Report

Current Biology

Extreme Genome and Nervous System Streamlining in the Invertebrate Parasite *Intoshia variabili*

Highlights

- *Intoshia variabili* has an extremely simplified nervous system
- *Intoshia variabili* has the smallest known metazoan genome
- Genome of *I. variabili* could be considered a model genome for a minimal Bilateria
- Reduction and simplification are the dominant modes of evolution in orthonectids

Authors

George S. Slyusarev,
Viktor V. Starunov,
Anton S. Bondarenko,
Natalia A. Zorina,
Natalya I. Bondarenko

Correspondence

n.bondarenko@spbu.ru

In Brief

Slyusarev et al. report on the morphology, genome structure, and content of the smallest Orthonectida species *Intoshia variabili*. This orthonectid with an extremely simplified nervous system has the smallest known metazoan genome, with one of the fewest numbers of genes reported.

Slyusarev et al., 2020, Current Biology *30*, 1292–1298
April 6, 2020 © 2020 Elsevier Ltd.
https://doi.org/10.1016/j.cub.2020.01.061

CellPress

CellPress

Current Biology
Report

Extreme Genome and Nervous System Streamlining in the Invertebrate Parasite *Intoshia variabili*

George S. Slyusarev,[1] Viktor V. Starunov,[1,2] Anton S. Bondarenko,[3] Natalia A. Zorina,[1] and Natalya I. Bondarenko[1,4,*]
[1]Department of Invertebrate Zoology, Faculty of Biology, St. Petersburg State University, Universitetskaya nab. 7/9, 199034 St. Petersburg, Russia
[2]Zoological Institute RAS, Universitetskaya nab. 1, 199034 St. Petersburg, Russia
[3]Faculty of Physics, Saint-Petersburg State University, Universitetskaya nab. 7/9, 199034 St. Petersburg, Russia
[4]Lead Contact
*Correspondence: n.bondarenko@spbu.ru
https://doi.org/10.1016/j.cub.2020.01.061

SUMMARY

Orthonectida is a small, rare, and in many aspects enigmatic group of organisms with a unique life cycle and a highly simplified adult free-living stage parasitizing various marine invertebrates [1, 2]. Phylogenetic relationships of Orthonectida have remained controversial for a long time. According to recent data, they are close to Annelida, specifically to Clitellata [3–5]. Several studies have shown that parasitism can not only lead to a dramatic reduction of the body plan and morphological structures but also affect organisms at the genomic level [6, 7]. Comparative studies of parasites and closely related non-parasitic species could clarify the genome reduction degree and evolution of parasitism. Here, we report on the morphology, genome structure, and content of the smallest known Orthonectida species *Intoshia variabili*, inhabiting the flatworm *Graffiellus croceus*. This orthonectid with an extremely simplified nervous system demonstrates the smallest known genome (15.3 Mbp) and one of the lowest reported so far gene numbers (5,120 protein-coding genes) among metazoans. The genome is extremely compact, due to a significant reduction of gene number, intergenic regions, intron length, and repetitive elements. The small genome size is probably a result of extreme genome reduction due to their parasitic lifestyle, as well as of simplification and miniaturization of the free-living stages. Our data could provide further insights into the evolution of parasitism and could help to define a minimal bilaterian gene set.

RESULTS AND DISCUSSION

Morphological Characteristics of *Intoshia variabili* Free-Living Generation

Intoshia variabili adult free-living stages have a highly simplified body plan. Female *I. variabili* has an oval body with a tapering anterior end (Figures 1A–1D and 2B). It is 80 μm long and 20 μm across at the midbody. The male has a cylindrical body about 35 μm long and 10 μm across (Figure 2C). The body of the free-living stage comprises a monolayered epithelium, muscle, nerve, and germ cells. The muscular system is made of 16 and 32 cells in males and females, respectively (Figures 2B and 2E). The longitudinal muscles are represented by longitudinal lateral muscle fibers and a dorsal and a ventral fiber running from the anterior end of the body up to the posterior. At the anterior end of the body, the lateral fibers bifurcate, coming in contact with ciliary cells as described elsewhere [8]. The circular muscle cells lie under the longitudinal fibers and form a compact group at the midbody. Orthonectid locomotion is mostly ciliary, although the muscular system, in our opinion, is related to reproductive function. First, during copulation, the musculature contraction in the male injects sperm into the female. After copulation, the contraction of the circular muscles allows larvae to egress from the female (Figure S1D).

Our cell counts suggest that there are about only 2 nerve cells in the male and 4 in the female (Figures S1F and S1G). In the male *I. variabili*, the antibody labeling against serotonin reveals only two neurons, located in the anterior third part of the body, on its dorsal side (Figures 2C and 2D). The neurons are situated in between the epithelial and the muscle cells and are in direct contact with them. Both cells are unipolar, and their processes run backward along the longitudinal muscles, nearly reaching the end of the body. No FMRF-amide-positive cells have ever been revealed with the antibody treatment. Surprisingly, the acetylated α-tubulin antibody labeling revealed only the ciliation and sperm flagella but never the nerve cells or their fibers (Figure S1G). The nervous system of the female consists of four serotoninergic cells and a receptor made of three cells (Figure 2) [9]. All of them are serotonin immunoreactive. The cells are multipolar, each anterior cell has a process running toward the frontal end of the body and each posterior cell a process running back along the longitudinal muscle fibers. All 4 cells are interconnected.

In another studied orthonectid, *I. linei*, the nervous system of both males and females comprises 6 serotonin-immunopositive cells. Thus, considering the smaller size of male and female *I. variabili*, we can suggest the reduction of the nervous system cell number due to miniaturization.

General Characteristics of the Smallest Genome among Metazoa

The genome of *I. variabili* was sequenced at 300-fold coverage and assembled into 15.3 Mbp on 239 scaffolds (Table 1). *Intoshia*

Table 1. Features of *Intoshia variabili* Genome and Comparison with Those of *Intoshia linei*

	Intoshia variabili	*Intoshia linei*
Size (Mbp)	15.3	41.6
Scaffolds	239	3,520
N50 (kbp)	227,919	26,268
GC%	27.9	25.9
Predicted protein-coding genes	5,120	8,724
Gene density (per Mb)	335	202
Protein-coding region content %	78.6	53.8
Mean gene length (bp)	2,343	3,043
Mean exon length (bp)	290	187
Mean length of intergenic region (bp)	620	1,549
Intron density (introns per gene)	5.11	5.07
Most common intron length	43	37
Mean intron length (bp)	111	342
Repetitive content %	5.26	27.7

Data for *I. linei* were obtained from NCBI: PRJNA316116.

variabili has the smallest genome among Metazoa described so far; its size is less than the genome of the plant-parasitic nematode *Pratylenchus coffeae* (19.7 Mbp; 6,712 genes), myxozoa *Kudoa iwatai* (22.5 Mbp; 5,533 genes), and orthonectids *I. linei* (43 Mbp; 8,724 genes) [6, 10, 11]. The genome of *I. variabili* has a small number of poorly represented repetitive elements, only 827,534 bp or 5.26% of the total assembly size relative to that reported for other metazoan genomes [12–15] (Table S1). The 19 distinct retrotransposon sequences representing only one family (long terminal repeats [LTRs]) were identified. The main content of the genome-repetitive elements accounted for simple repeats (2.4% of the genome) and unclassified ones (2% of the genome; Table S1). These values are less than the corresponding ones within *I. linei* genome [11].

The genome of *I. variabili* is predicted to encode 5,120 genes—one of the smallest reported gene numbers among metazoans and the second after the recently sequenced genome of *Dicyema japonicum* (5,021 genes) [6, 10, 11, 16]. The predicted genes with a mean length of 2.3 kb are intron rich in total account for 19% of the genome and contain an average of six exons. The introns tend to be reduced. Comparison of gene structure in two Orthonectida species reveals a downward trend in mean intron lengths from 342 bp in *I. linei* to 111 bp in *I. variabili* genes (Table 1). This is in good agreement with the observations on the correlation of intron size with genome size [17]. *Intoshia variabili* has one of the most gene-dense metazoan genomes currently known, where genes cover 78.6% of the genome at an average density of 335 genes per Mb and the median intergenic distance is 620 bp (Table 1) [6, 11, 18].

To identify gene gain and loss in *I. variabili*, orthologous relationships were predicted between orthonectids and 30 other species of eukaryotes. According to the analysis, *I. variabili* has 4,038 orthologous groups with 4,225 genes in common with all analyzed species. Among Annelida, *I. variabili* has 2,945 orthologous groups in common with *Helobdella robusta*, 3,145 with *Capitella teleta* and 3,644 with *I. linei* (Figure 3A). Comparative analyses revealed constancy in gene content between two Orthonectida species. The 711 orthologous groups are shared between Orthonectida species and are specific only for this group. The 849 and 2,591 taxonomically restricted or "orphan" protein-coding genes (PCGs) for *I. variabili* and *I. linei*, respectively, did not cluster with any other PCGs in the dataset, suggesting that they cannot be linked to their orthologs by sequence similarity alone. They represent putative species-specific genes that emerged in one of the two analyzed genomes, and because orthonectids diverged, these genes have quickly evolved and have undergone significant reduction within *I. variabili* genome. Of the entire *I. variabili* gene set, 2,835 genes have an ortholog (BLASTp e-value cutoff $\leq 10^{-5}$) belonging to one of the 279 known biological pathways (Kyoto Encyclopedia of Genes and Genomes [KEGG] database). The high-confidence gene set of *I. variabili* reveals an extensive reduction of the number of genes in overall key pathways (Figures 3B–3E).

Thus, the genome of *I. variabili* has been reduced on an unprecedented scale and has become extremely compact, predominantly due to reduction of the intergenic distances, short intronic sizes, elimination of repetitive elements, and extensive reduction of the gene number. As a result, gene density in this species became much higher than in *I. linei* and other metazoans. All our findings are consistent with the statement that genome size variation in eukaryotes is determined by the proportion of mobile elements, number of introns, and complexity of regulatory regions [19].

The Nervous System, Axon Guidance Molecules, and Their Receptors

Among Orthonectida, *I. variabili* possesses the simplest nervous system. This simplification is associated with significant gene losses. Thus, *I. variabili* genome lost half of the known genes related to neuronal development, axon guidance, synapse formation, and receptor diversity (Table S2). *Intoshia variabili* lost early neural patterning genes belonging to homeobox gene families, such as Pax3/7, Gbx, Hbn, and Rax. Genes potentially involved in the nervous system development, such as netrin, cadherins, integrins, IgSF-CAMs, ephrins, and ephrin receptors are present in *I. linei*, but ephrins and IgSF-CAMs were not identified in *I. variabili*. According to the Pfam analysis and reciprocal BLAST searches, only 29 hits of potassium, tetrameric sodium, and calcium ion channels were found in *I. variabili* although 42 hits in *I. linei* [11]. Also, 12 hits correspond to the innexin/pannexin-specific Pfam domain, which can function as electrical synapsis in excitable tissues. We found 9 genes for neurotransmitter-gated ion channels in *I. variabili*, which is also less than in *I. linei* (Table S3) [11]. Neuropeptide systems comprise only three peptide-signaling molecules (pedal-peptide-2, diazepam-binding inhibitor, and nicotinamide phosphoribosyltransferase) that operate within a structurally simplest nervous system of 2–4 neurons. This significant gene loss is in good accordance with the high degree of morphological simplification and reduction of the nervous system to several cells.

Homeobox and Other Regulatory Genes

Among Orthonectids, transcription factor family genes have undergone significant reduction. Similar to other parasitic organisms, *I. variabili* lost genes from 65 homeobox gene families

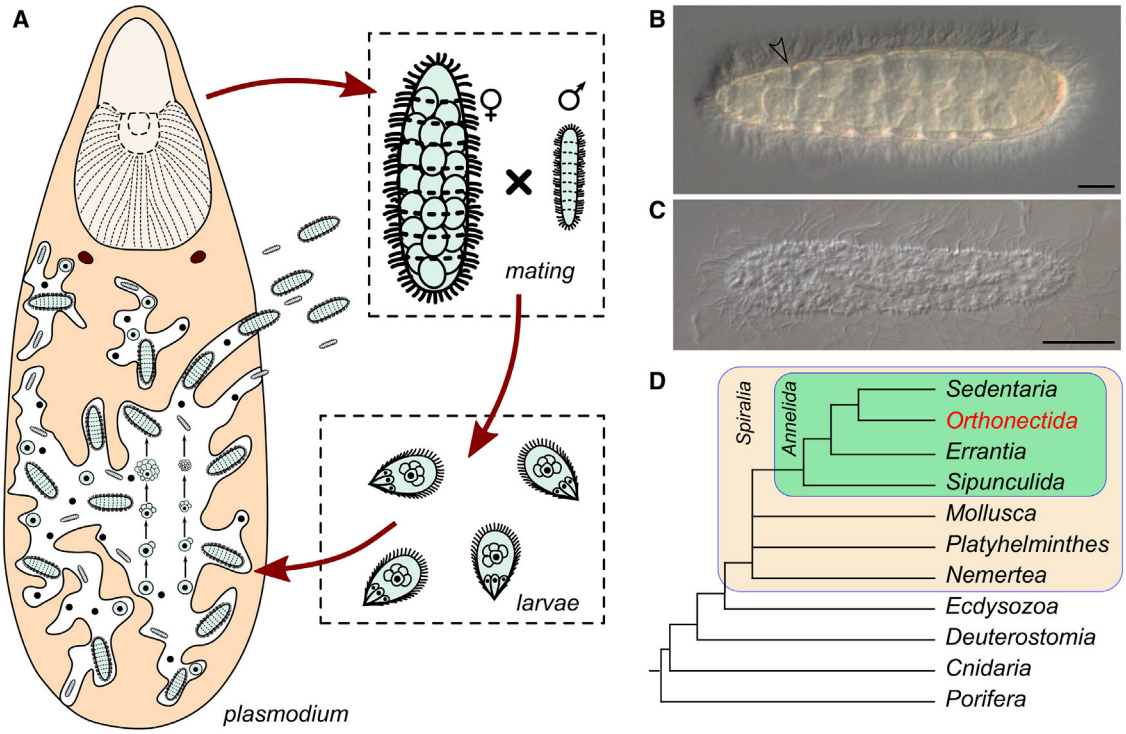

Figure 1. Life Cycle and Phylogeny of Orthonectida

(A) Schematic diagram of *I. variabili* life cycle with parasitic plasmodium stage and free-living adults and larvae.

(B and C) Adult free-living stages: (B) female and (C) male. Arrowhead labels the position of the cerebral ganglion in female. Scale bars, 10 μm.

(D) Phylogenetic position of Orthonectida as a member of Annelida, according to [4].

(Table S4) [20, 21]. For example, two out of the three ParaHox genes and at least ten Hox genes have been lost. The remaining three Hox genes are located on three different scaffolds, suggesting that they are not linked together in a single cluster. Moreover, repetitive elements occur around *I. variabili* Hox and ParaHox genes, whereas they are typically absent within Hox and ParaHox clusters, which can probably lead to genome rearrangements by facilitating non-homologous recombination [22]. Disorganized Hox cluster structure is also inherent not only for parasites, such as orthonectids, dicyemids, and tapeworms, but also for free-living species as a result of adaptation to an extreme habitat [11, 16, 20, 23]. Within ParaHox genes involved in the specification of a through-gut, *I. variabili* lost *gsx* and *pdx* genes while keeping *cdx*. Although the *cdx* gene was preserved in both species, Orthonectida lacks any gut [24].

The highly conserved Pax gene family is represented by only 3 out of 9 genes belonging to Pax2/5/8 and Pax4/6 subfamilies. All other key bilaterian genes for transcription factors and signaling molecule families are also poorly represented within *I. variabili* genome (Table S4). Two out of six major and universal signal transduction pathways in the cell, such as Wnt and Notch, have undergone a significant reduction of a large number of genes, whereas Hedgehog, steroid hormone receptor, and tyrosine kinase receptor pathways were not found within the genome of *I. variabili*. Together, these pathways provide basic machinery for cell fate transitions and underlie embryonic development, and their reduction has led to the highly simplified body organization of the free-living *I. variabili* stages.

Regulatory genes responsible for stem and germ cell differentiation, belonging to the *piwi* and *vasa* subfamilies, were also not detected in *I. variabili* genome. Despite this fact, orthonectids produce germ cells that undergo meiosis, followed by cellular differentiation into oocytes and sperm [2]. Other key genes belonging to the microRNA pathway and essential for microRNA (miRNA) biogenesis, post-transcriptional regulation of mRNAs, and therefore for correct development, such as one Argonaute gene from AGO family, Dicer, Drosha, and Pasha genes, were found.

Modification of the Metabolism

The high-confidence gene sets reveal extensive reductions in overall metabolic capability and an increased ability to absorb nutrients compared to that of other parasites (Figures 3B–3E) [11, 20]. The genome of *I. variabili* illustrates the absence of most genes encoding key enzymes for fatty acid biosynthesis. Only acetyl-coenzyme A (CoA) carboxylase, an enzyme that can be involved in fatty acid precursor production, was detected, wherein the set of genes encoding fatty acid elongation and degradation is largely complete. Moreover, genes coding for fatty acid binding and transport proteins were identified. Altogether, this suggests possible uptake of fatty acids from the host. *I. variabili* also contains a poor set of genes, required for glycolysis, tricarboxylic acid cycle, and the pentose phosphate pathway, which provides evidence for the alteration in the metabolic pathways of the parasite. This finding conforms to the data on the drastic changes in metabolic pathways, including the loss of peroxisomes in tapeworms and trematodes [25, 26].

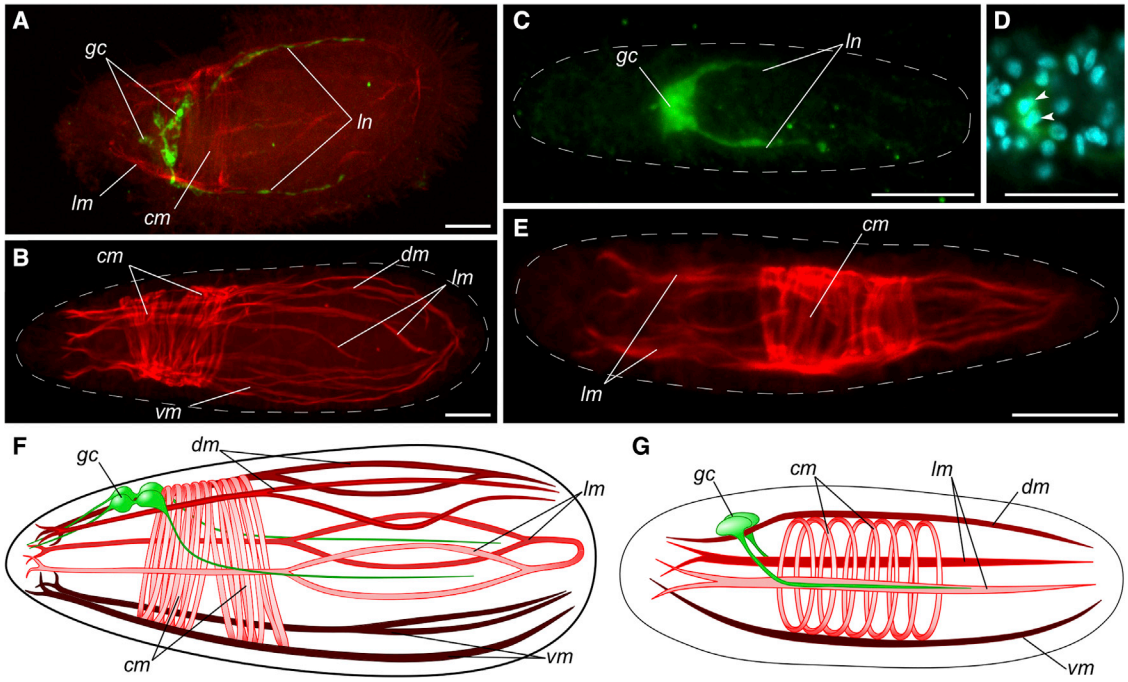

Figure 2. Muscular and Nervous Systems of *Intoshia variabili*, Confocal Laser Scanning Microscopy (CLSM)

(A) A female's muscular and nervous systems, dorsal view.

(B) A female's muscular system, lateral view.

(C) A male's serotoninergic nervous system.

(D) A male's nuclei revealed with DAPI staining; arrowheads point to the nuclei of the two nerve cells.

(E) A male's general view of the muscular system revealed by tetramethylrhodamine (TRITC)-phalloidin staining.

(F and G) Schemes of muscular and nervous systems in *Intoshia variabili*.

(F) A female, lateral view.

(G) A male, lateral view.

cm, circular muscles; dm, dorsal muscle; gc, ganglion cells; lm, longitudinal muscles; ln, lateral nerve; vm, ventral muscle. Scale bars, 10 μm.

Orthonectids have lost all marker protein genes PEX3, PEX10, PEX12, and PEX19 associated with the peroxisome [27]. They lack peroxisomes altogether, as seen in several other parasites [20, 28].

Basement Membrane

Intoshia variabili has cell-cell junctions between cells but lacks an underlying basement membrane or any described extracellular matrix [8, 24, 29]. Nidogen, fibronectin, fibrin, elastin, vitronectin, heparan sulfate proteoglycans, and matrilin-2-like genes are absent in the genome. However, the genome contains a reduced set of genes encoding putative extracellular matrix (ECM) proteins, such as collagen IV and 2 types of laminin. *I. variabili* genome encodes cell adhesion molecules belonging to 4 families, cytoskeletal linker proteins (paxillin, talin, and α- and β-catenin) and focal adhesion kinase. Genes encoding proteins that can modify ECM components and signaling molecules in the matrix, such as lysyl oxidases and the ADAM metalloproteinases (including the TACE family), are also present.

Thus, analyzing the genome of *I. variabili*, we observe a drastic reduction of genes in the main pathways (Figures 3B–3E). In most cases of gene reduction, orthonectids eliminate pathway-specific genes, which are useless in a parasitic lifestyle, while keeping genes with fundamental functions involved in multiple pathways. It is not clear how orthonectids overcome the physiological gaps of lost genes. Two key questions emerge from these data: what accounts for the simplicity of organization and such small genome size and are these features secondarily derived? Because it has been shown that Orthonectida are closely related to Annelida, it should be assumed that the small genome size is a result of extreme genome reduction, due to their transfer to the parasitic lifestyle that is similar to cestodes, parasitic nematodes, and myxozoans, as well as the miniaturization and extreme simplification of free-living sexual generation [3, 4, 6, 10, 15, 26, 30, 31].

STAR★METHODS

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

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Material collection
- METHOD DETAILS
  - Immunocytochemistry and confocal microscopy
  - Morphology and cell count of I. variabili male

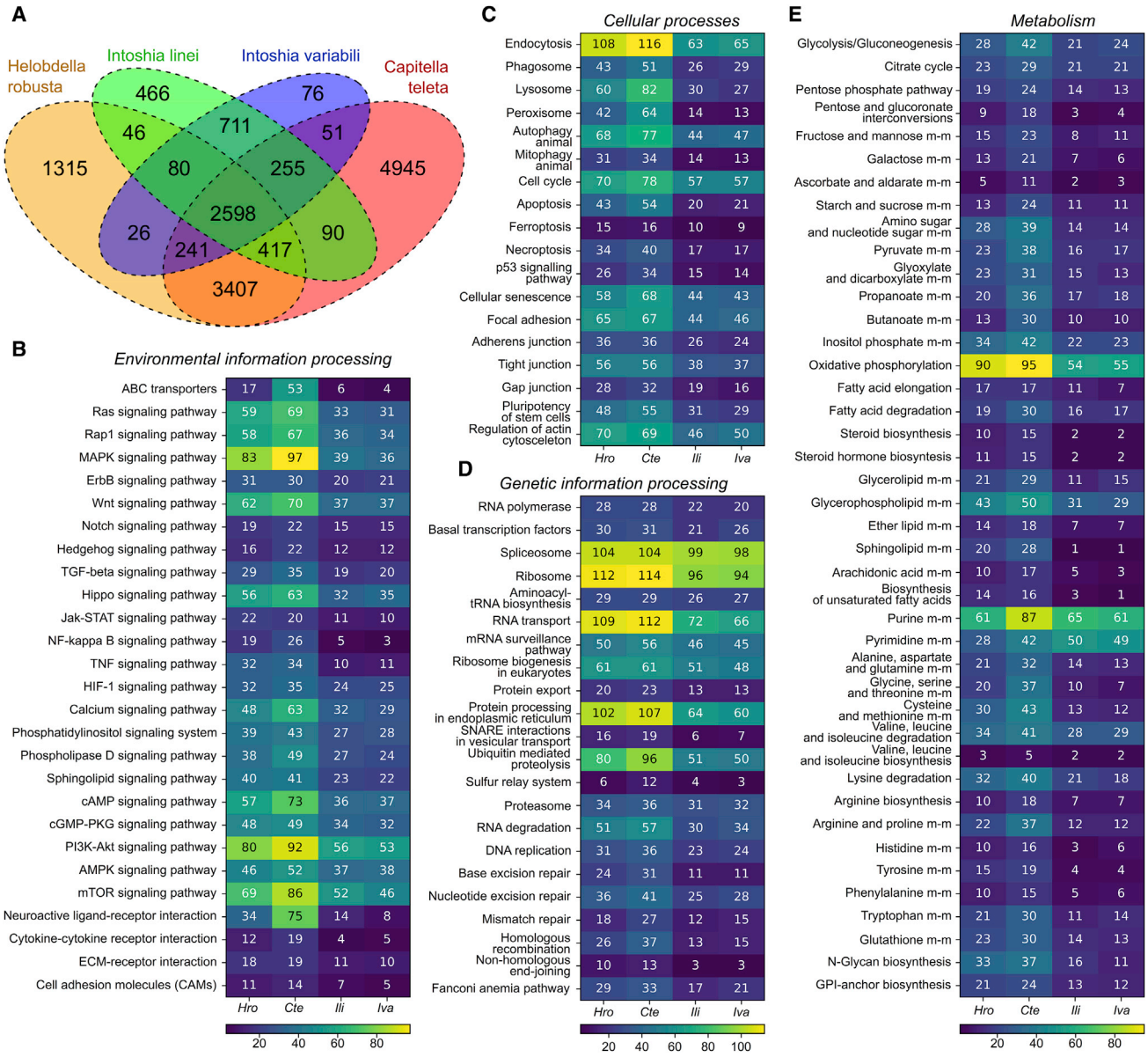

Figure 3. Venn Diagram and Heatmap Summarizing Annotation of *Intoshia variabili* Genome

(A) Venn diagram summarizing the overlapping homology between *Intoshia variabili* orthologous groups and those of *Intoshia linei* and two Annelida species. (B–E) Heatmap showing the conservation of individual metabolic pathways for sequenced Orthonectida with comparison to two Annelida species in terms of environmental information processing (B), cellular processes (C), genetic information processing (D), and metabolism (E). Each row indicates an individual metabolic pathway grouped by their superclass membership (defined by KEGG). Each column indicates species. Colored tiles indicate the level of conservation of each pathway within each species (see inset color key).

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2020.01.061.

ACKNOWLEDGMENTS

We are thankful to Dr. M. Makarov for providing accommodation and work facilities at the Barents Sea and to the staff of the Marine Biological Station of the Saint Petersburg State University for assistance in collecting material. The work was partially supported by the grant of Russian Foundation for Basic

Research (RFBR) no. 19-04-00218. Bioinformatics was supported with Russian Science Foundation grant 19-74-10013 to N.I.B. This study utilized equipment of the "Center for Microscopy and Microanalysis," "Center for Molecular and Cell Technologies," "CHROMAS," "Culture Collection of Microorganisms," "Biobank," and "Computing Centre SPbU" of Saint-Petersburg State University.

AUTHOR CONTRIBUTIONS

G.S.S. and V.V.S. provided the samples and performed confocal microscopy. N.I.B. isolated genomic DNA for sequencing. N.I.B., A.S.B., and N.A.Z. undertook the sequencing, assembly, and annotation of genomic data and planned and performed additional bioinformatic analyses. N.I.B. browsed genome. N.I.B., V.V.S., and N.A.Z. drafted and edited tables, figures, and Supplemental Information. Both N.I.B. and G.S.S. drafted and edited the manuscript and conceived and planned the project.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: September 27, 2019
Revised: November 21, 2019
Accepted: January 17, 2020
Published: February 20, 2020

REFERENCES

1. Kozloff, E.N. (1992). The genera of the phylum Orthonectida. Cah. Biol. Mar. *33*, 377–406.
2. Sliusarev, G.S. (2008). [Phylum Orthonectida: morphology, biology, and relationships to other multicellular animals]. Zh. Obshch. Biol. *69*, 403–427.
3. Schiffer, P.H., Robertson, H.E., and Telford, M.J. (2018). Orthonectids are highly degenerate annelid worms. Curr. Biol. *28*, 1970–1974.e3.
4. Bondarenko, N., Bondarenko, A., Starunov, V., and Slyusarev, G. (2019). Comparative analysis of the mitochondrial genomes of Orthonectida: insights into the evolution of an invertebrate parasite species. Mol. Genet. Genomics *294*, 715–727.
5. Zverkov, O.A., Mikhailov, K.V., Isaev, S.V., Rusin, L.Y., Popova, O.V., Logacheva, M.D., Penin, A.A., Moroz, L.L., Panchin, Y.V., Lyubetsky, V.A., and Aleoshin, V.V. (2019). Dicyemida and Orthonectida: two stories of body plan simplification. Front. Genet. *10*, 443.
6. Burke, M., Scholl, E.H., Bird, D.M., Schaff, J.E., Coleman, 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.
7. Wang, J., Gao, S., Mostovoy, Y., Kang, Y., Zagoskin, M., Sun, Y., Zhang, B., White, L.K., Easton, A., Nutman, T.B., et al. (2017). Comparative genome analysis of programmed DNA elimination in nematodes. Genome Res. *27*, 2001–2014.
8. Slyusarev, G.S., and Kristensen, R.M. (2003). Fine structure of the ciliated cells and ciliary rootlets of *Intoshia variabili* (Orthonectida). Zoomorphology *122*, 33–39.
9. Slyusarev, G.S. (1994). Fine structure of the female *Intoshia variabili* (Alexandrov & Sljusarev) (Mesozoa: Orthonectida). Acta Zoologica *75*, 311–321.
10. 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.
11. 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.
12. C. elegans Sequencing Consortium (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012–2018.
13. Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The genome sequence of *Drosophila melanogaster*. Science *287*, 2185–2195.
14. Jex, A.R., Liu, S., Li, B., Young, N.D., Hall, R.S., Li, Y., Yang, L., Zeng, N., Xu, X., Xiong, Z., et al. (2011). *Ascaris suum* draft genome. Nature *479*, 529–533.
15. Mitreva, M., Jasmer, D.P., Zarlenga, D.S., Wang, Z., Abubucker, S., Martin, J., Taylor, C.M., Yin, Y., Fulton, L., Minx, P., et al. (2011). The draft genome of the parasitic nematode *Trichinella spiralis*. Nat. Genet. *43*, 228–235.
16. Lu, T.-M., Kanda, M., Furuya, H., and Satoh, N. (2019). Dicyemid mesozoans: a unique parasitic lifestyle and a reduced genome. Genome Biol. Evol. *11*, 2232–2243.
17. Zhang, Q., and Edwards, S.V. (2012). The evolution mechanism of intron length size in amniotes: a role for powered flight? Genome Biol. Evol. *4*, 1033–1043.
18. Fernandez-Valverde, S.L., and Degnan, B.M. (2016). Bilaterian-like promoters in the highly compact *Amphimedon queenslandica* genome. Sci. Rep. *6*, 22496.
19. Lynch, M. (2007). Origins of Genome Architecture (Sinauer Associates).
20. Tsai, I.J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sánchez-Flores, A., Brooks, K.L., Tracey, A., Bobes, R.J., Fragoso, G., Sciutto, E., et al.; Taenia solium Genome Consortium (2013). The genomes of four tapeworm species reveal adaptations to parasitism. Nature *496*, 57–63.
21. Hahn, C., Fromm, B., and Bachmann, L. (2014). Comparative genomics of flatworms (platyhelminthes) reveals shared genomic features of ecto- and endoparastic neodermata. Genome Biol. Evol. *6*, 1105–1117.
22. Simmen, M.W., Leitgeb, S., Charlton, J., Jones, S.J., Harris, B.R., Clark, V.H., and Bird, A. (1999). Nonmethylated transposable elements and methylated genes in a chordate genome. Science *283*, 1164–1167.
23. Malmstrøm, M., Britz, R., Matschiner, M., Tørresen, O.K., Hadiaty, R.K., Yaakob, N., Tan, H.H., Jakobsen, K.S., Salzburger, W., and Rüber, L. (2018). The most developmentally truncated fishes show extensive Hox gene loss and miniaturized genomes. Genome Biol. Evol. *10*, 1088–1103.
24. Kozloff, E.N. (1969). Morphology of the orthonectid *Rhopalura ophiocomae*. J. Parasitol. *55*, 171–195.
25. Zarowiecki, M., and Berriman, M. (2015). What helminth genomes have taught us about parasite evolution. Parasitology *142* (*Suppl 1*), S85–S97.
26. International Helminth Genomes Consortium (2019). Comparative genomics of the major parasitic worms. Nat. Genet. *51*, 163–174.
27. Schlüter, A., Fourcade, S., Ripp, R., Mandel, J.L., Poch, O., and Pujol, A. (2006). The evolutionary origin of peroxisomes: an ER-peroxisome connection. Mol. Biol. Evol. *23*, 838–845.
28. Kaasch, A.J., and Joiner, K.A. (2000). Targeting and subcellular localization of *Toxoplasma gondii* catalase. Identification of peroxisomes in an apicomplexan parasite. J. Biol. Chem. *275*, 1112–1118.
29. Kozloff, E.N. (1971). Morphology of the orthonectid *Ciliocincta sabellariae*. J. Parasitol. *57*, 585–597.
30. Sommer, R.J., and Streit, A. (2011). Comparative genetics and genomics of nematodes: genome structure, development, and lifestyle. Annu. Rev. Genet. *45*, 1–20.
31. Kikuchi, T., Eves-van den Akker, S., and Jones, J.T. (2017). Genome evolution of plant-parasitic nematodes. Annu. Rev. Phytopathol. *55*, 333–354.
32. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.
33. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012).

SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.

34. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics *29*, 1072–1075.
35. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.
36. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map (SAM) format and SAMtools. Bioinformatics *25*, 2078–2079.
37. Parra, G., Bradnam, K., and Korf, I. (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics *23*, 1061–1067.
38. Smit, A.F.A., and Hubley, R. (2008–2015). RepeatModeler Open-1.0. http://www.repeatmasker.org.
39. Bao, Z., and Eddy, S.R. (2002). Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res. *12*, 1269–1276.
40. Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21* (*Suppl 1*), i351–i358.
41. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.
42. Stanke, M., Steinkamp, R., Waack, S., and Morgenstern, B. (2004). AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. *32*, W309–W312.
43. Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.
44. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
45. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240.
46. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.
47. Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35*, W182–W185.
48. Alexandrov, K.E., and Sljusarev, G.S. (1992). A new species of orthonectids, *Rhopalura variabeli* sp.n. (Mesozoa) from the turbellarian *Macrorynchus crocea*. Parazitologiia *16*, 347–351.
49. Slyusarev, G.S., Nesterenko, M.A., and Starunov, V.V. (2019). The structure of the muscular and nervous systems of the male *Intoshialinei* (Orthonectida). Acta Zoologica *100*, 451–458.
50. Slyusarev, G.S., and Manylov, O.G. (2001). General morphology of the muscle system in the female orthonectid, *Intoshia variabili* (Orthonectida). Cah. Biol. Mar. *42*, 239–242.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Monoclonal Anti-Tubulin, Acetylated antibody produced in mouse	Sigma	Cat# T-6793; RRID: AB_477585
5-HT (Serotonin) Rabbit Antibody	Immunostar	Cat# 20080; RRID: AB_572263
FMRF-amide (Cardio-excitatory Peptide) Antibody	Immunostar	Cat# 20091
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647	Invitrogen	Cat# A-31571; RRID: AB_162542
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Invitrogen	Cat# A-21206; RRID: AB_2535792
Chemicals, Peptides, and Recombinant Proteins		
paraformaldehyde (PFA)	Sigma-Aldrich	Cat# P6148
phosphate-buffered saline (PBS)	Ecoservice	Cat# B-60201
Triton X-100	Sigma-Aldrich	Cat# T8787
phalloidin	Sigma-Aldrich	Cat# P1951
Hoechst 33258	Invitrogen	Cat# H1398
DAPI	Carl Roth	Cat# 6843.1
Critical Commercial Assays		
PicoPure DNA Extraction Kit	Thermo Fisher	Cat# KIT0103
TruSeq Nano DNA Low Throughput Library Prep Kit	Illumina	Cat# 20015964
Deposited Data		
Raw sequence data	SRA	SRS4604627
Software and Algorithms		
FIJI	[32]	N/A
FastQC	N/A	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
SPAdes	[33]	http://cab.spbu.ru/software/spades/
QUAST	[34]	http://cab.spbu.ru/software/quast
BWA-MEM	[35]	http://bio-bwa.sourceforge.net
SAMtools	[36]	http://samtools.sourceforge.net
CEGMA	[37]	http://korflab.ucdavis.edu/datasets/cegma/
KrATER	N/A	https://github.com/mahajrod/KrATER
RepeatModeler	[38–41]	http://www.repeatmasker.org/RepeatModeler/
RepeatMasker	[38]	http://www.repeatmasker.org
Augustus	[42]	http://bioinf.uni-greifswald.de/augustus/
OrthoMCL	[43]	https://orthomcl.org/common/downloads/software/
BLAST	[44]	https://blast.ncbi.nlm.nih.gov/
InterProScan	[45]	https://www.ebi.ac.uk/interpro/
KEGG	[46, 47]	https://www.genome.jp/kegg/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Natalya Bondarenko (n.bondarenko@spbu.ru). This study did not generate new unique reagents.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Material collection

The orthonectid *Intoshia variabili* is a parasite of the flatworm *Graffiellus croceus* (Platyhelminthes: suborder Kalyptorhynchia). The flatworms were collected in August 2017, in the Barents Sea at the marine biological station Dalnie Zelentsi (69°07' N, 36°05' E). The flatworms were collected at the low tide by hand by placing clumps of *Fucus spp*. overgrown with various epiphytic algae in a bucket with seawater, shaking the algae vigorously in the water, and sieving the water through a fine sieve, c. 0.25 mm mesh. The infected flatworms were maintained in Petri dishes with filtered seawater at + 4°C. The emission of the free-living stages was carried out by raising the temperature of the water up to 10-15°C. A total of more than 100 free-living males and females of *I. variabili* were collected, 50 of them were used for morphological study, and about 50 were fixed in ethanol for further DNA isolation and genome sequencing.

METHOD DETAILS

Immunocytochemistry and confocal microscopy

Specimens of *I. variabili* were fixed for 8-12 hr in 4% paraformaldehyde (PFA) with 0.01 M phosphate-buffered saline (PBS) at + 4°C. After fixation, specimens were washed three times in PBS, containing 0.1% Triton X-100 (PBT) and then stored in PBS with 0.1% NaN_3 at + 4°C. Prior to immunostaining, specimens were washed in PBS with 0.1% Triton X-100 (PBT), then blocked in PBT containing 1% BSA for 2 hr and incubated in primary antibodies against acetylated α-tubulin (dilution 1:1,000 – 1:2,000) and serotonin or FMRF-amide (dilution 1:2,000) overnight at + 4°C. Afterward, specimens were washed three times in PBT for 15 min and incubated with secondary antibodies diluted 1:800 – 1:1,000 overnight at + 4°C. We used Alexa Fluor 488 Donkey Anti-Rabbit and Alexa Fluor 647 Donkey Anti-Mouse as secondary antibodies. To visualize muscular elements, after immunolabeling, the specimens were stained with TRITC conjugated phalloidin in dilution 1:100 for 2 hr. Cell nuclei were stained with Hoechst 33258 or DAPI. Stained specimens were immersed in Mowiol and mounted between two coverslips. Observations were made with Leica TCS SP5 or Leica TCS SPE laser confocal microscopes. Resulting stacks of images were processed using FIJI package [32].

Morphology and cell count of I. variabili male

DAPI staining combined with antibody treatment and confocal laser scanning microscopy (CLSM) allows for nuclei/cell quantification. Though the sperm nuclei clearly differ from all other nuclei in the male (See also Figure S1A), they hamper exact nuclei quantification. Therefore, only males that had already copulated and contain only a few sperm cells were taken into analysis (See also Figure S1B). The total cell number (sperm cells left out) is 125. The number of sperm cells is about 45-50 and varies from specimen to specimen. Epithelial ciliary and non-ciliary cells are revealed with silver impregnation and their number amounts to 101 [48]. The male has a genital pore on the ventral side of the body located in-between the epithelial cells. The cells comprising the genital pore (6 in number) are easily revealed with silver impregnation. The muscular system is made of 16 cells. Thus, the nervous system should consist of 2 cells only.

All inner space of the non-copulated male is filled with sperm cells, which form a compact mass. The sperm heads lie at the periphery of this mass, while the tails occupy its central part and are oriented along the longitudinal axis of the body (See also Figure S1C). The same orientation is typical of I. linei [49].

Morphology and cell count of I. variabili female

Cell quantification was done as in the males. The total cell number (oocytes left out) is 241 ± 2. There are 196 ± 2 epithelial cells, 6 genital pore cells and 32 muscle cells. There are only 7 cells left for the nervous system. Three of them are receptor-forming cells [9]. Therefore, there should be only 4 nervous cells.

All inner space of the female is filled with oocytes (Figures S1D and S1F), their number varies from 20 to 24. The nervous system is formed by 4 serotoninergic cells (2 anterior and 2 posterior) located in the front part of the body, (Figures 1A, 1C, and 2D–2F). The structure of the muscle system is described elsewhere, however, this description requires an amendment: the circular muscles are located under the longitudinal fibers, and not above them [50].

DNA isolation and sequencing

Total genomic DNA from 30 orthonectids was isolated using the PicoPure DNA Extraction Kit (Thermo Fisher) for single-cell isolation according to the recommendations of the manufacturer. DNA quantity assessments were performed with Qubit fluorometric quantification (Life Technologies). The genomic DNA libraries of *I. variabili* were constructed using the TruSeq library preparation protocol (Illumina) and sequenced on a HiSeq2000 sequencing system. The libraries have an estimated insert length of 326 bp. We obtained 62,8 M 100bp paired end reads. Quality control check on raw sequence data was performed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/, last accessed January 2018).

QUANTIFICATION AND STATISTICAL ANALYSIS

Genome assembly and quality evaluation

De novo assembly was implemented using SPAdes assembler with settings for Illumina paired-end reads dataset with initial k-mer of 21 and a maximum k-mer size of 99 [33]. The quality of assemblies was evaluated using QUAST [34]. The size of the assembly is 15,3 Mbp

in 239 scaffolds with an N50 of 227,919 bp. The assembly was evaluated by mapping the high-quality reads from the scaffolds by using BWA-MEM [35]. Approximately 90% of the reads could be mapped to the assembly, which covered 100% of the assembled sequences. The distribution of the sequencing depth at each position was calculated with SAMtools to assess the completeness of the assembled genome [36]. Relative completeness for *I. variavili* genome assembly was assessed using CEGMA that searches for the presence of 248 ultra-conserved CEGs [37]. The CEGMA results confirmed the homologs for 77.42% of the core eukaryotic genes in the assembly.

Estimation of genome size

Genome size was estimated by K-mer distribution analysis using KrATER (https://github.com/mahajrod/KrATER). Approximately 62 M high-quality paired-end reads (100 bp) was generated and used to determine the abundance of 23-nt K-mers.

Assessment of repeat content and annotation of non-coding RNA

The assembly was repeat masked using a custom library of *I. variabili* specific repetitive elements constructed with RepeatModeler v 1.0.8 [38–41]. Repetitive elements were classified using the 2014-01-31 RepeatMasker version of the RepBase-20170127 and Dfam_-Consensus-20170127 databases. The custom library of *I. variabili* repetitive elements created by RepeatModeler was used for masking repeats with RepeatMasker open-4.0.5. A total of 5.26% of the assembly was masked by RepeatMasker (See also Table S1) [38].

Gene prediction and protein orthology clustering

Ab initio gene prediction was carried out using Augustus v. 3.0.3 program [42]. Using the previously sequenced and annotated genome of *I. linei* as a reference, we created a training set of gene structures for Augustus and predicted putative coding regions and protein sequences for *I. variabili.* Contigs shorter than 500 bp were excluded from the annotation procedure. Annotation repeat-masked assembly returned 5,120 protein-coding genes. For comparative analyses, the predicted proteins of *I. variabili* were clustered using the OrthoMCL v2.0.9 pipeline with proteins from completely sequenced genomes of 28 Eukaryotes [43]. The proteomes were obtained from the following resources: NCBI's GenBank database (http://www.ncbi.nlm.nih.gov/genbank), reference proteomes of the UniProtKB database (ftp://ftp.uniprot.org/pub/databases), WormBase ParaSite online resource (http://parasite.wormbase.org/), The *Gyrodactylus salaris* Genome Project site (http://invitro.titan.uio.no/gyrodactylus), NHGRI Mnemiopsis Genome Project Portal (http://research.nhgri.nih.gov/mnemiopsis/), the comparative genomics platform for early branching metazoan animals (http://www.compagen.org/datasets.html). BLAST similarity search for OrthoMCL was carried out with an e-value cutoff of 1E-05, and the Markov clustering was performed with an inflation parameter of 1.5 [44]. The clustering returned 57,298 orthologous groups for all analyzed species. The predicted *I. variabili* proteins were clustered into 4,038 orthologous groups and contain a total of 4,271 *I. variabili* proteins or 83.5% of its whole proteome.

Functional annotation of coding genes

Initial functional annotation of *I. variabili* genome was performed using InterProScan 5.6 to search against PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, PIR SuperFamily and SUPERFAMILY databases, employing default settings [45]. Each predicted protein was assessed for a known functional ortholog, defined using the Kyoto Encyclopaedia of Genes and Genomes (KEGG) with best bi-directional hit method and a default bit score cutoff of 60 [46, 47]. A total of 2,833 *I. variabili* proteins were assigned a KEGG orthology. For comparison, the same procedure was also performed for the gene sets of *I. linei, Helobdella robusta*, and *Capitella teleta*. We estimated numbers of genes for transcription factors and signaling molecules in *I. variabili* genome to reveal the conserved genes relevant to development (See also Table S2). All gene models containing a homeobox domain, piwi domain and putative neurotransmitter receptor domains were identified using InterProScan (See also Tables S3 and S4). All InterProScan results were compared with KEGG annotation.

DATA AND CODE AVAILABILITY

All bioinformatic analyses were performed with open-source software in a Unix environment. All scripts used for analysis during this project are available at our GitHub repository for the project:

https://github.com/nilannik/scripts-for-genome-analysis. This genome project has been registered at NCBI under BioProject accession number PRJNA531602. Raw sequence data have been deposited at NCBI under SRA accession number SRS4604627.