Report

Current Biology

Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha

Highlights

- First genomes for the phylum Nematomorpha
- Nematomorph genomes lack ~30% of universal metazoan genes (BUSCOs)
- Widely conserved genes related to cilium formation have been lost in the phylum
- Chromosomes are largely rearranged compared to ancestral condition in animals

Authors

Tauana J. Cunha, Bruno A.S. de Medeiros, Arianna Lord, Martin V. Sørensen, Gonzalo Giribet

Correspondence

tcunha@fieldmuseum.org

In brief

Cunha et al. report the first genomes of the parasitic phylum Nematomorpha, worms known for manipulation of host behavior. Genomes show extensive chromosomal rearrangement, many lineage-specific orthologs of unknown function, and surprising loss of ~30% of conserved animal genes, providing the molecular basis for their lack of ciliary structures.

Cunha et al., 2023, Current Biology *33*, 1–8
August 21, 2023
https://doi.org/10.1016/j.cub.2023.07.003

CellPress

Current Biology

CellPress

Report

Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha

Tauana J. Cunha,[1,2,4,5,*] Bruno A.S. de Medeiros,[1,2] Arianna Lord,[1] Martin V. Sørensen,[3] and Gonzalo Giribet[1]
[1]Museum of Comparative Zoology, Department of Organismic and Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA
[2]Field Museum of Natural History, 1400 S DuSable Lake Shore Drive, Chicago, IL 60605, USA
[3]Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark
[4]Twitter: @tauanajc
[5]Lead contact
*Correspondence: tcunha@fieldmuseum.org
https://doi.org/10.1016/j.cub.2023.07.003

SUMMARY

Parasites may manipulate host behavior to increase the odds of transmission or to reach the proper environment to complete their life cycle.[1,2] Members of the phylum Nematomorpha (known as horsehair worms, hairworms, or Gordian worms) are large endoparasites that affect the behavior of their arthropod hosts. In terrestrial hosts, they cause erratic movements toward bodies of water,[3–6] where the adult worm emerges from the host to find mates for reproduction. We present a chromosome-level genome assembly for the freshwater *Acutogordius australiensis* and a draft assembly for one of the few known marine species, *Nectonema munidae*. The assemblies span 201 Mbp and 213 Mbp in length (N50: 38 Mbp and 716 Kbp), respectively, and reveal four chromosomes in *Acutogordius*, which are largely rearranged compared to the inferred ancestral condition in animals. Both nematomorph genomes have a relatively low number of genes (11,114 and 8,717, respectively) and lack a high proportion (~30%) of universal single-copy metazoan orthologs (BUSCO genes[7]). We demonstrate that missing genes are not an artifact of the assembly process, with the majority of missing orthologs being shared by the two independent assemblies. Missing BUSCOs are enriched for Gene Ontology (GO) terms associated with the organization of cilia and cell projections in other animals. We show that most cilium-related genes conserved across eukaryotes have been lost in Nematomorpha, providing a molecular basis for the suspected absence of ciliary structures in these animals.

RESULTS AND DISCUSSION

A chromosome-level assembly for Nematomorpha

Assemblies for two divergent species of hairworms (Figure 1) were built with 30–48× coverage of Nanopore long reads and polished with 230–320× Illumina data. *Acutogordius australiensis* was additionally scaffolded with 126× coverage of short-read Hi-C data. Raw data and assembly statistics for both genomes are shown in Table S1. For *Acutogordius*, we produced a chromosome-level assembly of 201 Mbp, 83.4% of which was assembled into four chromosome-scale scaffolds (scaffold N50 of 38 Mbp) (Figure 2). Although chromosome number had never been estimated for this species, this result is consistent with the haploid number of 1–8 chromosomes determined for other species in the same family.[8,9] The assembled chromosomes have lengths between 36.3 Mbp and 46.6 Mbp. The remaining 1,319 small scaffolds represent 16.6% of the genome, with the largest representing less than 0.07% of the assembly (Figure 2). The strong physical interactions shown in the contact map (bottom/right of Figure 2A) indicate that most of these scaffolds belong to one of the four chromosomes. For comparison, the draft assembly of *Acutogordius* before scaffolding with Hi-C data had an N50 of 230 Kbp and 2,368 contigs.

Despite limited tissue availability preventing us from obtaining Hi-C data for *Nectonema munidae*, the draft assembly has relatively high contiguity, with 1,061 scaffolds assembled into a 213 Mbp genome, and an N50 of 716.6 Kbp (Figure 2B). This is of greater quality than the draft assembly of *Acutogordius* before Hi-C mapping, which can be at least partially explained by variation in heterozygosity: 0.71% in *Nectonema* compared to a high level of 3.72% in *Acutogordius*, as estimated with GenomeScope.[10]

Genome architecture and annotation

Deep conservation of chromosomal linkages has been recognized across metazoan phyla, with 29 ancestral linkage groups (ALGs) inferred for Metazoa.[11] However, like in fruit flies and nematodes,[11] the genome of nematomorphs has been substantially rearranged into few chromosomes, and most orthologs from ALG origin are now dispersed throughout the genome of

Please cite this article in press as: Cunha et al., Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.07.003

Figure 1. Representatives of the two lineages of Nematomorpha

(A) Gordioidea: two live tangled individuals of *Gordionus violaceus* from Germany; photo by G. Giribet.

(B) Gordioidea: tail of a male specimen of *Acutogordius australiensis* from Australia; SEM by G. Giribet.

(C) Nectonematoidea: one individual of *Nectonema munidae* from Norway, body fragmented, collected from inside the squat lobster host; SEM by M. V. Sørensen.

Acutogordius (Figure 3A). Still, it is possible to recognize blocks of conserved synteny: two equivalences tie ALG_M to scaffold_3 and ALG_O2 to scaffold_4, scaffold_2 seems to be the product of an end-to-end fusion with mixing of ALG_A1a and ALG_L, and a centric insertion of ALG_I into ALG_D formed scaffold_1 (Figure 3A).

Repetitive elements comprise 49.41% and 53.47% of the genome of *Acutogordius* and *Nectonema* (Figure 3B), respectively. Interspersed repeats make up the majority of repetitive regions, with a high proportion of unclassified repeats (27%–34%), followed by DNA transposons (8%–16%). Prediction of protein coding genes revealed 11,114 and 8,717 genes in each of the assemblies. There is a positive relationship between number of genes and scaffold size, with the majority of predicted genes in *Acutogordius* being located on the four chromosome-level scaffolds, which contain 2,611, 2,775, 1,841, and 1,962 genes (Figure 3C). The number of genes in nematomorphs is notably lower than in other animals, and at the lower end of the range for parasitic worms (6,712–17,274 genes/species in nematodes and flatworms[12,13]). Other parasites with the smallest animal genomes ever recorded include myxozoan cnidarians (*Myxobolus cerebralis* at 22.5 Mpb and 5,533 genes[14]) and orthonectids (*Intoshia variabili* at 15.3 Mbp and 5,120 genes;[15] *I. linei* at 43.2 Mbp and ~9,000 genes[16,17]). In a comparison of representative protostomes, free-living taxa hold more species-specific orthogroups than parasitic ones (Figure S1), possibly reflecting a tendency of parasitic lineages toward smaller genomes. Still, nematomorphs reveal many lineage-specific orthologs (886 in *Acutogordius*, 449 in *Nectonema*), even when compared to other parasites (Figure S1). Future functional analyses across clades will help determine which genes lost or gained by Nematomorpha could be related to their parasitic lifestyle.

In addition to the low number of protein-coding genes, a reduction of non-coding DNA has been noted in some parasitic species (e.g., orthonectids,[16] nematodes,[13] and plants[18]). In nematomorphs, however, loss of genes does not seem to be accompanied by genome streamlining. The average number of introns per gene (7.1 in *Acutogordius*, 5.1 in *Nectonema*) is similar to that in other protostomes, regardless of lifestyle (Figure 3D). Furthermore, intron size and intergenic distances are larger than in free-living relatives (Figures 3E and 3F).

Of all predicted genes, functional annotations were obtained for 4,670 genes in *Acutogordius* and for 3,910 in *Nectonema*. Three regulatory muscle proteins have been identified on the body wall musculature of a Gordian worm based on immunocytochemistry,[19] of which caldesmon was unexpected for never having been detected in the genomes of invertebrates.[20] In our two nematomorph assemblies, no homolog was detected for caldesmon, supporting that results from electron microscopy imaging were likely a false positive from antibodies binding to other unknown proteins in the tissue.[19] For the other two proteins, we identified 5–6 genes with homology to calponin domains and 1–2 genes annotated as troponins.

Rampant loss of universal metazoan genes

Notably, evaluating our two nematomorph assemblies against a curated dataset of metazoan single-copy orthologs (BUSCO, benchmarking universal single-copy orthologs[7]) revealed that over 30% of such widely conserved genes are missing in the genomes of horsehair worms (Figure 2B). To investigate whether this could be an artifact of the assembly process, we first compared the sets of missing genes, finding that the missing orthologs in the two nematomorph genomes largely overlap (Figure 4A). If the quality of the assemblies was responsible for missing BUSCOs, we would expect missing genes to be randomly distributed in the genomes. Instead, we found that such a large identity overlap of 220 missing genes would be statistically implausible ($p < 2.2e{-}16$, $\chi^2 = 339.3$, df = 1).

We then considered the fact that, even with high-quality genomes, sequences that are highly derived compared to the reference clade might artificially display more missing BUSCOs due to strong sequence divergence preventing recognition of orthologs.[7] While this could be responsible for some of the genes missing in our assemblies, we note that other divergent phyla of Ecdysozoa with no representation in the reference Metazoa database show much lower proportions of missing BUSCOs, even with less contiguous assemblies, such as velvet worms (Onychophora, 3.7%)[21] and penis worms (Priapulida, 3%).[22] Even the model nematode *Caenorhabditis elegans*, well known for its small genome of 100 Mbp, has only 11% missing BUSCOs.[7] Evolutionary divergence alone is therefore not expected to be the cause of the large amount of undetected BUSCOs in nematomorphs.

Please cite this article in press as: Cunha et al., Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.07.003

CellPress

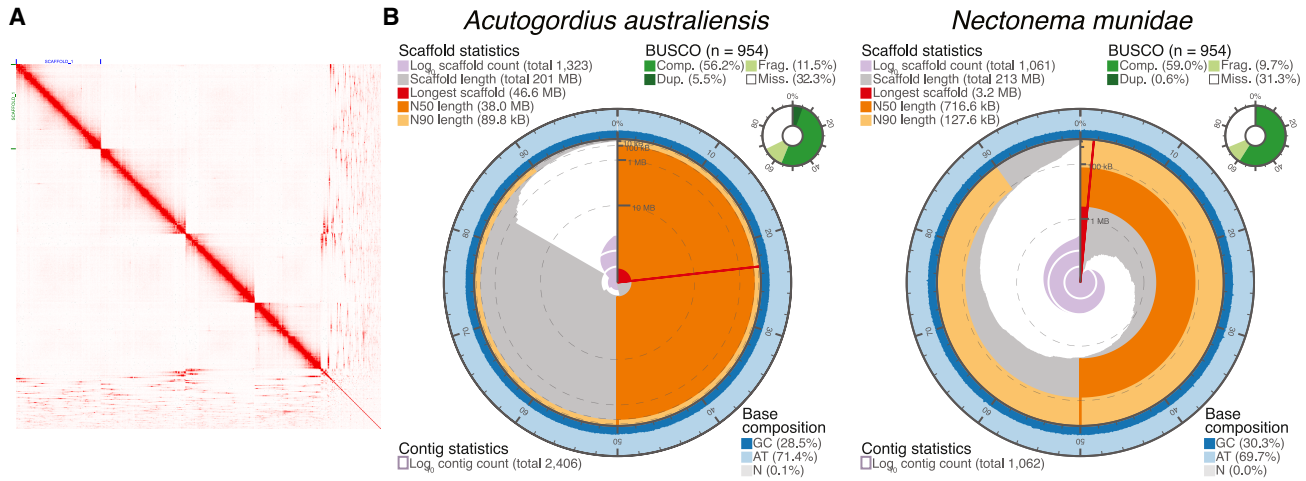

Figure 2. A chromosome-level assembly for Nematomorpha

(A) Hi-C contact map for the assembly of *Acutogordius australiensis*, showing 4 chromosome-scale scaffolds.

(B) Assembly statistics and BUSCO completeness for *A. australiensis* (C:56.2% [S:50.7%, D:5.5%], F:11.5%, M:32.3%) and *Nectonema munidae* (C:59.0% [S:58.4%, D:0.6%], F:9.7%, M:31.3%). Major differences in contiguity are noted in N50 values, as well as in the length of the longest scaffold. Nematomorph genomes lack a high proportion of universal metazoan orthologs (BUSCOs).

See also Table S1.

To investigate the nature of the missing universal genes, we did a functional enrichment analysis of Gene Ontology terms in the set of 220 missing genes compared to all 954 orthologs in the metazoan BUSCO dataset. We found that Biological Processes (BP) and Cellular Component (CC) terms that in other animals are associated with cilium and cell projection assembly are strongly over-represented in the set of genes missing in Nematomorpha (Figures 4B and S2A). Overall, 112 missing BUSCOs were significant for 43 terms (Figures S2B and S2C). Importantly, terms with the lowest p values in the enrichment analysis (red colors in Figure 4B) represent functions for which most genes in the background metazoan set are missing in nematomorphs (Figure 4C). For example, of the 14 genes associated with the top BP term (GO:0030030: cell projection organization) in the background set of metazoan BUSCOs, 13 are missing in both nematomorph genomes (Figures 4B and 4C; Table S2). The top 10 terms with the most significant loss of genes are all connected to cilium assembly, cell projection organization, and microtubule cytoskeleton (Figure 4C). Other enriched terms with related functions include organelle organization, cellular localization, and intracellular transport (Figure 4C).

Nematomorpha is one of the few animal phyla with no conclusive evidence for cilia.[23,24] Defined as organelles containing a microtubule-based cytoskeleton (axoneme) and supported by a basal body,[23,25] cilia are conserved at an even broader evolutionary level, sharing a similar ultrastructure across eukaryotes.[26,27] Mechano- and chemosensory perception, as well as male gamete motility, are some of the important roles played by ciliary structures across animals.[23] Nematomorphs, however, are known for aflagellate spermatozoa,[28] and despite ultrastructural studies spanning different species, developmental stages, and body regions,[29–32] only one uncertain instance of a microtubule-like feature in a cell projection has been recorded.[33,34] Our findings provide a molecular basis for the lack of cilia, with evidence that the genomic machinery responsible for making cilia no longer exists in the genomes of nematomorphs. The congruent gene loss in our two species, which represent the sister lineages Nectonematoidea (marine) and Gordioidea (freshwater),[35] further implies that this substantial genomic reduction very likely happened in the ancestor of modern nematomorphs.

In animals, cilia have a prominent role in sensory structures. A search for selected functional terms related to sensory genes in the genome annotations of *Acutogordius* and *Nectonema* resulted in less than 20 genes for each species. In ecdysozoan relatives, some of these gene families are extremely rich. For example, *C. elegans* boasts over 1,300 predicted G protein-coupled receptors (GPCRs),[36] and *D. melanogaster* about 200.[37] GPCRs are involved in a variety of physiological processes, including visual and olfactory senses. It is plausible that the absence of such genes in nematomorphs could be related to loss of sensory structures associated with cilia. While obligate parasitism has been tentatively linked to a reduction of sensory genes in some nematodes,[38] nematomorph adults and early larvae are free-living, presumably having some recognition mechanism for finding mates and hosts. Given the number of lineage-specific orthologs in nematomorphs, and of genes that cannot be annotated by homology, it is conceivable that the phylum might have unique sensory structures and pathways that remain to be characterized.

Long-read sequencing of museum specimens

The specimen of *Nectonema munidae* had been preserved in RNA*later* in −80°C for nearly 10 years, while specimens of *Acutogordius australiensis* were likewise preserved for under 1 year. Noteworthy laboratory observations are described to help inform future long-read sequencing of museum specimens. When extracting high-molecular-weight DNA, a high-salt method provided the highest yields and the longest fragments (Figure S3). Initial attempts to sequence a genomic library in the Nanopore MinION device resulted in very low pore occupancy and almost

Please cite this article in press as: Cunha et al., Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.07.003

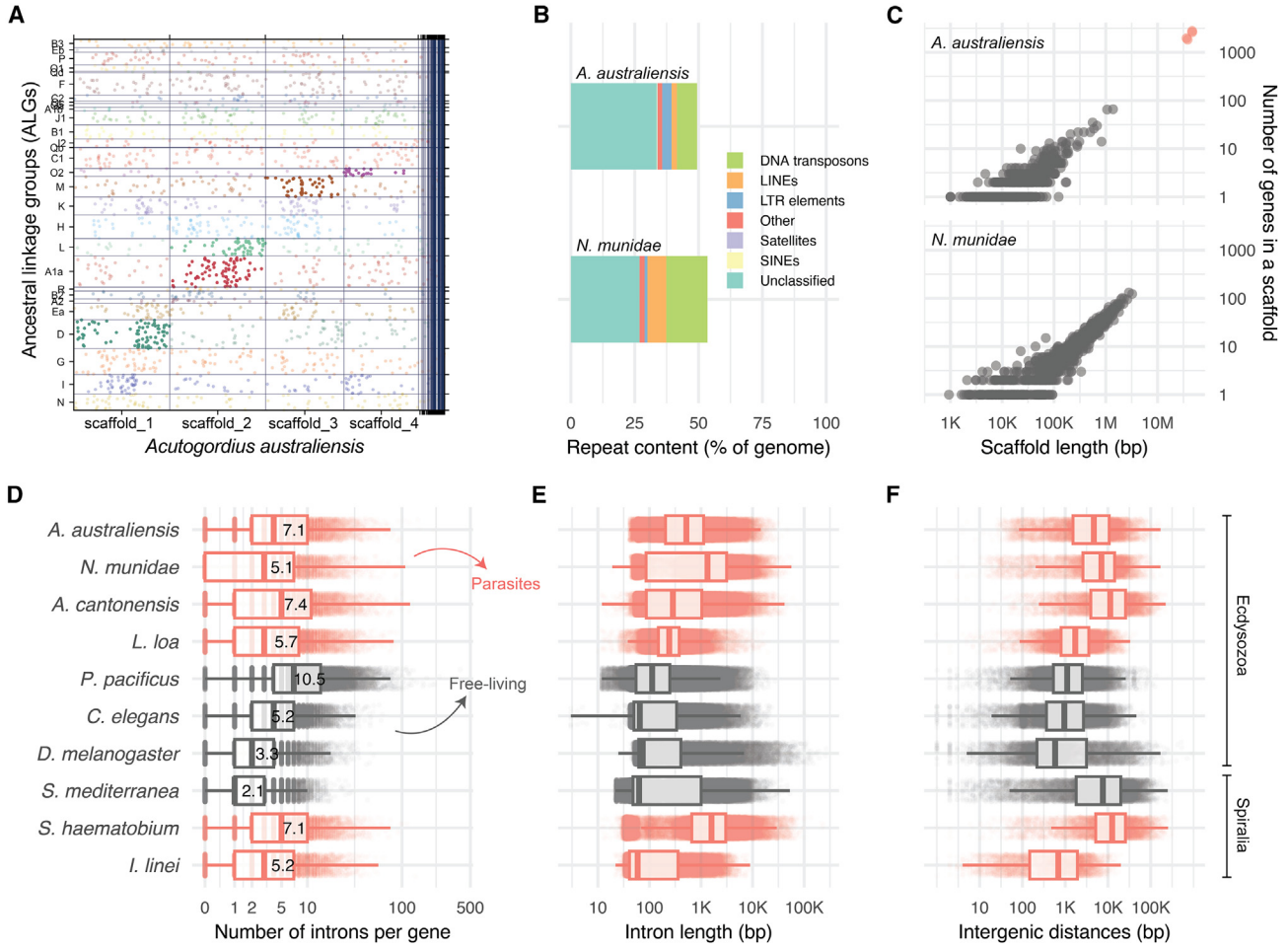

Figure 3. Genome architecture and annotation

(A) Syntenic relationships between the genome of *Acutogordius australiensis* and ancestral metazoan linkage groups (ALGs). Each of the 2,100 dots represents an ortholog, colored by ALG; opaque dots are significant orthologs in conserved linkage groups.

(B) Repetitive content of the two nematomorph genomes.

(C) Number of predicted genes per scaffold, with the four chromosome-scale scaffolds in salmon.

(D–F) Number of introns per gene, intron length, and intergenic distances for new nematomorph genomes and other representative parasitic and free-living species. The central line in boxplots represents the median; the average number of introns per gene is shown numerically. Nematomorpha, *Acutogordius australiensis* and *Nectonema munidae*; Nematoda, *Angiostrongylus cantonensis*, *Loa loa*, *Pristionchus pacificus*, *Caenorhabditis elegans*; Arthropoda, *Drosophila melanogaster*; Platyhelminthes, *Schmidtea mediterranea*, *Schistosoma haematobium*; Orthonectida, *Intoshia linei*.

no yield. We observed the same issue with other non-model invertebrate taxa, with initial pore occupancy as low as one active pore for a library made from mollusk tissue preserved in 95% EtOH at −20°C for about 5 years. We discovered that extending the DNA repair and end-prep step from the 5 min in the standard protocol to 30 min produced the expected optimal pore occupancy and high sequencing yield. We also used the PreCR Repair Mix (NEB) following DNA extraction as an additional measure to repair DNA. Applying the extended repair steps was successful for all taxa.[21,22]

Decline in pore occupancy throughout the run is a typical feature of ONT sequencing. We found that washing the flow cell when the rate of new output is reduced (at ~24 h) rescues unavailable pores and significantly increases yield (Figure S4). This entailed starting library preparation with up to 2 μg of input DNA whenever possible and splitting libraries into two aliquots of 400–600 ng, but smaller aliquots were also successful as a second load. For *Nectonema*, for example, one wash doubled the yield of the run to 10.3 Gb (Figure S3), with the second load containing about 200 ng of DNA.

Finally, a taxon-specific behavior was observed in handling nematomorph DNA. While the Short Read Eliminator kit (Circulomics; now Pacific Biosciences) was successfully used for size selection in other taxa we worked with, the DNA of *Acutogordius* consistently failed to precipitate, despite increased centrifugation time. We alternatively tried size selection with SPRI beads, but the DNA likewise remained in the supernatant despite extensive incubation times with the beads. It is unclear what caused these behaviors, possibly some undetected carryover from the DNA extractions. These size selection methods, therefore, were not applied in the final protocol, which nonetheless resulted in high-quality assemblies.

Please cite this article in press as: Cunha et al., Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.07.003

CellPress

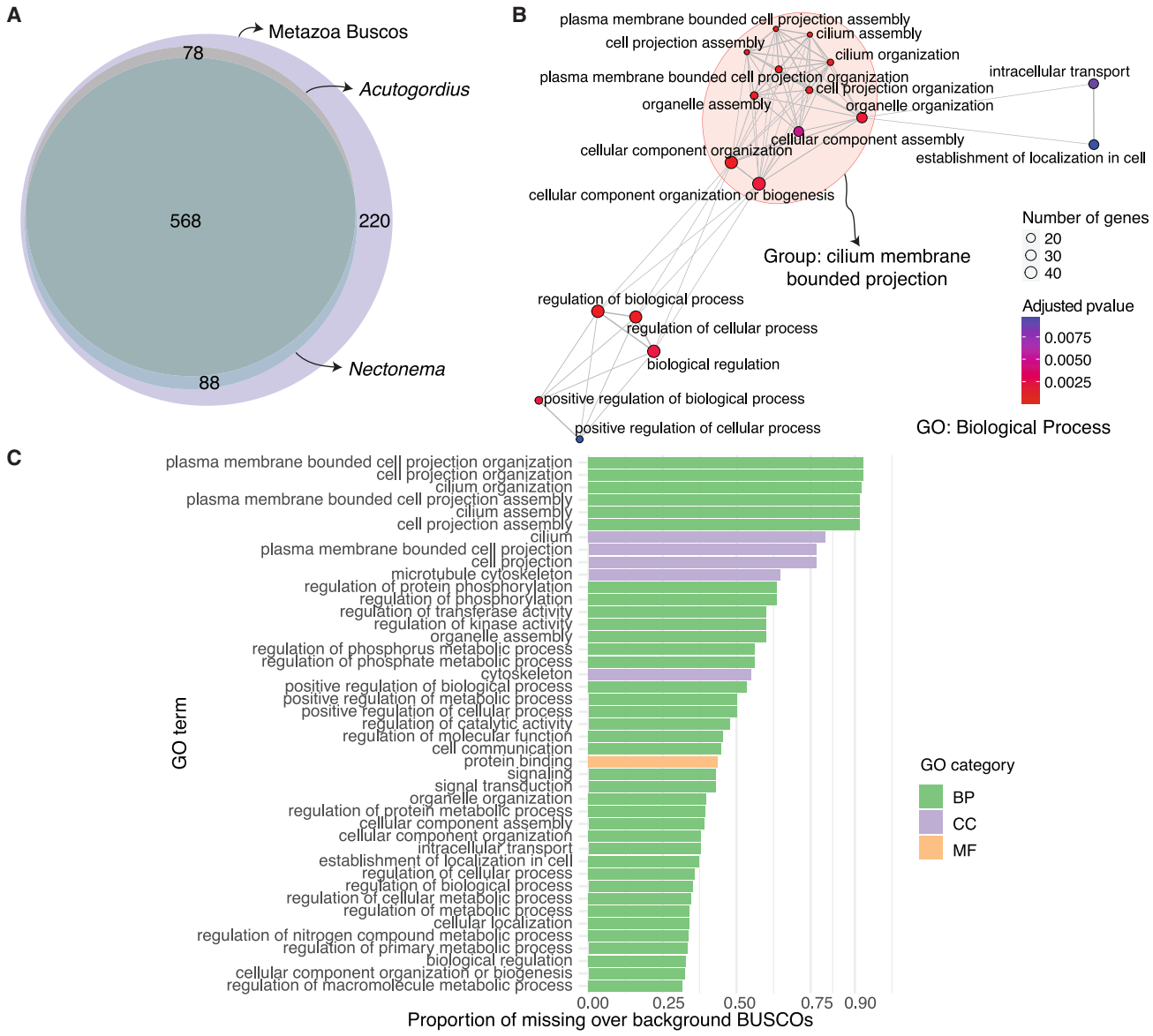

Figure 4. Rampant loss of universal metazoan genes in nematomorph genomes

(A) BUSCO genes present in the genome assemblies of *Acutogordius australiensis* and *Nectonema munidae*, in relation to 954 reference orthologs in the Metazoa dataset; 220 genes are missing in both species.

(B) Significant Gene Ontology (GO) terms in the set of missing BUSCOs. Each circle represents a Biological Process, colored by the statistical significance of the enrichment. Circle size represents the number of genes associated with the term. Shorter and thicker edges indicate higher similarity between terms.

(C) Degree of gene loss for significant GO terms, showing the proportion of genes missing out of all metazoan BUSCOs annotated with that term. Lower p values in (B) are correlated with higher proportion of missing genes for a GO term in (C).

See also Figure S2 and Table S2.

Gordioid horsehair worms are common parasites with notable effects on lotic systems, infecting a suite of host species in natural communities[39–41] and drastically altering food webs and the energy flow in forest-stream ecosystems.[5,42] They accelerate the development of paratenic hosts[43] and shape the behavior of definitive hosts,[3,6,44] in addition to being able to survive long periods in the digestive tract of predators that consume host species.[45] Despite their unique biology and ecological role, Nematomorpha is among the 10 animal phyla for which genomic representation has so far been neglected.[46] Here we present contiguous genomes for the two lineages of nematomorphs. We demonstrate extensive chromosomal rearrangement, a large number of lineage-specific orthologs of unknown function, and pervasive loss of universal metazoan genes, providing the molecular foundation that explains their lack of ciliary structures. These genomic resources will serve as an atlas of nematomorph gene repertoire, opening new avenues for investigating the genomic mechanisms

underlying parasitism, control of host behavior, and genome reduction.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Specimen data
- METHOD DETAILS
  - DNA extraction and sequencing
  - Estimation of genome size
  - Genome assembly
  - RNA sequencing
  - Genome annotation
  - Comparative analyses with other protostomes
  - Synteny
  - Functional enrichment

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.07.003.

ACKNOWLEDGMENTS

Computations were run on the FASRC Cannon cluster supported by the FAS Division of Science Research Computing Group at Harvard University, and on the cluster of the Grainger Bioinformatics Center at the Field Museum. T.J.C. was partially supported by a Field Museum Women's Board Postdoctoral Fellowship. Lab work was supported by internal funds from the Harvard Faculty of Arts and Sciences to G.G. and from the Harvard Department of Organismic and Evolutionary Biology to T.J.C. We thank Chris Laumer for sharing the high-salt DNA extraction protocol. Three reviewers provided feedback on this work, and we are grateful for their time and insights.

AUTHOR CONTRIBUTIONS

Conceptualization, G.G. and T.J.C.; methodology, T.J.C. and B.A.S.d.M.; formal analysis, T.J.C. and A.L.; investigation, T.J.C., B.A.S.d.M., and A.L.; resources, G.G. and M.V.S.; data curation, T.J.C.; writing – original draft, T.J.C.; writing – review & editing, all authors; supervision, G.G.; funding acquisition, G.G. and T.J.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 8, 2023
Revised: May 21, 2023
Accepted: July 3, 2023
Published: July 18, 2023

REFERENCES

1. Lafferty, K.D., and Shaw, J.C. (2013). Comparing mechanisms of host manipulation across host and parasite taxa. J. Exp. Biol. *216*, 56–66. https://doi.org/10.1242/jeb.073668.
2. Bhattarai, U.R., Doherty, J.-F., Dowle, E., and Gemmell, N.J. (2021). The adaptiveness of host behavioural manipulation assessed using Tinbergen's four questions. Trends Parasitol. *37*, 597–609. https://doi.org/10.1016/j.pt.2021.01.006.
3. Thomas, F., Schmidt-Rhaesa, A., Martin, G., Manu, C., Durand, P., and Renaud, F. (2002). Do hairworms (Nematomorpha) manipulate the water seeking behaviour of their terrestrial hosts? J. Evol. Biol. *15*, 356–361. https://doi.org/10.1046/j.1420-9101.2002.00410.x.
4. Sanchez, M.I., Ponton, F., Schmidt-Rhaesa, A., Hughes, D.P., Misse, D., and Thomas, F. (2008). Two steps to suicide in crickets harbouring hairworms. Anim. Behav. *76*, 1621–1624. https://doi.org/10.1016/j.anbehav.2008.07.018.
5. Sato, T., Watanabe, K., Kanaiwa, M., Niizuma, Y., Harada, Y., and Lafferty, K.D. (2011). Nematomorph parasites drive energy flow through a riparian ecosystem. Ecology *92*, 201–207. https://doi.org/10.1890/09-1565.1.
6. Obayashi, N., Iwatani, Y., Sakura, M., Tamotsu, S., Chiu, M.-C., and Sato, T. (2021). Enhanced polarotaxis can explain water-entry behaviour of mantids infected with nematomorph parasites. Curr. Biol. *31*, R777–R778. https://doi.org/10.1016/j.cub.2021.05.001.
7. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
8. Román-Palacios, C., Medina, C.A., Zhan, S.H., and Barker, M.S. (2021). Animal chromosome counts reveal a similar range of chromosome numbers but with less polyploidy in animals compared to flowering plants. J. Evol. Biol. *34*, 1333–1339. https://doi.org/10.1111/jeb.13884.
9. Makino, S. (1951). An Atlas of the Chromosome Numbers in Animals (Iowa State College Press). https://doi.org/10.5962/bhl.title.7295.
10. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204. https://doi.org/10.1093/bioinformatics/btx153.
11. Simakov, O., Bredeson, J., Berkoff, K., Marletaz, F., Mitros, T., Schultz, D.T., O'Connell, B.L., Dear, P., Martinez, D.E., Steele, R.E., et al. (2022). Deeply conserved synteny and the evolution of metazoan chromosomes. Sci. Adv. *8*, eabi5884, https://doi.org/10.1126/sciadv.abi5884.
12. International Helminth Genomes Consortium (2019). Comparative genomics of the major parasitic worms. Nat. Genet. *51*, 163–174. https://doi.org/10.1038/s41588-018-0262-1.
13. Burke, M., Scholl, E.H., Bird, D.M., Schaff, J.E., Colman, S.D., Crowell, R., Diener, S., Gordon, O., Graham, S., Wang, X., et al. (2015). The plant parasite *Pratylenchus coffeae* carries a minimal nematode genome. Nematology *17*, 621–637. https://doi.org/10.1163/15685411-00002901.
14. Chang, E.S., Neuhof, M., Rubinstein, N.D., Diamant, A., Philippe, H., Huchon, D., and Cartwright, P. (2015). Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. Proc. Natl. Acad. Sci. USA *112*, 14912–14917. https://doi.org/10.1073/pnas.1511468112.
15. Slyusarev, G.S., Starunov, V.V., Bondarenko, A.S., Zorina, N.A., and Bondarenko, N.I. (2020). Extreme genome and nervous system streamlining in the invertebrate parasite *Intoshia variabili*. Curr. Biol. *30*, 1292–1298.e3. https://doi.org/10.1016/j.cub.2020.01.061.
16. Mikhailov, K.V., Slyusarev, G.S., Nikitin, M.A., Logacheva, M.D., Penin, A.A., Aleoshin, V.V., and Panchin, Y.V. (2016). The genome of *Intoshia linei* affirms Orthonectids as highly simplified Spiralians. Curr. Biol. *26*, 1768–1774. https://doi.org/10.1016/j.cub.2016.05.007.
17. Giribet, G. (2016). Zoology: invertebrates that parasitize invertebrates. Curr. Biol. *26*, R537–R539. https://doi.org/10.1016/j.cub.2016.05.040.
18. Cai, L., Arnold, B.J., Xi, Z., Khost, D.E., Patel, N., Hartmann, C.B., Manickam, S., Sasirat, S., Nikolov, L.A., Mathews, S., et al. (2021). Deeply altered genome architecture in the endoparasitic flowering plant *Sapria himalayana* Griff. (Rafflesiaceae). Curr. Biol. *31*, 1002–1011.e9. https://doi.org/10.1016/j.cub.2020.12.045.

19. Meyer-Rochow, V.B., and Royuela, M. (2018). Immunocytochemically determined regulatory proteins, troponin, calponin and caldesmon, may occur together in the musculature of a Gordian worm (Ecdysozoa, Cycloneuralia, Nematomorpha). Zoomorphology *137*, 13–17. https://doi.org/10.1007/s00435-017-0375-6.

20. Steinmetz, P.R.H., Kraus, J.E.M., Larroux, C., Hammel, J.U., Amon-Hassenzahl, A., Houliston, E., Wörheide, G., Nickel, M., Degnan, B.M., and Technau, U. (2012). Independent evolution of striated muscles in cnidarians and bilaterians. Nature *487*, 231–234. https://doi.org/10.1038/nature11180.

21. Sato, S., Cunha, T.J., de Medeiros, B.A.S., Khost, D.E., Sackton, T.B., and Giribet, G. (2023). Sizing up the onychophoran genome: repeats, introns, and gene family expansion contribute to genome gigantism in *Epiperipatus broadwayi*. Genome Biol. Evol. *15*, evad021, https://doi.org/10.1093/gbe/evad021.

22. Lord, A., Cunha, T.J., de Medeiros, B.A.S., Sato, S., Khost, D.E., Sackton, T.B., and Giribet, G. (2023). Expanding on our knowledge of ecdysozoan genomes, a contiguous assembly of the meiofaunal priapulan *Tubiluchus corallicola*. Genome Biol. Evol. *15*, evad103, https://doi.org/10.1093/gbe/evad103.

23. Bezares-Calderón, L.A., Berger, J., and Jékely, G. (2020). Diversity of cilia-based mechanosensory systems and their functions in marine animal behaviour. Philos. Trans. R. Soc. Lond. B Biol. Sci. *375*, 20190376, https://doi.org/10.1098/rstb.2019.0376.

24. Brusca, R.C., Giribet, G., and Moore, W. (2023). Invertebrates, 4th Edition (Sinauer Associates and Oxford University Press).

25. Ishikawa, T. (2017). Axoneme structure from motile cilia. Cold Spring Harb. Perspect. Biol. *9*, a028076. https://doi.org/10.1101/cshperspect.a028076.

26. Mitchell, D.R. (2007). The evolution of eukaryotic cilia and flagella as motile and sensory organelles. Adv. Exp. Med. Biol. *607*, 130–140. https://doi.org/10.1007/978-0-387-74021-8_11.

27. Haimo, L.T., and Rosenbaum, J.L. (1981). Cilia, flagella, and microtubules. J. Cell Biol. *91*, 125s–130s. https://doi.org/10.1083/jcb.91.3.125s.

28. Sokolova, E.A., Zograf, J.K., and Yushin, V.V. (2022). Ultrastructure of spermatozoa of a hairworm *Gordionus alpestris* (Villot, 1885) (Nematomorpha, Chordodidae). Invertebr. Reprod. Dev. *66*, 88–96. https://doi.org/10.1080/07924259.2022.2039308.

29. Schmidt-Rhaesa, A. (1996). Ultrastructure of the anterior end in three ontogenetic stages of *Nectonema munidae* (Nematomorpha). Acta Zool. *77*, 267–278. https://doi.org/10.1111/j.1463-6395.1996.tb01271.x.

30. Eakin, R.M., and Brandenburger, J.L. (1974). Ultrastructural features of a Gordian worm (Nematomorpha). J. Ultrastruct. Res. *46*, 351–374. https://doi.org/10.1016/S0022-5320(74)90062-8.

31. Zanca, F., De Villalobos, C., Schmidt-Rhaesa, A., and Achiorno, C. (2006). Revision of the genus *Chordodes* (Gordiida, Nematomorpha) from Africa—I. Ultrastructural redescription of *Chordodes gariazzi* Camerano, 1902, *C. heinzei* Sciacchitano, 1937, *C. kolensis* Sciacchitano, 1933, *C. muelleri* Sciacchitano, 1937. J. Nat. Hist. *40*, 17–31. https://doi.org/10.1080/00222930600617898.

32. Eldarov, C.M., Vays, V.B., Vangeli, I.M., Averina, O.A., Efeykin, B.D., and Bakeeva, L.E. (2020). Morphometric analysis of the internal ultrastructure of mitochondria of muscle tissue in horsehair worm *Gordionus alpestris* (Nematomorpha). Biochem. Moscow. Suppl. Ser. A. *14*, 255–259. https://doi.org/10.1134/s199074782002004x.

33. Schmidt-Rhaesa, A. (2004). Ultrastructure of an integumental organ with probable sensory function in *Paragordius varius* (Nematomorpha). Acta Zool. *85*, 15–19. https://doi.org/10.1111/j.0001-7272.2004.00153.x.

34. Schmidt-Rhaesa, A. (2005). Morphogenesis of *Paragordius varius* (Nematomorpha) during the parasitic phase. Zoomorphology *124*, 33–46. https://doi.org/10.1007/s00435-005-0109-z.

35. Bleidorn, C., Schmidt-Rhaesa, A., and Garey, J.R. (2005). Systematic relationships of Nematomorpha based on molecular and morphological data. Invertebr. Biol. *121*, 357–364. https://doi.org/10.1111/j.1744-7410.2002.tb00136.x.

36. Thomas, J.H., and Robertson, H.M. (2008). The *Caenorhabditis* chemoreceptor gene families. BMC Biol. *6*, 42. https://doi.org/10.1186/1741-7007-6-42.

37. Brody, T., and Cravchik, A. (2000). *Drosophila melanogaster* G protein–coupled receptors. J. Cell Biol. *150*, F83–F88. https://doi.org/10.1083/jcb.150.2.f83.

38. Cardoso, J.C.R., Félix, R.C., Fonseca, V.G., and Power, D.M. (2012). Feeding and the rhodopsin family G-protein coupled receptors in nematodes and arthropods. Front. Endocrinol. *3*, 157. https://doi.org/10.3389/fendo.2012.00157.

39. Hanelt, B., and Janovy, J. (2005). Spanning the gap: experimental determination of paratenic host specificity of horsehair worms (Nematomorpha: Gordiida). Invertebr. Biol. *122*, 12–18. https://doi.org/10.1111/j.1744-7410.2003.tb00068.x.

40. Doherty, J.-F., Filion, A., and Poulin, R. (2022). Infection patterns and new definitive host records for New Zealand gordiid hairworms (phylum Nematomorpha). Parasitol. Int. *90*, 102598, https://doi.org/10.1016/j.parint.2022.102598.

41. Doherty, J.-F., and Poulin, R. (2022). Come with me if you want to live: sympatric parasites follow different transmission routes through aquatic host communities. Int. J. Parasitol. *52*, 293–303. https://doi.org/10.1016/j.ijpara.2021.11.009.

42. Sato, T., Egusa, T., Fukushima, K., Oda, T., Ohte, N., Tokuchi, N., Watanabe, K., Kanaiwa, M., Murakami, I., and Lafferty, K.D. (2012). Nematomorph parasites indirectly alter the food web and ecosystem function of streams through behavioural manipulation of their cricket hosts. Ecol. Lett. *15*, 786–793. https://doi.org/10.1111/j.1461-0248.2012.01798.x.

43. Doherty, J.-F., and Poulin, R. (2022). The return to land: association between hairworm infection and aquatic insect development. Parasitol. Res. *121*, 667–673. https://doi.org/10.1007/s00436-021-07410-6.

44. Ponton, F., Otálora-Luna, F., Lefèvre, T., Guerin, P.M., Lebarbenchon, C., Duneau, D., Biron, D.G., and Thomas, F. (2011). Water-seeking behavior in worm-infected crickets and reversibility of parasitic manipulation. Behav. Ecol. *22*, 392–400. https://doi.org/10.1093/beheco/arq215.

45. Ponton, F., Lebarbenchon, C., Lefèvre, T., Biron, D.G., Duneau, D., Hughes, D.P., and Thomas, F. (2006). Parasite survives predation on its host. Nature *440*, 756. https://doi.org/10.1038/440756a.

46. Hotaling, S., Kelley, J.L., and Frandsen, P.B. (2021). Toward a genome sequence for every animal: where are we now? Proc. Natl. Acad. Sci. USA *118*, e2109019118, https://doi.org/10.1073/pnas.2109019118.

47. Laumer, C.E., Fernández, R., Lemer, S., Combosch, D., Kocot, K.M., Riesgo, A., Andrade, S.C.S., Sterrer, W., Sørensen, M.V., and Giribet, G. (2019). Revisiting metazoan phylogeny with genomic sampling of all phyla. Proc. Biol. Sci. *286*, 20190831, https://doi.org/10.1098/rspb.2019.0831.

48. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

49. De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., and Van Broeckhoven, C. (2018). NanoPack: visualizing and processing long-read sequencing data. Bioinformatics *34*, 2666–2669. https://doi.org/10.1093/bioinformatics/bty149.

50. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546. https://doi.org/10.1038/s41587-019-0072-8.

51. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

52. Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. *27*, 737–746. https://doi.org/10.1101/gr.214270.116.

53. Guan, D., Mccarthy, S.A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics *36*, 2896–2898. https://doi.org/10.1093/bioinformatics/btaa025.

54. Krueger, F., James, F., Ewels, P., Afyounian, E., and Schuster-Boeckler, B. (2018). Trim Galore. https://doi.org/10.5281/zenodo.5127899.

55. Kundu, R., Casey, J., and Sung, W.-K. (2019). HyPo: super fast & accurate polisher for long read genome assemblies. Preprint at bioRxiv. https://doi.org/10.1101/2019.12.19.882506.

56. Challis, R., Richards, E., Rajan, J., Cochrane, G., and Blaxter, M. (2020). BlobToolKit – interactive quality assessment of genome assemblies. G3 (Bethesda) *10*, 1361–1374. https://doi.org/10.1534/g3.119.400908.

57. Zhou, C., McCarthy, S.A., and Durbin, R. (2023). YaHS: yet another Hi-C scaffolding tool. Bioinformatics *39*, btac808, https://doi.org/10.1093/bioinformatics/btac808.

58. Robinson, J.T., Turner, D., Durand, N.C., Thorvaldsdóttir, H., Mesirov, J.P., and Aiden, E.L. (2018). Juicebox.js provides a cloud-based visualization system for Hi-C data. Cell Syst. *6*, 256–258.e1. https://doi.org/10.1016/j.cels.2018.01.001.

59. Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., and Zdobnov, E.M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol. Biol. Evol. *38*, 4647–4654. https://doi.org/10.1093/molbev/msab199.

60. Challis, R. (2017). rjchallis/assembly-stats v17.02. https://doi.org/10.5281/zenodo.322347.

61. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117.

62. Smit, A., Hubley, R., and Green, P. (2015). RepeatMasker Open-4.0. http://www.repeatmasker.org.

63. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. *37*, 907–915. https://doi.org/10.1038/s41587-019-0201-4.

64. Brůna, T., Hoff, K.J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom. Bioinform. *3*, lqaa108. https://doi.org/10.1093/nargab/lqaa108.

65. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*, 637–644. https://doi.org/10.1093/bioinformatics/btn013.

66. Hoff, K.J., Lomsadze, A., Borodovsky, M., and Stanke, M. (2019). Whole-genome annotation with BRAKER. Methods Mol. Biol. *1962*, 65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.

67. Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinf. *7*, 62. https://doi.org/10.1186/1471-2105-7-62.

68. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment using DIAMOND. Nat. Methods *12*, 59–60. https://doi.org/10.1038/nmeth.3176.

69. Lomsadze, A., Burns, P.D., and Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res. *42*, e119. https://doi.org/10.1093/nar/gku557.

70. Cantalapiedra, C.P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol. Biol. Evol. *38*, 5825–5829. https://doi.org/10.1093/molbev/msab293.

71. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. *47*, D309–D314. https://doi.org/10.1093/nar/gky1085.

72. Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. *20*, 238. https://doi.org/10.1186/s13059-019-1832-y.

73. Schultz, D.T., Haddock, S.H.D., Bredeson, J.V., Green, R.E., Simakov, O., and Rokhsar, D.S. (2023). Ancient gene linkages support ctenophores as sister to other animals. Nature *618*, 110–117. https://doi.org/10.1038/s41586-023-05936-6.

74. Kriventseva, E.V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F.A., and Zdobnov, E.M. (2019). OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. *47*, D807–D811. https://doi.org/10.1093/nar/gky1053.

75. Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS A J. Integr. Biol. *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

76. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. Innovation *2*, 100141, https://doi.org/10.1016/j.xinn.2021.100141.

77. R Core Team (2023). R: a language and environment for statistical computing (R Foundation for Statistical Computing).

78. Spiridonov, S. (1984). Two new Nematomorpha species of the family Gordiidae. In Proceedings of the Zoological Institute USSR Academy of Sciences, *126*, pp. 97–101.

79. Schmidt-Rhaesa, A., and Geraci, C.J. (2006). Two new species of *Acutogordius* (Nematomorpha), with a brief review of literature data of this genus. Syst. Biodivers. *4*, 427–433. https://doi.org/10.1017/s1477200006001964.

80. Dong, C. (2017). Purification of HMW DNA from Fungi for long read sequencing. protocols.io. https://doi.org/10.17504/protocols.io.hbvb2n6.

81. Schalamun, M., Nagar, R., Kainer, D., Beavan, E., Eccles, D., Rathjen, J.P., Lanfear, R., and Schwessinger, B. (2019). Harnessing the MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant tissue from *Eucalyptus pauciflora*. Mol. Ecol. Resour. *19*, 77–89. https://doi.org/10.1111/1755-0998.12938.

82. Kumar, S., Jones, M., Koutsovoulos, G., Clarke, M., and Blaxter, M. (2013). Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front. Genet. *4*, 237. https://doi.org/10.3389/fgene.2013.00237.

83. Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Please cite this article in press as: Cunha et al., Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.07.003

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Acutogordius australiensis	Museum of Comparative Zoology	https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:152393
Nectonema munidae	Museum of Comparative Zoology	https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:153622
Chemicals, peptides, and recombinant proteins		
High salt DNA extraction	Donnan Laboratories, University of Liverpool	http://web.archive.org/web/20160207182543/https://www.liverpool.ac.uk/~kempsj/IsolationofDNA.pdf
TRIzol Reagent	Invitrogen	Cat#15596026
Critical commercial assays		
PreCR Repair Mix	New England Biolabs	https://www.neb.com/products/m0309-precr-repair-mix
KAPA HyperPlus Kit	Roche	https://sequencing.roche.com/us/en/products/group/kapa-hyperplus-kits.html
Arima Hi-C Kit	Arima Genomics	https://arimagenomics.com/products/genome-assembly-hic
KAPA HyperPrep Kit	Roche	https://sequencing.roche.com/us/en/products/group/kapa-hyperprep-kits.html
Genomic DNA by Ligation SQK-LSK109 v.14Aug2019	Oxford Nanopore Technologies	https://community.nanoporetech.com/protocols/gDNA-sqk-lsk109/v
EXP-WSH003 v.18Sep2019	Oxford Nanopore Technologies	https://community.nanoporetech.com/protocols/flow-cell-wash-kit-protocol/v
KAPA mRNA HyperPrep Kit	Roche	https://sequencing.roche.com/us/en/products/group/kapa-rna-hyperprep-kits.html
Deposited data		
Nanopore and Illumina raw data from genomic libraries of *Acutogordius australiensis* and *Nectonema munidae*	This paper	NCBI BioProject: PRJNA983812
Genome assemblies and gene annotations for *Acutogordius australiensis* and *Nectonema munidae*	This paper	https://doi.org/10.6084/m9.figshare.23419922; GenBank: JAUIRJ010000000; GenBank: JAUIRK010000000
RNA-seq data for *Acutogordius australiensis*	This paper	NCBI BioProject: PRJNA983812
RNA-seq data for *Nectonema munidae*	Laumer et al.[47]	NCBI SRX5417513
Genome assemblies and annotations from 8 protostomes	NCBI; WormBase ParaSite	GCF_000002985.6; GCA_000180635.4; GCA_009735665.1; GCF_000183805.2; GCA_000001215.4; GCF_000699445.3; GCA_001642005.1; SmedGD_v1.3
Software and algorithms		
Jellyfish v2.2.5	Marçais and Kingsford[48]	https://github.com/gmarcais/Jellyfish
GenomeScope	Vurture et al.[10]	http://qb.cshl.edu/genomescope
Guppy v4.5.2	Oxford Nanopore Technologies	https://community.nanoporetech.com/downloads
NanoPlot v1.29.0	De Coster et al.[49]	https://github.com/wdecoster/nanoplot
Flye v2.8.1	Kolmogorov et al.[50]	https://github.com/fenderglass/Flye
BWA v0.7.17	Li H.[51]	https://github.com/lh3/bwa
Racon v1.4.21	Vaser et al.[52]	https://github.com/lbcb-sci/racon

(*Continued on next page*)

Please cite this article in press as: Cunha et al., Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.07.003

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Medaka v1.3.2	Oxford Nanopore Technologies	https://github.com/nanoporetech/medaka
purge_dups v1.2.5	Guan et al.[53]	https://github.com/dfguan/purge_dups
TrimGalore	Krueger et al.[54]	https://github.com/FelixKrueger/TrimGalore
HyPo v1.0.3	Kundu et al.[55]	https://github.com/kensung-lab/hypo
BlobTools2, BlobToolKit v2.5.0	Challis et al.[56]	https://github.com/blobtoolkit/blobtoolkit
Mapping pipeline A160156_v02	Arima Genomics	https://github.com/ArimaGenomics/mapping_pipeline
YaHS v1.2a.2	Zhou et al.[57]	https://github.com/c-zhou/yahs
JuiceBox	Robinson et al.[58]	https://aidenlab.org/juicebox
BUSCO v5.4.5	Simão et al. and Manni et al.[7,59]	https://busco.ezlab.org
assembly-stats v17.02	Challis R.[60]	https://github.com/rjchallis/assembly-stats
RepeatModeler2 v2.0.2	Flynn et al.[61]	https://github.com/Dfam-consortium/RepeatModeler
RepeatMasker v.4.1.2	Smit et al.[62]	https://github.com/rmhubley/RepeatMasker
HISAT2 v2.1.0	Kim et al.[63]	https://github.com/DaehwanKimLab/hisat2
BRAKER2 v2.1.6	Brůna et al.; Stanke et al.; Hoff et al.; Stanke et al.; Buchfink et al. and Lomsadze et al.[64–69]	https://github.com/Gaius-Augustus/BRAKER
eggNOG-mapper v.2.1.7	Cantalapiedra et al. and Huerta-Cepas et al.[70,71]	https://github.com/eggnogdb/eggnog-mapper
OrthoFinder v2.5.4	Emms and Kelly[72]	https://github.com/davidemms/OrthoFinder
odp v0.3.0	Schultz et al.[73]	https://github.com/conchoecia/odp
OrthoDB v10 database	Kriventseva et al.[74]	https://v10.orthodb.org
clusterProfiler	Yu et al. and Wu et al.[75,76]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
R environment	R Core Team[77]	https://www.r-project.org
Custom bash, R, and python scripts	This paper	https://github.com/tauanajc/Cunha_etal_2023_CurrBiol; https://doi.org/10.6084/m9.figshare.23419922

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to Tauana J. Cunha (tcunha@fieldmuseum.org).

Materials availability

Specimens of *Acutogordius australiensis* and *Nectonema munidae* are deposited at the Museum of Comparative Zoology (https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:152393, https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:153622).

Data and code availability

- Nanopore and Illumina raw data are deposited in NCBI BioProject PRJNA983812, under the umbrella of the Global Invertebrate Genomics Alliance (GIGA) BioProject PRJNA649812. Assemblies and annotation files are available in Figshare (https://doi.org/10.6084/m9.figshare.23419922).
- A pipeline with all original code and scripts is provided as a repository in GitHub (https://github.com/tauanajc/Cunha_etal_2023_CurrBiol), as well as in Figshare (https://doi.org/10.6084/m9.figshare.23419922).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Please cite this article in press as: Cunha et al., Rampant loss of universal metazoan genes revealed by a chromosome-level genome assembly of the parasitic Nematomorpha, Current Biology (2023), https://doi.org/10.1016/j.cub.2023.07.003

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Specimen data

Specimens of *Acutogordius australiensis* and *Nectonema munidae* are deposited at the Museum of Comparative Zoology (https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:152393, https://mczbase.mcz.harvard.edu/guid/MCZ:IZ:153622; Table S1). *Nectonema munidae* was collected from inside the squat lobster host *Munida* sp. in 2010, in a well-known locality for this species in Norway. *Acutogordius australiensis* was so far known from a single male collected in Queensland, Australia.[78] Our specimens, collected in 2019 in New South Wales, match the original illustrations and the English diagnosis provided in a literature review of the genus.[79] The posterior end of the body has a broad post-cloacal crescent, semicircular and not extending over the base of the tail lobes, and with abundant cone-shaped spines in the region posterior to the crescent (Figure 1B). The tail region of a male was cut and mounted on bi-adhesive carbon tape on a SEM stub. The specimen was coated with 20 nm of Pt–Pd (80:20) in a HAR 050 EMS 300T D dual head sputter coater at the Center for Nanoscale Systems, Harvard University. Specimens were then imaged using a Zeiss FESEM Ultra Plus using an SE2 detector with an EHT target of 5 kV. All images are accessible through the voucher link.

METHOD DETAILS

DNA extraction and sequencing

For each specimen, high-molecular-weight DNA was extracted from a tissue clip using a high salt protocol from the Donnan Laboratories, University of Liverpool (the DNA pellet of *Acutogordius* was eluted overnight). The DNA was then repaired with the PreCR Repair Mix (New England Biolabs), and cleaned off contaminants with a chloroform cleanup.[80,81] The quantity and quality of resuspended DNA were measured with a Qubit fluorometer and a NanoDrop spectrophotometer (Thermo Fisher Scientific), and the fragment size distribution was assessed with a TapeStation Genomic DNA ScreenTape (Agilent Technologies) (Figure S3).

For short-read sequencing, a whole-genome shotgun library was prepared using the KAPA HyperPlus library preparation kit (Roche), starting with about 100 ng of input DNA. For *Acutogordius* only, a second library was prepared using the Arima Hi-C kit (Arima Genomics) and the Kapa HyperPrep kit (Roche), starting from a flash-frozen tissue clip of the same specimen. Libraries were dual-indexed and sequenced for paired ends of 150 bp in an Illumina NovaSeq S4 flow cell, pooled with unrelated libraries at the Bauer Core Facility at the Faculty of Arts and Sciences (FAS), Harvard University.

For long-read sequencing, libraries were prepared following the Genomic DNA by Ligation protocol from Oxford Nanopore Technologies (SQK-LSK109, v.14Aug2019), with the following modifications: increased incubation times for the DNA repair and end-prep step (30 min at each temperature); skipping the first bead cleanup; doubling the incubation times for the second bead cleanup. Two MinION flow cells were used for *Acutogordius*, and one for *Nectonema*. Whenever possible, we started library preparation with more than 1 μg of input DNA, so that the final library was concentrated enough to be split into two volumes. The second aliquot was loaded in the same flow cell in the middle of the sequencing run after a wash (ONT protocol EXP-WSH003, v.18Sep2019) to increase yield. Run statistics were visualized with NanoPlot v1.29.0[49] (Figure S4).

Estimation of genome size

Genome size and heterozygosity were estimated with a k-mer approach using Jellyfish v2.2.5[48] and GenomeScope.[10]

Genome assembly

ONT long-reads were basecalled with Guppy v4.5.2 and assembled with Flye v2.8.1.[50] Long reads were then mapped back to the assembly with BWA v0.7.17[51] and used to polish the assembly with Racon v1.4.21[52] and Medaka v1.3.2. Haplotigs and contig overlaps were removed with purge_dups v1.2.5,[53] and the assembly was polished with HyPo v1.0.3[55] using Illumina short-reads preprocessed with TrimGalore.[54] Contaminant contigs in the assembly were identified and removed with BlobTools2 in the BlobToolKit v2.5.0.[56,82] For the assembly of *Acutogordius*, we further mapped Hi-C reads to the draft assembly using Arima Genomics' pipeline A160156_v02 (https://github.com/ArimaGenomics/mapping_pipeline), followed by scaffolding with YaHS v1.2a.2.[57] The contact map was visualized in JuiceBox.[58,83] Completeness of the assemblies was evaluated with BUSCO v5.4.5 by comparison with the metazoa_odb10 database.[7] Snailplots for assembly statistics were plotted with assembly-stats v17.02.[60]

RNA sequencing

For *Acutogordius*, RNA was extracted from a second specimen with the TRIzol Reagent (Invitrogen). A library was prepared with the Kapa mRNA HyperPrep kit (Roche) and sequenced for 150 bp paired-end reads in an Illumina NovaSeq S4 flow cell at the Bauer Core Facility at Harvard University. We used TrimGalore[54] to trim raw reads of *Acutogordius* and of the published RNA-seq data of *Nectonema munidae* (SRX5417513[47]), derived from the same specimen used here for genome sequencing.

Genome annotation

For each species, a library of repetitive elements was built using RepeatModeler2 v2.0.2,[61] and repeats were masked in the respective assemblies using RepeatMasker v.4.1.2.[62] Cleaned RNA-seq reads were mapped to the assemblies with HISAT2 v2.1.0[63] and used to inform gene predictions during annotation of the masked assemblies with BRAKER2 v2.1.6.[64–69] Functional annotations were added to the final protein predictions using eggNOG-mapper v.2.1.7,[70,71] with settings to report only alignments equal or above an

identity threshold of 40%, and query and subject coverage fraction thresholds of 20%. To identify genes related to sensory organs, we screened the eggNOG output for the terms 'opsin', 'GPCR', '7 transmembrane receptor', 'olfactory', and 'visual'.

Comparative analyses with other protostomes

We compared the two new nematomorph genomes with other representative parasitic and free-living taxa, focusing on the distribution of non-coding regions in the genome, and on lineage-specific and shared orthologous genes. The following assemblies were downloaded from NCBI (unless otherwise noted): nematodes *Caenorhabditis elegans* (GCF_000002985.6), *Pristionchus pacificus* (GCA_000180635.4), *Angiostrongylus cantonensis* (GCA_009735665.1), and *Loa loa* (GCF_000183805.2); arthropod *Drosophila melanogaster* (GCA_000001215.4), platyhelminthes *Schmidtea mediterranea* (WormBase ParaSite: SmedGD_v1.3) and *Schistosoma haematobium* (GCF_000699445.3); orthonectid *Intoshia linei* (GCA_001642005.1). Orthologs were identified with OrthoFinder v2.5.4[72] using the fasta file of annotated proteins of each species (Figure S1). Introns and intergenic distances were calculated as the distance between exons or between transcripts for the longest isoform of each gene from GTF annotation files.

Synteny

Orthologs belonging to ancestral linkage groups (ALGs) were identified in the assembly of *Acutogordius australiensis* and visualized in Oxford dot plots with odp v0.3.0.[73] Orthologous relationships were determined by odp with reciprocal best diamond blastp matches, and significant interactions between chromosomes calculated with Fisher's exact test (opaque dots in the plot representing p values ≤ 0.05). To create the required .chrom input file of protein locations, we first used a custom script (gtf_to_gff.py) to convert our GTF annotation file to a GFF file. We then slightly modified the script NCBIgff2chrom.py from odp to process our file instead of those derived from NCBI.

Functional enrichment

The list of 954 BUSCO gene IDs from the metazoa_odb10 database[59] was used to obtain specific and ancestral Gene Ontology (GO) terms from the OrthoDB v10 database[74] (https://v10.orthodb.org/, accessed on February 14, 2023) with a custom python script (buscos_to_GO.py). We then performed a hypergeometric enrichment analysis using the 220 genes missing from both *Acutogordius australiensis* and *Nectonema munidae* as the gene set of interest, and the entire metazoan dataset as the background set of genes. The over-representation analysis was done for each of the three GO categories (Biological Process, Cellular Component, Molecular Function) with the *clusterProfiler* package[75,76] in R.[77]