Phylogenomic Insights into the Evolution and Origin of Nematoda

Xue Qing[1 #], Y. Miles Zhang[2#], Sidi Sun[1], Mohammed Ahmed[3], Wen-Sui Lo[4], Wim Bert[5], Oleksandr Holovachov[6], Hongmei Li[1*]

[1] Department of Plant Pathology, Nanjing Agricultural University, Nanjing 210095, China

[2] Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK

[3] School of Life Sciences, Liverpool University, Liverpool, UK

[4] Institute of Future Agriculture, Northwest A&F University, Yangling 712100, China

[5] Department of Biology, Ghent University, Ghent B-9000, Belgium

[6] Department of Zoology, Swedish Museum of Natural History, Stockholm SE-104 05, Sweden

[#] These authors contributed equally to this work

* Corresponding author, e-mail: lihm@njau.edu.cn

Running title: Phylogenomic of Nematoda

Abstract:

The phylum Nematoda represents one of the most cosmopolitan and abundant metazoans on Earth. In this study, we reconstructed the first phylogenomic tree for phylum Nematoda. A total of 54 genomes belongs to eight nematode orders were newly sequenced, providing the first representatives for orders Dorylaimida, Mononchida, Monhysterida, Chromadorida, Triplonchida, and Enoplida. The obtained phylogeny is well-resolved in most of clades, and topologies are similar across reconstruction parameters. Few conflicts were found within Rhabditida when using different metrics or between amino acid and nucleotides. The subclass Enoplia is placed as sister to the rest of Nematoda, thus supporting the hypothesis of a likely marine origin for the nematode ancestor. The Triplonchida is monophyletic but not well-supported, and Enoplida is paraphyletic. The taxa equipped with the stomatostylet form a monophyletic group, but the superfamily Aphelenchoidea does not form a monophyletic clade. Genera *Trichinella* and *Trichuris* shared a common ancestor ~170 mya, considerably later period compared to previous studies. All stomatostylet-bearing nematodes is suggested to have

existed ~250 mya, coinciding with the transition from the Permian to the Triassic period. The genus *Thornia* is placed outside of Dorylaimina and Nygolaimina, disagrees with all previous studies. Additionally, we tested the performance of whole genome amplification, and demonstrated it is a promising strategy in phylogenomic of microscopy eukaryotes. The presented study considerably expanded current nematode genome dataset and well-resolved phylogeny improved our understanding for the origin and evolution of Nematoda.

Keywords:

Nematodes, worm, phylogeny, genome sequencing, whole genome amplification, diversity

Introduction

The phylum Nematoda represents one of the most cosmopolitan, abundant, and diverse metazoans on Earth, with an estimated number of species extending to 10 million (Boucher and Lambshead, 1995; Coomans, 2002) (Figure 1). Free-living nematodes occupy all trophic levels in the food web and play a pivotal role in carbon cycling and energy flows (Ferris, 2010; Ingham et al., 1985; Thomas et al., 2022; van den Hoogen et al., 2019), and the genera *Caenorhabditis* and *Pristionchus* are two popular nematode models in molecular and evolutionary biology. Nematode parasites are a threat to not only the health of plants (e.g. *Meloidogyne* spp., *Heterodera* spp.), but also that of animals and humans (*e.g.* ascariasis, trichuriasis, and hookworm disease, filariasis). Resolving the nematode phylogeny is an important step in understanding the origin of parasitism and transition of lifestyles, from basic evolutionary biology to pathogen control and development of anthelmintic drugs (International Helminth Genomes Consortium, 2019). However, the high phenotypic plasticity (Coomans, 2002; Nadler, 2002), a near-complete of synapomorphic diagnostic characters, and the absence of an informative fossil record have prevented scientists from deriving a consistent evolutionary framework (Lorenzen, 1981; Coomans, 2000). The first molecular phylogeny was reconstructed based on 18S rRNA of 53 species and led to the recognition of three major lineages of nematodes: Dorylaimia, Enoplia, and Chromadoria (Blaxter et al.,1998; De Ley et al., 2002). However, despite its popularity as a molecular marker, 18S rRNA lacks the resolving power for deep phylogenies, i.e. the root of the nematode tree (Blaxter and Koutsovoulos, 2015). As a result, currently all published rRNA phylogenies lack supporting evidence for the

“backbone” of the tree, and many lineages remain poorly resolved (Ahmed and Holovachov, 2021).

By contrast, phylogenomic analyses have yielded statistically-robust and congruent results that have proven useful in resolving previously controversial phylogenetic relationships across evolutionary lineages (Misof et al., 2014; Jarvis et al., 2014; Telford et al., 2015; Williams et al., 2020). However, difficulties in obtaining high-quality DNA have posed a major challenge for omics studies on uncultured microscopic organisms. In the case of nematodes, pioneering work has been limited to studies on vertebrate parasitic taxa and model *Caenorhabditis* spp. (Blaxter & Koutsovoulos, 2015; International Helminth Genomes Consortium, 2019; Stevens et al., 2019), with little or no representation for species less than 1mm in length as well as for uncultured species. Recent advances in whole-genome amplification (WGA) have enabled single cells to generate sufficient DNA for sequencing, and these techniques have now been applied to various microscopic eukaryotic (Lepere et al.2011; Weisz et al., 2019; Sahraei et al., 2022; Lee et al., 2023), making it possible to sequence nematodes from a single individual at a phylum scale. The current paper focuses on sequencing the genomes of 55 nematode species covering eight orders using WGA, aiming to provide finer resolution for phylogeny and improved knowledge about patterns and processes of nematode evolution.

Figure 1. Diversity of free-living and parasitic species in Nematoda. 1-27: the head region, 28-35: body habitus. 1, 40:*Acrobeloides*; 2: *Acrobeles*; 3: *Bicirronema*, 4: *Teratocephalus*; 5: *Tricirronema*; 6: *Pristionchus*; 7: *Steinernema*; 8: *Prismatolaimus*, 9: Aphanolaimus, 10: *Alaimus*; 11: *Aphelenchoides*; 12: *Diphtherophora*; 13: *Tripyla*; 14: *Miconchus*; 15: *Mylonchulus*; 16: *Ironus*; 17: *Aporcella*; 18, 38: *Helicotylenchus*; 19, 34: *Meloidogyne*; 20: *Trichodorus*; 21: *Sphaerolaimus*; 22, 35: *Greeffiella*; 23: *Eudorylaimus*; 24: *Oesophagostomum*; 25: *Ancylodtoma*; 26: *Strongylus*; 27:

Bunostomum; 28: *Croconema*; 29: *Tylenchulus*; 30: Draconematidae; 31, 37: *Desmoscolex*; 32, 36: *Tricoma*; 33: *Heterodera*; 39: *Etamphidelus*; 41: *Malenchus*. 7: insect parasite; 18-20, 29, 33, 34, 38: plant parasites; 24-27: vertebrate parasites; 1-6, 8-17, 23, 39-41: terrestrial free-living species; 21, 22, 28, 30-32, 35-37 marine free-living species.

Material & Methods

DNA extraction, sequencing, and genome assembly

Nematodes were isolated from soil or marine sediments and extracted using a Baermann funnel. One to five adult fresh worms were mounted on temporary slide and morphologically identified under the Olympus BX51 microscope (Olympus Optical, Japan). The slide was then open and the worms were cut into pieces and transferred with a needle into a 200-µl PCR tube. Genomic DNA extraction and amplification were performed with the REPLI-g Single Cell Kit (Qiagen, Germany). In addition to the whole genome amplification (WGA) method, DNA was directly extracted from 20,000 fresh individuals for three species (*Aphelenchoides blastopthorus*, *Aphelenchoides smolae* and *Ditylenchus* sp.) using the Ezup Column Animal Genomic DNA Purification Kit (Sangon Biotech, China), in order to evaluate the performance of WGA. The TruSeq DNA Sample Prep Kit (Illumina Inc. CA, USA) was used for library preparation according to the manufacturer's instructions. Sequencing was preformed using the Illumina NovaSeq platform. Quality control was performed in FASTP (Chen et al., 2018), reads less than 50 bp, Q-score less than 20, or reads with N base number > 3 were removed. Filtered reads were assembled using SPAdes v3.15.5 with the default parameters (Bankevich, A., Nurk, S., et al. 2012). To avoid potential noise, we removed fragments with total length < 3k and coverage < 15. The remaining sequences were subjected to a microbial contamination prediction analysis using Kraken2 (Wood et al., 2019) to identify and eliminate any fragments stemming from microbial contaminants.

Phylogenomic analysis

The phylogeny hypothesis was reconstructed using newly generated data supplemented with additional genomes downloaded from GenBank, and a total of 133 Nematoda and six outgroups (two Tardigrada, two Arthropoda, one Onychophora, and one Priapulida) were included. The complete and single copy ortholog genes were extracted using BUSCO (Simão, F.A., Waterhouse, R.M., et al. 2015) using the Nematode ortholog database OrthoDB v10, and translated into amino acids (AA). We selected 50% and 70% complete nucleotide and AA matrices for downstream analyses to balance between the number of loci and data completeness, which included 1505 and 304 loci, respectively. The matrices were aligned using the L-INS-I

option in MAFFT v7.490 (Katoh, K. and Toh, H. 2008), trimmed using TrimAl (Capella-Gutiérrez, S., Silla-Martínez, J.M., et al. 2009), and concatenated using FASconCAT-G v1.05.1 (Kück, P. and Longo, G.C. 2014). We further removed potentially misaligned regions using SpruceUp (Borowiec, M.L. 2019) at 97% lognormal criterion, and performed symmetry tests (Naser-Khdour, S., Minh, B.Q., et al. 2019) to remove partitions that failed at symmetry tests for stationarity and homogeneity assumptions using IQ-TREE v.2.2.0 (Minh, B.Q., Schmidt, H.A., et al. 2020).

The following analyses were performed on both matrices using IQ-TREE: 1) Partitioned by locus with best model selected using ModelFinder (Kalyaanamoorthy, S., Minh, B.Q., et al. 2017), and with ultrafast bootstrap (UF2, Hoang, D.T., Chernomor, O., et al. 2017) and SH-like approximate likelihood ratio test (SH-aLRT, Guindon, S., Dufayard, J.-F., et al. 2010) as support values; 2) Individual gene trees with best model selected using ModelFinder; 3) Gene and site concordance factors (gCF, sCF, respectively) using steps 1 and 2 following (Minh, B.Q., Hahn, M.W., et al. 2020); 4) GHOST model to account for heterotachy (Crotty, S.M., Minh, B.Q., et al. 2019); and finally, 5) the 70% AA matrix was also analysed using the posterior mean site frequency model (PMSF, class C20, C40, and C60) which has been designed to correct for long-branch-attraction by modeling site heterogeneity (Wang, H.C., Minh, B.Q., et al. 2018), this was not done for the 50% matrix due to the computational limits required for such a large dataset.

We also analysed both datasets using the coalescent-based method implemented in ASTRAL-III (Zhang, C., Rabiee, M., et al. 2018), using the gene trees generated in step 2 above as input. We collapsed the nodes with <10% support using Newick Utilities (Junier, T. and Zdobnov, E.M. 2010), and support was evaluated using local posterior probability (LPP, Sayyari, E. and Mirarab, S. 2016).

Molecular dating and ancestral state reconstruction

Divergence dating was performed using MCMCTree using PAML v4.9j (Yang, Z. 2007), using approximate likelihood calculation and ML estimation of branch lengths to reduce the computational burden. Hessian matrices were calculated using the LG substitution model and the independent rates clock mode. Ten nodes (the root and nine internodes) were selected for calibration based on existing fossil records (Supplementary table S1). Two independent runs were performed with 5 million generation burnins and 50 million generations using the HKY model, and convergence was assessed using Tracer v1.7 (Rambaut, A., Drummond, A.J., et al.

2018). The tree was visualized using MCMCtreeR (Puttick, M.N. 2019) in R v4.2.0 (R Core Team 2022).

Results

Whole genome amplification and genome assembly

A total of 54 species from terrestrial and marine environments were newly sequenced. The raw assemblies ranged in size from 43.67 Mb to 839.99 Mb, comprising contigs spanning from 15.99 Kb to 2.69 Mb. Genome sizes decreased after removing potential contaminations, with lengths ranging from 2.90 MB to 189.63 Mb, and contig counts varying from 484 to 33537 (Supplementary table S2). The BUSCO analysis yielded completeness percentages ranging from 7.65% to 81.86% for 55 assembled genomes. We observed smaller size and more fragmented genomes (higher contig numbers) when using the WGA method compared to direct DNA extraction without amplification from fresh nematodes (Supplementary Figures 1, 2). Further BUSCO evaluation indicated varying performance results, *i.e.* both methods generated a similar genome completeness in *Ditylenchus* sp., a more complete *A. blastopthorus* and *A. smolae* genome with direct extraction and WGA, respectively (Supplementary Figure 3). The assembled contigs were annotated and plotted according to their GC content and coverage (Supplementary Figure 4), revealing that direct extraction introduced more bacterial contaminations than the WGA approach.

Evolution of Nematoda

From the assembled genomes, we extracted 1486 loci with 496,704 bp for the 50% complete matrix, and 303 loci with 76,524 bp for the 70% complete matrix. The reconstructed phylogenies have similar topologies in most cases regardless the metrics, datasets, and reconstruction methods (Figure 2). The major difference was found in placements of the Rhabditomorpha+Diplogasteromorpha and Spiruromorpha clades between the nucleotide and AA in both 50p and 70p comparisons. Additionally, clades of *Panagrellus*/*Halicephalobus*/*Propanagrolaimus* and *Rotylenchulus*/*Hoplolaimus*/*Helicotylenchus* were also slightly changed among phylogenies (Supplementary Figures 5-9). The subclass Enoplia primarily consists of taxa inhabiting marine environments. Placing the class Enoplia as sister to the rest of Nematoda in our analysis supports the hypothesis of a likely marine origin for the nematode ancestor. This subclass comprises two orders, Enoplida and Triplonchida. Notably, the representation of taxa in the latter group has significantly improved compared to transcriptomic analyses (Smythe et al. 2019; Ahmed et al. 2022). The monophyly of Triplonchida is not well-supported. Enoplida is

paraphyletic due to *Ironus dentifurcatus* being part of a lineage that is sister to the rest of Enoplida. This is consistent with the previous finding of Ahmed et al. (2022) who recovered Enoplida as paraphyletic, with *Bathylaimus* sp. (Triplonchida) and *Tobrilus* sp. (Enoplida) forming a sister branch to the rest of Enoplida.

There are five orders within Dorylaimia represented in this analysis, all of which are terrestrial. This clade was named in Blaxter et al. (1998) as clade I. A lineage consisting of the vertebrate parasitic orders Trichinellida and Dioctophymatida forms a sister branch with a clade comprising Dorylaimida, Mononchida and Mermithida. The orders Dorylaimida, Mononchida and Trichinellida, which are represented by more than one taxon, all form well-supported monophyletic groups, consistent with previous 18S rRNA-based and phylogenomic analyses (van Megen et al. 2009; Smythe et al. 2019; Ahmed et al. 2022). Apart from largely represented Rhabditida, three early-diverging Chromadoria orders are represented in this study, Chromadorida, Monhysterida and Plectida, with the first two being mostly marine. Within Rhabditida, all three suborders are strongly supported. The three infraorders representing Spirurina also received robust support. The suborders, Rhabditomorpha and Diplogasteromorpha within Rhabditina are likewise well-supported, although Bunonematomorpha was not represented. Even though the analysis accounted for all three infraorders of Tylenchina, of the majority of the taxa belonged to the infraorder Tylenchomorpha. In congruence with previous 18S rRNA and transcriptomic analyses of Tylenchina, Strongyloidoidea, one of the superfamilies of Panagrolaimomorpha, diverged first from the rest of Tylenchina, with two separate branching events–first with Steinernematidae, and then the Alloionematidae + Strongyloididae lineage. The remaining crown members of the tree formed a well-supported branch, within which Panagrolaimidae, another family of Panagrolaimomorpha, nested. Consistent with other phylogenomic studies, but contradictory to 18S rRNA data, Tylenchomorpha was found to be monophyletic, supporting the single evolutionary origin of the stomatostylet, although the superfamily Aphelenchoidea does not form a monophyletic clade within Tylenchomorpha.

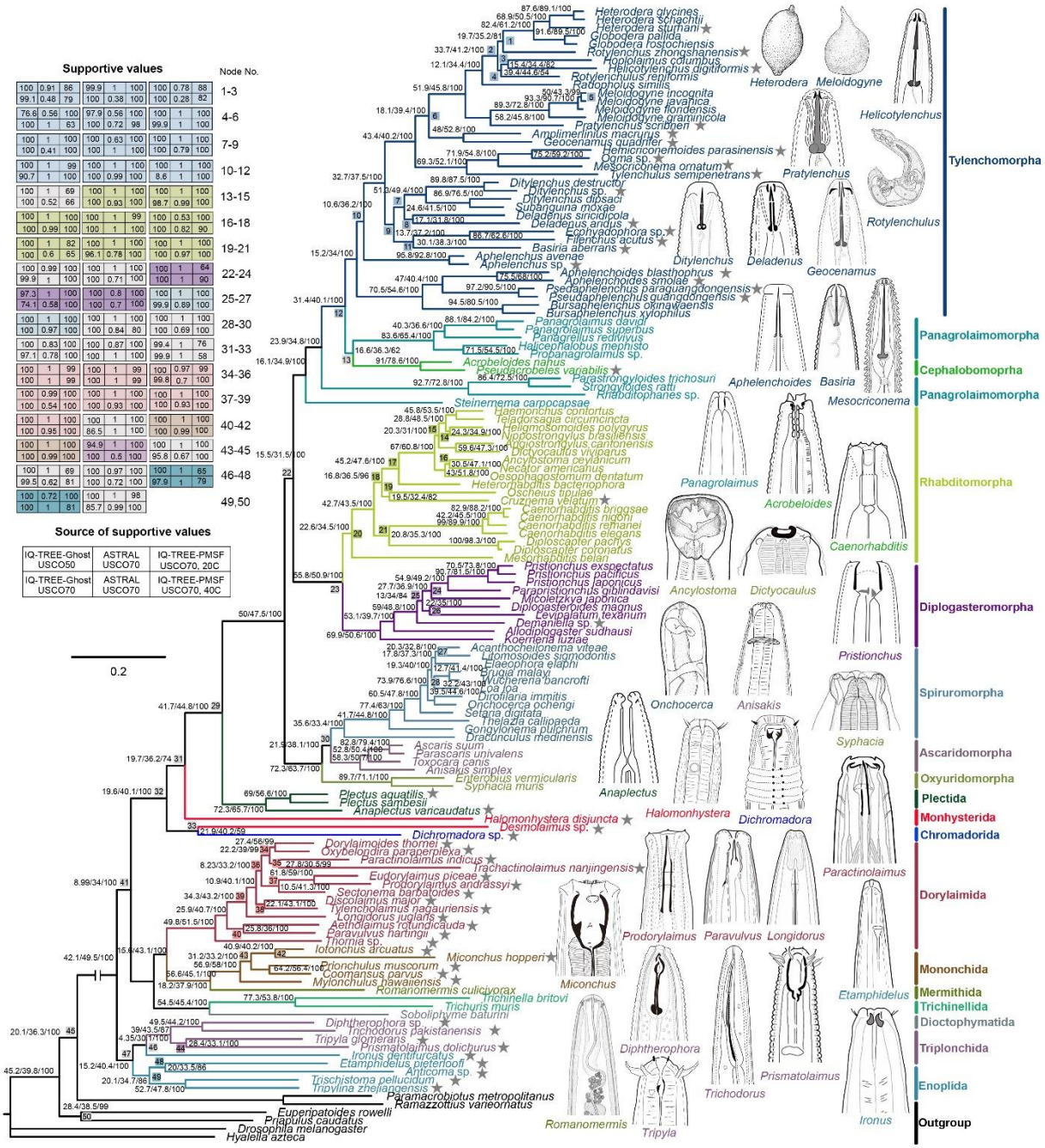

Figure 2. Phylogeny of Nematoda inferred from amino acids matrix USCO70 using the site-heterogeneous likelihood PMSF_C60 model implemented in IQ-TREE. Node labels show gCFs/sCFs/UFBoot. Node supports from other analyses are given in the table at left top corner, with the corresponding node numbers indicated by the colored squares at tree node. The nodes without corresponding nodes number suggest they were fully supported in all analyses. The newly sequenced species were indicated with asterisk.

Molecular dating of Nematoda

The divergence time dating, based on the ten calibrated nodes, traces the divergence date for Nematoda from its sister group Tardigrada in the Cambrian period (471-561 mya) (Figure 3). Within Nematoda, the most recent common ancestor of all Enoplia dated back to a period

spanning the beginning of the Silurian to the end of the Devonian. It also suggests a similar time for the common ancestor of all Chromadoria. All Dorylaimia had their most recent common ancestor ~ 380 mya, during the latter part of the Devonian through to the mid-Permian. Our analysis indicates that *Trichinella* and *Trichuris* shared a common ancestor ~170 mya, which is a considerably later period compared ~275 mya and ~281 mya, respectively reported in Zarlenger et al. (2006) and Korhonen et al. (2016). The most recent common ancestor of all stomatostylet-bearing nematodes is suggested to have existed ~250 mya, coinciding with the transition from the Permian to the Triassic period.

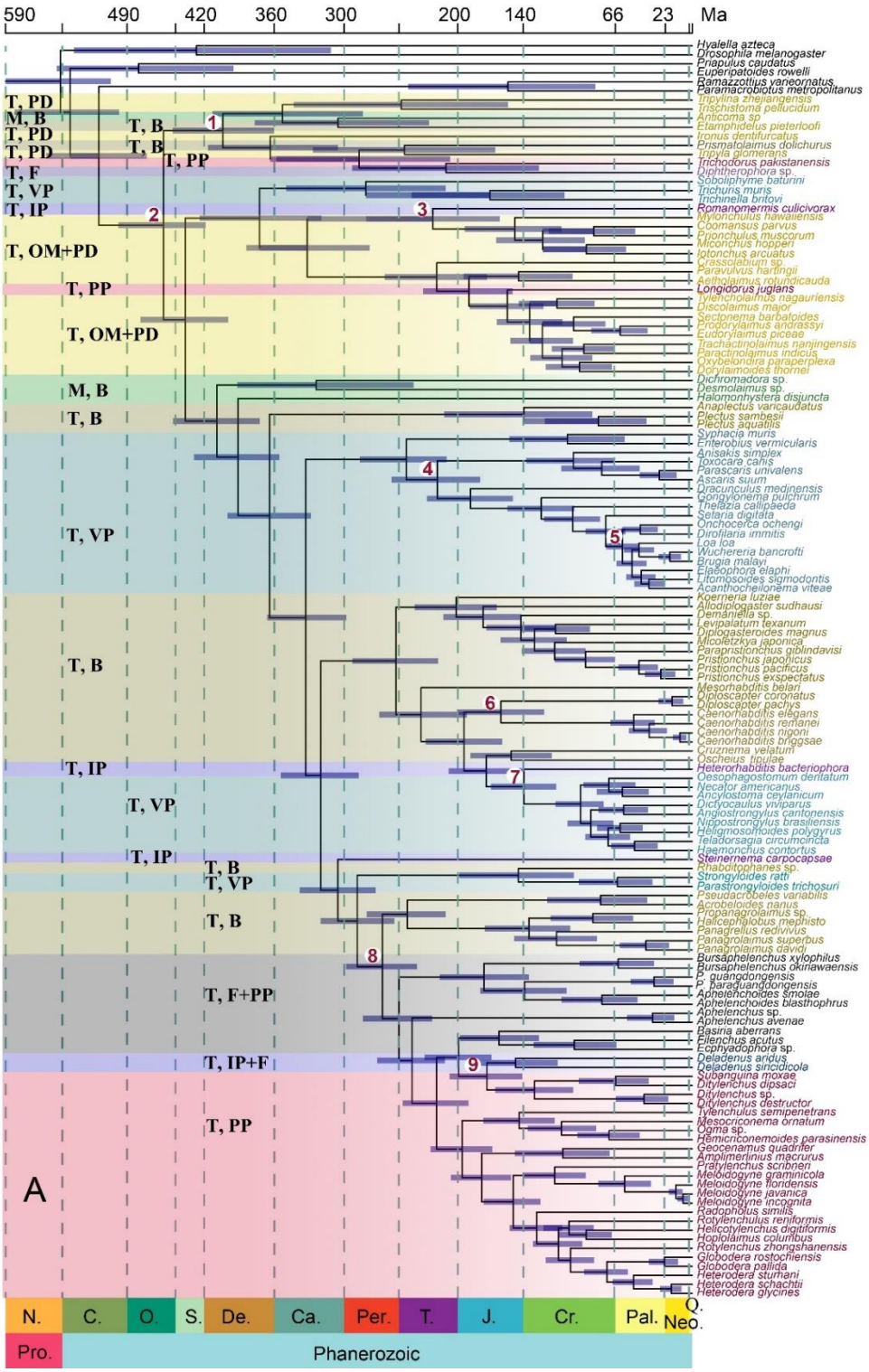

Figure 3. The Nematoda divergence time estimation performed by using MCMCTree on the PMSF_C60 topology. Node numbers and node bars represent 95% CIs of the estimated divergence times integrated from all MCMC runs. Number at nodes were constrained with the fossils records listed in supplementary table S1. Abbreviation in geologic time scale: Q. = Quaternary, Neo. = Neogene, Pal.= Paleogene, Cr. = Cretaceous, J. = Jurassic, T. = Triassic, Per. = Permian, Ca. = Carboniferous, De. = Devonian, S. = Silurian, O. = Ordovician, C. = Cambrian, N. = Neoproterozoic, Pro. = Proterozoic. Abbreviation in habitats and feeding type:

T= terrestrial, M = marine, PP = plant-parasitic, B = bacterivores, IP = invertebrate parasitic, F = fungivorous, VP = vertebrate parasitic, OM = omnivorous, PD = predatory.

Discussion

Evolution of early-branching Nematoda

The general topology of the tree agrees well results of previously-published analyses obtained using combined genomic and transcriptomic datasets for different sets of species (Smythe at al., 2019, Ahmed at al., 2022), with Enoplia (Clade II of Blaxter et al., 1998) diverging first from the rest of the nematodes, and Dorylaimia (Clade I of Blaxter et al., 1998) being close second, in full agreement with the scarce morphological and embryological evidence summarised by Holterman et al. (2006). Within Enoplia, all included species are represented by newly-sequenced genomes, and although none of the obtained genera/families are identical to earlier datasets, the resultant topology agrees in basic details with the transcriptome-only phylogenies (Smythe at al., 2019, Ahmed at al., 2022). However, the bootstrap support of several nodes within Enoplida is suboptimal: Triplonchida is resolved as monophyletic, while Enoplida as paraphyletic. This agrees with morphological evidence for the monophyly of Triplonchida suggested by De Ley & Blaxter (2001) but contradicts the phylogenies based on 18S rRNA gene (summarised in Ahmed & Holovachov, 2021). Enoplia (Clade II) is one of the most morphologically-diverse groups of nematodes, so there are no easily-testable morphology-based phylogenies. Enoplia are also likely to comprise the first nematode lineages that made the transition from marine to freshwater habitats in the Devonian-Carboniferous periods several times independently (Figure 2). Future sampling of additional genomes from all lineages within Enoplia will not only help resolve relationships within the group, but will also shed light on how colonisation of freshwater habitats took place in the early Palaeozoic.

The tree topology within Dorylaimia (Clade I) not only agrees fully with previously published datasets (Smythe at al., 2019, Ahmed at al., 2022), but considerably expanded dataset of Mononchida and especially Dorylaimida also allows further comparison with morphology- and rRNA-based phylogenies. Within Dorylaimida, the genus *Thornia* branches off first, which disagrees with morphological evidence (Peña-Santiago, 2014). The sole available 18S rRNA sequence of *Thornia* is highly divergent and its position is therefore unreliable (Fig. S1. in Holterman et al. 2019). The remaining Dorylaimida phylogeny is more in agreement with previous studies: Nygolaimina (represented here by *Paravulvus* and *Aetholaimus*) and Dorylaimina (all other taxa) are sister clades. The two families represented by more than one

species are Actinolaimidae (*Paractinolaimus* and *Trachactinolaimus*) and Qudsianematidae (*Eudorylaimus* and *Discolaimus*), are herein resolved as monophyletic and polyphyletic respectively, in agreement with previous rRNA-based analyses (summarised in Ahmed & Holovachov, 2021). The entire Clade I is represented by limno-terrestrial (Mononchida and Dorylaimidae) and parasitic (Trichinellida, Dioctophymatida and Mermithida) species. Of the latter, only Trichinellida includes considerable number of parasites inhabiting marine vertebrate species. Molecular dating suggests that the Clade I diverged off from the rest of Nematoda in the Devonian-Carboniferous, but the phylogeny of the group is not detailed enough to hypothesize about the transition between marine, freshwater, and terrestrial habitats within this lineage. Mermithida, parasitoids or insects and other (mostly) terrestrial arthropods are represented by single species in the current analysis; molecular dating suggests a very broad timeframe for the origin of the clade, spanning from Permian to Jurassic, much later than the period in which their arthropod hosts have diversified on land. Similarly, the plant-parasitic lineage within the Clade I is represented by a single species, *Longidorus juglans*, and molecular dating suggests it split off from non-parasitic dorylaims in Jurassic-Cretaceous, at about the same time their hosts, flowering plants. However, more data on the evolution of both hosts and parasites is necessary to make in-depth cophylogenetic study.

Evolution of plant-parasitic Tylenchina

Taxa closely related to plant-parasitic lineages have long been a focal point of interest in understanding the origins of plant parasitism (Bert et al., 2008; Holterman et al., 2017; Ahmed et al., 2020). Most rRNA-based phylogenies indicate that Cephalobomorpha shares the most recent common ancestry with Tylenchomorpha (Blaxter et al., 1998; Holterman et al., 2006; Bert et al., 2008; van Megen et al., 2009). However, our analyses are consistent with the analyses of Ahmed et al. (2020), which positions both Cephalobomorpha and Panagrolaimomorpha as sister taxa to Tylenchomorpha. The taxonomic position of aphelenchs (Aphelenchoidea: Aphelenchoididae + Aphelenchidae) relative to the tylenchs (Tylenchomorpha without aphelenchs) has been a subject of especially extensive debate. Despite the evident morphological similarities between Aphelenchidae and Aphelenchoididae, the latter has been placed outside the Tylenchomorpha based on rRNA-based phylogenies. They were separated from Aphelenchidae, which were found to be sister to the remaining members of Tylenchomorpha (van Megen et al., 2009). This placement was hypothesized as resulting from long branch attraction and/or elevated AT-content (De Ley and Blaxter, 2002; Holterman et al., 2006) and their monophyletic grouping could not be significantly rejected (Bert et al.,

2016). In contrast, mitogenome trees have consistently shown aphelenchs as monophyletic and positioned away from other Tylenchomorpha (Kern et al., 2020). Current analyses support the hypothesis that is most plausible based on morphology, namely that taxa equipped with the stomatostylet (the tylenchid and aphelenchid feeding apparatus) form a monophyletic group, a supposition that is congruent with the findings of Ahmed et al. (2020). Nonetheless, aphelenchs are still depicted as paraphyletic and Aphelenchidae are, similar to rRNA-based phylogenies, more closely related to tylenchs than to the morphologically more similar Aphelenchoidea.

Limitations of molecular dating of nematode phylogeny

Although the number of known nematode fossils is much larger than for some other soft-bodied invertebrate phyla (Poinar et al., 2011), their usefulness for molecular dating is limited. The large majority of known nematode fossils originate from various fossilised resins (amber and copal), are relatively recent and, due to the nature of the fossilisation, are strongly biased towards insect-associated and terrestrial nematodes. The next most common category also includes relatively recent fossils of vertebrate parasites found in coprolites and mummified remains. Only few fossils date back to Paleozoic, where the majority of nematode lineages are thought to have diverged. Comparing the dates of origin of parasitic nematodes and their respective hosts we find both agreements and disagreements. For example, the possible date of origin of Longidoridae and various Tylenchomorpha, parasites of flowering plants, matches that of their hosts during the Jurassic-Cretaceous period. On the other hand, the possible date of divergence of plant-parasitic Triplonchida, represented in our analysis by single *Trichodorus pakistanensis*, is overestimated (Permian to early Cretaceous) and predates the origin of their hosts, flowering plants (Casper et al., 2020). Similarly, the possible date of divergence of animal parasitic Ascaridomorpha is underestimated in our analysis (Triassic) compared to cophylogenetic study using fossil data from their vertebrate hosts (Carboniferous) (Li et al., 2018). The dating of the split between *Trichinella* and *Trichuris* is also inconsistent between our and previous analyses (Zarlenger et al, 2006; Korhonen et al. 2016). These examples not only expose the limitations of the fossils when dating the nematode phylogeny, but highlight the need for a multidisciplinary approach, combining fossil evidence, co-phylogenetic studies, and mutation accumulation rates.

Whole genome amplification and microscopic eukaryotes phylogenomic

Microscopic eukaryotes, such as Protozoan, Nematoda, Platyhelminthes, and Gastrotricha, are abundant, diverse and fill critical ecological roles across every ecosystem on Earth, and yet they were poorly known across subjects ranging from biodiversity and, evolution, to genomics

(Wardle, 2006; Bik et al., 2012). The lack of omics study in these groups can in part be explained by their low DNA quantity due to their small body size. Indeed, the high-throughput sequencing protocols require large amounts of input DNA (*e.g.* several micrograms), typically thousands of times more than that which is contained within a single microscopic individual. Zhang et al. (2019) proposed WGA in genome sequencing and phylogenomic of small animals, and two pioneering studies have used this method based on Illumina and PacBio platform to generate nematodes genomes (Lee et al., 2023; Doyle et al., 2019), although no study had yet evaluated its performance using microscopic species. In this study, we demonstrated that WGA generated smaller assembly and more fragmented contigs, probably due to the bias in the amplification process (Sabina et al., 2015). Albeit reduced, the reduced quality of results remains high enough for to allow studies flexible in genome quality, *e.g.* our nematode phylogenomic study. As a result, our study supports WGA as being a powerful tool in phylogenomic, population genomic and metagenomic of microscopic eukaryotes.

Acknowledgements: We thank Prof. Wilfrida Decraemer and Laboratory of Parasitology, Ghent University for providing nematode images. We acknowledge the Smithsonian Institution High Performance Cluster (https://doi.org/10.25572/SIHPC) for providing computational resources and support that have contributed to the research results reported in this publication. X.Q. was funded by National Natural Science Foundation of China (32001876). Y.M.Z. was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 101024056.

Supplementary material:

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.b2rbnzsnt

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May; 19(5):455–477. https://doi.org/10.1089/cmb.2012.0021 Epub 2012 Apr 16. PMID: 22506599

Blaxter M, Koutsovoulos G. The evolution of parasitism in Nematoda. Parasitology. 2015 Feb;142 (Suppl 1):S26–S39. https://doi.org/10.1017/S0031182014000791Epub 2014 Jun 25. PMID: 24963797 PMCID: PMC4413787

Borowiec ML. Spruceup: fast and flexible identification, visualization, and removal of outliers from large multiple sequence alignments. J Open Source Softw. 2019 Oct; 4(42):1635. https://doi.org/10.21105/joss.01635

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009 Aug; 25(15):1972–1973. https://doi.org/10.1093/bioinformatics/btp348 Epub 2009 Jun 8. PMID: 19505945 PMCID: PMC2712344

van der Kooi CJ, Ollerton J. The origins of flowering plants and pollinators. Science. 2020 Jun; 368(6497):1306–1308. https://doi.org/10.1126/science.aay3662 PMID: 32554579

Coomans, A. (2000). Nematode systematics: past, present and future. Nematology 2, 3–7.

Crotty SM, Minh BQ, Bean NG, Holland BR, Tuke J, Jermiin LS, von Haeseler A. GHOST: Recovering Historical Signal from Heterotachously Evolved Sequence Alignments. Syst Biol. 2020 Mar; 69(2):249–264. https://doi.org/10.1093/sysbio/syz051 PMID: 31364711

De Ley P, Blaxter ML. Systematic position and phylogeny. In: Lee DL, editor. The biology of Nematodes. CRC Press; 2002. p. 1–30. https://doi.org/10.1201/b12614. ISBN: 978-0-415-27211-7

Doyle SR, Sankaranarayanan G, Allan F, Berger D, Jimenez Castro PD, Collins JB, Crellen T, Duque-Correa MA, Ellis P, Jaleta TG, et al. Evaluation of DNA extraction methods on individual helminth egg and larval stages for whole-genome sequencing. Front Genet. 2019 Sep;10:826. https://doi.org/10.3389/fgene.2019.00826 PMID: 31616465 PMCID: PMC6764475

Gentekaki E, Kolisko M, Gong Y, Lynn D. Phylogenomics solves a long-standing evolutionary puzzle in the ciliate world: The subclass Peritrichia is monophyletic. Mol Phylogenet Evol. 2017 Jan; 106:1–5. https://doi.org/10.1016/j.ympev.2016.09.016 Epub 2016 Sep 19. PMID: 2765972

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010 May; 59(3):307–321. https://doi.org/10.1093/sysbio/syq010 Epub 2010 Mar 29. PMID: 20525638

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. Mol Biol Evol. 2018 Feb; 35(2):518–522. https://doi.org/10.1093/molbev/msx281 PMID: 29077904

International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. Nat Genet. 2019 Jan; 51(1):163–174. https://doi.org/10.1038/s41588-018-0262-1 Epub 2018 Nov 5. PMID: 30397333 PMCID: PMC6349046

Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics . 2010 Jul; 26(13):1669–1670. https://doi.org/10.1093/bioinformatics/btq243 Epub 2010 May 13. PMID: 20472542 PMCID: PMC2887050

Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017 Jun; 14(6):587–589. https://doi.org/10.1038/nmeth.4285 Epub 2017 May 8. PMID: 28481363 PMCID: PMC5453245

Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 2008 Jul; 9(4):286–298. https://doi.org/10.1093/bib/bbn013 Epub 2008 Mar 27. PMID: 18372315

Kern EMA, Kim T, Park J-K. The mitochondrial genome in nematode phylogenetics. Front Ecol Evol. 2020 August; 8:250. https://doi.org/10.3389/fevo.2020.00250

Korhonen PK, Pozio E, La Rosa G, Chang BC, Koehler AV, Hoberg EP, et al. Phylogenomic and biogeographic reconstruction of the *Trichinella* complex. Nat Commun. 2016 Feb;7:10513. https://doi.org/10.1038/ncomms10513 PMID: 26830005 PMCID: PMC4740406

Kück P, Longo GC. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front Zool. 2014 Nov;11(1):81. https://doi.org/10.1186/s12983-014-0081-x PMID: 25426157 PMCID: PMC4243772

Lepere C, Demura M, Kawachi M, Romac S, Probert I, Vaulot D. Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations. FEMS Microbiol Ecol. 2011 Jun; 76(3):513–523. https://doi.org/10.1111/j.1574-6941.2011.01072.x Epub 2011 Mar 16. PMID: 21348885

Li L, Lü L, Nadler SA, Gibson DI, Zhang L-P, Chen H-X, Zhao W-T, Guo Y-N. Molecular phylogeny and dating reveal a terrestrial origin in the early carboniferous for ascaridoid nematodes. Syst Biol. 2018 Sep; 67(5):888–900. https://doi.org/10.1093/sysbio/syy018 PMID: 29528459

Lorenzen, S. (1981). Entwurf eines phylogenetischen Systems der freilebenden Nematoden (Outline of a phylogenetic system for the free-living nematodes). Veröff. Inst. Meeresforsch. Bremerh. Suppl. 7, 1–472.

Minh BQ, Hahn MW, Lanfear R. New methods to calculate concordance factors for phylogenomic datasets. Mol Biol Evol. 2020 Sep; 37(9):2727–2733. https://doi.org/10.1093/molbev/msaa106 PMID: 32365179 PMCID: PMC7475031

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020 May; 37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015 PMID: 32011700 PMCID: PMC7182206

Naser-Khdour S, Minh BQ, Zhang W, Stone EA, Lanfear R. The prevalence and impact of model violations in phylogenetic analysis. Genome Biol Evol. 2019 Dec; 11(12):3341–3352. https://doi.org/10.1093/gbe/evz193 PMID: 31536115 PMCID: PMC6893154

Ocaña KACS, Dávila AMR. Phylogenomics-based reconstruction of protozoan species tree. Evol Bioinform Online. 2011; 7:107–121. https://doi.org/10.4137/EBO.S6861 Epub 2011 Jul 31. PMID: 21863127 PMCID: PMC3153122

Poinar GO. The evolutionary history of nematodes as revealed in stone, amber and mummies. Nematology Monographs and Perspectives.,Volume 9. (Hunt D, Perry R N. eds.). 429 pp. Leiden, The Netherlands : Koninklijke Brill NV, 2011. ISBN: 978-90-04-17521-1

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2018. Available online at https://www.R-project.org/.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. 2018 Sep; 67(5):901–904. https://doi.org/10.1093/sysbio/syy032 PMID: 29718447 PMCID: PMC6101584

Puttick, MN, 2019. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics, 35(24), 5321–5322. https://doi.org/10.1093/bioinformatics/btz554 PMID: 31292621

Sabina J, Leamon JH. Bias in whole genome amplification: Causes and considerations. Methods Mol Biol. 2015; 1347:15–41. https://doi.org/10.1007/978-1-4939-2990-0_2 PMID: 26374307

Sahraei SE, Sanchez-Garcia M, Montoliu-Nerin M, Manyara D, Bergin C, Rosendahl S, Rosling A. Whole genome analyses based on single, field collected spores of the arbuscular mycorrhizal fungus *Funneliformis geosporum*. Mycorrhiza. 2022 Nov; 32(5–6):361–371. https://doi.org/10.1007/s00572-022-01091-4 Epub 2022 Sep 26. PMID: 36161535 PMCID: PMC9560946

Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. Mol Biol Evol. 2016 Jul; 33(7):1654–1668. https://doi.org/10.1093/molbev/msw079. Epub 2016 Apr 15. PMID: 27189547 PMCID: PMC4915361

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015 Oct; 31(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351 Epub 2015 Jun 9. PMID: 26059717

Stevens L, Félix MA, Beltran T, Braendle C, Caurcel C, Fausett S, Fitch D, Frézal L, Gosse C, Kaur T, Kiontke K, Newton MD, et al. Comparative genomics of 10 new *Caenorhabditis* species. Evol Lett. 2019 Apr; 3(2):217–236. https://doi.org/10.1093/10.1002/evl3.110 PMID: 31007946 PMCID: PMC6457397

Telford MJ, Budd GE, Philippe H. Phylogenomic insights into animal evolution. Curr Biol. 2015 Oct; 25(19):R876–R887. https://doi.org/10.1016/j.cub.2015.07.060 PMID: 26439351

Viney M. How can we understand the genomic basis of nematode parasitism? Trends Parasitol. 2017 Jun; 33(6):444–452. https://doi.org/10.1016/j.pt.2017.01.014 Epub 2017 Mar 6. PMID: 28274802 PMCID: PMC5449551

Wang HC, Minh BQ, Susko E, Roger AJ. 2018. Modeling site heterogeneity with posterior mean site frequency profiles accelerates accurate phylogenomic estimation. Syst Biol. 2018 Mar; 67(2):216–235. https://doi.org/10.1093/sysbio/syx068 PMID: 28950365

Weisz F, Lalle M, Nohynkova E, Sannella AR, Dluhošová J, Cacciò SM. Testing the impact of Whole Genome Amplification on genome comparison using the polyploid flagellated *Giardia duodenalis* as a model. Exp Parasitol. 2019 Dec; 207:107776. https://doi.org/10.1016/j.exppara.2019.107776 Epub 2019 Oct 16. PMID: 31628895

Williams TA, Cox CJ, Foster PG, Szöllősi GJ, Embley TM. Phylogenomics provides robust support for a two-domains tree of life. Nat Ecol Evol. 2020 Jan; 4(1):138–147. https://doi.org/10.1038/s41559-019-1040-x. Epub 2019 Dec 9. PMID: 31819234 PMCID: PMC6942926

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol . 2019 Nov 28;20(1):257. https://doi.org/10.1186/s13059-019-1891-0. PMID: 31779668 PMCID: PMC6883579

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007 Aug;24(8):1586–1591. doi: 10.1093/molbev/msm088. Epub 2007 May 4. PMID: 17483113

Lee Y-C, Ke H-M, Liu Y-C, Lee H-H, Wang M-C, Tseng Y-C, Kikuchi T, Tsai I J. Single-worm long-read sequencing reveals genome diversity in free-living nematodes, Nucleic Acids Res. 2023 Aug; 51(15):8035–8047. https://doi.org/10.1093/nar/gkad647. PMID: 37526286 PMCID: PMC10450198

Zarlenga DS, Rosenthal BM, La Rosa G, Pozio E, Hoberg EP. Post-Miocene expansion, colonization, and host switching drove speciation among extant nematodes of the archaic genus *Trichinella*. Proc Natl Acad Sci U S A. 2006 May; 103(19):7354–7359. https://doi.org/10.1073/pnas.0602466103. Epub 2006 May 1. PMID: 16651518 PMCID: PMC1464345

Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018 May; 19(Suppl 6):153. https://doi.org/10.1186/s12859-018-2129-y PMID: 29745866 PMCID: PMC5998893

Zhang F, Ding Y, Zhu CD, Zhou X, Orr MC, Scheu S, Luan YX. Phylogenomics from low-coverage whole-genome sequencing. Methods Ecol Evol. 2019 Apr; 10(4):507-517. https://doi.org/10.1111/2041-210X.13145