

Investigating the Evolutionary Diversity of Nematodes from a Genomic Perspective

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln



vorgelegt von
Joseph Kirangwa
aus Uganda
2024

Berichterstatter (Gutachter):

Dr. Philipp Schiffer

Prof. Dr. Siegfried Roth

Dissertationsgesuch: 15.04.2024

Disputation: 05.07.2024

Abstract

The phylum Nematoda comprises a vast array of species with diverse lifestyles. However, our understanding of the genomic mechanisms underlying nematode diversity remains limited. Through extensive genome comparisons among nematodes and related taxa, this thesis has leveraged the newly sequenced high-quality genomes to investigate the evolutionary plasticity of nematode Hox gene complements and genomic loci arrangements. Hox genes are central to metazoan body plan formation, patterning and evolution, playing a critical role in cell fate decisions early in embryonic development in invertebrates and vertebrates. While the archetypical Hox gene cluster consists of members of nine ortholog groups (HOX1-HOX9), arrayed in close linkage in the order in which they have their anterior-posterior patterning effects, nematode Hox gene sets do not fit this model. The *Caenorhabditis elegans* Hox gene set is not clustered and contains only six Hox genes from four of the ancestral groups. The pattern observed in *C. elegans* is not typical of the phylum, and variation in orthologue set presence and absence and in genomic organization has been reported. Recent advances in genome sequencing have resulted in the availability of many novel genome assemblies in Nematoda, especially from taxonomic groups that had not been analyzed previously. I have explored Hox gene complements in high-quality genomes of 80 species from all major Clades of Nematoda to understand the evolution of this key set of body pattern genes and especially to probe the origins of the “dispersed” Hox cluster observed in *C. elegans*. I found that nematodes can have Hox genes from up to six orthology groups. While nematode Hox “clusters” are often interrupted by unrelated genes, I not only identified species in which the cluster is intact and not dispersed but also Hox gene losses, inversions, and translocations within various nematode lineages. This study emphasizes the diversity of Hox gene evolution within the phylum.

DNA methylation influences genome evolution by directly affecting transposon silencing, gene expression, genome stability, and mutation rates. DNA methylation has only been scarcely studied in nematodes. Moreover, its study in newly accessible high-quality nematode genomes remains limited. Here I have investigated the presence of genomic signatures of DNA methylation in two phyla, Nematoda and Nematomorpha. I found variations in the conservation of critical DNA methylation enzymes in Nematoda, revealing the dynamic nature of DNA methylation as an evolving epigenetic mechanism. While certain nematode species like *C. elegans* lack the DNA methylation system, others exhibit variations of DNA Methyltransferases (DNMTs) similar to mammalian DNMT3 and DNMT1 enzymes which are important enzymes responsible for introducing

and maintaining genomic DNA methylation. I found this modification to be widespread in some early-branching nematode lineages, including both animal parasitic and free-living species from Clade I, and absent in others like Clade II Enoplean nematodes. Similarly, limnoterrestrial nematomorphs from the superfamily Gordioidea had a full complement of DNMTs responsible for DNA methylation, whereas marine *Nectonema munidae* showed its absence. This investigation contributes to the existing body of evidence, emphasizing the presence of DNA methylation as an ancient regulatory mechanism in nematodes. Furthermore, it provides evidence of secondary loss of DNA methylation in model organisms like *C. elegans* and other nematodes belonging to Clades II, III, IV and V, underscoring the evolutionary dynamics of this regulatory process within the Nematoda.

Indeed nematodes are abundant and diverse but also include many parasitic species. Previous molecular phylogenetic studies have revealed parasitism of plants and animals has occurred independently many times in nematode evolution. The recent availability of high-quality nematode genomes has enabled the analysis of genetic data from various nematode species. This data holds the potential to enhance our understanding of genetic signatures associated with parasitic lifestyles. Within nematode parasites, there are also parasitoids, which kill their host to complete their lifecycle. Mermithidae and Nematomorpha are parasitoids sharing commonalities in their lifestyle: immature stages infect arthropod hosts, manipulate hosts to induce water-seeking behavior, and emerge as free-living adults, often killing their host. Some of these species are of great economic importance, being evaluated as biological control agents against mosquito vectors responsible for diseases like malaria, and other insect pests, but with scarce genomic resources currently available. Nematomorpha, despite being closely related to Nematoda, has received insufficient attention in genomic research, leading to gaps in our understanding of their diverse genetic makeup. Therefore, I investigated the genetic features encoded in the genomes of both parasitoid taxa to identify similarities and parallels linked to their ecological lifestyles. I performed a comparative analysis of 12 genomes, comprising parasitoid, parasitic and free-living worms. The investigation revealed genomic signatures in parasitoid species, including expanded gene families enriched in neural transmission modulation, likely linked to the known host manipulation that both mermithids and nematomorphs exert on their hosts. The analysis also uncovered a diverse array of conserved transposable element superfamilies across both lineages. The findings from this study provide valuable insights into the potential genomic adaptations associated with parasitoidism in nematode and nematomorph worms. The identification of expanded gene families and conserved transposable element su-

peramilies sheds light on the molecular underpinnings of their unique biological traits. Additionally, the core set of orthologs specific to parasitoid worms offers new avenues for understanding the evolution of parasitism within these groups of organisms. This comparative genomics resource serves as a significant asset for the research community, facilitating further investigations into the biology and ecology of parasitoid mermithids and nematomorph worms.

Graphical Abstract

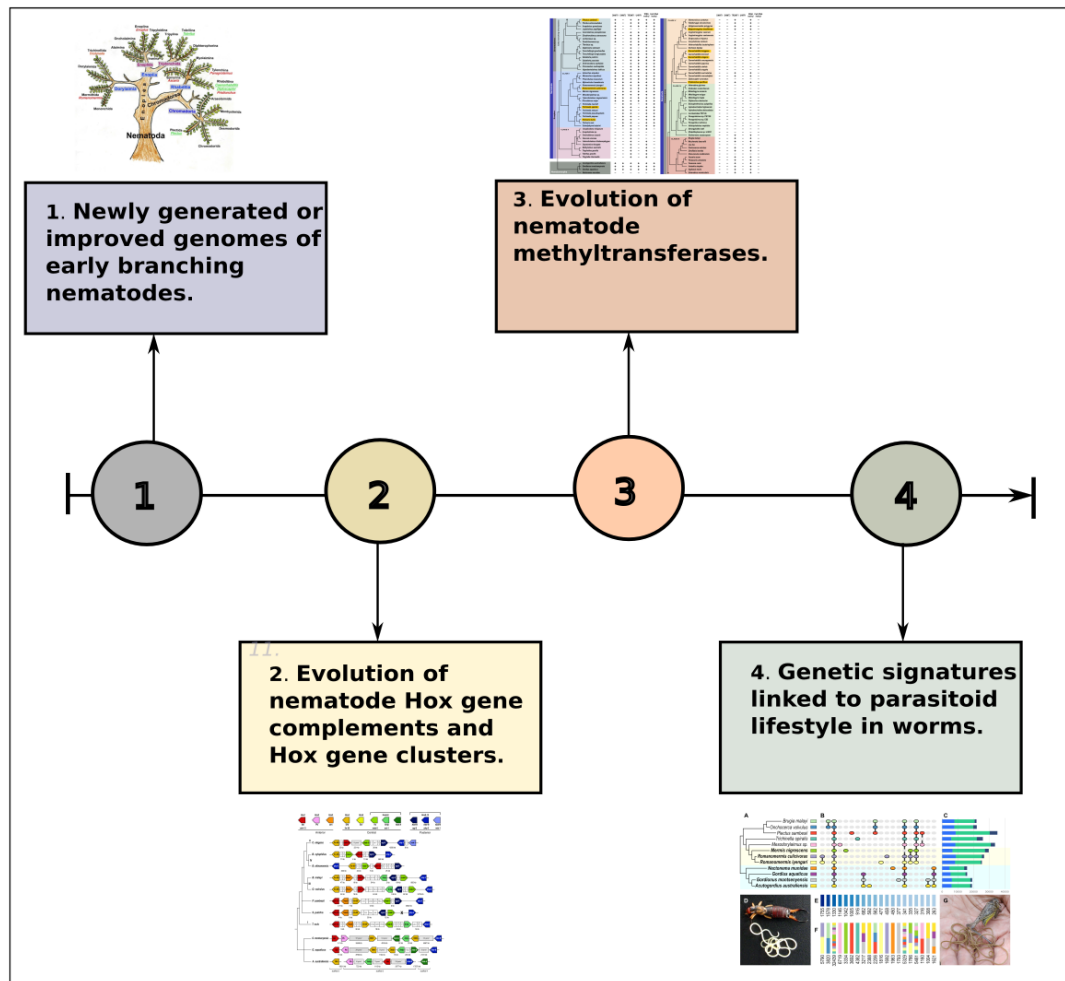


Figure 1: The work of this thesis is summarized. (1) High-quality genomes were initially generated, (2) Then followed the study of Hox gene evolution, (3) Evolution of DNA methylation across the Nematoda. (4) Study of genetic signatures encoded in worm parasitoids linked to their ecological lifestyle.

Contents

1	BACKGROUND	1
1.1	The diverse phylum Nematoda	3
1.1.1	Major Characteristics of Chromadorea	9
1.1.2	Key Characteristics of Enoplea	11
1.1.3	Description of nematode body structures	12
1.1.4	Diversity of nematode in embryonic development and re-productive cycles, feeding and survival strategies	14
1.1.5	The conserved simple nematode body plan	16
1.1.6	Nematoda as an important model for investigating developmental evolution	17
1.2	Evolution of nematode genomes	18
1.2.1	Nematode genome size evolution	18
1.2.2	Evolution of Nematode chromosome number and size	18
1.2.3	The evolution of gene content in nematodes	19
1.2.4	The evolution of HOX genes	21
1.2.5	The evolution of critical genes for the DNA methylation machinery	25
1.2.6	The evolution of gene content in parasitic nematodes	26
1.3	Neglected lineages in Nematoda	29
2	METHODS	30
2.1	Establishment of high-quality genomic resources	30
2.1.1	Nematode sample collection and processing	31
2.1.2	Extraction of live nematode specimens	32
2.1.3	Morphological and molecular identification of nematodes	34
2.1.4	Culturing of <i>Mononchus aquaticus</i> , <i>Aphelenchus avenae</i> and <i>Tobrilus gracilis</i>	36
2.1.5	Establishment of single and pooled worm HMW DNA extraction	37
2.1.6	Long read sequencing	39
2.2	Long-read assembly of nematode genomes	40
2.3	Nematode Hox gene family analysis	49
2.3.1	Hox cluster annotation and gene identification	50
2.3.2	Non-Hox gene interruptions within Hox clusters	51
2.3.3	microRNA interruptions within Hox clusters	51
2.4	Genomic signatures of DNA methylation in the phylum Nematoda	51
2.4.1	Genome data acquisition	51

2.4.2	Identification of DNMTs in nematode genomes	52
2.5	Analysis of genetic features encoded in the genomes of worm par- asitoid taxa	52
2.5.1	Orthologous cluster identification and visualization	53
2.5.2	Gene family analysis	53
2.5.3	Repeat sequence analysis	54
3	RESULTS	55
3.1	Genomes of early branching nematode species from long-read se- quencing	55
3.1.1	<i>Romanomermis</i> genomes had a high level sequence repetitions	55
3.1.2	The genome assembly of a predatory nematode <i>Mononchus</i> <i>aquaticus</i> contains a full DNMT methylation toolkit	56
3.1.3	The genome assembly of a free-living parthenogenetic ne- matode <i>Tripylella intermedia</i>	56
3.1.4	The genome assembly of the only free-living freshwater sex- ual nematode <i>Tobrilus gracilis</i>	57
3.2	Evolutionary plasticity in Nematode Hox gene complements and genomic loci arrangement	57
3.2.1	HOX gene evolution in nematodes is largely characterized by consecutive losses	57
3.2.2	The analysis pipeline reliably recovers <i>C. elegans</i> HOX loci	58
3.2.3	Hox6-8 (Antp) loci variation in Clade V (including an in- group Strongylomorpha=Strongyoidea) crown group nema- todes	58
3.2.4	Hox3 gene loss in Tylenchina and Spirurina	58
3.2.5	Hox gene complements in early-branching chromadoreans .	58
3.2.6	An independent loss of ant-1 locus in Enoplida, and a single HOX9-13 (php-3) in Enoplean nematodes	59
3.2.7	Nematomorpha have at least six Hox loci	59
3.2.8	nob-1 is a recent duplication within <i>Caenorhabditis</i>	61
3.2.9	Hox cluster organization in the Nematoda	61
3.2.10	Unique characteristics of Hox gene organization in <i>Bur-</i> <i>saphelenchus</i>	62
3.2.11	Some members of Spirurina (Clade III) display Hox clusters containing a greater complement of Hox genes.	63
3.2.12	Hox cluster organization in basal Chromadorea	63
3.2.13	Hox loci in <i>Trichuris</i> are linked together in <i>CiS</i> and adhere to the model order	63

3.2.14	Hox cluster arrangement in Nematomorpha	63
3.2.15	Hox clusters in nematodes are interrupted by non-Hox genes	64
3.2.16	Hox clusters in nematodes are interrupted by miRNAs . .	64
3.3	Evolution of DNA methylation genes across the Nematoda	65
3.3.1	Distinct patterns of DNMT and TRDMT1 presence in Eno- plean nematodes (Clades I-II)	66
3.3.2	Presence of DNMT1 in Chromadorean nematodes (includ- ing Clades III-V)	67
3.3.3	DNA methyltransferases are conserved in Gordioidea ne- matomorphs	67
3.4	Comparative genomics of parasitoid lifestyle as exemplified by Mermithi- dae and Nematomorpha	67
3.4.1	Orthologous clusters in parasitoid worms	67
3.4.2	Mermithid orthologous clusters	68
3.4.3	Nematomorpha orthologous clusters	69
3.4.4	Expanded gene families in mermithids and nematomorphs	70
3.4.5	Mermithid expanded and contracted gene family clusters .	70
3.4.6	Nematomorphs expanded and contracted gene families . .	71
3.4.7	Transposable elements	72
4	DISCUSSION	75
4.1	Genomes of early branching nematode species	75
4.2	Evolutionary plasticity in Nematode Hox gene complements and genomic loci arrangement	76
4.3	Evolution of DNA methylation genes across the Nematoda	81
4.4	Comparative genomics of mermithida and nematomorpha	84
5	CONCLUSIONS AND FUTURE PERSIPECTIVES	88
6	ACKNOWLEDGEMENTS	91
7	DEDICATION	93
8	DECLARATION	94
	References	95
	Curriculum Vitae	125

List of Figures

1	Thesis graphical abstract.	
2	Cladogram showing nematodes Clades relevant to this work . . .	5
3	Global map depicting total nematode abundance	6
4	The general life cycle of nematodes.	15
5	Nematode processing workflow	31
6	Collection site for <i>Tobrilus gracilis</i>	32
7	Extraction of live nematode specimen	34
8	Nematode 18S PCR cycle	35
9	Establishment of single and pooled worm HMW DNA	38
10	Fragment length profile of extracted HMW DNA	39
11	General workflow for the assembly nematode draft genomes using PacBio HiFi long reads.	40
12	Theoretical k-mer frequency	42
13	Hifiasm algorithmic outline	43
14	Hi-C overview	47
15	General workflow for primary annotation of draft genomes	48
16	Hox gene identification workflow	49
17	Pipeline for orthologous cluster analysis of worm parasitoids . . .	53
18	Nematoda and Nematomorpha Hox gene complements	60
19	nob-1 Hox gene complement across <i>Caenorhabditis</i>	61
20	Nematode Hox gene clusters	62
21	Nematoda DNA methylation evolution	66
22	Orthologous cluster analysis of parasitoid worms	68
23	Changes in Gene family expansion and contraction	70
24	TE and repeat divergence landscapes	73
25	Relative amounts of TE superfamilies	74

List of Tables

1	Long-read Enoplean nematode genomes	55
2	Hox cluster features of Nematoda and Nematomorpha	64
3	Mermithid orthologous clusters	69
4	Nematomorphs orthologous clusters	69
5	Mermithid expanded gene family clusters	71
6	Nematomorph expanded gene family clusters	72

1 BACKGROUND

Nematoda is a species-rich and diverse phylum, best known for its model species *Caenorhabditis elegans*, used, among a myriad of other topics, to study early embryonic development, gene regulation and evolutionary adaptation. Nematodes as a whole have a remarkably conservative body plan (Aboobaker & Blaxter, 2003; Schulze & Schierenberg, 2011) contrasted by exceptionally high ecological and genomic diversity (Hodda, Peters, Traunspurger, et al., 2009; Nigon & Félix, 2018; “Comparative genomics of the major parasitic worms”, 2019; Stein et al., 2003). Consequently, the knowledge we gained from studying *C. elegans* can’t always be directly applied to other nematode species. In this thesis, I will use three examples to show that the best-studied animal is not the most representative of its phylum, and that such fundamental biological processes as early embryonic development (HOX genes) and epigenetic gene regulation (DNA methylation) in early branching nematodes are more similar to those in other animal phyla, and that new model species are crucial for understanding of nematodes incredible adaptability to exploit various environs, including bodies of other organisms (parasitism).

Changes in developmental processes are fundamental to the emergence of morphological variations, and these alterations are linked to genetic modifications inherited across generations (Hughes & Kaufman, 2002; Carroll, 2008). The identification of these genetic alterations facilitates insights into the evolution of animal body plans, albeit a challenging endeavor, especially when morphological shifts occurred millions of years ago (Mulhair & Holland, 2024). A promising approach is to investigate genes shared among different taxa, particularly those pivotal in development, often referred to as the developmental toolkit. This gene toolkit, a conserved set of genes found across all bilaterian organisms, plays fundamental roles in directing body axis formation and determining cell fate during embryonic development (Ryan & Baxeavanis, 2007). These genes have been repurposed multiple times to generate a diverse array of species-specific traits (Carroll, 2000; Abzhanov, Protas, Grant, Grant, & Tabin, 2004; Gillis & Hall, 2016). Among the components of this toolkit are well-known signaling pathways like Hedgehog, Wnt, and Notch, as well as transcription factors such as homeobox genes (Carroll, 2005; Erwin, 2009) which together orchestrate development in bilaterian lineages. Hox genes form a core part of the developmental toolkit genes. They are a set of homeobox transcription factors, that play a pivotal role in patterning the anteroposterior axis. In *Drosophila*, these genes are organized into two clusters on the same chromosome, with their spatial arrangement precisely reflecting their expression pattern (Mulhair & Holland, 2024). By reconstructing the ancestral

protostome Hox cluster based on gene complements in existing phyla, it appears that at least nine orthologous groups have maintained their genomic organization across various species (Aboobaker & Blaxter, 2003, 2003). While traditionally considered highly conserved, tightly clustered, and spatially aligned across all animals, the genomic organization of Hox genes can sometimes deviate from this canonical pattern (Duboule, 2007). For example, the Hox cluster in *C. elegans* does not exhibit conservation or clustering (Aboobaker & Blaxter, 2003, 2003; Duboule, 2007; Aboobaker & Blaxter, 2010). Instead, *C. elegans* possess a diminished and scattered gene complement resulting from a gradual loss of Hox genes over its evolutionary trajectory (Baker & Woollard, 2019). However, **how representative is the reduced Hox cluster of *C. elegans* for nematodes in general?** This thesis investigated the Hox gene complements of a wide range of newly-sequenced, high-quality genomes from diverse non-model nematodes to understand the evolution of this key set of body patterning genes.

Epigenetic regulations, which encompass chromatin-based regulatory mechanisms without altering the DNA sequence, such as chromatin remodeling, histone modification, and DNA methylation, are crucial for the precise and stable control of gene expression and genomic regulation (Berger, Kouzarides, Shiekhata, & Shilatifard, 2009). DNA cytosine methylation is a widespread epigenetic modification observed across a variety of eukaryotic organisms (Zemach & Zilberman, 2010), and seems to play a role in suppressing transposable elements (TEs) and repetitive sequences, maintaining genome stability, influencing development, and controlling transcription (Zemach & Zilberman, 2010; Finnegan, Peacock, & Dennis, 2000; Weber & Schübeler, 2007). DNA methylation has been observed in gene bodies, promoters, and repetitive elements in examined vertebrates and plants. In contrast, most invertebrate lineages exhibit DNA methylation preferences within gene bodies (Zemach & Zilberman, 2010; Z. Su, Han, & Zhao, 2011). However, initial studies did not find DNA methylation in the model *C. elegans* which led to the notion that DNA methylation was absent in Nematodes. Nonetheless, DNA methylation of repeats in previously analyzed nematodes stands out as a widely preserved function, likely originating in the common ancestor of nematodes. Methylation within the bodies of transcribed protein-coding genes has been lost in the lineage leading to *T. spiralis* and *T. muris*, as well as in *P. sambesii*. In *R. culicivora*, methylation occurs in only a minority of genes, indicating a not universally conserved feature of DNA methylation in nematodes (Rošić et al., 2018). The presence of methylated promoters is commonly understood as a signal for gene repression, while methylation within transposable elements (TEs) is viewed as a protective mechanism against their harmful spread within genomes (Z. Su et al., 2011). The regulation of gene expression during development has

been demonstrated to aid in the transition between various life cycle stages of parasitic nematode *Trichinella spiralis* (Gao et al., 2012). DNA methylation has been lost several times in the course of animal evolution, including in lineages leading to *Drosophila* and *C. elegans* (Zemach & Zilberman, 2010). This thesis revisits the evolution of the DNA methylation machinery in non-model nematode genomes that were missing in previous studies to answer the question: **to what extent is the absence of DNA methylation, as observed in *C. elegans*, is conserved across Nematoda, particularly in light of recent high-quality genomic data from previously underrepresented nematode species?** The aim here was to investigate the evolution of genomic signatures of DNA methylation in two closely related Nematoda and Nematomorpha phyla.

Nematodes are undoubtedly very successful in exploiting other organisms, both animals and plants. They have repeatedly evolved parasitic lifestyles during their evolutionary history (M. L. Blaxter et al., 1998). Among them, Mermithidae are distinguished by killing their host in the process of completing their life cycle, unlike other nematode parasites. In this way, they are similar to the phylum Nematomorpha, and are united under the term "parasitoid". Commonalities in the lifestyle of mermithids and nematomorphs include: the immature stages infect arthropod hosts, species from both phyla can manipulate their host behavior, and both have a final free-living non-feeding adult reproductive stage, usually killing their host upon emergence. Some of these species are of great economic importance, being evaluated as biological control agents against mosquito vectors responsible for diseases like malaria, and against insect pests, but with scarce genomic resources currently available. Nematomorphs, despite being closely related to Nematoda, have received insufficient attention in genomic research, leading to gaps in our understanding of their diverse genetic makeup. Another question tackled in this thesis is: **which genetic signatures are conserved or divergent between parasitoid worms, and how do these traits relate to their respective ecological lifestyles?** This thesis investigated the genetic features encoded in the genomes of both parasitoid taxa to identify similarities and parallels linked to their ecological lifestyles and to establish new model systems.

1.1 The diverse phylum Nematoda

Nematodes, commonly known as roundworms, represent a remarkably diverse group of organisms (De Ley & Blaxter, 2002; Z.-Q. Zhang, 2013) (Figure 2). Nematodes are an important group of metazoans, they thrive in diverse environments worldwide, ranging from polar ice and deep sea to hot deserts (Sapir, 2021). They inhabit nearly all terrestrial and aquatic environments freely, while

also parasitizing the majority of animals and plants. While most nematodes are free-living and feed on bacteria, fungi, or prey on other nematodes, the phylum Nematoda also includes parasites that infect vertebrates (Anderson, 2000), invertebrates (Luo et al., 2023), and plants (Bird, Jones, Opperman, Kikuchi, & Danchin, 2015). Plant parasitic nematodes alone cause approximately USD 173 billion (Elling, 2013) in annual economic damage to global agriculture. Nematode parasites cause harm to both animal and human health. Such parasitic nematodes are an urgent global burden infecting more than a billion people, causing significant economic losses and ultimately impeding socioeconomic advancement (Roeber, Jex, & Gasser, 2013; Elling, 2013; Viney, 2017; Garcia-Bustos, Sleebs, & Gasser, 2019). However, nematodes also serve vital functions in ecosystems, linking different components of soil food webs, influencing plant growth, acting as bioindicators in evaluations of biodiversity and biomonitoring efforts and contributing significantly to nutrient cycling (Gingold, Moens, & Rocha-Olivares, 2013; Van Den Hoogen et al., 2019; Y. Zhang et al., 2020; Devi, 2020; G. W. Yeates & Boag, 2004; P. J. D. Lamshead, 2004). Despite the significance of nematodes as both free-living organisms and parasites affecting humans, livestock, and crops, and notwithstanding over a century of thorough investigation, only about 30,000 species being officially described (Hodda, 2022), while estimates suggest there could be anywhere from half a million to over ten million species (P. Lamshead, 1993; P. J. D. Lamshead & Boucher, 2003; M. Blaxter, 2011).

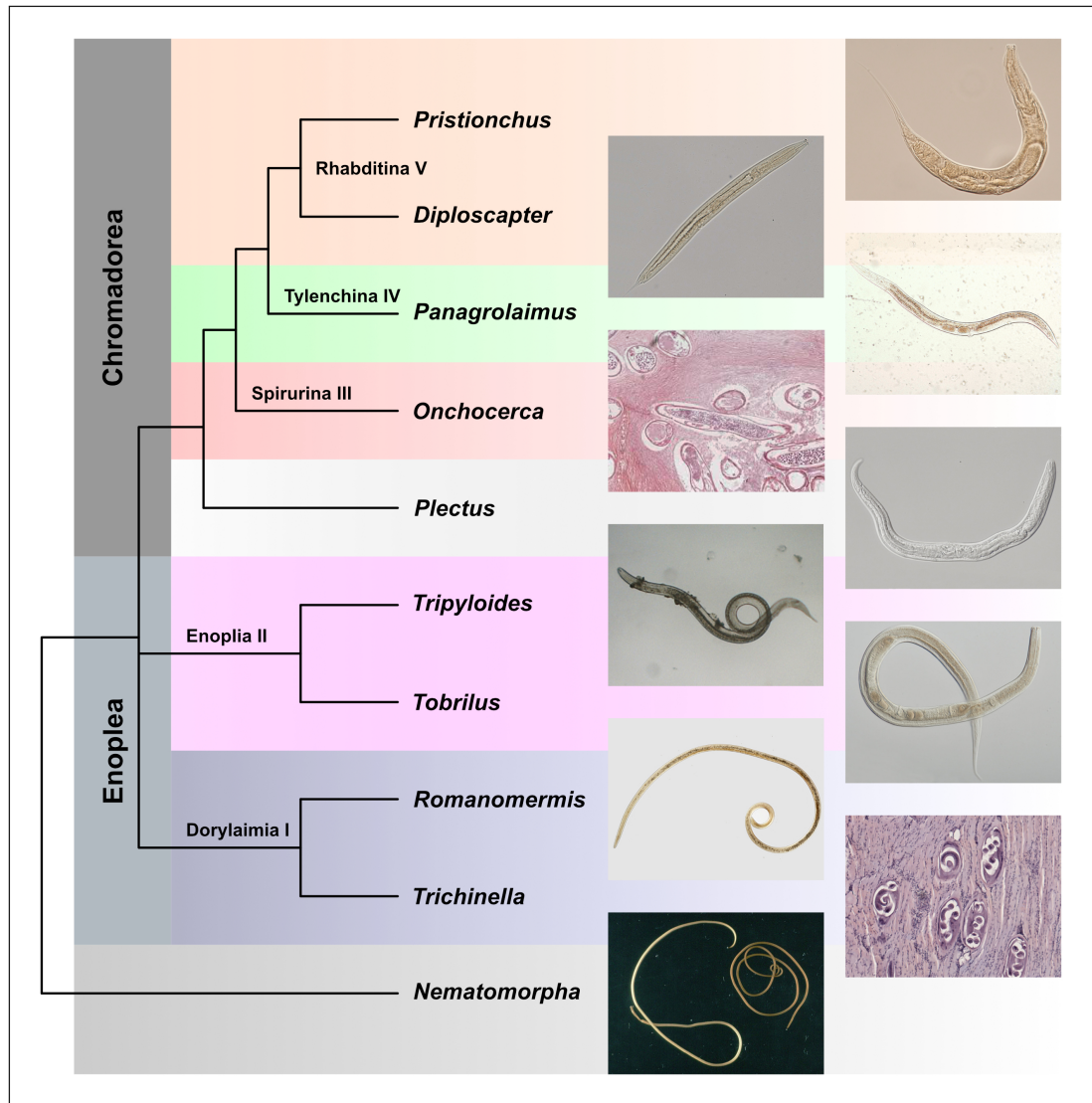


Figure 2: Cladogram of the Phylum Nematoda showing Clades relevant to this work. The clade naming convention was introduced by *Blaxter et al.* in 1998.

Consequently, a significant portion of nematode diversity remains unexplored. "If all the matter in the universe except the nematodes were swept away, our world would still be dimly recognizable, ... We should find its mountains, hills, vales, rivers, lakes, and oceans represented by a thin film of nematodes." This quote from Cobb (Cobb, 1914) refers to the remarkable ubiquity of free-living nematodes on Earth.

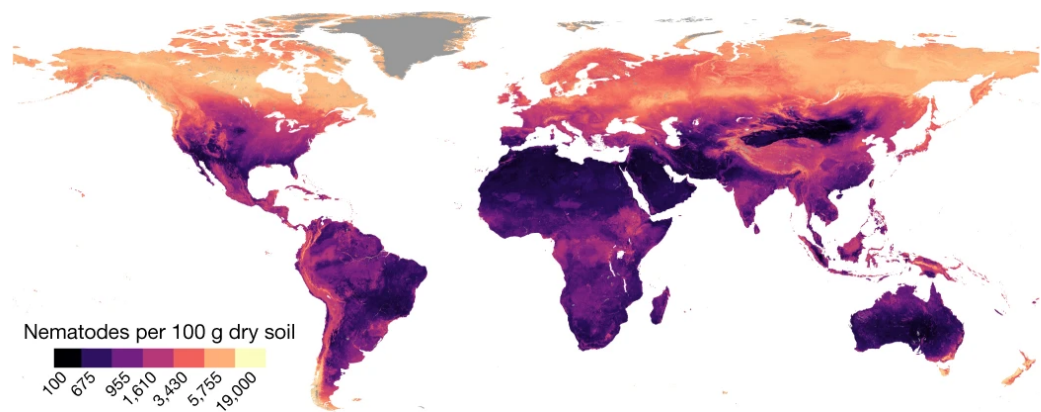


Figure 3: Global map depicting total nematode abundance per unit area (m^2), adjusted for variations in soil bulk density attributed to high organic matter content. Adapted from Johan van den Hoogen (2019) (Van Den Hoogen et al., 2019)

Historically, more research has been conducted on parasitic than free-living nematodes, so the actual diversity of free-living nematodes is presumed to be substantially underestimated (Appeltans et al., 2012). Free-living nematodes do not parasitize animals or plants but live in the space between sediment and soil particles. Despite the ubiquitous presence of free-living nematodes in various limno-terrestrial habitats, including those with unfavorable or extreme conditions that preclude many other invertebrates from living in, they have received less attention compared to their marine and soil-inhabiting counterparts. Freshwater nematodes have been relatively understudied, with only a fraction of their diversity documented in these environments. The subclass Dorylaimia dominates freshwater habitats, while some exclusively freshwater taxa in the subclass Enoplia show extreme levels of endemism. The subclass Chromadoria contains half of all freshwater nematode families, with Monhysterida and Plectida also commonly reported. Research on freshwater nematodes is biased regionally, particularly lacking data from the southern hemisphere compared to European ecosystems. The majority of records come from a single biogeographic region. Discussion on nematode endemism is still premature because, apart from Lake Baikal, the idea of ancient lakes serving as centers of speciation for nematodes is limited. Previous discoveries reveal a high abundance and diversity of nematodes in hidden freshwater bodies, underground calcrete formations, and stromatolite pools, potentially harboring numerous new taxa (Abebe, Decraemer, & De Ley, 2008). Limited attention to comprehensive surveys of freshwater nematodes, compounded by in-

herent biases and uneven sampling, have hindered our understanding of their diversity and evolution. Notably, ancient lakes are crucial as natural laboratories for speciation (Rossiter & Kawanabe, 2000), warranting focused examination of their nematode populations to advance freshwater nematode biogeography.

The evolution of parasitism in nematodes has occurred multiple times independently within free-living groups. Thus, pinpointing a single genetic change or set of changes enabling parasitic exploitation seems unlikely. Instead, a variety of genetic adaptations to parasitism is expected. Present-day nematode genomes contain ancient remnants of their shared lineage and recent markers of their evolutionary paths. Additionally, they may harbor unique, perhaps chance, alterations influenced by current conditions and adaptations. Some nematode groups, whether parasitic or not, may exhibit distinctive genome compositions and structures compared to the rest of the phylum. It would be a mistake to focus solely on genetic adaptations in parasitic species, and comparing parasitic species with distantly related free-living ones could be misleading. Conversely, if members of a parasitic group share a unique genomic trait, it would be insightful to investigate whether this trait is functionally linked to parasitism by considering its presence or absence in other nematode groups (M. L. Blaxter, 2003).

Among the four parasitic nematodes genomes: *Haemonchus contortus*, *Meloidogyne hapla*, *Brugia malayi*, and *Trichinella spiralis* (used for comparison as model species in this thesis), transition to parasitism likely necessitated adaptations of genes inherited from their free-living ancestors (M. L. Blaxter, 2003). For instance, modifications of nutrient-acquisition genes, such as digestive enzymes and secreted hydrolases similar to those found in *C. elegans*, likely played a crucial role in this evolutionary process (Geary & Thompson, 2001). However, surviving immunological attacks, essential for parasitic nematodes, remain a perplexing challenge. Many genes implicated in evading host defenses in nematodes are either secreted or cuticle proteins, since the cuticle serves as the primary site of interaction between the nematode and its environment (Davis, Hussey, & Baum, 2004; Maizels, Blaxter, & Scott, 2001). Notably, secreted glutathione peroxidase (GPX-1), the main soluble surface glycoprotein of filarial nematodes, is hypothesized to contribute to immune evasion (Zvelebil, Tang, Cookson, Selkirk, & Thornton, 1993). Moreover, genes involved in host immune evasion or recognition often undergo rapid amino acid substitutions, indicative of positive selection, as observed in *B. malayi* GPX-1 (Zvelebil et al., 1993).

SCP/TAPS proteins have been suggested to engage in suppressing neutrophil and platelet functions, alongside other activities involved in modulating the immune system during hookworm infections (Del Valle, Jones, Harrison, Chadderdon, & Cappello, 2003). Gene duplication, a process wherein one duplicate

maintains its original function while the other assumes a new role, could serve as a mechanism for the emergence of such genes (V. L. Hunt et al., 2016). For instance, the *alt* gene family found in filarial nematodes, which is associated with initiating infection, possesses only a single ortholog in *C. elegans* (Gomez-Escobar et al., 2002). Additionally, novel genes adapted for parasitism might have arisen through horizontal gene transfer or *de novo*. Plant parasitic nematodes, for example, appear to have acquired "parasitism-related genes" from bacteria in their environment (Bird, Opperman, & Davies, 2003). Furthermore, examining the expression patterns of *C. elegans* orthologs could help identify potential genes involved in parasitism. In *Haemonchus contortus* and *Brugia malayi*, the infective stage of the life cycle corresponds to the third larval (L3) developmental stage (M. L. Blaxter, 2003). Interestingly, in *C. elegans*, the L3 developmental stage represents an alternative pathway, known as the dauer larva, adopted during food scarcity. Thus, identifying orthologs of *C. elegans* genes expressed in the dauer larva stage may offer insights into pinpointing genes involved in infection in *Brugia* and *Haemonchus* (Bürglin, Lobos, & Blaxter, 1998; J. Wang & Kim, 2003).

Progress in molecular biology techniques has facilitated an objective and empirical examination of the evolution of Nematoda. The initial molecular phylogenetic framework of the phylum was established (M. L. Blaxter et al., 1998) by utilizing sequences of the nuclear ribosomal small subunit (SSU) gene. However, the analysis primarily focused on terrestrial and economically significant parasitic species lacking representation from the complete spectrum of taxa found in marine and freshwater environments. Subsequent phylum-wide investigations (Aleshin, Kedrova, Milyutina, Vlad'yehenskaya, & Petrov, 1998; Kampfer, Sturmbauer, & Ott, 1998; Litvaitis, Bates, Hope, & Moens, 2000) introduced additional marine species, yet many prominent clades remained inadequately sampled. Subsequent studies revised the classification of the phylum Nematoda by incorporating molecular data from additional species, supplemented by morphological data to aid in the placement of taxa lacking SSU sequences (De Ley & Blaxter, 2002, 2004). However, the classification system remained primarily grounded in terrestrial and parasitic taxa. Subsequent enhancements of the molecular phylogenetic analysis originally conducted by De Ley and Blaxter (De Ley & Blaxter, 2002) were performed by supplementing the nematode SSU dataset with sequences from marine taxa that were previously not adequately represented, along with additional groups from terrestrial and parasitic environments (Holterman et al., 2006; Meldal et al., 2007; van Megen et al., 2009).

Based on the first phylogenetic analysis using the small subunit RNA genes of nematodes, the phylum Nematoda was divided into three major lineages and five

clades (Figure 2) with each including parasitic species: Dorylaimia (clade I), Enoplia (clade II), and Chromadoria (clades III–V). Within Chromadoria, there are subgroups like Spirurina (clade III), Tylenchina (clade IV), and Rhabditina (clade V), as well as earlier evolutionary branches such as Araeolaimida, Chromadorida, Desmodorida, Monhysterida, and Plectida (M. L. Blaxter et al., 1998). Additional data and alternative phylogeny methods do not fundamentally change the basic topology of the nematode phylogenetic tree. Despite the use of classical methods and partial genomic sequences, which historically may have provided limited resolution, the consistent signal present in genome-scale datasets reinforces our understanding of nematode relationships and offers clearer insights into the phylogenetic resolution of internal Clades (De Ley, 2006; Smythe, Holovachov, & Kocot, 2019; Ahmed et al., 2022; Y.-C. Lee et al., 2023). Both molecular and morphological evidence suggest Enoplean nematodes are most closely related to the common ancestor of nematodes, whereas Chromadorea encompasses species that are more phylogenetically derived (De Ley, 2002; Holterman et al., 2006; Ahmed et al., 2022).

1.1.1 Major Characteristics of Chromadorea

The Chromadoria encompass several lineages that have expanded their habitat range beyond that of most Enoplia and many Dorylaimia. While commonly found in marine sediments, Chromadoria have also thrived in terrestrial environments, including those prone to rapid dehydration and rehydration, such as mosses, lichens, and extremely arid or cold soils. Throughout the Chromadoria, the structure of the cuticle has undergone significant evolutionary changes, sometimes resulting in elaborate decorative patterns. A notable adaptation within the order Rhabditida has been the development of a chemically impermeable cuticle, contributing to their success as parasites, colonizers, and extremophiles. Additionally, the pharynx in Chromadoria exhibits greater structural diversity compared to Enoplia or Dorylaimia, often featuring one or more rounded muscular bulbs that allow for more compact body designs. Free-living Chromadoria tend to be smaller on average than Enoplia and Dorylaimia, with many species specializing in rapid reproduction and bacterial feeding. Several lineages within the Chromadoria have independently evolved curved, swiveling teeth, used for various feeding strategies such as preying on diatomaceous algae or slicing through nematode cuticles. Notably, certain bacterivorous species of Monhysterida and Plectida are the most distant relatives of *C. elegans* capable of being cultured efficiently using methods similar to those for *C. elegans* (Ley & Mundo-Ocampo, 2004).

In the order Rhabditida, a significant development has been the emergence

of a modified juvenile stage specifically adapted for long-term survival. While a true dauer stage is infrequently observed outside of the suborder Rhabditina, which includes *C. elegans*, a precursor to the dauer stage, referred to as a "proto-dauer," is found in several other lineages, including certain panagrolaims and the peculiar genus *Myolaimus* (De Ley, 2006). The emergence of the proto-dauer in the early evolutionary history of Rhabditida likely laid the groundwork for subsequent specialization into the non-feeding, highly modified, and widely dispersing dauer stage characteristic of Rhabditina. True dauers often possess the ability to actively seek out and hitch rides on larger animals, facilitating their dispersal and invasion of the internal organs of other organisms. In Rhabditida, at least three lineages have independently undergone significant evolution leading to major radiations of zooparasitic species. A fourth lineage has not only given rise to zooparasites of animals, but has also diversified into the most varied group of parasites of plants and feeders on fungi among nematodes – the tylenchs. Equipped with a protrusible stomatosylet, which resembles but is distinct from the odontostyle of dorylaims and the onchiostyle of trichodorids, tylenchs have been confirmed through SSU sequences to be closely related to the morphologically dissimilar cephalobs (Siddiqi et al., 1980). As a result, both groups are now classified under the suborder Tylenchina, illustrating another significant change driven by recent advances in phylogenetic understanding. A notable trait supporting the monophyly of Tylenchina is the axis determination in the early zygote, which does not rely on sperm entry in these nematodes, unlike in Rhabditina (Goldstein, Frisse, & Thomas, 1998). Parthenogenesis appears to be more prevalent in Tylenchina compared to hermaphroditism (Goldstein et al., 1998).

Similar to zooparasitic Rhabditida, certain parasitic tylenchs have an infective juvenile stage in their life cycle. However, unlike Rhabditida, most tylenchs and cephalobs do not rely on a single enduring stage for dispersal, but instead, they possess the capability to endure harsh conditions throughout much of their life cycle. This adaptability has enabled them to successfully compete with dorylaims even in extremely dry and cold terrestrial environments. A key factor contributing to their success may be the continual postembryonic expression of traits initially evolved during the proto-dauer stage (De Ley, 2006). As a result, tylenchs and cephalobs exhibit dauer-like characteristics throughout their life cycles. While tylenchs and cephalobs display adaptations for surviving in harsh environments, panagrolaims exhibit a broader spectrum of ecological roles within the Tylenchina lineage. The panagrolaims, constituting the third major lineage within Tylenchina, represent a diverse group comprising free-living opportunists, fermentation specialists, insect pathogens, and animal parasites. While there exist numerous morphological intermediates between panagrolaims and cephalobs,

a definitive molecular resolution is yet to be achieved.

In the suborder Rhabditina, diversity is organized into four main groups: strongylids, diplogasterids, bunonematids, and rhabditids. The rhabditids encompass a variety of species, including *C. elegans*, each differing in morphological details of their male and female reproductive systems. Traditionally classified separately due to their significance as animal parasites and their morphological intricacy as adults, strongylids originated from within the rhabditids, evident from SSU and phylogenomics sequences (Sudhaus & Fitch, 2001; Ahmed et al., 2022), juvenile morphology, and male genital characteristics. Diplogasterids are distinguished by a shift in pharyngeal pumping function to the median bulb, accompanied by muscle and valve reduction in the basal bulb. Although this arrangement appears similar to that of Tylenchina, molecular evidence suggests it is purely convergent (M. L. Blaxter et al., 1998). Notably, diplogasterids encompass diverse organisms such as *Pristionchus pacificus*, known for its dimorphism in feeding and predatory behavior (Von Lieven & Sudhaus, 2000). Diplogasterids are closely associated with species displaying typical rhabditid morphology (Sudhaus & Fitch, 2001), as well as the enigmatic bunonematids (von Lieven, 2002). These nematodes exhibit highly distinctive anatomical features including a 90-degree switch in symmetry of external morphology, compared to typical nematodes' internal bilateral symmetry.

1.1.2 Key Characteristics of Enoplea

Nematodes, found in various habitats and environments, span a wide range of ecological niches, from deep-sea sediments to dry deserts, and include both bacteria-eating forms and parasites with complex life cycles. While many groups within the Nematoda occupy different ecological niches, none of them alone encompasses the full range of habitats found in the phylum, particularly within the Enoplea. This suggests potential constraints on the evolutionary adaptations within each nematode taxon, possibly due to limitations in gene evolution rates, ecophysiological constraints, or competitive exclusion from habitats occupied by other taxa (De Ley, 2006). The non-marine distribution and current diversity of Dorylaimia suggest their potential role as early colonizers of freshwater and terrestrial habitats. While numerous Dorylaimia (Capillariidae) species parasitize marine fish, notably absent among the free-living lineages are species adapted to marine environments, with only a few exceptions discovered in slightly brackish Bothnia Bay.

The early origin hypothesis may explain the significant diversity within this subclass, which includes both free-living species like Doryaimida and Mononchida, as well as notable animal parasites such as Mermithida and Trichinellida (Smythe

et al., 2019; Ahmed et al., 2022). Mermithids exhibit a unique lifecycle trait of leaving the host before reaching adulthood, a characteristic otherwise observed only in the phylum Nematomorpha and some insects. The existence of such distinctive biological features suggests a much greater past diversity within this nematode lineage. Within Dorylaimia, the order Dorylaimida stands out as particularly speciose, comprising large predators/omnivores and the plant-parasitic family Longidoridae, some of which transmit plant viruses. The evolutionary radiation of dorylaims appears to be driven, in part, by the functional diversification of the odontostyle (a needle-like specialized feeding structure used for puncturing and extracting food items) (De Ley, 2006) which potentially correlate with niche partitioning.

The subclass Enoplia exhibit notable diversity in marine habitats (Smythe et al., 2019; Ahmed et al., 2022), with several lineages also found in freshwater sediments and moist soils. Certain lineages within Enoplia demonstrate adaptability across marine, freshwater, and terrestrial environments, suggesting enhanced osmotic tolerance compared to early Dorylaimia. Many Enoplian clades comprise large predators with distinct teeth or hooks, alongside photosensitive structures like eyespots. Enoplia are of particular interest phylogenetically due to the presence of ancestral features within nematodes, such as highly indeterminate modes of development and retention of the nuclear envelope in mature spermatozoa (Schierenberg, 2005). While Enoplia have not extensively colonized extreme terrestrial environments or evolved animal parasitism except for *Echinomermella* and Marimermithidae, the order Triplonchida has undergone significant diversification in soil habitats, including *Trichodorus* nematodes, which are not only plant parasites and virus vectors but exhibit similarities with dorylaims in various aspects, such as possessing a protrusible feeding tooth called an onchiostyle.

1.1.3 Description of nematode body structures

Nematodes are non-segmented animals with usually long, narrow and threadlike bodies (D. Hunt, Palomares-Rius, et al., 2012). Their anatomical arrangement resembles a tube within a tube, wherein the intestinal and gonadal structures are enclosed by the body wall comprising dorsal and ventral longitudinal muscles, epidermis, and a cuticle (Kiontke & Fitch, 2013). Between these inner and outer layers lies a pressurized, fluid-filled pseudocoel used as a hydrostatic skeleton. This configuration facilitates the undulating movement of nematodes in sinusoidal waves while lying on one side. Moreover, this design likely imposes a significant constraint on evolutionary modifications to this basic body plan. While the hydrostatic skeleton is a shared characteristic among all vermiform creatures, nematodes have refined it to a high degree. From this perspective,

nematodes can be viewed as having achieved a streamlined form of perfection. Consequently, compared to other highly successful phyla like arthropods or vertebrates, nematodes exhibit a restricted degree of morphological diversity (Kiontke & Fitch, 2013). Nematodes exhibit a cylindrical, bilaterally symmetrical body structure and possess digestive, nervous, excretory, and reproductive systems, although they lack circulatory and respiratory systems (D. Hunt et al., 2012). Their body is unsegmented and covered by a thick cuticle, providing protection from the external environment and serving as a flexible exoskeleton. Beneath the epidermis lies a layer of longitudinal muscle cells that extends along the entire body, facilitating nematode movement. The oral cavity typically consists of a cuticle-lined capsule, which may feature modifications like bristles, plates, or tooth-like structures. Particularly in parasitic nematodes, including plant parasites, there exists a pointed needle-like stylet used by the organism to puncture its prey. When penetrating host plant tissues, nematodes utilize the hollow stylet, often assisted by the secretion of various proteins. In plant hosts, these proteins may include enzymes that degrade cell walls, while in vertebrate parasitic nematodes, proteases and hyaluronidases play a role in penetrating and moving within host cells during feeding (Perry & Moens, 2013). At the anterior end of the body, nerves emanate from a dense circular nerve ring encircling the pharynx, which functions as a rudimentary brain. Variations in morphological features such as body shape, presence or absence of stylets, position of the vulva, tail configurations, presence of spicules, and cuticular patterns have traditionally served as distinguishing characteristics for identifying nematodes morphologically (Perry & Moens, 2013).

The cuticle is differentiated and is normally moulted four times. Cuticular spines and scales are common and likely serve to allow stability or traction within substrates. The digestive tract runs from a mouth, centered anteriorly, to a subterminal anus. Posterior to the anus is a terminal pore, sometimes opening through a cuticular tube that is supplied with adhesive material by a caudal gland. The mouth opens into a buccal capsule, which is followed by and may be incorporated within a muscular pharynx or oesophagus. The pharynx is triangular and the pharyngeal lumen Y-shaped. Typically, six lobes or labia surround the mouth. Each labium has two sensory sensillae to form inner and outer rings of six labial sensillae each. The sensillae lie in dorsolateral, lateral and ventrolateral positions. Four cephalic sensillae lie outside of the labial sensillary rings in dorsolateral and ventrolateral positions however variations in these arrangements occur. The sensillae function as anterior sense organs and they are served by six dorsolateral, lateral and ventrolateral nerve strands that split so as to supply two or three sensillae each. These strands connect with a nerve ring that encircles

the pharynx. The anterior end of a nematode thus has a triradial or hexaradial symmetry. Nematode excretion involves unique renette cells that connect to a ventral pore that is found near the pharyngeal midgut border. The body-wall muscles form a longitudinal cylinder. Small nematodes are compact and essentially acoelomate. The gut is often lined by the basal laminae of the epithelia but mesoderm cells sometimes occur on the gut and there is not a complete mesodermal tissue sheath (D. Hunt et al., 2012; Perry & Moens, 2013).

1.1.4 Diversity of nematode in embryonic development and reproductive cycles, feeding and survival strategies

Embryonic studies have unveiled distinct developmental features among nematode species and higher taxa, which correlate with their phylogenetic placement (Dolinski, Baldwin, & Thomas, 2001; Skiba & Schierenberg, 1992; Malakhov, 1986; Voronov, 1999; Goldstein, 2001; Schierenberg, 2001; Lahl, Halama, & Schierenberg, 2003). These investigations have not only revealed interspecies variations but also intraspecies differences. For instance, plasticity in pattern formation has been observed in *Acrobelloides nanus* (clade IV) (Skiba & Schierenberg, 1992) and *Diploscapter coronatus* (clade V) (Lahl, Schulze, & Schierenberg, 2009), while *Pristionchus pacificus* (clade V) exhibits postembryonic mouth dimorphism (Bento, Ogawa, & Sommer, 2010; Kiontke & Fitch, 2010). Studies on *A. nanus* (clade IV) have demonstrated a high regulative potential, showcasing a hierarchy of somatic cell fate transformations following early embryo cell ablation (Wiegner & Schierenberg, 1999). Enoplean nematodes present even more striking peculiarities. In clade II (*Enoplus brevis*, *Pontonema vulgaris*), asymmetric cleavages and distinct cell lineages are initially absent, with only a gut lineage present (Malakhov, 1986; Voronov & Panchin, 1998). The development of *Tobrilus diversipapillatus* (clade II) is characterized by a prominent coeloblastula, a developmental trait previously thought to be absent in nematodes (Schierenberg, 2005). Compared to *C. elegans*, *Romanomermis culicivorax* (clade I) exhibits significant differences in embryonic polarity establishment, pattern formation, programming of somatic founders, and cell lineage complexity (Schulze & Schierenberg, 2008, 2009). Thus, the embryonic development of nematodes displays a far greater diversity than what is suggested by the essentially invariant development of *C. elegans* and its close relatives (Schulze & Schierenberg, 2011).

Nematodes display a vast array of reproductive strategies, encompassing both sexual and asexual modes (Singaravelu & Singson, 2011). Sexually reproducing nematodes typically rely on the fusion of sperm and oocytes to form zygotes. Many sexually reproducing nematodes exhibit gonochorism, where distinct male and female individuals are present. In contrast, certain nematodes exhibit andro-

dioecy, where the majority of the population consists of hermaphrodites (capable of retaining self-fertilized eggs within their uterus until hatching) with occasional males, as observed in *C. elegans*. Additionally, some nematodes engage in asexual reproduction known as parthenogenesis, seen in species like *Meloidogyne hapla* (Liu, Thomas, & Williamson, 2007) and *Strongyloides* (Streit, 2008). Parthenogenesis is widespread among species found in shallow water environments or those experiencing recurrent cycles of drying and wetting (Wharton, 2012).

Nematodes typically undergo embryonic development followed by four developmental stages, alternately referred to as larval (L1–L4) or juvenile (J1–J4), separated by moults (D. L. Lee, 2002a). In many parasitic (animal and plant) species first one or two molts can happen inside the egg. In some *Plectus* one of the juvenile stages is apparently lost. However, variations exist, including the presence of alternative forms such as dauer juveniles (Figure 4), capable of enduring extended periods of starvation and adverse conditions. Some stages of the life cycle can be parasitic, with varying minimal generation times and life expectancies (Anderson, 2000; Eckert, Friedhoff, Zahner, & Deplazes, 2005; Gems & Riddle, 2000). Life cycle stages often serve as a mechanism for surviving environmental fluctuations. For instance, different life stages may exhibit specific preferences for food sources (G. Yeates, 1987), and feeding strategies can be adjusted based on the availability of food (Riera & Hubas, 2003). Juveniles of mononchids are likely bacterial feeding because they are too small to be predatory.

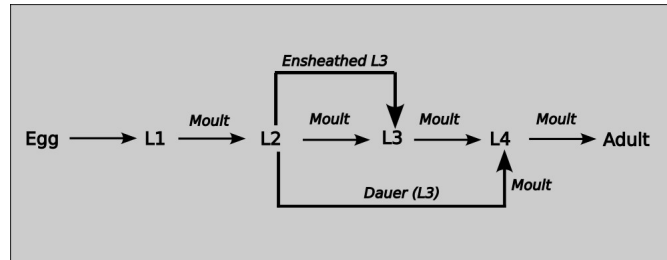


Figure 4: The general life cycle of nematodes. Certain species generate a modified L3 larva known as a resistant dauer larva, while others produce an obligatory ensheathed L3. Both variants undergo developmental arrest and necessitate appropriate stimuli to resume development.

The nematode *Caenorhabditis elegans* is one of the most extensively studied free-living nematodes. Sydney Brenner selected it in 1965 as a model organism for investigating animal development, genetics, and behavior (Brenner, 1973, 1974; Hodgkin & Brenner, 1977; J. Sulston & Hodgkin, 1988; WB, 1988). *C. elegans* was also the first multicellular organism to have its genome fully sequenced and annotated. This microbivorous nematode resides in soil environments in various regions of the world. Its life cycle spans approximately 3 days, and under optimal

conditions, it typically lives for around 2 weeks (Johnson & Wood, 1982; Riddle, Blumenthal, Meyer, & Priess, 1997; D. L. Lee, 2002b). *C. elegans* exhibits two genders: hermaphrodites and males. The adult hermaphrodite, structurally female, reproduces through self-fertilization. However, hermaphrodites cannot fertilize each other (Wood, 1988). Males occur spontaneously and infrequently, capable of fertilizing hermaphrodites. Hermaphrodites are protandrous, with self-produced sperm stored in the spermatheca for fertilizing their oocytes. A self-fertilized hermaphrodite lays approximately 300 eggs during its reproductive lifespan (Wood, 1988). Despite producing more oocytes than sperm, mating with a male enables the hermaphrodite to potentially generate over 1000 offspring (Riddle et al., 1997). Embryogenesis in *C. elegans* occurs rapidly at 25 degrees Celsius, completing in 14 hours. Subsequently, the fertilized egg hatches into the L1 larval stage, progressing through four larval stages to adulthood within 36 hours in favorable conditions (Riddle et al., 1997). When population density increases and food supply diminishes, conditions become unsuitable for successful reproduction. At this point, development halts at the second moult, and larvae enter a dauer stage (Riddle et al., 1997), waiting for the return to favorable conditions to complete its life cycle.

1.1.5 The conserved simple nematode body plan

Even though nematodes possess a conserved and relatively simple body plan, they adapt by changing their physiology and biology, underlined by very diverse genomes. Despite closely related nematode species exhibiting unexpectedly large genetic differences (Riddle et al., 1997), their distinctive body plan seems to have limited the extent of morphological diversification, unlike what is observed in other phyla such as arthropods or chordates. The anatomical structure of nematodes stands as a remarkable outcome of evolutionary processes, showcasing a harmonious integration of form and function that results in architecturally captivating design. Despite seemingly possessing a simple and conserved body plan, nematodes exhibit significant variations in structural patterns. These morphological adaptations likely mirror the diverse ecological roles and the array of environmental conditions within which different species have evolved and adapted for optimal functioning. The conserved nematode body plan is not only a testament to the elegance of evolution but also a reflection of the nuanced interplay between their form, function, and the dynamic ecological landscapes they inhabit.

1.1.6 Nematoda as an important model for investigating developmental evolution

Nematoda has emerged as a valuable model for investigating the evolution of development. The possession of external embryonic development, along with the optical transparency of eggs in many species and the rapid pace of development, generally enables the analysis of cellular processes *in vivo*. However, it's worth noting that these advantages are primarily pronounced in species such as *C. elegans*, rather than being universally applicable. The observed variations in early embryogenesis among nematode species suggest that, while strict constraints influence crucial developmental steps, there is greater adaptability in cellular behaviors leading to these steps (Schulze & Schierenberg, 2011).

Extensive examination of *C. elegans* embryogenesis, particularly concerning asymmetric cell division and axis formation (J. E. Sulston, Schierenberg, White, & Thomson, 1983), initially led to the assumption that early cell lineages and blastomere arrangements closely resembled the well-established model seen in *Ascaris* (Boveri, 1899) and other studied representatives of the phylum (Skiba & Schierenberg, 1992). However, subsequent analyses, considering a broader range of species revealed significant variations in critical steps of early embryogenesis (Goldstein, 2001). Despite these variations being described descriptively, little is known about their molecular or genetic foundations. The intriguing question of why early embryogenesis displays greater variability than later phases prompts two potential explanations: neutral evolution, where variations arise without adaptive value, or the presence of diverse developmental strategies aimed at enhancing fitness. The modular nature of development, often used to explain increasing integration over time, may not fully account for the fast-developing nematodes like those following the *C. elegans* type of development.

To connect ontogeny and phylogeny, embryonic variations are proposed as heuristic phylogenetic markers, offering independent insights beyond traditional parameters like morphology and nucleic acid molecules. Investigations into axis specification (Goldstein et al., 1998), cleavage patterns, blastomere arrangement (Dolinski et al., 2001), germline behavior, and gastrulation (Schierenberg & Lahl, 2004; Schierenberg, 2005) have contributed to tracing the evolutionary trajectory of embryonic diversity in nematodes (Schierenberg, 2006).

1.2 Evolution of nematode genomes

1.2.1 Nematode genome size evolution

The evolution of nematode genomes encompasses a wide spectrum of genome sizes across the phylum Nematoda. Most nematodes possess genomes with sizes ranging from 20 to 500 Mb (Gregory et al., 2007; Kumar, Koutsovoulos, Kaur, & Blaxter, 2012), such as *C. elegans*, *Haemonchus contortus* with a genome size of 53 Mb (Morand, 2003), *Trichinella spiralis* with a size of 64 Mb (Mitreva et al., 2011), *Romanomermis culicivorax* with a size of 300 Mb (Guiglielmoni, Villegas, Kirangwa, & Schiffer, 2024). Conversely, there are exceptions like *Parascaris univalens* with a genome size of approximately 2100 Mb (Niedermaier & Moritz, 2000) and *Bursaphelenchus mucronatus* with a genome of around 30 Mb (Morand, 2003). The factors driving changes in genome size among nematodes are multifaceted. The compact nature of nematode genomes may be attributed to a high rate of large, spontaneous deletions and potential selection for deletions (Witherspoon & Robertson, 2003; Denver, Morris, Lynch, & Thomas, 2004). The genome of *C. briggsae*, for instance, is slightly larger than that of *C. elegans* due to an increased amount of repetitive DNA (Stein et al., 2003). Insights into the relative significance of deletions versus insertions and proliferation of repeats can be gained by comparing genomes of closely related nematodes. Furthermore, genome size may reflect gene count to some extent. For example, despite the difference in size between the genomes of *C. briggsae* and *C. elegans*, they both harbor a similar number of genes. Similarly, the *Brugia malayi* genome, despite its comparable size to *Caenorhabditis* genomes, contains a similar number of genes. On the other hand, the genomes of *S. ratti* and *S. stercoralis*, approximately 43 Mb in size, are relatively small compared to other nematodes. Despite their smaller size, the total protein-coding content of each genome is comparable to that of other nematodes. This discrepancy in genome size is attributed to the significant loss of introns and shorter intergenic regions (V. L. Hunt et al., 2016).

1.2.2 Evolution of Nematode chromosome number and size

Nematode genome evolution also involves changes in chromosome number and structure. Most nematodes have haploid chromosome numbers of $n=4-12$ (Walton et al., 1959; Coghlan, 2005), with significant variation observed. Some species stand out as having the lowest haploid number ($n=1$). The holocentric characteristic of many nematode chromosomes could promote the formation of stable chromosome fusions. This is exemplified by a select few nematode species, including *Diploscapter pachys*, *Diploscapter coronatus*, and *Parascaris univalens*, each of

which possesses only a single pair of chromosomes (Niedermaier & Moritz, 2000; Fradin et al., 2017; Hiraki et al., 2017). Alternatively, the polyploid species in the Tylenchomorpha exhibit high chromosome counts. Numerous Tylenchomorphs, such as *Meloidogyne*, exhibit parthenogenesis, a reproductive mode where unfertilized eggs give rise to offspring. Species utilizing this reproductive strategy appear to be prone to polyploidization (Otto & Whitton, 2000). The significant diversity in chromosome numbers and the corresponding alterations in reproductive strategies suggest that cytogenetic evolution played a pivotal part in the speciation of root-knot nematodes and plant-parasitic nematodes overall (Janssen, Karssen, Topalović, Coyne, & Bert, 2017). The majority of rhabditines typically exhibit a haploid chromosome count of 5 to 6 (M. Blaxter, 2000). For instance, both *C. elegans* and *C. briggsae* have a haploid chromosome number of 6, despite having experienced approximately 4000 rearrangements in their chromosomes since their evolutionary divergence (Stein et al., 2003). The absence of fissions or fusions implies that there might be a preference for maintaining a stable chromosome count within the Rhabditina (Coghlan, 2005). The *C. elegans* and *C. briggsae* genomes share conserved segments, suggesting ancient linkage groups (Stein et al., 2003). Different evolutionary patterns exist between the arms and centers of nematode chromosomes, with arms evolving more rapidly due to higher recombination rates and mutations (Barnes, Kohara, Coulson, & Hekimi, 1995; elegans Sequencing Consortium*, 1998). Chromosome organization also influences gene regulation, with essential genes and operons concentrated in chromosome centers (Blumenthal et al., 2002; Kamath et al., 2003). There seems to be a discernible organization of *C. elegans* genes across chromosomes based on their functional roles. Essential genes tend to be predominantly located in the central regions of chromosomes I, II, and III, whereas genes associated with worm behavior and morphology are clustered on chromosomes II and X (Kamath et al., 2003). This chromosomal arrangement potentially facilitates the co-regulation of genes within shared chromatin domains (Kamath et al., 2003). Furthermore, the presence of operons and co-expressed gene clusters contributes to chromosomal evolution. Operons stabilize gene order and are concentrated in chromosome centers, influencing the rate of rearrangements (Blumenthal et al., 2002).

1.2.3 The evolution of gene content in nematodes

The evolution of gene content in nematodes reveals fascinating insights into their genomic diversity and adaptive strategies. The sequencing ESTs from 30 nematode species resulted in the identification of approximately 94,000 genes across approximately 60,000 families (Parkinson et al., 2004). Only around 15% of approximately 94,000 genes are universally conserved across all four studied ne-

matode clades (Rhabditina, Tylenchina, Spirurina, Dorylaimia), suggesting their involvement in core metabolic or structural pathways. Additionally, about 1300 genes are nematode-specific yet widespread among nematodes, potentially playing crucial roles in body plan development and evolutionary history (Parkinson et al., 2004). The nematode gene space likely encompasses a much larger gene repertoire than the estimated 60,000 gene families. Firstly, it's probable that numerous genes were overlooked because of expression at low levels and restricted expression in specialized cells, which are challenging to detect through an EST survey. Additionally, only 30 species were examined out of the currently known approximately 27,000 nematode species. Moreover, the sampled species did not include marine and freshwater nematodes, which means their specific proteomes were substantially underrepresented. Considering that the first nematode is believed to have been a marine organism, possibly an enoplid, it's plausible that the oceans harbor even greater diversity in gene content (Poinar, 1983; M. L. Blaxter, 2003; Coghlan, 2005).

Gene families consist of genes that share similar structures and functions, and they can expand or contract over evolutionary periods, which may influence species variations (Henikoff et al., 1997; Mushegian, Garey, Martin, & Liu, 1998). Therefore, the examination of gene family expansions and contractions constitutes a crucial aspect of genomic investigation (M.-Y. Chen et al., 2021). Gene family proliferation and loss significantly contribute to the evolutionary trajectory of nematode genomes. Notably, following its divergence from *C. briggsae*, *C. elegans* has exhibited a notable proliferation of chemoreceptors (Stein et al., 2003). This expansion of chemoreceptor genes in *C. elegans* suggests a potential adaptation to a distinct ecological niche. Chemoreceptors play crucial roles in detecting environmental cues, aiding in locating food sources, and avoiding predators, pathogens, and toxins (Troemel, 1999). The duplication and divergence of additional chemoreceptor genes in *C. elegans* likely facilitated its ability to exploit new ecological niches compared to its relative, *C. briggsae*. Conversely, *C. elegans* has also lost several genes present in other nematodes, such as a DNA methyltransferase gene found in *P. pacificus* (Gutierrez & Sommer, 2004). Initially, Dnmt2 orthologs were presumed to function as DNA methylating enzymes due to their significant sequence similarity with enzymes classified under the "Dnmt" family, known for their role in catalyzing cytidine methylation at the carbon-5 position, resulting in the formation of 5-methylcytosine (5mC) in DNA. Despite this, experiments using Dnmt2 preparations failed to demonstrate robust DNA methyltransferase activity comparable to other members of the Dnmt family. Through an extensive investigation into target DNA sequences across various genomes, Dnmt2 enzymes were shown to function as tRNA transferases (Goll et

al., 2006), catalyzing the methylation of position 38 in tRNA^{Asp}GUC⁽¹⁾ to produce 5-methylcytosine commonly abbreviated as m⁵C in the RNA modification field (Jeltsch et al., 2017).

The model *C. elegans* possesses numerous species-specific genes, with about 1000 genes lacking homologs in *C. briggsae* (Stein et al., 2003). These genes may have rapidly diverged or arisen *de novo* via chromosomal rearrangements (Long, 2001). Interestingly, a survey of ESTs from various nematode species suggests that 30–50% of genes in each species may be species-specific, indicating extensive genomic diversity (Parkinson et al., 2004). Duplication events, chromosomal reorganizations, and the activity of transposable elements have been identified as contributing factors in the emergence of new genes (Betrán & Long, 2002; Ganko, Bhattacharjee, Schliekelman, & McDonald, 2003; Long, 2001).

Although *C. elegans* serves as a superb model nematode and its genome, enriched with extensive annotations, it's important to note that *C. elegans* cannot fully represent all nematode genomes. This species is recognized as significantly derived within the Nematoda phylum (M. L. Blaxter et al., 1998), and it does not possess numerous genes shared among other nematodes and other organisms within the Metazoa group (Wasmuth, Schmid, Hedley, & Blaxter, 2008). For instance, *C. elegans* possess unusual genomic characteristics, such as a reduced Hox cluster, the absence of a Hedgehog signaling system and the loss TRDMT1 tRNA aspartic acid methyltransferase 1 gene (Gutierrez & Sommer, 2004). Additionally, nematode genomes exhibit a wide range of sizes, spanning from 20 megabases to 500 megabases (Gregory et al., 2007). Sequenced nematode genomes vary significantly, with examples including *Meloidogyne hapla* at 54 megabases and *Ascaris suum* at 273 megabases (Opperman et al., 2008; Jex et al., 2011). Moreover, intriguing genomic characteristics have been observed in other species. These include chromatin diminution in *Ascaris suum* and other ascaridids, where the germline possesses a larger genome than the soma (Müller, Bernard, & Tobler, 1996), aneuploid triploidy in the *Meloidogyne incognita* genome (Abad et al., 2008), and the presence of obligate, vertically-transmitted symbiont alphaproteobacterial *Wolbachia* and their genomes within the cells of numerous filarial nematodes (Fenn et al., 2006).

1.2.4 The evolution of HOX genes

Hox genes stand out in evolutionary-developmental biology for their profound influence, particularly in shaping body plans and animal development through patterning the formation of the anterior-posterior axis (Akam et al., 1994; Akam, 1998, 2000). They produce regulatory proteins that bind to DNA and share a common 60 amino acid domain called the homeodomain. Typically, these genes

are grouped in a small genomic segment and are present in most animal species. Through comparing HOX genes across different phyla, previous research identified eight groups of orthologues, which are genes with similar functions inherited from a common ancestor. These orthologue groups regulate the anterior-posterior patterning along the body axis during early embryo development, with their arrangement on the chromosome reflecting this patterning (De Rosa et al., 1999). For instance, genes like *hox-1/labial* orthologues control the formation of anterior structures and are located at one end of the HOX cluster, while genes like *AbdB/hox-9-13* orthologues control the development of posterior structures and are positioned at the opposite end of the cluster. This arrangement, known as a linkage group, has remained remarkably consistent throughout the evolution of multicellular animals. The presence of HOX genes in unsegmented animals like nematodes, which lack clear evidence of segmental patterning, initially posed a surprise given their traditional association with segmented or metameric structures. However, it appears that HOX genes serve as foundational molecular structures shaping the body plan across all animal species. In animals considered early branching on the evolutionary scale, such as nematodes, these genes are thought to contribute to anterior-posterior patterning, although not in the context of repeated segments (M. L. Blaxter, 2003).

Hox genes are crucial regulators of body patterning and development in bilaterian animals, controlling the identity and position of body structures along the anterior-posterior axis (Slack, Holland, & Graham, 1993). These body patterning genes were initially discovered in *Drosophila melanogaster* fruit flies through the striking phenotypes they gave rise to when genes were mutated (Lewis, 1978). In these homeotic phenotypes, one part of the body along the anterior-posterior axis is transformed into the likeness of another part of the body, such as Antennapedia mutations that transform fly antennae into legs (Struhl, 1981). In many bilaterian animals, these genes are organized into clusters in the genome, with the order of the genes reflecting their spatial expression during embryonic development (called collinearity) (Lewis, 1978). The last common ancestor of protostomes and deuterostomes is thought to have had a cluster of at least seven Hox genes characterized by a common transcriptional orientation and by collinearity in the order of the genes and their expression domains along the anterior-posterior axis (Garcia-Fernandez, 2005; Hejnol & Dunn, 2016; Gaunt, 2022).

Vertebrate Hox gene clusters have a distinctive structural and transcriptional organisation. This is true of all of the Hox cluster copies present following the whole genome duplications in vertebrates (Holland, Garcia-Fernández, Williams, & Sidow, 1994; Putnam et al., 2008). Despite the generally larger size of vertebrate genomes, vertebrate Hox genes are very densely packed and Hox cluster

span is much smaller than in many invertebrates. The clusters have low repeat densities, and while protein-coding genes unrelated to the Hox family are excluded, small RNA genes that regulate cluster expression can be interspersed between Hox genes. Indeed, clustering is strongly associated with the mechanisms of cross- and co-regulation of Hox gene expression (Wagner, Amemiya, & Ruddle, 2003; Darbellay et al., 2019).

Clustering and collinearity are not always conserved. Numerous animal taxa have managed to escape the constraint of having a tightly regulated, compact and complete Hox cluster. In several bilaterian phyla, Hox genes have undergone significant evolutionary changes including gene duplication, loss, and gain as well as changes in cluster organization and expression patterns (Ferrier & Minguillon, 2003). Some taxa have Hox clusters that are less tightly organized but still intact when compared to vertebrate counterparts, including the cephalochordates (*Branchiostoma*) (Minguillón et al., 2005), sea urchins (Cameron et al., 2006), and insects (*Apis* (Dearden et al., 2006; Consortium et al., 2006; Negre & Ruiz, 2007), *Anopheles* (Holt et al., 2002; Powers et al., 2000) and *Tribolium castaneum* (Kulakova, 2018; Beeman, 1987)). The Hox cluster in *Drosophila melanogaster* is split into two subclusters (Kaufman, Lewis, & Wakimoto, 1980). Splitting has occurred repeatedly and independently in other drosophilids, at different positions along the cluster (Negre & Ruiz, 2007; Mulhair et al., 2023). Disaggregation of the Hox cluster can be observed in tunicate chordates *Ciona intestinalis* (solitary tunicate) and *Oikopleura dioica* (appendicularian) where Hox loci are present but not tightly linked in the genome (Ikuta, Yoshida, Satoh, & Saiga, 2004; Seo et al., 2004).

By comparing the complements and developmental roles of Hox genes in different groups, it is possible to propose models of how the evolution of these genes has influenced the morphological diversity within various metazoan phyla. Loss of Hox genes can lead to body plan simplification while gene duplication can result in the acquisition of novel developmental processes (Van Auken, Weaver, Edgar, & Wood, 2000; F. W. Smith et al., 2016). Indeed the body plan of an animal can be a consequence of Hox gene loss as thought for the tardigrade *Hypsibius exemplaris* (Aboobaker & Blaxter, 2010; F. W. Smith et al., 2016). Hox gene loss is a common feature in Tardigrada, where a reduced complement of seven HOX genes including a local duplication of Hox9-13 was found (Yoshida et al., 2017).

Nematomorphs are the closest relatives to Nematoda and are all parasitoids in arthropods. Both groups have relatively simple and vermiform body plan. A previous survey of Hox gene content using both PCR surveys and transcriptome analysis in *Paragordius* revealed a Hox gene complement of five ancestral orthologue groups. The presence of Hox2 (Pb) in *Paragordius* yet lost in the

nematode counterparts suggested the loss in the common ancestor of nematodes (Aboobaker & Blaxter, 2003; Yoshida et al., 2017).

In nematodes, HOX genes have always been a special case since early work in the model worm *C. elegans*, and later on in the satellite model *P. pacificus*, showed a lack of some of the canonical, bilaterian genes, and a loss of the clustering and colinearity in expression along the anterior-posterior body axis. Past research has focused on establishing whether or not the *C. elegans* HOX complement is archetypical for the phylum. While more HOX genes were found in some species, the lack of fully sequenced and chromosomally assembled genomes across the tree impeded any comprehensive analysis of HOX gene evolution across the phylum. The *C. elegans* Hox gene set has been shown to differ substantially from counterparts in other animal phyla. The six Hox genes include representatives of only four core Hox orthology groups. The posterior pair consists of *nob-1* (also known as "no backend") and *php-1* (referred to as "posterior HOX paralogue"), which belong to the posterior hox9-10/AbdB-like group. The central pair includes *egl-5* (abbreviated for "egg-laying defective," another gene similar to AbdB-like posterior genes) and *mab-5* (abbreviated for "male abnormal," belonging to the hox6-8/ftz/Lox5 group). The anterior pair comprises *ceh-13* (known as "*C. elegans* homeobox gene," an orthologue of *labial/hox-1*) and *lin-39* (referred to as "lineage defect," an orthologue of *Scr/hox-4*). The genes are linked but largely unclustered: they are distributed over an approximately >4 Mb span of *C. elegans* chromosome III, with up to 44 unrelated genes interspersed (Van Auken et al., 2000; Gutierrez, Knoch, Witte, & Sommer, 2003). While there is some residual collinearity, the HOX1 and HOX4 orthologues of *ceh-13* and *lin-39* are inverted compared to the relative order of HOX6 and HOX9. In recent years it has become evident that many features of nematode embryonic development (including timings of cell division, cell lineage, and gene expression) are highly variable across the phylum (Schulze & Schierenberg, 2011; Schiffer et al., 2018), and, like nematode genomes, evolve rapidly (Schiffer et al., 2019; Stevens et al., 2019; Aboobaker & Blaxter, 2003; Kumar et al., 2012). While Hox gene evolution has been well studied in *C. elegans* and a taxonomically biased sample of parasitic species from both Clade I and III (Aboobaker & Blaxter, 2003), our knowledge of Hox gene repertoire, Hox Cluster organisation and linkage in other nematode lineages, representing most of the genomic diversity of the phylum, is limited. Determination of the full complement of Hox genes in recently available nematode genomes across the Nematoda is fundamental for understanding the link between the genomic and morphological evolution of this phylum.

1.2.5 The evolution of critical genes for the DNA methylation machinery

Epigenetic modifications regulate gene expression without changing the underlying DNA sequence. DNA methylation serves as a significant epigenetic marker in numerous eukaryotic organisms. It involves the chemical modification of genomic DNA, accomplished by adding a methyl group to specific nucleotide bases. In animals, DNA methylation commonly takes place at CG sites, initiated by the enzyme DNA methyltransferase 3 (DNMT3) and sustained by the maintenance methyltransferase DNMT1 (Goll & Bestor, 2005; X. Cheng & Blumenthal, 2008; Lyko, 2018). In vertebrates, DNMT3 primarily functions during embryonic development (Z. D. Smith & Meissner, 2013). However, the traditional notion of distinct roles for DNMT enzymes has been questioned, especially the idea of clear task separation (Jeltsch & Jurkowska, 2014; Lyko, 2018). DNMT1 in addition to its role in maintaining DNA methylation might possess alternative functions beyond DNA methylation regulation, although these roles remain largely unexplored (Yarychivska, Shahabuddin, Comfort, Boulard, & Bestor, 2018; Schulz et al., 2018). Investigating these alternative functions has proven challenging, particularly due to the severe repercussions associated with DNMT1 or DNMT3 knock-outs in human embryonic stem cells or mouse embryos, such as cell death or embryonic lethality (Liao et al., 2015). Similar to DNMT1 and DNMT3, DNMT2 is a methyltransferase enzyme that acts on tRNA^{Asp} DNA, but it does not play a role in shaping the DNA methylome (Goll et al., 2006). Previously, DNMT2 was thought to function as a DNA methyltransferase; however, subsequent research revealed its substrate specificity for tRNAs. Specifically, DNMT2 methylates cytosine C38 of tRNA(Asp) in humans, classifying it as an RNA methyltransferase rather than a DNA methyltransferase (Goll et al., 2006). Therefore, DNMT2, though being the most highly phylogenetically conserved of any of the DNMT enzymes, is relatively poorly characterized and is thought to play a larger role in RNA processing (Schaefer & Lyko, 2010).

Epigenetic inheritance in mammals relies partially on the consistent transmission of DNA methylation patterns during development. Previous research demonstrated the essential role of the protein UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1) in preserving DNA methylation. Throughout the S phase, UHRF1 co-localizes with the maintenance DNA methyltransferase protein DNMT1. This co-localization suggests that UHRF1 directly interacts with DNMT1, thereby anchoring it to chromatin. Additionally, UHRF1 possesses a methyl DNA binding domain, known as the SRA (SET and RING associated) domain, which exhibits a strong preference for binding to hemimethylated CG

sites – the natural substrate for DNMT1. These findings imply that UHRF1 may play a role in recruiting DNMT1 to hemimethylated DNA, facilitating the accurate maintenance of DNA methylation (Bostick et al., 2007; Ren, Gao, & Song, 2018).

Numerous research endeavors have explored the uneven spread of DNA methylation within insect lineages (Field, Lyko, Mandrioli, & Prantera, 2004; Bewick, Vogel, Moore, & Schmitz, 2017; Provataris, Meusemann, Niehuis, Grath, & Misof, 2018), as well as in other arthropods (de Mendoza, Pflueger, & Lister, 2019; Gatzmann et al., 2018). These investigations revealed that the majority of insect orders have retained varying levels of DNA methylation (Engelhardt, Scheer, Stadler, & Prohaska, 2022). Additionally, while DNA methylation is commonly observed in most non-bilaterian metazoans, exceptions such as placozoans have been noted, emphasizing the complexity of epigenetic regulation across different evolutionary lineages (de Mendoza, Hatleberg, et al., 2019; X. Xu et al., 2019). The ancestral lineage leading to *C. elegans* underwent a loss of DNMT1 and DNMT3 approximately 350 million years ago (Rošić et al., 2018). Notably, in the nematode lineage leading to the mammalian parasites *Trichinella spiralis* and *Trichuris muris*, DNMT3 is retained while DNMT1 is lost, a configuration uncommon among eukaryotes. However, investigation into the occurrence across a broad range of recently available nematode and nematomorph species genomes is limited. This knowledge gap hinders our ability to fully understand the evolutionary dynamics and functional implications of DNA methylation in these organisms.

1.2.6 The evolution of gene content in parasitic nematodes

Our thorough understanding of the biology of the model organism *Caenorhabditis elegans* could lead to misconceptions, as we might forget that it represents just one nematode species. Nematodes exhibit remarkable success in both parasitic and free-living forms. Their wide range of parasitic behaviors and the variety of hosts they infect indicate their inherent tendency toward parasitism and their ability to adapt to diverse, often harsh, environments. The study of their genes and genomes has been significantly advanced by the research on *C. elegans*. Nonetheless, possessing the whole genome sequence of this free-living nematode does not directly reveal genes associated with parasitism; rather, each nematode genome is a complex combination of shared traits and unique evolutionary developments. Expansion of genome sequencing for parasitic nematodes including species from throughout the phylum, facilitates the identification of groups of genes that show evolutionary trends, hinting at roles related to parasitism—particularly those involved in nutrient uptake in new hosts and in circumventing the immune re-

sponses of these hosts. However, within the phylum Nematoda, the subgroup of parasitoid nematodes, although sharing similarities with parasitic nematodes, possess distinct characteristics in their life cycle and interactions with hosts. Unlike typical parasites, parasitoids often kill their hosts (Godfray, 1994), which makes their study particularly intriguing from an evolutionary standpoint.

Mermithid nematodes exhibit a parasitic lifestyle during their larval stage within invertebrates, while the adult stage is characterized by a free-living, non-feeding existence. Mermithids (a family of the phylum Nematoda) and Nematomorphs (a phylum called horsehair worms or hairworms) represent two distinct groups of worms that exhibit intriguing resemblances in their parasitoid lifestyles and the alterations in behavior they induce in their definitive hosts. Parasitoids are a subgroup of parasites that "develop on or within another organism (=host), derive nourishment from it, and ultimately cause its demise during or at the end of this development" (modified after (Eggleton & Gaston, 1990; Maggenti, Maggenti, & Gardner, 2005)). The parasitoid lifestyle is better known in insects (Eggleton & Belshaw, 1992) but there are important biological differences between them and parasitoid worms. In insects, the free-living adult feeds on a different source of nutrients and occupies a different ecological niche, compared to its parasitoid larvae. Thus, for parasitoid insects, killing the host imposes no cost on the immediate survival of individuals. The biology of parasitoid nematodes and nematomorphs is different. Adult mermithids and nematomorphs, as well as their invasive juveniles, are free-living and do not feed at all, overcoming the disadvantage of killing the host. The parasitoid juveniles of both mermithids and nematomorphs source their nutrition from a variety of arthropod species in a similar way (de Valdez, 2005; Mishina et al., 2023; Sato et al., 2012). The cuticle of parasitoid juveniles undergoes moulting within the host, and the epidermis endures a certain degree of degeneration (Schmidt-Rhaesa, 2005), allowing for transcuticular feeding by the haemocoel-dwelling parasitoid juvenile (Rutherford, Webster, & Barlow, 1977; Skaling & MacKinnon, 1988; Poinar Jr & Hess, 1977), while its digestive system reduces in complexity (Schmidt-Rhaesa, 2005). Accumulated nutrients are stored in the trophosome, a modified intestine, in Mermithidae and parenchyma in Nematomorpha. This energy store allows free-living non-feeding adults to engage in mating and egg-laying for a long period, producing huge numbers of progeny, after emerging from the host (Hanelt, Thomas, & Schmidt-Rhaesa, 2005; Nickle, 1972). Some Mermithidae and Nematomorpha are known to manipulate the behavior of their hosts by inducing water-seeking in terrestrial arthropods that harbor those worm species which complete their development in the aquatic environment (Maeyama, Terayama, & Matsumoto, 1994; Vance, 1996; Poulin, 1998; Thomas et al., 2002; Ponton et al., 2011; Al-

lahverdipour et al., 2019; R. E. H. Herbison, Evans, Doherty, & Poulin, 2019). In nematomorphs, this manipulation may involve alterations in neurotransmitter levels and the regulation of the host's nervous system through parasite-produced molecules like serotonin, dopamine, and octopamine (in *Paragordius tricuspidatus*) (Thomas et al., 2003) and Wnt family proteins (in *Paragordius tricuspidatus* and *Spinochordodes tellinii*) (D. Biron et al., 2005, 2006). In mermithids, the species *Thaumamermis zealandica* induces an increase in haemolymph osmolality in its semi-terrestrial crustacean host *Talorchestia quoyana* – this alteration may create a requirement for water in the host, providing a potential explanation for the observed water-seeking behavior (Williams, Poulin, & Sinclair, 2004). Another mermithid species, *Mermis nigrescens*, induces erratic or hyperactive behavior in the European earwig *Forficula auricularia*, potentially through initial modulation of Mtp α (Mitochondrial trifunctional protein α subunit) and subsequent manipulation of neuronal connections and neurotransmitters, ultimately leading to a hydrophilic state. Clathrin emerges as a key protein in this manipulation process (R. Herbison et al., 2019). Regulation in host brains, coupled with predominant proteomic changes in hosts infected by large worms, and shared functional characteristics of proteins regulated in mermithid and hairworm-infected hosts, especially in axon or dendrite and synapse modulation, strongly support the notion of adaptive manipulation of host behavior.

Substantial differences between mermithids and nematomorphs relate to their free-living stages. The infective juveniles of mermithids are actively seeking their host and use digestive enzymes to penetrate the body wall and make their way to the haemocoel (Shamseldean & Platzer, 1989), whereas the larvae of nematomorphs are semi-sessile and rely on the intermediate paratenic host to reach the definitive host (Hanelt et al., 2005). Mermithid adults are armed with a variety of sensory structures and likely engage them in host-finding (Robinson, Baker, & Heald, 1990), mate-finding and oviposition (Jay Burr & Babinszki, 1990; Burr, Wagar, & Sidhu, 2000). Our knowledge of the receptors present in Nematomorpha is limited – existing ultrastructural studies provide minimal insights into their functionality and role (Schmidt-Rhaesa, 2004; Sokolova, Zograf, & Yushin, 2022). From the phylogenetic point of view, the adoption of a parasitoid lifestyle in all extant species of Nematomorpha suggests that this characteristic likely originated early in the evolutionary history of the group. Within the phylum Nematoda, a parasitoid lifestyle is not unique and originated independently in four different lineages: Mermithidae, Marimermithidae, Benthimermithidae and *Neocamacolaimus* (Miljutin, 2013; Tchesunov et al., 2023; Holovachov & Boström, 2014) but only Mermithidae were evolutionary successful, with about one thousand known species infesting a broad variety of hosts.

Given the significant representation of parasitic organisms among diverse species, their emergence across various phyla underscores a vast biological spectrum (Poulin, 2011). Notably, taxa such as mermithids and nematomorphs exhibit convergence in functionality and phenotype. We investigate how this convergence is demonstrated at the genomic level. Genomic resources for both Nematomorphs and Mermithids are limited. The sizes of a few published genomes range from 250-500 Mb and the genomes overall exhibit a reduced gene content when compared to metazoan orthologs. (Cunha, de Medeiros, Lord, Sørensen, & Giribet, 2023; Eleftheriadi et al., 2023; Guiglielmoni, Villegas, Kirangwa, & Schiffer, 2023; Bhattarai, Poulin, Gemmell, & Dowle, 2022; Kirangwa et al., 2023). A comprehensive comparative analysis of these genomes within an evolutionary framework remains an unexplored aspect.

1.3 Neglected lineages in Nematoda

Nematodes have played crucial roles in research endeavors spanning developmental biology (J. E. Sulston et al., 1983), genome biology (elegans Sequencing Consortium*, 1998; Stein et al., 2003), evolutionary genomics (Cutter, Dey, & Murray, 2009), neurobiology (White, Southgate, Thomson, Brenner, et al., 1986), aging (Crittenden et al., 2003), health (M. C. Wang, O'Rourke, & Ruvkun, 2008), and parasitology (Brooker, 2010). The blueprint of an organism is its native DNA. It places the respective species within the context of the Tree of Life and helps to identify an individual as a unique organism. A large fraction of free-living nematodes from terrestrial and aquatic environments remains poorly studied or even unknown at the genome scale. This bias is obvious and widely recognized in previous nematode genome research studies (Kumar et al., 2012; Mitreva et al., 2006; Smythe et al., 2019; "Comparative genomics of the major parasitic worms", 2019; Parkinson et al., 2004), which have focused on mainly nematode parasites, the model nematode *C. elegans* and other closely related free-living bacterial feeders. The challenge with this approach where only specific groups are continuously studied outside of broader phylogenetic context is that it leads to a limited view of species-specific nematode biology while various free-living nematode species are unique in terms of morphology (Bogale, Baniya, & DiGennaro, 2020), local adaptations to various ecological lifestyles (Schratzberger, Holterman, van Oevelen, & Helder, 2019), genomic disparity (De Ley, 2006), reproductive modes (Schiffer et al., 2019) but also in their embryonic developmental patterns (Schierenberg, 2005). The enhanced transcriptome sequencing of free-living nematodes improved the resolution of deep nematode lineages, despite that the dataset remained heavily skewed towards parasites (Smythe et al.,

2019) and lacked various free-living Enoplean nematodes. The neglect was partly due to many free-living nematodes in marine and freshwater environments being extremely hard to culture under laboratory conditions, which was needed to generate enough biological material for the current state-of-the-art long-read sequencing methods. Pooling a few individuals can potentially increase both haplotype diversity and overall heterozygosity which would make genome assembly challenging (Adams et al., 2020). To overcome these challenges, it is important to establish laboratory cultivation techniques for non-model nematodes spanning the entire Nematoda phylum. Furthermore, the utilization of ultra-low input sequencing methods, as exemplified in recent studies by (Schneider et al., 2021; Priest, Orellana, Huettel, Fuchs, & Amann, 2021), which required as little as 5 ng of input DNA, along with the adoption of low-input High-Molecular-Weight DNA extraction protocols tailored for long-read sequencing, are crucial measures for enhancing high-quality genomic sampling across the entire phylum. **High-quality genomic resources from early-branching Enoplea have been enhanced through the work of this thesis.**

2 METHODS

2.1 Establishment of high-quality genomic resources

The process of establishing high-quality genomic resources for understudied nematode species in early-branching Nematoda involves several crucial steps, from nematode sampling to genome assembly and annotation. Here, I describe the methods including sampling and preservation, morphological and molecular identification of key nematode species (Figure 5), isolation of High Molecular Weight DNA, genome assembly methods, and downstream analyses used to address the research questions outlined in this thesis.

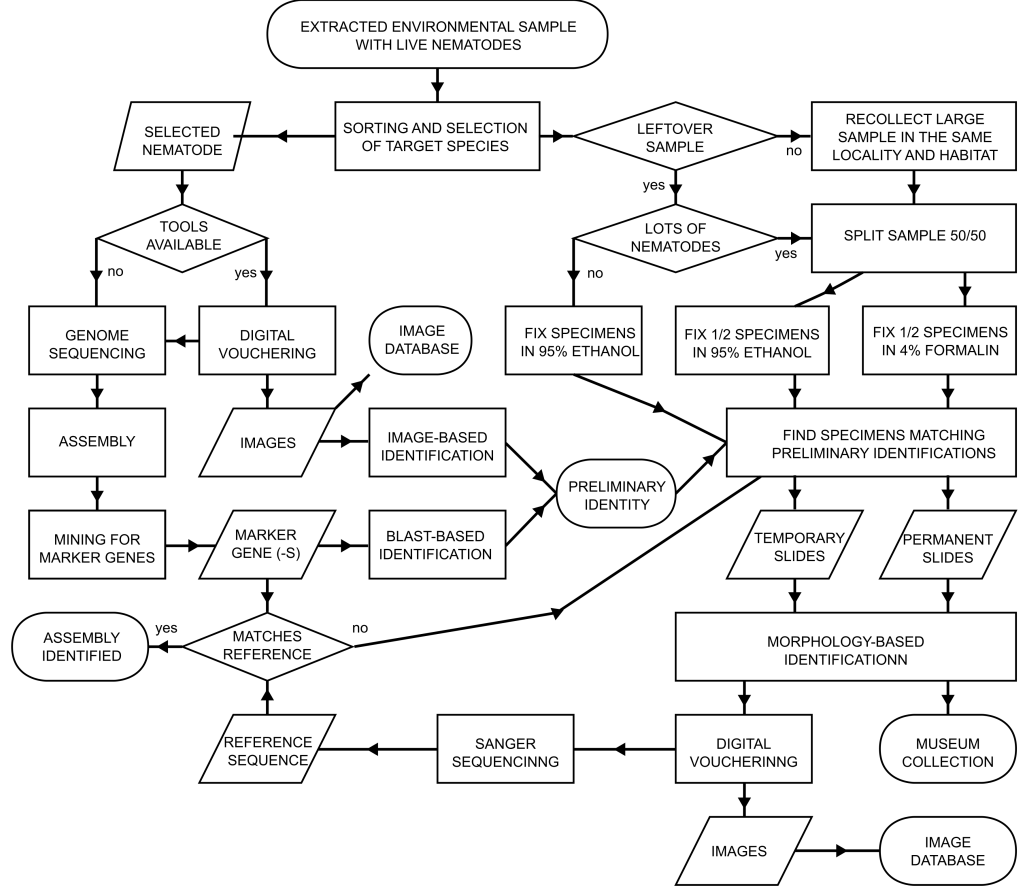


Figure 5: Pipeline from extraction of live nematode specimen to identification.

2.1.1 Nematode sample collection and processing

The sampling efforts extended to source two species, namely *Tobrilus gracilis* and *Tripylella intermedia*, with specimens collected from distinct locations in Germany and Sweden, respectively. Additionally, specimens of *Romanormis culicivora*, *Romanormis iyengari* and *Mononchus aquaticus* (live cultures) were obtained through collaboration with partners based in the United States. All these species belong to the class Enoplea and are genetically distant from a model species *C. elegans*.

Tobrilus gracilis is a freshwater nematode species collected from the sandy substrates along the shores of the pond which is located in Stadtwald, Cologne. These nematodes are particularly abundant during spring and frequently found within detritus particles. They have a distinct morphology and unique intermittent wiggling motion and can be easily separated from other freshwater nematode species. *Tripylella intermedia*, another free-living nematode species collected from the moss material supplied by Dr. Holovachov Oleksandr from the Swedish Nat-

ural History Museum in Stockholm. Both species belong to the subclass Enoplia (Clade II). *Romanomermis culicivorax* and *Romanomermis iyengari* worms were supplied by Prof. Dr. Edward Platzer at the University of California Riverside in the USA. Both mermithids are parasitoids of mosquitoes. *Mononchus aquaticus* are free-living predatory nematodes isolated from Monmouth, Illinois supplied by Dr. Nathan Schroeder. All three species belong to the subclass Dorylaimia (Clade 1).



Figure 6: *Tobrilus gracilis* was collected from the sandy substrates along the shores of the pond located in Stadtwald, Cologne.

Collection of freshwater and terrestrial nematodes included the sampling of soil, sediment, or litter from specific habitats where target species are known to thrive. For example, a site characterized by organic decay, such as decomposing wood, is recognized as favorable for various species of Rhabditidae and Panagrolaimidae. Sandy sediments of rivers and lakes are known to be inhabited by *Tobrilus*, while arboreal mosses have very specific nematode fauna. 200-500 ml of the chosen substrate was transported to the laboratory in a plastic bag or container. For cases where immediate extraction was not feasible, samples were stored at 4°C for up to several weeks. This approach allows to preservation of live nematodes until DNA extraction can be commenced, particularly suitable for species that can not be maintained in artificial conditions.

2.1.2 Extraction of live nematode specimens

Terrestrial nematodes were extracted from soil and moss using the Whitehead tray extraction method (Whitehead & Hemming, 1965). Despite its limitations in quantitative reliability, this method proved cost-effective, rapid, and efficient in yielding large numbers of viable nematodes. The process involved the use

of three essential components: a large plastic tray, a wide mesh, and thin but durable paper towels able to withstand long times being submerged in water and retain their integrity (Figure 7). The size of the tray was adjusted according to the volume of material to be processed, with larger trays allowing for the extraction of greater quantities in a single run. First, the mesh was positioned in the tray and lightly moistened with water. A single layer of paper was then placed to cover the bottom and sides of the tray, creating a semi-permeable barrier that actively moving nematodes could pass through. A thin, uniform layer of soil was evenly spread over the paper surface. Clean water was gently poured into the tray until the soil was adequately moistened but not submerged. The tray was then left undisturbed overnight and up to 48 hours, allowing actively moving nematodes to gradually descend through the paper and accumulate at the bottom of the tray while the substrate remained on the paper. After a few hours, a substantial number of nematodes can be collected at the bottom of the tray, ready for further processing. To concentrate the suspended nematodes, the water was carefully poured out of the tray over a fine sieve with a mesh size of 25 micrometers or less. The nematodes retained on the sieve were then washed off into a small beaker or a counting dish. To initiate the fixation process, nematodes were carefully transferred into a 4-5% formaldehyde solution, where they were left submerged for one day. During this time, the formaldehyde penetrates the nematode tissues, halting any metabolic activities and preserving the specimens. Following the fixation process, the next step involved transferring the nematodes into a solution composed of 5% glycerol and 95% ethanol. This ensured the long-term stability of the nematode specimens for further analysis. The combination of glycerol and ethanol enables the maintenance of the structural integrity of the nematodes while preventing degradation over time.



Figure 7: Extraction of live nematode specimen

2.1.3 Morphological and molecular identification of nematodes

The identification involved mounting live nematode specimens on temporary slides, after which each nematode was identified using a Zeiss Axioscope light microscope equipped with differential interference contrast optics. Digital images of each nematode were captured using a CCD camera allowing for the retention of images for potential reevaluation of specimen identity. The identification process relies on the morphological features unique to each nematode group. *Panagrolaimus*, for instance, exhibits distinctive characteristics such as a cylindrical or barrow-shaped stoma and protruding vulva in females, while males possess a conoid tail with bilateral papilla. *Diploscapter*, commonly found in soil or high organic decomposing substrates, is identified by its cylindrical stoma with bilateral hooks, annulated cuticle, and elongated conoid tail in females. *Acrobeloides*, thriving in soil environments, present diagnostic features like an annulated cuticle with variable lateral incisures or alae, and a single ovary with flexures in females. *Tobrilus*, a freshwater nematode, showcases sensory structures like supplements, aiding in species identification, along with a funnel-shaped stoma and sub-ventral teeth. *Enoplus*, primarily found in marine environments, exhibits eye spots or ocelli at the anterior end, a buccal cavity with three jaws and setae, and a short conoid tail in males.

After identifying the nematodes based on morphological characteristics, DNA was isolated from each worm using the Quick-DNA Miniprep Plus Kit from Zymo Research. For amplifying SSU rDNA, two partially overlapping fragments were targeted. This amplification employed a combination of three universal primers and one nematode-specific primer (1912R), intentionally included to prevent am-

plification of non-target eukaryotic SSU rDNA, such as from fungal spores adhering to the nematode cuticle. To target the first fragment, either primer 988F (5'-ctcaaagattaagccatgc-3') or primer 1096F (5'-ggtaattctggagctaatac-3') was paired with primer 1912R (5'-tttacggtcagaactaggg-3'). The second fragment was amplified using primers 1813F (5'-ctgcgtgagaggtgaaat-3') and 2646R (5'-gctaccttggtacgactttt-3') (Holterman et al., 2006). PCR reactions were conducted in a final volume of 25 l, comprising 3 l of a 100-fold diluted crude DNA extract, 0.1 M of each PCR primer, and Ready-To-Go PCR beads. The PCR was performed according to the profile in (Figure 8). Subsequently, the PCR products were cleaned using the Exo-CIP™ Rapid PCR Cleanup kit following the manufacturer's protocol. Finally, the cleaned PCR products underwent standard Sanger sequencing, and the resulting sequences were assembled into contigs using Geneious Prime 2024.0 (Kearse et al., 2012). Species identity was confirmed by BLAST against the 18S-NemaBase database (Gattoni et al., 2023), a custom and curated reference database dedicated to 18S rRNA sequences from nematodes. The process for identifying *Tobrilus gracilis* involved first extracting the 18S rRNA sequence from its assembled genome using bedtools v2.31.1 (Quinlan & Hall, 2010). Subsequently, BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990) was utilized for the identification process. The same strategy was used for the identification of *Mononchus aquaticus*.

	Step	Temperature	Time
Initial denaturation	1	94 °C	5 min
Denaturation	2	94 °C	30 sec
Annealing	3	45 °C	30 sec
Extension	4	72 °C	70 sec
	5	Goto 2, 4X	
Denaturation	6	94 °C	30 sec
Annealing	7	54 °C	30 sec
Extension	8	72 °C	70 sec
	9	Goto 6, 34X	
Final extension	10	72 °C	5 min
	11	4 °C	∞

Figure 8: Nematode 18S PCR cycle program.

In cases where nematode samples possess very low input DNA, which often happens due to loss during the DNA extraction, whole genome amplification of a single nematode was employed as a solution. For this purpose, the Direct WGA Kit from Jena Bioscience, Germany, was used, following the manufacturer's recommendations. This kit facilitated the direct amplification of genomic DNA from a single nematode without the need for DNA purification steps. The primary en-

zyme utilized in this process was Phi29 DNA polymerase, which employs a multiple displacement mechanism to isothermally amplify genomic DNAs from the samples. The amplified DNA was subsequently suitable for various applications, including library construction and Oxford Nanopore sequencing.

2.1.4 Culturing of *Mononchus aquaticus*, *Aphelenchus avenae* and *Tobrilus gracilis*

To culture predatory *Mononchus aquaticus*, Soil Extract Agar media was prepared: Garden soil was combined with deionized water at a ratio of 130g per 1L, allowing it to settle overnight. Afterward, the liquid portion was carefully separated and filtered through a Miracloth. The resulting filtered liquid was then mixed with agar to achieve a final concentration of 1.5% along with 0.005 g of cholesterol, followed by heating to boil the solution and dissolve the agar. The agar media was prepared under sterile conditions to minimize contamination and optimize growth conditions for predatory nematodes during culturing. Following preparation, the soil extract agar was sterilized via autoclaving to ensure sterility. Subsequently, the sterilized agar media was poured into sterile petri plates (100mm x 15mm) at approximately 20-25 mL per dish and left to cool overnight to solidify. To provide food for predatory *Mononchus aquaticus*, portions of agar from old *Aphelenchus avenae* culture plates were placed onto both the crude and uncrude agar plates, or the surface of agar from *Aphelenchus avenae* culture plates was rinsed with deionized water and poured onto the prepared agar plates. Adult *Mononchus aquaticus*, typically three or six individuals, were then transferred onto the plates already seeded with its prey species *Aphelenchus avenae*, which were subsequently sealed with parafilm and maintained at a temperature of 22 degrees Celsius.

In the cultivation of *Aphelenchus avenae*, the process involved mixing 39 g of Potato dextrose agar and 5 g of agar-agar in 1000ml of distilled water, with the addition of 25 mg of antibiotic. After autoclaving at 120°C, the agar mixture was poured into petri dishes (approximately 20 to 25 mL per dish, 100mm x 15mm) and left to solidify. Subsequently, the plates were inoculated with *Monilinia fructicola* fungus, which serves as a food source for *Aphelenchus avenae*. The Potato dextrose agar medium used for *Monilinia fructicola* culture is versatile, supporting the growth of various fungi, including yeasts and molds. *Aphelenchus avenae* nematodes were then transferred onto the same media plates containing actively growing *Monilinia fructicola*. Finally, the plates were sealed with parafilm and incubated at a temperature of 22 degrees Celsius.

To keep *Tobrilus gracilis* in the laboratory, water from its native freshwater habitat was blended with agar to attain a final concentration of 1.5%. The

mixture was boiled to ensure complete dissolution. Subsequently, the solution was poured into the petri dishes, approximately 20-25 mL per dish (100mm x 15mm), and allowed to cool overnight until solidified. To maintain moisture, drops of freshwater from the habitat were added to the surface of the media in the petri dishes. Nematodes were then introduced onto the plates, which were sealed with parafilm and stored at 8°C for 4-6 months.

2.1.5 Establishment of single and pooled worm HMW DNA extraction

The progress in long-read single-molecule sequencing has significantly eased the unambiguous assembly of complete genomes, enabling the exploration of previously inaccessible regions and uncovering their genuine complexity. Nonetheless, despite these advancements, genomic DNA extraction remains one of the most challenging hurdles in genome sequencing. To fully harness the capabilities of long-read sequencing, it is imperative to obtain high molecular weight genomic DNA of adequate quality and quantity. The HMW DNA extraction (Figure 9) established in this thesis is a modification of Picogram input multimodal sequencing protocol (Laumer, 2023) for the rapid and cost-effective extraction of high-molecular-weight gDNA from single and pooled nematodes. The modification mainly involved introducing isopropanol into the high-concentration salt solution, which induces the larger DNA molecules to precipitate out of the solution. In contrast, the smaller RNA fragments remain soluble. This technique was validated using five nematode species, resulting in a mean read length of 50 kb and a maximum read length of 160 kb (Figure 10). Following is a detailed description of the methodology I established and used from HMW DNA extraction (Figure 9).

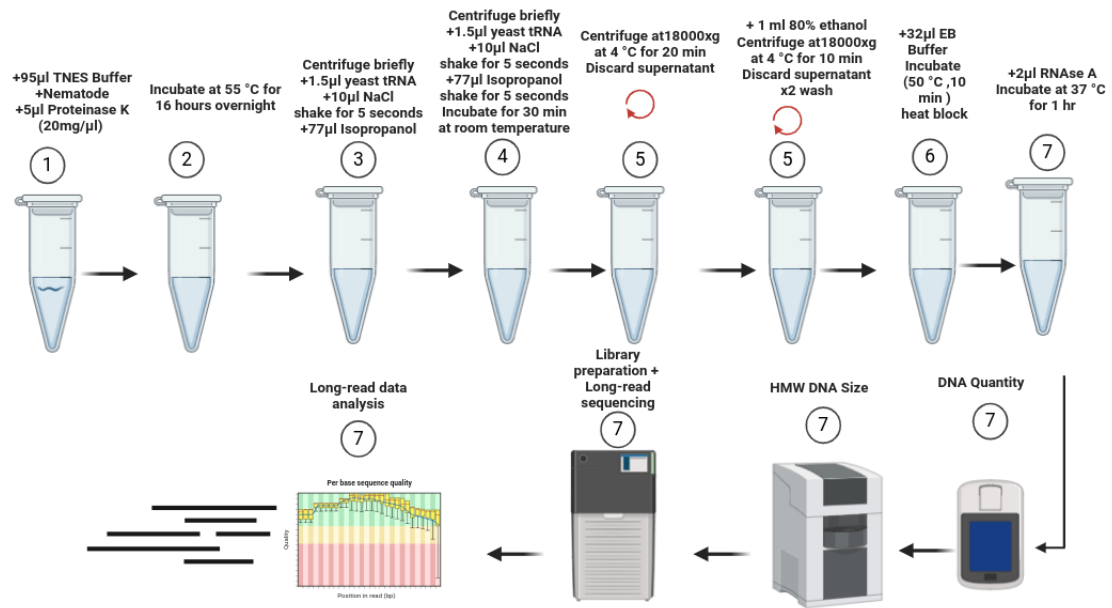


Figure 9: Shows workflow from HMW DNA extraction, Quality control of isolated DNA, long read sequencing to sequence read analysis.

Homogenization of a single nematode with freeze-and-thaw cycles.

1 or 10 nematodes were added to a microcentrifuge tube containing 95 μL of lysis buffer (100 mM Tris-HCL pH 8, 50 mM EDTA pH 8 (0.5 M), 0.5 M NaCl (stock 5M), 1% SDS (10% SDS). Next, the microcentrifuge was placed in a liquid nitrogen for 10 minutes and, after freezing it was moved to a heat block warmed to 37 °C for another 10 minutes. Sufficient homogenization of the nematode can be achieved with only three freeze-and-thaw cycles.

Lysis and overnight incubation. The lysis mix was supplemented with 5 μL of proteinase K (20 mg/ml) and incubated at 55 °C for 16 h on a heat block. During the first 30 min of the lysis, the sample was mixed manually by inverting the tube twice every 5 min.

gDNA precipitation and elution. After overnight incubation, 1.5 μL of yeast tRNA was added to the tube and mixed by gently inverting the tube. This was followed by the addition of 10 μL of 5 M sodium chloride and the solution was mixed gently by inverting the tube. Next, 77 μL of isopropanol was added and mixed gently by inverting the tube. The mix was incubated for 30 min at room temperature. The gDNA was pelleted by centrifugation at 18,000xg at 4 °C for 20 min. The supernatant was discarded, and the pellet of gDNA was washed with 1 ml of freshly prepared 70% ethanol. The gDNA was pelleted by centrifugation at 18,000 g for 10 min, and the supernatant was discarded. This step was repeated for a total of two ethanol washes. The pellet of gDNA was

air-dried for 5 min and resuspended in 32 μL of elution buffer (10 mM Tris pH 8, 1 mM EDTA pH 8). The sample was left overnight at 4 $^{\circ}\text{C}$ to uncoil the eluted gDNA. The following day, 2 μL of RNase A (QIAGEN 10 mg/ml) was added to degrade the Ribonucleic acids. This was followed by incubation for 30 min at 37 $^{\circ}\text{C}$.

Genomic DNA quantification and size verification. The gDNA was quantified with Quantus Fluorometer (Promega) and stored at 4 $^{\circ}\text{C}$ for short-term storage. The fragment sizes were measured with a Femto Pulse System (Advanced Analytical Technologies Inc., USA) using the Genomic DNA 165 kb Kit, FP-1002-0275 (Advanced Analytical Technologies Inc., USA).

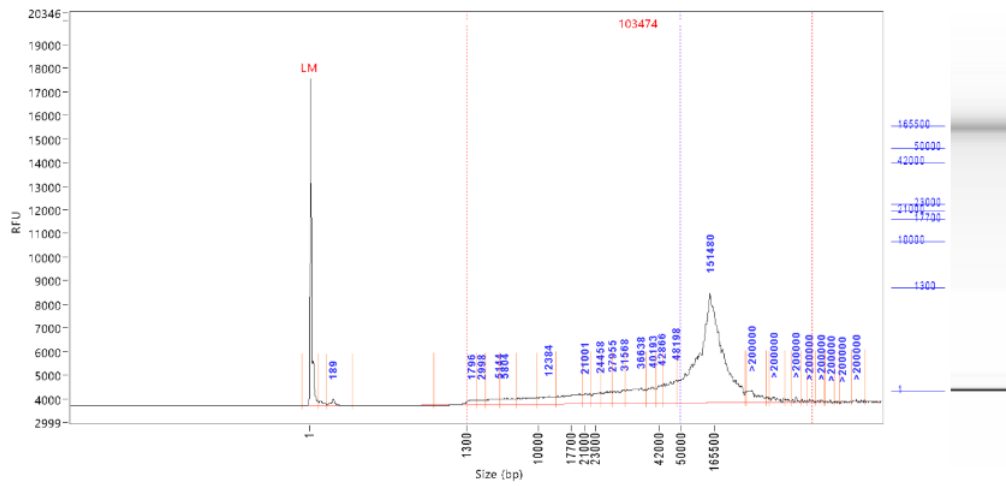


Figure 10: Femto Pulse profile showing integrity and size of *R. culicivora* gDNA obtained using the described modified salting-out protocol.

2.1.6 Long read sequencing

Following the isolation of high molecular weight (HMW) DNA and its subsequent quality assessment, the next critical stage involves library preparation and sequencing. The HiFi libraries were created using the Express 2.0 Template kit from the PacBio (Menlo Park, CA, USA), and then sequenced on a Sequel II/Sequel IIe instrument by the West German Genome Center (WGCG) with a movie time of 30 hours. The generation of HiFi reads was conducted using SMRT Link (v10) from Pacific Biosciences (Menlo Park, CA, USA), with default parameters.

2.2 Long-read assembly of nematode genomes

Below, I will describe the steps utilized for the long-read assembly of nematode genomes generated in the current thesis (Figure 11).

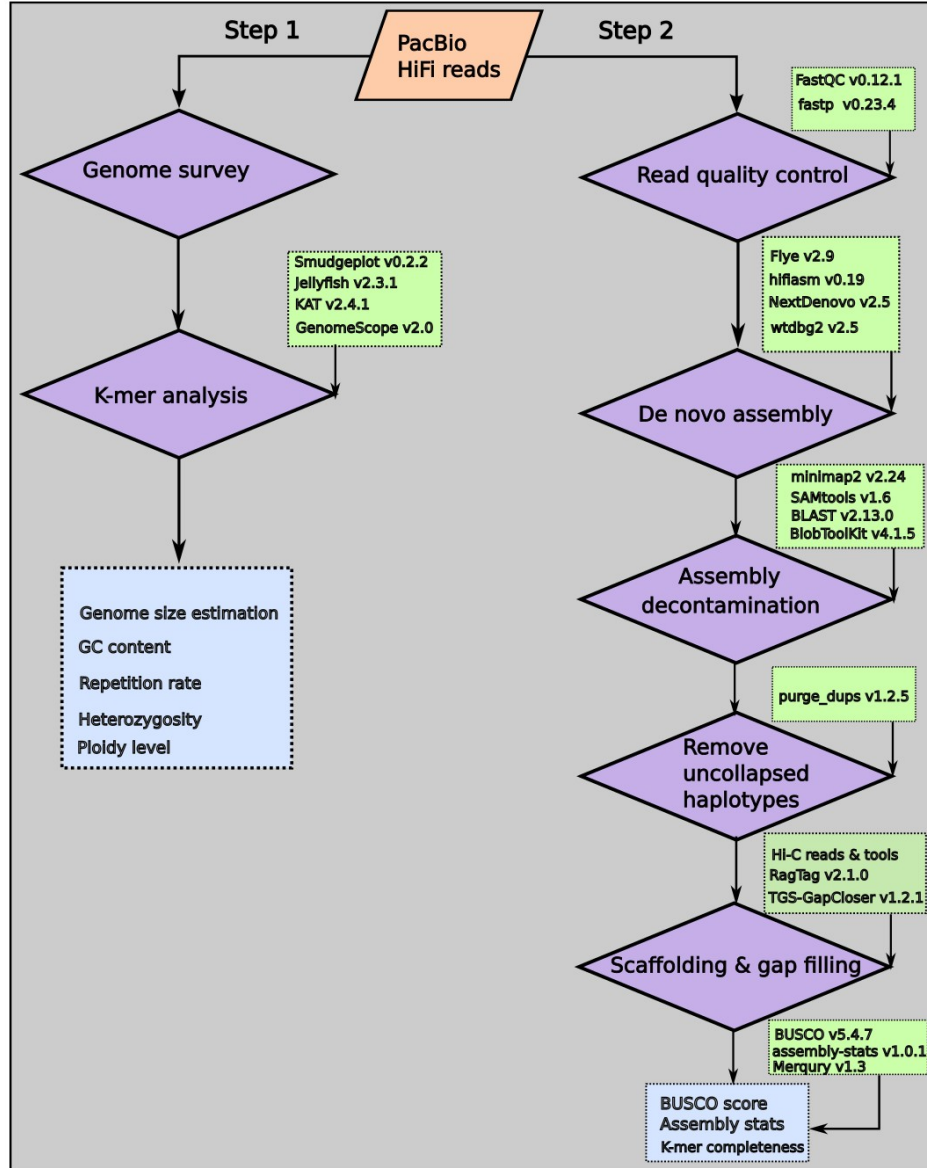


Figure 11: General workflow for the assembly nematode draft genomes using PacBio HiFi long reads.

Long read quality assessment. Before proceeding with genome analysis, it's essential to perform quality control checks on the raw data. This ensures that the data is free from any biases or issues that could affect its interpretation. The first step is assessing the quality of sequence read data. FastQC v0.12.1 (Simons, 2010) and Fastp v0.23.4 (S. Chen, Zhou, Chen, & Gu, 2018) are two commonly

used tools for this purpose. FastQC provides detailed quality control assessments, while Fastp is a comprehensive FASTQ data preprocessor, capable of removing adapter contamination, filtering low-quality reads, and correcting inaccurately represented bases. FastQC was used with parameter: (`fastqc input.fastq`) while Fastp was used with (`fastp -i input.fastq -o output.fastq`).

Genome size, heterozygosity and repeat percentage inference. This step involved the analysis of k-mer frequencies within raw sequencing reads to estimate key genome attributes such as size. k-mer refers to all the possible subsequences of length k that are contained within a string of nucleotides, such as a genome assembly or collection of sequencing reads. (Figure 12) illustrates the estimation of genome size using k-mer counting, heterozygosity, and repetitiveness. A plot of k-mer frequencies can serve as an indicator of heterozygosity in an organism. K-mers originating from heterozygous sites typically exhibit approximately half the sequencing coverage compared to those from homozygous regions, resulting in an intermediate peak or shoulder at roughly half the genome coverage peak (Figure 12B). The prominence of this peak (or shoulder) corresponds to the level of heterozygosity in the sample. Likewise, k-mers originating from repetitive regions typically exhibit a significantly higher representation compared to the average coverage, manifesting as a pronounced high-frequency shoulder towards the right side of the frequency plot (Figure 12B). K-mer counting and plots can be produced using tools such as Jellyfish v2.3.1 (Marçais & Kingsford, 2011) with option (`-m 21`), GenomeScope (Ranallo-Benavidez, Jaron, & Schatz, 2020) and KAT (Mapleson, Garcia Accinelli, Kettleborough, Wright, & Clavijo, 2017) were used with default parameters.

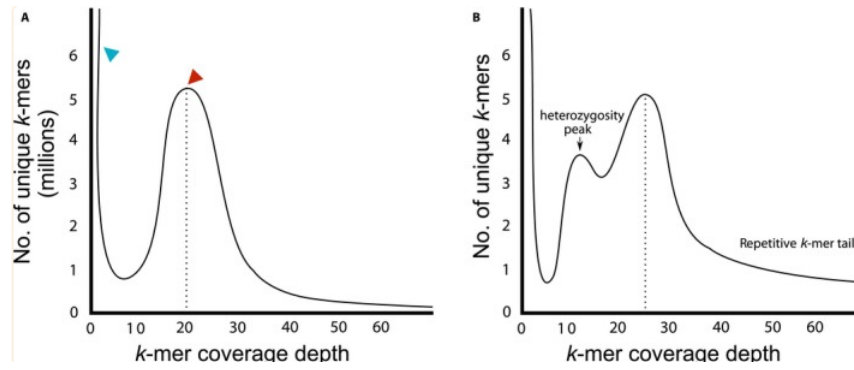


Figure 12: Theoretical k-mer frequency diagrams illustrating a genome with low heterozygosity (A) and a genome with high levels of heterozygosity and repetition (B). The x-axis displays the frequency of occurrence of a specific k-mer in the sequence reads, while the y-axis represents the count of unique k-mers at that frequency. For instance, the red arrow indicates the presence of 5 million unique k-mers occurring precisely 20 times. The initial peak near 1 (blue arrow) primarily arises from sequencing errors, resulting in numerous low-frequency unique k-mers. The subsequent peak (red arrow) denotes the sequencing coverage, indicating that a significant portion of k-mers were sequenced 20 times. The genome size estimation can then be derived by dividing the total count of k-mers (the area under the curve) by the sequencing coverage. (Image: Fay-Wei Li et al. in 2018). <https://doi.org/10.1002/aps3.1030>.

De novo assembly. The following tools were thus tested during the genome assembly process.

Flye v2.9.3 (Kolmogorov, Yuan, Lin, & Pevzner, 2019) is a rapid de novo assembler that combines overlap-layout-consensus principles with de Bruijn graphs, specifically optimized for long-read data. It was used with a parameter (`--pacbio-hifi`). Wtdbg2 v2.5 (Ruan & Li, 2020) is a novel de novo assembler tailored for long reads, employing a fuzzy Bruijn graph methodology. It utilizes an all-versus-all raw read alignment strategy to aid assembly. It was used with option (`-ax map-pb`). NextDenovo v2.5.2 (Hu et al., 2023) is a de novo assembler designed for long reads using a string graph approach. It employs a "correct-then-assemble" strategy, with the exception that it does not include a correction step for PacBio HiFi reads. However, it offers notable advantages in terms of reduced computing resources and storage requirements. It was tested with parameter (`read_type = hifi`). NextDenovo consistently produced draft assemblies of sizes similar to the expected genome sizes.

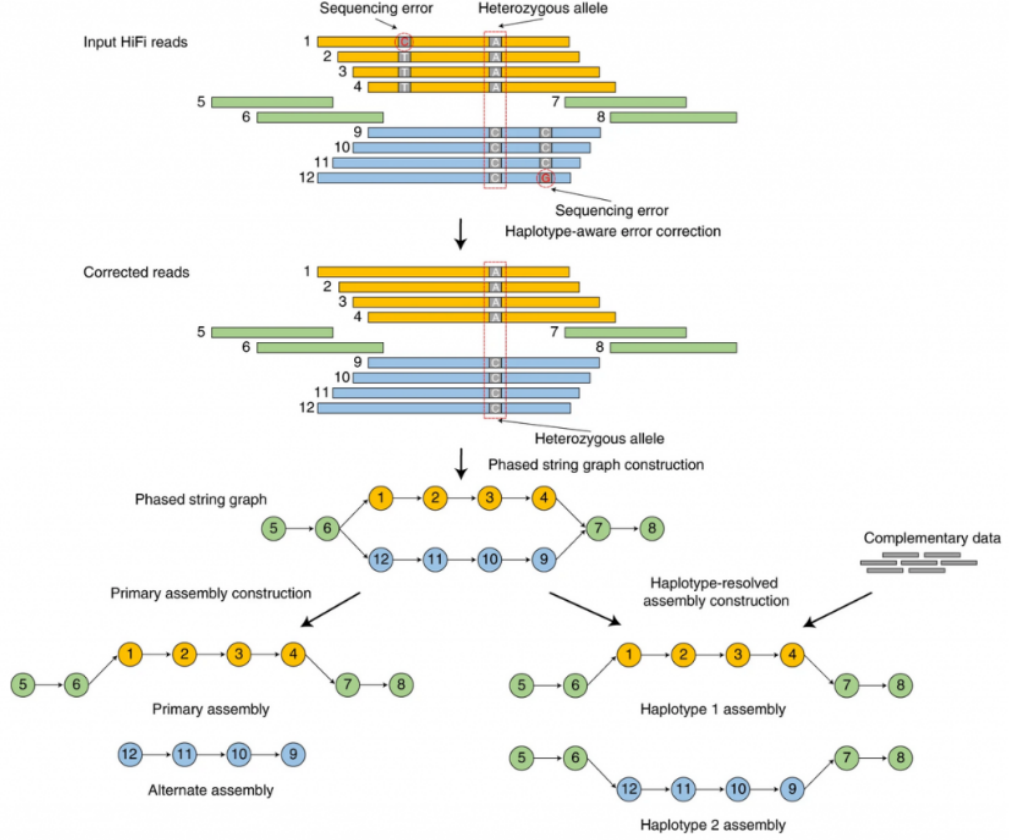


Figure 13: Hifiasm algorithmic outline. The orange and blue bars depict reads containing heterozygous alleles, indicating local phasing information, whereas the green bars originate from homozygous regions devoid of any heterozygous alleles. Within the phased string graph, each vertex represents a HiFi read identified by its unique ID. Connections between vertices signify overlaps between the reads they represent. Hifiasm initially conducts haplotype-aware error correction to rectify sequence errors while retaining heterozygous alleles. Subsequently, it constructs a phased assembly graph using the local phasing details obtained from the corrected reads. Within this graph, only reads originating from the same haplotype are interconnected. Leveraging additional data that provides global phasing insights, hifiasm produces fully phased assemblies for each haplotype based on the assembled graph. Additionally, hifiasm can generate an unphased primary assembly exclusively with HiFi reads. This primary assembly delineates phased blocks that are discernible with HiFi reads but does not maintain phasing coherence between two phased blocks. *Image: Nat Methods* 18, 170–175 (2021) (H. Cheng et al., 2021).

Hifiasm v0.19 (H. Cheng et al., 2021) is a rapid de novo assembler specialized for long reads, particularly from Oxford Nanopore or PacBio sequencing platforms. It employs long-read data for haplotype-aware error correction and to generate haplotype-resolved genome assemblies. The algorithm begins by constructing a graph based on the long-read data, capturing genomic variations

such as heterozygous variants and structural variations. Subsequently, it utilizes a greedy algorithm to traverse the graph, iteratively resolving haplotypes and yielding highly contiguous assemblies (Figure 13). During usage, it was configured with the parameter **(-l 3)**.

Genome quality assessment. This step involves basic assembly metrics and assembly completeness, to evaluate draft assemblies and select the most suitable candidate draft assembly for further analysis. In general, evaluating the quality of draft assemblies typically involved assessing the assembly size and ensuring its alignment with the estimated genome size, if accessible. The N50 metric is a commonly employed assembly metric to gauge assembly contiguity, representing the length of the shortest contig for which longer or equally sized contigs cover at least half of the assembly. Assembly quality assessment was performed using BUSCO v5.4.7 (Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015) for evaluating assembly completeness with parameters: **(-l nematoda_odb10 and metazoa_odb10)** and **(-m genome)** and basic metrics provided by QUAST v5.2.0 (Gurevich, Saveliev, Vyahhi, & Tesler, 2013) with default parameters.

Assembly contamination screening. Reconstructing genomes from sequence data generated by different instruments, regardless of the species, can be complicated by contaminant DNA. Whether introduced during sample processing or co-extracted alongside the target DNA, insufficient care during the assembly process can lead to a final assembled genome containing a mixture of data from multiple species. Such assemblies can distort sequence-based biological inference and, if deposited in public databases, may be used in downstream analyses by users unaware of this underlying issue. BlobToolKit v4.3.5 citechal-lis2020blobtoolkit employs GC content proportion and coverage as two primary dimensions to visualize contigs or scaffolds within an assembly. Additionally, it integrates taxonomic assignments to facilitate the distinction between signal (=target species) and noise (=contaminant) to differentiate sequences originating from various sources. Typically, discerning bacterial contamination in metazoan assemblies is relatively straightforward due to their distinct GC content. However, the challenge escalates when the target metazoan genome is mixed with other eukaryotes or metazoans, particularly if these species are absent from existing databases. Chromosome-level assemblies mitigate the risk of contamination by allowing downstream analyses to focus solely on sequences anchored to the main scaffolds. Moreover, leveraging Hi-C data enables the segregation of sequences from different species through the identification of their lack of trans interactions (Guiglielmoni, Rivera-Vicéns, Koszul, & Flot, 2022). Blobtools was used with parameters: **(blobtools create --fasta --cov --hits --taxdump)** and de-contamination parameters: **(blobtools filter --param --fasta --output)**

Long read mapping. Sequence mapping, a subset of alignment, is crucial in the genome assembly process for identifying the approximate origin of query sequences within the target genome. Mapping sequences to a reference facilitates assessment of the coverage and distribution of reads, identification of potential misassemblies, and validation of the accuracy of the assembly. Minimap2 v2.28 (Li, 2018) is an aligner optimized for mapping short reads, Oxford Nanopore, and PacBio reads to a reference sequence. It is used to generate coverage information from mapping sequencing reads back to the assembly. Minimap2 was used with parameters: (`minimap2 -ax map-hifi ref.fa pacbio-ccs.fq.gz > aln.sam`)

Removal of low-coverage, duplicated contigs. Purge_dups v1.2.6 (Guan et al., 2020) was used to detect and eliminate duplicate haplotypes and heterozygous overlaps in regions with high heterozygosity during genome assembly. It effectively reduces heterozygous duplications and enhances assembly continuity while preserving assembly completeness. Utilizing sequence similarity and read depth information, Purge_dups offers a robust solution to improve assembly quality and accuracy. Unlike other tools like Purge_haplotigs (Roach, Schmidt, & Borneman, 2018) and HaploMerger2 (Huang, Kang, & Xu, 2017), Purge_dups automatically distinguishes between haploid and diploid coverage depths, eliminating the need for manual read-depth cutoff specification. Additionally, Purge_dups considers both sequence similarity and read depth information, providing a comprehensive approach to haplotig identification and overlap resolution. This ensured a more accurate and reliable assembly, making Purge_dups a preferred choice for addressing heterozygosity-related challenges in genome assembly. Purge_dups was used with parameters: (`run_purge_dups.py config.json bin_dir spid -p bash`).

Scaffolding and Gap filling. Gap-filling represents a crucial final stage in genome assembly, especially for large genomes (>100 Mb). Initially, assembly algorithms generate contigs, which are contiguous sequences of overlapping reads without ambiguities, with scaffolding then linking these contigs into longer sequences using specialized sequencing data. While mate-pair libraries historically served this purpose, long-read technologies like Pacific Biosciences' RS II or Sequel platforms have become common for scaffolding, particularly the high-fidelity (HiFi) reads produced exclusively by PacBio technology, offering precision comparable to short reads and Sanger sequencing. However, the scaffolding process often leaves gaps, unidentified sequences represented by N-characters between contigs. Subsequent gap-filling aims to resolve these gaps, crucial for large and complex eukaryotic genomes. Gap-closing algorithms, like TGS-GapCloser v1.2.1 (Alonge et al., 2022), provide a streamlined approach using low-coverage, error-prone long

reads to close these gaps effectively, utilizing four main steps: detecting gaps, acquiring candidates from long read alignments, error correction of alternative sub-long reads, and closing the gaps by selecting the highest-scoring candidates. Additionally, tools like RagTag v2.1.0 (Alonge et al., 2022) offer advancements in homology-based genome assembly correction and scaffolding, introducing features like the "patch" and "merge" tools to enhance genome assembly further. TGS-GapCloser was used with parameters: (`tgsgapcloser --scaff --reads --output --racon --tgstype pb --thread`). RagTag was utilized with the following parameters: (`ragtag.py scaffold ref.fasta query.fasta`). This approach was applied to scaffold the assembly of *R. culicivora* where both PacBio HiFi and Nanopore reads were available. The NextDenovo PacBio HiFi contigs (`query.fasta`) were scaffolded using the assembly of hifiasm PacBio HiFi + Nanopore contigs (`ref.fasta`) serving as the reference.

Additionally, Long-range connection data, including Hi-C contact maps, are commonly utilized to assemble chromosome-scale scaffolds from contigs. Genome scaffolding with Hi-C integrates additional data from Arima Hi-C or Dovetail OmniC kits. In Hi-C technology, chromatin interactions are detected both within individual chromosomes and between them, allowing for scaffold assembly at the scale of full chromosomes. This technique involves several steps. Firstly, protein/DNA complexes are covalently crosslinked using formaldehyde. Next, the chromatin is digested with a restriction enzyme, and the resulting ends are marked with a biotinylated nucleotide. Subsequently, the DNA within the crosslinked complexes is ligated, forming chimeric DNA molecules. After ligation, biotin is removed from the ends of linear fragments, and the molecules are fragmented to reduce their overall size. Following fragmentation, molecules containing internal biotin incorporation are isolated using streptavidin-coated magnetic beads and modified for deep sequencing (Rao et al., 2014; Belton et al., 2012) (Figure 14). Hi-C interaction heat maps are commonly used to evaluate genome assemblies, where assembly errors appear at chromatin interaction breakpoints. Various scaffolding software tools have been created specifically for building chromosome-scale assemblies using Hi-C data, including instaGRAAL (Baudry et al., 2020), YaHS (Zhou, McCarthy, & Durbin, 2023), SALSA2 (Ghurye et al., 2019), 3D-DNA (Dudchenko et al., 2017), and pin_hic (Guan et al., 2021), each with its own set of constraints and influenced by factors such as genome complexity, repeat content, Hi-C library preparation, and sequencing coverage (Ghurye et al., 2019; Kadota et al., 2020; Guan et al., 2021). However, YaHS performs a more precise assessment of contig connections, producing genome assemblies with superior accuracy and contiguity, showing greater resilience to assembly errors. During the scaffolding process of *R. culicivora* with Hi-C reads, YaHS was tested with the

following parameters: `(yahs contigs.fa hic-to-contigs.bam)`.

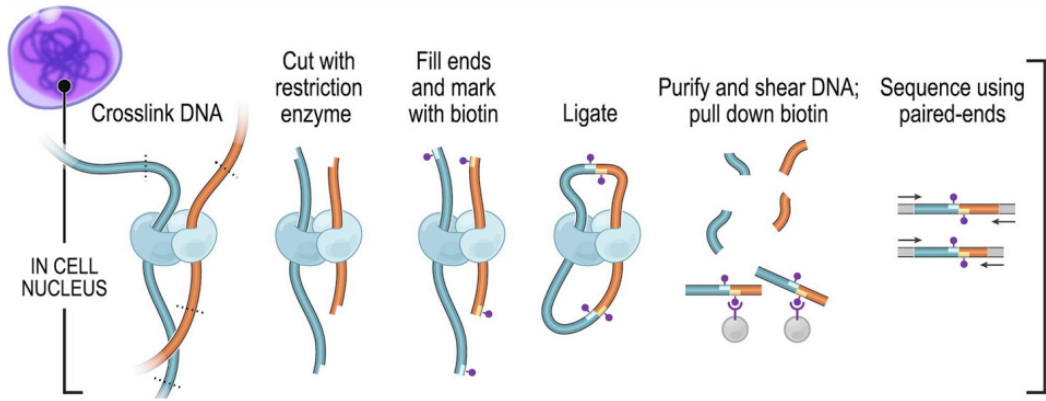


Figure 14: Hi-C overview. The procedure starts with cross-linking cells using formaldehyde to create bonds between physically adjacent DNA regions. Subsequently, the cells undergo lysis, and a restriction enzyme is applied to cleave the chromatin into numerous DNA fragments. Following purification and fragmentation of the DNA, the biotinylated fragments are isolated from the solution using streptavidin beads, and a library is prepared and sequenced. Analysis of the resulting paired-end short-read sequences yields a matrix of interactions between various DNA regions. *Image source: Suhas S.P. Rao et al., 2014, Figure 1A.* <https://doi.org/10.1016/j.cell.2014.11.021>.

Genome assembly polishing. Polishing, which is the process of rectifying errors within contigs, entails aligning reads to the assembled contigs to eliminate artifacts, thereby enhancing the accuracy and overall consensus accuracy of the draft assembled sequence. Errors like base substitutions, insertions, and deletions may occur during assembly due to factors like sequencing errors or repetitive regions. Tools such as Pilon (Walker et al., 2014), Racon (Vaser, Sović, Nagarajan, & Šikić, 2017), NextPolish (Hu, Fan, Sun, & Liu, 2020), HyPo (Kundu, Casey, & Sung, 2019) are commonly used for error correction in genome assemblies. While most tools rely on short-read alignment, ntEdit (Vaser et al., 2017) utilizes a k-mer approach, and NextPolish combines both strategies. These algorithms identify errors by analyzing read pileups, generating consensus sequences, or detecting variants. However, correcting heterozygous regions, especially in diploid genomes, poses challenges. Hypo-G (Aury & Istace, 2021) excels in integrating phasing data from high-quality reads to refine genome assemblies, particularly for diploid and heterozygous genomes. In the polishing process of the genome assemblies, Hypo-G was employed with parameters: `(hapog.py --genome -b -o -t )`.

Genome annotation process The subsequent crucial stage in rendering a genome accessible to a broader audience involves annotating it for genes, repet-

itive segments, and additional regulatory and non-coding areas. Genome annotation was performed for *R. culicivora* and *R. iyengari*. The genome annotation pipeline involved the annotation of repeats using the Extensive De novo TE Annotator (EDTA) pipeline v2.0.1 (Ou et al., 2019), employing sensitive parameters (`-sensitive 1 -anno 1`). This comprehensive pipeline integrates predictions from various tools such as LTRharvest (Ellinghaus, Kurtz, & Willhoeft, 2008), LTR_FINDER (?), LTR_retriever (Ou & Jiang, 2018), HelitronScanner (Xiong, He, Lai, Dooner, & Du, 2014), Generic Repeat Finder shi2019generic, TIR-learner (W. Su, Gu, & Peterson, 2019) and generates a transposable element library with RepeatModeler (Flynn et al., 2020). The resulting hardmasked assemblies underwent conversion to a softmasked assembly where repeat elements are written in lower case. This process makes the repeats in the genome invisible to annotation tools. Repeats can significantly impact gene predictions, making it advisable to mask them in the genome before making predictions. For RNA-seq analysis, reads are trimmed using Trim Galore v0.6.10 and mapped to the softmasked assembly using hisat2 v2.2.1 (Kim, Paggi, Park, Bennett, & Salzberg, 2019). The mapped reads were then utilized in BRAKER v3.0.3 (Gabriel et al., 2023) with specific parameters (`-gff3 -UTR off`) for subsequent annotation. The general annotation pipeline can be visualized (Figure 15).

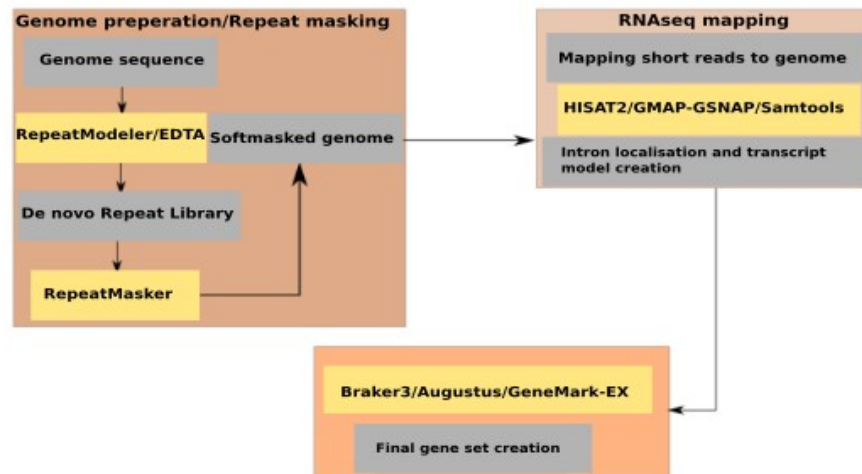


Figure 15: General workflow for primary annotation of draft genomes.

2.3 Nematode Hox gene family analysis

To investigate the Hox gene complements of a wide range of newly-sequenced, high-quality genomes from diverse non-model nematodes to understand the evolution of this key set of body patterning genes, here Hox gene complements analysis and HOX cluster identification methods are described.

In addressing the challenges posed by the unique genomic characteristics of many non-model nematodes, particularly the lack of pre-existing gene models and closely related species with well-annotated genomes in recently available species, we employed the Bitacora pipeline (Vizueta, Sánchez-Gracia, & Rozas, 2020) to facilitate comprehensive gene family annotation within genome assemblies. This pipeline utilizes a combination of BLAST and HMMER searches to identify putative candidate gene regions, whether already annotated or not, and integrates evidence from all searches to generate new gene models. The pipeline outputs a new structural annotation file (GFF) along with the sequences of the identified genes (See Figure 16 for visualized analysis pipeline).

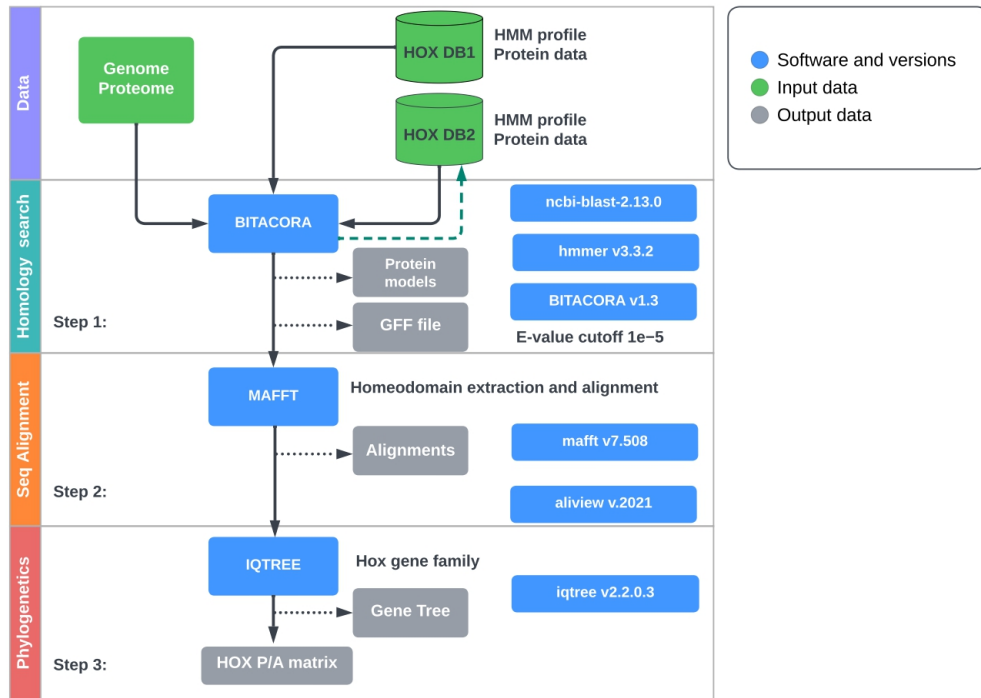


Figure 16: Hox gene identification workflow. The initial inputs of genome or proteome in fasta format and Hox gene reference profiles from the HOX database are colour-coded green. Software tools (and versions) are colour-coded blue while output files are colour-coded grey. This pipeline does not include non-Hox gene and miRNAs identification.

We collated a reference dataset of *C. elegans*, and *Branchiostoma floridae*

(amphioxus) Hox homeodomain proteins from HomeoDB (Zhong, Butts, & Holland, 2008). In the framework of the highly conserved Hox orthologues, the choice of these species allowed us to search with the reduced Hox gene complement of *C. elegans* as an ingroup and the full amphioxus set as an outgroup. We performed a sequence similarity search using the BITACORA (Vizueta et al., 2020) gene family analysis pipeline in both genome and protein mode with default parameters (e-value of 10^{-5}). The output of the first round of searching was used to generate an HMMER (Finn, Clements, & Eddy, 2011) search profile containing orthologous Hox genes from multiple nematode species, including recently available genomes without genome annotation. Final predicted proteins and corresponding gene annotation files for each genome were generated through a second round of Hox gene orthologue search in the BITACORA environment using the HMM models. The predicted proteins were aligned with MAFFT (Katoh & Standley, 2013) and corresponding homeodomain sequences were visualised and extracted using Aliview (Larsson, 2014). Phylogenetic trees were reconstructed using the homeodomain protein sequences by maximum likelihood implemented in IQTREE2 (Minh et al., 2020) to find clusters of Hox gene orthologs in different species. The final trees were visualized using Figtree (Andrew Rambaut, 2006). (See Figure 16 for visualized analysis pipeline).

2.3.1 Hox cluster annotation and gene identification

After the identification of HOX homeodomains, we extracted the respective scaffolds or contigs for annotation of complete Hox genes. To search for potential clusters of Hox genes in assembled genomes, genomic scaffolds or contigs containing Hox genes were extracted from the respective assemblies using seqtk (v1.4) (Li, Heng, 2012). I used AUGUSTUS (v3.5.0) (Hoff & Stanke, 2013) to predict complete protein models in contigs and scaffolds. To improve the accuracy of gene prediction AUGUSTUS was trained on *C. elegans* and a reference set of Hox gene profiles from *C. elegans* and *B. floridae* as hints for Hox gene finding. A database of predicted proteins from AUGUSTUS was constructed for search using BLASTP (v2.9.0) (Camacho et al., 2009) with an e-value of 10^{-5} using a query file containing a reference set of Hox protein sequences. Hox gene loci were manually inspected if they were in clusters. I define a Hox cluster as a group of four or more Hox genes spanning genomic loci with a size less than the size of the loosely connected *C. elegans* Hox genes (i.e less than 0.3 kb) on the same contig or scaffold. To validate the presence of Hox clusters, I used HbxFinder pipeline available at GitHub (<https://github.com/PeterMulhair/HbxFinder>) to identify homeobox genes in the respective contigs or scaffolds. Hox gene cluster organization and strand orientation maps from the AUGUSTUS output GFF files were drawn using

Affinity Designer 2 (<https://affinity.serif.com/en-gb/learn/designer/desktop/>).

2.3.2 Non-Hox gene interruptions within Hox clusters

Genomic contigs or scaffolds containing Hox genes extracted from the genome assembly in the previous step were used for structural annotation using AUGUSTUS (v3.5.0) (Hoff & Stanke, 2013) and functional using eggNOG-mapper v2 (Cantalapiedra, Hernández-Plaza, Letunic, Bork, & Huerta-Cepas, 2021), leveraging pre-computed eggNOG v5.0 clusters and phylogenies to assign orthologous groups and infer functional annotations. Non-Hox genes present within the Hox cluster were extracted and counted from the respective GFF structural annotation files.

2.3.3 microRNA interruptions within Hox clusters

To examine the microRNAs within the Hox cluster, I used *Caenorhabditis elegans* microRNA references from miRBase (Kozomara, Birgaoanu, & Griffiths-Jones, 2019), a comprehensive resource for microRNA sequences and annotations. Initially, custom nucleotide databases were built for each species' Hox cluster. Following this, I performed searches against these custom databases using established mature miRNAs from *Caenorhabditis elegans*, setting the e-value threshold to 10⁻⁵ in BLASTN (Camacho et al., 2009) analysis.

2.4 Genomic signatures of DNA methylation in the phylum Nematoda

To investigate the presence of genomic signatures of DNA methylation and to understand the evolution of this regulatory mechanism across the Nematoda, the following methodology was used.

2.4.1 Genome data acquisition

Our goal was to include in the analysis a comprehensive and well-balanced but still bioinformatically manageable selection of nematode species from across the entire phylum. Moreover, I specifically included additional species for those genera where DNA methylation was studied in the past, namely *Caenorhabditis* and *Trichinella*. I included only the genomes with above-average completeness, estimated using BUSCO Nematoda odb10. Many of the genomes included in this analysis have not been available until now, especially for free-living Enoplia (Clade II), Dorylaimia (Clade I) and early-branching Chromadoria (outside Clade III-V) lineages. These new draft genomes were produced at the Darwin

Tree of Life (Sanger Institute, UK), the WormLab laboratory (the University of Cologne, Germany) and the Department of Plant Pathology (Nanjing Agricultural University, China). In addition, I included four genomes from the sister phylum Nematomorpha, namely *Acutogordius australiensis*, *Nectonema munidae*, *Gordionus montsenyensis*, and *Gordius aquaticus*.

2.4.2 Identification of DNMTs in nematode genomes

Nematode genomes were searched for Cytosine methyl-transferase domains using BITACORA pipeline (Vizueta et al., 2020). The analysis steps were similar as visualized in A reference dataset of *Homo sapiens* DNMT1, 2 and 3 was used to annotate potential methyltransferases in recently available nematode genomes. The reference set of DNMTs together with UHRF1 were obtained from UniProt, which is a comprehensive and freely accessible resource of protein sequence and functional information (“UniProt: the universal protein knowledgebase in 2023”, 2023). A sequence similarity search was conducted using the BITACORA (Vizueta et al., 2020) in both genome and protein mode with default parameters (e-value of 10⁻⁵). The output of the first round of searching was used to generate an HMMER (Finn et al., 2011) search profile containing orthologous methyltransferase genes from multiple nematode species. Final predicted proteins and corresponding gene annotation files for each genome were generated through a second round of orthologue search in the BITACORA environment using the HMM models. To generate clusters of DNMT gene families based on their evolutionary relationships, the predicted proteins were aligned with MAFFT (Katoh, Misawa, Kuma, & Miyata, 2002) and corresponding homeodomain sequences were visualized and extracted using Aliview (Larsson, 2014). Phylogenetic trees were reconstructed using the conserved domains of DNA methyltransferase protein sequences by maximum likelihood implemented in IQTREE2 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) to find clusters of methyltransferase gene orthologs in different species. The final trees were visualized using Figtree (Andrew Rambaut, 2006). Similar methods were used as in (Figure 16)

2.5 Analysis of genetic features encoded in the genomes of worm parasitoid taxa

To investigate the genetic features encoded in the genomes of two independently evolved worm parasitoid lineages (Mermithidae and Nematomorpha), the following steps were conducted.

2.5.1 Orthologous cluster identification and visualization

Identifying orthologous clusters is crucial for comparative genomic studies as it facilitates the examination of evolutionary relationships between genes across diverse species. Orthologous gene identification was conducted using OrthoFinder (Emms & Kelly, 2019) integrated into OrthoVenn3 pipeline (Sun et al., 2023), employing an E-value cutoff of $1e-5$ for comparing all protein similarities and an inflation value of 1.5 to generate orthologous clusters via the Markov clustering algorithm. Gene Ontology term annotations for individual orthologous clusters were further performed (Figure 17 analysis pipeline).

2.5.2 Gene family analysis

To gain insight into the evolutionary dynamics of genes within both Mermithidae and Nematomorpha, we assessed the expansion and contraction of gene ortholog clusters. Computational analysis of gene family evolution was conducted using CAFE5 (Mendes, Vanderpool, Fulton, & Hahn, 2020) software, defining expansion and contraction based on differences in cluster size between ancestors and each current species. Furthermore, to elucidate the genetic factors or mechanisms underlying phenotypic traits in species, the gene families were annotated with GO terms (Figure 17).

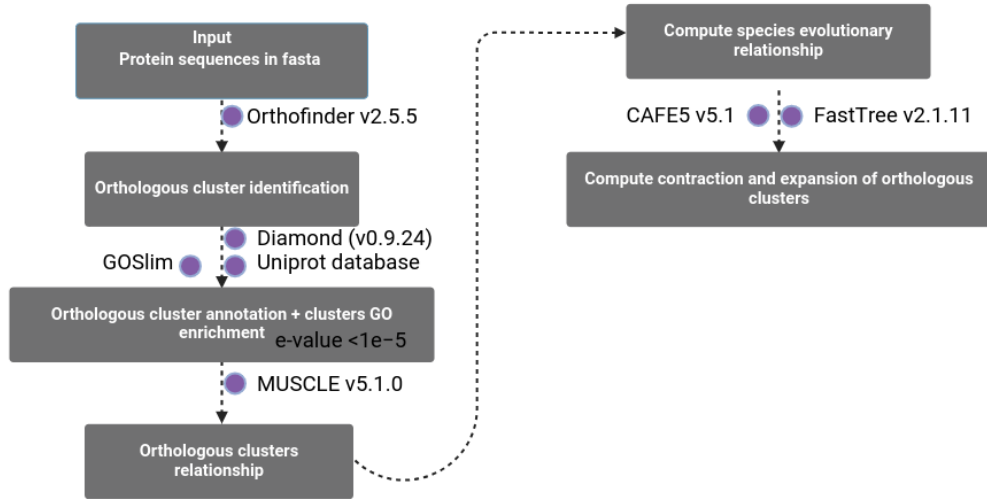


Figure 17: Integrated pipeline for orthologous cluster identification, phylogenetic analysis allowing for the inference of evolutionary relationships among species based on their orthologous clusters, gene family contraction and expansion analysis. The steps in the analysis are shown using rectangles. The analysis tools are shown in purple circles.

2.5.3 Repeat sequence analysis

To assess the content of transposable elements (TE) within the Mermithidae and Nematomorpha, first, a TE and repeat library was built using EDTA v2.1.3 (Ou et al., 2019) with parameters **—sensitive 1 —anno 1**. TEs were then annotated and classified to superfamily level using the FasTE pipeline (Ellen A. Bell, 2022). Output files were parsed with the R script “RMTrips”, contained in FasTE. A custom R script “RMTrips output to divsum-file format” was used, renaming the TE superfamilies according to Wicker-classification (Wicker et al., 2007), and converting the data format to further plot the TE divergence landscapes with the “Plot Kimura Distance” R script. Stacked barplots were plotted in R.

3 RESULTS

3.1 Genomes of early branching nematode species from long-read sequencing

The limited availability of high-quality genomes from early branching Nematoda lineages, particularly Enopleans, has posed a significant challenge in conducting comprehensive phylum-wide genome-scale evolutionary analyses. This scarcity has hindered our understanding of the deeper phylogenetic relationships within the phylum (Smythe et al., 2019), as well as the evolution of crucial developmental gene families such as HOX genes across the Nematoda. While the genome of *Romanomermis culicivorax* represents an improvement over previously sequenced short-read genomes (Schiffer et al., 2013), the following genomes (Figure ??) presented in this thesis are entirely novel and have been assembled using long-read sequencing technologies. Assembly metrics for these newly constructed genomes, including *R. culicivorax*, *R. iyengari*, *Tobrilus gracilis*, *Mononchus aquaticus*, and *Tripylella intermedia*, are summarized in Table 1.

Table 1: Assembly evaluation of five early-branching nematode genomes generated. BUSCO databases used: nematoda_odb10 and metazoa_odb10.

Assembly metrics	<i>R. culicivorax</i>	<i>R. iyengari</i>	<i>M. aquaticus</i>	<i>T. intermedia</i>	<i>T. gracilis</i>
Assembly size	359 Mb	304 Mb	212 Mb	217 Mb	148 Mb
N50	1000 kb	326 kb	694 kb	700 kb	199 kb
BUSCOs	68.4%	63.3%	86.3%	90.1%	84.9%
Single copy orthologs	65.2%	52.0%	75.4%	72.3%	64.9%
Duplicated orthologs	3.2%	11.3%	10.9%	17.8%	20.0%
Fragmented orthologs	9.2%	2.6%	1.9%	1.8%	1.2%
BUSCOs	39.4%	45.0%	64.2%	63.3%	57.8%
Single copy orthologs	37.3%	36.6%	54.3%	35.4%	44.8%
Duplicated orthologs	2.1%	8.4%	9.9%	27.9%	13.0%
Fragmented orthologs	4.1%	3.0%	2.9%	3.4%	2.9%
GC level (%)	36.0	37.4	46	36	48
Repeats masked (%)	68.2	53.9	9.8	21.4	17.0

3.1.1 *Romanomermis* genomes had a high level sequence repetitions

Two genomes from the genus *Romanomermis* were sequenced, belonging to an early-branching Clade I of the Nematoda. For *Romanomermis culicivorax* which was re-assembled, both Nanopore and PacBio HiFi sequence data were generated, yielding 5.7 Gb and 37.5 Gb of reads, respectively, with respective N50 values of 15.9 kb and 12.6 kb. The final assembly size for *Romanomermis culicivorax* reached 359 Mb, with an N50 of 1.0 Mb. Overall BUSCO scores indicated 68.4% completeness for metazoa_odb10, with 3.2% duplicated orthologs and 9.2% fragmented orthologs. Assessment for nematoda_odb10 orthologs revealed 39.4% completeness, with 2.1% duplicated orthologs and 4.1% fragmented orthologs. Additionally, 68.2% of the genome was identified as repetitive, with 16,689 annotated genes and overall BUSCO scores of 77.6% (Metazoa) and 56.2%

(Nematoda). Comparatively, the genome of *Romanomermis iyengari* had lower contiguity because only HiFi data was used in assembly construction. Access to material for Oxford Nanopore sequencing was not enough to proceed with sequencing. This genome, the second in the genus to be sequenced, exhibited a 304 Mb assembly size with a 37.4% GC content and a scaffold N50 of 326 kb, with the longest scaffold spanning over 3.9 Mb. BUSCO scores for Metazoa and Nematoda were 52.0% and 45.0% completeness, respectively, with duplicated orthologs at 11.3% and 8.4%, and fragmented orthologs at 2.6% and 3.0%, respectively (Table 1).

3.1.2 The genome assembly of a predatory nematode *Mononchus aquaticus* contains a full DNMT methylation toolkit

Mononchus aquaticus is a free-living nematode with a predatory lifestyle. It belongs to early-branching Enoplean nematodes within the subclass Dorylaimia and family Mononchidae. The genome assembly was generated from PacBio HiFi data. The assembly size is 212 Mb with N50 of 694 kb. Orthology analysis revealed 75.4% BUSCO completeness for the Metazoa database and 64.2% for Nematoda (Table 1). Downstream comparative analysis of this genome together with published nematode genomes across the Nematoda revealed a full complement of DNA methylation transferases including DNMT1 and DNMT3 (Figure 21).

3.1.3 The genome assembly of a free-living parthenogenetic nematode *Tripylella intermedia*

Parthenogenetic ancestral nematodes, particularly from Enoplea, are scarcely represented in genome-scale evolutionary studies of sex. *Tripylella intermedia* is a free-living predatory nematode with a parthenogenetic lifestyle belonging to Clade II of the Nematoda, in the subclass Enoplia and family Tripylidae. The genome assembly for this species was generated from PacBio HiFi data. The assembly size is 217 Mb with N50 of 700 kb. Orthology analysis revealed 72.3% BUSCO completeness for Metazoa and 63.3% for Nematoda (Table 1). Downstream comparative analysis of this genome together with published nematode genomes across the Nematoda revealed an absence of DNA methylation Methyltransferases including DNMT1 and DNMT3 (Figure 21) however with the presence of DNMT2 which is a tRNA-methyltransferase (Figure 21).

3.1.4 The genome assembly of the only free-living freshwater sexual nematode *Tobrilus gracilis*

Freshwater free-living nematodes are particularly understudied in genome-scale research compared to their terrestrial and parasitic counterparts. The draft genome of *Tobrilus gracilis* was constructed using PacBio HiFi data, and possesses an assembly size of 148 Mb with an N50 value of 199 kb. Orthology analysis indicated a BUSCO completeness of 64.9% for Metazoa and 57.8% for Nematoda (Table 1). Subsequent comparative analyses involving this genome, along with other published nematode genomes across the Nematoda, unveiled the absence of DNA methyltransferase, such as DNMT1 and DNMT3. However, similar to other Clade II Enopleans examined, tRNA-methyltransferase was found to be present (Figure 21).

3.2 Evolutionary plasticity in Nematode Hox gene complements and genomic loci arrangement

To unravel the evolutionary trajectory of nematode Hox genes, particularly to uncover the origins of the "dispersed" cluster seen in *C. elegans*, a comprehensive analysis of high-quality genomes from 80 species representing all major clades of Nematoda was conducted. Utilizing a homology-based search strategy, Hox gene complements across these species were assembled and annotated. Nematodes exhibit Hox genes belonging to up to six orthology groups. Although Hox "clusters" in nematodes are frequently disrupted by unrelated genes, certain species harbor an intact, non-dispersed Hox cluster.

3.2.1 HOX gene evolution in nematodes is largely characterized by consecutive losses

Hox genes play critical roles early in embryonic development in determining cell fates so that the correct body parts are formed in the right places from nematodes to vertebrates. While it is possible to develop interim catalogues of Hox gene repertoires by directed PCR, transcriptomics or draft genome sequencing, to validate presence-absence patterns and explore synteny relationships, high-quality genome sequences are required. Across the analyzed nematode species the identified loci were attributable to six Hox orthology groups; HOX1 (arthropod *lab*, *C. elegans* *ceh-13*), HOX3 (*zen*), HOX4 (*Dfd*, *lin-39*), HOX6-8 (*ftz*, *mab-5*), HOX6-8 (*Antp*), HOX9-13 (*AbdB*, *egl-5/php-3/nob-1*). I did not find loci that were classifiable as HOX2 (*pbscr*) and the *Ubx*/*AbdA* subtypes of HOX6-8 loci in any of the nematode genomes analyzed. The maximum number of Hox loci in a single

species was seven (in members of Spirurina / Clade III), and the minimum number of genes was four (*Oncholaimus oxyuris*) (Figure 18). Loci attributed to HOX orthogroups HOX9-13 showed variable presence across the nematode phylogeny (Figure 18).

3.2.2 The analysis pipeline reliably recovers *C. elegans* HOX loci

The model *C. elegans* has six Hox genes belonging to four core Hox orthology groups (*ceh-13* from the HOX1 group, *lin-39* from HOX4, *mab-5* from HOX6-8 and *egl-5*, *php-3* and *nob-1* from HOX9). The analysis pipeline (Figure 16) reliably recovered all the *C. elegans* expected Hox loci (Figure 18).

3.2.3 Hox6-8 (*Antp*) loci variation in Clade V (including an ingroup Strongylomorpha=Strongyoidea) crown group nematodes

Nematodes belonging to the Clade V form a diverse group exhibiting various lifestyles, ranging from being free-living bacterial feeding or predatory to obligate parasites in vertebrates and invertebrates. *C. elegans* misses Hox6-8 (*Antp*) loci, two of the early-branching *Caenorhabditis* members possess Hox6-8 (*Antp*). Hox6-8 (*Antp*) is absent in derived members of *Caenorhabditis*. Hox6-8 () is present in the crown lineage Strongyloidea, except for *Teladorsagia circumcincta* and *Nippostrongylus brasiliensis*. In all the members of Clade V included in the analysis, all *C. elegans* Hox genes are present except for *nob-1* (Figure 18).

3.2.4 Hox3 gene loss in Tylenchina and Spirurina

During the investigation, I noted that all the examined members within the Tylenchina clade (Clade IV) lacked HOX3 loci. Within the Spirurina clade (Clade III), specific instances of Hox3 loss were identified in *Syphacia muris*, *Enterobius vermicularis*, and *Dracunculus medinensis*, which are illustrated in Figure 18. There was a single loss of *ant-1* in Tylenchina, except in the most early branching *Steinernema carpocapsae*, as depicted in Figure 18. However, despite the presence of *ant-1* loci in all Spirurina species included in the study, *Syphacia muris* and *Anisakis simplex* were exceptions where this locus was found to be absent. Furthermore, there were separate instances of Hox9-13 (*egl-5*) loss among the members of the Tylenchoidea crown group that includes root-knot and cyst nematodes parasites of plants.

3.2.5 Hox gene complements in early-branching chromadoreans

Previous investigations into the Hox gene complement among early-branching members in Chromadorea were limited due to insufficient genomic resources, im-

peding a comprehensive understanding of Hox gene evolutionary patterns (Yoshida et al., 2017). I have expanded the analysis by utilizing newly available draft genomes in early-branching Chromadorea. Hox3 loss was found in (*Terschellingia longicaudata*, *Daptonema setosum*, and *Axonolaimus paraspinosus*). Furthermore, the Antp-like locus was absent in *Plectus sambesii*, *Sabatieria pulchra*, and *Eleutherolaimus stenosoma*. Additionally, there were independent cases of Hox9-13 (*egl-5*) loss in *Leptolaimus papilliger*, *Daptonema setosum*, and *Eleutherolaimus stenosoma* in early-branching Chromadorea (Figure 18). Other members of this Clade collectively possessed a total of seven Hox genes, as depicted in Figure 18.

3.2.6 An independent loss of ant-1 locus in Enoplida, and a single HOX9-13 (*php-3*) in Enoplean nematodes

The Clades I and II (Enoplea) class encompasses the earliest-branching lineages within the Nematoda. Previous investigations into the Hox gene composition of Enoplea were limited due to the lack of genomic data for most of the non-parasitic Enoplean lineages. In this study, nine genomes from Enoplia (Clade II) and nine genomes from Dorylaimia (or Clade I) were utilized. Initially, the findings revealed the presence of a collective total of six Hox gene loci in most Enoplean members (depicted in Figure 18). Among these, a solitary HOX9-13 orthologue was identified, which was classified as *php-3*. However, *Oncholaimus oxyuris*, a species with only four Hox orthologue groups, lacked any HOX9-13 loci, specifically the one (*php-3*) that would be anticipated. It's important to note that this assembly originated from a single specimen amplified for long-read sequencing, and the absence might stem from residual assembly issues rather than actual loss (as shown in Figure 18). The *egl-5* locus was absent in both Enoplea (Clade II) and Dorylaimia (Clade I) species. Furthermore, there were independent losses of the Antp-like locus within sampled Clade II (Enoplida) members among Enoplia (Figure 18).

3.2.7 Nematomorpha have at least six Hox loci

Nematomorpha, the sister phylum of Nematoda (Bleidorn, Schmidt-Rhaesa, & Garey, 2002), represents a group of parasitoid worms commonly known as horse-hair worms or Gordian worms. Hox gene complements from four recently available genomes included: Hox loci corresponding to Hox1 (*lab*), Hox2 (*Pb*), Hox4 (*Dfd*), Hox6-8 (*Antp*), Hox6-8 (*Abd-A*) and Hox9-13 (*Abd-B*). I did not find Hox loci corresponding to Hox3 (*zen*) and Hox6-8 (*Ftz*) (Figure 18). There was a duplication of Hox4 (*Dfd*) in *G. aquaticus* and *G. montsenyensis* (Figure 18).

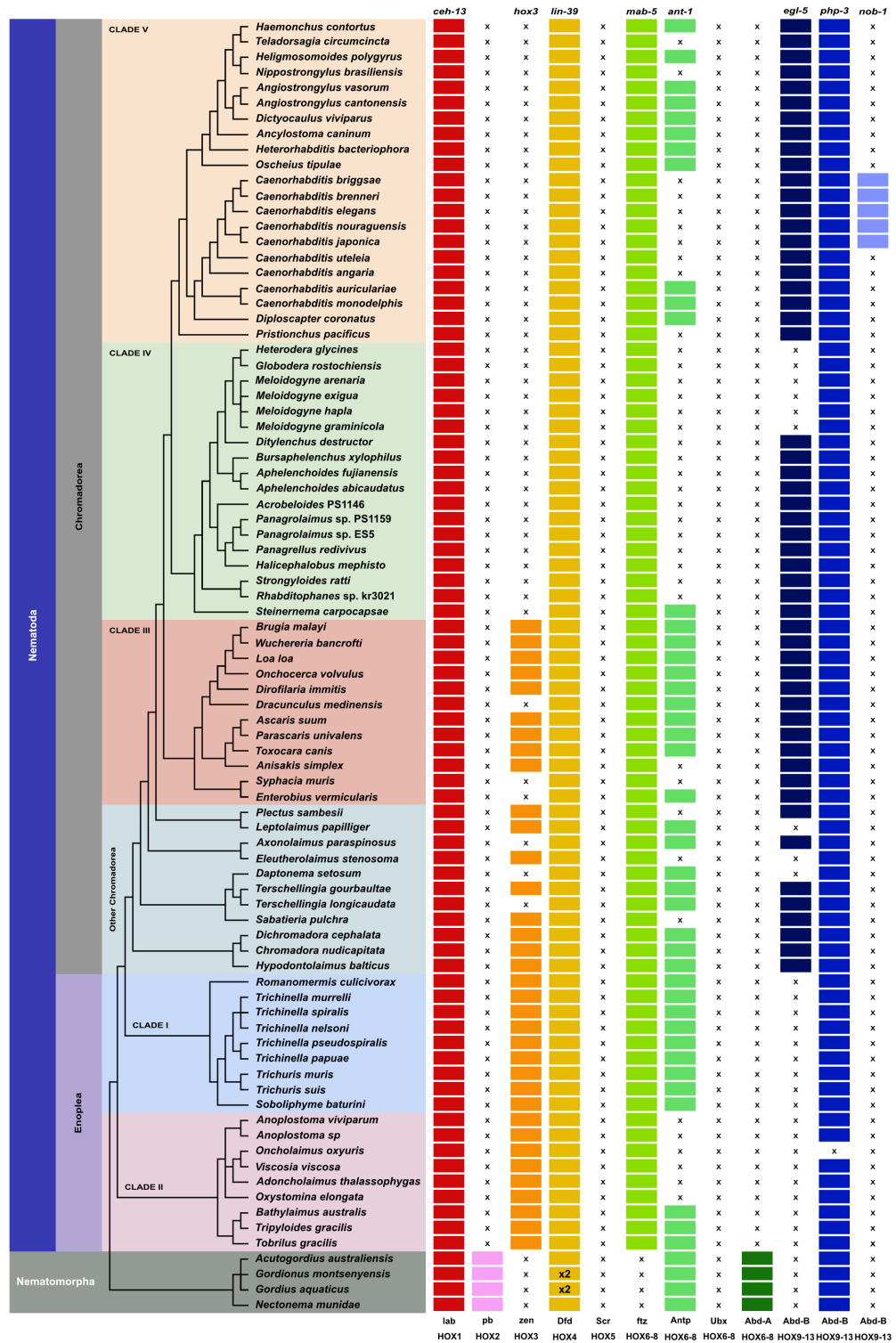


Figure 18: Nematoda and Nematomorpha Hox gene complements. Genes belonging to the same orthology group (HOX1-HOX9) are coloured similarly. X indicates the absence of the orthologue. x2 indicated the duplication event. The cladogram on the left shows the phylogenetic relationships of the analysed nematode species based on (Ahmed & Holovachov, 2021; Ahmed et al., 2022)

3.2.8 nob-1 is a recent duplication within *Caenorhabditis*

I expanded the scope of the study to investigate the Hox gene complements in diverse *Caenorhabditis* nematodes beyond *C. elegans* (Figure 19). In the analyzed members of *Caenorhabditis*, the Hox gene repertoire resembled that of *C. elegans*, with an exception: the detection of a recent nob-1 gene duplication in species belonging to the "elegans" and "japonica" subgroups including *C. briggsae*, *C. brenneri*, *C. nauraguensis*, and *C. japonica* (Figure 19). The nob-1 gene was absent in the earlier branching *Caenorhabditis* members of the "drosophilae" subgroup (Figure 19).

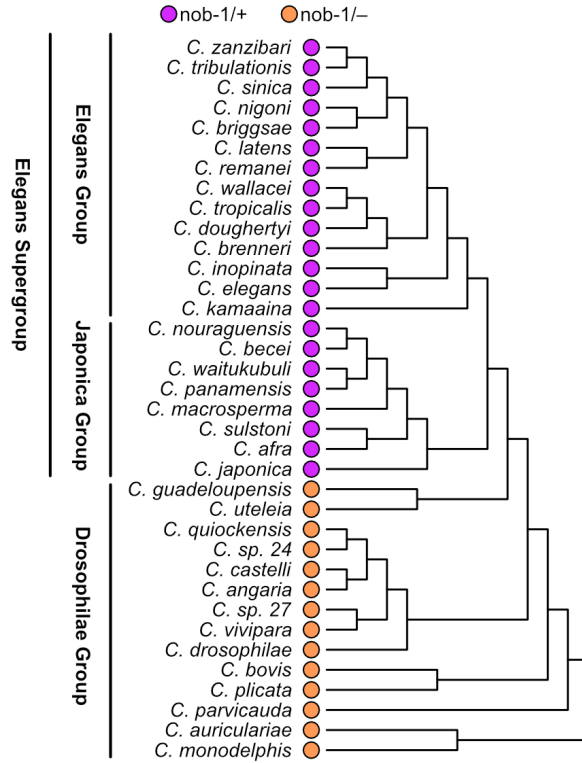


Figure 19: nob-1 Hox gene complement across *Caenorhabditis* lineage.

3.2.9 Hox cluster organization in the Nematoda

Clustering and collinearity are prominent features of HOX genes in a wide spectrum of species. In contrast to the loose clustering of Hox genes in three loci pairs on chromosome III in *C. elegans*, the analysis of nematode genomes in the present study reveals more condensed Hox gene clusters. These clusters exhibit variations in the number of genes, the sequence of gene arrangement, and the orientation of transcription. Furthermore, there are interruptions within these Hox gene clusters by the presence of non-Hox genes.

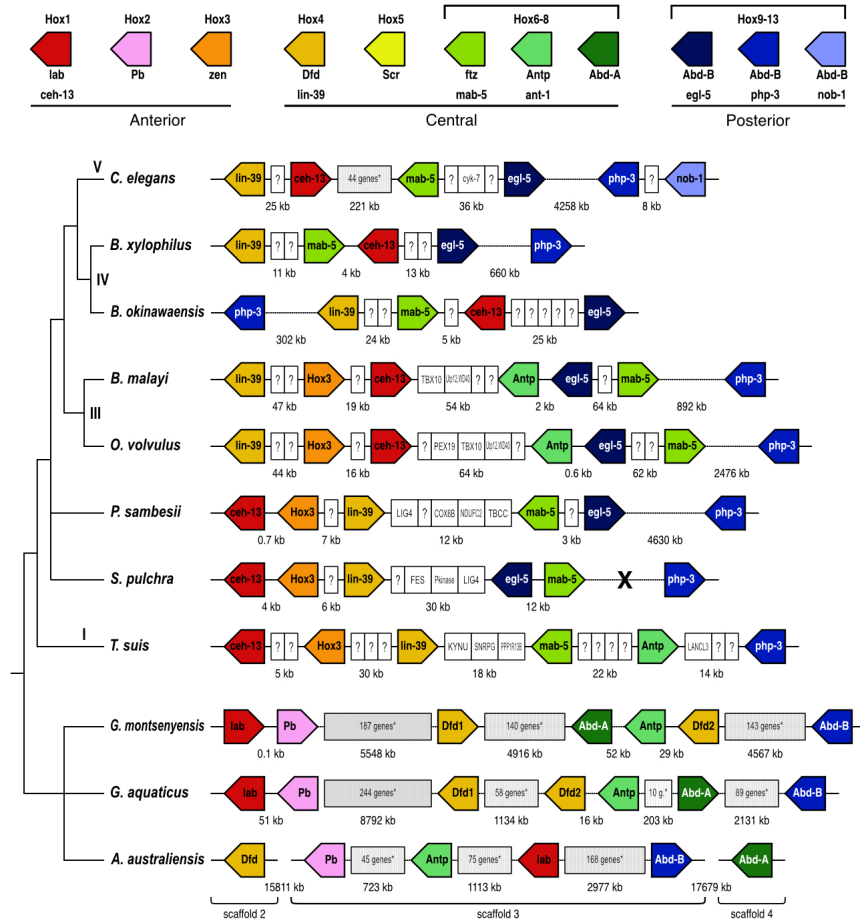


Figure 20: Hox gene clusters for eight selected species of Nematoda and all three studied species of Nematomorpha showing synteny relationships of Hox cluster loci. Identified Hox genes and their respective transcriptional orientations are indicated by arrow boxes which are coloured by orthology group. Dashed horizontal lines indicate cluster breakage in scaffolded genomes. Lineage-specific duplications include paralog groups Hox6-8 and Hox9-13. A cross indicates a break that resulted from gaps in the un-scaffolded draft genome. The cladogram on the left summarises the relationships between species.

3.2.10 Unique characteristics of Hox gene organization in *Bursaphelenchus*

A comprehensive investigation into clustering, gene order, and orientation within Hox gene sets across the phylum was conducted. The analysis revealed potential HOX clusters in several species. Notably, in the two *Bursaphelenchus* species, four out of the five Hox loci were found to be closely linked, with a separation of less than 40 kb in *B. xylophilus*, while the HOX9-13 locus, *php-3*, was separated by 302 or 660 kb. Within this core set of four loci, the genes maintained the same relative order, but in *B. okinawensis*, the core cluster exhibited an inversion concerning the position of *php-3* compared to its arrangement in *B. xylophilus*.

The order of genes in the core did not align with the expected HOX1-HOX6-8 ordering where in both species *lin-39* and *mab-5* are anterior to *ceh-13* (Figure 20).

3.2.11 Some members of Spirurina (Clade III) display Hox clusters containing a greater complement of Hox genes.

Both *Brugia malayi* and *Onchocerca volvulus* possess expanded Hox clusters containing seven linked Hox genes, of which six are closely clustered, and both species have a greater number of Hox genes compared to most other members of the Nematoda group. This core cluster comprises six members (*ceh-13*, Hox-3, *lin-39*, *mab-5*, *ant-1*, *egl-5*), while the HOX9-13 *php-3* locus is situated at a distance of 0.9 Mb (*B. malayi*) or 2.5 Mb (*O. volvulus*). Although the gene order within the core group remains consistent between the two species, it does not adhere to the typical HOX1-HOX6-8 structure, and there are no differences in the transcriptional orientation of the Hox loci between these species (Figure 20).

3.2.12 Hox cluster organization in basal Chromadorea

In *Plectus sambesii* and *Sabatiera pulchra* (Chromadorea) clusters containing five Hox genes, the homologues of *ceh-13*, Hox-3, , *mab-5* and *ant-1*, were identified. The cluster in *P. sambesii* was particularly compressed (less than 40 kb) (Figure 20). The clusters differed by an inversion of the *mab-5* and *egl-5* pair, but in *P. sambesii* the order of the loci corresponds to the model structure, albeit the transcriptional order does not correspond (Figure 20). The HOX9-13 locus *php-3* was located on a different genomic scaffold in *Plectus sambesii* (Figure 20).

3.2.13 Hox loci in *Trichuris* are linked together in *CiS* and adhere to the model order

In the instance of *Trichuris suis*, a member of the Dorylaimia (Clade I) lineage, a cluster covering about 100 kb was detected, possessing six Hox loci that were interconnected and transcribed from distinct DNA strands (Figure 20). The Hox gene order follows the ancestral cluster arrangement without inversions between anterior, central and posterior paralogs. The six Hox genes, *ceh-13*/Hox1, Hox3, *lin-39*/Hox4, *mab-5*/Hox6-8, *Antp*/Hox6-8 and *php-3*/Hox9-10 belong to five ancestral orthologous groups (Figure 20).

3.2.14 Hox cluster arrangement in Nematomorpha

The Hox genes in *Gordionus montsenyensis* exist as a fragmented cluster (Figure 20); anterior paralogs (Hox1(*lab*) and Hox2(*Pb*)) are separated from the three

central group paralogs (Hox4 *Dfd*), Hox6-8 (*Antp*) and Hox6-8 (*Abd-A*)) by approximately 5.5 Mb (Figure 20). The posterior Hox9-13 (*Abd-B*) is further away on the same scaffold by more than 4 Mb (Figure 20). The order of the genes in the core does reflect the expected HOX1-HOX6-8 ordering except that HOX6-8/*AbdA* and Hox6-8/*Antp* genes are inverted with respect to the order found in the ancestral cluster. Hox genes in *Gordius aquaticus* are linked on scaffold 5 (Figure 20). *Acutogordius australiensis* has members of Hox genes present in 3 separate genomic scaffolds. Scaffold_2 contains Hox4 (*Dfd*), scaffold_3 contains 4 Hox loci, scaffold_4 contains Hox6-8 (*Abd-A*).

3.2.15 Hox clusters in nematodes are interrupted by non-Hox genes

Although vertebrate genomes are typically larger, the arrangement of vertebrate Hox genes is notably compact, with Hox clusters spanning a smaller region compared to many invertebrates. These clusters exhibit low repeat densities and exclude protein-coding genes unrelated to the Hox family. One of the few gene types that are found inside the clusters is small RNA genes (microRNAs) that have roles in Hox gene regulation. This suggests a strong constraint on the Hox cluster in vertebrates, likely because of the intimate co-regulation of the Hox loci. I explored whether similar constraints act on nematode Hox clusters. I identified non-Hox genes between Hox loci in nematode species analyzed (Table 2). Where Hox loci were tightly clustered, the number of genes was reduced, but even in the smallest clusters (*B. xylophilus*, *P. sambesii*, *T. muris*), non-Hox loci were common (Table 2) and (Figure 20).

Table 2: Hox cluster features of Nematoda and Nematomorpha from available high-quality genomes.

Taxon	Hox loci	Genome size (Mb)	Genomic size (Mb)	Cluster loci (kb)	Number non-HOX
<i>C. elegans</i>	ChrIII	100	14	282	49
<i>B. xylophilus</i>	BXYJ5_Chr3	78	13	28	4
<i>B. okinawensis</i>	BOKI_Chr3	70	12	54	8
<i>B. malayi</i>	Bm_v4_Chr1_scaffold_001	90	15	186	8
<i>O. volvulus</i>	OVOC_OM3	96	17	187	10
<i>P. sambesii</i>	scaffold_1	187	22	29	7
<i>S. pulchra</i>	ptg0000571	109	1	52	5
<i>T. suis</i>	T_suis	64	4	89	15
<i>A. australiensis</i>	scaffold_3	201	38	—	288
<i>G. montsenyensis</i>	SUPER_6	288	32	—	471
<i>Gordius aquaticus</i>	Scaffold_5	288	26	—	402

3.2.16 Hox clusters in nematodes are interrupted by miRNAs

MicroRNAs play pivotal roles in post-transcriptional regulation, orchestrating mRNA degradation or translational repression. I found miRNA presence within diverse nematode HOX clusters: dpu-miR-993 was identified in both *B. malayi*

and *O. volvulus*; meanwhile, *T. suis*, *P. sambesii*, and *S. pulchra* exhibited dpu-miR-993. Notably, no microRNAs were detected in the HOX clusters of *B. xylophilus* and *B. okinawaensis*. Additionally, the analysis of *C. elegans* revealed the presence of cel-miR-231-3p, consistent with prior findings of the cel-miR-231 microRNA family's localization between the *lin-39* and *ceh-13* Hox loci in *C. elegans* (Kuntz et al., 2008).

3.3 Evolution of DNA methylation genes across the Nematoda

An exhaustive investigation was conducted to explore the evolution of crucial enzymes involved in DNA methylation. This study utilized homology-based techniques to identify cytosine DNA methyltransferases (DNMTs), including DNMT2 (also known as Human TRDMT1 or tRNA methyltransferase), across a range of published and newly acquired nematode genomes spanning equally distributed across the entire phylum, supplemented with four nematomorph species as outgroups. Additionally, the complement of UHRF1 domains essential for DNA methylation maintenance was investigated (Figure 21).

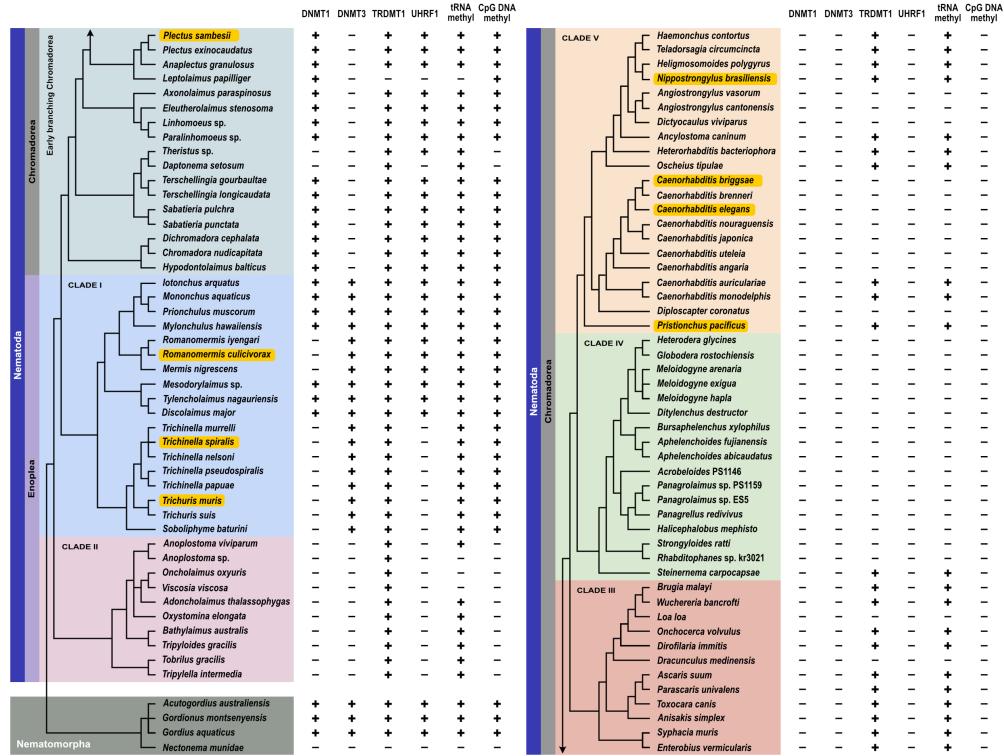


Figure 21: Nematoda and Nematomorpha DNMT complements. The cladogram on the left shows the phylogenetic relationships of the analysed nematode species based on (Ahmed et al., 2022; Ahmed & Holovachov, 2021; Qing et al., 2023). Species where either DNMT1, DNMT2 or DNMT3 presence has been previously investigated shown in yellow.

3.3.1 Distinct patterns of DNMT and TRDMT1 presence in Enoplean nematodes (Clades I-II)

The Enoplea class encompasses ancestral lineages within the phylum Nematoda, with Clades I-II representing major constituents of Enopleans. In Clade I, DNMT1 was present in all free-living species belonging to Mononchida (represented by *Myionchulus hawaiiensis*, *Prionchulus muscorum*, *Mononchus aquaticus* and *Iotonchus arquatus*) and Dorylaimida (*Discolaimus major*, *Tylencholaimus nagauriensis* and *Mesodorylaimus* sp.). DNMT1 was absent in parasitic Mermithidae (*Mermis nigrescens*, *Romanomermis culicivora* and *Romanomermis iyen-gari*), *Trichinella*, *Soboliphyme baturini*, *Trichuris muris* and *Trichuris suis*. DNMT3 and TRDMT1 were both identified in all analyzed Clade I species. However, UHRF1 was absent in studied nematode species from *Trichinella* and *Soboliphyme* (see Figure 21). Interestingly, all Clade II nematodes analyzed lacked both DNMT1 and DNMT2, yet they universally possessed TRDMT1, a tRNA (cytosine-5-)-methyltransferase. Additionally, HRF1 was not detected in any Clade II nematodes 21).

3.3.2 Presence of DNMT1 in Chromadorean nematodes (including Clades III-V)

In early branching Chromadorean species, DNMT1 was consistently present, except for *Daptonema setosum* and *Theristus* sp. while DNMT3 was absent across all these species. Moreover, *Daptonema setosum* lacked TRDMT1, while both TRDMT1 and UHRF1 were absent in *Leptolaimus papilliger*, as illustrated in Figure 21. In Clade III, a distinctive pattern emerged where TRDMT1 (DNMT2) was identified in 10 species, whereas DNMT1 and DNMT3 were conspicuously absent (Figure 21). Within Clade IV, none of the sixteen species examined exhibited DNMTs or UHRF1; however, TRDMT1 (DNMT2) was detected in basally branching *Steinernema carpocapsae* (Figure 21). In Clade V nematodes, which encompass well-studied model species such as *Caenorhabditis elegans* and the satellite model *Pristionchus pacificus*, 11 instances of TRDMT1 (DNMT2) loss were detected, while TRDMT1 was present in 10 species. Notably, DNMT1, DNMT3, and UHRF1 were absent (Figure 21).

3.3.3 DNA methyltransferases are conserved in Gordioidea nematomorphs

The small phylum Nematomorpha was represented by merely four species with recently available genomes, spanning both limnoterrestrial (Gordioidea) and marine (Nectonematoidea) lineages. All examined nematomorphs from the limnoterrestrial Gordioidea exhibited the presence of DNMT1, TRDMT1 (DNMT2), DNMT3, and UHRF1. However, the only marine species, *Nectonema munidae*, showed no evidence of DNMT1, DNMT3, or tRNA (cytosine-5-)-methyltransferase enzymes. Additionally, HRF1 was not identified in this species (Figure 21).

3.4 Comparative genomics of parasitoid lifestyle as exemplified by Mermithidae and Nemathomorpha

3.4.1 Orthologous clusters in parasitoid worms

Ortholog cluster identification was employed to investigate the genetic signatures encoded in the genomes of both parasitoid taxa to finding similarities and parallels linked to their ecological lifestyles and establish new model systems. This approach allowed the comparison of evolutionary relationships between genes across different species. After the identification and functional annotation of orthologous clusters, a set of 335 highly conserved orthologous clusters were identified to be unique to mermithids and 263 to Nematomorpha respectively (Figure 22).

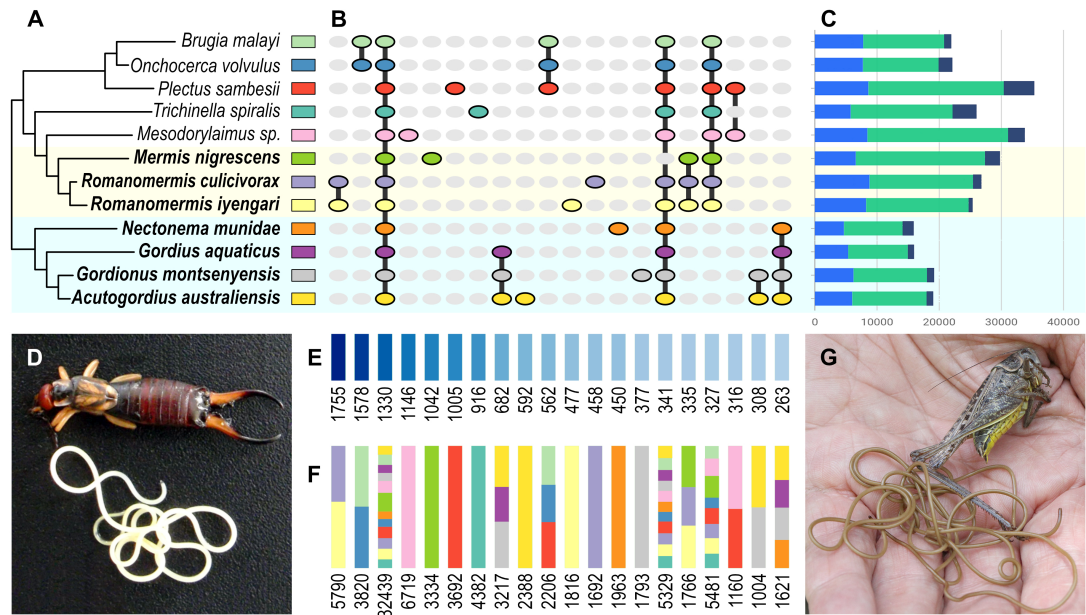


Figure 22: Orthologous cluster analysis: **(A)** Phylogenetic relationships between studied species (color label for B and F); **(B)** Distribution of shared and unique orthologous gene clusters among studied species, colors correspond to species in A; **(C)** Number of protein sequences (blue), orthologous clusters (green) and singletons (dark purple); **(D)** Mermithid *Mermis nigrescens* emerging from its host (photograph by Bronwen Presswell, reproduced with permission); **(E)** Number of shared orthologous clusters for each subset of species or the number of unique clusters; **(F)** Combined number of proteins in shared and unique clusters, colored segments represent relative proportion of proteins for each species, colors correspond to species in A; **(G)** Nematomorph emerging from its host (photograph by Alastar Rae, reproduced here under CC BY-SA 2.0 license).

3.4.2 Mermithid orthologous clusters

Mermithid orthologous clusters were over-represented for 40 Gene Ontology (GO) terms. Here the top 20 enriched GO terms based on $e\text{-value} < 1e5$ are presented (3), however, additional enriched functional signatures were observed including regulation of DNA methylation (GO:0044030), G-protein coupled receptor activity (GO:0004930), chloride channel activity (GO:0005254). A comprehensive list of all enriched GO terms is provided as supplementary material (File_S2_Mermithids_orthologs).

Table 3: Top 20 GO functional terms enriched in Mermithid gene family orthologous clusters.

GO ID	Description	p value
GO:0007165	signal transduction	1.02E-69
GO:0006511	ubiquitin-dependent protein catabolic process	2.85E-53
GO:0016192	vesicle-mediated transport	2.40E-37
GO:0015031	protein transport	6.25E-33
GO:0007283	spermatogenesis	1.81E-23
GO:0007601	visual perception	1.63E-22
GO:0046872	metal ion binding	2.41E-21
GO:0008270	zinc ion binding	1.80E-20
GO:0006508	proteolysis	3.63E-19
GO:0016567	protein ubiquitination	5.38E-17
GO:0006099	tricarboxylic acid cycle	1.13E-13
GO:0006468	protein phosphorylation	3.3E-12
GO:0006814	sodium ion transport	9.83E-11
GO:0034765	regulation of ion transmembrane transport	2.39E-8
GO:0005634	nucleus	4.61E-8
GO:0097225	sperm midpiece	4.62E-8
GO:0004867	serine-type endopeptidase inhibitor activity	1.05E-7
GO:0007224	smoothed signaling pathway	5.74E-7
GO:0046676	negative regulation of insulin secretion	4.67E-7
GO:0034976	response to endoplasmic reticulum stress	1.83E-5

3.4.3 Nematomorpha orthologous clusters

Nematomorpha orthologous gene clusters were over-represented for 76 Gene Ontology (GO) terms. The top 20 enriched orthologous clusters were enriched in GO terms shown in (Table 4). Other enriched orthologous clusters beyond the top 20 were enriched in sodium ion transport (GO:0006814), response to toxic substance (GO:0009636), response to starvation (GO:0042594), and sensory perception of sound (GO:0007605). A complete list is shown in the supplementary material (File_S3_Nematomorphs_orthologs).

Table 4: Top 20 GO functional terms enriched in Nematomorph gene family orthologous clusters.

GO ID	Description	p value
GO:0007165	signal transduction	5.72E-142
GO:0006511	ubiquitin-dependent protein catabolic process	5.19E-87
GO:0016192	vesicle-mediated transport	1.11E-62
GO:0016021	integral component of membrane	7.67E-62
GO:0006355	regulation of transcription, DNA-templated	2.54E-59
GO:0015031	protein transport	4.36E-56
GO:0007275	multicellular organismal development	2.10E-49
GO:0008270	zinc ion binding	3.45E-47
GO:0006508	proteolysis	1.10E-46
GO:0016032	viral process	6.19E-41
GO:0016055	Wnt signaling pathway	2.62E-39
GO:0006468	protein phosphorylation	7.69E-36
GO:0006357	regulation of transcription from RNA polymerase II promoter	2.89E-34
GO:0007601	visual perception	7.13E-34
GO:0016567	protein ubiquitination	4.72E-26
GO:0006486	protein glycosylation	1.38E-21
GO:0030433	ER-associated ubiquitin-dependent protein catabolic process	2.45E-20
GO:0006886	intracellular protein transport	5.58E-17
GO:0071805	potassium ion transmembrane transport	7.80E-17
GO:0034765	regulation of ion transmembrane transport	3.67E-16

3.4.4 Expanded gene families in mermithids and nematomorphs

In Mermithidae 226 expanded gene families and 114 contracted gene families were identified in *Mermis nigrescens*, 89 expanded gene families and 122 contracted gene families in *Romanomermis culicivorax* and 96 expanded gene families and 145 contracted gene families in *Romanomermis iyengari* (Figure 23). However, within the lineage leading to the Mermithidae family, 169 gene families underwent expansion and 40 gene families contracted (Figure 23).

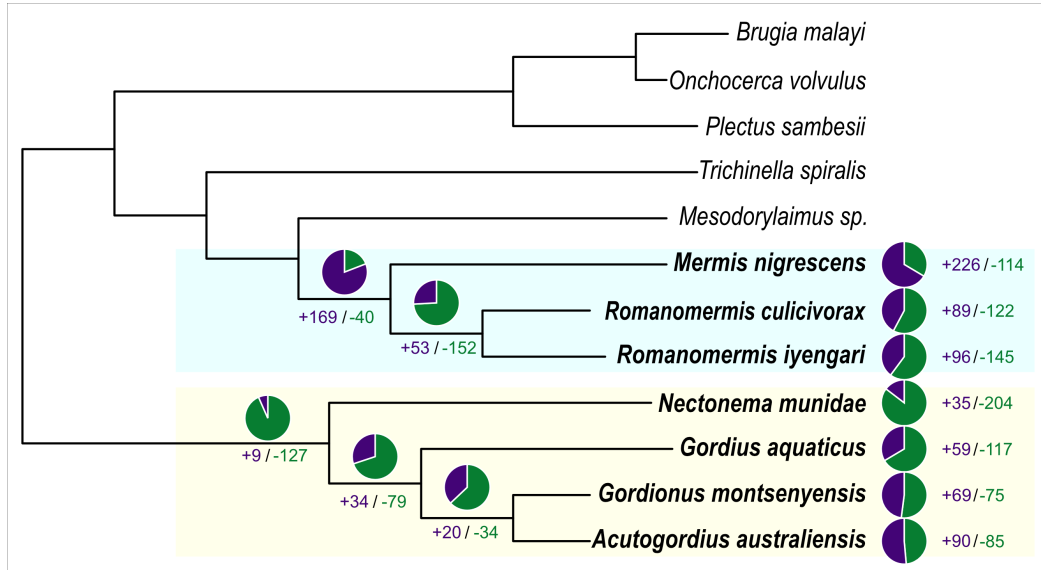


Figure 23: Pie charts show gene family size. The number of contracted (green) and expanded (Purple) gene families. The Cladogram on the left shows the evolutionary history of gene families and differences between species included in the analysis.

3.4.5 Mermithid expanded and contracted gene family clusters

Table 5 displays the functional annotation with GO terms of the expanded gene families predominantly enriched in mermithid nematodes. Though the top 20 are shown, mermithids exhibited enrichment in various other signatures, including norepinephrine transport (GO:0015874), motor activity (GO:0003774), neurotransmitter transport (GO:0006836), response to toxic substance (GO:0009636), response to oxidative stress (GO:0006979), regulation of synapse structural plasticity (GO:0051823), telomere maintenance via recombination (GO:0000722), response to starvation (GO:0042594), spermatogenesis (GO:0007283), DNA recombination and GABA-A receptor activity (GO:0004890). They are shown in the supplementary material (File_S4_mermithid_expanded_gene_families). The top five contracted gene families were enriched in the structural constituent of the

cuticle (GO:0042302), protein heterodimerization activity (GO:0046982), nucleosome assembly (GO:0006334), macromolecular complex binding (GO:0044877), peptidyl-proline hydroxylation to 4-hydroxy-L-proline (GO:0018401). A comprehensive list of all contracted gene families is given in the supplementary material (File_S6_mermithid_contracted_gene_families).

Table 5: Top 20 GO functional terms enriched in Mermithid expanded gene family clusters.

GO ID	Description	p value
GO:0019346	transsulfuration	9.73E-82
GO:00304311	sleep	1.45E-56
GO:0006412	translation	2.20E-58
GO:0006511	ubiquitin-dependent protein catabolic process	3.21E-51
GO:0015267	channel activity	4.63E-49
GO:0005198	oligosaccharide metabolic process	3.99E-47
GO:0005198	structural molecule activity	3.84E-43
GO:0005993	trehalose catabolic process	1.31E-42
GO:0005991	trehalose metabolic process	1.27E-40
GO:0007017	microtubule-based process	3.41E-40
GO:0016185	synaptic vesicle budding from presynaptic membrane	4.59E-38
GO:0016021	integral component of membrane	1.74E-36
GO:0015144	carbohydrate transmembrane transporter activity	1.02E-32
GO:0018107	peptidyl-threonine phosphorylation	2.84E-32
GO:0006556	S-adenosylmethionine biosynthetic process	6.67E-32
GO:0001505	regulation of neurotransmitter levels	9.84E-32
GO:0006833	water transport	6.67E-31
GO:0003774	motor activity	5.77E-29
GO:0007016	cytoskeletal anchoring at plasma membrane	8.69E-29
GO:0045332	phospholipid translocation	8.72E-29

3.4.6 Nematomorphs expanded and contracted gene families

In nematomorphs, the expanded gene families were most enriched in GO terms shown in Table 6. Additionally, several others were significantly enriched in for example norepinephrine transport (GO:0015874) and neurotransmitter transport (GO:0006836), along with DNA-mediated transposition (GO:0006313), spermatogenesis (GO:0007283). Further details on these findings can be found in the supplementary material (File_S5_nematomorphs_expanded_gene_families). The top five contracted gene families based on e-value $< 1e5$ were enriched in structural constituent of the cuticle (GO:0042302), microtubule-based process (GO:0007017), translation (GO:0006412), channel activity (GO:0015267) and oligosaccharide metabolic process (GO:0009311). A full list of all contracted gene families is in supplementary material (File_S7_nematomorphs_contracted_gene_families).

Table 6: Top 20 GO functional terms enriched in Nematomorph expanded gene family clusters.

GO ID	Description	p value
GO:0006508	proteolysis	0.0
GO:0032197	transposition, RNA-mediated	6.90E-52
GO:0015074	DNA integration	1.27E-50
GO:0060012	synaptic transmission, glycinergic	5.59E-47
GO:0005328	neurotransmitter:sodium symporter activity	1.99E-43
GO:0004252	serine-type endopeptidase activity	2.25E-34
GO:0006979	response to oxidative stress	4.38E-30
GO:1901216	positive regulation of neuron death	1.49E-27
GO:0015838	amino-acid betaine transport	1.68E-26
GO:0006874	cellular calcium ion homeostasis	2.67E-24
GO:0048771	tissue remodeling	8.60E-19
GO:0009636	response to toxic substance	8.78E-17
GO:0051919	positive regulation of fibrinolysis	1.55E-15
GO:0061537	glycine secretion, neurotransmission	1.05E-14
GO:0007586	digestion	3.51E-14
GO:0006310	DNA recombination	1.06E-11
GO:0051610	serotonin uptake	1.30E-10
GO:0006457	protein folding	1.44E-8
GO:0075732	viral penetration into host nucleus	5.88E-7
GO:0040017	positive regulation of locomotion	4.08E-6

3.4.7 Transposable elements

TE and repeat divergence landscapes are useful to infer recent and past activity of TEs, as TEs accumulate mutations and degrade over time. The Kimura substitution level is a measure of divergence to the TE consensus sequence. Highly diverged sequences are likely to have lost the ability to transpose and become inactive remnants. In this analysis, the profiles of *Gordionus montsenyensis*, *Mesodorylaimus sp.*, *Nectonema munidae* and *Romanomermis iyengari* display a history of high TE activity and even recent bursts, only *Gordionus montsenyensis* shows a decreasing TE activity. *Plectus sambesii* and *Trichinella spiralis* reveal striking recent activity, while *Brugia malayi*, *Mermis nigrescens* and *Romanomermis culicivorax* show intermediate increases in recent activity. *Acutogordius australiensis*, *Brugia malayi* and *Onchocerca volvulus* display low TE activity. Among the Mermithidae family, *R. iyengari* exhibits greater TE activity compared to both *M. nigrescens* and *R. culicivorax*. Conversely, *M. nigrescens* possess higher activity than *R. culicivorax* (Figure 24).

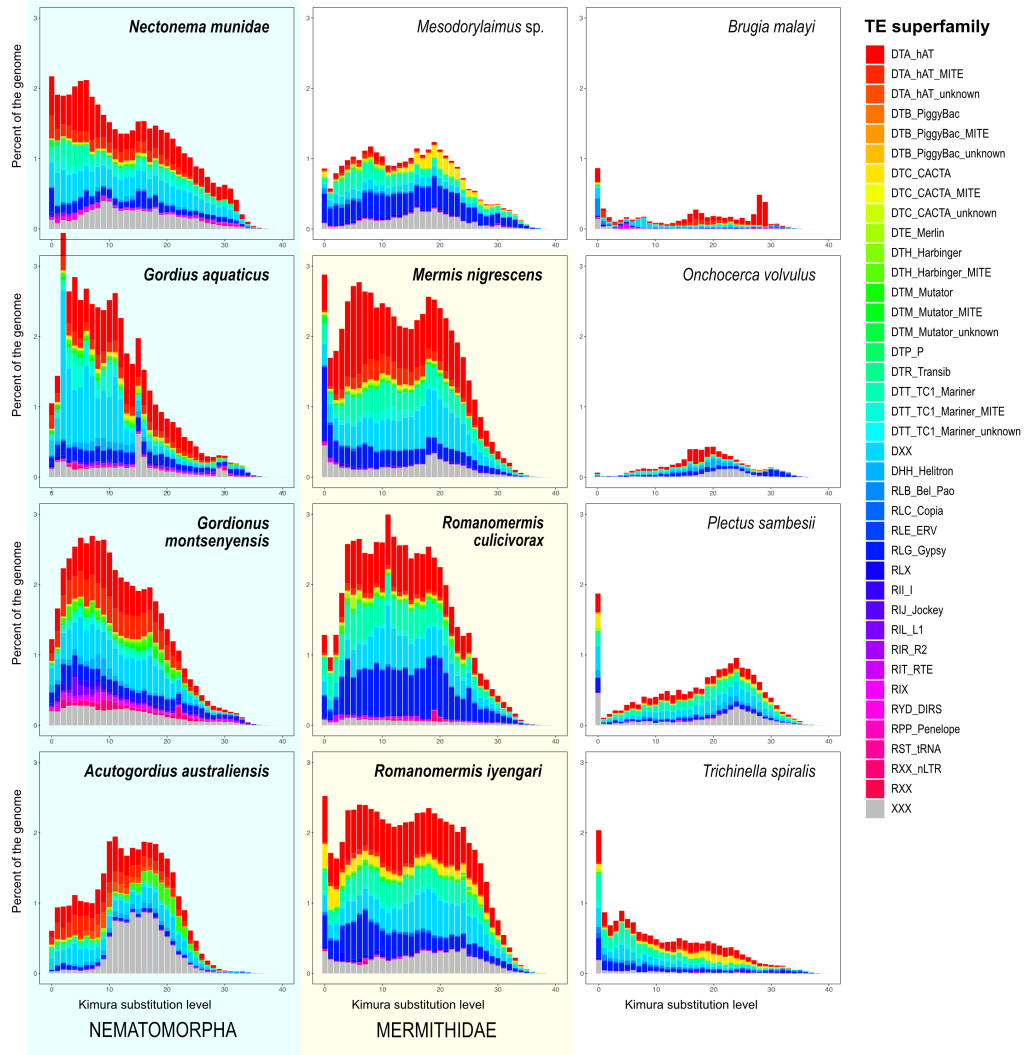


Figure 24: TE and repeat divergence landscapes showing TE portions as a percent of the genome (y-axis) over Kimura substitution levels (x-axis) obtained via the “FasTE” pipeline. Legend for TE superfamily colours is shown to the right.

All studied species of Nematomorpha have a comparable genome size between 200-300 Mb, however, *Nectonema munidae* shows an inflated amount of TEs, compared to other species. In the mermithids, *Romanomermis culcivorax* and *Mermis nigrescens* show amounts of TEs of 4.04% and 6.77% respectively, within the range of 2-16% for the other species, while *Romanomermis iyengari* displays a highly inflated amount of TEs of 53.9% (Figure 25). Comparing the composition of TEs it becomes evident that DNA elements of the superfamilies *hAT*, *CACTA*, *Mutator*, *Transib* and *Mariner* are conserved between the species, while also containing a large portion of not further discriminated DNA elements. Further, all genomes contain Helitron elements and the composition of retroelements is dominated by the superfamily *Gypsy* (Figure 25).

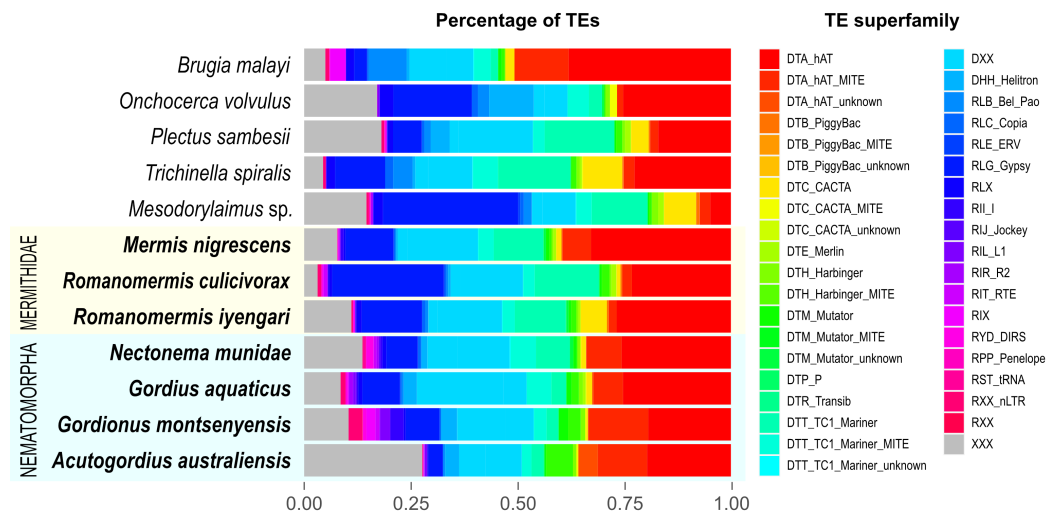


Figure 25: Stacked barplots show relative amounts of TEs within the genomes of species included in the analysis.

4 DISCUSSION

4.1 Genomes of early branching nematode species

A high-quality genome is essential for an in-depth understanding of the biology and evolutionary dynamics of a species. While advancement in long-read sequencing has made genome generation more feasible, the challenge lies in obtaining high-quality DNA from a diverse array of non-model nematodes across the phylum. This challenge is particularly pronounced in microscopic nematodes that cannot be cultured, resulting in sampling bias in genomic sequencing and the lack of valuable biological information for phylum-wide genetic and evolutionary studies. However, the innovative protocols such as PacBio ultra-low input sequencing, the establishment of HMW DNA extraction protocols and establishment of new cultures in the lab enabled the generation of high-fidelity data for *Tobrilus gracilis*, *Mononchus aquaticus* and *Tripylella intermedia* in particular (Table 1). PacBio HiFi sequencing with ultra-low input yielded data of over 30 GB, despite being derived from only a small number of individuals, resulting in high-quality draft assemblies. This method, based on amplification, proves advantageous when the input amount of DNA of a species is minuscule, particularly applicable to nematodes that cannot be cultured. However, it is crucial to acknowledge that amplification protocols may introduce biases in the sequencing data. To gain a better understanding of the implications of such biases on assemblies, future experiments could involve generating PacBio HiFi reads from a diverse pool of individuals without amplification (Guiglielmoni et al., 2024). Incorporating long-range sequencing techniques, such as chromosome conformation capture, will likely facilitate the scaffolding of these high-quality assemblies into potential chromosome candidates, enabling a more in-depth exploration of genome structures. Further improvements of these genomes will also require a combination of both PacBio HiFi and Nanopore long reads to boost the contiguity.

Indeed we used this approach to generate recent long-read assembly for *Romanomermis culicivorax* (Guiglielmoni et al., 2024). This genome represents a significant enhancement over earlier short-read-based assembly (Schiffer et al., 2013), offering increased contiguity, improved resolution of repetitive regions, and more accurate gene annotation (Guiglielmoni et al., 2024). The important improvement in this assembly lies in the enhanced resolution of repetitive regions, revealing a broader array of transposable elements. This observation underscores the importance of long-read assemblies in uncovering previously overlooked genome segments.

The BUSCO scores for both *R. culicivora* and *R. iyenagri* against the Nematoda lineage is relatively low, with values of only 39.4% and 45.0%, respectively. Notably, the Nematoda reference dataset is composed of seven nematode species, among which only one is an early-branching nematode, *Trichinella spiralis*. This scarcity of representation of early-branching nematodes likely contributes to the decreased BUSCO completeness observed in genomes of early-branching nematode species compared to those of higher clades, highlighting a taxonomic bias in currently available quality control tools, which are not truly representative of the diversity of the phylum Nematoda. Additionally, ortholog divergence may also play a role in the lower BUSCO scores, warranting further investigation. The genomic data generated in this study will serve as a valuable resource for guiding future sequencing endeavors focused on understudied nematodes, ultimately aiding in the completion of available genome assemblies to chromosome-scale and addressing gaps in our understanding of nematode genomics.

4.2 Evolutionary plasticity in Nematode Hox gene complements and genomic loci arrangement

Hox genes are key players in the structuring of animal body plans during development. The developmental role of Hox genes has been well studied in *C. elegans* and the satellite model nematode species *Pristionchus pacificus* (Bürglin & Ruvkun, 1993; Van Auken et al., 2000; Aboobaker & Blaxter, 2003; Grandien & Sommer, 2001). Our knowledge of the Hox gene repertoire, Hox Cluster organisation, and linkage in most nematode lineages was limited until now, with data restricted to some parasitic species scattered across Nematoda (Aboobaker & Blaxter, 2003). Most published nematode genomic resources came from Clades I, III, IV and V often missing a higher representation of free-living nematodes in Clade II and early branching Chromadorea (M. Blaxter, 2011; Ahmed et al., 2022). Determination of the full complement of Hox genes in recently available nematode genomes across the phylum is fundamental to understanding the link between the genomic and morphological evolution of these animals. I have re-visited the Hox gene complements of Nematoda not only to understand the evolution of this key set of body pattern genes using newly available high-quality genomes but also to classify the loci into their orthologue groups as well as to look for linkage between loci, non-Hox genes and miRNAs interspersing HOX genes in a broader phylogenetic context.

Nematodes are renowned for their simple and highly conserved body plan. However, at the sequence level, they have been shown to possess rapid homeodomain sequence and genome evolution which might be a general feature of

Nematoda Hox genes (Aboobaker & Blaxter, 2003). I observe that it is difficult to fully prove the absence of a gene and that the missing Hox genes in some genomes could indeed be truly missing from these genomes or they could be a result of extreme homeobox sequence divergence such that the sequence similarity search methods used in this study could not detect extreme sequence variability. In *C. elegans*, a reduced Hox gene complement (6 genes in 4 orthology groups) has been reported (Aboobaker & Blaxter, 2003). This study confirms this reduction (Figure 18), illustrating the robustness of the analysis methods. In *Syphacia muris*, *Enterobius vermicularis*, as well as the common ancestor of Clade IV-V nematodes, an independent loss of Hox3 had been reported (Yoshida et al., 2017). This loss is confirmed in this study as well (Figure 18). This further supports the fidelity of the analysis methods employed here for detecting gene presence and absence in assembled genomes.

We have expanded the availability of Hox gene complements for a fairly large number of nematode species that were previously missing due to a lack of genomic resources. The Hox complement across the phylum Nematoda is shaped by key gene losses. We found a recent duplication of *nob-1* in the model genus. This case illustrates the potential for new in-paralogues to evolve in certain lineages. In *C. elegans*, loss of *nob-1* and *php-3* functions leads to severe abnormalities during embryonic development, affecting both the arrangement of posterior structures and the movement of the posterior hypodermis. Additionally, this disruption results in changes from posterior to anterior cell identities and ultimately leads to non-viability (Van Auken et al., 2000). Further examinations are required to establish if these roles persist across different *Caenorhabditis* species that possess *nob-1*.

We found the presence of Hox3 in Clades I-III of Nematoda and the absence in both Clade IV and V of Nematoda and Nematomorph genomes analyzed here (Figure 18). The pattern of presence and absence in Nematoda can be explained by an independent derivation in the respective Clades. Within the insect counterparts, orthologs of Hox3 (*zerknüllt*, *zen*) have undergone repeated functional alterations, resulting in the relinquishment of their conventional Hox role and subsequent reassignment of their functional domain from embryonic to extraembryonic tissues. All known *zen* genes have been found to participate in the development of extraembryonic membranes (EEMs). These EEMs serve to shield embryos from external environmental challenges, and their formation has enabled insects to deposit eggs in diverse habitats, ultimately facilitating their successful colonization of terrestrial environments (Hughes, Liu, & Kaufman, 2004). Functional investigations are necessary to clarify the function of Hox3 in nematodes that possess this gene.

Previous analysis using transcriptome datasets and PCR surveys identified loss of *ant-1* locus in the Enoplea species *Enoplos brevis*, the Chromadorea *Plectus sambesi*, *Syphacia muris*, the ancestor of Tylenchomorpha, the diplogasteromorph *Pristionchus pacificus*, and the ancestor of *Caenorhabditis* (Yoshida et al., 2017). We now certify this pattern of presence-absence using currently available high-quality genomes from these species and further extend the loss of *ant-1* locus in the pinworm *Enterobius vermicularis* (Clade III). In *Brugia malayi*, it has been demonstrated that the antennapedia gene’s positioning occurs after the *Bm-egl-5* Hox gene. Moreover, their homeodomain exons are alternatively joined to the same 5’ exon through *cis* splicing and this arrangement might signify an intermediate stage in the potential loss of Hox genes due to redundancy in other nematode species (Aboobaker & Blaxter, 2003). Furthermore, *ant-1* was identified in *Caenorhabditis monodelphis* in contrast to prior PCR survey-based studies that reported its absence (Yoshida et al., 2017). The presence of *ant-1* exhibited variability in Clade V nematodes (Figure 18). Nevertheless, additional investigation is required to substantiate this distribution pattern.

We found *Hox2/pb*, *Hox5/scr* and *Hox6-8/Ubx/AbdA* missing in Nematoda (Figure 18). The missing genes in newly sequenced nematode genomes further validate the previous findings from PCR-based studies (Yoshida et al., 2017; Aboobaker & Blaxter, 2003). The finding of *Hox2/pb* and *Hox6-8/AbdA* in Nematomorpha (Figure 18), which are closest relatives to Nematoda, indicates that these Hox loci were lost secondarily in the lineage leading to Nematoda. Recent work suggests much more gene loss has occurred in metazoan genome evolution than previously thought (Fernández & Gabaldón, 2020; Guijarro-Clarke, Holland, & Paps, 2020; Cunha et al., 2023), and Hox gene loss has been observed in several other lineages (Yoshida et al., 2017). The absence of *Hox5/Scr* in both Nematoda and Nematomorpha not only suggests an inheritance characteristic from the common ancestor of all extant Nematoidea (a clade uniting Nematoda and Nematomorpha) but also a potential example of evolutionary conservation.

Why some Hox genes were lost early on in the evolution of the phylum and others were lost or duplicated later on remains unclear. We found that some nematodes possess *Hox3* and antennapedia-like Hox gene *ant-1* which belong to orthology groups present in other metazoan animals but are absent in *C. elegans* and some of its relatives (Figure 18). The loss may be due to a more specialized function that is only needed in certain nematode species or developmental contexts. For instance, *Hox3* or *Hox6-8/Antp* may be involved in the development of specialized structures that are unique to those non-model nematode species and if these structures are lost during evolution, the respective Hox genes may also become dispensable and eventually lost. In contrast, other Hox genes such

as *ceh-13* which is critical for embryogenesis and influences cell fate specification beyond the anterior body part (Van Auken et al., 2000; Tihanyi et al., 2010) and *php-3* which regulates specification of the tail (Zheng, Lee, & Pham, 2022) might have more essential functions that are needed for proper development of nematodes. These genes are less likely to be lost because their loss would have a more significant impact on the nematode’s development and survival. Therefore, the Hox complement in Nematoda is shaped by gene loss and might have played an important role in the evolution of their simplified body plan.

While the *C. elegans* Hox cluster is dispersed, we found intact and not dispersed Hox clusters in several genomes of non-model nematode species (Figure 20). We found rearrangement, translocation, and disruption by inversion with respect to the transcriptional orientation of the Hox genes in the clusters of analyzed nematode genomes (Figure 20). This pattern of mixed arrangement reflects a highly dynamic and complex history of nematode Hox cluster evolution and organization similar to a situation observed in *Drosophila melanogaster* where the Hox cluster is broken into parts (the Antennapedia (ANTC) and bithorax (BXC) complexes). It exhibits aberrations in the transcriptional orientation of some genes as well as both interspersed genes of independent origin and Hox-derived genes that have evolved novel developmental functions (Ferrier & Minguillon, 2003). Additional Hox cluster rearrangements have also been reported in other *Drosophila* species (Negre & Ruiz, 2007; Mulhair et al., 2023; Von Allmen et al., 1996; Negre, Ranz, Casals, Cáceres, & Ruiz, 2003) as well as in the silk moth *Bombyx mori* (Yasukochi et al., 2004). The rearrangement of Hox genes in the Cluster of *C. elegans* has also been reported before (Aboobaker & Blaxter, 2003). This further suggests that the physical orientation and rearrangement of the Hox genes in the clusters are not crucial for normal nematode development and body plan conservation. The dynamics in the organization of the nematode Hox clusters could be further evidence that strong evolutionary constraints are rather acting on the highly conserved functional domains of the nematode Hox genes to maintain the conserved body plan. Hox gene rearrangements may drive evolutionary changes causing alterations in the pattern of Hox gene expression during development. They could facilitate the evolution of new morphological structures that are characteristic of different species thereby contributing to the process of nematode speciation. As the complete genome sequences of more nematode species become available, the molecular phylogenetics of Hox cluster evolution will become clearer with an emphasis on understanding the evolution of the regulatory modules that partition Hox expression domains.

We identified non Hox genes between Hox loci (Table 2) which may act as Hox cluster co-regulator candidates. Previous works identified T-box proteins as

being particularly amplified in Rhabditid nematodes (Okkema, 2017; Papaioannou, 2014). T-box genes encompass a group of transcription factors that possess a strongly conserved DNA-binding domain called the T-domain, and they are associated with the control of various developmental processes in multicellular organisms. I found TBX10 in *Brugia malayi* and *Onchocerca volvulus* parasitic nematodes (Figure 20). Tbx10 has been shown to have retained ancestral roles in specific developmental processes, such as ventral somite specification in the tail region in mouse embryogenesis (Bush, Maltby, Cho, & Jiang, 2003). Additionally, the presence of "foreign" genes in Hox clusters of nematodes indicates that the distances between Hox genes are quickly evolving (Dillman et al., 2015).

The identification of potentially unified Hox clusters in species occupying phylogenetic positions close to or in the early branching Clades is an indicator that perhaps the common ancestor of all nematodes had a single Hox cluster of at least six Hox genes as observed in Trichuridae (*Trichuris suis*) and *Plectus sambesii* and that these clusters have undergone remodification in different nematode lineages as a result of gene loss and Hox paralog-specific duplication. The Nematoda Hox cluster organization is highly dynamic. If present at all, the arrangement of genes in a cluster can be interrupted, shuffled, or even completely disintegrated. Further knowledge on topological islands in the analyzed nematode Hox gene clusters herein await an investigation.

MicroRNAs are known to play crucial roles in post-transcriptional regulation of gene expression, often by targeting mRNAs for degradation or translational repression (Bartel, 2004). The presence of asu-miR-993-3p in *Brugia malayi* and *Onchocerca volvulus*, dpu-miR-993 in *Plectus sambesii*, *Sabatieria pulchra* and *Trichuris suis* Hox clusters underscore their conservation within closely related nematode species. This conservation of microRNAs implies functional importance. These findings are consistent with previous studies highlighting microRNA presence in *C. elegans* (Ambros, Lee, Lavanway, Williams, & Jewell, 2003; Lim et al., 2003). The microRNA family cel-miR-231 has been previously reported in *ceh-13/lin-39* Hox locus (Kuntz et al., 2008). While these microRNAs have been identified within nematode Hox clusters, their specific regulatory targets and functional implications within non-model nematode development remain to be elucidated.

In contrast to the nematode body plan's conservation, essential bilaterian regulators like Hox genes have undergone significant evolutionary flexibility. The loss of some of these genes in the lineage leading to the Ur-Nematode might have played a role in roundworm body plan evolution. While the *C. elegans* Hox cluster is dispersed, I found species including *Plectus sambesii* and *Trichuris suis* in which the cluster is intact and not dispersed but interrupted with non-Hox

genes nonetheless. The rearrangement of Hox genes in nematode Hox clusters may have underlying effects including the breakage of the usual collinearity of Hox gene expression in Nematoda. Further research regarding nematode Hox gene complements that I have identified in the present study should seek to understand whether and where the colinear expression is retained in the phylum. As more species become amenable to molecular methods (Schiffer et al., 2018, 2019; Hellekes et al., 2023), this can be done with tools like *in situ* hybridization, single-cell RNA-Seq studies, and CRISPR-guided gene knockouts. In summary, the findings suggest the potential for new and exciting work on Hox genes and regulation of Hox clusters in a broad sampling of Nematoda. New lineages with unified Hox gene clusters will likely emerge as more complete high-quality genomes become available.

4.3 Evolution of DNA methylation genes across the Nematoda

The current thesis investigated the evolution of genomic signatures of DNA methylation in two closely related phyla, Nematoda and Nematomorpha. Within the Chromadorea, particularly in Clade V, an absence of DNMT1 and DNMT3 was found. However, the presence of tRNA methyltransferase was observed in *Haemonchus contortus*, *Teladorsagia circumcincta*, *Heligmosomoides polygyrus*, *Nippostrongylus brasiliensis*, *Ancylostoma caninum*, *Heterorhabditis bacteriophora*, *Caenorhabditis auriculariae*, *Caenorhabditis monodelphis*, *Oscieus tipulae*, and *Pristionchus pacificus*, while it was absent in other species. Additionally, UHRF1 was found to be absent in all investigated Clade V species. In Clade IV nematodes, tRNA methyltransferase was detected solely in *Steinernema carpocapsae*, with the remaining species lacking DNMT1, DNMT3, tRNA methyltransferase, and UHRF1. Within Clade III, DNMT1 and DNMT3 were absent, but tRNA methyltransferase was present in all the species except in *Loa loa* and *Dracunculus medinensis*. Notably, among early-branching Chromadorea, a greater number of species were found to harbor DNMT1, while DNMT3 was notably absent. Moreover, tRNA methyltransferase and UHRF1 were commonly observed. In Dorylaimia (Clade I), DNMT3 and tRNA methyltransferase were present across all species, although DNMT1 and UHRF1 were not consistently detected. Conversely, within Enoplia (Clade II), DNMT1, DNMT3, and UHRF1 were absent, while tRNA methyltransferase was consistently present across all species (Figure 21).

This investigation represents the most extensive analysis of DNA methylation across a wide phylogenetic range within Nematoda and Nematomorpha. While

previous studies have explored DNA methylation in various Nematoda species, they have tended to concentrate on specific lineages (Gao et al., 2012; Rošić et al., 2018). In this investigation, I broadened these lineages to include closely related species and genera. However, I also greatly expanded the dataset by incorporating both published and recently generated nematode genomes from all nematode Clades, along with genomes from the closely related phylum Nematomorpha. The aim was to explore the evolution of DNA methyltransferase enzymes which are critical for a functional methylation machinery in as large and diverse set of species as reasonably possible. Notably, the presence of DNA methylation has not been explored in Nematomorpha, nor in early-branching Nematoda from Clade II or in paraphyletic early-branching Chromadoria.

The findings revealed the independent loss of specific DNMTs across multiple lineages (Figure 21). Proving the absence of an enzyme is challenging. There are several factors contributing to the potential failure in detecting a DNMT. A primary reason is the possibility of working with a low-quality, fragmented genome assembly. Thus, I utilized data from closely related species and employed parsimonious methods within the phylogenetic framework to confidently identify instances of loss: the absence of a DNMT in an entire clade of related species strongly suggests an actual loss event in their most recent common ancestor. When a homolog is absent in a species with well-assembled, high-quality genomes, especially in established model organisms like *Caenorhabditis elegans*, it enhances the credibility of a negative search result, thus affirming the robustness of the analysis methods used. In the present study, critical DNMTs (DNMT1 and DNMT3) were not identified in nematodes belonging to Clades III-V. This corroborates previous analyses of a limited number of nematode species within these same clades, which also failed to find DNMTs (Engelhardt et al., 2022; Rošić et al., 2018). Most of the genomes from these clades are highly contiguous and complete, negating the possibility of detection failure of the analysis pipeline. Concurrently, several studies have confirmed the absence of DNA methylation in many of these species, and no reports of DNMT enzymes have been documented. Hence, it is reasonable to infer that DNA methyltransferases are absent in nematodes belonging to Clades III, IV, and V (Rošić et al., 2018).

In early branching Chromadorea, DNMT1 was present while DNMT3 was absent in *Plectus sambesii*, *Sabatieria pulchra*, *Eleutherolaimus stenosoma*, *Plectus exinocaudatus*, *Anaplectus granulosus*, *Sabatieria pulchra*, *Sabatieria punctata*, *Axonolaimus paraspinosus*, *Chromadora nudicapitata*, *Dichromadora cephalata*, *Linhomoeus sp.*, *Paralinhomoeus sp.*, *Hypodontolaimus balticus*, *Terschellingia goubaultae*, *Terschellingia longicaudata* and *Leptolaimus papilliger*. (Figure 21). However, both DNMT1 and DNMT3 were absent in *Daptonema setosum* and

Theristus sp.. The presence of DNMT1 and DNMT2, with DNMT2 predominantly methylating tRNA3, has been documented in *Plectus sambesii* (Rošić et al., 2018; Engelhardt et al., 2022). This work expands the number of species with DNMT1 present (Figure 21) in early-branching Chromadorea.

tRNA-methyltransferase (DNMT2) was present in ancestral species from both Clade II and Clade I of Nematoda, as well as in sister species from Nematomorpha. A recent loss of a DNA methyltransferase-2 gene (*dnmt-2*) now a tRNA-methyltransferase, in the well-studied model organism *C. elegans* has been previously described (Gutierrez & Sommer, 2004). Notably, a *dnmt-2*-like gene has been identified in the satellite model organism *Pristionchus pacificus*, another free-living nematode that diverged from *C. elegans* approximately 200–300 million years ago. The findings in the current study corroborate this finding, suggesting that DNMT2 was likely part of an ancestral DNA methylation system in nematodes. However, through gene loss and sequence divergence, DNA methylation appears to have been abolished in *C. elegans* and related nematode species.

In Dorylaimia (clade I), DNMT3 methylation has been previously reported in *Trichinella spiralis* and *Romanomermis culicivorax* (Gao et al., 2012; Rošić et al., 2018). Both DNMT1 and DNMT3 were found in 7 other species including *Mononchus aquaticus*, *Prionchulus muscorum*, *Itochus arquatus*, *Myelonchulus hawaiiensis*, *Tylencholaimus nagauriensis*, *Mesodorylaimus* sp., *Discolaimus major* (Figure 21). The absence of DNMT1 in *Trichuris muris* and *Trichuris suis* and *Trichinella spiralis* has been previously reported (Rošić et al., 2018). This work validates the absence in these species and others including *Trichinella murrelli*, *Trichinella nelsoni*, *Trichinella pseudospiralis*, *Trichinella papuae* and *Soboliphyme baturini*. While DNMT1 has been previously reported in *Romanomermis culicivorax* (Rošić et al., 2018), all mermithids nematodes investigated in the current study did not possess DNMT1 methyltransferase (Figure 21). The absence of DNMT1 in the three mermithid nematodes (Figure 21) investigated in the current study could indicate intraspecific genetic variation within mermithid populations. Different strains or populations of mermithid nematodes may exhibit variability in the presence or absence of DNMT1. It's conceivable that over time, genetic changes and adaptations specific to each mermithid lineage may have led to the loss or modification of certain genes, including DNMT1. Additionally, differences in environmental factors, such as habitat or host preferences, could have exerted selective pressures contributing to the observed variation in gene presence or absence among these nematode species. Further comparative genomic analyses and functional studies may provide additional insights into the evolutionary processes underlying these differences.

I observed the presence of DNMT1, DNMT3, DNMT2, and UHRF1 in lim-

noterrestrial Gordiidae nematomorphs, while these proteins were absent in the marine nematomorph *Nectonema munidae*. However, it is important to note that the absence of these proteins in *Nectonema munidae* does not definitively confirm a loss, especially considering it is just one marine species analyzed. Therefore, while these findings suggest potential differences in the DNA methylation machinery among limnoterrestrial and marine nematomorph lineages, further investigation is needed to understand the evolutionary dynamics underlying these observations.

The presence of DNMT3 and DNMT1 signifies the likely importance of epigenetic mechanisms in the respective nematode species. This work presents a comprehensive overview of DNA methylation evolution in nematodes and nematomorphs, offering a solid foundation for future research avenues. One potential direction for further exploration could involve functional studies to elucidate the specific roles of identified DNMTs and UHRF1 domains in regulating gene expression, development, and response to environmental cues across different nematode species. Additionally, comparative genomic analyses with other animal phyla could provide insights into the evolutionary origins and diversification of DNA methylation machinery. Furthermore, investigating the interplay between DNA methylation and other epigenetic modifications, such as histone modifications and non-coding RNAs, could unveil intricate regulatory networks governing genome function and phenotype determination in nematodes.

4.4 Comparative genomics of mermithida and nematomorpha

Significant expansions and contractions in gene families across all analyzed species of the family Mermithidae (*Mermis nigrescens*, *Romanomermis culicivorax*, and *Romanomermis iyengari*) were identified. The analysis of nematomorph species revealed varying patterns of expansions and contractions in gene families in studied species (*Acutogordius australiensis*, *Gordionus montsenyensis*, *Gordius aquaticus*, and *Nectonema munidae*) but at a lesser scale than in mermithids. Both combined, however, provide interesting insight into their evolutionary genomic dynamics in general and point towards gene families that may be associated with their parasitoid lifestyle, most notably in controlling the behavior of their hosts, but also gene families possibly involved in defense mechanisms against the host's immune response.

To complete their life cycle, some parasites must influence the behavior of their hosts. This phenomenon is a notable illustration of the "extended phenotype", wherein the genes of one organism manifest phenotypic effects in an-

other organism. I found enriched GO terms associated with norepinephrine transport, neurotransmitter transport, and regulation of neurotransmitter activity in Mermithidae. Their genomes exhibit an expansion of gene families that play crucial roles in regulating neurotransmitter levels, motor activity, synaptic transmission, neurotransmitter:sodium symporter activity, GABA-A receptor activity, synaptic vesicle budding from the presynaptic membrane, and the regulation of synapse structural plasticity. Some of the expanded gene families in the common ancestor of the nematomorphs analyzed here were also associated with neurotransmitter transport. Previous studies identified dysregulated proteins linked to axon/dendrite and synapse modulation that was consistent across hosts of both Mermithidae and Nematomorpha, indicating that neuronal manipulation likely contributes to the induction of positive hydrotaxis (R. Herbison et al., 2019). Furthermore, neurotransmitter titers were shown to have changed in crickets infected with nematomorph parasites (Thomas et al., 2003). Proteomics analysis of the biochemical changes occurring in the cricket *Nemobius sylvestris* and a grasshopper *Meconema thalassinum* when manipulated by the hairworms *Paragordius tricuspidatus* and *Spiniochordodes tellinii* respectively identified parasitoid-produced molecules from the Wnt family, which likely affect the host's central nervous system development. Differential protein expression in the manipulated insects suggested involvement in neurogenesis, circadian rhythm, and neurotransmitter activities (D. Biron et al., 2006, 2005). All of this confirms a genomic capacity of mermithids and nematomorphs to manipulate neuronal connections of their host, possibly either by altering neurotransmitter release in hosts or by producing neuromodulators/neurotransmitters themselves, in a similar way as described for parasitoid wasps (Gal, Rosenberg, & Libersat, 2005; Libersat, Delago, & Gal, 2009) and tapeworms (Øverli, Pall, Borg, Jobling, & Winberg, 2001).

A less direct way to control the behavior of its host may be used by the parasitic mermithid *Thaumamermis zealandica*, either instead of or in addition to the neural pathways discussed above. The nematode induces an increase in haemolymph osmolality in parasitized *Bellorchestia quoyana* (formerly *Talorchestia*) comparing to unparasitized amphipods (Williams et al., 2004). This alteration may trigger a water-seeking behaviour in the host, providing a potential explanation for their observed tendencies to dig deeper into water saturated sand that the nematode requires to emerge from the host. This analysis suggests that the last common ancestor of mermithid genomes had expanded gene families associated with potassium ion transmembrane transport and iron ion transport. However, further studies are required to validate the connection between observed patterns of mermithid genomes and the manipulation of haemolymph parameters in their hosts. Most descriptions of parasite-induced changes in host behavior

rely on observational reports, with limited experimentally confirmed examples of parasite genes directly causing these changes (van Houte, Ros, & van Oers, 2013; D. G. Biron & Loxdale, 2013).

The expansion of gene families associated with various functions such as response to toxic substances, norepinephrine transport, DNA integration, proteolysis, DNA recombination, amino-acid betaine transport, transposition, RNA-mediated mechanisms, and neurotransmitter:sodium symporter activity was observed in both Mermithidae and Nematomorpha, indicating their vital roles in the biology or ecology of both groups of worm parasitoids. Furthermore, this suggests parallel evolution, driven by similar ecological pressures or selective forces, leading to convergent adaptations in phylogenetically distant lineages. These gene families likely contribute to key aspects of the parasitoid lifestyle, such as host immune evasion and manipulation of host physiology (Adamo & Webster, 2013; Quicke, Ghafouri Moghaddam, & Butcher, 2023; Del Giudice, 2019). Parasitoids frequently encounter various host defenses, including the production of toxic substances as part of host immune responses, thus gene families involved in detoxification and responses to such toxins are crucial for parasitic worms to withstand and overcome these challenges. For instance, studies have highlighted the significance of detoxification enzymes like glutathione S-transferases in facilitating the metabolic adjustments required for the endoparasitoid wasp *Pteromalus puparum* to adapt to papilionid butterfly pupae (G. Xu et al., 2020).

There were some differences observed in genomes between Mermithidae and Nematomorpha. For example, both mermithids and nematomorphs accumulate resources during the infective juvenile stage and cease feeding as adults, indicating a shared life history strategy. However, while nematomorphs exhibit enrichment of gene families in response to starvation, indicating potential adaptations to periods of nutrient scarcity, such enrichment is not observed in mermithids. This difference might suggest divergent metabolic strategies between these two groups, warranting further investigation to elucidate the underlying mechanisms and ecological implications. The sensorium in Nematomorpha is practically unknown and the question of how hairworms find mates for internal insemination has never been satisfactorily answered (Sokolova et al., 2022). The discovery of expanded gene families enriched in GO terms associated with sensory perception of sound (vibrations) offers an interesting avenue to explore. Mermithids, on the other hand, are fully equipped with a variety of sensory structures, including photoreceptors, although there is limited understanding regarding the expression of phototaxis in different mermithid species. I found an expanded gene family related to visual perception in mermithids which might mirror the host-seeking behavior and presence of photosensitive structures in *Mermis nigrescens* (Mohamed, Burr, &

Burr, 2007), while information regarding *Romanomermis* remains elusive.

TE content displays a wide variation among the studied species, and the underlying reasons for this diversity remain unclear. Particularly, *Nasonia* parasitoid wasp exhibits notably high TE content, yet the factors driving this phenomenon remain elusive (Wurm & Keller, 2010). The variability of TE landscapes within and across insect orders suggests a complex interplay influenced by factors beyond phylogenetic relatedness, possibly including horizontal transfer. While the impact of insect traits on TE landscapes remains uncertain, recent discoveries suggest TEs play crucial roles in insect adaptations, aging, and antiviral immunity. Thus, TEs are increasingly recognized as essential symbiotic elements in insects, with potentially both harmful and beneficial effects depending on the context (Gilbert, Peccoud, & Cordaux, 2021). Therefore, this analysis focuses on determining TE content in the species that are available, aiming to establish foundational data for further investigations.

There are differences in the copy number of TE families between different studied genomes as shown in the (Figure 25). The mermithid genomes have varying completeness with *R. culicivora*x and *M. nigrescens* possessing much better contiguity than the genome of *R. iyengari*. Detection and annotation of TEs are sensitive to the quality and contiguity of the assembly. The draft assembly of *Nectonema munidae* has a relatively high contiguity 213 Mbp genome, and an N50 of 716.6 Kbp and that of *Romanomermis iyengari* shows a 336 Mbp genome and an N50 of 152 Kbp, still needing improvement in terms of contiguity and could be the source of inflated TE amounts. Nevertheless, the detection of the TE superfamily composition is not impeded.

Understanding the conserved TE superfamilies within a lineage sheds light on potentially active TEs within the respective group. From this analysis it can be derived that the superfamilies of *hAT*, *CACTA*, *Mutator*, *Transib*, *Mariner*, *Helitron* and *Gypsy* persisted successfully among all groups while *Onchocerca volvulus* and *Brugia malayi* succeeded in eliminating whole superfamilies like *PiggyBac*, *ERV*, *Jockey* and *SINES* and downregulating the TE activity overall. Apart from *Nectonema munidae*, it seems that the amounts and activities of TEs are decreasing proportionately in the Nematomorphs. On the contrary, they increase in Mermithids, and while generally suppressed in activity and low in amount in the comparative in-group of *Brugia malayi*, *Onchocerca volvulus* and *Plectus sambesii*, TEs also exhibit a recent burst in activity, most strikingly in *Plectus sambesii*. *Mermis nigrescens* is the only species where *DIRS* elements were detected.

5 CONCLUSIONS AND FUTURE PERSPECTIVES

This thesis introduces high-quality genomes of *Tobrilus gracilis*, *Mononchus aquaticus*, *Romanomermis culicivorax*, *Romanomermis iyengari* and *Tripylella intermedia*, thereby enriching the genomic resources available for underrepresented nematode lineages. Moving forward, further refinement of these genomes except *Romanomermis culicivorax*, will be necessary. This includes integrating PacBio HiFi and Nanopore long reads, as well as employing long-range sequencing techniques like chromosome conformation capture, to enhance their structural resolution to the chromosome level. It is crucial to obtain extensive high quality data from numerous nematode species to study how genes, gene families, and genome organization evolve throughout the Nematoda. Having high-quality genomes from various nematodes helps us understand how genetic changes affect major evolutionary shifts in the phylum and also shows the extent of genomic differences within the group. Nonetheless, the genomes produced as part of this thesis provide sufficient information for in-depth analysis of several fundamental biological questions.

Hox gene evolution and Hox cluster dynamics in nematodes highlight the remarkable evolutionary flexibility of these essential bilaterian regulators, within the nematode body plan. The loss and rearrangement of these genes, particularly observed in the lineage leading to the Ur-Nematode, may have played significant roles in shaping the roundworm body plan. The findings from here, including the identification of intact and uninterrupted Hox clusters in certain nematode species, underscore the need for further research to elucidate the implications of these rearrangements, including their effects on the collinearity of Hox gene expression. With the advent of advanced molecular techniques, such as *in situ* hybridization, single-cell RNA-Seq studies, and CRISPR-guided gene knockouts, future investigations have the potential to uncover novel insights into the regulation of Hox clusters across a broad spectrum of the Nematoda. As high-quality genomes continue to expand, the anticipation grows for the emergence of new lineages harboring unified Hox gene clusters, promising novel directions in nematode evolutionary developmental biology. The presence of "undispersed" Hox clusters in nematodes challenges prevailing notions of Hox cluster evolution, hinting at the preservation of ancestral traits long considered lost. Exploring Hox cluster evolution in nematodes holds the potential to uncover distinctive regulatory mechanisms, shedding light on developmental gene regulation and the evolutionary trajectory of the conserved nematode body plan. This thesis contributes to

our understanding of the diversity and complexity of genomic organization within the phylum Nematoda. The Hox cluster genomic pattern observed in *C. elegans* is not representative of the entire phylum. I highlight the wide variability in orthologous gene sets and their genomic arrangements across different nematode species. Specifically, the investigation underscores the importance of gene losses in shaping the Hox gene complements of nematodes and highlights the influence of inversions and translocations on the arrangement of nematode Hox clusters. These findings shed light on the dynamic nature of nematode genomes and provide valuable insights into the evolutionary processes driving genomic diversity in this phylum.

This thesis has made substantial strides in unraveling the evolutionary landscape of methyltransferase enzymes across a diverse spectrum of nematode species, including those belonging to the early branching Chromadorea and Enoplea, which have only recently had their genomes sequenced. A key revelation from this investigation is the absence of DNA methylation in the Clade II Enopleans studied, replaced instead by the presence of tRNA methyltransferase. These findings carry significant implications for understanding the regulatory mechanisms orchestrated by tRNA methylases in non-model nematodes. Furthermore, the absence of DNMT1 and DNMT3 observed in Clades II to V, contrasted with their presence in Clade I and the early-branching Chromadorea, adds significant depth to our understanding of the intricate variability of DNA methylation within the diverse nematode phylum. Finally, this study offers insight into which nematodes possess the capacity for DNA and tRNA methylation, providing a basis for further investigation into their epigenetic regulatory mechanisms.

Whereas comparative analysis of parasitoid genome sequences alone cannot fully explain parasitoid biology and interactions with hosts, our findings offer foundational insights into the genomic factors driving parasitoid worm adaptation, laying the groundwork and allowing future research to focus on particular gene families. Further, the identification and characterization of transposable element superfamilies within nematomorph and mermithid genomes provide a valuable starting point for a more in-depth exploration of their genomic impact. The significance of studies of gene family evolution and identification of particular gene families associated with specific functions extends beyond mere theoretical insights. They may hold practical value for addressing real-world challenges, particularly in the realm of vector control strategies. Both mermithids and nematomorphs have strong negative effects on the population density of their hosts, in addition to the manipulation of their host's dispersal behavior discussed above. Not covered in this particular study, but no less important is the ability of mermithids to affect host reproduction success by sterilization, inducing changes in secondary

sex characters (intersexes) and mating behavior (Vance, 1996; Ya'cob et al., 2021; Muñoz-Muñoz, Ramoneda, Pagès, Pujol, & Talavera, 2016; Rowell, 2000). Specifically, uncovering the mechanisms of parasitoid-induced changes in dispersal behavior and reproductive development has a promising potential for applying the knowledge obtained in developing targeted and environmentally friendly methods for controlling insect pests. By leveraging the understanding of molecular factors involved in parasitoid-host interactions, we may uncover innovative approaches to, for example, curbing the transmission of malaria, offering a sustainable and effective means of combating a global health threat. The identified molecular signatures among genomes of parasitoid taxa provide valuable insights into the genomic dynamics shaping their biology. Overall, this work contributes significantly to our understanding of parasitoidism in nematodes and nematomorphs, offering a valuable resource for further research in this field.

The evolutionary investigations conducted in this thesis, delving into the evolution of Hox gene complements, the dynamics of nematode Hox clusters, and the evolution of key enzymes for the DNA methylation machinery across nematodes, have been made possible by the availability of high-quality genomes. These genomes, encompassing both those generated within this study and those sourced from the Darwin Tree of Life (DTOL) project at the Sanger Institute, serve as the cornerstone for these critical inquiries.

6 ACKNOWLEDGEMENTS

There are many people who have helped me on my way (and sometimes helped me find what way that was in the first place). **Philipp Schiffer**, supervisor and nerdy friend. Thank you so much for always being there and having the almost unreal quality of always making me see things clearly and making me feel better in any situation. Without you, I would not have gone the Nematode genomics and evolutionary development route to begin with. Maybe you are the reason I thought a science career could be fun and have formed a foundation of what I think a good research team should be like. I can not imagine a better supervisor. Thank you to my mentors, **Oleksandr Holovachov** and **Siegfried Roth** for your excellent support and continuous helpfulness. I have learned a lot from you. Thanks for all the discussions on both personal and professional development. I am sometimes awestruck by your ability to always provide great and on-point feedback.

I have had the absolute pleasure of getting to know a whole lot of amazing people over the past few years. **Tarja Hoffmeyer**, I will not forget the times we spent together setting up the laboratory in the middle of the pandemic. I have learned a lot from you.

Nadège Guiglielmoni, you have been there all along, supporting and helping me in both experimental and bioinformatics analysis of nematode genomes whenever needed.

Laura Villegas, your happy outlook on life and aspirations for mastering everything you come across are inspiring and I am overjoyed to have met you. It was always fun working in the laboratory and attending conferences with you.

I extend my heartfelt gratitude to **Viktoria Hellekes**, whose unwavering support and empathetic presence have been a source of immense inspiration throughout my doctoral journey. You provided me with a constant reservoir of encouragement and understanding. The open dialogue we have maintained, touching upon both academic and daily life stresses, has created a space where I feel truly inspired.

All the people connected to **Ann-Marie Waldvogel's group** and **Jens Bast's group** at the University of Cologne, you contribute to why it is so much fun to go to work every day. You all made it so much fun whenever we had joint meetings

discussing both serious and geeky stuff. The time with you guys taught me about research life in general and your special humor just inspired me to never accept not having fun in the workplace.

Oliver Karst, I am grateful for your assistance with the technical aspects of the laboratory, particularly during the challenging times of the COVID-19 pandemic. Your consistent helpfulness and friendliness, even when I posed repetitive questions, have been invaluable. Thank you for your unwavering support.

To my collaborators, I am grateful to have met and worked with you over the years. It is my firm belief that collaboration creates something greater than the sum of its parts, and working with you has reinforced that belief. Special mentions to **Sanger Institute, Hinxton Cambridge**, Mark Blaxter, Erna King and Lewis Stevens. **Max Plank institute for plant breeding** Bruno H, **Genomics and Transcriptomics Laboratory (GTL)** Kerstin Becker, **Cologne Center for Genomics (CCG)** Christian Becker.

I owe a great deal of my sanity to my best friends **Elly Isabirye, Bob Lutaya, Kazibwe Joseph** and **Lutimba Stuart**, with whom we received the Swedish Institute Scholarship at the time when starting as masters students fresh from Uganda to Sweden. We became a family and ever since then we have shared happy memories. I can not imagine what the past four years would have been without you. Spending time with you was the best creative outlet I could have had when I needed to focus on something else. Thank you for being part of my extended family.

I am very grateful for my family and your ability to always listen to me even if the only clear thing is that I am not making much sense. Thank you **Rev. Fr. Eugene Kalyango**, thank you for teaching me the Godly values of love, gratefulness, generosity, forgiveness, courageousness and compassion above everything.

My lovely wife **Faith Githiri**, thank you for being you and for your unwavering support. Sharing my life with you is fantastic and I would marry you again in a heartbeat.

To my siblings **Adriane Ddamulira, Prossy Nanyange** and **JaneFrancis Nakuya** you are always there no matter what and thank you for standing with me in all situations.

7 DEDICATION

In profound reverence and everlasting gratitude, this dissertation is dedicated to my late beloved parents, **Edward Mayanja** and **JaneFrancis Nabulya**, whose unwavering love, sacrifices, and indomitable spirits have been my guiding lights throughout this academic journey.

To my father, whose wisdom and unyielding support have been the cornerstone of my intellectual pursuits. You were my first teacher, imparting not just knowledge but instilling in me a passion for learning that has fueled my academic aspirations. Your relentless work ethic, moral fortitude, and unspoken encouragement have shaped my character and propelled me forward, even in the face of adversity.

To my mother, in the poignant final moments before she departed from this world, she uttered heartfelt cries, a tearful refrain that echoed with profound emotion. "My children, my children". These words, saturated with love and concern, linger in the air as an enduring testament to the depth of her maternal devotion. Amid her transition, she bestowed upon us a poignant reminder of the cherished bond that will forever connect us, transcending the boundaries between life and eternity. Her love and nurturing spirit provided the fertile ground for my dreams to take root and flourish. Her sacrifices were the silent force that allowed me to pursue higher education with unwavering determination. Her resilience in the face of challenges has been a source of inspiration, demonstrating the power of grace and tenacity.

8 DECLARATION

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten – noch nicht veröffentlicht worden ist sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Ordnung zur Sicherung guter wissenschaftlicher Praxis und zum Umgang mit wissenschaftlichem Fehlverhalten der Universität zu Köln gelesen und sie bei der Durchführung der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht.

Teilpublikationen:

03.11.2024



.....

Date / Signature

References

- Abad, P., Gouzy, J., Aury, J.-M., Castagnone-Sereno, P., Danchin, E. G., Deleury, E., . . . others (2008). Genome sequence of the metazoan plant-parasitic nematode *meloidogyne incognita*. *Nature biotechnology*, 26(8), 909–915.
- Abebe, E., Decraemer, W., & De Ley, P. (2008). Global diversity of nematodes (nematoda) in freshwater. *Freshwater Animal Diversity Assessment*, 67–78.
- Aboobaker, A., & Blaxter, M. (2003). Hox gene evolution in nematodes: novelty conserved. *Current opinion in genetics & development*, 13(6), 593–598.
- Aboobaker, A., & Blaxter, M. (2010). The nematode story: Hox gene loss and rapid evolution. In *Hox genes: Studies from the 20th to the 21st century* (pp. 101–110). Springer.
- Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., & Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in darwin’s finches. *Science*, 305(5689), 1462–1465.
- Adamo, S. A., & Webster, J. P. (2013). *Neural parasitology: how parasites manipulate host behaviour* (Vol. 216) (No. 1). Company of Biologists.
- Adams, M., McBroome, J., Maurer, N., Pepper-Tunick, E., Saremi, N. F., Green, R. E., . . . Corbett-Detig, R. B. (2020). One fly—one genome: chromosome-scale genome assembly of a single outbred *drosophila melanogaster*. *Nucleic acids research*, 48(13), e75–e75.
- Ahmed, M., & Holovachov, O. (2021). Twenty years after de ley and blaxter—how far did we progress in understanding the phylogeny of the phylum nematoda? *Animals*, 11(12), 3479.
- Ahmed, M., Roberts, N. G., Adediran, F., Smythe, A. B., Kocot, K. M., & Holovachov, O. (2022). Phylogenomic analysis of the phylum nematoda: conflicts and congruences with morphology, 18s rna, and mitogenomes. *Frontiers in Ecology and Evolution*, 9, 769565.
- Akam, M. (1998). Hox genes, homeosis and the evolution of segment identity: no need for hopeless monsters. *International Journal of Developmental Biology*, 42, 445–451.

-
- Akam, M. (2000). Arthropods: developmental diversity within a (super) phylum. *Proceedings of the National Academy of Sciences*, 97(9), 4438–4441.
- Akam, M., Averof, M., Castelli-Gair, J., Dawes, R., Falciani, F., & Ferrier, D. (1994). The evolving role of hox genes in arthropods. *Development*, 1994(Supplement), 209–215.
- Aleshin, V. V., Kedrova, O. S., Milyutina, I. A., Vlad'yehenskaya, N., & Petrov, N. B. (1998). Relationships among nematodes based on the analysis of 18s rRNA gene sequences: molecular evidence for monophyly of chromadorian and secernentian nematodes. *Russian Journal of Nematology*, 6, 175–184.
- Allahverdi-pour, H. H., Talaei-Hassanloui, R., Karimi, J., Wang, Y., Rochlin, I., & Gaugler, R. (2019). Behavior manipulation of mosquitoes by a mermithid nematode. *Journal of invertebrate pathology*, 168, 107273.
- Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., ... Soyk, S. (2022). Automated assembly scaffolding using ragtag elevates a new tomato system for high-throughput genome editing. *Genome biology*, 23(1), 258.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, 215(3), 403–410.
- Ambros, V., Lee, R. C., Lavanway, A., Williams, P. T., & Jewell, D. (2003). MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Current Biology*, 13(10), 807–818.
- Anderson, R. C. (2000). *Nematode parasites of vertebrates: their development and transmission*. Cabi.
- Andrew Rambaut. (2006). *FigTree*. Retrieved from <http://tree.bio.ed.ac.uk/software/figtree/>
- Appeltans, W., Ah Yong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., ... others (2012). The magnitude of global marine species diversity. *Current biology*, 22(23), 2189–2202.
- Aury, J.-M., & Istace, B. (2021). Hapo-g, haplotype-aware polishing of genome assemblies with accurate reads. *NAR genomics and bioinformatics*, 3(2), lqab034.
- Baker, E. A., & Woollard, A. (2019). How weird is the worm? evolution of the developmental gene toolkit in *Caenorhabditis elegans*. *Journal of developmental biology*, 7(4), 19.

-
- Barnes, T., Kohara, Y., Coulson, A., & Hekimi, S. (1995). Meiotic recombination, noncoding dna and genomic organization in *caenorhabditis elegans*. *Genetics*, *141*(1), 159–179.
- Bartel, D. P. (2004). Micrnas: genomics, biogenesis, mechanism, and function. *cell*, *116*(2), 281–297.
- Baudry, L., Guiglielmoni, N., Marie-Nelly, H., Cormier, A., Marbouty, M., Avia, K., ... others (2020). instagraal: chromosome-level quality scaffolding of genomes using a proximity ligation-based scaffold. *Genome Biology*, *21*, 1–22.
- Beeman, R. W. (1987). A homoeotic gene cluster in the red flour beetle. *Nature*, *327*(6119), 247–249.
- Belton, J.-M., McCord, R. P., Gibcus, J. H., Naumova, N., Zhan, Y., & Dekker, J. (2012). Hi-c: a comprehensive technique to capture the conformation of genomes. *Methods*, *58*(3), 268–276.
- Bento, G., Ogawa, A., & Sommer, R. J. (2010). Co-option of the hormone-signalling module dafachronic acid–daf-12 in nematode evolution. *Nature*, *466*(7305), 494–497.
- Berger, S. L., Kouzarides, T., Shiekhatar, R., & Shilatifard, A. (2009). An operational definition of epigenetics. *Genes & development*, *23*(7), 781–783.
- Betrán, E., & Long, M. (2002). Expansion of genome coding regions by acquisition of new genes. *Genetica*, *115*, 65–80.
- Bewick, A. J., Vogel, K. J., Moore, A. J., & Schmitz, R. J. (2017). Evolution of dna methylation across insects. *Molecular biology and evolution*, *34*(3), 654–665.
- Bhattarai, U. R., Poulin, R., Gemmell, N. J., & Dowle, E. (2022). Genome assembly and annotation of the mermithid nematode *mermis nigrescens*. *bioRxiv*, 2022–11.
- Bird, D. M., Jones, J. T., Opperman, C. H., Kikuchi, T., & Danchin, E. G. (2015). Signatures of adaptation to plant parasitism in nematode genomes. *Parasitology*, *142*(S1), S71–S84.
- Bird, D. M., Opperman, C. H., & Davies, K. G. (2003). Interactions between bacteria and plant-parasitic nematodes: now and then. *International journal for parasitology*, *33*(11), 1269–1276.

-
- Biron, D., Marché, L., Ponton, F., Loxdale, H., Galéotti, N., Renault, L., ... Thomas, F. (2005). Behavioural manipulation in a grasshopper harbouring hairworm: a proteomics approach. *Proceedings of the Royal Society B: Biological Sciences*, 272(1577), 2117–2126.
- Biron, D., Ponton, F., Marché, L., Galeotti, N., Renault, L., Demey-Thomas, E., ... Thomas, F. (2006). ‘suicide’ of crickets harbouring hairworms: a proteomics investigation. *Insect Molecular biology*, 15(6), 731–742.
- Biron, D. G., & Loxdale, H. D. (2013). Host–parasite molecular cross-talk during the manipulative process of a host by its parasite. *Journal of Experimental Biology*, 216(1), 148–160.
- Blaxter, M. (2000). Genes and genomes of *necator americanus* and related hookworms. *International journal for parasitology*, 30(4), 347–355.
- Blaxter, M. (2011). Nematodes: the worm and its relatives. *PLoS biology*, 9(4), e1001050.
- Blaxter, M. L. (2003). Nematoda: genes, genomes and the evolution of parasitism. *Adv Parasitol*, 54, 101–195.
- Blaxter, M. L., De Ley, P., Garey, J. R., Liu, L. X., Scheldeman, P., Vierstraete, A., ... others (1998). A molecular evolutionary framework for the phylum nematoda. *Nature*, 392(6671), 71–75.
- Bleidorn, C., Schmidt-Rhaesa, A., & Garey, J. R. (2002). Systematic relationships of nematomorpha based on molecular and morphological data. *Invertebrate Biology*, 121(4), 357–364.
- Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., ... others (2002). A global analysis of *caenorhabditis elegans* operons. *Nature*, 417(6891), 851–854.
- Bogale, M., Baniya, A., & DiGennaro, P. (2020). Nematode identification techniques and recent advances. *Plants*, 9(10), 1260.
- Bostick, M., Kim, J. K., Estève, P.-O., Clark, A., Pradhan, S., & Jacobsen, S. E. (2007). Uhrf1 plays a role in maintaining dna methylation in mammalian cells. *Science*, 317(5845), 1760–1764.
- Boveri, T. (1899). Die entwicklung von *ascaris megalocephala* mit besonderer rücksicht auf die kernverhältnisse. (*No Title*).

-
- Brenner, S. (1973). The genetics of behaviour. *British medical bulletin*, 29(3), 269–271.
- Brenner, S. (1974). The genetics of caenorhabditis elegans. *Genetics*, 77(1), 71–94.
- Brooker, S. (2010). Estimating the global distribution and disease burden of intestinal nematode infections: adding up the numbers—a review. *International journal for parasitology*, 40(10), 1137–1144.
- Bürglin, T. R., Lobos, E., & Blaxter, M. L. (1998). Caenorhabditis elegans as a model for parasitic nematodes. *International journal for parasitology*, 28(3), 395–411.
- Bürglin, T. R., & Ruvkun, G. (1993). The caenorhabditis elegans homeobox gene cluster. *Current Opinion in Genetics & Development*, 3(4), 615–620.
- Burr, A., Wagar, D., & Sidhu, P. (2000). Ocellar pigmentation and phototaxis in the nematode mermis nigrescens: changes during development. *Journal of Experimental Biology*, 203(8), 1341–1350.
- Bush, J. O., Maltby, K. M., Cho, E.-S., & Jiang, R. (2003). The t-box gene tbx10 exhibits a uniquely restricted expression pattern during mouse embryogenesis. *Gene expression patterns*, 3(4), 533–538.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). Blast+: architecture and applications. *BMC bioinformatics*, 10, 1–9.
- Cameron, R. A., Rowen, L., Nesbitt, R., Bloom, S., Rast, J. P., Berney, K., ... others (2006). Unusual gene order and organization of the sea urchin hox cluster. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 306(1), 45–58.
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). egg-nog-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular biology and evolution*, 38(12), 5825–5829.
- Carroll, S. B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. *Cell*, 101(6), 577–580.
- Carroll, S. B. (2005). Evolution at two levels: on genes and form. *PLoS biology*, 3(7), e245.

-
- Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, 134(1), 25–36.
- Chen, M.-Y., Teng, W.-K., Zhao, L., Hu, C.-X., Zhou, Y.-K., Han, B.-P., . . . Shu, W.-S. (2021). Comparative genomics reveals insights into cyanobacterial evolution and habitat adaptation. *The ISME Journal*, 15(1), 211–227.
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics*, 34(17), i884–i890.
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods*, 18(2), 170–175.
- Cheng, X., & Blumenthal, R. M. (2008). Mammalian dna methyltransferases: a structural perspective. *Structure*, 16(3), 341–350.
- Cobb, N. A. (1914). *Nematodes and their relationships*. US Government Printing Office.
- Coghlan, A. (2005). Nematode genome evolution. *WormBook*, 1.
- Comparative genomics of the major parasitic worms. (2019). *Nature genetics*, 51(1), 163–174.
- Consortium, H. G. S., et al. (2006). Insights into social insects from the genome of the honeybee apis mellifera. *Nature*, 443(7114), 931.
- Crittenden, S. L., Eckmann, C. R., Wang, L., Bernstein, D. S., Wickens, M., & Kimble, J. (2003). Regulation of the mitosis/meiosis decision in the caenorhabditis elegans germline. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 358(1436), 1359–1362.
- Cunha, T. J., de Medeiros, B. A., Lord, A., Sørensen, M. V., & Giribet, G. (2023). Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic nematomorpha. *Current Biology*, 33(16), 3514–3521.
- Cutter, A. D., Dey, A., & Murray, R. L. (2009). Evolution of the caenorhabditis elegans genome. *Molecular biology and evolution*, 26(6), 1199–1234.
- Darbellay, F., Bochaton, C., Lopez-Delisle, L., Mascrez, B., Tschopp, P., Delpretti, S., . . . Duboule, D. (2019). The constrained architecture of mammalian hox gene clusters. *Proceedings of the National Academy of Sciences*, 116(27), 13424–13433.

-
- Davis, E. L., Hussey, R. S., & Baum, T. J. (2004). Getting to the roots of parasitism by nematodes. *Trends in parasitology*, 20(3), 134–141.
- Dearden, P. K., Wilson, M. J., Sablan, L., Osborne, P. W., Havler, M., McNaughton, E., ... others (2006). Patterns of conservation and change in honey bee developmental genes. *Genome Research*, 16(11), 1376–1384.
- De Ley, P. (2002). The biology of nematodes. (*No Title*), 1.
- De Ley, P. (2006). A quick tour of nematode diversity and the backbone of nematode phylogeny. *WormBook: The Online Review of C. elegans Biology [Internet]*.
- De Ley, P., & Blaxter, M. (2002). Systematic position and phylogeny. In *The biology of nematodes* (pp. 1–30). CRC Press.
- De Ley, P., & Blaxter, M. L. (2004). A new system for nematoda: combining morphological characters with molecular trees, and translating clades into ranks and taxa. In *Proceedings of the fourth international congress of nematology, 8-13 june 2002, tenerife, spain* (pp. 633–653).
- Del Giudice, M. (2019). Invisible designers: brain evolution through the lens of parasite manipulation. *The Quarterly Review of Biology*, 94(3), 249–282.
- Del Valle, A., Jones, B. F., Harrison, L. M., Chadderdon, R. C., & Cappello, M. (2003). Isolation and molecular cloning of a secreted hookworm platelet inhibitor from adult *ancylostoma caninum*. *Molecular and biochemical parasitology*, 129(2), 167–177.
- de Mendoza, A., Hatleberg, W. L., Pang, K., Leininger, S., Bogdanovic, O., Pflueger, J., ... others (2019). Convergent evolution of a vertebrate-like methylome in a marine sponge. *Nature ecology & evolution*, 3(10), 1464–1473.
- de Mendoza, A., Pflueger, J., & Lister, R. (2019). Capture of a functionally active methyl-cpg binding domain by an arthropod retrotransposon family. *Genome research*, 29(8), 1277–1286.
- Denver, D. R., Morris, K., Lynch, M., & Thomas, W. K. (2004). High mutation rate and predominance of insertions in the *caenorhabditis elegans* nuclear genome. *Nature*, 430(7000), 679–682.
- De Rosa, R., Grenier, J. K., Andreeva, T., Cook, C. E., Adoutte, A., Akam, M., ... Balavoine, G. (1999). Hox genes in brachiopods and priapulids and protostome evolution. *Nature*, 399(6738), 772–776.

-
- de Valdez, M. R. W. (2005). *Parasitoid-induced behavioral alterations of aedes aegypti mosquito larvae infected with mermithid nematodes (nematoda: Mermithidae)*. Colorado State University.
- Devi, G. (2020). Nematodes as environmental indicator. *International Journal of Current Research*, 12, 11699–11705.
- Dillman, A. R., Macchietto, M., Porter, C. F., Rogers, A., Williams, B., Antoshechkin, I., ... others (2015). Comparative genomics of steinernema reveals deeply conserved gene regulatory networks. *Genome biology*, 16, 1–21.
- Dolinski, C., Baldwin, J., & Thomas, W. (2001). Comparative survey of early embryogenesis of secernentea (nematoda), with phylogenetic implications. *Canadian Journal of Zoology*, 79(1), 82–94.
- Duboule, D. (2007). The rise and fall of hox gene clusters.
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., ... others (2017). De novo assembly of the aedes aegypti genome using hi-c yields chromosome-length scaffolds. *Science*, 356(6333), 92–95.
- Eckert, J., Friedhoff, K. T., Zahner, H., & Deplazes, P. (2005). Lehrbuch der parasitologie für die tiermedizin. *Schweizer Archiv für Tierheilkunde*, 147(8), 358–358.
- Eggleton, P., & Belshaw, R. (1992). Insect parasitoids: an evolutionary overview. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 337(1279), 1–20.
- Eggleton, P., & Gaston, K. J. (1990). " parasitoid" species and assemblages: Convenient definitions or misleading compromises? *Oikos*, 417–421.
- Eleftheriadi, K., Guiglielmoni, N., Salces-Ortiz, J., Vargas-Chavez, C., Martínez-Redondo, G. I., Gut, M., ... Fernández, R. (2023). The genome sequence of the montseny horsehair worm, gordionus montsenyensis sp. nov., a key resource to investigate ecdysozoa evolution. *bioRxiv*, 2023–06.
- elegans Sequencing Consortium*, C. (1998). Genome sequence of the nematode c. elegans: a platform for investigating biology. *Science*, 282(5396), 2012–2018.
- Ellen A. Bell, C. O. S. M. L. Y. M. I. T., Christopher L. Butler. (2022). Transposable element annotation in non-model species: The benefits of species-specific repeat libraries using semi-automated edta and deepte de novo pipelines. *Molecular Ecology Resources*, 22, 823–833.

-
- Elling, A. A. (2013). Major emerging problems with minor meloidogyne species. *Phytopathology*, *103*(11), 1092–1102.
- Ellinghaus, D., Kurtz, S., & Willhoeft, U. (2008). Ltrharvest, an efficient and flexible software for de novo detection of ltr retrotransposons. *BMC bioinformatics*, *9*, 1–14.
- Emms, D. M., & Kelly, S. (2019). Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, *20*, 1–14.
- Engelhardt, J., Scheer, O., Stadler, P. F., & Prohaska, S. J. (2022). Evolution of dna methylation across ecdysozoa. *Journal of molecular evolution*, *90*(1), 56–72.
- Erwin, D. H. (2009). Early origin of the bilaterian developmental toolkit. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1527), 2253–2261.
- Fenn, K., Conlon, C., Jones, M., Quail, M. A., Holroyd, N. E., Parkhill, J., & Blaxter, M. (2006). Phylogenetic relationships of the wolbachia of nematodes and arthropods. *PLoS pathogens*, *2*(10), e94.
- Fernández, R., & Gabaldón, T. (2020). Gene gain and loss across the metazoan tree of life. *Nature ecology & evolution*, *4*(4), 524–533.
- Ferrier, D. E., & Minguillon, C. (2003). Evolution of the hox/parahox gene clusters. *International Journal of Developmental Biology*, *47*(7-8), 605–611.
- Field, L., Lyko, F., Mandrioli, M., & Prantera, G. (2004). Dna methylation in insects. *Insect molecular biology*, *13*(2), 109–115.
- Finn, R. D., Clements, J., & Eddy, S. R. (2011). Hmmer web server: interactive sequence similarity searching. *Nucleic acids research*, *39*(suppl_2), W29–W37.
- Finnegan, E. J., Peacock, W. J., & Dennis, E. S. (2000). Dna methylation, a key regulator of plant development and other processes. *Current opinion in genetics & development*, *10*(2), 217–223.
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). Repeatmodeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457.

-
- Fradin, H., Kiontke, K., Zegar, C., Gutwein, M., Lucas, J., Kovtun, M., ... others (2017). Genome architecture and evolution of a unichromosomal asexual nematode. *Current Biology*, 27(19), 2928–2939.
- Gabriel, L., Brna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., & Stanke, M. (2023). Braker3: Fully automated genome annotation using rna-seq and protein evidence with genemark-etp, augustus and tsebra. *bioRxiv*.
- Gal, R., Rosenberg, L. A., & Libersat, F. (2005). Parasitoid wasp uses a venom cocktail injected into the brain to manipulate the behavior and metabolism of its cockroach prey. *Archives of Insect Biochemistry and Physiology: Published in Collaboration with the Entomological Society of America*, 60(4), 198–208.
- Ganko, E. W., Bhattacharjee, V., Schliekelman, P., & McDonald, J. F. (2003). Evidence for the contribution of ltr retrotransposons to *c. elegans* gene evolution. *Molecular biology and evolution*, 20(11), 1925–1931.
- Gao, F., Liu, X., Wu, X.-P., Wang, X.-L., Gong, D., Lu, H., ... others (2012). Differential dna methylation in discrete developmental stages of the parasitic nematode trichinella spiralis. *Genome biology*, 13, 1–13.
- Garcia-Bustos, J. F., Sleebs, B. E., & Gasser, R. B. (2019). An appraisal of natural products active against parasitic nematodes of animals. *Parasites & Vectors*, 12(1), 1–22.
- Garcia-Fernandez, J. (2005). Hox, parahox, protohox: facts and guesses. *Heredity*, 94(2), 145–152.
- Gattoni, K., Gendron, E. M., Sandoval-Ruiz, R., Borgemeier, A., McQueen, J. P., Shepherd, R. M., ... Porazinska, D. L. (2023). 18s-nemabase: Curated 18s rna database of nematode sequences. *Journal of Nematology*, 55(1).
- Gatzmann, F., Falckenhayn, C., Gutekunst, J., Hanna, K., Raddatz, G., Carneiro, V. C., & Lyko, F. (2018). The methylome of the marbled crayfish links gene body methylation to stable expression of poorly accessible genes. *Epigenetics & chromatin*, 11, 1–12.
- Gaunt, S. J. (2022). Seeking sense in the hox gene cluster. *Journal of Developmental Biology*, 10(4), 48.
- Geary, T. G., & Thompson, D. P. (2001). Caenorhabditis elegans: how good a model for veterinary parasites? *Veterinary parasitology*, 101(3-4), 371–386.

-
- Gems, D., & Riddle, D. L. (2000). Genetic, behavioral and environmental determinants of male longevity in *caenorhabditis elegans*. *Genetics*, *154*(4), 1597–1610.
- Ghurye, J., Rhie, A., Walenz, B. P., Schmitt, A., Selvaraj, S., Pop, M., ... Koren, S. (2019). Integrating hi-c links with assembly graphs for chromosome-scale assembly. *PLoS computational biology*, *15*(8), e1007273.
- Gilbert, C., Peccoud, J., & Cordaux, R. (2021). Transposable elements and the evolution of insects. *Annual Review of Entomology*, *66*, 355–372.
- Gillis, J. A., & Hall, B. K. (2016). A shared role for sonic hedgehog signalling in patterning chondrichthyan gill arch appendages and tetrapod limbs. *Development*, *143*(8), 1313–1317.
- Gingold, R., Moens, T., & Rocha-Olivares, A. (2013). Assessing the response of nematode communities to climate change-driven warming: a microcosm experiment. *PLoS One*, *8*(6), e66653.
- Godfray, H. C. J. (1994). *Parasitoids: behavioral and evolutionary ecology* (Vol. 12). Princeton University Press.
- Goldstein, B. (2001). On the evolution of early development in the nematoda. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *356*(1414), 1521–1531.
- Goldstein, B., Frisse, L. M., & Thomas, W. K. (1998). Embryonic axis specification in nematodes: evolution of the first step in development. *Current Biology*, *8*(3), 157–160.
- Goll, M. G., & Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.*, *74*, 481–514.
- Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C.-L., Zhang, X., ... Bestor, T. H. (2006). Methylation of trnaasp by the dna methyltransferase homolog dnmt2. *Science*, *311*(5759), 395–398.
- Gomez-Escobar, N., Gregory, W. F., Britton, C., Murray, L., Corton, C., Hall, N., ... Maizels, R. M. (2002). Abundant larval transcript-1 and-2 genes from *brugia malayi*: diversity of genomic environments but conservation of 5 promoter sequences functional in *caenorhabditis elegans*. *Molecular and biochemical parasitology*, *125*(1-2), 59–71.

-
- Grandien, K., & Sommer, R. J. (2001). Functional comparison of the nematode hox gene *lin-39* in *C. elegans* and *P. pacificus* reveals evolutionary conservation of protein function despite divergence of primary sequences. *Genes & Development*, 15(16), 2161–2172.
- Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch, I. J., ... Bennett, M. D. (2007). Eukaryotic genome size databases. *Nucleic acids research*, 35(suppl_1), D332–D338.
- Guan, D., McCarthy, S. A., Ning, Z., Wang, G., Wang, Y., & Durbin, R. (2021). Efficient iterative hi-c scaffolder based on n-best neighbors. *BMC bioinformatics*, 22, 1–16.
- Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., & Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 36(9), 2896–2898.
- Guiglielmoni, N., Rivera-Vicéns, R., Koszul, R., & Flot, J.-F. (2022). A deep dive into genome assemblies of non-vertebrate animals. *Peer Community Journal*, 2.
- Guiglielmoni, N., Villegas, L., Kirangwa, J., & Schiffer, P. (2023). Revisiting genomes of non-model species with long reads yields new insights into their biology and evolution. *bioRxiv*, 2023–10.
- Guiglielmoni, N., Villegas, L. I., Kirangwa, J., & Schiffer, P. H. (2024). Revisiting genomes of non-model species with long reads yields new insights into their biology and evolution. *Frontiers in Genetics*, 15, 1308527.
- Guijarro-Clarke, C., Holland, P. W., & Paps, J. (2020). Widespread patterns of gene loss in the evolution of the animal kingdom. *Nature Ecology & Evolution*, 4(4), 519–523.
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). Quast: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075.
- Gutierrez, A., Knoch, L., Witte, H., & Sommer, R. J. (2003). Functional specificity of the nematode hox gene *mab-5*.
- Gutierrez, A., & Sommer, R. J. (2004). Evolution of *dnmt-2* and *mbd-2*-like genes in the free-living nematodes *pristionchus pacificus*, *caenorhabditis elegans* and *caenorhabditis briggsae*. *Nucleic Acids Research*, 32(21), 6388–6396.
- Hanelt, B., Thomas, F., & Schmidt-Rhaesa, A. (2005). Biology of the phylum nematomorpha. *Advances in parasitology*, 59, 243–305.

-
- Hejnol, A., & Dunn, C. W. (2016). Animal evolution: Are phyla real? *Current Biology*, 26(10), R424–R426.
- Hellekes, V., Claus, D., Seiler, J., Illner, F., Schiffer, P. H., & Kroiher, M. (2023). Crispr/cas9 mediated gene editing in non-model nematode panagrolaimus sp. ps1159. *Frontiers in Genome Editing*, 5, 1078359.
- Henikoff, S., Greene, E. A., Pietrokovski, S., Bork, P., Attwood, T. K., & Hood, L. (1997). Gene families: the taxonomy of protein paralogs and chimeras. *science*, 278(5338), 609–614.
- Herbison, R., Evans, S., Doherty, J.-F., Algie, M., Kleffmann, T., & Poulin, R. (2019). A molecular war: convergent and ontogenetic evidence for adaptive host manipulation in related parasites infecting divergent hosts. *Proceedings of the Royal Society B*, 286(1915), 20191827.
- Herbison, R. E. H., Evans, S., Doherty, J.-F., & Poulin, R. (2019). Let’s go swimming: mermithid-infected earwigs exhibit positive hydrotaxis. *Parasitology*, 146(13), 1631–1635.
- Hiraki, H., Kagoshima, H., Kraus, C., Schiffer, P. H., Ueta, Y., Kroiher, M., . . . Kohara, Y. (2017). Genome analysis of diploscapter coronatus: insights into molecular peculiarities of a nematode with parthenogenetic reproduction. *BMC genomics*, 18, 1–18.
- Hodda, M. (2022). Phylum nematoda: a classification, catalogue and index of valid genera, with a census of valid species. *Zootaxa*, 5114(1), 1–289.
- Hodda, M., Peters, L., Traunspurger, W., et al. (2009). Nematode diversity in terrestrial, freshwater aquatic and marine systems. *Nematodes as environmental indicators*. CAB International, 45–93.
- Hodgkin, J. A., & Brenner, S. (1977). Mutations causing transformation of sexual phenotype in the nematode caenorhabditis elegans. *Genetics*, 86(2), 275–287.
- Hoff, K. J., & Stanke, M. (2013). Webaugustus—a web service for training augustus and predicting genes in eukaryotes. *Nucleic acids research*, 41(W1), W123–W128.
- Holland, P. W., Garcia-Fernández, J., Williams, N. A., & Sidow, A. (1994). Gene duplications and the origins of vertebrate development. *Development*, 1994(Supplement), 125–133.

-
- Holovachov, O., & Boström, S. (2014). Swedish Plectida (Nematoda). Part 6. *Neocamacolaimus parasiticus* gen. n., sp. n. *Zootaxa*, 3821(5), 538–550.
- Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., ... others (2002). The genome sequence of the malaria mosquito *Anopheles gambiae*. *science*, 298(5591), 129–149.
- Holterman, M., van der Wurff, A., van den Elsen, S., van Megen, H., Bongers, T., Holovachov, O., ... Helder, J. (2006). Phylum-wide analysis of ssu rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown clades. *Molecular biology and evolution*, 23(9), 1792–1800.
- Hu, J., Fan, J., Sun, Z., & Liu, S. (2020). Nextpolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*, 36(7), 2253–2255.
- Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A. O., Liang, F., ... others (2023). An efficient error correction and accurate assembly tool for noisy long reads. *bioRxiv*, 2023–03.
- Huang, S., Kang, M., & Xu, A. (2017). Haplomerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics*, 33(16), 2577–2579.
- Hughes, C. L., & Kaufman, T. C. (2002). Hox genes and the evolution of the arthropod body plan 1. *Evolution & development*, 4(6), 459–499.
- Hughes, C. L., Liu, P. Z., & Kaufman, T. C. (2004). Expression patterns of the rogue hox genes *hox3/zen* and *fushi tarazu* in the apterygote insect *Thermobia domestica*. *Evolution & development*, 6(6), 393–401.
- Hunt, D., Palomares-Rius, J., et al. (2012). General morphology and morphometrics of plant-parasitic nematodes. *Practical Plant Nematology*, 25–64.
- Hunt, V. L., Tsai, I. J., Coghlan, A., Reid, A. J., Holroyd, N., Foth, B. J., ... others (2016). The genomic basis of parasitism in the strongyloides clade of nematodes. *Nature genetics*, 48(3), 299–307.
- Ikuta, T., Yoshida, N., Satoh, N., & Saiga, H. (2004). *Ciona intestinalis* hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proceedings of the national Academy of Sciences*, 101(42), 15118–15123.
- Janssen, T., Karssen, G., Topalović, O., Coyne, D., & Bert, W. (2017). Integrative taxonomy of root-knot nematodes reveals multiple independent origins of mitotic parthenogenesis. *PLoS One*, 12(3), e0172190.

-
- Jay Burr, A., & Babinszki, C. P. (1990). Scanning motion, ocellar morphology and orientation mechanisms in the phototaxis of the nematode *Caenorhabditis elegans*. *Journal of Comparative Physiology A*, 167, 257–268.
- Jeltsch, A., Ehrenhofer-Murray, A., Jurkowski, T. P., Lyko, F., Reuter, G., Ankri, S., ... Helm, M. (2017). Mechanism and biological role of *dnmt2* in nucleic acid methylation. *RNA biology*, 14(9), 1108–1123.
- Jeltsch, A., & Jurkowska, R. Z. (2014). New concepts in dna methylation. *Trends in biochemical sciences*, 39(7), 310–318.
- Jex, A. R., Liu, S., Li, B., Young, N. D., Hall, R. S., Li, Y., ... others (2011). *Ascaris suum* draft genome. *Nature*, 479(7374), 529–533.
- Johnson, T. E., & Wood, W. B. (1982). Genetic analysis of life-span in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences*, 79(21), 6603–6607.
- Kadota, M., Nishimura, O., Miura, H., Tanaka, K., Hiratani, I., & Kuraku, S. (2020). Multifaceted hi-c benchmarking: what makes a difference in chromosome-scale genome scaffolding? *Gigascience*, 9(1), giz158.
- Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., ... others (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using *RNAi*. *Nature*, 421(6920), 231–237.
- Kampfer, S., Sturmbauer, C., & Ott, J. (1998). Phylogenetic analysis of rDNA sequences from adenophorean nematodes and implications for the adenophorea-secernentea controversy. *Invertebrate Biology*, 29–36.
- Katoh, K., Misawa, K., Kuma, K.-i., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic acids research*, 30(14), 3059–3066.
- Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772–780.
- Kaufman, T. C., Lewis, R., & Wakimoto, B. (1980). Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homoeotic gene complex in polytene chromosome interval 84a-b. *Genetics*, 94(1), 115–133.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... others (2012). Geneious basic: an integrated and extendable desktop software

-
- platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649.
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nature biotechnology*, 37(8), 907–915.
- Kiontke, K., & Fitch, D. H. (2010). Phenotypic plasticity: different teeth for different feasts. *Current Biology*, 20(17), R710–R712.
- Kiontke, K., & Fitch, D. H. (2013). Nematodes. *Current Biology*, 23(19), R862–R864.
- Kirangwa, J., Laetsch, D. R., King, E., Stevens, L., Blaxter, M., Holovachov, O., & Schiffer, P. (2023). Revisiting hox gene evolution and hox cluster linkage across nematoda. *bioRxiv*, 2023–10.
- Kolmogorov, M., Yuan, J., Lin, Y., & Pevzner, P. A. (2019). Assembly of long, error-prone reads using repeat graphs. *Nature biotechnology*, 37(5), 540–546.
- Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). mirbase: from microRNA sequences to function. *Nucleic acids research*, 47(D1), D155–D162.
- Kulakova, M. (2018). Evolutionary trends in hox cluster genes utilization: Whether common genes play by general rules? *Paleontological Journal*, 52, 1663–1671.
- Kumar, S., Koutsovoulos, G., Kaur, G., & Blaxter, M. (2012). Toward 959 nematode genomes. In *Worm* (Vol. 1, pp. 42–50).
- Kundu, R., Casey, J., & Sung, W.-K. (2019). Hypo: super fast & accurate polisher for long read genome assemblies. *Biorxiv*, 2019–12.
- Kuntz, S. G., Schwarz, E. M., DeModena, J. A., De Buysscher, T., Trout, D., Shizuya, H., ... Wold, B. J. (2008). Multigenome dna sequence conservation identifies hox cis-regulatory elements. *Genome research*, 18(12), 1955–1968.
- Lahl, V., Halama, C., & Schierenberg, E. (2003). Comparative and experimental embryogenesis of plectidae (nematoda). *Development genes and evolution*, 213, 18–27.
- Lahl, V., Schulze, J., & Schierenberg, E. (2009). Differences in embryonic pattern formation between caenorhabditis elegans and its close parthenogenetic relative diploscapter coronatus. *International Journal of Developmental Biology*, 53(4).

-
- Lambshead, P. (1993). Recent developments in marine benthic biodiversity reserch. *Oceanis*, 19, 5–24.
- Lambshead, P. J. D. (2004). Marine nematode biodiversity.
- Lambshead, P. J. D., & Boucher, G. (2003). Marine nematode deep-sea biodiversity—hyperdiverse or hype? *Journal of Biogeography*, 30(4), 475–485.
- Larsson, A. (2014). Aliview: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, 30(22), 3276–3278.
- Laumer, C. (2023). Picogram input multimodal sequencing (pimms). *Protocols.io*. Retrieved from <https://doi.org/10.17504/protocols.io.rm7vzywy51x1/v1> doi: 10.17504/protocols.io.rm7vzywy51x1/v1
- Lee, D. L. (2002a). Cuticle, moulting and exsheathment. *The biology of nematodes*, 171–209.
- Lee, D. L. (2002b). Life cycles. *The biology of nematodes*, 61–72.
- Lee, Y.-C., Lee, H.-H., Ke, H.-M., Liu, Y.-C., Wang, M.-C., Tseng, Y.-C., ... Tsai, I. J. (2023). Single worm long read sequencing reveals genome diversity in free-living nematodes. *bioRxiv*, 2023–04.
- Lewis, E. B. (1978). A gene complex controlling segmentation in drosophila. *Nature*, 276(5688), 565–570.
- Ley, P. d., & Mundo-Ocampo, M. (2004). Cultivation of nematodes.
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100.
- Li, Heng. (2012). *Seqtk*. Retrieved from <https://github.com/lh3/seqtk>
- Liao, J., Karnik, R., Gu, H., Ziller, M. J., Clement, K., Tsankov, A. M., ... others (2015). Targeted disruption of dnmt1, dnmt3a and dnmt3b in human embryonic stem cells. *Nature genetics*, 47(5), 469–478.
- Libersat, F., Delago, A., & Gal, R. (2009). Manipulation of host behavior by parasitic insects and insect parasites. *Annual Review of Entomology*, 54, 189–207.
- Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., ... Bartel, D. P. (2003). The micrnas of caenorhabditis elegans. *Genes & development*, 17(8), 991–1008.

-
- Litvaitis, M. K., Bates, J. W., Hope, W. D., & Moens, T. (2000). Inferring a classification of the adenophorea (nematoda) from nucleotide sequences of the d3 expansion segment (26/28s rdna). *Canadian Journal of Zoology*, 78(6), 911–922.
- Liu, Q. L., Thomas, V. P., & Williamson, V. M. (2007). Meiotic parthenogenesis in a root-knot nematode results in rapid genomic homozygosity. *Genetics*, 176(3), 1483–1490.
- Long, M. (2001). Evolution of novel genes. *Current opinion in genetics & development*, 11(6), 673–680.
- Luo, C., Poinar, G. O., Xu, C., Zhuo, D., Jarzembowski, E. A., & Wang, B. (2023). Widespread mermithid nematode parasitism of cretaceous insects. *Elife*, 12, e86283.
- Lyko, F. (2018). The dna methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics*, 19(2), 81–92.
- Maeyama, T., Terayama, M., & Matsumoto, T. (1994). The abnormal behavior of colobopsis sp.(hymenoptera: Formicidae) parasitized by mermis(nematoda) in papua new guinea. *Sociobiology*, 24(2), 115–120.
- Maggenti, M. A. B., Maggenti, A. R., & Gardner, S. L. (2005). Online dictionary of invertebrate zoology: C. Armand R. Maggenti Online Dictionary of Invertebrate Zoology, 18.
- Maizels, R. M., Blaxter, M. L., & Scott, A. L. (2001). Immunological genomics of brugia malayi: filarial genes implicated in immune evasion and protective immunity. *Parasite Immunology*, 23(7), 327–344.
- Malakhov, U. (1986). Nematodes: Structure, development, classification and phylogeny.
- Mapleson, D., Garcia Accinelli, G., Kettleborough, G., Wright, J., & Clavijo, B. J. (2017). Kat: a k-mer analysis toolkit to quality control ngs datasets and genome assemblies. *Bioinformatics*, 33(4), 574–576.
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770.
- Meldal, B. H., Debenham, N. J., De Ley, P., De Ley, I. T., Vanfleteren, J. R., Vierstraete, A. R., ... others (2007). An improved molecular phylogeny of the nematoda with special emphasis on marine taxa. *Molecular phylogenetics and evolution*, 42(3), 622–636.

-
- Mendes, F. K., Vanderpool, D., Fulton, B., & Hahn, M. W. (2020). Cafe 5 models variation in evolutionary rates among gene families. *Bioinformatics*, *36*(22-23), 5516–5518.
- Miljutin, D. (2013). Order Marimermithida Rubtzov 1980, emend. tchesunov 1995. In *Nematoda* (pp. 345–350).
- Minguillón, C., Gardenyes, J., Serra, E., Castro, L. F. C., Hill-Force, A., Holland, P. W., ... Garcia-Fernández, J. (2005). No more than 14: the end of the amphioxus hox cluster. *International journal of biological sciences*, *1*(1), 19.
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). Iq-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, *37*(5), 1530–1534.
- Mishina, T., Chiu, M.-C., Hashiguchi, Y., Oishi, S., Sasaki, A., Okada, R., ... others (2023). Massive horizontal gene transfer and the evolution of nematomorph-driven behavioral manipulation of mantids. *Current Biology*, *33*(22), 4988–4994.
- Mitreva, M., Jasmer, D. P., Zarlenga, D. S., Wang, Z., Abubucker, S., Martin, J., ... others (2011). The draft genome of the parasitic nematode trichinella spiralis. *Nature genetics*, *43*(3), 228–235.
- Mitreva, M., Wendl, M. C., Martin, J., Wylie, T., Yin, Y., Larson, A., ... McCarter, J. P. (2006). Codon usage patterns in nematoda: analysis based on over 25 million codons in thirty-two species. *Genome Biology*, *7*, 1–19.
- Mohamed, A. K., Burr, C., & Burr, A. J. (2007). Unique two-photoreceptor scanning eye of the nematode mermis nigrescens. *The Biological Bulletin*, *212*(3), 206–221.
- Morand, S. (2003). Flow cytometry for parasite nematode genome size measurement. *Molecular & Biochemical Parasitology*, *128*, 91–93.
- Mulhair, P. O., Crowley, L., Boyes, D. H., Harper, A., Lewis, O. T., Holland, P. W., ... others (2023). Diversity, duplication, and genomic organization of homeobox genes in lepidoptera. *Genome research*, *33*(1), 32–44.
- Mulhair, P. O., & Holland, P. W. (2024). Evolution of the insect hox gene cluster: Comparative analysis across 243 species. In *Seminars in cell & developmental biology* (Vol. 152, pp. 4–15).

-
- Müller, F., Bernard, V., & Tobler, H. (1996). Chromatin diminution in nematodes. *Bioessays*, 18(2), 133–138.
- Muñoz-Muñoz, F., Ramoneda, J., Pagès, N., Pujol, N., & Talavera, S. (2016). Is the morphology of culicoides intersexes parasitized by mermithid nematodes a parasite adaptation? a morphometric approach to culicoides circumscriptus (diptera: Ceratopogonidae). *Journal of Invertebrate Pathology*, 135, 1–9.
- Mushegian, A. R., Garey, J. R., Martin, J., & Liu, L. X. (1998). Large-scale taxonomic profiling of eukaryotic model organisms: a comparison of orthologous proteins encoded by the human, fly, nematode, and yeast genomes. *Genome Research*, 8(6), 590–598.
- Negre, B., Ranz, J. M., Casals, F., Cáceres, M., & Ruiz, A. (2003). A new split of the hox gene complex in drosophila: relocation and evolution of the gene labial. *Molecular biology and evolution*, 20(12), 2042–2054.
- Negre, B., & Ruiz, A. (2007). Hom-c evolution in drosophila: is there a need for hox gene clustering? *TRENDS in Genetics*, 23(2), 55–59.
- Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). Iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, 32(1), 268–274.
- Nickle, W. (1972). A contribution to our knowledge of the mermithidae (nematoda). *Journal of Nematology*, 4(2), 113.
- Niedermaier, J., & Moritz, K. B. (2000). Organization and dynamics of satellite and telomere dnas in ascaris: implications for formation and programmed breakdown of compound chromosomes. *Chromosoma*, 109(7), 439–452.
- Nigon, V. M., & Félix, M.-A. (2018). History of research on c. elegans and other free-living nematodes as model organisms. *WormBook: The Online Review of C. elegans Biology [Internet]*.
- Okkema, P. G. (2017). The remarkably diverse family of t-box factors in caenorhabditis elegans. *Current Topics in Developmental Biology*, 122, 27–54.
- Opperman, C. H., Bird, D. M., Williamson, V. M., Rokhsar, D. S., Burke, M., Cohn, J., ... others (2008). Sequence and genetic map of meloidogyne hapla: A compact nematode genome for plant parasitism. *Proceedings of the National Academy of Sciences*, 105(39), 14802–14807.
- Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annual review of genetics*, 34(1), 401–437.

-
- Ou, S., & Jiang, N. (2018). Ltr_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant physiology*, 176(2), 1410–1422.
- Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J. R., Hellinga, A. J., ... others (2019). Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology*, 20(1), 1–18.
- Øverli, Ø., Pall, M., Borg, B., Jobling, M., & Winberg, S. (2001). Effects of schistocephalus solidus infection on brain monoaminergic activity in female three-spined sticklebacks gasterosteus aculeatus. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1474), 1411–1415.
- Papaioannou, V. E. (2014). The t-box gene family: emerging roles in development, stem cells and cancer. *Development*, 141(20), 3819–3833.
- Parkinson, J., Mitreva, M., Whitton, C., Thomson, M., Daub, J., Martin, J., ... others (2004). A transcriptomic analysis of the phylum nematoda. *Nature genetics*, 36(12), 1259–1267.
- Perry, R. N., & Moens, M. (2013). *Plant nematology*. Cabi.
- Poinar, J., GO. (1983). The natural history of nematodes.
- Poinar Jr, G. O., & Hess, R. (1977). Romanomermis culicivorax: morphological evidence of transcuticular uptake. *Experimental Parasitology*, 42(1), 27–33.
- Ponton, F., Otálora-Luna, F., Lefevre, T., Guerin, P. M., Lebarbenchon, C., Duneau, D., ... Thomas, F. (2011). Water-seeking behavior in worm-infected crickets and reversibility of parasitic manipulation. *Behavioral Ecology*, 22(2), 392–400.
- Poulin, R. (1998). Evolution and phylogeny of behavioural manipulation of insect hosts by parasites. *Parasitology*, 116(S1), S3–S11.
- Poulin, R. (2011). The many roads to parasitism: a tale of convergence. *Advances in parasitology*, 74, 1–40.
- Powers, T. P., Hogan, J., Ke, Z., Dymbrowski, K., Wang, X., Collins, F. H., & Kaufman, T. C. (2000). Characterization of the hox cluster from the mosquito anopheles gambiae (diptera: Culicidae). *Evolution & Development*, 2(6), 311–325.

-
- Priest, T., Orellana, L. H., Huettel, B., Fuchs, B. M., & Amann, R. (2021). Microbial metagenome-assembled genomes of the fram strait from short and long read sequencing platforms. *PeerJ*, 9, e11721.
- Provataris, P., Meusemann, K., Niehuis, O., Grath, S., & Misof, B. (2018). Signatures of dna methylation across insects suggest reduced dna methylation levels in holometabola. *Genome biology and evolution*, 10(4), 1185–1197.
- Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., ... others (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature*, 453(7198), 1064–1071.
- Qing, X., Zhang, Y. M., Sun, S., Ahmed, M., Lo, W.-S., Bert, W., ... Li, H. (2023). Phylogenomic insights into the evolution and origin of nematoda. *bioRxiv*, 2023–12.
- Quicke, D. L., Ghafouri Moghaddam, M., & Butcher, B. A. (2023). Dietary challenges for parasitoid wasps (hymenoptera: Ichneumonoidea); coping with toxic hosts, or not? *Toxins*, 15(7), 424.
- Quinlan, A. R., & Hall, I. M. (2010). Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842.
- Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). Genomescope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *Nature communications*, 11(1), 1432.
- Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., ... others (2014). A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, 159(7), 1665–1680.
- Ren, W., Gao, L., & Song, J. (2018). Structural basis of dnmt1 and dnmt3a-mediated dna methylation. *Genes*, 9(12), 620.
- Riddle, D. L., Blumenthal, T., Meyer, B. J., & Priess, J. R. (1997). *C. elegans* ii.
- Riera, P., & Hubas, C. (2003). Trophic ecology of nematodes from various microhabitats of the roscoff aber bay (france): importance of stranded macroalgae evidenced through $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$. *Marine Ecology Progress Series*, 260, 151–159.
- Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC bioinformatics*, 19, 1–10.

-
- Robinson, A. F., Baker, G. L., & Heald, C. M. (1990). Transverse phototaxis by infective juveniles of *agamermis* sp. and *hexamermis* sp. *The Journal of parasitology*, 147–152.
- Roeber, F., Jex, A. R., & Gasser, R. B. (2013). Impact of gastrointestinal parasitic nematodes of sheep, and the role of advanced molecular tools for exploring epidemiology and drug resistance-an australian perspective. *Parasites & vectors*, 6(1), 1–13.
- Rošić, S., Amouroux, R., Requena, C. E., Gomes, A., Emperle, M., Beltran, T., ... others (2018). Evolutionary analysis indicates that dna alkylation damage is a byproduct of cytosine dna methyltransferase activity. *Nature genetics*, 50(3), 452–459.
- Rossiter, A., & Kawanabe, H. (2000). *Ancient lakes: biodiversity, ecology and evolution*. (Vol. 31).
- Rowell, C. (2000). Presumptive mermithid-induced intersex individuals in the neotropical grasshopper genus *drymophilacris* descamps 1976. *Journal of Orthoptera Research*, 31–35.
- Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. *Nature methods*, 17(2), 155–158.
- Rutherford, T., Webster, J., & Barlow, J. (1977). Physiology of nutrient uptake by the entomophilic nematode *mermis nigrescens* (mermithidae). *Canadian Journal of Zoology*, 55(11), 1773–1781.
- Ryan, J. F., & Baxevanis, A. D. (2007). Hox, wnt, and the evolution of the primary body axis: insights from the early-divergent phyla. *Biology Direct*, 2, 1–11.
- Sapir, A. (2021). Why are nematodes so successful extremophiles? *Communicative & Integrative Biology*, 14(1), 24–26.
- Sato, T., Egusa, T., Fukushima, K., Oda, T., Ohte, N., Tokuchi, N., ... Lafferty, K. D. (2012). Nematomorph parasites indirectly alter the food web and ecosystem function of streams through behavioural manipulation of their cricket hosts. *Ecology Letters*, 15(8), 786–793.
- Schaefer, M., & Lyko, F. (2010). Solving the dnmt2 enigma. *Chromosoma*, 119, 35–40.
- Schierenberg, E. (2001). Three sons of fortune: early embryogenesis, evolution and ecology of nematodes. *Bioessays*, 23(9), 841–847.

-
- Schierenberg, E. (2005). Unusual cleavage and gastrulation in a freshwater nematode: developmental and phylogenetic implications. *Development genes and evolution*, 215, 103–108.
- Schierenberg, E. (2006). Embryological variation during nematode development. *WormBook*, 1.
- Schierenberg, E., & Lahl, V. (2004). Embryology and phylogeny of nematodes. In *Proceedings of the fourth international congress of nematology, 8-13 june 2002, tenerife, spain* (pp. 667–679).
- Schiffer, P. H., Danchin, E. G., Burnell, A. M., Creevey, C. J., Wong, S., Dix, I., ... others (2019). Signatures of the evolution of parthenogenesis and cryptobiosis in the genomes of panagrolaimid nematodes. *IScience*, 21, 587–602.
- Schiffer, P. H., Kroiher, M., Kraus, C., Koutsovoulos, G. D., Kumar, S., R Camps, J. I., ... others (2013). The genome of romanomermis culicivox: revealing fundamental changes in the core developmental genetic toolkit in nematoda. *BMC genomics*, 14, 1–16.
- Schiffer, P. H., Polsky, A. L., Cole, A. G., Camps, J. I., Kroiher, M., Silver, D. H., ... others (2018). The gene regulatory program of acrobelloides nanus reveals conservation of phylum-specific expression. *Proceedings of the National Academy of Sciences*, 115(17), 4459–4464.
- Schmidt-Rhaesa, A. (2004). Ultrastructure of an integumental organ with probable sensory function in paragordius varius (nematomorpha). *Acta Zoologica*, 85(1), 15–19.
- Schmidt-Rhaesa, A. (2005). Morphogenesis of paragordius varius (nematomorpha) during the parasitic phase. *Zoomorphology*, 124, 33–46.
- Schneider, C., Woehle, C., Greve, C., D’Haese, C. A., Wolf, M., Hiller, M., ... Huettel, B. (2021). Two high-quality de novo genomes from single ethanol-preserved specimens of tiny metazoans (collembola). *GigaScience*, 10(5), giab035.
- Schratzberger, M., Holterman, M., van Oevelen, D., & Helder, J. (2019). A worm’s world: Ecological flexibility pays off for free-living nematodes in sediments and soils. *BioScience*, 69(11), 867–876.
- Schulz, N. K., Wagner, C. I., Ebeling, J., Raddatz, G., Diddens-de Buhr, M. F., Lyko, F., & Kurtz, J. (2018). Dnmt1 has an essential function despite the

absence of cpg dna methylation in the red flour beetle *tribolium castaneum*. *Scientific reports*, 8(1), 16462.

Schulze, J., & Schierenberg, E. (2008). Cellular pattern formation, establishment of polarity and segregation of colored cytoplasm in embryos of the nematode *romanomermis culicivorax*. *Developmental biology*, 315(2), 426–436.

Schulze, J., & Schierenberg, E. (2009). Embryogenesis of *romanomermis culicivorax*: an alternative way to construct a nematode. *Developmental Biology*, 334(1), 10–21.

Schulze, J., & Schierenberg, E. (2011). Evolution of embryonic development in nematodes. *EvoDevo*, 2, 1–17.

Seo, H.-C., Edvardsen, R. B., Maeland, A. D., Bjordal, M., Jensen, M. F., Hansen, A., ... others (2004). Hox cluster disintegration with persistent anteroposterior order of expression in *oikopleura dioica*. *Nature*, 431(7004), 67–71.

Shamseldean, M., & Platzer, E. (1989). *Romanomermis culicivorax*: penetration of larval mosquitoes. *Journal of invertebrate pathology*, 54(2), 191–199.

Siddiqi, M., et al. (1980). The origin and phylogeny of the nematode orders tylenchida thorne, 1949 and aphelenchida n. ord. In *Helminthological abstracts. series b. plant nematology* (Vol. 49, pp. 143–170).

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.

Simons, A. (2010). A quality control tool for high throughput sequence data. *A quality control tool for high throughput sequence data*, 10, f1000research.

Singaravelu, G., & Singson, A. (2011). New insights into the mechanism of fertilization in nematodes. *International review of cell and molecular biology*, 289, 211–238.

Skaling, B., & MacKinnon, B. (1988). The absorptive surfaces of *Nectonema* sp.(nematomorpha: Nectonematoidea) from *Pandalus montagui*: histology, ultrastructure, and absorptive capabilities of the body wall and intestine. *Canadian journal of zoology*, 66(2), 289–295.

Skiba, F., & Schierenberg, E. (1992). Cell lineages, developmental timing, and spatial pattern formation in embryos of free-living soil nematodes. *Developmental biology*, 151(2), 597–610.

-
- Slack, J. M., Holland, P. W., & Graham, C. F. (1993). The zootype and the phylotypic stage. *Nature*, *361*(6412), 490–492.
- Smith, F. W., Boothby, T. C., Giovannini, I., Rebecchi, L., Jockusch, E. L., & Goldstein, B. (2016). The compact body plan of tardigrades evolved by the loss of a large body region. *Current Biology*, *26*(2), 224–229.
- Smith, Z. D., & Meissner, A. (2013). Dna methylation: roles in mammalian development. *Nature Reviews Genetics*, *14*(3), 204–220.
- Smythe, A. B., Holovachov, O., & Kocot, K. M. (2019). Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evolutionary Biology*, *19*(1), 1–15.
- Sokolova, E. A., Zograf, J. K., & Yushin, V. V. (2022). Ultrastructure of spermatozoa of a hairworm *gordionus alpestris* (villot, 1885)(nematomorpha, chordodidae). *Invertebrate Reproduction & Development*, *66*(2), 88–96.
- Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., . . . others (2003). The genome sequence of *caenorhabditis briggsae*: a platform for comparative genomics. *PLoS biology*, *1*(2), e45.
- Stevens, L., Félix, M.-A., Beltran, T., Braendle, C., Caurcel, C., Fausett, S., . . . others (2019). Comparative genomics of 10 new *caenorhabditis* species. *Evolution Letters*, *3*(2), 217–236.
- Streit, A. (2008). Reproduction in strongyloides (nematoda): a life between sex and parthenogenesis. *Parasitology*, *135*(3), 285–294.
- Struhl, G. (1981). A homoeotic mutation transforming leg to antenna in *drosophila*. *Nature*, *292*(5824), 635–638.
- Su, W., Gu, X., & Peterson, T. (2019). Tir-learner, a new ensemble method for tir transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. *Molecular plant*, *12*(3), 447–460.
- Su, Z., Han, L., & Zhao, Z. (2011). Conservation and divergence of dna methylation in eukaryotes: new insights from single base-resolution dna methylomes. *Epigenetics*, *6*(2), 134–140.
- Sudhaus, W., & Fitch, D. (2001). Comparative studies on the phylogeny and systematics of the rhabditidae (nematoda). *Journal of nematology*, *33*(1), 1.
- Sulston, J., & Hodgkin, J. (1988). *The nematode caenorhabditis elegans*.

-
- Sulston, J. E., Schierenberg, E., White, J. G., & Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental biology*, 100(1), 64–119.
- Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., & Wang, Y. (2023). Orthovenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Research*, gkad313.
- Tchesunov, A. V., Nikolaeva, O. V., Rusin, L. Y., Sanamyan, N. P., Panina, E. G., Miljutin, D. M., ... others (2023). Paraphyly of Marimermithida refines primary routes of transition to parasitism in roundworms. *Zoological Journal of the Linnean Society*, 197(4), 909–923.
- Thomas, F., Schmidt-Rhaesa, A., Martin, G., Manu, C., Durand, P., & Renaud, F. (2002). Do hairworms (nematomorpha) manipulate the water seeking behaviour of their terrestrial hosts? *Journal of Evolutionary Biology*, 15(3), 356–361.
- Thomas, F., Ulitsky, P., Augier, R., Dusticier, N., Samuel, D., Strambi, C., ... Cayre, M. (2003). Biochemical and histological changes in the brain of the cricket *Nemobius sylvestris* infected by the manipulative parasite *Paragordius tricuspidatus* (nematomorpha). *International Journal for Parasitology*, 33(4), 435–443.
- Tihanyi, B., Vellai, T., Regős, Á., Ari, E., Müller, F., & Takács-Vellai, K. (2010). The *C. elegans* hox gene *ceh-13* regulates cell migration and fusion in a non-colinear way. implications for the early evolution of hoxclusters. *BMC developmental biology*, 10(1), 1–14.
- Troemel, E. R. (1999). Chemosensory signaling in *C. elegans*. *Bioessays*, 21(12), 1011–1020.
- Uniprot: the universal protein knowledgebase in 2023. (2023). *Nucleic acids research*, 51(D1), D523–D531.
- Van Auken, K., Weaver, D. C., Edgar, L. G., & Wood, W. B. (2000). *Caenorhabditis elegans* embryonic axial patterning requires two recently discovered posterior-group hox genes. *Proceedings of the National Academy of Sciences*, 97(9), 4499–4503.
- Vance, S. A. (1996). Morphological and behavioural sex reversal in mermithid-infected mayflies. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 263(1372), 907–912.

-
- Van Den Hoogen, J., Geisen, S., Routh, D., Ferris, H., Traunspurger, W., Wardle, D. A., ... others (2019). Soil nematode abundance and functional group composition at a global scale. *Nature*, 572(7768), 194–198.
- van Houte, S., Ros, V. I., & van Oers, M. M. (2013). Walking with insects: molecular mechanisms behind parasitic manipulation of host behaviour. *Molecular ecology*, 22(13), 3458–3475.
- van Megen, H., van den Elsen, S., Holterman, M., Karssen, G., Mooyman, P., Bongers, T., ... Helder, J. (2009). A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal dna sequences. *Nematology*, 11(6), 927–950.
- Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome research*, 27(5), 737–746.
- Viney, M. (2017). How can we understand the genomic basis of nematode parasitism? *Trends in parasitology*, 33(6), 444–452.
- Vizueta, J., Sánchez-Gracia, A., & Rozas, J. (2020). *Bitacora: A comprehensive tool for the identification and annotation of gene families in genome assemblies*. Wiley Online Library.
- Von Allmen, G., Spierer, A., Karch, F., Bender, W., Gyurkovics, H., Lewis, E., et al. (1996). Splits in fruitfly hox gene complexes. *Nature*, 380(6570), 116.
- von Lieven, A. F. (2002). The sister group of the diplogastrina (nematoda). *Russian Journal of Nematology*, 10(2), 127–138.
- Von Lieven, A. F., & Sudhaus, W. (2000). Comparative and functional morphology of the buccal cavity of diplogastrina (nematoda) and a first outline of the phylogeny of this taxon. *Journal of Zoological Systematics and Evolutionary Research*, 38(1), 37–63.
- Voronov, D. A. (1999). The embryonic development of pontonema vulgare (enoplida: Oncholaimidae) with a discussion of nematode phylogeny. *Russian Journal of Nematology*, 7, 105–114.
- Voronov, D. A., & Panchin, Y. V. (1998). Cell lineage in marine nematode enoplus brevis. *Development*, 125(1), 143–150.
- Wagner, G. P., Amemiya, C., & Ruddle, F. (2003). Hox cluster duplications and the opportunity for evolutionary novelties. *Proceedings of the National Academy of Sciences*, 100(25), 14603–14606.

-
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., ... others (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one*, 9(11), e112963.
- Walton, A. C., et al. (1959). Some parasites and their chromosomes. *Journal of Parasitology*, 45(1), 1–20.
- Wang, J., & Kim, S. K. (2003). Global analysis of dauer gene expression in *Caenorhabditis elegans*.
- Wang, M. C., O'Rourke, E. J., & Ruvkun, G. (2008). Fat metabolism links germline stem cells and longevity in *C. elegans*. *Science*, 322(5903), 957–960.
- Wasmuth, J., Schmid, R., Hedley, A., & Blaxter, M. (2008). On the extent and origins of genic novelty in the phylum nematoda. *PLoS neglected tropical diseases*, 2(7), e258.
- WB, B. S. W. (1988). The nematode *Caenorhabditis elegans*. *Cold Spring Harb. Laboratory*, 1, 988.
- Weber, M., & Schübeler, D. (2007). Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Current opinion in cell biology*, 19(3), 273–280.
- Wharton, D. A. (2012). *A functional biology of nematodes*. Springer Science & Business Media.
- White, J. G., Southgate, E., Thomson, J. N., Brenner, S., et al. (1986). The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Philos Trans R Soc Lond B Biol Sci*, 314(1165), 1–340.
- Whitehead, A., & Hemming, J. (1965). A comparison of some quantitative methods of extracting small vermiform nematodes from soil. *Annals of applied Biology*, 55(1), 25–38.
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., ... others (2007). A unified classification system for eukaryotic transposable elements. *Nature reviews genetics*, 8(12), 973–982.
- Wiegner, O., & Schierenberg, E. (1999). Regulative development in a nematode embryo: a hierarchy of cell fate transformations. *Developmental biology*, 215(1), 1–12.

-
- Williams, C., Poulin, R., & Sinclair, B. (2004). Increased haemolymph osmolality suggests a new route for behavioural manipulation of talorchestia quoyana (amphipoda: Talitridae) by its mermithid parasite. *Functional Ecology*, 685–691.
- Witherspoon, D. J., & Robertson, H. M. (2003). Neutral evolution of ten types of mariner transposons in the genomes of caenorhabditis elegans and caenorhabditis briggsae. *Journal of molecular evolution*, 56, 751–769.
- Wood, W. B. (1988). Determination of pattern and fate in early embryos of caenorhabditis elegans. *The Molecular Biology of Cell Determination and Cell Differentiation*, 57–78.
- Wurm, Y., & Keller, L. (2010). Parasitoid wasps: from natural history to genomic studies. *Current Biology*, 20(5), R242–R244.
- Xiong, W., He, L., Lai, J., Dooner, H. K., & Du, C. (2014). Helitronscanner uncovers a large overlooked cache of helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences*, 111(28), 10263–10268.
- Xu, G., Teng, Z.-W., Gu, G.-X., Guo, L., Wang, F., Xiao, S., ... others (2020). Genomic and transcriptomic analyses of glutathione s-transferases in an endoparasitoid wasp, pteromalus puparum. *Archives of Insect Biochemistry and Physiology*, 103(2), e21634.
- Xu, X., Li, G., Li, C., Zhang, J., Wang, Q., Simmons, D. K., ... others (2019). Evolutionary transition between invertebrates and vertebrates via methylation reprogramming in embryogenesis. *National Science Review*, 6(5), 993–1003.
- Yarychkivska, O., Shahabuddin, Z., Comfort, N., Boulard, M., & Bestor, T. H. (2018). Bah domains and a histone-like motif in dna methyltransferase 1 (dnmt1) regulate de novo and maintenance methylation in vivo. *Journal of Biological Chemistry*, 293(50), 19466–19475.
- Yasukochi, Y., Ashakumary, L. A., Wu, C., Yoshido, A., Nohata, J., Mita, K., & Sahara, K. (2004). Organization of the hox gene cluster of the silkworm, bombyx mori: a split of the hox cluster in a non-drosophila insect. *Development genes and evolution*, 214, 606–614.
- Ya’cob, Z., Low, V. L., Tan, T. K., Noor-Izwan, A., Lourdes, E. Y., Ramli, R., ... Adler, P. H. (2021). Sexually anomalous individuals of the black fly simulium trangense (diptera: Simuliidae) infected with mermithid parasites (nematoda: Mermithidae). *Parasitology Research*, 120, 1555–1561.

-
- Yeates, G. (1987). Nematode feeding and activity: the importance of development stages. *Biology and fertility of soils*, 3, 143–146.
- Yeates, G. W., & Boag, B. (2004). Background for nematode ecology in the 21st century. *Nematology: advances and perspectives*, 1, 406–437.
- Yoshida, Y., Koutsovoulos, G., Laetsch, D. R., Stevens, L., Kumar, S., Horikawa, D. D., ... others (2017). Comparative genomics of the tardigrades *hypsibius dujardini* and *ramazzottius varieornatus*. *PLoS biology*, 15(7), e2002266.
- Zemach, A., & Zilberman, D. (2010). Evolution of eukaryotic dna methylation and the pursuit of safer sex. *Current Biology*, 20(17), R780–R785.
- Zhang, Y., Li, S., Li, H., Wang, R., Zhang, K.-Q., & Xu, J. (2020). Fungi–nematode interactions: Diversity, ecology, and biocontrol prospects in agriculture. *Journal of Fungi*, 6(4), 206.
- Zhang, Z.-Q. (2013). Animal biodiversity: An update of classification and diversity in 2013. in: Zhang, z.-q.(ed.) animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (addenda 2013). *Zootaxa*, 3703(1), 5–11.
- Zheng, C., Lee, H. M. T., & Pham, K. (2022). Nervous system-wide analysis of hox regulation of terminal neuronal fate specification in *caenorhabditis elegans*. *PLoS Genetics*, 18(2), e1010092.
- Zhong, Y.-F., Butts, T., & Holland, P. W. (2008). Homeodb: a database of homeobox gene diversity. *Evolution & development*, 10(5), 516–518.
- Zhou, C., McCarthy, S. A., & Durbin, R. (2023). Ychs: yet another hi-c scaffolding tool. *Bioinformatics*, 39(1), btac808.
- Zvelebil, M. J., Tang, L., Cookson, E., Selkirk, M. E., & Thornton, J. M. (1993). Molecular modelling and epitope prediction of gp29 from lymphatic filariae. *Molecular and biochemical parasitology*, 58(1), 145–153.

Curriculum Vitae

Joseph Kirangwa

Wittekindstr 6, 50937, Koeln

Email: jkirangw@uni-koeln.de **Mobile:** +49015140364677



EDUCATION

Ph.D studies in Nematode Biodiversity Evolutionary Genomics

University of Cologne, Germany

Until: May 2024

Master of Science with major in Bioinformatics

University of Skövde and Karolinska Institute, Sweden

Graduated: June 2017

Biomedical Laboratory Technology

Makerere University, Uganda

Graduated: April 2013

EXPERIENCE

Graduate Research Assistant

Institute of Zoology

University of Cologne, Germany

June 2020 - May 2024

- Nematode sampling, identification and laboratory culture
- Setting-up protocols for HMW DNA extraction and RNA extraction
- Whole genome amplification from single microscopic nematode, 18S and 28S rRNA molecular barcode PCR
- Illumina, Arima Hi-C, PacBio HiFi and Oxford nanopore library preparation and sequencing (Miseq and MinION)
- Genome assembly, scaffolding, decontamination and annotation using long read input sequences
- Teaching and supervising students in molecular genomics laboratory work and bioinformatics processes

Laboratory Technologist

Center for Translational Microbiome Research

Karolinska Institute, Sweden

June 2017 - December 2019

- DNA extraction and 16S rRNA sequencing (MiSeq sequencing)
- Illumina TruSeq DNA library preparation
- Development of bioinformatics analysis pipeline: BACTpipe (Bacterial Genome quality control, assembly and annotation)

Laboratory Technologist

Medical Research Council

MRC/UVRI & LSHTM Research Unit, Uganda

April 2013 - December 2015

- Proviral HIV-1 DNA extraction and Sanger sequencing (ABI 3500 Genetic Analyser)
- Profiling HIV-1 drug resistance mutations
- Supervision of student interns

TRAININGS

Meiofauna Diversity and Taxonomy Workshop

Dauphin Island Sea Lab (DISL), Alabama, USA

May 2023

HiC library preparation training

Metazoa Phylogenomics Lab, Institute of Evolutionary Biology, Barcelona, Spain

May 2022

Next Generation Sequencing training

West German Genome Center, Düsseldorf, Germany

September 2021

Establishment of HMW DNA extraction methods for single nematode

LOEWE Centre for Translational Biodiversity Genomics, Senckenberg laboratory, Frankfurt, Germany

July 2021

Nematode sampling, preservation and identification

The Swedish Museum of Natural History, Stockholm, Sweden

June 2020

CONFERENCES

24th International *C. elegans* conference, Glasgow, Scotland

Poster, June 2023

Evolutionary plasticity in nematode HOX gene complements and genomic loci arrangement

Long read sequencing conference, Uppsala, Sweden

Poster, November 2022

Chromosomal scale genome assembly of *Romanomermis culicivorax*

114th Annual meeting of the German Zoological Society, Bonn, Germany

Poster, September 2022

Revisiting Hox gene complements across the Nematoda

Reference Genomes for Biodiversity Symposium, Cologne, Germany

Poster, September 2022

First results from the Deutscher ReferenzGenom Atlas (DeRGA) pilot project

TECHNICAL SKILLS AND INTERESTS

- Programming languages: Bash scripting, Python, R
- Databases: ENA, Uniprot, NCBI SRA
- Other Tools: Nextflow, Git, Github, Docker, Galaxy
- Areas of Interest: Nematode Genomics, Transcriptomics, Epigenomics, Bioinformatics

SCHOLARSHIP AWARDS

- MSc funding award from Swedish Institute Scholarships for Global Professionals
- BSc funding award from The Muljibhai Madhvani Foundation Scholarship