E. (2017). Differences in the genetic control of early egg development and reproduction between *C. elegans* and its parthenogenetic relative *D. coronatus*. *Evodevo* *8*, 1–17.
- Kuwabara, P.E., and Shah, S. (1994). Cloning by synteny: identifying *C. briggsae* homologues of *C. elegans* genes. *Nucleic Acids Res.* *22*, 4414–4418.
- Laetsch, D.R., and Blaxter, M.L. (2017). KinFin: software for taxon-aware analysis of clustered protein sequences. *G3 (Bethesda)* *7*, 3349–3357.
- Lee, R.Y.N., Howe, K.L., Harris, T.W., Arnaboldi, V., Cain, S., Chan, J., Chen, W.J., Davis, P., Gao, S., Grove, C., et al. (2018). WormBase 2017: molting into a new stage. *Nucleic Acids Res.* *46*, D869–D874.
- Lefort, V., Desper, R., and Gascuel, O. (2015). FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* *32*, 2798–2800.
- Lewis, S.C., Dyal, L.A., Hilburn, C.F., Weitz, S., Liao, W.-S., Lamunyon, C.W., and Denver, D.R. (2009). Molecular evolution in *Panagrolaimus* nematodes: origins of parthenogenesis, hermaphroditism and the Antarctic species *P. davidi*. *BMC Evol. Biol.* *9*, 15.
- Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* *13*, 2178–2189.
- Linard, B., Allot, A., Schneider, R., Morel, C., Ripp, R., Bigler, M., Thompson, J.D., Poch, O., and Lecompte, O. (2015). OrthoInspector 2.0: software and database updates. *Bioinformatics* *31*, 447–448.
- Lucas-Lledó, J.I., and Lynch, M. (2009). Evolution of mutation rates: phylogenomic analysis of the photolyase/cryptochrome family. *Mol. Biol. Evol.* *26*, 1143–1153.
- Lunt, D.H. (2008). Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. *BMC Evol.* *8*, 194.
- Lunt, D.H., Kumar, S., Koutsovoulos, G., and Blaxter, M.L. (2014). The complex hybrid origins of the root knot nematodes revealed through comparative genomics. *PeerJ* *2*, e356.
- Lutes, A.A., Neaves, W.B., Baumann, D.P., Wieggraabe, W., and Baumann, P. (2010). Sister chromosome pairing maintains heterozygosity in parthenogenetic lizards. *Nature* *464*, 283–286.
- Lynch, M., Bürger, R., Butcher, D., and Gabriel, W. (1993). The mutational meltdown in asexual populations. *J. Hered.* *84*, 339–344.
- Lynch, M. (1984). Destabilizing hybridization, general-purpose genotypes and geographic parthenogenesis. *Q. Rev. Biol.* *59*, 257–290.
- McGill, L.M., Shannon, A.J., Pisani, D., Félix, M.-A., Ramlöv, H., Dix, I., Wharton, D.A., and Burnell, A.M. (2015). Anhydrobiosis and freezing-tolerance: adaptations that facilitate the establishment of *Panagrolaimus* nematodes in polar habitats. *PLoS One* *10*, e0116084–24.
- Neiman, M., Meirmans, S., and Meirmans, P.G. (2009). What can asexual lineage age tell us about the maintenance of sex? *Ann. N. Y. Acad. Sci.* *1168*, 185–200.
- Nowell, R.W., Almeida, P., Wilson, C.G., Smith, T.P., Fontaneto, D., Crisp, A., Micklem, G., Tunnacliffe, A., Boschetti, C., and Barraclough, T.G. (2018). Comparative genomics of bdelloid rotifers: insights from desiccating and nondesiccating species. *PLoS Biol.* *16*, e2004830.
- Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T.D., et al. (2008). Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. U S A* *105*, 14802–14807.
- Partridge, M., and Murphy, D.J. (2009). Roles of a membrane-bound caleosin and putative peroxygenase in biotic and abiotic stress responses in *Arabidopsis*. *Plant Physiol. Biochem.* *47*, 796–806.
- Rancurel, C., Legrand, L., and Danchin, E. (2017). Alieness: rapid detection of candidate horizontal gene transfers across the tree of life. *Genes* *8*, 248.
- Rearson, W., Chakrabortee, S., Pereira, T.C., Tyson, T., Banton, M.C., Dolan, K.M., Culleton, B.A., Wise, M.J., Burnell, A.M., and Tunnacliffe, A. (2010). Expression profiling and cross-species RNA interference (RNAi) of desiccation-induced transcripts in the anhydrobiotic nematode *Aphelenchus avenae*. *BMC Mol. Biol.* *11*, 6.
- Rosic, S., Amouroux, R., Requena, C.E., Gomes, A., Emperle, M., Beltran, T., Rane JKm Linnett, S., Selkirk, M.E., Schiffer, P.H., Bancroft, A.J., et al. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nat. Genet.* *50*, 1–12.
- Schiffer, P.H., Nsah, N.A., Grotehusmann, H., Kroiher, M., Loer, C., and Schierenberg, E. (2014). Developmental variations among *Panagrolaimus* nematodes indicate developmental system drift within a small taxonomic unit. *Dev. Genes Evol.* *224*, 183–188.
- Schiffer, P.H., Polsky, A.L., Cole, A.G., Camps, J.I.R., Kroiher, M., Silver, D.H., Grishkevich, V., Anavy, L., Koutsovoulos, G., Hashimshony, T., and Yanai, I. (2018). The gene regulatory program of *Acroboloides nanus* reveals conservation of phylum-specific expression. *Proc. Natl. Acad. Sci. U S A* *115*, 201720817–201724464.
- Shannon, A.J., Brown, J.A., Boyd, J., Fitzpatrick, A., and Burnell, A.M. (2005). The anhydrobiotic

potential and molecular phylogenetics of species and strains of *Panagrolaimus* (Nematoda, Panagrolaimidae). *J. Exp. Biol.* 208, 2433–2445.

Signorovitch, A., Hur, J., Gladyshev, E., and Meselson, M. (2015). Allele sharing and evidence for sexuality in a mitochondrial clade of bdelloid rotifers. *Genetics* 200, 581–590.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212.

Srinivasan, J., Dillman, A.R., Macchietto, M.G., Heikkinen, L., Lakso, M., Fracchia, K.M., Antoshechkin, I., Mortazavi, A., Wong, G., and Sternberg, P.W. (2013). The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics* 193, 1279–1295.

Streit, A., Li, W., Robertson, B., Schein, J., Kamal, I.H., Marra, M., and Wood, W.B. (1999). Homologs of the *Caenorhabditis elegans*

masculinizing gene *her-1* in *C. briggsae* and the filarial parasite *Brugia malayi*. *Genetics* 152, 1573–1584.

Thorne, M.A.S., Kagoshima, H., Clark, M.S., Marshall, C.J., and Wharton, D.A. (2014). Molecular analysis of the cold tolerant Antarctic nematode, *Panagrolaimus davidi*. *PLoS One* 9, e104526.

Tyson, T., O'Mahony Zamora, G., Wong, S., Skelton, M., Daly, B., Jones, J.T., Mulvihill, E.D., Elsworth, B., Phillips, M., Blaxter, M., and Burnell, A.M. (2012). A molecular analysis of desiccation tolerance mechanisms in the anhydrobiotic nematode *Panagrolaimus superbus* using expressed sequenced tags. *BMC Res. Notes* 5, 68.

Van Der Kooij, C.J., Matthey-Doret, C., and Schwander, T. (2017). Evolution and comparative ecology of parthenogenesis in haplodiploid arthropods. *Evol. Lett.* 1, 304–316.

Vrijenhoek, R.C., and Davis Parker, E. (2009). Geographical parthenogenesis: general purpose genotypes and Frozen Niche variation. In *Lost*

Sex, I. Schön, M. Koen, and P. van Dijk, eds. (Springer Science & Business Media), pp. 99–131.

Welch, D.B.M., and Meselson, M. (2000). Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. *Science* 288, 1211–1215.

Yamashita, M., Jiang, J., Onozato, H., Nakanishi, T., and Nagahama, Y. (1993). A tripolar spindle formed at meiosis I assures the retention of the original ploidy in the gynogenetic triploid crucian carp, *Ginbuna carassius auratus langsdorfii*. *Dev. Growth Differ.* 35, 631–636.

Yoshida, Y., Koutsovoulos, G., Laetsch, D.R., Stevens, L., Kumar, S., Horikawa, D.D., Ishino, K., Konie, S., Kunieda, T., Tomoita, M., et al. (2017). Comparative genomics of the tardigrades *Hypsibius dujardini* and *Ramazzottius varieornatus*. *PLoS Biol.* 15, e2002266.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829.

**ISCI, Volume 21**

## **Supplemental Information**

### **Signatures of the Evolution of Parthenogenesis and Cryptobiosis in the Genomes of Panagrolaimid Nematodes**

**Philipp H. Schiffer, Etienne G.J. Danchin, Ann M. Burnell, Christopher J. Creevey, Simon Wong, Ilona Dix, Georgina O'Mahony, Bridget A. Culleton, Corinne Rancurel, Gary Stier, Elizabeth A. Martínez-Salazar, Aleksandra Marconi, Urmi Trivedi, Michael Kroiher, Michael A.S. Thorne, Einhard Schierenberg, Thomas Wiehe, and Mark Blaxter**

## Supplementary Results and Discussion

### Repeats

Combining RepeatModeller and RepeatMasker results we found only low proportions of repeats in the *Panagrolaimus* and *Propanagrolaimus* genomes. *P. sp.* ES5 had 7.96%, *P. superbis* 8.43 %, *P. sp.* PS1159 6.84% and *Prop. sp.* JU765 9.10% masked bases (we did not evaluate the fragmented *P. sp.* DAW1 genome). These data are similar to the proportions reported for the *P. redivivus* genome (7.1%) (Srinivasan et al. 2013), but considerably less than observed for the *C. elegans* (16.5%) and *C. briggsae* (22.4%) genomes. Using the *k*-mer based approach we identified not only genomic regions that had already been identified and masked by the RepeatModeler and RepeatMasker pipeline, but also additional candidates. This increased the total repeat proportion in the genomes by 1-2% to between 8-10% in total.

	Exact p-value    KS-statistic	Exact p-value    KS-statistic
<i>P. sp.</i> PS1159	0.021    16.49	0.015    34.18
<i>P. sp.</i> DAW1	0.0024    36.83	0.036    3.32
	<i>P. sp.</i> ES5	<i>P. sp.</i> JU765

*Supplementary Table 1: Results from Kolmogorov-Smirnov test comparing the variant distributions in triploid parthenogenetic (PS1159, P. sp. DAW1) and diploid amphimictic (ES5) and hermaphroditic (JU765) species.*

### Transposons

We detected a helitron transposon in a *P. redivivus* small heat shock protein gene. Helitrons are eukaryotic transposons that are predicted to amplify by a rolling-circle mechanism (Kapitonov and Jurka 2001). There may be as many as 560 helitrons in the *C. elegans* genome, located in heterochromatin and in the gene-poor chromosome arms. However, helitrons have been identified in only six *C. elegans* protein coding genes (Eki et al. 2007) and our BLASTp searches failed to find any additional helitron insertions in *C. elegans* coding sequences. By contrast helitron insertions appear to be quite abundant in *Panagrolaimus* protein coding genes. We identified an expansion of helitron Pfams in our InterProScan analysis (supplementary Fig. 9) and additional BLASTP screens indicated the

presence of 27 independent insertions in PS1159 coding sequences and 49 helitron insertions in *P. superbus* genes.

## Developmental Systems

We have previously shown that developmental system drift (DSD) (True and Haag 2001) has modified the gene regulatory networks (GRNs) of endoderm and mesoderm formation between *Panagrolaimus* and *Propanagrolaimus* (Schiffer et al. 2014). Specifically, the expression pattern of the *skn-1* orthologue, a key gene in determining *C. elegans* cell fates, changes between these genera: *skn-1* is expressed during germline specification in *Panagrolaimus*, but not *Propanagrolaimus* (Schiffer et al. 2014). In the current study, we enhanced our analyses by combining the stringent clustering of OrthoMCL with OrthoInspector, which is able to detect more remotely-connected paralogues. For example, in the endoderm/mesoderm induction pathway an *elt-2* orthologue (which is required in *C. elegans* for initiating and maintaining terminal differentiation of the intestine) was not identified in any species outside *C. elegans* by OrthoMCL, but is found by OrthoInspector. However, we found that many genes from the endoderm and mesoderm differentiation pathways in *C. elegans* are missing in all tylenchine (Clade IV) species (supplementary fig. 7) even when using OrthoInspector. The missing genes take part in intermediate switches in the endoderm and mesoderm GRNs. Panagrolaimidae and other tylenchine species are missing not only the *med* and *end* genes, but also have no orthologues of *tbx-35* (which encodes a T-box transcription factor) and *ceh-51* (encoding a homeodomain transcription factor), all of which act downstream of *skn-1* (Maduro and Rothman 2002).

The sex determination pathway has special relevance for the evolution of parthenogenesis, as it is expected that parthenogens will rapidly lose male function to avoid the cost of male production in the absence of the benefit of recombination. We found that many components of the *C. elegans* sex determination pathway were absent either from the panagrolaimids, or from most of the other tylenchine species analysed (Supplementary Excel file Developmental Orthologues). Again, we found a difference between OrthoMCL and OrthoInspector. While neither detected orthologues of the F-Box protein *fog-2* in any of the tylenchine species (or *A. suum*), only OrthoInspector identified the huge number of F-Box in-paralogues known to be present in *C. elegans* (Schiffer et al. 2016). The global picture for the regulation of the panagrolaimid sex determination pathway thus resembled that of the endoderm/mesoderm GRN with upstream (*sex-1*) and downstream players (*fem-2*, *tra-1*) conserved, while intermediate switches were missing (e.g. *her-1*, *tra-2*) (Supplementary Excel file Developmental Orthologues). We found no orthologues of *sdc-*

genes, which act in *C. elegans* dosage compensation, in the tylenchine species. We did find orthologues of *mab-3* (*Drosophila* Doublesex) and related proteins *dmd-4*, *dmd-5*, which are important for male formation in *C. elegans*. Thus, some genes in the canonical sex determination pathway appear to be conserved and could be targets for molecular knock-out studies to investigate the pattern of loss of residual male function in the parthenogenetic taxa.

Sperm function is expected to degrade in parthenogens. *C. elegans* and most other nematode sperm utilize a distinct, non-flagellar motility system involving members of the Major Sperm Protein family (MSP). We identified expressed MSP genes in the *Panagrolaimus* species, including the parthenogens. *C. elegans spe-41* encodes a calcium channel which is crucial for fertilisation in *C. elegans* (Xu and Sternberg 2003), and calcium channels are generally important for fertilisation in animals (Stricker 1999). We found an orthologue of *spe-41* expressed in the parthenogenetic species. We hypothesise that in a hybrid system, and even more so in a polyploid one, imperfect dosage compensation and chromatin-based silencing could lead to expression 'leakage' of 'male-genes', which could then act in the induction of development. It has however been shown that *C. elegans* MSPs can have a role in somatic cells (Schmitz et al. 2007; Schwartz et al. 2012). Thus, similar to the observation of MSP expression in parthenogenetic *A. nanus* and *D. coronatus* (Heger et al. 2010) our finding does not necessarily imply a role of this gene in the reproduction of parthenogenetic panagrolaims. Equally, it is possible that *spe-41* could have a so far undiscovered role in other than sperm cells. It will thus be necessary to conduct differential single cell-transcriptomics analyses and gene expression visualisation studies in the parthenogenetic and closely related male-female species to study the role of these and other genes in asexual reproduction.

### **DNA repair and parthenogenesis**

Efficient DNA repair is particularly crucial for anhydrobiotic species, since exposure to frequent desiccation can lead to DNA strand breaks (Hespeels et al. 2014), and could be equally important under parthenogenesis to avoid mutational meltdown by Muller's ratchet (Muller 1964; Lynch et al. 1993). We identified components of the *C. elegans* DNA repair system in the tylenchine species, and contrasted amphimictic and parthenogenetic *Panagrolaimus* species. We found that most loci were either universally present or absent across Tylenchina (Supplementary Excel file Developmental Orthologues). One exception is *mlh-1*, involved in mismatch repair, which is apparently absent in the parthenogenetic

*Panagrolaimus* species, but present in the amphimictic representatives and in other tylenchines. To explore and independently validate this pattern of absence, we used figmap (Curran et al. 2014) to build a gene model based on motifs in the *C. elegans* gene and screened the *P. sp.* PS1159 genome. We found one candidate in this genome which was paralogous to the orthologues of *mlh-1* in *P. sp.* ES5, *P. superbus*, and the outgroup. We identified a *Panagrolaimus*-specific region at the N-terminal end of the protein, as well as residues specific to the parthenogenetic species. A splits network (not shown) constructed for the proteins also indicates *Panagrolaimus*-specific divergence. Thus, while the general pattern of presence or absence for orthologues in comparison to the model organism does not allow for inferences about loss or gain of function in the panagrolaims, changes in single genes identifiable in our dataset open avenues for future studies into the molecular underpinnings of parthenogenesis.

### **Genomic response to cryptobiosis**

We found several gene families that could play a role in survival under cryptobiosis to be inflated in number in the *Panagrolaimus* species (also see main text for enrichment results). In particular, small Heat Shock Proteins (supplementary figs. 3, 8), as well as Ubiquitin family genes, BTB/POZ domain containing genes, HSP70 genes, C-type lectins, Piwi domain containing genes, and DEAD/DEAH box containing genes (supplementary fig. 8). Among this group of inflated gene families the HSP70s are important in the protein folding machinery and stress response mechanisms. Similarly, ubiquitin family genes and small Heat Shock proteins, which have been reported expressed in desiccation-tolerant *Aphelenchus avenae* nematodes (Karim et al. 2009), could play a role in response to cryptobiosis. Their inflation in number in comparison to other clade IV species, especially the *Propanagrolaimus* and *Panagrellus* species are thus promising candidates for an involvement in the adaptation to survival under extreme conditions in *Panagrolaimus*.

Present in all domains of life, C-type lectins are a diverse group of proteins involved in protein-protein, protein-lipid, or protein-nucleic acid interactions. In fish, type II antifreeze proteins are derived from C-type lectins (see (Davies 2014)), although what role they play in nematodes specifically in relation to cryptobiotic function is unclear. However, *P. sp.* DAW1 show an increased transcription of a C-type lectin gene when the nematode is undergoing intracellular freezing (Thorne et al. 2017), suggesting the importance of the gene in the process.

## Loci under positive selection in *Panagrolaimus*

We identified genes that could be under positive selection by comparing dN/dS ratios in species with different reproductive modes and comparing cryptobiotic with non-cryptobiotic panagrolaims (supplementary fig. 9). Homologues of the human DNA polymerase zeta (REV3L) were identified as being under positive selection in parthenogenetic compared to obligate outcrossing *Panagrolaimus*. These might act in DNA repair, maintaining genome integrity in the nematodes. An orthologue of *C. elegans denn-4*, a likely GDP/GTP exchange factor in MAPK pathways, was also possibly under positive selection the parthenogenetic *Panagrolaimus*. Comparing *Panagrolaimus* to other taxa, orthologues of *C. elegans fcd-2*, which acts on cross-linked DNA and DNA damage repair (Vermeulen 2015), had a dN/dS >1, as did helicases similar to *C. elegans* WBGene00010061, an orthologue of human TTF2 (transcription termination factor 2). The TTF2 orthologue will be of interest for follow-up studies, as the human gene is thought to be involved in DNA repair (Hara et al. 1999). In comparison to the parthenogenetic species *fcd-2* had a signature of selection in the sexual species. Similarly, the *Panagrolaimus* orthologue of the *C. elegans* fatty acid and retinol-binding protein *far-6* (Garofalo et al. 2003) appeared to be under positive selection. InterProScan and GO annotation of loci under selection showed significant enrichment of functions such as carbohydrate transport and metabolism, cell wall/membrane/envelope biogenesis, and replication, recombination and repair (supplementary fig. 10, Supplementary Excel file dNdS). Loci contributing to "replication, recombination and repair" annotations include several connected to DNA integrity and processing of nucleic acids (supplementary fig. 9).

## Transparent Methods

### Nematode strains and culture

Nematodes were acquired from collaborators or through sampling by one of us (E.S.) and cultured in the laboratory on low-salt agar plates (Lahl et al. 2003). *Panagrolaimus* sp. PS1159 is an unnamed parthenogenetic species, isolated in North Carolina by J. Millar. *Panagrolaimus* sp. ES5 is obligately outcrossing, and was isolated from dried blackberry twigs by E. Schierenberg in Bornheim, Germany. *Panagrolaimus superbus* DF5050 is also a male-female species, and was isolated from a gull's nest on the island of Surtsey, Iceland by B. Sohlenius. *Propanagrolaimus* sp. JU765, a protandrous hermaphrodite, was isolated by M.-A. Felix in Guangxi, China from a rice paddy. *Panagrolaimus* sp. DL137 was isolated by



D. Denver in Corvallis, Oregon from soil on University Campus, and *Panagrolaimus* sp. PS1579 was isolated by M.-A. Felix in Huntington Gardens, California. Cultures were usually kept at 15°C, but to quickly increase worm populations before DNA extraction, the culture temperature was increased to 25°C. To minimise confounding factors due to population divergence (or even the mixing of samples on one plate) single worms were picked, propagated for a few generations and then sequenced. Stabilates are stored anhydrobiotically at the Universität zu Köln. The parthenogenetic *P.* sp. PS1159 and the hermaphroditic *Prop.* sp. JU765 were bottlenecked for >30 generations in several single offspring lines, while the other 3 species (*P.* sp. ES5, *P. superbus*, *P.* sp. DAW1) were held in large populations and propagated by the transfer of several worms at a time. *Panagrolaimus* sp. DAW1 (previously designated *Panagrolaimus* sp. CB1 and *P. davidi* [Wharton et al. 2017]), was collected from the McMurdo Sound region, Antarctica by D. Wharton. Other tylenchine nematode genomes analysed were obtained from WormBase ([www.wormbase.org](http://www.wormbase.org)).

### **Genome sizes and chromosome numbers**

We estimated genome sizes in several isolates of *Panagrolaimus* using Feulgen image analysis densitometry (FIAD, following the protocols in (Hardie et al. 2016)), integrated optical densities (IODs) from at least 14 nuclei per individual worm were compared with erythrocytes of *G. domesticus* (C-value =1.25 pg) under x100 magnification (immersion oil, nD = 1.5150). The DNA content in picograms was converted to megabases using the formula 1 Mb = 1.022 \* 10<sup>-3</sup> pg or 1 pg = 978 Mb (Dolezel et al. 2003). These measurements were complemented by RT-D-PCR assay (Wilhelm 2003) based on orthologues of *C. elegans* locus ZK682.5 (WBGene00022789), which has been identified as a nematode-wide single copy gene (Mitreva et al. 2011). Chromosomes were counted, after DAPI staining, in oocytes and 1-cell embryos. For this, gravid nematodes were picked from plates and transferred to M9 buffer and dissected with an insect needle to release gonadal tubes and embryos. These were transferred with a capillary tube into a small drop of 2µg/ml DAPI (Sigma) on a microscope slide, squashed under a coverslip in the stain and the preparation was sealed with nail varnish. After 10-15 min staining the preparations were analysed under an Olympus FluoView1000 confocal microscope with a 60x (NA 1.35) oil objective. Karyotypes were taken as valid only when confirmed in 10 independent chromosome sets. See supplementary fig. 1.

## Genome Assembly

We constructed genome assemblies from Illumina short read libraries. After running first pass assemblies with the CLC Assembly Cell v.4.2 we applied the khmer pipeline (Brown et al. 2012) to digitally normalise read coverage on ES5, *P. superbus*, JU765 and PS1159 data. We identified and removed contaminating bacterial sequences using the blobtools approach (Kumar et al. 2013). The cleaned reads were reassembled using the Velvet assembler (Zerbino and Birney 2008) exploring different k-mer sizes. We then employed RNA-Seq derived mRNA predictions (see below) to scaffold the genomes with the SCUBAT pipeline (<https://github.com/elswob/SCUBAT>). For *P. superbus* the genome input to SCUBAT originated from a CLC assembly, as this proved better than the tested Velvet assemblies. For *P. sp.* DAW1, where the original assembly (Thorne et al. 2014) was of low completeness (39% complete KOGs as assessed by CEGMA), we (re-)assembled the genomes with SPAdes (Bankevich et al. 2012) and the redundans pipeline (Pryszcz and Gabaldón 2016). We also used SPAdes and the redundans pipeline to assemble the *P. sp.* PS1579 genome. Assembly qualities were evaluated with custom scripts. We utilized GNU Parallel (Tange 2011) in various steps of the assembly pipeline and in downstream analyses.

To assemble mitochondrial genomes contigs generated in the main genome assemblies were extended with Illumina reads using IMAGE (Tsai et al. 2010), then aligned to reference with ABACAS (Assefa et al. 2009) and, finally, any remaining gaps in the sequence filled with GapFiller (Boetzer and Pirovano 2012). All assembled mitochondrial genomes were annotated using the MITOS2 (Bernt et al. 2013) online pipeline. Additionally, the boundaries of protein coding genes (PCGs) in the new genomes were manually curated using the ORF finder tool (<https://www.ncbi.nlm.nih.gov/orffinder/>) and BLAST+ searches, while the tRNA predictions made by MITOS2 were verified using the online version of ARWEN (Laslett and Canbäck 2008). See supplementary fig 4 for a comparison of the mitochondrial genomes.

## Transcriptome assembly

To aid gene annotation and to confirm expression of important genes we sequenced transcriptomes of mixed life cycle stages in *P. sp.* PS1159, *P. sp.* PS1579, *P. sp.* ES5, and *Prop. sp.* JU765 using Illumina TruSeq RNA-Seq. Raw reads were adaptor and quality trimmed with Trimmomatic (versions  $\leq 0.32$ ) before assembly with Trinity (Haas et al. 2013). We further sequenced the *de novo* transcriptomes of *P. sp.* DL137 (see also (Schiffer et al. 2014)). For *P. superbus* a transcriptome enriched for nematodes in the anhydrobiotic stage was generated using the Roche 454 platform and SMART cDNA synthesis. We assessed

the completeness of our transcriptome assemblies using CEGMA and also with BUSCO (through gVolante [Nishimura et al. 2017]) using the eukaryotic gene set as reference.

### **Repetitive DNA**

We used a two-pronged approach to screen for repetitive elements in the final draft genomes. First we used RepeatModeler (open-1.0; <http://www.repeatmasker.org/RepeatModeler.html>) to identify and RepeatMasker (v. open-4.0.5) (Smit 2015) to mask genomes using the BLAST-based search engine. In addition we used *LTRharvest* (Ellinghaus et al. 2008) implemented in the GenomeTools analysis system (Gremme et al. 2013) to screen for transposons missed by RepeatModeler, but did not detect any additional transposons.

As genomes assembled from short insert libraries may collapse repeats into a single contig, we screened the genomes with a second, k-mer based assay to identify regions with high base coverage in the raw data that could be repetitive. Following a protocol established by Dan Bolser (unpublished; <https://github.com/dbolser/PGSC/blob/master/kmer-filter/README>), which relies on 'tallymer' from the *GenomeTools* package, we counted the frequency of all unique and repeated k-mers of sizes 10 - 50 bases. For each genome we then picked a k-mer size close to the value where frequency curves plateaued and counted the occurrence of k-mers at coverage levels 10 - 50. We identified the sequences in the genome corresponding to these k-mers and analysed the base coverage for these genomic regions based on CLC mappings with the aid of BEDtools (Quinlan and Hall 2010). We compared these empirical coverage distributions with genome wide coverage and selected an additional set of potential repeats to add to the RepeatModeler set. This unified set was then fed back into RepeatMasker for a second round of masking.

### **Gene Prediction**

We used Augustus (v.3.0.1) (Stanke and Waack 2003) for gene prediction, employing an iterative approach to generate the most credible predictions. Augustus was trained with CEGMA-predicted genes. Wherever RNA-Seq data were available (*P. sp.* ES5, *P. sp.* PS1159, *Prop. sp.* JU765), it was assembled based on alignment to the genome, using gsnap (Wu and Nacu 2010), which has been shown to perform well in finding correct splice sites (Engström et al. 2013). These mappings then served as evidence for Augustus in a second round of gene prediction. For the first round of Augustus training in *P. superbus* we used exonerate (Slater and Birney 2005) to align predicted *P. sp.* PS1159 proteins, as no large-scale Illumina RNA-Seq data were available. We then mapped findorf (Krasileva et al.

2013) predicted ORFs (see below) from the 454-sequencing derived *P. superbus* transcriptome using BLAT (Kent 2002) and created Augustus evidence from this. For *P. sp. DAW1* no transcriptomic data were available at the time of our analyses, and we used the *P. sp. PS1159* species model in Augustus along with the *P. sp. DAW1* CEGMA genes as hints. We implemented a best hit BLAST approach using different taxonomic databases to identify likely contamination in predicted genes. Candidate contamination was removed from the predicted gene sets and from the corresponding contigs in the genome assemblies. In the ES5 genome we identified an excess of *E. coli* contamination and employed mugsy (Angiuoli et al. 2011) for full genome alignments between the nematode and *E. coli* to identify and remove the corresponding contigs. Contamination was also additionally screened for in the HGT assay.

Open reading frames were predicted from *de novo* transcriptome assemblies of *P. superbus* (454-sequencing derived data, as well as DL137 and PS1579 (both high coverage Illumina RNA-Seq) using findorf. The findorf approach involved comparison of transcripts with proteomes from other nematode species (*Brugia malayi*, *Caenorhabditis briggsae*, *Caenorhabditis elegans*, *P. sp. ES5*, *Prop. sp. JU765*, *Meloidogyne hapla*, *Panagrellus redivius*, *P. sp. PS1159*, and preliminary data from *Plectus sambesii*) using BLASTX to identify frameshifts and premature stop codons, and hidden Markov models (HMMs) to identify Pfam domains to infer ORF start positions. In total, findorf inferred 21,381 *P. superbus* ORFs, 34,868 *P. sp. DL137* ORFs and 47,754 *P. sp. PS1579* ORFs.

Codon usage for the predicted genes was analysed with GCUA (McInerney 1998) and codonW (v.1.4.4; <http://codonw.sourceforge.net>). We also used codonW to compare codon usage between genes putatively gained through HGT and the genomic background of the host nematodes. We used tRNAscan-SE (v1.3.1) (Lowe and Eddy 1997) to identify tRNA genes, selecting the implemented Cove probabilistic RNA prediction package as the search engine.

### **Proteome Annotation**

Putative domains in 13 nematode proteomes were inferred using InterProScan (v5.7-48.0) (Jones et al. 2014). We analysed all our *Panagrolaimus* species (including both transcriptomic and genomic data for *P. superbus*), our *Propanagrolaimus* outgroup, and the two remote outgroup species *Caenorhabditis elegans* and *Ascaris suum*. A summary of the Pfam domains annotation is deposited on the accompanying genome hubs webpage for the *Panagrolaimus* genomes.

## Orthologous Proteins

We inferred orthology between proteins from our newly sequenced panagrolaimids and a set of outgroup species from the Tylenchina (Clade IV of Nematoda *sensu* Blaxter (Blaxter et al. 1998) using OrthoMCL (v.2.0.8) (Li et al. 2003), OrthoInspector (v.2.11) (Linard et al. 2015), and OrthoFinder (Emms and Kelly 2015). We also included data from the cephalobid species *Acrobeloides nanus* (Schiffer et al. 2018), as well as the model *C. elegans* (Rhabditina; Clade V) and *Ascaris suum* (Spirurina; Clade III). To detect orthologs to key *C. elegans* developmental genes (see section on GRNs) we primarily relied on the extensively tested and robust OrthoMCL pipeline based on NCBI blast, which we validated with OrthoInspector. The latter found additional divergent orthologous, which were hidden to OrthoMCL. OrthoFinder based on Diamond (Buchfink et al. 2014) BLAST was only used in the GRAMPA (Thomas et al. 2016) analysis, since the output contains gene trees for each cluster (see section on ploidy below). In cases where OrthoMCL and OrthoInspector grossly disagreed or the presence/absence pattern across species appeared inconsistent in the GRN analysis, we directly screened genomes for the presence of a gene using the HMM profile figmap pipeline (Curran et al. 2014).

While our use of additional orthology finding programs and the figmap pipeline added confidence to the inference of absence of genes, it is still possible that orthologous genes were not detected due to rapid sequence evolution in Nematoda. Thus, we wanted to implement an additional test: It had previously been shown that using a synteny approach some particularly fast evolving genes in the sex determination pathway of *Caenorhabditis* species can be identified, as collinearity is maintained (Kuwabara and Shah 1994, Streit et al. 1999, Haag et al. 2002). Consequently, we performed a synteny analysis with MCScanX (Wang et al. 2012). For this we compared the genomes of *Panagrolaimus* PS1159, *Propanagrolaimus* JU765, which had the best contiguity of our newly assembled genomes and included the *Bursaphelenchus xylophilus* genome as an independent control. We used lenient (non-standard) parameters (evalue:1e-03 (default: 1e-5); match size: 2 (d: 5); maximum distance: 10 (d: 5)) in MCScanX, and as a proof of principle also compared the genomes of *C. elegans*, *C. briggsae*, and *C. remanei*.

## Phylogenetic analyses

To increase confidence in the phylogenetic positions of species in our analyses we extended an existing phylogeny (Lewis et al. 2009) by adding all 452 proteins predicted by the CEGMA pipeline. We constructed individual alignments for each cluster of orthologous CEGMA proteins from our species and outgroups with Clustal Omega (v1.2) (Sievers et al.

2011) and then used trimAl (v1.4.rev15) (Capella-Gutiérrez et al. 2009) to exclude ambiguous regions. A supermatrix combining all trimmed alignments was then constructed with the aid of phyutility (v2.2.6) (Smith and Dunn 2008).

Phylogenies were inferred using the MPI version of PhyloBayes (1v.4f) (Lartillot et al. 2013) with the CAT model implemented therein and RAxML (v7.7.2) (Stamatakis 2006) using the VT model under the GAMMA parameter, as predicted by Prottest (v3.2) (Darriba et al. 2011). We let RAxML automatically stop the bootstrap replicates sampling needed to reach convergence and it found 56 trees after 861 bootstraps. Bayesian analyses in the MPI version of PhyloBayes did not converge on a single tree after more than 8,000 generations in four parallel chains. This was due to *P. sp. DAW1* swapping between two positions within the group of parthenogenetic *Panagrolaimus* species. However, tree topologies of both methods are otherwise congruent. Phylogenetic inferences for single proteins, or groups of orthologous (e.g. sHSP), were conducted with Clustal Omega, trimAl, protest3, and RAxML as described for the CEGMA KOGs. We employed Jalview (Waterhouse et al. 2009) and SeaView (Galtier et al. 1996) to visualise alignments, and used SplitsTree4 (Huson and Bryant 2005) to explore relationships in gene families. The phylogenetic inference was carried out using the CHEOPS cluster at the University of Köln.

### **Assessment of ploidy**

We first mapped reads against the assembled small subunit ribosomal sequences in each species and identified variants. This suggested the presence of a significant "minor allelic variant" in the parthenogenetic species. Both the hermaphroditic species *Prop. sp. JU765*, expected to be diploid, and the potentially polyploid parthenogenetic strain *P. sp. PS1159*, were inbred through 30 generations of single offspring propagation; from this we expected genomes to be largely homogenised. In contrast, *P. sp. DAW1* was expected to be more heterozygous as large populations from multiple plates were used for DNA extraction. Including the parthenogens *Acrobelloides nanus* (cephalobida, clade IV) and *Plectus sambesii* (plectica, clade III), where genomes (but not of sexual sister species) became recently available (Rosic et al. 2018; Schiffer et al. 2018), into this analysis, we aimed to gain a more general insight into the origin of parthenogenesis in *Nematoda*. We extracted all Augustus predicted genes for each species and mapped RNA-Seq read sets against the coding sequences using the CLC mapper (v.5.0), requiring a sequence identity of 90% and a read length threshold of 80%. We performed the same mapping approach with genomic reads against repeat masked and unmasked genomes. We used BamTools (Barnett et al. 2011) (v.2.3.0) to sort the mappings, SAMtools (v1.0-13) (Li et al. 2009) to create mpileups,

VarScan (v2.3.6) (Koboldt et al. 2012) to call variants, and collated the minor variant frequency spectrum (folded site spectrum) for each species. We then implemented a two sample Kolmogorov Smirnov (KS) test available in the Julia (Bezanson et al. 2012) programming language to analyse whether the observed variant frequencies are different under the null hypotheses that the frequencies are sampled from the same distribution (Supplementary Table 1): we first sampled 20,000 data points from each distribution and directly compared these with the KS test; next we ran 50,000 bootstrap replicates on 20,000 randomly mixed samples per distribution, and then measured the exact p-value in the bootstrapped distribution.

For the GRAMPA analysis testing allo- vs. auto-polyploidy we used all ~15k FastME trees from running Orthofinder mid-point rooted with NOTUNG (v2.8.1.7) (Chen et al. 2000). To validate these results we also used a more comprehensive initial dataset including the DL137 and PS1579 proteomes and selected 2313 groups of orthologues from OrthoMCL that had a 2:1 ratio of parthenogenetic to sexual species. For each group we built an alignment with clustal-omega, inferred a protein tree with RAxML and ran GRAMPA. This resulted in the same topology as the OrthoFinder FastME based GRAMPA tree shown in Fig 1.

### **Estimation of divergence times**

We used ANDI to calculate pairwise evolutionary distances between three species of *Panagrolaimus* (*P. superbus*, *P. sp.* ES5, and *P. sp.* PS1159) and two outgroups (*Propanagrolaimus sp.* JU765, and *Panagrellus redivivus*) (see supplementary table 2). We excluded the parthenogenetic *P. sp.* DAW1 from this analysis since the assembly span of 118 Mb and our variant calling (see above) indicate this assembly to contain duplicate regions when compared to the quasi-haploid assembly of PS1159 (85Mb). ANDI is designed for comparisons of entire genomes, and does not require homologous sequences or genes as input. To convert ANDI-distances into divergence measured in generations or in absolute time we used the divergence of *Caenorhabditis briggsae* and *Caenorhabditis sp.* 5 as a reference. The *Caenorhabditis* species are separated by about 171 M generations (i.e. 85.5 M generations since their last common ancestor (Cutter 2008)). Generation time measurements were conducted by Isabel Goebel at the University of Köln (unpublished thesis “Untersuchungen zum Lebenszyklus von Nematoden mit unterschiedlichem Fortpflanzungsmodus”) under supervision of PHS and ES. Briefly, nematodes from parthenogenetic strains *P. sp.* PS1159, and *P. sp.* PS1579, and the sexual species *P. superbus*, *P. sp.* ES5, and *Prop. sp.* JU765 were cultured at room temperature on small *E.*

*coli*-seeded agar plates. These were monitored every few hours during early development and then every day during adult life. Across all species assayed, generation times (Vancoppenolle et al. 1999) averaged 8 days under our laboratory conditions. Since there is some uncertainty in the generation length in nature we estimated timings based on 4, 8, 16, 32, and 50 days per generation (corresponding to around 91, 46, 231, 111 and 7 generations per year, respectively).

### **Identification of conserved protein domains with significantly different abundances in *Panagrolaimid* and other nematodes**

We identified Pfam domain annotations in the proteomes of our species set using InterProScan. Statistical differences in abundance of individual Pfam domains was analysed using a two-sided Fisher's exact test (Null Hypothesis Significance Testing - NHST), corrected for multiple testing using the Benjamin-Hochberg algorithm (for false discovery rate) implemented in the R (R Core Team) statistical language. We also analysed transcriptome assemblies from *P. sp.* DL137 (for which we have no genome) and *P. sp.* PS1159 (to retrieve proteins potentially missed in the genome-wide annotations).

To support and further investigate the results of the NHST with an independent method, we employed classifier analysis in a support vector machine (SVM) framework to address two pivotal questions:

- (1) Do the patterns of annotation frequency derived from InterProScan correspond to defining differences between species groups?
- (2) Are the annotations that together form separate and clearly identifiable patterns for each group associated with functions that differ between them?

Given a number of samples, each belonging to one of  $x$  possible classes, a classifier attempts to learn the underlying sample-class mapping. Classes are labels with (in the present case) three possible values (*Panagrolaimus*, *Propanagrolaimus*, *Panagrellus*). Samples were  $n$ -dimensional vectors of annotation presence and absence. The classifier was trained on a set of samples, including an equal number of representatives from both classes and then tested on its ability to predict the class of an independent sample. We used 30 randomly drawn annotations from each representative species as dimensions in the classifier and ran 2000 iterations, each time selecting a different subset of training samples and test species. We tested variations relying on 50 or 100 dimensions (annotations), but found that these did not change the overall pattern of results. The analyses based on thirty dimensions are presented.



Classification of Panagrolaimidae achieved 55 % accuracy (standard error of the mean [SEM] = 0.006; chance level classification = 50%) by an SVM based on randomly selecting 30 Pfam domains in each of 2000 iterations. The same approach yielded 58% mean accuracy (SEM = 0.006) classification for the Outgroup; the difference between classification performance for Panagrolaimidae and Outgroup was not statistically significant in a two-samples t-test ( $T(1999) < 1$ ), showing that the classifier was not biased towards one class. More importantly, classification performance was significantly better than control classification (where sample-class mapping in the training set was permuted). Based on a permuted class-sample mapping in the training set, the classifier achieved a mean accuracy for Panagrolaimidae of 49 % (SEM = 0.006) and of 49% for the Outgroup (SEM = 0.008). The comparison between the real classifier incorporating true sample-to-class mappings in the training set and the permuted classifier was statistically significant for Panagrolaimidae:  $T = 7$ ,  $p < 0.001$ ,  $df = 1999$ ; and for the outgroup:  $T = 7.4$ ,  $p < 0.001$ ,  $df = 1999$  (see supplementary fig. 5). During classification, we recorded annotations that were most often part of successful classification iterations. These lists were then intersected with results from NHST analysis and interpreted with reference to identifying differences in the biology of each species group.

### **Detection of Horizontal Gene Transfers**

To detect candidate horizontal gene transfers (HGT), we used Alieness (Rancurel et al. 2017) to calculate an Alien Index (AI) (Gladyshev et al. 2008; Flot et al. 2013). Briefly, all Panagrolaimidae predicted proteins were compared against the NCBI's nr protein database using BLAST (Altschul et al. 1990) with an E-value threshold of  $1e^{-3}$  and no SEG filtering. BLAST hits were parsed to retrieve associated taxonomic information, using the NCBI taxonomy database. For every Panagrolaimidae protein returning at least one hit in both metazoan and non-metazoan species, we calculated the AI:

$$AI = \ln(\text{best metazoan E-value} + E^{-200}) - \ln(\text{best non-metazoan E-value} + E^{-200})$$

When either no significant metazoan or non-metazoan BLAST hit was found, a penalty E-value of 1 was automatically assigned. To detect HGT events that took place in an ancestor of Panagrolaimidae or its close relatives, BLAST hits to Panagrolaimoidea (TaxID: 55746) and Aphelenchina (TaxID: 1182516) were excluded from the calculation of the AI. An AI > 0 indicates a better hit to a non-metazoan species than to a metazoan species and thus a possible acquisition *via* HGT. An AI > 30 corresponds to a difference of magnitude  $1.0e^{-10}$  between the best non-metazoan and best metazoan E-values and is taken as strong indication of a HGT event. All *Panagrolaimidae* proteins that returned an AI > 0 and that

aligned with  $\geq 70$  % identity to a non-metazoan protein were considered as possible contaminants and were discarded from the analysis. Potential candidates were further validated by identifying gene models that contained spliceosomal introns, were surrounded by *bona fide* nematode genes, had mapped RNA-Seq reads (TPM estimates were calculated with RSEM [Li and Dewey 2011] incorporating bowtie2 [Langmead and Salzberg 2012] mapping) and had codon usages similar to bulk genome values (supplementary fig. 6).

### **Reconstruction of the timing of gene acquisitions *via* HGT**

We used Mesquite (v3.01) (Maddison and Maddison 2018) to reconstruct the timing of gene acquisition *via* HGT. By adding orthology information to HGT candidates, we built a matrix of presence/absence of each HGT candidate across the different *Panagrolaimidae* species. We mapped this matrix to the species phylogeny, including two additional outgroups (*Acrobelloides nanus* and *Bursaphelenchus xylophilus*). Based on presence/absence information, Mesquite was used to trace back ancestral presence / absence at each node, using a parsimony model. When ancestral presence/absence of a gene family at a given node was equally parsimonious, we arbitrarily considered the family as present, because secondary loss of a gene acquired by HGT ancestrally is intuitively more likely than proposing multiple independent HGT events. HGT candidates that were species-specific were considered as acquired specifically in this species.

### **Functional analysis of HGT candidates**

Pfam annotations were retrieved for all HGT candidate proteins, and domains that were conserved among the set of candidate HGT proteins of several *Panagrolaimidae* species identified (Supplementary Excel file HGT – PFAM). We also predicted functions based on the presence of Pfam domains, using pfam2go (available at [geneontology.org/external2go/pfam2go](http://geneontology.org/external2go/pfam2go)) and custom perl scripts. To make Gene Ontology annotations comparable, we mapped raw GO terms to the generic GOSlim ontology. We used the GOSlimViewer, developed as part of AgBase (McCarthy et al. 2006), to map terms to the GOSlim ontology (Supplementary Excel file HGT – GOSlim).

### **Mitochondrial genome assembly and annotation**

To assemble mitochondrial genomes contigs generated in the main genome assemblies were extended with Illumina reads using IMAGE (Tsai et al. 2010), then aligned to reference

with ABACAS (Assefa et al. 2009) and, finally, any remaining gaps in the sequence filled with GapFiller (Boetzer and Pirovano 2012).

All assembled mitochondrial genomes were annotated using the MITOS2 (Bernt et al. 2013) online pipeline. Additionally, the boundaries of protein coding genes (PCGs) in the new genomes were manually curated using the ORF finder tool (<https://www.ncbi.nlm.nih.gov/orffinder/>) and BLAST+ searches, while the tRNA predictions made by MITOS2 were verified using the online version of ARWEN (Laslett and Canbäck 2008). See supplementary Fig 4 for a comparison of the mitochondrial genomes.

### **Codon usage**

Codon usage for the predicted genes was analysed with GCUA (McInerney 1998) and codonW (v.1.4.4; <http://codonw.sourceforge.net>). We also used codonW to compare codon usage between genes putatively gained through HGT (see below) and the genomic background of the host nematodes. We used tRNAscan-SE v1.3.1 (Lowe and Eddy 1997) to identify tRNA genes, selecting the implemented Cove probabilistic RNA prediction package as the search engine.

### **Signatures of adaptive evolution**

We used Crann (Creevey and McInerney 2003) to explore signatures of selection in panagrolaim genes. Over 20,000 OrthoMCL derived orthology cluster protein alignments, built using clustal-omega, were reverse-translated into DNA alignments with trimAl and SAMtools. We retained 7,335 alignments that contained at least one parthenogenetic and one male female *Panagrolaimus* species. Further filtering of short and outlier sequences within each of these gene family alignments, using amino acid versions of the sequence alignments, removed poorly-defined clusters. If the length of any sequence was less than half the length of the overall alignment it was removed. For all remaining sequences in each alignment, two filtering steps were carried out: A distance matrix was constructed from the amino acid versions of the alignments using ProtDist from the PHYLIP (Felsenstein 1993) package and the JTT model of evolution was used to construct a neighbour-joining tree with neighbour from the PHYLIP package. Using custom scripts, the average taxon-taxon path length on the tree was calculated, and outlier sequences which had a significantly greater average path length were identified using R. These sequences were removed from both the DNA and amino acid alignments. The second filter checked for saturation in the nucleotide versions of the alignment. This was carried out by building two trees with PAUP for each alignment, the first using a P distance and the second using the HKY model. The taxon-

taxon length across both trees were calculated as before and both calculations combined for visualization as a scatter plot, using R. Any alignments exhibiting signatures of saturation following visual inspection of the plots were removed from both the DNA and amino acid alignments. The resulting filtered DNA alignments were input to Crann to calculate all pairwise rates of non-synonymous substitutions per non-synonymous site (dN) and synonymous substitution per synonymous site (dS). The average of all pairwise dN/dS ratios between the following categories of species were calculated: (1) sexually-reproducing *Panagrolaimus* species versus parthenogenetic *Panagrolaimus* species; (2) all *Panagrolaimus* species, versus nearest outgroups; (3) all species in analysis versus all. See supplementary fig. 10. The rates of dN/dS were sorted into categories by their functions (Tatusov et al. 1997) (supplementary fig. 9).

## Supplementary References

- Altschul SF, Fish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Molec. Biol.* 3:403-410.
- Angiuoli SV, Dunning H JC, Salzberg SL, Tettelin H. 2011. Improving pan-genome annotation using whole genome multiple alignment. *BMC Bioinformatics* 12:272.
- Assefa S, Keane TM, Otto TD, Newbold C, Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* 25:1968–1969.
- Bankevich A, Nurk S, Antipov D, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19:455–477.
- 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.
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phyl. Evol* 69:313–319.
- Bezanson J, Karpinski S, Shah VB, Edelman A. 2012. Julia: A Fast Dynamic Language for Technical Computing. *arXiv*.
- Blaxter ML, De Ley P, Garey JR, Le Guédard M, Hanano A, Heintz D, Ehrling J, Herrfurth C, Feussner I, Bessoule J-J. 1998. A molecular evolutionary framework for the phylum Nematoda. *Nature* 392:71–75.
- Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol* 13:R56.
- Brown CT, Howe A, Zhang Q, Pyrkosz AB, Brom TH. 2012. A single pass approach to reducing sampling variation, removing errors, and scaling de novo assembly of shotgun sequences. *arXiv*.
- Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND. *Nat Methods.* 12:59–60.
- Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
- Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol.* 7:429–447.
- Curran DM, Gilleard JS, Wasmuth JD. 2014. Figmap: a profile HMM to identify genes and bypass troublesome gene models in draft genomes. *Bioinformatics* 30:3266–3267.
- Cutter AD. 2008. Divergence times in *Caenorhabditis* and *Drosophila* inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol* 25:778–786.
- Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–1165.
- Davies PL. 2014. Ice-binding proteins: a remarkable diversity of structures for stopping and starting ice growth. *Trends in Biochemical Sciences* 39:548–555.
- Eki T, Ishihara T, Katsura I, Hanaoka F. 2007. A genome-wide survey and systematic RNAi-based characterization of helicase-like genes in *Caenorhabditis elegans*. *DNA Res.* 14:183–199.
- Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.
- Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:E9–E13.
- Engström PG, Steijger T, Sipos B, Grant GR, Kahles A, The RGASP Consortium, Räscher G, Goldman N, Hubbard TJ, Harrow J, Guigó R, Bertone P. 2013. Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat Methods.* 10:1185–1191.
- Felsenstein J. 1993. *Phylogenetic Inference Programs (PHYLIP)*.
- Flot J-F, Hespels B, Li X, Noel B, Arkhipova I, Danchin EGJ, Hejnol A, Henrissat B, Koszul R, Aury J-M, Barbe V, Barthelemy R-M, Bast J, Bazykin GA, Chabrol O, Couloux A, Da Rocha M, Da Silva C, Gladyshev E, Gouret P, Hallatschek O, Hecox-Lea B, Labadie K, Lejeune B, Piskurek O, Poulain J, Rodriguez F, Ryan JF, Vakhrusheva OA, Wajnberg E, Wirth B, Yushenova I, Kellis M, Kondrashov AS, Mark Welch DB, Pontarotti P, Weissenbach J, Wincker P, Jaillon O, Van Doninck K. 2013. Genomic evidence for ameiotic evolution in the bdelloid rotifer *Adineta vaga*. *Nature* 500:1–5.
- Galtier N, Gouy M, Gautier C. 1996. SEAVIEW and PHYLO\_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 12:543–548.

- Garofalo A, Rowlinson M-C, Amambua NA, et al. 2003. The FAR protein family of the nematode *Caenorhabditis elegans*. *Journal of Biological Chemistry* 278:8065–8074.
- Gladyshev EA, Meselson M, Arkipova IR. 2008. Massive horizontal gene transfer in bdelloid rotifers. *Science* 320:1210–1213.
- Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)* 10:645–656.
- Haag ES, Wang S, Kimble J. 2002. Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*. *Curr Biol* 12:2035–2041.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. *Nat Protoc.* 8:1494–1512.
- Hara, R, C P Selby, M Liu, D H Price, and A Sancar. 1999. Human transcription release factor 2 dissociates RNA polymerases I and II stalled at a cyclobutane thymine dimer. *Journal of Biological Chemistry* 274: 24779–24786. doi:10.1074/jbc.274.35.24779.
- Hardie DC, Gregory TR, Hebert PDN. 2016. From pixels to picograms: A beginners' guide to genome quantification by Feulgen image analysis densitometry. *Journal of Histochemistry & Cytochemistry* 50:735–749.
- Heger, Peter, Michael Kroiher, Nsah Ndifon, and Einhard Schierenberg. 2010. Conservation of MAP kinase activity and MSP genes in parthenogenetic nematodes. *BMC Dev Biol*: 51. doi: 10.1186/1471-213X-10-51.
- Hespeels B, Knapen M, Hanot-Mambres D, Heuskin AC, Pineux F, Lucas S, Koszul R, van Doninck K. 2014. Gateway to genetic exchange? DNA double-strand breaks in the bdelloid rotifer *Adineta vaga* submitted to desiccation. *J of Evol Biol* 27:1334–1345.
- Huson DH, Bryant D. 2005. Estimating Phylogenetic Trees and Networks Using SplitsTree 4. [www.splitstree.org](http://www.splitstree.org)
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240.
- Karim N, Jones JT, Okada H, Kikuchi T. 2009. Analysis of expressed sequence tags and identification of genes encoding cell-wall-degrading enzymes from the fungivorous nematode *Aphelenchus avenae*. *BMC Genomics* 2009 10:1 10:525.
- Kent WJ. 2002. BLAT---The BLAST-Like Alignment Tool. *Genome Res* 12:656–664.
- Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22: 568–576.
- Krasileva KV, Buffalo V, Bailey P, Pearce S, Ayling S, Tabbita F, Soria M, Wang S, Consortium IWGS, Akhunov E, Uauy C, Dubcovsky J. 2013. Separating homeologs by phasing in the tetraploid wheat transcriptome. *Genome Biol* 14:R66.
- 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.
- Kuwabara PE, Shah S. 1994. Cloning by synteny: identifying *C. briggsae* homologues of *C. elegans* genes. *Nucleic Acids Res* 22:4414–4418.
- Lahl V, Halama C, Schierenberg E. 2003. Comparative and experimental embryogenesis of Plectidae (Nematoda). *Dev Genes Evol.* 213:18–27.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359.
- Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic Biol* 62:611–615.
- Laslett D, Canbäck B. 2008. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24:172–175.
- Lewis SC, Dyal LA, Hilburn CF, Weitz S, Liao W-S, Lamunyon CW, Denver DR. 2009. Molecular evolution in *Panagrolaimus* nematodes: origins of parthenogenesis, hermaphroditism and the Antarctic species *P. davidi*. *BMC Evol Biol* 9:15.

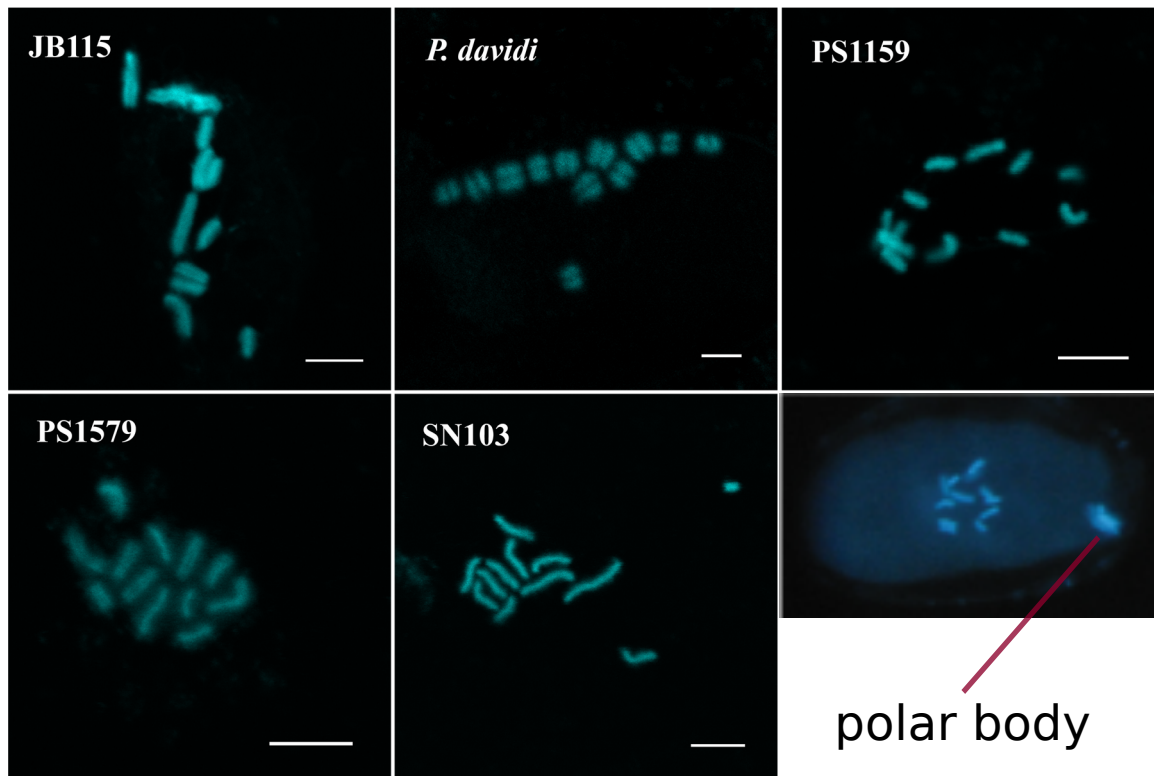
- Li B, Dewey CN. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323.
- Li H, Handsaker B, Wysoker A, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
- Li L, Stoeckert CJ, Roos DS. 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res* 13:2178–2189.
- Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964.
- Lynch M, Bürger R, Butcher D, Gabriel W. 1993. The mutational meltdown in asexual populations. *J Heredity* 84:339–344.
- Linard B, Allot A, Schneider R, Morel C, Ripp R, Bigler M, Thompson JD, Poch O, Lecompte O. 2015. OrthoInspector 2.0: Software and database updates. *Bioinformatics* 31:447–448.
- Maddison, W. P. and D.R. Maddison. 2018. Mesquite: a modular system for evolutionary analysis. Version 3.51. <http://www.mesquiteproject.org>.
- McCarthy FM, Wang N, Magee GB, Nanduri B, Lawrence ML, Camon EB, Barrell DG, Hill DP, Dolan ME, Williams WP, Luthe DS, Bridges SM, Burgess SC. 2006. AgBase: a functional genomics resource for agriculture. *BMC Genomics* 2006 7:229.
- McInerney JO. 1998. GCUA: general codon usage analysis. *Bioinformatics* 14:372–373.
- Mitreva M, Jasmer DP, Zarlenga DS, Wang Z, Abubucker S, Martin J, Taylor CM, Yin Y, Fulton L, Minx P, Yang S-P, Warren WC, Fulton RS, Bhonagiri V, Zhang X, Hallsworth-Pepin K, Clifton SW, McCarter JP, Appleton J, Mardis ER, Wilson RK. 2011. The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat Genetics*. 43:228–235.
- Muller HJ. 1964. The relation of recombination to mutational advance. *Mut. Research* 1:2–9.
- Nishimura O, Hara Y, Kuraku S. 2017. gVolante for Standardizing Completeness Assessment of Genome and Transcriptome Assemblies. *Bioinformatics* 22: 3635–3637.
- Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res* 44:e113–e113.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842.
- Rancurel C, Legrand L, Danchin E. 2017. Alienness: rapid detection of candidate horizontal gene transfers across the tree of life. *Genes* 8:248.
- Rosic S, Amouroux R, Requena CE, Gomes A, Emperle M, Beltran T, Rane JK, Linnett S, Selkirk ME, Schiffer PH, Bancroft AJ, Grencis RK, Jeltsch A, Hajkova P, Sarkies P. 2018. Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nat Genetics*. 50:1–12.
- Schiffer PH, Nsah NA, Grotehusmann H, Kroiher M, Loer C, Schierenberg E. 2014. Developmental variations among Panagrolaimid nematodes indicate developmental system drift within a small taxonomic unit. *Dev. Genes Evol.* 224:183–188.
- Schiffer PH, Gravemeyer J, Rauscher M, Wiehe T. 2016. Ultra large gene families: a matter of adaptation or genomic parasites? *Life* 6.
- Schiffer PH, Polsky AL, Cole AG, Camps JIR, Kroiher M, Silver DH, Grishkevich V, Anavy L, Koutsovoulos G, Hashimshony T, Yanai I. 2018. The gene regulatory program of *Acroboloides nanus* reveals conservation of phylum-specific expression. *Proc Natl Acad Sci* 1:201720817–201724464.
- Schmitz C, Kinge P, Hutter H. 2007. Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive *Caenorhabditis elegans* strain Nre-1(Hd20) Lin-15b(Hd126). *Proc Natl Acad Sci* 104: 834–839.
- Schwarz EM, Kato M, Sternberg PW. 2012. Functional transcriptomics of a migrating cell in *Caenorhabditis elegans*. *Proc Natl Acad Sci* 109: 16246–16251. doi:10.1073/pnas.1203045109.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soeding J, Thompson JD, Higgins DG. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7:539–539.
- Smit AFA. RepeatMasker Open-4.0. 2015. [www.repeatmasker.org](http://www.repeatmasker.org).
- Smith SA, Dunn CW. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 5:715–716.

- Srinivasan J, Dillman AR, Macchietto MG, et al. 2013. The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics* 193:1279–1295.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22:2688–2690.
- Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19:215–ii225.
- Stricker SA. 1999. Comparative biology of calcium signaling during fertilization and egg activation in animals. *Dev Biol* 211:157–176.
- Streit A, Li W, Robertson B, Schein J, Kamal IH, Marra M, Wood WB. 1999. Homologs of the *Caenorhabditis elegans* masculinizing gene *her-1* in *C. briggsae* and the filarial parasite *Brugia malayi*. *Genetics* 152:1573–1584.
- Tange O. 2011. Gnu parallel—the command-line power tool. *The USENIX Magazine*
- Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* 278:631–637.
- Thomas GWC, Ather SH, Hahn MW. 2016. Gene-tree reconciliation with MUL-trees to resolve polyploidy events. *Systematic Biol.* 66:1007–1018.
- Thorne MAS, Kagoshima H, Clark MS, Marshall CJ, Wharton DA. 2014. Molecular analysis of the cold tolerant Antarctic nematode, *Panagrolaimus davidi*. *PLoS ONE* 9:e104526.
- Thorne MAS, Seybold A, Marshall C, Wharton D. 2017. Molecular snapshot of an intracellular freezing event in an Antarctic nematode. *Cryobiology* 75:117–124.
- True JR, Haag ES. 2001. Developmental system drift and flexibility in evolutionary trajectories. *Evolution & Development* 3:109–119.
- Tsai IJ, Otto TD, Berriman M. 2010. Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. *Genome Biol* 11:R41.
- Vancoppenolle B, Borgonie G, Coomans A. 1999. Generation times of some free-living nematodes cultured at three temperatures. *Nematology* 1:15–18.
- Wang Y, Tang H, Debarry JD, Tan X, Jingping L, Wang X, Tae-ho L, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* 40:e49–e49.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.
- Wharton, DA, Marshall CJ, Egeter B. 2017. Comparisons between two Antarctic nematodes: cultured *Panagrolaimus* sp. DAW1 and field-sourced *Panagrolaimus davidi*. *Nematology*:19, 533–542.
- Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 26:873–881.
- Xu XZS, Sternberg PW. 2003. A *C. elegans* sperm TRP protein required for sperm-egg interactions during fertilization. *Cell* 114:285–297.
- Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* 18:821–829.

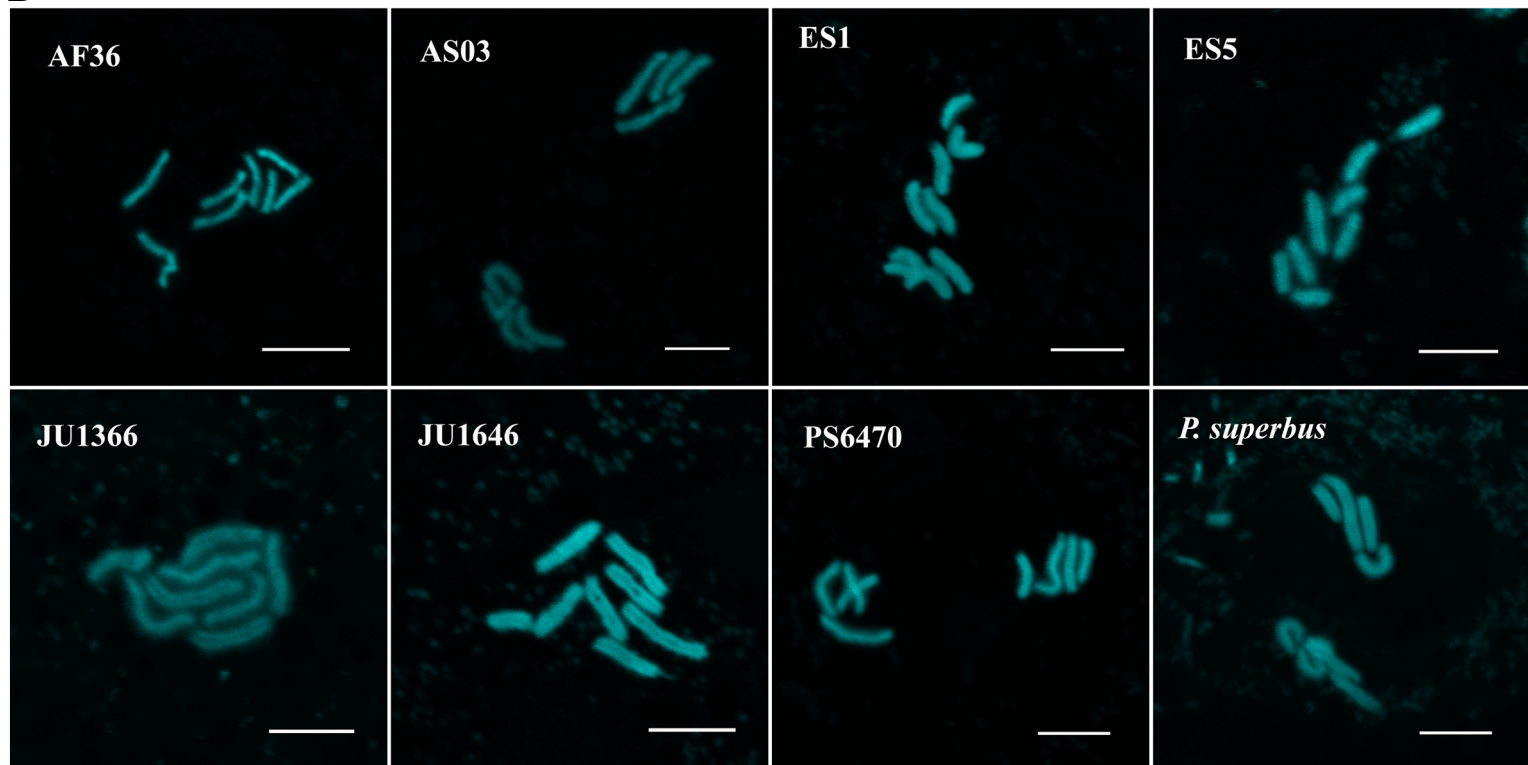


# Supplementary Figure 1

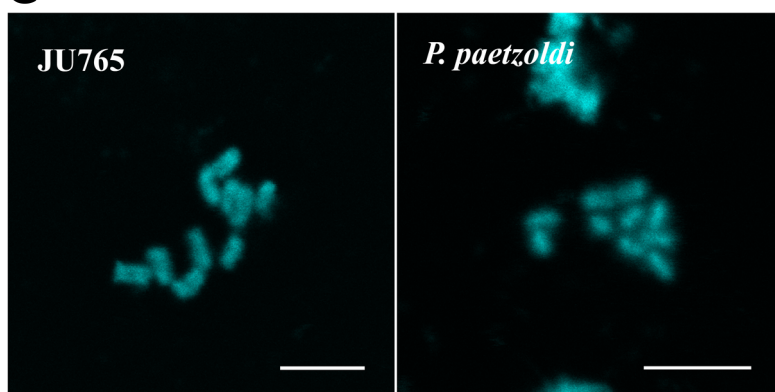
A



B

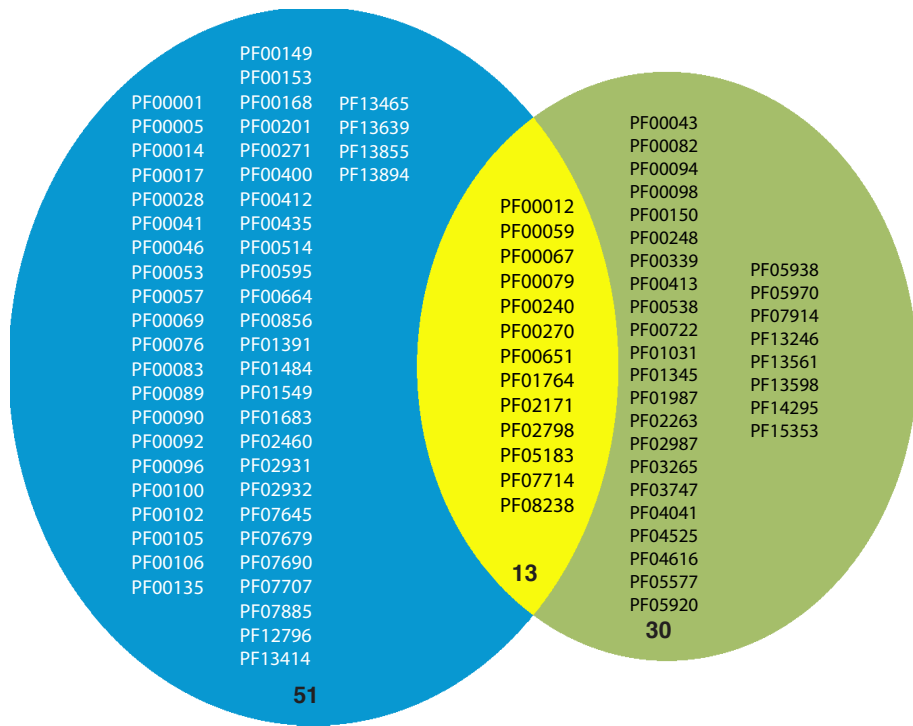


C



scale bar = 5  $\mu$ m

# Supplementary Figure 2



*Panagrolaimus vs. Propanagrolaimus & Panagrellus*

*Panagrolaimidae vs. Outgroup*

# Supplementary Figure 3

clade V  
clade IV  
clade III

○ *C. elegans*

□ *P. pacificus*

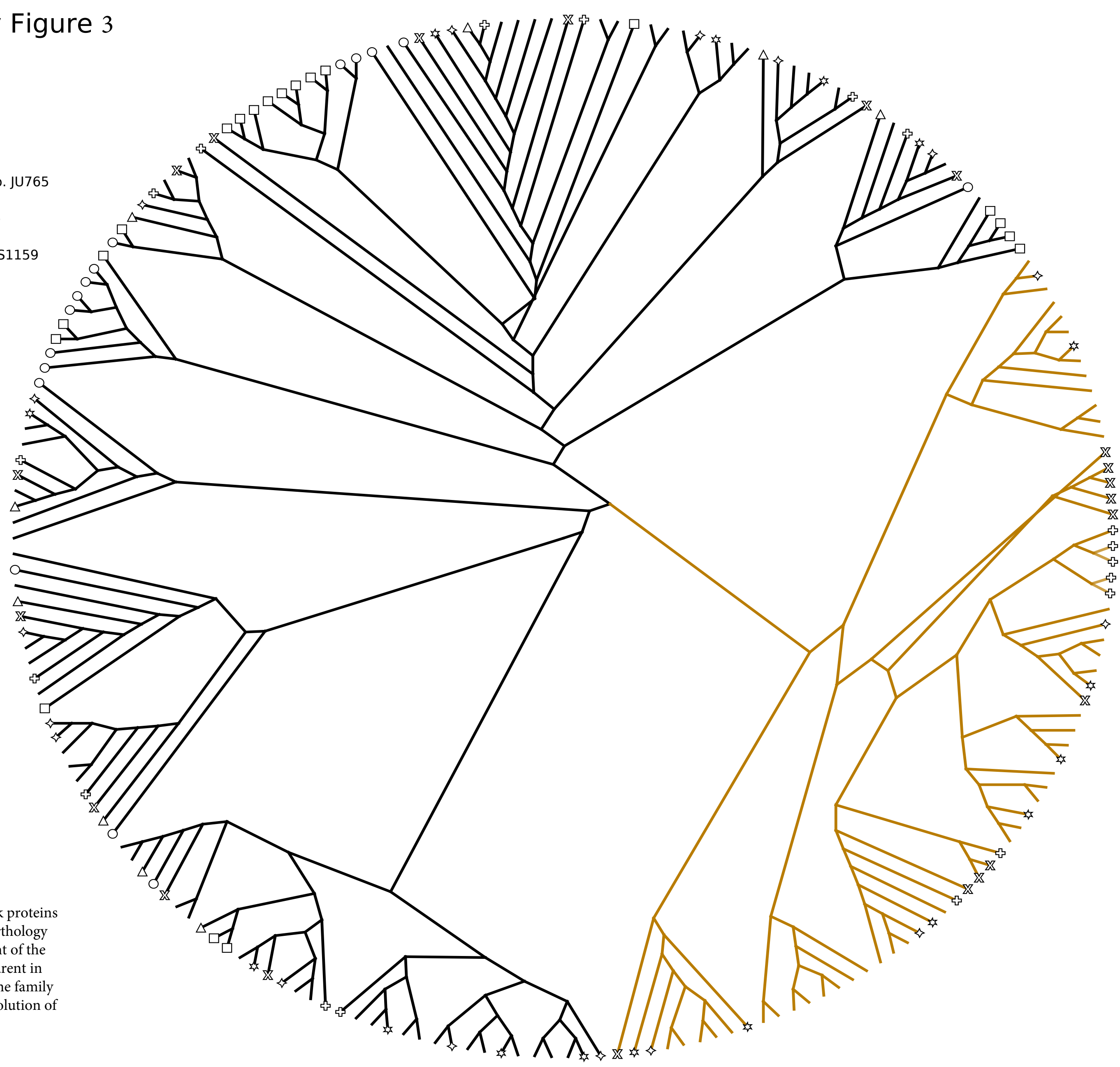
⊕ *P. redivivus*

⌘ *Propanagrolaimus* sp. JU765

◇ *Panagrolaimus* sp. ES5

☆ *Panagrolaimus* sp. PS1159

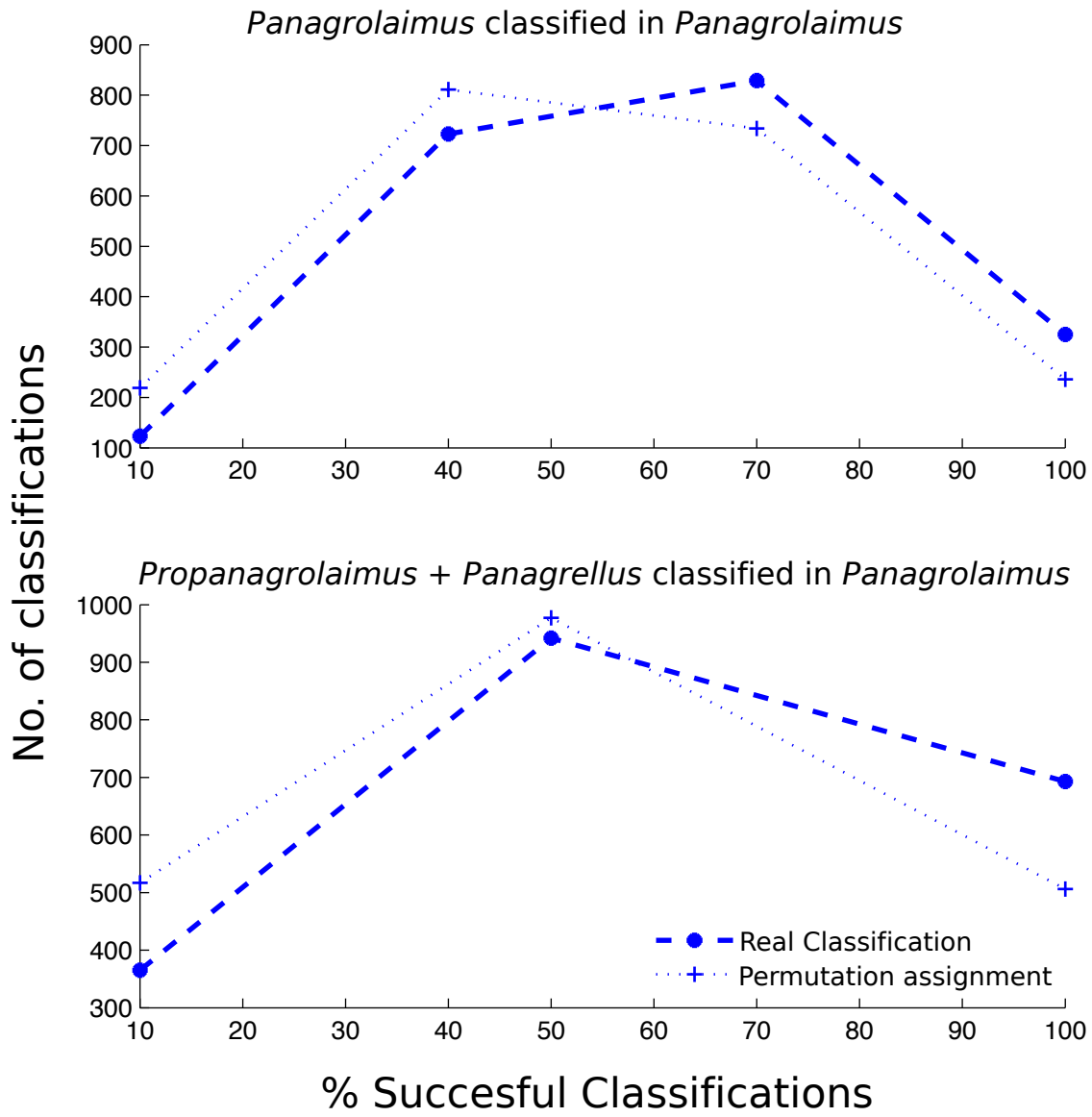
△ *A. suum*



Supplementary Figure 3:  
RAxML tree of small heat-shock proteins  
in the species analysed in our orthology  
screen. An inflation independent of the  
known one in *C. elegans* is apparent in  
clade iv species. The inflated gene family  
might be pre-adaptive to the evolution of  
cryptobiosis in *Panagrolaimus*.

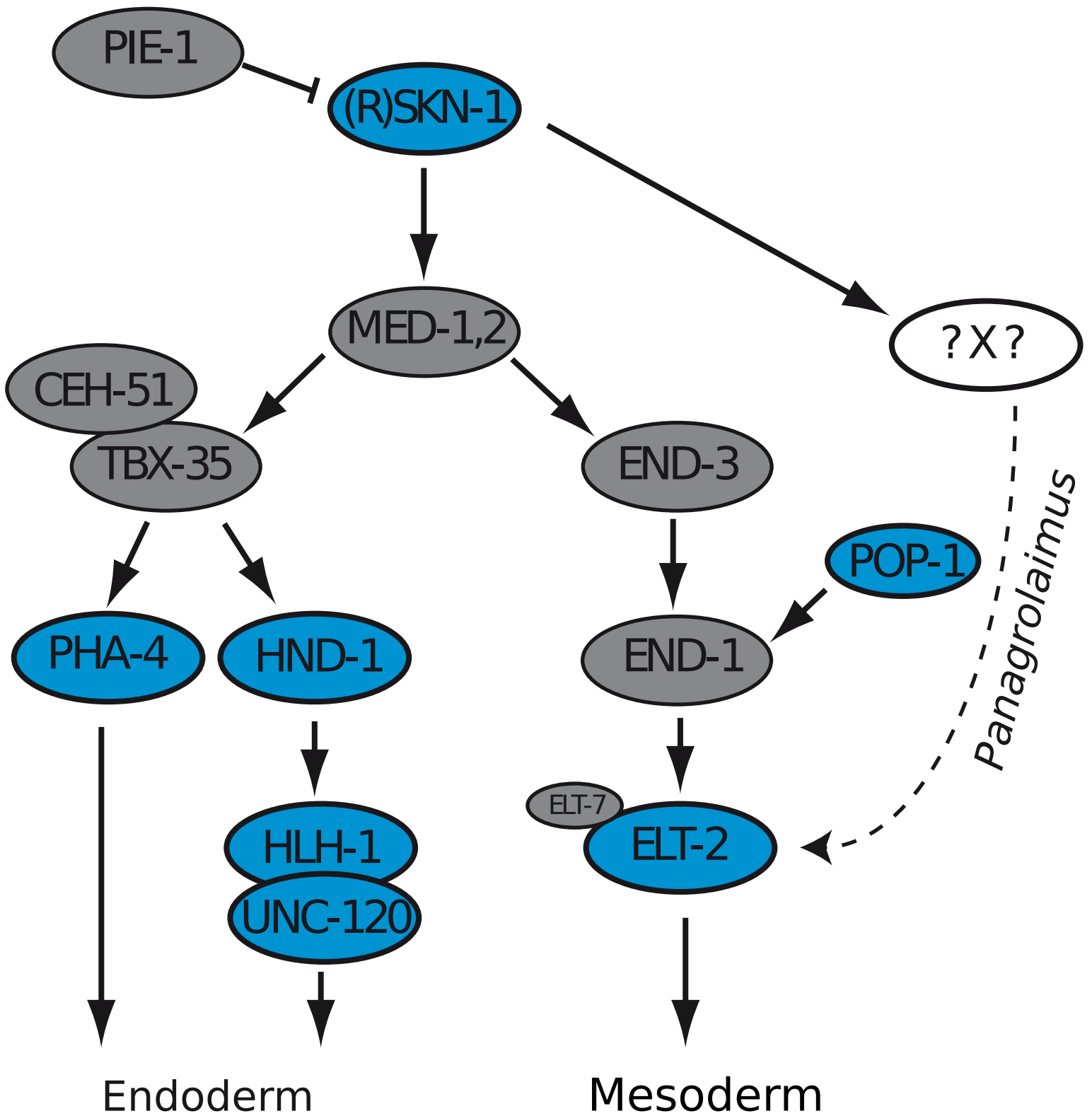


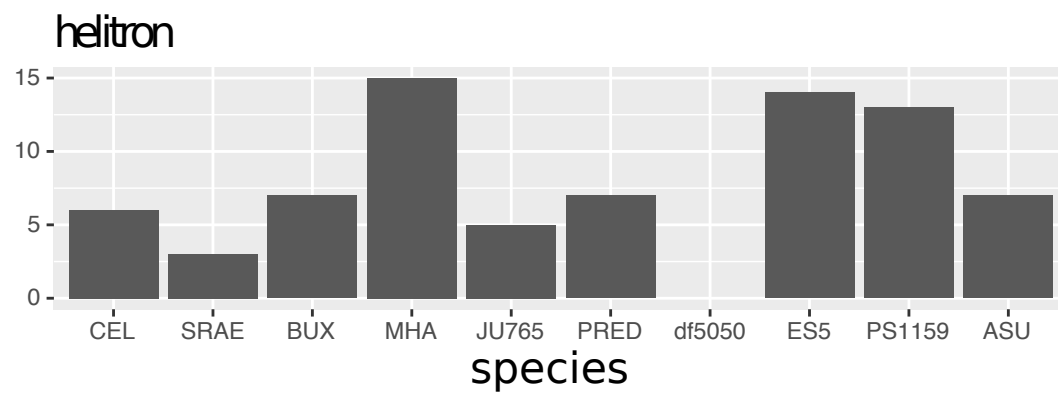
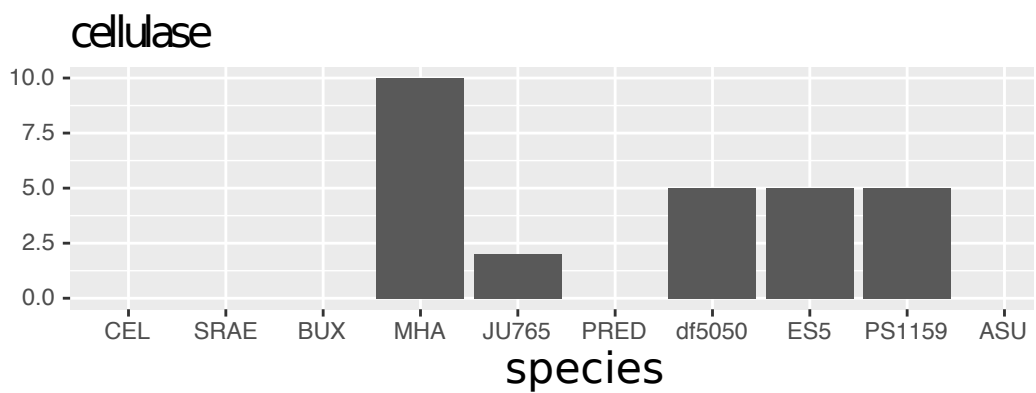
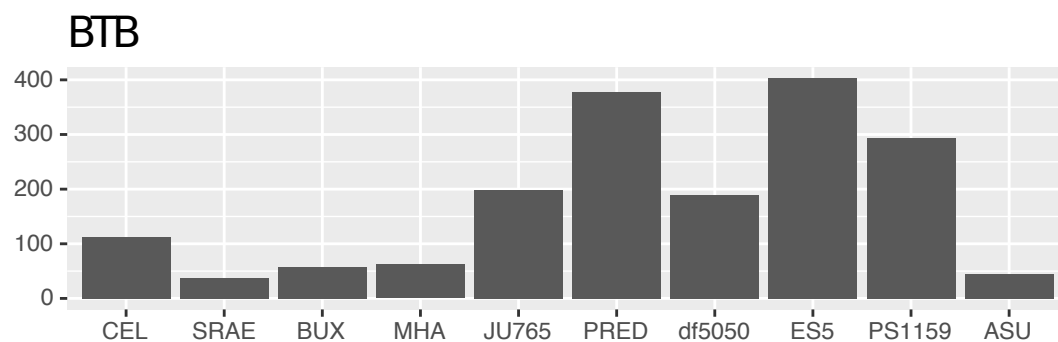
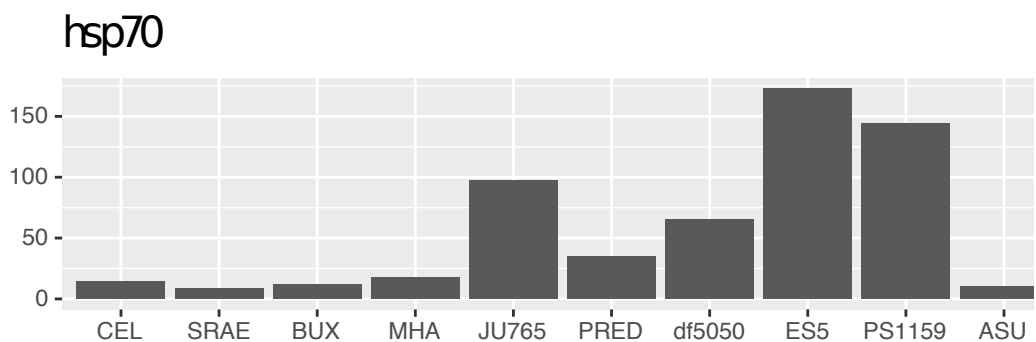
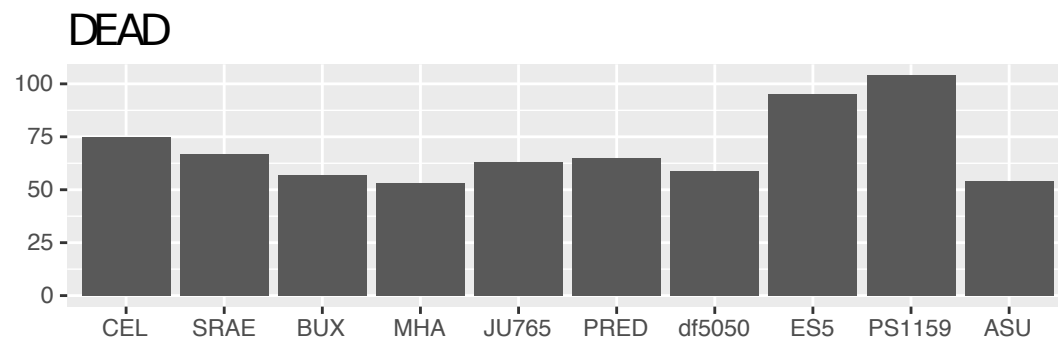
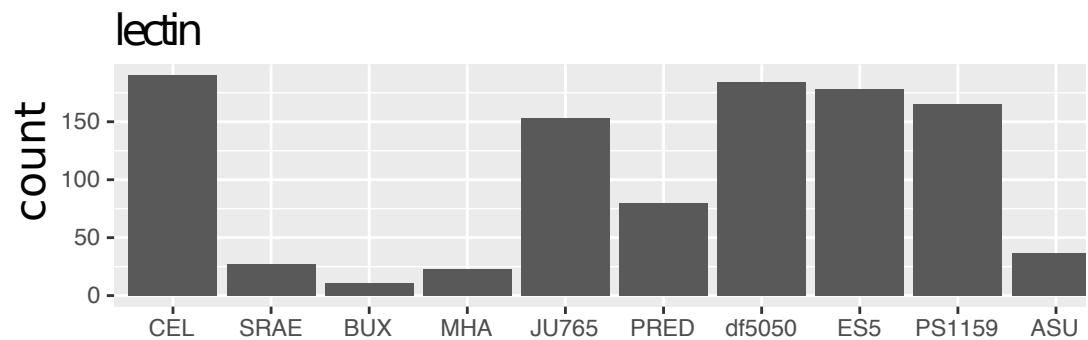
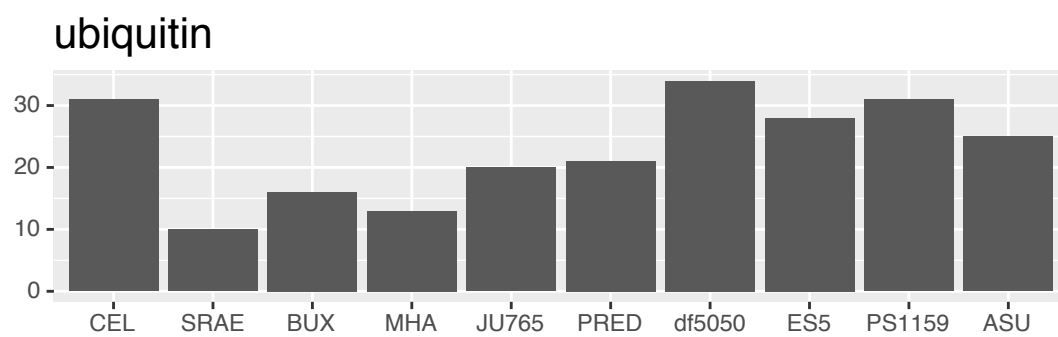
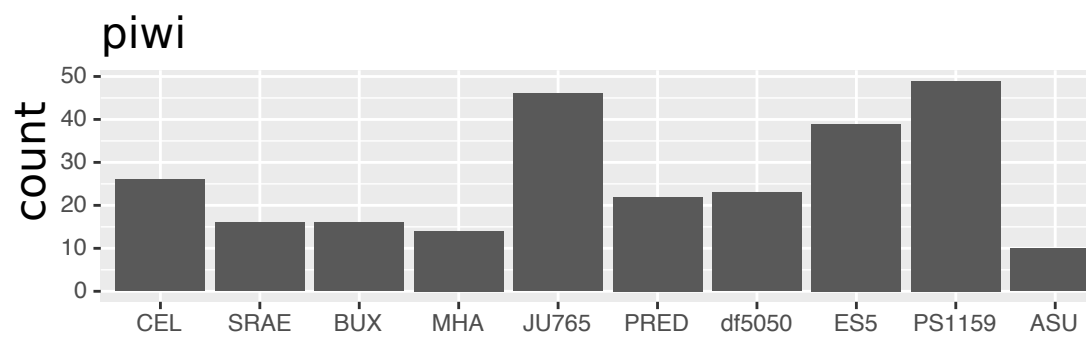
# Supplementary Figure 5





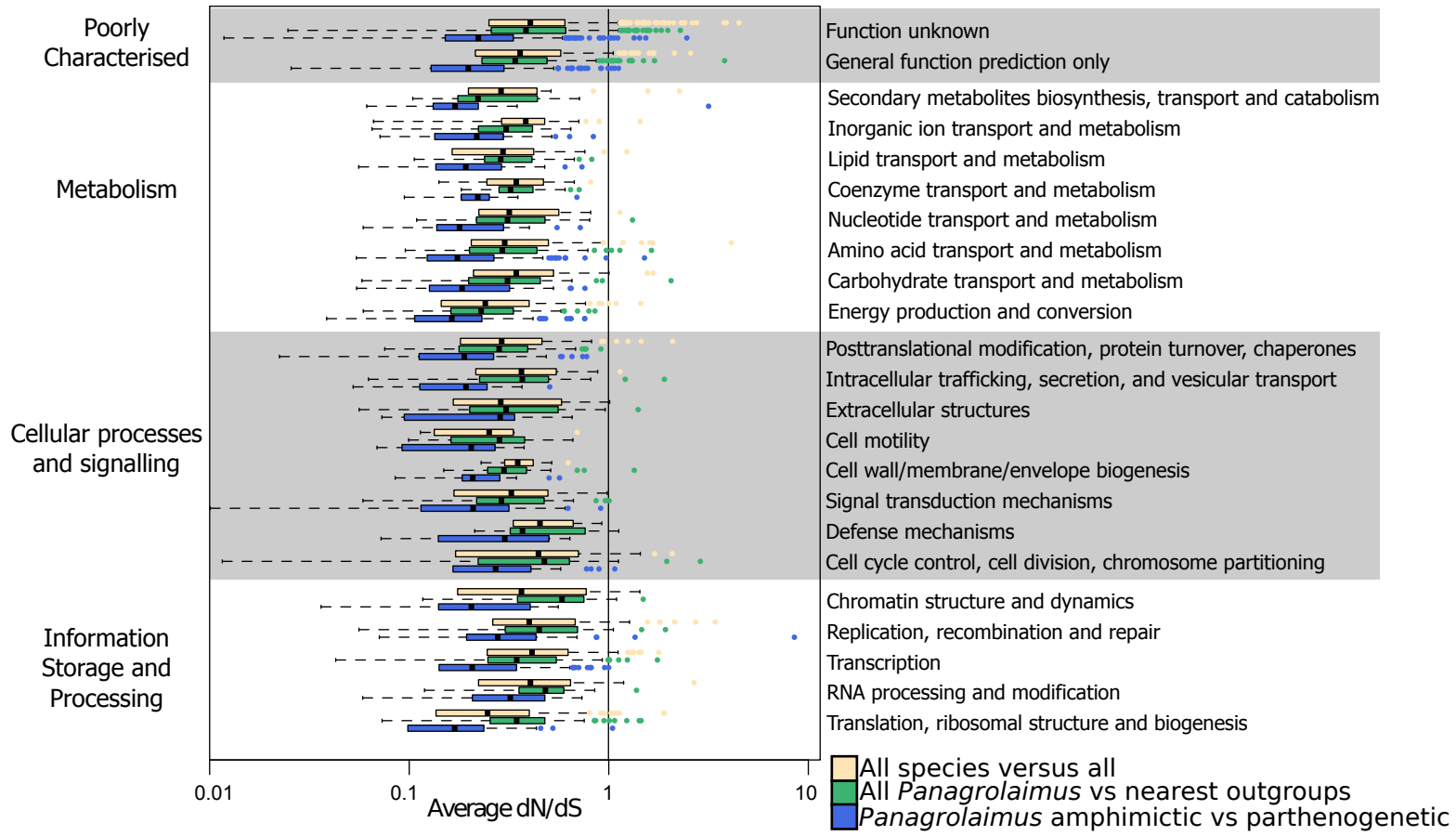
Supplementary Figure 7



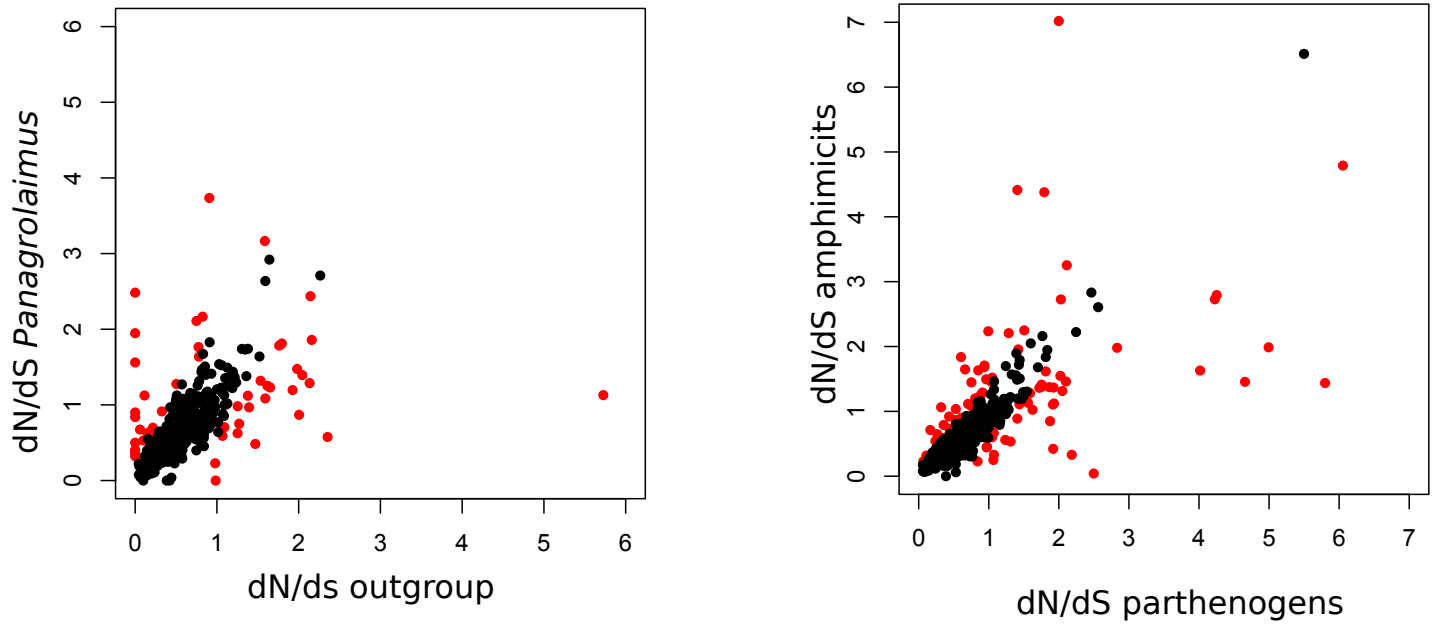




# Supplementary Figure 9



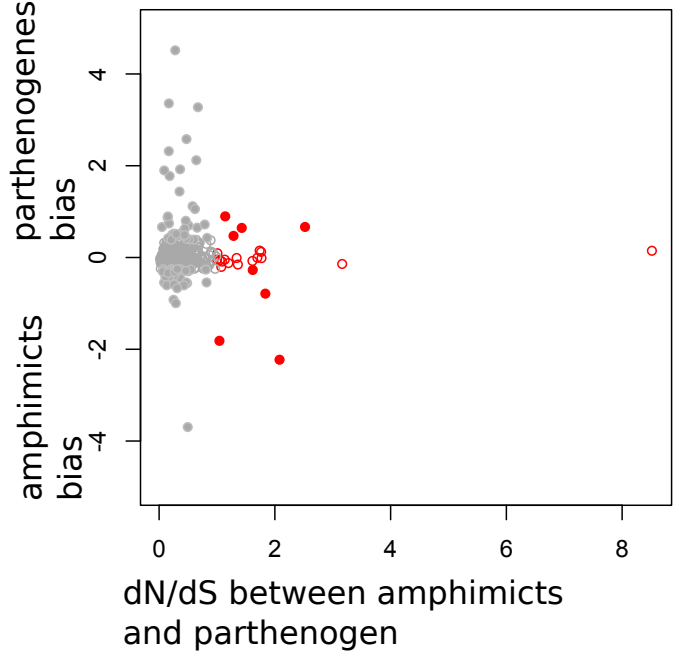
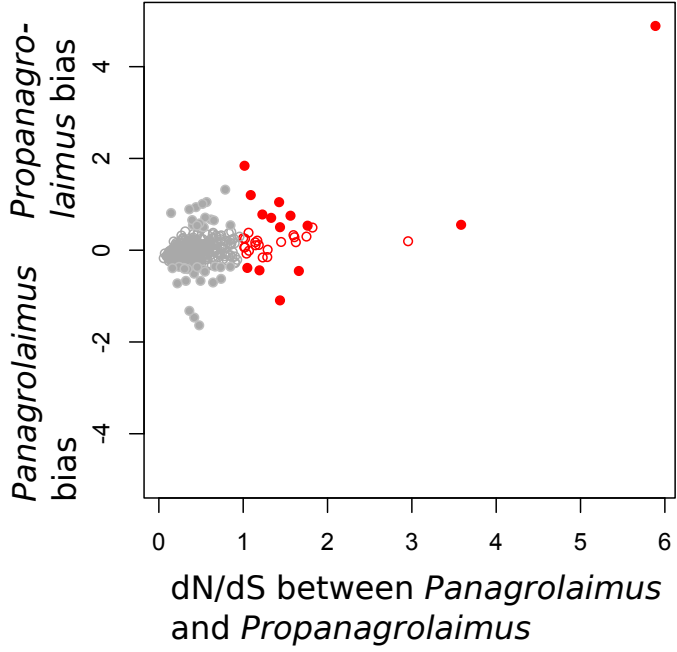
# Supplementary Figure 10A



Supplementary Figure 10:

A. Plots of dN/dS values showing outlier genes (in red) for comparisons between cryptobiotic *Panagrolaimus* species and non-cryptobiotic outgroups and between parthenogenetic and amphictic *Panagrolaimus* species.

# Supplementary Figure 10B



Supplementary Figure 10:

B. Plots displaying genes with bias towards a given taxon in both comparisons displayed in A.

## Supplementary Figure Captions

### Supplementary Figure 1:

Karyograms of parthenogenetic (A) and amphimictic (B) *Panagrolaimus* species, and outgroup species (C). All analysed amphimictic *Panagrolaimus* were found to have  $2n=8$ , while the parthenogenetic had ( $xn=$ ) 12 chromosomes. The *Propanagrolaimus* outgroup species had  $2n=10$  chromosomes, which is in line with the number in *Panagrellus redivivus*. Scale bar represents 5  $\mu\text{m}$ . We observed the parthenogenetic *Panagrolaimus* species to have one polar body at the end of oocyte development, confirming meiosis.

### Supplementary Figure 2:

Using a Fisher's exact test (green) and a Support Vector Machine classification (blue) we compared Pfam annotations. Firstly (left), to find genes potentially important in *Panagrolaimus* specific traits (e.g. desiccation tolerance) we tested species in the genus *Pangrolaimus* against (non-desiccating) *Propanagrolaimus* and *Panagrellus* (combined). Secondly (right), to search for panagrolaimidae traits we combined *Pangrolaimus* with *Propanagrolaimus* and *Panagrellus* and tested against outgroup species. Overall, we found several protein families associated with cryptobiosis, e.g HSPs, helicases, and C-type lectins, to be important for *Panagrolaimus* biology. See supplementary Excel table for Pfam description.

### Supplementary Figure 3:

RAxML tree of small heat-shock proteins in the species analysed in our orthology screen. An inflation independent of the known one in *C. elegans* is apparent in clade IV species. The inflated gene family might be pre-adaptive to the evolution of cryptobiosis in *Panagrolaimus*.

### Supplementary Figure 4:

Maps of the complete mitochondrial genomes of *Panagrolaimus superbus*, *Panagrolaimus* sp. ES5, *Panagrolaimus* sp. PS1159 and *Propanagrolaimus* sp. JU765. The *P. sp.* DAW1 mt genome could not be fully assembled, and PS1579 (not shown) remained in 4 fragments. The non-coding and unresolved portions of the genome are represented by continuous grey and shaded regions, respectively. All genes are encoded in the same direction (as indicated by arrow heads). Proteins and rRNA genes have standard nomenclature, tRNA genes are designated by single-letter abbreviations. Two tRNA genes are present for leucine (L1 and L2) and serine (S1 and S2). Note the different placement of the *cox* genes between PS1159 and *P. sp.* DAW1, and the other species, and the *nad5* gene between *P. superbus* and *P. sp.* ES5.

### Supplementary Figure 5:

Histograms on the percent of correct classifications across all iterations of the classifier and classifier control permutation.

- A. In this comparison of *Panagrolaimus* species each classification attempt can assign the correct label to either none, one, two, or three worms (samples). Comparisons show that the classifier trained on real sample-class mappings outperforms the control permutation significantly (trained on random assignment of samples to class labels), with a larger number of above-chance classifications (i.e., correct categorizations for 2, or 3 out of three worms).
- B. In the comparison of outgroup species and *Panagrolaimus* classification at chance level corresponds to correctly classifying one out of two worms. Accordingly, this is where the control permutation analysis peaks, whereas the classifier trained on correct class-to-sample mapping significantly outperforms the control, classifying both worms correctly in a larger number of iterations.

**Supplementary Figure 6:**

Plots showing the distribution of codon usage in *Panagrolaimus* sp. PS1159 genes identified as HGT candidates (orange) in comparison to genes identified as stemming from contamination (yellow) with aid of our AlienIndex analysis, and the genomic background (green). HGT candidates appear to be more similar in their codon usage to the genomic background, than the ones identified as contamination. Similar patterns have been observed in the other species and are indicative of an integration into the genomic background of the laterally acquired genes.

**Supplementary Figure 7:**

Many orthologues of *C. elegans* developmental GRNs could not be detected in either *Panagrolaimus* nor other clade IV (nor clade III) species in our analysis, indicating non-orthologues gene displacement and high evolutionary turn-over. One example displayed here, is the *C. elegans* endo-mesoderm specification GRN (missing genes in grey). It is possible that the induction of mesoderm formation was under direct influence of SKN-1, and particular WNT-pathway control, in species ancestral to clades IV and V. See Supplementary Results and Discussion for details.

**Supplementary Figure 8:**

Key gene families that were found to be inflated in *Panagrolaimus* species. Numbers are based on a screen of PFAM domains obtained by InterProScan followed by enrichment analysis (Fisher's test) comparing with *Propanagrolaimus* and *Panagrellus* as outgroups. Among others the displayed families have a potential role in the evolution of cryptobiosis in the *Panagrolaimus* species. Species abbreviations are: CEL – *C. elegans* (clade V), SRAE – *S. ratti*, BUX – *B. xylophilus*, MHA – *M. hapla*, JU765 – *Propanagrolaimus* sp. JU765, PRED – *P. redivivus*, DF5050 – *P. superbus*, ES5 – *P. sp.* ES5, PS1159 – *P. sp.* PS1159 (all preceding clade IV), ASU – *A. suum* (clade III).

**Supplementary Figure 9:**

Average rates of evolution (dN/dS) are shown for comparisons between parthenogenic and gonochoristic *Panagrolaimus* (in blue), *Panagrolaimus* and the closest outgroups

(*Propanagrolaimus* and *Panagrellus*) (in green), and between all species (in yellow) grouped by functional categories (see Supplementary Excel file dNdS). We find that while the majority of genes have a ratio of  $< 1$  there are notable exceptions across several functional categories, both between parthenogenetic and gonochoristic species and between panagrolaimids and their closest relatives. Genes in these categories could play roles in the evolution of cryptobiosis and potentially also in the evolution of parthenogenesis.

**Supplementary Figure 10:**

- A. Plots of dN/dS values showing outlier genes (in red) for comparisons between cryptobiotic *Panagrolaimus* species and non-cryptobiotic outgroups and between parthenogenetic and amphictic *Panagrolaimus* species.
- B. Plots displaying genes with bias towards a given taxon in both comparisons displayed in A.

**Supplementary Table 2:** Pairwise divergence time estimates (in Myr) in *Panagrolaimus* spp. assuming a range of 4 to 50 days per generation. For calibrating divergence time we used the species pair *C. briggsae* and *C. sp5* a per base substitution rate of 0.1322 (calculated with Andi). Furthermore, we assumed a divergence time of 171M generations between these two species.

Note: divergence times from the common ancestor (Figure 2 B) are obtained by multiplying the numbers in this Table by the factor 0.5.

				4 days	8 days	16 days	32 days	50 days
Species A	repro	Species B	repro	Myr (4d/g = 91.2 gen/yr)	Myr (8 days/gen = 45.6 gen/yr)	Myr (16d/g = 22.8 gen/yr)	Myr (32d/g = 11.4 gen/yr)	Myr (50 days/gen = 7.3 gen/yr)
<i>P. superbus</i>	A	ES5	A	1.01	2.03	4.06	8.12	12.69
PS1159	P	ES5	A	1.32	2.65	5.29	10.58	16.53
PS1159	P	<i>P. superbus</i>	A	1.37	2.74	5.48	10.95	17.12
ES5	A	<i>P. redivivus</i>	A	1.34	2.68	5.35	10.71	16.73
PS1159	P	<i>P. redivivus</i>	A	1.49	2.98	5.95	11.91	18.61
<i>P. superbus</i>	A	<i>P. redivivus</i>	A	1.86	3.71	7.43	14.86	23.21
ES5	A	JU765	H	2.40	4.79	9.59	19.18	29.96
PS1159	P	JU765	H	2.42	4.83	9.66	19.32	30.19
<i>P. superbus</i>	A	JU765	H	2.52	5.05	10.10	20.20	31.56
<i>P. redivivus</i>	A	JU765	H	3.26	6.51	13.02	26.05	40.70