BMC Genomics

RESEARCH Open Access

Chromosome-level genome assembly of the nematophagous flatworm *Luticola nematophagus*: revealing molecular adaptations for predation and its biocontrol potential against nematode diseases

Chongtao Guo[1,2,3†], Wang Yin[1,2,3,4†], Zhouqiong Zhang[1,2,3], Renju Deng[1,2,3*], Shouhui Pan[5], Hai Zhang[5], Hao Cen[5], Quan Zhang[5], Fei Dai[5†], Congyong Li[5] and Qing Yang[5]

Abstract

Background The soil-dwelling Rhabdocoela flatworm, *Luticola nematophagus*, recently identified as a specialized predator of plant-parasitic nematodes such as *Meloidogyne incognita*, demonstrates considerable potential as a biocontrol agent for agricultural nematode diseases. This study presents the first chromosome-level genome assembly of *L. nematophagus*, achieved by integrating PacBio HiFi long-read sequencing with Hi-C chromatin conformation capture.

Results The assembled genome is approximately 899.92 Mb with a scaffold N50 of 231.83 Mb, comprising 50.08% repetitive sequences and 45,225 predicted protein-coding genes. Genome completeness was evaluated using CEGMA (94.76%) and BUSCO (85.49%) assessments. Comparative genomic analysis revealed notable expansions in gene families related to neural signaling, environmental adaptation, and nematode predation, including synaptic vesicle components, metalloendopeptidases, antioxidant enzymes, transmembrane transporters, peptidases, lipases, and detoxification mechanisms. These expansions suggest that *L. nematophagus* has evolved molecular adaptations favorable for thriving in soil ecosystems and for a specialized predation strategy, possibly aiding in the digestion of nematode cuticles and the neutralization of toxins. Interestingly, unlike Triclad flatworms such as *Dugesia japonica* and *Schmidtea mediterranea*, *L. nematophagus* has lost regenerative capabilities, instead allocating resources towards reproductive adaptations through self-fertilization and high fecundity. Controlled pot trials have experimentally validated its effectiveness in suppressing tomato root-knot nematode disease, underscoring its high reproductive output as critical for maintaining predator populations needed for effective biocontrol.

†Chongtao Guo and Wang Yin contributed equally to this work.

*Correspondence:
Renju Deng
xiaoju63@163.com

Full list of author information is available at the end of the article

Conclusions This research provides essential genomic resources for understanding the evolutionary mechanisms of nematode predation within Platyhelminthes and promotes the development of *L. nematophagus* as a scalable biocontrol agent for managing nematode infections.

Keywords *Luticola nematophagus*, Rhabdocoela, Chromosome-level genome assembly, Nematode predation, Biocontrol, Comparative genomics

Background

Although genomic resources are readily available for regenerative planarians, such as *Schmidtea mediterranea*, predatory species including *Luticola nematophagus*, which exhibit unique adaptations for nematode biocontrol, are yet to be genetically characterized. This paucity of genetic data hinders both evolutionary research and practical applications in biocontrol. Plant-parasitic nematodes, notably the root-knot nematode *Meloidogyne incognita*, cause substantial agricultural losses worldwide, thereby threatening food security and inflicting economic damages amounting to billions annually [1–3]. Biological control emerges as a promising sustainable alternative to chemical nematicides, yet the discovery and development of effective, scalable, and environmentally benign biocontrol agents face considerable challenges [4, 5]. Recently identified, the soil-dwelling flatworm *L. nematophagus* (Platyhelminthes: Rhabditophora: Rhabdocoela) has shown great promise as a biocontrol agent [6]. This species demonstrates specific predatory behavior toward plant-parasitic nematodes (Fig. 1A), and preliminary pot trials have shown significant efficacy in controlling tomato root-knot nematode disease (Fig. S1 E-I). Importantly, *L. nematophagus* possesses two critical traits for effective biocontrol: adaptability to soil environments and a high reproductive rate through self-fertilization, which facilitates large-scale cultivation.

Flatworms (Platyhelminthes: Rhabditophora) are an early-diverging group of triploblastic, bilaterally symmetrical animals [7], with a global distribution across tens of thousands of diverse species [8, 9]. Species such as *Dugesia japonica*, *S. mediterranea* (Tricladida), and *Macrostomum lignano* (Macrostomida) have been established as model organisms, largely due to their remarkable regenerative abilities, driven by abundant pluripotent neoblasts [10–12]. Studies utilizing these models have provided profound insights into stem cell biology, regeneration mechanisms, neurodevelopment, and diseases relevant to humans [13–15]. Over approximately 40 million years of evolution, flatworms have developed effective predatory capabilities, enabling them to adapt to diverse and fluctuating ecological niches [16, 17].

In addition to their regenerative capacities, flatworms serve as active predators that contribute to the management of various invertebrate pests in aquatic and terrestrial ecosystems [18], including mosquitoes, snails, slugs, and termites [19–21]. For example, the planarian *Girardia anceps* (Tricladida: Dugesiidae) has been documented to effectively reduce populations of mosquito larvae, including those of *Aedes aegypti* and *Culex pipiens* [22]. Despite extensive research into regeneration, the molecular mechanisms underlying flatworm predation and their ecological interactions remain underexplored, particularly in non-model species that show significant biocontrol potential. The literature over the past decade reports sparingly on the use of flatworms for pest management, and fundamental questions about their predation mechanisms and environmental adaptability continue to persist [21]. A primary obstacle is the absence of comprehensive genomic resources for these non-model, biocontrol-relevant flatworm species.

In contrast to more generalist predators, *L. nematophagus* demonstrates remarkable specificity toward plant-parasitic nematodes, underscoring its potential as a targeted biocontrol agent. However, the molecular underpinnings of this specialization remain elusive, owing to the lack of genomic resources, a situation that epitomizes the existing knowledge gap. Recently characterized, the Rhabdocoelan flatworm genus and species *L. nematophagus*, which inhabits soils infested with nematodes [6], possesses a distinct set of characteristics that differentiates it from conventional model planarians: (1) Absence of regenerative capability: Unlike species in the Tricladida and Macrostomida, *L. nematophagus* exhibits a loss of regenerative ability, likely attributable to a paucity of neoblasts. (2) Soil adaptation: It has evolved to thrive in moist soil environments. (3) Specialized predation on nematodes: This species actively preys on nematodes, specifically targeting plant-parasitic varieties such as *M. incognita*. (4) High reproductive potential through self-fertilization: By allocating resources predominantly toward reproduction rather than regeneration, *L. nematophagus* achieves efficient reproduction, facilitating mass production with potential annual yields of up to 200 million eggs. This unique amalgamation of traits not only positions *L. nematophagus* as an outstanding candidate for the biological control of nematode diseases but also as a compelling model for studying the evolutionary trade-offs among regeneration, predation specialization, and environmental adaptation. These characteristics collectively underscore its biocontrol capabilities, yet they also highlight the need for a chromosome-level genome to elucidate: (1) the genetic trade-offs between the loss of regeneration and reproductive adaptations, (2) the

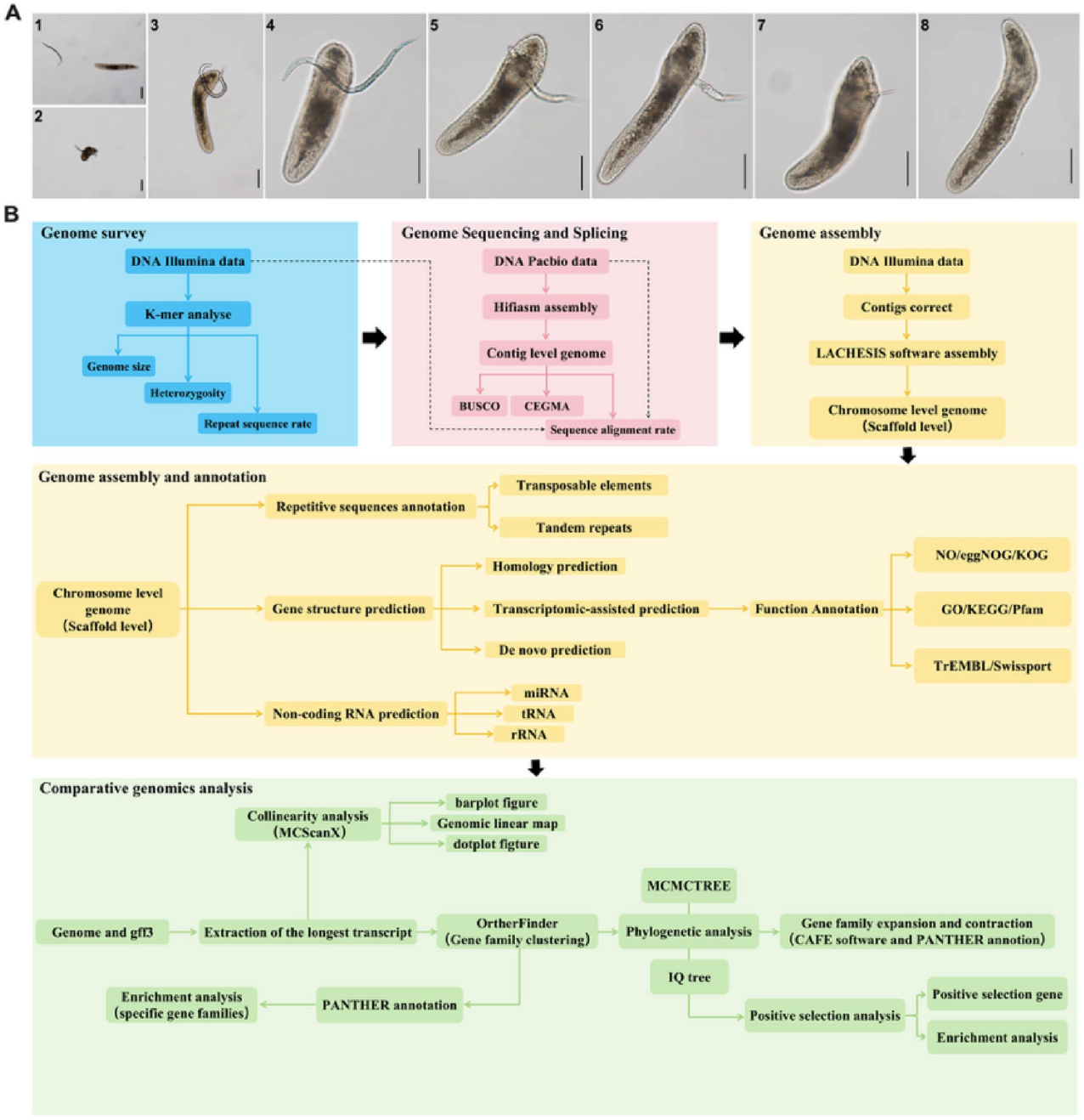

Fig. 1 Images recording of *L.nematophagus* preying on *M.incognita* and the workflow of this study. **A** The process of *L. nematophagus* preying on *M. incognita*. 1 for the discovery and targeting process. 2–3 for the process of *L. nematophagus* pharyngeal muscles extending and binding *M. incognita*. 4–6 for the process of *L. nematophagus* gradually devouring *M. incognita*. 7–8 for the completion of the *L. nematophagus* predation process, and looking for the next prey. For the operational steps of the observation process, please refer to the Supplementary Materials. **B** The panes with light blue, pink, orange, and light green represent the processes of genome survey, genome sequencing and splicing, genome assembly and annotation, and comparative genomics analysis, respectively

mechanisms enabling survival in soil environments, and (3) the strategies underlying nematode-specific predation.

To effectively advance both fundamental evolutionary biology and the development of applied biocontrol, it is imperative to elucidate the molecular genetic basis underpinning the key traits of *L. nematophagus*. These traits include its mechanisms for nematode predation, adaptation to complex soil environments, and the evolutionary trade-off that favors high fecundity over regeneration. While complete genome assemblies are available for model species such as *D. japonica* [23, 24],

S. mediterranea [25], and *M. lignano* [26], most Rhabdocoela species, which exhibit limited tissue repair capabilities [27, 28], lack published genome sequences. Similarly, the absence of a high-quality, chromosome-level genome assembly for *L. nematophagus* represents a significant obstacle. To address this deficiency and unlock the potential of this novel biocontrol agent, we generated the first chromosome-level genome assembly of *L. nematophagus*. We employed an integrated approach that combined PacBio HiFi long-read sequencing with Hi-C chromatin conformation capture technology. This methodology was chosen to overcome the challenges posed by its complex genome. A comparative genomic analysis was also conducted using selected species. The workflow of this study is illustrated in Fig. 1B. This high-quality genomic resource is foundational for several critical research objectives: (1) Deciphering the molecular adaptations that underlie its specialized nematode predation strategy and resilience to soil environmental conditions. (2) Investigating the genetic basis of its unique life-history traits, which include loss of regeneration, self-fertilization, and high fecundity. (3) Identifying candidate genes and pathways, such as those encoding proteases, lipases, detoxification enzymes, and neural signaling components, that could contribute to its biocontrol function. (4) Ultimately, facilitating the development of *L. nematophagus*-based strategies for sustainable management of plant-parasitic nematodes. This study not only fills a critical void in genomic resources for non-model, predatory flatworms but also provides pivotal insights into the evolution of predation and environmental adaptation within the Platyhelminthes, specifically the Rhabdocoela.

Results

Genome size estimation through survey analysis

A 350 bp library was constructed from the genomic DNA extracted from *L. nematophagus*. This library was then sequenced on the Illumina platform using 150 bp paired-end reads (PE 150). The paired-end sequencing data were evaluated based on GC distribution statistics and quality values (Q20, Q30), and were subsequently filtered to yield high-quality data (clean reads). These clean reads were employed for various tasks, including genome size estimation, genome assembly, GC content analysis, and heterozygosity rate determination. A total of 72.28 Gb of high-quality data was obtained, achieving an overall sequencing depth of approximately 89.71×, with Q20 and Q30 ratios of 96.97% and 91.96%, respectively. The GC content was calculated to be approximately 40.03% as shown in Table S1-1.

In the K-mer analysis, the average K-mer depth, indicative of the peak depth, was calculated to be 65.31 (Fig. 2A). The total count of K-mers derived from sequencing was 57,379,410,366. After the exclusion of K-mers with abnormal depths, 52,618,537,557 K-mers were used to estimate the genome length, which was determined to be approximately 805.66 Mb. Based on the K-mer distribution, the estimated repeat sequence content was approximately 67.59%, the estimated heterozygosity rate was around 0.07%.

Genome assembly

A PacBio library was constructed from genomic DNA and sequenced, yielding approximately 53.16 Gb of clean data. This amount of data corresponds to approximately 65X coverage, as illustrated in Fig. 2B. Figures 2C-D, and 2E depict the distribution of GC content and contig coverage depth, respectively. The assembly of reads into contigs resulted in a mean coverage depth of 57.91X, with an N50 length of 19.69 Kb and an average read length of 18.2 Kb, as detailed in Table S1-2. Additionally, read length distribution statistics are provided in Table S1-3. Initially, the total length of the genome contig sequence was 947.06 Mb, with a contig N50 of 5.11 Mb following

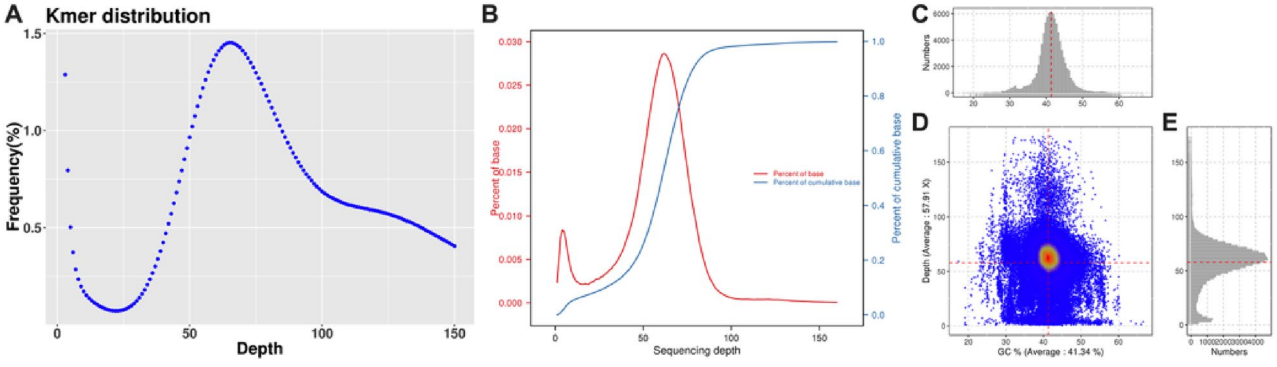

Fig. 2 The results of Survey, Hifi sequencing depth, and GC content. **A** The k-mer distribution graph for k = 21. **B** Sequencing depth of PacBio reads. The horizontal axis represents sequencing depth, and the vertical axis represents the proportion of base pairs at that sequencing depth to the total number of base pairs. **C** GC content distribution graph. **D** The scatterplot based on GC distribution and depth of coverage information of contigs, where colour shades are used to reflect the density of points in the scatterplot. The horizontal axis represents the GC content, and the vertical axis represents the coverage depth. **E** Distribution of contig coverage depth

assembly with Hifiasm. Sequences potentially representing contamination, such as plastid or mitochondrial DNA, were identified and excluded through BLAST searches against relevant databases. Subsequent to this decontamination process, the total genomic contig length was revised to 899.89 Mb, with a contig N50 of 5.49 Mb, as reported in Table 1.

Table 1 also indicates that out of 458 core eukaryotic genes (CEGs), 434 were successfully assembled using CEGMA, achieving a completion rate of 94.76%. Of the 248 highly conserved CEGs analyzed, 237 were successfully assembled, reflecting a completion rate of 95.56%. The BUSCO v5.4.7 assessment, employing the infecta_odb10-m genome, indicated a genome integrity of 85.49%. This assessment detailed the following proportions: complete single-copy core genes at 21.96%, complete multi-copy core genes at 63.53%, fragmented core genes at 7.06%, and deletions of core genes at 7.45%. The total number of genes in the core gene set was 255. To assess the completeness of genome assembly and the uniformity of sequencing coverage, the alignment of Illumina sequencing data with the *L. nematophagus* genome achieved a second-generation alignment rate of 98.67%, with an average depth of 74 and coverage rates of 99.82% for 1X and 95.18% for 20X. Similarly, the alignment of PacBio HiFi sequencing data with the *L. nematophagus* genome yielded a third-generation alignment rate of 98.86%, with an average depth of 60 and coverage rates of 99.99% for 1X and 92.35% for 20X. The results indicated that the genome assembly was relatively complete, with high integrity and sequencing uniformity.

Genome annotation and assessment

As indicated in Table S1-4, the Hi-C library underwent sequencing on the Illumina high-throughput sequencing platform, leading to the generation of 138.37 Gb of clean data. The GC content of this library was measured at 42.72%, with an N base proportion of 0.02%. Quality scores (Q-scores) for Q20 and Q30 reached 97.42% and 93.45%, respectively. Analysis of the Hi-C library, conducted using HiC-Pro (v2.10.0), yielded 340,056,553 unique read pairs. Of these, 221,972,677 pairs, or 65.28%, constituted valid Hi-C data that aligned with the genome. The categorization of read types is detailed in Table S1-5. Employing Hi-C-assisted genome assembly, a total of 895.47 Mb was mapped onto four chromosomes. This mapping was accomplished following Hi-C assembly adjustments and manual heatmap refinements, constituting 99.51% of the final assembled genome size, which was 899.92 Mb. From these chromosome-mapped sequences,

Table 1 Genome assembly and annotation statistics of *L. nematophagus*

Genome assembly		Gene annotation	
Accession Number	**JBRFHS000000000**	Number of genes	45,225
Total length of sequencing reads (Gb)	53.16 Gb	Protein-coding genes	40,737
Average sequencing depth	65.0 ×	Mean gene length (bp)	4139.21
Average length of sequencing reads (Kb)	18.2	Exons length (bp)	84,554,519
Number of scaffolds	401	Mean exon length (bp)	1869.64
Total length of scaffolds (Mb)	899.92	Exons number	262,424
N50 scaffold length (Mb)	231.83	Mean exon number per gene	5.8
N90 scaffold length (Mb)	165.63	CDS length (bp)	78,459,823
The longest scaffold length (Mb)	243.83	Mean CDS length (bp)	1734.88
Number of contigs	677	CDS number	259,415
Total length of contigs (Mb)	899.89	Mean CDS number per gene	5.74
N50 contig length (Mb)	5.49	Introns length (bp)	102,641,253
N90 contig length (Mb)	0.995	Mean intron length (bp)	2269.57
The longest contig length (Mb)	23.88	Introns number	217,199
Number of gaps	276	Mean intron number per gene	4.8
Total length of gaps (Kb)	27.6	BUSCO completeness (%)	92.16
Assembly size (Mb)	**899.92**	Transposable elements (%)	33.56
GC (%)	41.34	Tandem repeats (%)	16.52
Gaps (%)	0	Repetitive elements Size (%)	50.08
Number of assembled chromosomes	4	Number of ncRNAs	6740
Chromosome-level contig N50 (Mb)	5.49	Number of rRNA	770
Chromosome-level scaffold N50 (Mb)	231.83	Number of miRNA	0
CEGMA	94.76%	Number of snRNA	0
BUSCO completeness (%)	85.49%	Number of tRNA	5970
BioProject	PRJNA1308648	Number of pseudogene	647
BioSample	SAMN50709787	Number of motifs	1904
		Number of domains	57,069

856.23 Mb, or 95.62% of the anchored sequence, were successfully ordered and oriented as shown in Table S1-6. The final genome assembly was achieved after implementing Hi-C error correction, assembling accessory chromosomes, and eliminating redundancies. This genome assembly comprises 899.92 Mb with a contig N50 of 5.49 Mb and a scaffold N50 of 231.83 Mb, as recorded in Table 1.

To evaluate the Hi-C assembly results, the assembled genome was divided into bins of 300,000 bp each. The number of Hi-C read pairs between each pair of bins was quantified, serving as an indicator of interaction intensity. As depicted in Fig. 3A, four distinct chromosome groups were identified. Within each group, interaction intensity is markedly higher along the diagonal, indicating robust interactions between adjacent sequences (diagonal), in contrast to weaker interactions across non-adjacent sequences (off-diagonal regions). Moreover, the Circlize software facilitated the creation of circular plots that illustrate gene density, repeat sequence density, GC content, and collinearity, as shown in Fig. 3B.

Based on Hi-C clustering and subgenome phasing analysis, we determined that *L. nematophagus* is an allotetraploid with a karyotype of 2n = 4x = 8. The four chromosomes were assigned to two subgenomes: A01 and A02 belong to subgenome A, while B01 and B02 belong to subgenome B, with homoeologous relationships between A01/B01 and A02/B02.

To assess the completeness of the chromosome-level assembly, we searched for telomeric repeats. The canonical telomeric repeat motif AACCCT was identified at both ends of chromosome B01, specifically at positions 1–10,000 bp and 214,965,325–214,975,325 bp. This finding suggests that chromosome B01 approaches telomere-to-telomere (T2T) completeness. However, the current genome assembly still contains 276 gaps totaling 27.6 Kb (Table 1), and telomeric repeats were not detected on other chromosomes (A01, A02, B02). Therefore, while this assembly represents a high-quality chromosome-level reference, it has not yet achieved a full T2T status. The identification of telomeric repeats on B01 provides a foundation for future gap-closing efforts to obtain a complete T2T genome.

The genome annotation encompasses repetitive sequences, coding genes, and functional annotations, including databases such as NR, eggNOG, GO, KEGG,

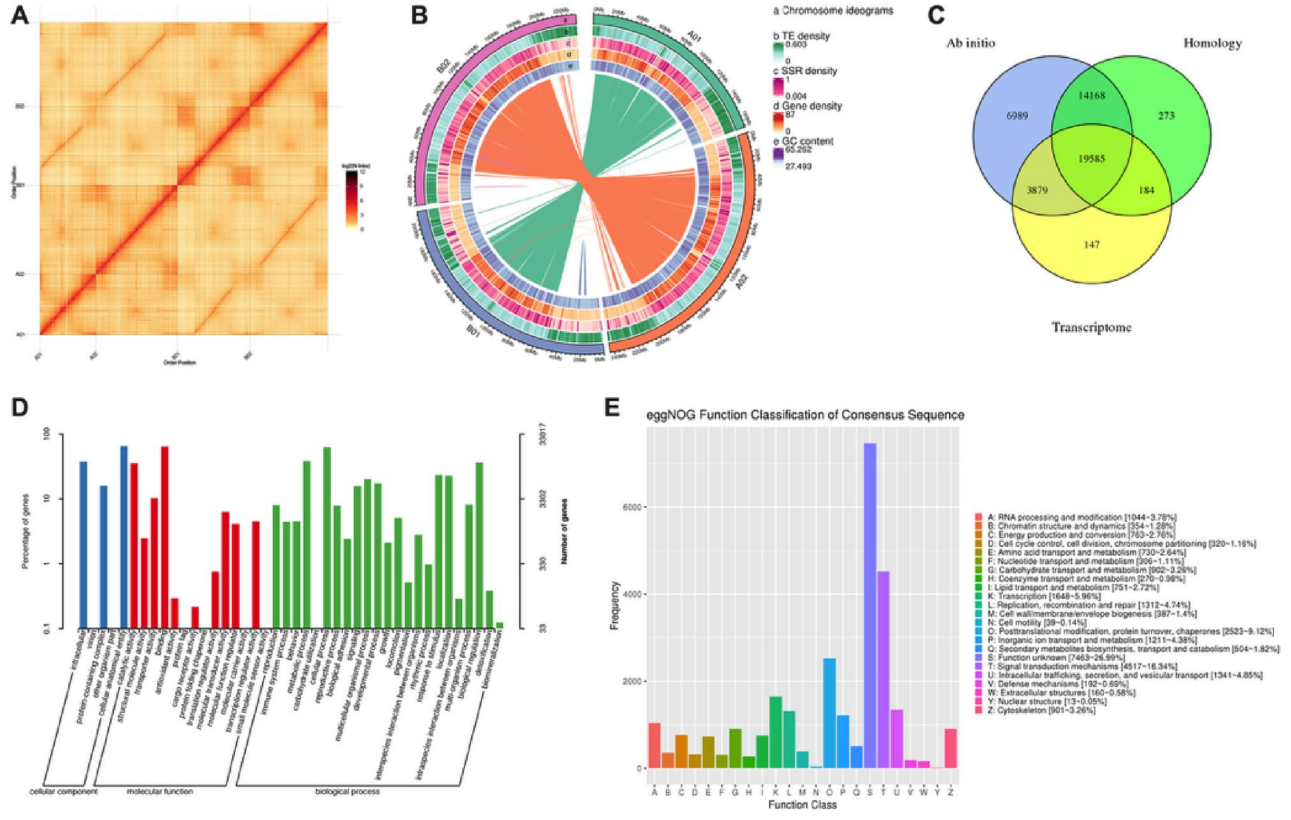

Fig. 3 Overview of the assembly and annotation of the *L. nematophagus* genome. **A** Hi-C assembly chromosome interaction heatmap. 4 distinct chromosome groups can be clearly distinguished from each other. **B** Genomic circle diagrams used to display gene density, repeat sequence density, GC content, and collinearity situations. **C** The distribution graph of genes from Ab initio, Homology, and transcriptome after integration with the EVM software. **D** Classification statistics of genes classified and annotated by the GO database. **E** Classification statistics after annotation of gene functions in the eggNOG database

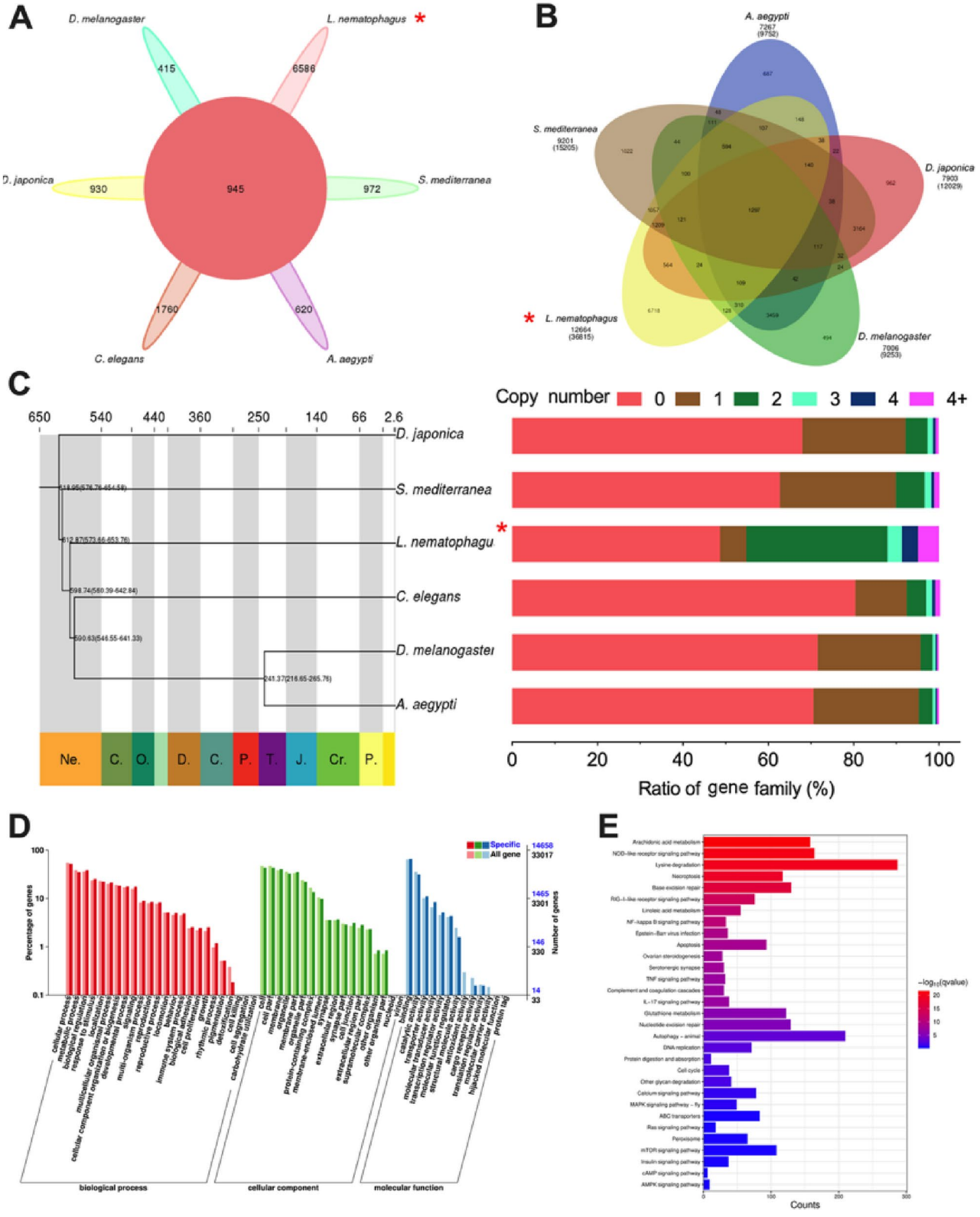

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Gene family clustering and phylogenetic analysis. **A** The clustered gene family diagrams of the six species involved in comparative genomics analysis, namely *D. melanogaster*, *D. japonica*, *A. aegypti*, *S. mediterranea*, *C. elegans*, and *L. nematophagus*. The middle circle represents the number of gene families common to all species, and the sides represent the number of gene families specific to each species. Note: The sizes of the petals in this diagram are determined by the visualization software's layout algorithm and do not represent the relative proportions of gene family counts. Please refer to the numerical labels and Table S2-2 for accurate gene family statistics. **B** Venn diagram of gene family clustering. The number below the species name is the total number of gene families, with the corresponding number of genes in parentheses; the number in the Venn diagram is the number of gene families. *C. elegans* is not included in this diagram due to the visualization software's limitation of displaying a maximum of 5 species. Please refer to Supplementary Figure S2 for a visualization including all six species. **C** The phylogenetic tree with divergence times and a plot of the copy number distribution of each species across all gene families. The times on the evolutionary tree on the left indicate the time of divergence supported by the 95% highest posterior density (HPD). At the bottom of the tree is geological time, and at the top of the tree are absolute ages in millions of years, shaded to define each geological period. Common geological periods include: Neoproterozoic, Cambrian, Ordovician, Silurian, Devonian, Carboniferous, Permian, Triassic, Triassic, Jurassic, Cretaceous, Paleogene and Neogene. The graph on the right shows the number of gene families containing 0 copies, 1 copy, 2 copies, 3 copies, 4 copies, and more than 4 copies per species. **D** GO enrichment analysis of unique gene families in *L. nematophagus*. The horizontal axis represents the annotation information of the GO term, the left vertical axis shows the proportion of unique genes under this GO term among all unique genes, and the right side shows the number of unique genes under this GO term. **E** KEGG enrichment analysis of the unique gene family in *L. nematophagus*. The vertical axis represents the KEGG pathway annotation information, the horizontal axis represents the number of genes under that KEGG pathway, and the color represents the -log10 of the corresponding q-value

TrEMBL, Swiss-Prot, KOG, Pfam, and InterPro. It also includes pseudogenes and annotations for non-coding RNAs. TEs within the repetitive sequences constitute approximately 301,977,592 base pairs, or 33.56% of the genome, as noted in Table S1-7. Additionally, tandem repeat sequences account for 148,700,161 base pairs, or 16.52% of the genome, as listed in Table S1-8. According to Table S1-9, we identified 45,225 coding genes through a combination of homology-based, de novo, and transcriptome-based predictions. The integration of these genes, facilitated by EVM and supported by the three aforementioned methods, is detailed in Fig. 3C. This figure confirms that most genes are validated by at least two methods, underscoring the high accuracy of the predictions. The prediction accuracy is further evaluated in Table S1-10, which compares genetic characteristics such as gene length, exon length and count, intron length and count, and coding sequence (CDS) length with those of selected species. The completeness of gene prediction was assessed using the BUSCO (v5.2.2) tool, which identified 235 conserved core genes within the predicted set, representing 92.16% of BUSCO genes and indicating a high level of completeness. Non-coding gene prediction, employing multiple software tools, resulted in the identification of 5,970 tRNAs, 770 rRNAs, and no detected snoRNAs or miRNAs, as shown in Table S1-11. Subsequent software comparisons and corrections led to the prediction of 647 pseudogenes, documented in Table S1-12. The predicted gene sequences were annotated using the aforementioned databases, with 90.08% of genes successfully annotated, as detailed in Table S1-13. Specifically, the GO database annotated 33,017 genes, accounting for 73.01% of all coding genes. This analysis allows for the classification of genes into three main categories: cellular component, molecular function, and biological process, as depicted in Fig. 3D. The KEGG database annotated 32,198 genes, covering 71.20% of the genome, while the KOG database annotated 24,697 genes, representing 54.61%. The Pfam database annotated 35,089 genes, accounting for 77.59% of the total. The SwissProt database annotated 22,276 genes (49.26%), and the TrEMBL database annotated 39,015 genes (86.27%). The eggNOG database annotated 28,542 genes, representing 63.11% of the coding genes, with 26.99% (7,643 genes) encoding proteins of unknown function and 16.34% (4,517 genes) encoding proteins with predicted signaling mechanisms, as illustrated in Fig. 3E. Furthermore, analysis revealed 1,904 modal motifs and 57,069 structural domains, as listed in Table S1-14. The genome assembly has been deposited in the NCBI under accession number JBRFHS000000000, and the annotation data have been uploaded to the Figshare platform (https://doi.org/10.6084/m9.figshare.30000730.v1).

Gene family identification, phylogenetic, and gene family expansion analysis

To investigate the genomic basis of the unique predatory and adaptive traits of *L. nematophagus*, we performed gene family clustering and comparative genomic analyses with selected model species representing key evolutionary lineages.

Our analysis identified 24,690 gene families. Table S2-2 provides the statistical details of gene family clustering for each species. Figures 4A-B indicate that a total of 945 gene families were common across all species. *L. nematophagus* displayed 12,664 gene families, which represent 51.30% of the total. This species also possessed 6,585 species-specific orthogroups comprising 22,436 genes, accounting for 49.60% of the total gene count. These figures are relatively higher compared to those of the other species analyzed. It is important to note that the identification of species-specific orthogroups can be sensitive to the quality of genome assemblies and annotations, as well as the parameters used for orthology inference. Therefore, these numbers should be interpreted as estimates that provide a foundation for further exploration

rather than absolute counts. As the visualisation software can display a maximum of five species, *C. elegans* is not included in Fig. 4B. For a visualisation incorporating all six species, please refer to Supplementary Figure S2 and Table S2-2.

In *L. nematophagus*, the distribution of gene family copy numbers was characterized by a notably higher proportion of families with 2 copies (33%) and more than 4 copies (4.9%) compared to other species. The distribution was as follows: 48.7% had 0 copies, 6.2% had 1 copy, 33% had 2 copies, 3.4% had 3 copies, 3.8% had 4 copies, and 4.9% had more than 4 copies. The gene counts for families with 1, 2, 3, 4, and more than 4 copies were 1,525, 16,312, 2,550, 3,716, and 12,712, respectively, as shown in Fig. 4C. GO and KEGG enrichment analyses were conducted for species-specific gene families as depicted in Figs. 4D-E. In the GO biological process category, 295 genes were associated with DNA-mediated transposition, 225 with the stress-activated protein kinase signaling cascade, and 116 with inositol phosphate ceramide metabolism, as detailed in Table S3-1. The enrichment analysis highlighted distinct gene functions related to lipid metabolism, including activities of lipases and phospholipases, as well as those of serine and collagen. Additionally, genes related to antioxidant activity, oxygen response, digestive system development, detoxification, drug response, reproduction, and embryonic development were identified, as shown in Figs. S3-7. For cellular components, 251 genes were linked to the membrane side, 272 to the plasma membrane protein complex, 163 to the cytoplasmic side of the membrane, and 177 to the receptor complex, as outlined in Table S3-2. The enrichment analysis underscored that the functions of specific gene families were associated with transmembrane transport, signal transduction, antioxidant activity, detoxification, and protein degradation, as illustrated in Figs. S8-9. In the molecular function category, 266 genes were related to transposase activity, 254 to carbohydrate binding, and 178 and 120 to monosaccharide binding and mannosyltransferase activity, respectively, as listed in Table S3-3. The enrichment analysis demonstrated that specific gene families were enriched in activities such as phospholipase, protease and peptidase, transmembrane transport, signal transduction, and antioxidant and oxidoreductase activities, as shown in Fig. S10-12. In the KEGG analysis, 287 genes were enriched in the lysine degradation pathway, 158 in arachidonic acid metabolism, 164 in the NOD-like receptor signaling pathway, and 117 and 130 in necrotic apoptosis and base excision repair, respectively, as recorded in Table S3-4. KEGG pathway enrichment analysis, depicted in Fig. S13, revealed that specific gene families were involved in energy substance metabolism, protein digestion and absorption, oxidative stress, and various signaling pathways.

The phylogenetic tree depicted in Fig. 4C was constructed using 1,373 single-copy gene sequences that were compared and concatenated. This tree shows strong bootstrap support across all nodes, with values of 100%, indicating high confidence in the inferred topology. The evolutionary timeline suggests that *D. japonica*, *S. mediterranea*, *L. nematophagus*, and *C. elegans* diverged during the Neoproterozoic era, approximately 650 to 540 million years ago. *L. nematophagus* diverged later than *D. japonica* and *S. mediterranea*, around 599 million years ago, but earlier than *C. elegans*, which diverged 591 million years ago, and significantly earlier than *D. melanogaster* and *A. aegypti*, both of which diverged around 241 million years ago.

The evolutionary tree of the species, which illustrates the contraction and expansion of gene families, is presented in Fig. 5A. An analysis of gene family dynamics in *L. nematophagus* throughout its evolution revealed 51 gene family expansions and one contraction, specifically in the mannose phosphatidylinositol ceramide synthetase gene, as detailed in Table S2-3. These changes in gene families were analyzed using GO and KEGG enrichment analyses through Cluster Profile, as shown in Figs. 5B-C. The GO analysis identified expansions in biological processes (Table S3-5) including 42 genes linked to cellular drug response, 24 to ammonium ion response, and 15 to membrane protein hydrolysis. The functions of these expanded gene families were enriched in areas such as lipid metabolism, neural signaling, reproduction, water homeostasis, detoxification, behavior, antioxidant activity, response to oxygen, and response to various stimuli, as detailed in Figs. S14-17. In cellular components (Table S3-6), expansions included 29 genes related to synaptic vesicles, 15 to voltage-gated potassium channel complexes, and 11 to both essential and intrinsic synaptic vesicle membrane components. The functions of these expanded gene families, as depicted in Fig. S18, were enriched in reproduction, embryonic development, transmembrane transport, and signal transduction. Molecular function analysis (Table S3-7) revealed expansions of 83 genes in metalloendopeptidase activity, 15 in metallopeptidase activity, 35 in aspartate-type endopeptidase activity, and 18 in carboxyhydrolase activity. These expanded gene families were enriched in functions including antioxidant activity, response to oxygen, phospholipase activity, lipid metabolism, peptidase activity, transmembrane transport, signal transduction, detoxification, and drug response, as shown in Figs. S19-20. The KEGG analysis (Table S3-8 and Fig. S21) identified significant expansions of 17 genes within the pentose-glucuronide interconversion pathway, 46 in lysine degradation, and 29 in both the Notch signaling and nucleotide excision repair pathways, as well as in the cascade signal transduction pathway. These pathways are crucial

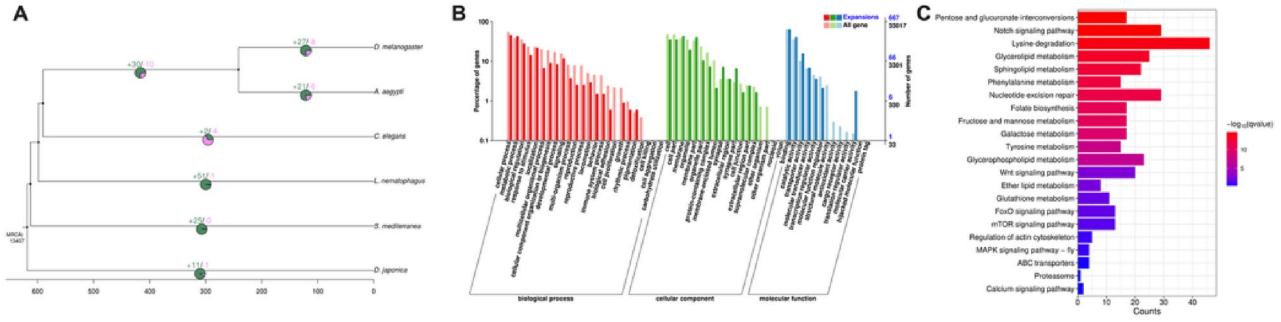

Fig. 5 Analysis of gene family expansion and contraction. **A** The species evolutionary tree with contraction and expansion of gene families drawn using CAFE v4.2 software. The '+' represents the number of expanding gene families at the node, and the '-' represents the number of contracting gene families at the node. The pie charts represent the proportion of contracting and expanding gene families in the corresponding branches. The size of circles in this figure is determined by the visualisation software's layout algorithm and does not represent the number of expanded/contracted gene families. For the precise number of expanded/contracted gene families across species, please refer to the figures and Table S2-3 (**B**) GO enrichment analysis of expanding gene families in *L. nematophagus*. The horizontal axis represents the annotation information of the GO term, the left vertical axis shows the proportion of expanding genes under this GO term among all expanding genes, and the right side shows the number of expanding genes under this GO term. **C** KEGG enrichment analysis of the expanding gene family in *L. nematophagus*. The vertical axis represents the KEGG pathway annotation information, the horizontal axis represents the number of genes under that KEGG pathway, and the color represents the -log10 of the corresponding q-value

for embryonic development, maintenance of adult stem cells, tissue homeostasis, and repair. As listed in Table S2-4, only one gene, XspB01G048080, was significantly positively selected in *L. nematophagus*.

Discussion

The genome of *L. nematophagus*, encompassing 899.92 Mb with a Scaffold N50 of 231.83 Mb, has been assembled at the chromosome level for the first time. This species, a newly identified genus and species of Rhabdocoela, was discovered in soil affected by root-knot nematode disease in Anshun, Guizhou Province [6]. It shows significant potential for the biological control of plant-parasitic nematode diseases. Our study underscores several distinctive features that set this species apart from traditional model planarian species:

(1) Unlike model planarians that rely on stem cells for regeneration, *L. nematophagus* lacks regenerative capabilities. Instead, it exhibits effective reproduction through self-fertilization (Figs. S1 A-D), contrasting sharply with the regenerative mechanisms of Tricladida and Macrostomida planarians such as *D. japonica*, *S. mediterranea*, and *M. lignano*. (2) Adaptation to soil ecosystems in *L. nematophagus* is marked by the expansion of gene families related to transmembrane transport, antioxidant activity, detoxification, and oxygen response, which may reflect its evolutionary adjustments to terrestrial habitats. (3) Predatory behavior towards plant-parasitic nematodes is another key trait (Fig. 1A); preliminary pot experiments have confirmed its efficacy in managing tomato root-knot nematode disease, as illustrated in Figs. S1 E-I. (4) The species also demonstrates the capability for large-scale laboratory cultivation. For instance, under our standard rearing conditions (e.g., 100 transparent plastic culture tanks equipped with aerator pumps (80 L/min flow rate, 0.030 MPa pressure), 60 cm in length, 40 cm in width, and 15 cm in height, under dark conditions at 25 °C), it can produce up to 200 million eggs annually. This high fecundity addresses a significant challenge in producing sufficient biological material for nematode control.

The potential of *L. nematophagus* for controlling root-knot nematode disease is further underscored by the sequencing and analysis of its chromosome-level genome in this study. This research provides the first chromosome-level reference genome of *L. nematophagus*, with a genome size of 899.92 Mb and a Scaffold N50 of 231.83 Mb. This size is intermediate between the genomes of *D. japonica* (1.13 Gb) and *S. mediterranea* (840.2 Mb) as reported in the existing literature [23, 25]. Earlier genome assemblies for planarians, such as the first draft of *D. japonica* (Scaffold N50 = 27.7 kb) [24] and the initial *S. mediterranea* assembly (dd_Smed_g4, Scaffold N50 = 3.85 Mb) [15], were highly fragmented due to the limitations of short-read sequencing. However, recent advances in long-read sequencing and Hi-C technologies have led to substantial improvements. For instance, the latest *S. mediterranea* assembly has achieved chromosome-scale (schMedS3h1, N50 = 270.2 Mb) [25]. While the genome of *D. japonica* has been significantly improved to a highly contiguous, scaffold-level assembly (N50 = 652.52 kb) [23]. The chromosome-level assembly of *L. nematophagus* (Scaffold N50 = 231.83 Mb) achieves contiguity comparable to that of the latest high-quality references. The enhanced continuity achieved in this assembly is attributed to the utilization of an integrated sequencing strategy combining Hi-C with multi-platform sequencing (Illumina and PacBio HiFi), which effectively mitigates assembly fragmentation commonly exacerbated by a high proportion of repetitive sequences (50.08%).

In contrast, *S. mediterranea*, despite employing PacBio combined with Chicago technology, faces challenges due to its higher repeat content (61.7%) and the presence of large Gypsy elements from the Burro family [15]. These findings highlight the necessity of integrating long-read sequencing technologies with three-dimensional genome mapping to effectively resolve complex, highly repetitive genomes.

In addition to high contiguity, the repetitive sequence content in *L. nematophagus* constitutes 50.08% of its genome, positioning it intermediately among the published genomes of flatworms. This percentage surpasses the values reported for *D. japonica* [24] and certain Macrostomum species [29], yet it remains lower than those documented for *S. mediterranea* [15], *S. polychroa* [30], and another assembly of *D. japonica* [23]. Similarly, the TE content in *L. nematophagus*, at 33.56%, is less than that observed in *S. mediterranea*, *S. polychroa*, *S. nova*, and *S. lugubris* [25]. This substantial genomic repetitiveness, primarily attributed to TEs, appears to have facilitated gene family innovation in *L. nematophagus*, including notable expansions in pathways related to neuropeptide signaling, detoxification, and predation. In contrast, the elevated rate of BUSCO duplication observed in *M. lignano* is linked to a recent whole-genome duplication rather than to TE proliferation, underscoring potential differences in the mechanisms that drive repeat content expansion and their functional implications across different planarian lineages [29]. The repetitive genomic landscape has likely supported the noted gene family innovations, particularly the expansions in metalloendopeptidases (83 genes) and synaptic vesicle components (29 genes). These expansions may play important roles in the degradation of nematode cuticle and could contribute to the enhancement of prey detection capabilities.

L. nematophagus exhibits the highest gene count (45,225) among the planarians reported to date, marked by significant enrichment in three principal functional modules. Firstly, it possesses an abundance of predator-specific enzymes such as metalloendopeptidase, phospholipase, and lipase. These enzymes likely facilitate the degradation of nematode cuticles, exemplified by those of *M. incognita*. Secondly, there is a notable expansion of gene families associated with detoxification, environmental response, water homeostasis, and ABC transporters, which may enhances its capacity to adapt to diverse soil environments. Thirdly, the organism displays a significant expansion in gene families involved in neural signal transduction, including those related to lysine degradation and the Notch signaling pathway. This expansion is distinct from the regeneration-associated pathways, such as Wnt and FGFR, observed in *S. mediterranea* [25]. The focus on gene families associated with predator specialization and environmental adaptation in *L. nematophagus* is functionally analogous to the patterns observed in Tricladida (*S. mediterranea* and *D. japonica*); however, the specific pathways expanded differ notably, such as lysine degradation and neural signaling in *L. nematophagus* versus Wnt signaling and stem cell maintenance in Tricladida [31]. This divergence likely stems from the prolonged adaptation of the Rhabdocoela lineage to predatory habitats. The genome of *L. nematophagus* also shows a significant expansion in gene families related to neural signaling, response to environmental stimuli, and drug response, suggesting a complex gene regulatory network potentially essential for predation behavior and environmental perception in soil. In contrast, the genome of *M. lignano* primarily contains simple tandem repeats (approximately 75%) and exhibits low levels of CpG methylation (approximately 2.5%) [26], indicating that epigenetic regulation may play a role in sustaining the high proliferative activity of neoblasts in this species. Furthermore, while *S. mediterranea* lacks essential metabolic genes such as FASN and MAD1/2 [15], other species like *D. japonica* [24] and *S. polychroa* [30] have retained these genes, suggesting that *L. nematophagus* may possess complete metabolic pathways conducive to thriving in predatory environments. Nevertheless, additional genetic functional analyses are required to confirm this hypothesis. Unlike typical model planarians, although the KEGG enrichment analysis of *L. nematophagus* revealed a significant expansion in genes associated with stem cell homeostasis pathways, including Notch and Wnt, these organisms have not demonstrated tissue regeneration capabilities.

Genomic analysis has disclosed a substantial expansion in gene families pertinent to neural signaling, stress response, and peptidase activity in *L. nematophagus*. Notably, the count of metalloendopeptidase genes has risen to 83. This augmentation may provide a molecular basis for the distinct ecological characteristics of the species, including its affinity for soil environments and its predatory behavior towards nematodes [6]. For an exhaustive overview of the whole-genome alignment data comparing *L. nematophagus* with other planarian model species, refer to Table 2, which includes direct links to the NCBI database and publication DOIs for each reference genome. Detailed whole-genome information for species within the class Rhabditophora, as cataloged in the NCBI database, can be found in Table S4.

Further comparative genomic analyses have demonstrated significant enlargements in gene families related to neural signaling, such as those associated with synaptic vesicles, responses to environmental stimuli including antioxidant and oxygen-responsive genes, and transposase activity in *L. nematophagus*. These genes show functional enrichment in areas such as

Table 2 List of whole genome sequence information of *L. nematophagus* and model planarians

Order	Rhabdocoela	Tricladida		Macrostomida
Species	***L. nematophagus***	*S. mediterranea*	*D. japonica*	*M. lignano*
Genome assembly	JBRFHS000000000	GCA_045838265.1		GCA_002269645.1
Genome size (Mb)	**899.92**	840.2	1,128.2	764.4
Number of scaffolds	401	662	10,017	5269
Scaffold N50	231.83 Mb	270.2 Mb	652.52 kb	246.2 kb
Number of contigs	677	833	10,017	5,974
Contig N50	5.49 Mb	9.6 Mb	248.44 kb	215.3 kb
GC percent	41.34	29.5	24.1	46
Genome coverage	65.0 ×	44.0 ×	79.0 ×	180.0 ×
Assembly level	Chromosome	Chromosome	Scaffold	Scaffold
Reference DOI		https://doi.org/10.1038/s41467-024-52380-9	https://doi.org/10.1016/j.ygeno.2022.110293	
NCBI Database Link		https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_045838265.1/		https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_002269645.1/

The bold texts indicate the chromosome-level genome size of *Luticola nematophagus*. DOIs refer to the primary publications describing each genome assembly. NCBI database links provide direct access to the deposited genome assemblies

transmembrane transport, lipid metabolism, exemplified by lipase activity, and proteolysis, specifically concerning metalloendopeptidase activity. This enrichment indicates a potential role of these genes in the sophisticated predation strategy of *L. nematophagus*, which involves the recognition, capture, and digestion of nematodes. The evolution of these flatworms from a model focused on regeneration to one emphasizing predatory capabilities has been clarified. The analysis of gene family expansions, including those involved in the Notch signaling pathway and the nucleotide excision repair pathway as indexed in the KEGG database, reveals that *L. nematophagus* possesses specific adaptive traits for maintaining tissue homeostasis and responding to environmental stresses. These genetic adaptations provide a foundation for the species' effective predation on nematodes within soil ecosystems.

Over a century ago, researchers established that the majority of Rhabdocoela species have lost their capacity for tissue regeneration and repair [27, 28], a characteristic also observed in *L. nematophagus*. Unlike *D. japonica*, *L. nematophagus* not only lacks regenerative capabilities but also exhibits self-fertilization, which leads to a reallocation of resources toward this exclusive reproductive strategy. This adaptation is likely connected to its evolutionary progression, enhancing population growth to flourish in complex soil environments. Genomic analysis reveals a marked expansion in gene families associated with reproduction and embryonic development (Figs S15 and 18). Niels V. S. et al. clarified the transition of species within the Typhloplanidae from marine origins to freshwater and terrestrial habitats [32]. *L. nematophagus*, a novel genus and species within the Typhloplanidae, resides in moist soil [6]. Essential factors enabling *C. elegans* and soil-dwelling arthropods to prosper in soil encompass the MAPK pathways, antioxidant systems, hypoxia tolerance, and osmoregulation-related genes [33–35]. Whole-genome analysis of *L. nematophagus* has uncovered gene expansions in areas such as detoxification, homeostasis, oxidative stress response, stimulus sensitivity, transmembrane transport, and signal transduction (Figs. S14-20). These gene expansions may contribute to *L. nematophagus*' adaptability to soil stresses, including desiccation, hypoxia, and exposure to external substances. The identification of expanded synaptic vesicle genes (29 genes, Figs. S14-17) and the significant enrichment of neural signal transduction-related gene families suggest a potential role in sensory perception and motor coordination, which could be relevant for navigating complex soil environments and detecting prey. However, as synaptic vesicle components are fundamental to neuronal function across all animals, their expansion may also reflect broader neural requirements unrelated to predation. *L. nematophagus* consumes nematodes entirely using its muscular pharynx [6], a process that requires the degradation of the complex nematode cuticle, which consists of lipids, proteins (including collagen), glycoproteins, and chitin [36–38]. Interestingly, nematode-trapping fungi such as *Dactylella dianchiensis*, *Arthrobotrys oligospora*, and *Monacrosporium microcaphoides* employ a convergent strategy, secreting an array of enzymes (e.g., proteases, collagenases, phospholipases) to penetrate this barrier [39–41]. Our genomic analysis not only aligns with this convergent strategy for breaching the nematode cuticle but also identifies specific gene families associated with phospholipase activity and

reveals significant expansions in gene families involved in peptidase (particularly metalloendopeptidase) activity and lipid metabolism. The particular and expanded genes encoding enzymes that degrade crucial components of the nematode cuticle (e.g., lipases/phospholipases for lipids, peptidases for proteins/collagen) suggest a potentially vital role for these enzymes in the predation process of *L. nematophagus*. Additionally, expansions in neural signal transduction genes may support the sensory detection of nematode prey and the precise motor coordination necessary for efficient engulfment. Taken together, these genetic characteristics are strongly implicated in the recognition, ingestion, and particularly the digestive breakdown of nematodes by *L. nematophagus*.

Planarians display predatory behaviors towards small invertebrates, which coexist with their extraordinary capabilities for tissue repair and regeneration. The research conducted by Robert E. Ogren [18] at Wilkes University in 1995 delineates the predatory strategies of planarians. These strategies include hunting, attacking, and capturing prey through mechanisms such as mucus adherence, pharyngeal expansion, toxic secretion, and physical constriction, followed by the secretion of digestive fluids to facilitate feeding. Similarly, *L. nematophagus* targets root-knot nematodes, progressively engulfing and digesting the entire organism via the pharynx, yet refrains from absorbing the nematode's internal fluid due to its diminutive size (Fig. 1A). This mode of predation bears resemblance to that of the aquatic Rhabdocoela flatworm, *Mesostoma ehrenbergii*, which consumes small invertebrates like *Daphnia magna* and *A. aegypti* using discernibly different methods [42]. Predatory behavior in flatworms typically manifests as a combination of active predation and scavenging, attraction to chemical cues, and neurotoxin production [18]. Similar to how *Giardia anceps* has been observed to regulate mosquito populations [22], *L. nematophagus* demonstrates potential in controlling infestations of root-knot nematodes. However, various flatworm species exhibit diverse tactics in hunting and consuming prey according to the type encountered [43]. Despite these observations, there remains a paucity of comprehensive research exploring the complex interactions between flatworms and their prey.

L. nematophagus exhibits remarkable convergent evolution in the central module of predator digestion and detoxification. The expansion of lipase and peptidase gene families in *L. nematophagus* parallels that observed in *Tenodera sinensis* [44], promoting the efficient degradation of their high-protein prey, nematodes for *L. nematophagus* and insects for *T. sinensis*. This similarity suggests that convergent evolution may be driven by the demands of predatory diets [44]. Additionally, *L. nematophagus* shows an expansion in the lipase gene family and enhanced metabolism of arachidonic acid, which likely contributes to osmoregulation in soil humus. In terms of detoxification and signal transduction, there is a notable expansion in the ABC transporters, glutathione metabolism, and genes associated with oxidative stress response. These expansions are comparable to those in the P450/GST gene family observed in predatory mites [45, 46] and *Arma custos* [47], potentially reflecting a coordinated adaptation to environmental and dietary toxins. However, the mechanisms of expansion vary; for example, *A. custos* has significantly expanded the P450/GST gene family to counteract toxins from a diverse range of prey [47], whereas *L. nematophagus* primarily may enhance antioxidant mechanisms and transmembrane transport to mitigate xenobiotic impacts in the soil. In the realm of neural signal perception and prey capture, *L. nematophagus* possesses an expanded array of genes related to neurosignal transduction and digestive enzymes, including metalloendopeptidases and those involved in lipid metabolism. These genes have functions analogous to venom proteins in *A. custos* [47], which facilitate the immobilization and digestion of prey. In contrast to the visually oriented *T. sinensis* [44], *L. nematophagus* relies on genes responsive to environmental stimuli, such as stress-activated MAPK, to adapt to sensory perception in dimly lit soil environments. The evolutionary mechanisms of the genome in *L. nematophagus* significantly diverge from those of insect predators. While *T. sinensis* utilizes recent LTR retroposon bursts to facilitate gene expansion (e.g., trypsin, ABC transporter) [44], *L. nematophagus* predominantly employs partial gene duplication or paleo-whole genome duplication. This results in a more compact genome, which is maintained within the stable soil niche. Notably, a substantial portion (49.6%) of the expanded gene families in *L. nematophagus* are species-specific, a proportion markedly higher than that seen in model insects. This distinction may be attributed to unique adaptations linked to its soil-dwelling lifestyle, autogenous breeding behavior, and specialized predation on nematodes. Niche differentiation profoundly influences the functional evolution of *L. nematophagus*. Unlike terrestrial natural enemies such as ladybugs and mantises, *L. nematophagus* exhibits expansions in sphingolipid metabolism, antioxidant defense mechanisms, and cuticle-degrading enzymes (proteases/lipases). These expansions may represent adaptations that help address the challenges posed by oxidative stress in soil, digestion of nematode cuticles, and osmotic pressure regulation in humus-rich environments. The specific prey type (nematodes) necessitates the enrichment of cuticle-degrading enzymes, while predators of aphids (e.g., *Aphidoletes aphidimyza*) extend their chemoreceptors to detect plant volatiles, supporting a prey-sensing coevolutionary model [48].

The distinctive genomic characteristics of *L. nematophagus* confer unique practical benefits, particularly through the presence of genes encoding metalloendopeptidase and phospholipase. This discovery offers a promising avenue for the development of environmentally benign agents for the management of plant-parasitic nematodes. Additionally, the capacity of *L. nematophagus* for self-reproduction facilitates its large-scale propagation, while its adaptability to soil environments enhances its persistence and population growth in these habitats. Consequently, these traits amplify the predatory effectiveness against plant-parasitic nematodes, thus enhancing the efficacy of strategies for nematode disease prevention and control.

Although our findings provide significant insights, it is important to acknowledge the limitations of this study. First, the functional roles of the expanded gene families we identified remain speculative. These include metalloendopeptidases, neural signaling components, detoxification genes, and positively selected genes (XspB01G048080). Their proposed involvement in nematode predation, soil adaptation, and loss of regeneration requires rigorous experimental validation using CRISPR/Cas9 or RNAi. Without such functional data, the precise molecular mechanisms underlying these key traits will remain elusive.

Second, our comparative genomic analysis was constrained by the limited availability of high-quality chromosome-level genomes. To date, no closely related Rhabdocoela species or a few other non-model predatory flatworms have been sequenced. This scarcity prevented us from conducting comprehensive synteny analyses or identifying deeply conserved regulatory elements. For example, it hindered our ability to systematically contrast regeneration pathways (e.g., Wnt/FGFR) with predation pathways or to fully resolve the genomic structural changes associated with the evolution of nematode-specific predation within Platyhelminthes.

Third, although we successfully annotated an abundance of tRNAs and rRNAs, we did not explore the functional roles of non-coding RNAs (ncRNAs). The apparent absence of snoRNAs and miRNAs in our annotations is noteworthy but requires further investigation. This represents a significant gap in our understanding of potential post-transcriptional regulatory mechanisms that may govern predation or stress responses in *L. nematophagus*.

Fourth, our assessment of soil adaptation was primarily based on genomic evidence. Currently, empirical data on the specific physiological and behavioral mechanisms that enable survival under variable natural soil conditions are lacking. These include responses to desiccation, extreme temperatures, and hypoxia. Furthermore, we have not yet evaluated the stability of its predatory efficacy across different soil types or under different agricultural management practices.

Finally, the ecological impact and safety profile of *L. nematophagus* as a biocontrol agent remain unclear. Critical issues require further investigation, including its potential effects on non-target soil fauna such as beneficial nematodes, microarthropods, and earthworms. Its long-term persistence in the field after release also remains unknown. These unresolved questions pose significant challenges for the practical application of this species.

Addressing these limitations will require targeted efforts in future research. Building upon this chromosome-level genome assembly, functional validation of expanded gene families—such as metalloendopeptidases, synaptic vesicle components, and detoxification enzymes—using CRISPR/Cas9 or RNAi should be prioritized to move from genomic correlations to mechanistic understanding of predation, soil adaptation, and reproductive traits. In parallel, expanding comparative genomic resources to include additional Rhabdocoela species and other predatory flatworms will enable robust synteny analyses and clarify the evolutionary trajectory of nematode-specific predation and regeneration loss. From an applied perspective, comprehensive field trials are needed to evaluate long-term biocontrol efficacy and environmental safety of *L. nematophagus* across diverse agricultural conditions, while optimization of artificial breeding protocols informed by reproduction-associated genes could enhance cost-effective mass rearing. Finally, exploring synthetic biology approaches—such as engineering rhizosphere microbes with key *L. nematophagus*-derived cuticle-degrading enzymes—may offer alternative self-sustaining strategies for nematode management within the plant root zone.

In conclusion, this study presents the first chromosome-level genome assembly for the Rhabdocoela flatworm, *L. nematophagus*, thereby addressing a critical gap in genomic resources for non-model flatworms. Our analysis deciphers the genetic architecture underlying its specialization as a soil-dwelling nematode predator, potentially driven by expansions in gene families related to neural signaling and environmental response. These findings establish a crucial molecular link between its unique adaptations—soil persistence, nematode predation efficiency, and high reproductive output through self-fertilization—and its promising potential as a novel biological control agent against plant-parasitic nematodes, as preliminarily validated in our experiments. The genomic resource and mechanistic insights provided herein lay a solid foundation for future research into flatworm predatory biology and pave the way for the development of sustainable, flatworm-based biocontrol strategies.

Conclusion

This study establishes the inaugural chromosome-level genome assembly for the nematophagous flatworm *L. nematophagus*, providing insights into the molecular foundations of its key adaptations: (1) Expansions in gene families associated with neural signaling, such as synaptic vesicles, and peptidases, including metalloendopeptidases, may enhance its specialized capabilities for nematode predation; (2) Genes for detoxification and transmembrane transport may facilitate its survival in soil environments; (3) Self-fertilization paired with high fecundity may offset the loss of regenerative abilities. These findings deepen our comprehension of evolutionary trade-offs within Platyhelminthes and underscore the significant potential of *L. nematophagus* as a biocontrol agent, as preliminarily validated by its efficacy against root-knot nematodes.

Materials and methods

Collection of Eggs from *L. nematophagus* for Whole Genome Sequencing

An adult *L. nematophagus* was placed in a 9 cm petri dish containing sterile water and fed with *Panagrellus redivivus* until it laid eggs, according to Focken's method [49]. After the eggs were laid, they were transferred to a new dish. This feeding process was repeated until a total of 100 eggs were collected. These eggs were then transferred to a larger culture basin and reared until a population size sufficient to produce approximately 100,000 eggs was achieved. A total of 100,000 eggs from this culture were collected using a 350-mesh screen and cleaned with sterile water, then promptly frozen in liquid nitrogen for storage at -80 ℃. DNA and RNA extracted from these egg samples were utilized for genome survey, HiFi, Hi-C, and transcriptome sequencing-assisted gene annotation.

DNA and RNA extractions, library construction, and sequencing

Genomic DNA and total RNA were extracted by Biomarker Technologies Co., Ltd (Beijing, China) using the sodium dodecyl sulfate (SDS) and TRIzol methods, respectively [50, 51]. The concentration and purity of DNA (A260/A280 and A260/A230 ratios) were assessed using a NanoDrop spectrophotometer, and the concentration of double-stranded DNA was quantified using a Qubit® 2.0 Fluorometer. The integrity of the DNA was evaluated via 1% agarose gel electrophoresis. Similarly, the concentration and purity of RNA were determined using NanoDrop, and its integrity was assessed with an Agilent 2100 Bioanalyzer.

Validated genomic DNA was fragmented to the target size of 350 bp by ultrasonication. A short-insert (350 bp) paired-end sequencing library was then constructed following the manufacturer's instructions (Illumina), which included end-repair, A-tailing, adapter ligation, size selection, and PCR amplification. The size and quantification of the library fragments were assessed using Qseq400 and Qubit. The qualified libraries were bound to Primer and Polymerase (both from PacBio, USA) using the PacBio Binding kit (PacBio, USA). The final reaction products were purified using AMpure PB Beads (PacBio, USA). Subsequently, the paired-end 150 bp (PE 150) sequences of the library were processed on the Sequel III (PacBio, USA) sequencer for online sequencing. The data generated from the sequencing were evaluated for GC distribution statistics, Q20 and Q30 values, and filtered to obtain high-quality data (clean reads).

For transcriptome-assisted gene annotation, mRNA was enriched by poly(A) selection using Oligo(dT) magnetic beads (Vazyme Biotech Co., Ltd.) and subsequently fragmented with the provided fragmentation buffer. The mRNA then served as a template for synthesizing the first strand of cDNA. After the synthesis of the second strand, the double-stranded cDNA was purified using AMPure XP beads (Vazyme Biotech Co., Ltd.). This purified double-stranded cDNA underwent end repair, A-tailing, adapter ligation, and target fragment selection using NEBNext End Prep (End Repair Reaction Buffer and End Prep Enzyme Mix). The cDNA library was then enriched through PCR amplification. Upon construction, the library's concentration and insert size were assessed using Qubit 2.0 and Agilent 2100. The concentration and quality of the library were further quantified accurately using Q-PCR. A library with a PE150 insert was constructed and sequenced on the Illumina NovaSeq 6000 platform.

K-mer analysis for genome survey

K-mer analysis was conducted on Illumina paired-end sequencing data (350 bp insert size) to estimate genome size, heterozygosity, and repeat sequence rate. The genome size was predicted using the formula G = K_num / peak_depth, where K_num denotes the total number of K-mers and peak_depth represents the depth of K-mers. A total of 72.28 Gb of clean reads were used to construct a K-mer distribution map. The proportion of Q20 sequencing data was 96.97%. The K-mer (k = 21) depth was calculated using Jellyfish software (v2.2.7) [52]. The heterozygosity rate and repeat sequence rate were estimated using GenomeScope software (v1.0) [53].

Genome assembly

The low-quality data from sequencing were filtered to obtain 2,920,733 reads. The high-accuracy HiFi reads (high-fidelity reads) were then assembled using Hifiasm v. 0.19 software (https://arxiv.org/abs/2008.01237) [54]. The assembly process with Hifiasm can be divided into three main steps. The first step involves identifying and

correcting haplotype errors and performing phasing of heterozygous SNPs. The second step entails constructing an assembly graph, specifically a phasing string graph, with reads serving as vertices and overlapping regions as edges. The third step involves generating assembly sequences. The final version of the genome was obtained by using NCBI's NT (nucleotide sequences) and mitochondrial database (https://fp.ncbi.nlm.nih.gov/refseq/release/mitochondrion/) to filter out NT contamination and mitochondrial sequences. The completeness and uniformity of the assembled genome's sequencing coverage were evaluated based on four key aspects. Firstly, Illumina reads were mapped onto the genome using Burrows-Wheeler Aligner software (https://bio-bwa.sourceforge.net/) [55–57]. Secondly, HiFi reads were compared with the assembled genome using Minimap2 software (https://github.com/lh3/minimap2) [58]. Thirdly, CEGMA v2.5 (Core Eukaryotic Genes Mapping Approach), which contains 458 conserved eukaryotic core genes and 248 highly conserved CEGs, was used to assess the completeness of the assembled genome with default parameters [59]. Finally, BUSCO v3 (Benchmarking Universal Single-Copy Orthologs), comprising 1,658 conserved insect genes, was executed using data from the INSECTA database (OrthoDB v9) to quantitatively assess the completeness of the assembled genome [60].

Hi-C

A single lane of the Illumina NovaSeq 6000 (150 bp PE reads) was used to construct Hi-C (High-throughput/Resolution Chromosome Conformation Capture) fragment libraries with insert sizes ranging from 300 to 700 bp, utilizing the Biomarker Hi-C Library Prep Kit for Illumina [61]. From the Hi-C library, 138.37 Gb of clean read pairs were obtained and subsequently mapped to the polished *L. nematophagus* genome using the bwa version 0.7.10 software [57] with default parameters. Reads whose mates mapped to different contigs were used for Hi-C-associated scaffolding. Hi-C-Pro v2.10.0 was employed to filter out invalid reads, including those resulting from self-ligation, non-ligation, PCR amplification, random breaks, and extreme fragments [62]. The LACHESIS software [63] was used to correct and assemble the Hi-C contigs with the parameters: CLUSTER_MIN_RE_SITES = 389; CLUSTER_MAX_LINK_DENSITY = 2; ORDER_MIN_N_RES_IN_TRUNK = 249; ORDER_MIN_N_RES_IN_SHREDS = 241. Then we used SubPhaser (v1.2.5) software [64] for subgenomic typing analysis. This process ultimately resulted in a chromosomal-level, high-quality assembly of the *L. nematophagus* genome.

Telomere analysis

Telomeric repeats were identified using the TeloExplorer module of quarTeT (v1.2.5r4) [65] with the parameter '-m 100'. The genome was scanned for tandem repeats characteristic of telomeres, and the presence of the repeat motif AACCCT at chromosome ends was examined.

Annotation of repetitive sequences

The annotation of transposable elements (TEs) and tandem repeats was conducted using specific workflows. TEs were identified through a blend of homology-based and de novo methods. A de novo repeat library was created using RepeatModeler (http://www.repeatmasker.org/RepeatModeler/) [66], which automatically executes two de novo repeat finding programs: RECON [67] and RepeatScout [68]. Full-length long terminal repeat retrotransposons (fl-LTR-RTs) were identified using both LTRharvest (v1.5.9) [69] and LTR_finder (v2.8) [70]. LTR_retriever [71] was used to produce high-quality, intact fl-LTR-RTs and a non-redundant LTR library. A non-redundant, species-specific TE library was constructed using this de novo TE sequence library alongside the known Dfam (v3.5) database [72]. The final TE sequences in the *L. nematophagus* genome were identified and classified by a homology search against this library using RepeatMasker (v4.12) [73]. Tandem repeats were annotated using Tandem Repeats Finder (TRF 409) [74], and microsatellites were identified with the Microsatellite identification tool (MISA v2.1) [75].

Annotation of protein-coding gene

Three methods were integrated to annotate protein-coding genes in the genome: de novo prediction, homology search, and transcript-based assembly. The de novo gene models were predicted using two ab initio gene-prediction software tools: Augustus (v3.1.0) [76] and SNAP [77]. In the homology-based approach, GeMoMa (v1.7) [78] was used with a reference gene model from four other species: *Caenorhabditis elegans*, *D. japonica*, *M. lignano*, and *S. mediterranea*. For the transcript-based prediction, RNA-sequencing data were mapped to the reference genome using Hisat (v2.1.0) [79] and assembled with Stringtie (v2.1.4) [80]. GeneMarkS-T (v5.1) [81] was utilized to predict genes based on the assembled transcripts. The PASA (v2.4.1) [82] software predicted genes using the unigenes and full-length transcripts from PacBio (ONT) sequencing, which were assembled by Trinity (v2.11) [83]. Gene models from these different approaches were combined using the EVM software (v1.1.1) [84] with the weight assignments listed in Table 3 for each evidence source.

After EVM integration, we performed several post-processing steps to refine the gene models. First, the EVM results were compared with the individual

Table 3 Weight assignments for different evidence sources in EVM integration

Evidence source	Category	Weight
PROTEIN (OTHER)	PROTEIN	50
PROTEIN (GeMoMa)	PROTEIN	50
AUGUSTUS	ABINITIO_PREDICTION	0.3
SNAP	ABINITIO_PREDICTION	0.3
GeMoMa (ab initio)	ABINITIO_PREDICTION	0.3
OTHER (OTHER_PREDICTION)	OTHER_PREDICTION	100
alignAssembly-PASA	TRANSCRIPT	50
RNAseq assembly	TRANSCRIPT	20

Weights reflect the relative confidence assigned to each evidence type during gene model integration in EVM (v1.1.1)

homology-based and transcriptome-based predictions using BLAST searches. Gene models present in the individual predictions but absent from the EVM output were added to the final set if they met the following criteria: sequence identity ≥ 50%, and both query and target coverage ≥ 80%. Second, loci lacking start codons or stop codons, containing ambiguous nucleotides (Ns), or harboring transposon-related protein domains were filtered out. Third, the filtered gene models were updated using PASA (v2.4.1) [82] with the parameters '-c alignAssembly.config --CPU 12 -A -L' to incorporate untranslated regions (UTRs) and alternative splicing information where transcript evidence supported such structures.

Functional annotation of protein-coding genes

Gene functions were inferred based on the best matches from alignments to the National Center for Biotechnology Information (NCBI) Non-Redundant (NR), EggNOG [85], KOG, TrEMBL [86], InterPro, and Swiss-Prot [86] protein databases, using Diamond BLASTP (v0.9.29.130) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database with an E-value threshold of 1E-3 [87]. Protein domains were annotated using InterProScan (v5.34-73.0) [88], which relies on the InterPro protein databases. The motifs and domains within gene models were identified using the PFAM databases [89]. Gene Ontology (GO) IDs for each gene were obtained from TrEMBL, InterPro, and EggNOG.

Annotation of non-coding RNA genes

The tRNAscan-SE (v1.3.1) [90] algorithm with default parameters was employed to identify tRNA genes, which are adaptor molecules composed of RNA that link the three-letter genetic code in messenger RNA (mRNA) with the twenty-letter amino acid code in proteins. For rRNA identification, barrnap (v0.9) [91] with default parameters was used to identify genes associated with rRNA. snoRNAs, a class of small RNA molecules that guide chemical modifications of other RNAs, primarily ribosomal RNAs, transfer RNAs, and small nuclear RNAs, were also cataloged. MiRNAs and snRNAs were identified using the Infernal (v1.1) [92] software against the Rfam (v14.5) [93] database with default parameters.

Pseudogene prediction

Pseudogenes typically share sequences similar to those of functional genes but may have lost their biological function due to genetic mutations, such as insertions and deletions. After masking the predicted functional genes, we used the GenBlastA (v1.0.4) [94] program to scan the entire genomes. We then analyzed putative candidates by searching for premature stop codons and frameshift mutations using GeneWise (v2.4.1) [95].

Gene family identification

For comparative genomic analyses, we selected a range of species to represent key evolutionary lineages and functional contrasts. *Drosophila melanogaster* and *Aedes aegypti* were chosen as well-characterized arthropod outgroups. *C. elegans* serves as a representative nematode, which is both a model organism and the prey of *L. nematophagus*. *D. japonica* and *S. mediterranea* were included as model planarians from the order Tricladida, representing the regenerative capabilities and genomic resources within Platyhelminthes that contrast sharply with *L. nematophagus* (order Rhabdocoela). To cluster families of protein-coding genes, we used proteins derived from the longest transcripts of each gene from *L. nematophagus* and other selected species mentioned above. The genomic information for these species, used in our comparative genomic analysis, is listed in Table S2-1. We employed the OrthoFinder software (v2.4.0) [96] to compare the protein-coding sequences from the *L. nematophagus* genome with those from the other species. The gene families identified were annotated using the PANTHER V14 database [97].

Phylogenetic analysis

To reconstruct the phylogenetic relationships, we used the 1,373 single-copy ortholog genes identified by OrthoFinder. Protein sequences of these orthologs were aligned using the MAFFT (v7.205) [98] program with the parameters '--localpair --maxiterate 1000', which is suitable for high-accuracy alignments. The corresponding coding DNA sequence (CDS) alignments were generated from the protein alignments using pal2nal (v14) [99] to preserve the reading frame. Conserved and reliably aligned regions were extracted using Gblocks (v0.91b) [100] with the parameters '-t = c -b5 = h', which removes poorly aligned positions and divergent regions. The resulting codon alignments were concatenated into a supergene matrix using an in-house Perl script. Phylogenetic inference was performed using IQ-TREE (v2.2.0) [101] under the maximum likelihood (ML) criterion. The best-fit substitution model was automatically selected using the

ModelFinder Plus option ('-m TEST') [102], which identified LG + F+I+G4 as the optimal model based on the Bayesian Information Criterion (BIC). Branch support was assessed with 1,000 ultrafast bootstrap replicates ('--bb 1000') [103].

Divergence time estimation

Divergence times were estimated using the MCMCTREE program implemented in the PAML package (v4.9i) [104]. To obtain an ultrametric tree required for subsequent gene family evolution analysis in CAFE, we calibrated the molecular clock using three fossil-based constraints derived from the TimeTree database [105] (http://www.timetree.org/):D. *japonica* vs. *A. aegypti*: 543.0–670.1 million years ago (Mya); *C. elegans* vs. *D. melanogaster*: 555.2–941.0 Mya; *D. melanogaster* vs. Aedes aegypti: 216.6–265.7 Mya. These calibration points were applied as soft bounds at the corresponding nodes in the phylogenetic tree. The MCMCTREE analysis was run with the following settings: model = 3 (HKY85), clock = 3 (independent rates), and burn-in = 10,000,000 iterations, followed by 100,000 samples collected every 100 iterations. Convergence was checked by examining the effective sample sizes (ESS) of the parameters. The resulting time-calibrated tree was used as input for CAFE analysis.

Gene family expansion and contraction analysis

Based on the divergence times and phylogenetic relationships obtained from the time-calibrated tree generated by MCMCTREE, we utilized CAFE (v4.2) [106] to analyze gene family expansion and contraction. CAFE employs a random birth and death model to explore gene gains or losses across specified phylogenetic trees. We calculated a conditional p-value for each gene family, considering those with a conditional p-value of less than 0.05 as exhibiting an accelerated rate of gene gain or loss. The gene families in *L. nematophagus* that expanded or contracted (p-value ≤ 0.05) were mapped to KEGG pathways for functional enrichment analysis. This analysis used hypergeometric test algorithms, and the Q-value (False Discovery Rate, FDR) was estimated to adjust the p-values using the R package (https://github.com/StoreyLab/qvalue).

Positively selected genes

We selected five species from the phylogenetic tree to estimate the ratio (ω) of non-synonymous (Ka) to synonymous (Ks) nucleotide substitutions using the PAML (v4.9i) package [104]. After obtaining high-quality alignments of the relevant sequences, we compared various evolutionary models within a likelihood framework based on the species trees. A branch-site model was implemented to determine the average ω across the tree (ω0), the ω of the branch under examination (ω2), and the ω of all other branches (ω1).

Statistical analysis

Data from the pot experiments were collected using Excel 2018, analyzed through one-way ANOVA with GraphPad Prism 10 software and charted.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-026-12863-z.

Supplementary Material 1.

Supplementary Material 2.

Supplementary Material 3.

Supplementary Material 4.

Acknowledgements

We thank Rui Zuo for discovering Luticola nematophagus and for his initial efforts in isolating and cultivating it. We are deeply grateful to the team of Academician Zhang Keqin at Yunnan University for providing the Meloidogyne incognita materials. We also acknowledge Beijing Biomake Biotechnologies Co., Ltd. for their contribution to the sequencing strategy and the execution of Illumina and PacBio sequencing, Survey, Hi-C, and comparative genomics analysis included in this manuscript.

Code Availability

No custom code was used in this study. Data analyses were performed using the standard bioinformatics tools specified in the methods.

Authors' contributions

Chongtao Guo and Wang Yin conceived and designed the study, created the figures, and wrote the initial draft, contributed equally to this work and should be considered co-first authors. Zhouqiong Zhang was responsible for the reproduction and imaging of **L. nematophagus** , as well as the expansion of cultures and collection of egg samples. Renju Deng revised and edited the final manuscript. Shouhui Pan provided funding for the work. Hai Zhang, Hao Cen, and Quan Zhang participated in the large-scale reproduction of **L. nematophagus** . Fei Dai, Congyong Li, and Qing Yang conducted the pot experiments. All authors provided feedback on previous versions of the manuscript and approved the final version.

Funding

This study was co-funded by the following three research projects:
1. The Guizhou Provincial Science and Technology Projects (QKHJC-[2024] youth076)
2. The Youth Fund (Natural Science) Project of Guizhou Academy of Agricultural Sciences ([2024]04)
3. The major Scientific and Technological Project of the Anshun Branch of Guizhou Tobacco Company (2023ASXM04).

Data availability

The genome assembly and raw sequencing data generated in this study have been deposited in the NCBI database under the following accession numbers: Genome assembly: JBRFHS000000000 (BioProject: PRJNA1308648, BioSample: SAMN50709787)Hi-C sequencing data: SRR36270746 (BioProject: PRJNA1372647, BioSample: SAMN53604513)Transcriptome sequencing data: SRR36264464 (BioProject: PRJNA1372277, BioSample: SAMN53539511) Survey sequencing data: SRR36264486 (BioProject: PRJNA1372274, BioSample: SAMN53539462)PacBio HiFi sequencing data: SRR36264478 (BioProject: PRJNA1372271, BioSample: SAMN53539452)The genome annotation data are available on Figshare: [https://doi.org/10.6084/m9.figshare.30000730.v1]All other data supporting the findings of this study are available within the article and its supplementary files.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Biotechnology Institute of Guizhou Province, Guiyang, Guizhou, China
[2]Guizhou Key Laboratory of Agricultural Biotechnology, Guiyang, Guizhou, China
[3]Ministry of Agriculture and Rural Affairs Key Laboratory of Crop Genetic Resources and Germplasm Innovation in Karst Region, Guiyang, Guizhou, China
[4]Guizhou Zhikang Biotechnology Company, Guizhou, China
[5]Anshun Branch of Guizhou Tobacco Company, Anshun, Guizhou, China

Received: 26 November 2025 / Accepted: 13 April 2026
Published online: 23 April 2026

References
1. Onkendi EM, Kariuki GM, Marais M, Moleleki LN. The threat of root-knot nematodes (*Meloidogyne* spp.) in Africa: a review. Plant Pathol. 2014;63(4):727–37. https://doi.org/10.1111/ppa.12202.
2. Khan A, Khan A, Ali A, Fatima S, Siddiqui MA. Root-Knot Nematodes (*Meloidogyne* spp.): Biology, Plant-Nematode Interactions and Their Environmentally. Benign Manage Strategies. 2023;75(6):2187–205. https://doi.org/10.1007/s10343-023-00886-5.
3. Ratnadass A, Fernandes P, Avelino J, Habib R. Plant species diversity for sustainable management of crop pests and diseases in agroecosystems: a review. Agron Sustain Dev. 2012;32(1):273–303. https://doi.org/10.1007/s13593-011-0022-4.
4. Li G, Liu T, Whalen JK, Wei Z. Nematodes: an overlooked tiny engineer of plant health. Trends in Plant Science. Trends Plant Sci. 2024;29(1):52–63. https://doi.org/10.1016/j.tplants.2023.06.022.
5. Afzal A. Revolutionizing nematode management to achieve global food security goals - An overview. Heliyon. 2024;10(3):e25325. 2024. https://doi.org/10.1016/j.heliyon.2024.e25325
6. Zuo R, Zhu J, Zhang Z, Pan S, Yin W, Guo C, et al. *Luticola nematophagus* gen. et sp. nov. (Platyhelminthes: Rhabdocoela), a Terrestrial Ascophorine Flatworm From China and Its Biological Characteristics. Zoolog Sci. 2025;42(4). https://doi.org/10.2108/zs240111.
7. Rink JC. Stem cell systems and regeneration in planaria. Dev Genes Evol. 2013;223(1–2):67–84. https://doi.org/10.1007/s00427-012-0426-4.
8. Sluys R, Riutort M. Planarian Diversity and Phylogeny. In: Rink JC, editor. Planarian Regeneration. New York, NY: Springer New York; 2018. pp. 1–56. https://doi.org/10.1007/978-1-4939-7802-1_1.
9. Vila-Farré M, Rink C. The Ecology of Freshwater Planarians. In: Rink JC, editor. Planarian Regeneration. New York, NY: Springer New York; 2018. pp. 173–205. https://doi.org/10.1007/978-1-4939-7802-1_3.
10. Tellakula NS, Collins E-MS, Kristan WB. Regeneration in planarians modifies behavioral switching. iScience. 2025;28(3):111947. https://doi.org/10.1016/j.isci.2025.111947.
11. Dai X, Li X, Tyshkovskiy A, Zuckerman C, Cheng N, Lin P, et al. Regeneration leads to global tissue rejuvenation in aging sexual planarians. Nat Aging. 2025;5(5):780–98. https://doi.org/10.1038/s43587-025-00847-9.
12. Wudarski J, Egger B, Ramm SA, Schärer L, Ladurner P, Zadesenets KS, et al. The free-living flatworm *Macrostomum lignano*. EvoDevo. 2020;11:5. https://doi.org/10.1186/s13227-020-00150-1.
13. Rossi L, Salvetti A. Planarian stem cell niche, the challenge for understanding tissue regeneration. Semin Cell Dev Biol. 2019;87:30–6. https://doi.org/10.1016/j.semcdb.2018.03.005.
14. Rabiasz A, Ziętkiewicz E. *Schmidtea mediterranea* as a Model Organism to Study the Molecular Background of Human Motile Ciliopathies. Int J Mol Sci. 2023;24(5):4472. https://doi.org/10.3390/ijms24054472.
15. Grohme MA, Schloissnig S, Rozanski A, Pippel M, Young GR, Winkler S, et al. The genome of *S. mediterranea* and the evolution of cellular core mechanisms. Nature. 2018;554(7690):56–61. https://doi.org/10.1038/nature25473.
16. Poinar G. A rhabdocoel turbellarian (Platyhelminthes, Typhloplanoida) in Baltic amber with a review of fossil and sub-fossil platyhelminths. Invertebr Biol. 2003;122:308–12. https://doi.org/10.1111/j.1744-7410.2003.tb00095.x.
17. Mogi M. Insects and other invertebrate predators. J Am Mosq Control Assoc. 2007;23:93–109. https://doi.org/10.2987/8756-971X(2007)23%255B93:IAOIP%255D2.0.CO;2.
18. Ogren RE. Predation behaviour of land planarians. Hydrobiologia. 1995;305(1):105–11. https://doi.org/10.1007/bf00036370.
19. Jones HD, Cumming MS. Feeding behaviour of the termite-eating planarian *Microplana termitophaga* (Platyhelminthes: Turbellaria: Tricladida: Terricola) in Zimbabwe. J Zool. 1998;245:53–64. https://doi.org/10.1111/j.1469-7998.1998.tb00071.x.
20. Boll PK, Leal-Zanchet AM. Predation on invasive land gastropods by a Neotropical land planarian. J Nat His. 2015;49:983–94. https://doi.org/10.1080/00222933.2014.981312.
21. Verma M, Mishra P, Verma AK. Potential biological agents for control of mosquitoes. Int J Mosq Res. 2023;10:19–27. https://doi.org/10.22271/23487941.2023.v10.i5a.692.
22. Tranchida MC, Pelizza SA, Micieli MV, Maciá A. Consequences of the introduction of the planarian *Girardia anceps* (Tricladida: Dugesiidae) in artificial containers with larvae of the mosquitoes *Aedes aegypti* and *Culex pipiens* (Diptera: Culicidae) from Argentina. Biol Control. 2014;71:49–55. https://doi.org/10.1016/j.biocontrol.2014.01.005.
23. Tian Q, Guo Q, Guo Y, Luo L, Kristiansen K, Han Z, et al. Whole-genome sequence of the planarian *Dugesia japonica* combining Illumina and PacBio data. Genomics. 2022;114(2):110293. https://doi.org/10.1016/j.ygeno.2022.110293.
24. An Y, Kawaguchi A, Zhao C, Toyoda A, Sharifi-Zarchi A, Mousavi SA, et al. Draft genome of *Dugesia japonica* provides insights into conserved regulatory elements of the brain restriction gene nou-darake in planarians. Zoological Lett. 2018;4:24. https://doi.org/10.1186/s40851-018-0102-2.
25. Ivanković M, Brand JN, Pandolfini L, Brown T, Pippel M, Rozanski A, et al. A comparative analysis of planarian genomes reveals regulatory conservation in the face of rapid structural divergence. Nat Commun. 2024;15(1):8215. https://doi.org/10.1038/s41467-024-52380-9.
26. Wasik K, Gurtowski J, Zhou X, Ramos OM, Delás MJ, Battistoni G, et al. Genome and transcriptome of the regeneration-competent flatworm, *Macrostomum lignano*. Proc Natl Acad Sci USA. 2015;112(40):12462–7. https://doi.org/10.1073/pnas.1516718112.
27. Grosbusch AL, Bertemes P, Kauffmann B, Gotsis C, Egger B. Do Not Lose Your Head over the Unequal Regeneration Capacity in Prolecithophoran Flatworms. Biology. 2022;11(11):1588. https://doi.org/10.3390/biology11111588.
28. Egger B, Gschwentner R, Rieger R. Free-living flatworms under the knife: past and present. Dev Genes Evol. 2007;217(2):89. https://doi.org/10.1007/s00427-006-0120-5.
29. Wiberg RAW, Brand JN, Viktorin G, Mitchell JO, Beisel C, Schärer L. Genome assemblies of the simultaneously hermaphroditic flatworms *Macrostomum cliftonense* and *Macrostomum hystrix*. G3. 2023;13(9):9. https://doi.org/10.1093/g3journal/jkad149
30. Póti Á, Szüts D, Vermezovic J. Mutational profile of the regenerative process and de novo genome assembly of the planarian *Schmidtea polychroa*. Nucleic Acids Res. 2024;52(4):1779–92. https://doi.org/10.1093/nar/gkad1250.
31. Vila-Farré M, Rozanski A, Ivanković M, Cleland J, Brand JN, Thalen F, et al. Evolutionary dynamics of whole-body regeneration across planarian flatworms. Nat Ecol Evol. 2023;7(12):2108–24. https://doi.org/10.1038/s41559-023-02221-7.
32. Steenkiste NV, Tessens B, Willems W, Backeljau T, Jondelius U, Artois T. A Comprehensive Molecular Phylogeny of Dalytyphloplanida (Platyhelminthes: Rhabdocoela) Reveals Multiple Escapes from the Marine Environment and Origins of Symbiotic Relationships. PLoS ONE. 2013;8(3):e59917. https://doi.org/10.1371/journal.pone.0059917.
33. Villani MG, Allee LL, Díaz A, Robbins PS. Adaptive strategies of edaphic arthropods. Annu Rev Entomol. 1999;44:233–56. https://doi.org/10.1146/annurev.ento.44.1.233.
34. Zhao YL, Wang DY. Formation and Regulation of Adaptive Response in Nematode *Caenorhabditis elegans*. Oxid Med Cell Longev. 2012;2012:1–6. https://doi.org/10.1155/2012/564093.

35. Braendle C, Milloz J, Felix A. Mechanisms and Evolution of Environmental Responses in *Caenorhabditis elegans*. Curr Top Dev Biol. 2008;80:171–207. https://doi.org/10.1016/S0070-2153(07)80005-6.
36. Silverman MA, Blaxter ML, Link CD. Biochemical Analysis of *Caenorhabditis elegans* Surface Mutants. J Nematol. 1997;29(3):296–305. https://doi.org/10.1006/jipa.1997.4682.
37. McKEAN PG, Pritchard DI. The action of a mast cell protease on the cuticular collagens of *Necator americanus*. Parasite Immunol. 2010;11(3):293–7. https://doi.org/10.1111/j.1365-3024.1989.tb00667.x.
38. Spiegel Y, McClure MA. The Surface Coat of Plant-Parasitic Nematodes: Chemical Composition, Origin, and Biological Role—A Review. J Nematol. 1995;27(2):127–34.
39. Yang J, Tian B, Liang L, Zhang K-Q. Extracellular enzymes and the pathogenesis of nematophagous fungi. Appl Microbiol Biotechnol. 2007;75(1):21–31. https://doi.org/10.1007/s00253-007-0881-4.
40. Huang X, Zhao N, Zhang K. Extracellular enzymes serving as virulence factors in nematophagous fungi involved in infection of the host. Res Microbiol. 2004;155(10):811–6. https://doi.org/10.1016/j.resmic.2004.07.003.
41. Yang J, Liang L, Li J, Zhang KQ. Nematicidal enzymes from microorganisms and their applications. Appl Microbiol Biotechnol. 2013;97(16):7081–95. https://doi.org/10.1007/s00253-013-5045-0.
42. Wrona FJ, Koopowitz H. Behavior of the rhabdocoel flatworm *Mesostoma ehrenbergii* in prey capture and feeding. Hydrobiologia. 1998;383(1):35–40. https://doi.org/10.1023/A:1003439827920.
43. Cseh A. Observations on food preference of Neotropical land planarians (Platyhelminthes), with emphasis on *Obama anthropophila*, and their phylogenetic diversification. ZOOLOGIA. 2017;34:e12622. https://doi.org/10.3897/zoologia.34.e12622.
44. Ruizhong Y, Boying Z, Zekai L, Xingzhou M, Xiaohan S, Qiuyu Q, et al. The chromosome-level genome of Chinese praying mantis *Tenodera sinensis* Mantodea: Mantidae reveals its biology as a predator. Gigascience. 2022;12:giad090. https://doi.org/10.1093/gigascience/giad090.
45. Zhang YX, Chen X, Wang JP, Zhang ZQ, Wei H, Yu HY, et al. Genomic insights into mite phylogeny, fitness, development, and reproduction. BMC Genomics. 2019;20(1):954. https://doi.org/10.1186/s12864-019-6281-1.
46. Liu R, Zhu R, Guo J, Yi T, Jin D. High-quality reference genome of predatory mite *Neoseiulus californicus* McGregor (Acari: Phytoseiidae) provides insights into its biological traits and potential RNAi off-target effects. Pest Manag Sci. 2025;81(3):1638–56. https://doi.org/10.1002/ps.8570.
47. Wang Y, Luo Y, Ge Y, Liu S, Liang W, Wu C, et al. Chromosome-level genome assembly of the predatory stink bug *Arma custos*. Sci Data. 2024;11(1):417. https://doi.org/10.1038/s41597-024-03270-8.
48. Shen X, Jin J, Zhang G, Yan B, Yu X, Wu H, et al. The chromosome-level genome assembly of *Aphidoletes aphidimyza* Rondani (Diptera: Cecidomyiidae). Sci Data. 2024;11(1):785. https://doi.org/10.1038/s41597-024-03614-4.
49. Schlechtriem C, Ricci M, Focken U, Becker K. Mass produced nematodes *Panagrellus redivivus* as live food for rearing carp larvae: preliminary results. Aquac Res. 2004;35(6):547–51. https://doi.org/10.1111/j.1365-2109.2004.01049.x.
50. Chen H, Rangasamy M, Tan SY, Wang H, Siegfried BD. Evaluation of Five Methods for Total DNA Extraction from Western Corn Rootworm Beetles. PLoS ONE. 2010;5:e11963. https://doi.org/10.1371/journal.pone.0011963.
51. Liu S-Y, Rink JC. Total RNA Isolation from Planarian Tissues. In: Rink JC, editor. Planarian Regeneration. New York, NY: Springer New York; 2018. pp. 259–65. https://doi.org/10.1007/978-1-4939-7802-1_6.
52. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k* -mers. Bioinformatics. 2011;27(6):764–70. https://doi.org/10.1093/bioinformatics/btr011.
53. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33(14):2202–4. https://doi.org/10.1093/bioinformatics/btx153.
54. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–5. https://doi.org/10.1038/s41592-020-01056-5.
55. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26(5):589–95. https://doi.org/10.1093/bioinformatics/btp698.
56. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics. 2009;25(14):1754–60. https://doi.org/10.1093/bioinformatics/btp324.
57. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv e-prints. 2013. https://doi.org/10.48550/arXiv.1303.3997. 1–3.
58. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100. https://doi.org/10.1093/bioinformatics/bty191.
59. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics. 2007;23(9):1061–7. https://doi.org/10.1093/bioinformatics/btm071.
60. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2. https://doi.org/10.1093/bioinformatics/btv351.
61. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;159(7):1665–80. https://doi.org/10.1016/j.cell.2014.11.021.
62. Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. 2015;16:259. https://doi.org/10.1186/s13059-015-0831-x.
63. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat Biotechnol. 2013;31(12):1119–25. https://doi.org/10.1038/nbt.2727.
64. Jia K, Wang Z, Wang L, Li G, Zhang W, Wang X, et al. SUBPHASER: a robust allopolyploid subgenome phasing method based on subgenome-specific *k* -mers. New Phytol. 2022;235(2):801–9. https://doi.org/10.1111/nph.18173.
65. Lin Y, Ye C, Li X, Chen Q, Wu Y, Zhang F, et al. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic Res. 2023;10(8):uhad127. https://doi.org/10.1093/hr/uhad127.
66. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020;117(17):9451–7. https://doi.org/10.1073/pnas.1921046117.
67. Bao Z, Eddy SR. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. Genome Res. 2002;12(8):1269–76. https://doi.org/10.1101/gr.88502.
68. Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21(Suppl 1):i351–8. https://doi.org/10.1093/bioinformatics/bti1018.
69. Ellinghaus D, Kurtz S, Willhoeft U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 2008;9:18. https://doi.org/10.1186/1471-2105-9-18.
70. Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35(Web Server):W265–8. https://doi.org/10.1093/nar/gkm286.
71. Ou S, Jiang N, LTR_retriever. A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol. 2018;176(2):1410–22. https://doi.org/10.1104/pp.17.01310.
72. Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, et al. Dfam: a database of repetitive DNA based on profile hidden Markov models. Nucleic Acids Res. 2012;41(Database issue):D70–82. https://doi.org/10.1093/nar/gks1265.
73. Tarailo-Graovac M, Chen N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Curr Protoc Bioinf. 2009;25. 4.10.1–4.10.14.
74. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80. https://doi.org/10.1093/nar/27.2.573.
75. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: a web server for microsatellite prediction. Bioinformatics. 2017;33(16):2583–5. https://doi.org/10.1093/bioinformatics/btx198.
76. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve *de novo* gene finding. Bioinformatics. 2008;24(5):637–44. https://doi.org/10.1093/bioinformatics/btn013.
77. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59. https://doi.org/10.1186/1471-2105-5-59.
78. Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. 2016;44:e89. https://doi.org/10.1093/nar/gkw092.
79. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60. https://doi.org/10.1038/nmeth.3317.

80. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5. https://doi.org/10.1038/nbt.3122.
81. Tang S, Lomsadze A, Borodovsky M. Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. 2015;43(12):e78–78. https://doi.org/10.1093/nar/gkv227.
82. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31(19):5654–66. https://doi.org/10.1093/nar/gkg770.
83. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Nat Biotechnol. 2011;29(7):644–52. https://doi.org/10.1038/nbt.1883. Full-length transcriptome assembly from RNA-Seq data without a reference genome.
84. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9(1):R7. https://doi.org/10.1186/gb-2008-9-1-r7.
85. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47(D1):D309–14. https://doi.org/10.1093/nar/gky1085.
86. Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003;31(1):365–70. https://doi.org/10.1093/nar/gkg095.
87. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 2016;44(D1):D457–62. https://doi.org/10.1093/nar/gkv1070.
88. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40. https://doi.org/10.1093/bioinformatics/btu031.
89. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, et al. Pfam: clans, web tools and services. Nucleic Acids Res. 2006;34(Database issue):D247–51. https://doi.org/10.1093/nar/gkj149.
90. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25(5):955–64. https://doi.org/10.1093/nar/25.5.955.
91. Yamamoto K, Yoneda Y, Makino A, Tanaka Y, Meng XY, Hashimoto J, et al. Draft Genome Sequence of *Bryobacteraceae* Strain F-183. Microbiol Resour Announc. 2022;11(1):e00453–21. https://doi.org/10.1128/mra.00453-21.
92. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29(22):2933–5. https://doi.org/10.1093/bioinformatics/btt509.
93. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 2004;33(Database issue):D121–4. https://doi.org/10.1093/nar/gki081.
94. She R, Chu JS, Wang K, Pei J, Chen N, GenBlastA. Enabling BLAST to identify homologous gene sequences. Genome Res. 2009;19(1):143–9. https://doi.org/10.1101/gr.082081.108.
95. Birney E, Clamp M, Durbin R, GeneWise. Genomewise Genome Res. 2004;14(5):988–95. https://doi.org/10.1101/gr.1865504.
96. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. https://doi.org/10.1186/s13059-019-1832-y.
97. Mi H, Muruganujan A, Ebert D, Huang X, Thomas D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019;47(D1):D419–26. https://doi.org/10.1093/nar/gky1038.
98. Katoh K, Asimenos G, Toh H. Multiple Alignment of DNA Sequences with MAFFT. Methods Mol Biol. 2009;537:39–64. https://doi.org/10.1007/978-1-59745-251-9_3.
99. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 2006;W609–12. https://doi.org/10.1093/nar/gkl315. 34(Web Server issue.
100. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. Mol Biol Evol. 2000;17(4):540–52. https://doi.org/10.1093/oxfordjournals.molbev.a026334.
101. Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015;32(1):268–74. https://doi.org/10.1093/molbev/msu300.
102. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14(6):587–9. https://doi.org/10.1038/nmeth.4285.
103. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap. Mol Biol Evol. 2018;35(2):518–22. https://doi.org/10.1093/molbev/msx281.
104. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics. 1997;13(5):555–6. https://doi.org/10.1093/bioinformatics/13.5.555.
105. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, et al. TimeTree 5: An Expanded Resource for Species Divergence Times. Mol Biol Evol. 2022;39(8):msac174. https://doi.org/10.1093/molbev/msac174.
106. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating Gene Gain and Loss Rates in the Presence of Error in Genome Assembly and Annotation Using CAFE 3. Mol Biol Evol. 2013;30(8):1987–97. https://doi.org/10.1093/molbev/mst100.

Publisher's note