

Genome of the fatal tapeworm *Sparganum proliferum* uncovers mechanisms for cryptic life cycle and aberrant larval proliferation

Taisei Kikuchi¹[✉], Mehmet Dayi^{1,2}, Vicky L. Hunt³, Kenji Ishiwata⁴, Atsushi Toyoda⁵, Asuka Kounosu¹, Simo Sun¹, Yasunobu Maeda¹, Yoko Kondo⁶, Belkisyole Alarcon de Noya⁷, Oscar Noya⁷, Somei Kojima⁸, Toshiaki Kuramochi⁹ & Haruhiko Maruyama¹

The cryptic parasite *Sparganum proliferum* proliferates in humans and invades tissues and organs. Only scattered cases have been reported, but *S. proliferum* infection is always fatal. However, *S. proliferum*'s phylogeny and life cycle remain enigmatic. To investigate the phylogenetic relationships between *S. proliferum* and other cestode species, and to examine the mechanisms underlying pathogenicity, we sequenced the entire genomes of *S. proliferum* and a closely related non-life-threatening tapeworm *Spirometra erinaceieuropaei*. Additionally, we performed larvae transcriptome analyses of *S. proliferum* plerocercoid to identify genes involved in asexual reproduction in the host. The genome sequences confirmed that the *S. proliferum* has experienced a clearly distinct evolutionary history from *S. erinaceieuropaei*. Moreover, we found that nonordinal extracellular matrix coordination allows asexual reproduction in the host, and loss of sexual maturity in *S. proliferum* are responsible for its fatal pathogenicity to humans. Our high-quality reference genome sequences should be valuable for future studies of pseudophyllidean tapeworm biology and parasitism.

¹Faculty of Medicine, University of Miyazaki, Miyazaki, Japan. ²Forestry Vocational School, Duzce University, Duzce, Turkey. ³Department of Biology and Biochemistry, University of Bath, Bath, UK. ⁴Department of Tropical Medicine, The Jikei University School of Medicine, Tokyo, Japan. ⁵Comparative Genomics Laboratory, Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Shizuoka, Japan. ⁶Division of Medical Zoology, Department of Microbiology and Immunology, Faculty of Medicine, Tottori University, Yonago, Japan. ⁷Institute of Tropical Medicine, Central University of Venezuela, Maracay, Caracas, Venezuela. ⁸Department of Clinical Laboratory Medicine, Chiba-Nishi General Hospital, Matsudo City, Chiba, Japan. ⁹Department of Zoology, National Museum of Nature and Science, Tsukuba, Ibaraki, Japan. ✉email: taisei_kikuchi@med.miyazaki-u.ac.jp

The cryptic parasite *Sparganum proliferum* was first identified in a 33-year-old woman in Tokyo in 1904. The patient's skin was infected with numerous residing cestode larva of unknown taxonomy. Ijima et al.¹ originally designated the parasite as *Plerocercoides prolifer*, and considered it a pseudophyllidean tapeworm in the plerocercoid larval stage. In 1907, an extremely similar human case was reported by Stiles in Florida, USA, and the responsible parasite was renamed *S. proliferum*². Clinical symptoms and post-mortem findings indicate that *S. proliferum* proliferates in humans and invades various organs and tissues, including the skin, body walls, lungs, abdominal viscera, lymph nodes, blood vessels, and the central nervous system, leading to miserable disease prognosis^{3,4}. Not many cases have been reported to date but the infection was fatal in all reported cases (reviewed in ref. 5).

There was a postulation about the origin of this parasite. Some parasitologists considered it to be a new species of pseudophyllidean tapeworm, whereas others suspected that *S. proliferum* was a virus-infected or aberrant form of *Spirometra erinaceieuropaei*, based on morphological similarities^{6,7}. DNA sequence analyses of mitochondrial NADH dehydrogenase subunit III, mitochondrial transfer RNA (tRNA), cytochrome oxidase subunit I, and nuclear succinate dehydrogenase iron-sulfur protein subunit (*sdhB*) genes suggested that *S. proliferum* is a closely related but distinct species to *S. erinaceieuropaei*^{8,9}. However, the adult stage of *S. proliferum* has not been observed and the precise taxonomic relationships of *S. proliferum* with other worms remain unclear because most of the submitted sequences are of mitochondria origin and include only a limited number of genes. Recently, adult forms of tapeworms were isolated from wild felids and reported that they showed high-DNA sequence similarity to *S. sparganum*¹⁰. However, this study was based on only a few mitochondrial genes and more evidence is required to determine the precise taxonomic classification of this parasite.

In addition to taxonomic considerations, the pathogenicity of *S. proliferum* and its mechanisms of proliferation and invasion in mammalian hosts are of considerable interest. In principle, plerocercoids of pseudophyllidean tapeworms (spargana), including those of *S. erinaceieuropaei* and other *Spirometra* species, do not proliferate asexually, but migrate through subcutaneous connective tissues, causing only non-life threatening sparganosis (non-proliferative sparganosis). Other organs, such as the lungs and liver or the central nervous system, may be niches for these worms but are not commonly described. Symptoms of non-proliferative sparganosis are mainly caused by the simple mass effect⁵. The life cycles of other pseudophyllidean species including those belonging to the genera *Dibothriocephalus*, *Diplogonoporus* and *Schistocephalus* are analogous to *Spirometra* spp.¹¹, except for the spectrum of second intermediate hosts, and asexual proliferation of plerocercoids, which have not previously been reported. Asexual proliferation of larvae and the destruction of host tissues are characteristic of cyclophyllidean tapeworms such as *Echinococcus*, which proliferates asexually by generating a characteristic germinative layer in a hydatid cyst form^{12,13}. In another cyclophyllidean tapeworm, *Mesocestoides*, asexual multiplication is achieved by longitudinal fission^{12,14}. In contrast, the pseudophyllidean *Sparganum* plerocercoid undergoes continuous branching and budding after invading the human body by an unidentified route, and produces vast numbers of progeny plerocercoids.

To clarify the phylogenetic relationship of the enigmatic parasite *S. proliferum* with other cestode species and investigate the underlying pathogenic mechanisms, we sequenced its entire genome as well as that of newly isolated *S. erinaceieuropaei* (it should be noted that recent studies suggested that Asian *S. erinaceieuropaei* can be taxonomically distinguished from

European *S. erinaceieuropaei*^{15,16}, but we retain the name *S. erinaceieuropaei* for the Japanese isolate). We also performed transcriptome analyses of *S. proliferum* plerocercoid larvae to identify genes that are involved in asexual proliferation in the host. Those analyses revealed its phylogeny and gene evolution that contribute to the proliferation and pathogenicity of *S. proliferum*.

Results

Genomic features of *S. proliferum* and *S. erinaceieuropaei*. We sequenced the *S. proliferum* genome using multiple insert-length sequence libraries (Supplementary Table 1) and compiled a 653.4-Mb assembly of 7388 scaffolds with N50 of 1.2 Mb. The *S. erinaceieuropaei* genome was assembled into 796 Mb comprising 5723 scaffolds with N50 of 821 kb. These assembly sizes were 51.9% and 63.2% of the previously published *S. erinaceieuropaei* genome (UK isolate)¹⁷. CEGMA and BUSCO report the percentage of highly conserved eukaryotic gene families that are present as full or partial genes in assemblies and nearly 100% of core gene families are expected in most eukaryote genomes. BUSCO analyses showed that 88.1% and 88.5% of core gene families were represented in *S. proliferum* and *S. erinaceieuropaei* genome assemblies, respectively, higher than or comparable to other previously published tapeworm genomes (Table 1). CEGMA completeness values for *S. proliferum* and *S. erinaceieuropaei* were slightly lower than those from BUSCO analyses. Low CEGMA completeness was also seen in other pseudophyllidean tapeworm genomes, including *S. erinaceieuropaei* UK isolate, *Dibothriocephalus latus*, and *Schistocephalus solidus* (Table 1 and Supplementary Data 1). Low CEGMA completeness values of these two genome assemblies could indicate pseudophyllidean-specific loss or high sequence/structure divergence of the genes that are conserved in other eukaryotic taxa. The percentage of duplicated genes in BUSCO analyses and average numbers of CEGs (CEGMA core genes) were 1.2% and 2.8% for *S. proliferum*, and 1.3% and 10.5% for *S. erinaceieuropaei*, respectively. To generate the genome assemblies we used the Haplomerger tool¹⁸ to collapse heterozygous haplotypes, which reduced the assembly sizes from 717.0 to 653.4 Mb and 968.1 to 796.0 Mb for *S. proliferum* and *S. erinaceieuropaei*, respectively. However, K-mer analyses of Illumina short reads estimated haploid genome sizes of 582.9 and 530.1 Mb for the two species (Supplementary Fig. 1a). These results indicate that the final assemblies roughly represent the haploid genome sizes of these tapeworms, but they still contain some heterozygous haplotypes and/or overestimated gap sizes especially in *S. erinaceieuropaei*. Ploidies were inferred from heterozygous K-mer pairs and were diploid for both species (Supplementary Fig. 1b).

The genomes of *S. proliferum* and *S. erinaceieuropaei* are highly repetitive, with ~53% of the total assembly length composed of repeats in both genomes (Supplementary Fig. 2 and Table 2). Long-interspersed nuclear elements (LINEs) occupy ~30% for both *S. proliferum* and *S. erinaceieuropaei*. These LINEs predominantly comprise the three types (Penelope, RTE-BovB, and CR1), which are also abundant in other pseudophyllidean genomes (Supplementary Fig. 2).

A total of 18,919 genes were predicted in *S. proliferum* assemblies, about 3,000 fewer than for *S. erinaceieuropaei* (22,162), which is a similar number to the number of genes in other pseudophyllidean genomes, but greater than the cyclophyllidean genomes (Table 1). In studies of the *S. erinaceieuropaei* UK isolate¹⁷, the gene number (>39,000) was likely overestimated due to fragmentation and redundancy in the assembly. In addition, we predicted ~4,000 and ~6,000 transposon proteins (e.g., gag-pol polyproteins and reverse transcriptases) for the *S. proliferum* and *S. erinaceieuropaei* assemblies. We cannot compare these

Table 1 Statistics of genome assemblies.

	<i>Sparganum proliferum</i> (v2.2)	<i>Spirometra erinaceiropaei</i> (v2.0)	<i>Spirometra erinaceiropaei</i> (UK) (WBPS13)	<i>Dibothriocephalus latius</i> (WBPS13)	<i>Schistocephalus solidus</i> (WBPS13)	<i>Hymenolepis microstoma</i> (WBPS15) ^a	<i>Taenia solium</i> (WBPS13)	<i>Echinococcus multilocularis</i> (WBPS15) ^a
Assembly size (Mb)	653.4	796.0	1258.7	531.4	539.4	168.9	122.4	115.0
Num. scaffolds	7388	5723	482,608	140,336	56,778	7	11,237	1217
Average (kb)	88.4	139.0	2.6	3.8	10.0	24,134	11.2	94.5
Largest scaff (kb)	8099	5490	90	80	595	43,032	740	20,116
N50 (kb)	1242	821	5	7	32	25,815	68	13,762
N90 (kb)	110	167	1	2	5	17,547	5	2924
Gaps (kb)	51020	77,788	128,163	38,407	22,091	6487	164	16,567
Cegma completeness	61.7/81.5	58.5/80.2	29.4/45.9	49.6/65.3	76.6/87.9	88.7/88.7	87.1/90.7	93.5/93.5
complete/partial (%)	1.1/1.3	1.1/1.3	1.8/2.2	1.5/1.6	1.2/1.3	1.1/1.1	1.2/1.2	1.1/1.1
Average CEG number	72.0/88.1	71.9/88.5	33.6/37.3	38.1/53.8	70.3/86.2	79.3/n.t. ^b	72.6/85.5	76.2/n.t. ^b
complete/partial	2.8	10.5	14.9	1.7	2.0	0.9	3.3	
BUSCO completeness (Metazoa dataset/Eukaryota dataset)								
% of duplicated BUSCOs	18,919	22,162	39,039	19,648	18,323	10,134	12,304	10,663
Num. coding genes	579.0	554.0	198.0	261.0	472.0	838.0	613.0	725.0
Coding gene size (median; aa)								
Num. transposon proteins	4124	6128	518	318	1905	58	177	83

^a WormBase ParaSite Release 13 (WBPS13) was used for the analyses in the paper.
^b Not tested.

Table 2 Statistics of repeats in genomes.

	<i>Sparganum proliferum</i> (v2.2)		<i>Spirometra erinaceiropaei</i> (v2.0)	
	Num element	% in bp	Num element	% in bp
SINEs:	36,816	1.24	35,223	1.09
LINEs:	601,034	31.63	653,492	31.49
LINE/Penelope	197,550	9.95	198,272	10.41
LINE/RTE-BovB	211,620	10.47	192,085	11.18
LINE/CR1	164,121	10.11	40,013	9.37
LTR element:1	115,379	4.93	152,710	6.49
LTR/Gypsy	89,911	3.79	52,938	1.81
DNA element:	41,513	1.75	125,893	4.07
DNA/CMC-EnSpm	4908	0.29	48,074	0.69
DNA/TcMar-Tc1	14,579	0.63	4007	0.53
Small RNA:	1964	0.09	4603	0.14
Satellites:	12,814	0.62	4931	0.17
Simple repeat:	85,865	1.60	72,020	0.41
Low complexity:	3278	0.03	3945	0.03
Unclassified:	416,273	21.10	507,009	21.55
TOTAL	1,314,936 (52.94%)		1,559,826 (52.87%)	

numbers directly with those of other cestode genome assemblies due to differences in gene prediction strategies, but those numbers are considerably high compared to other cestode genomes (Table 1), which is consistent with the results that *S. proliferum* and *S. erinaceiropaei* genome are highly repetitive. Further, the presence of a large number of intact coding sequences for transposon proteins indicates that those transposable elements are still active or were active until recently in the cestode genomes.

Phylogenetic placement of *S. proliferum*. Phylogenetic relationships of *S. proliferum* with other cestode species were inferred from 205 single-copy orthologues (Fig. 1 and Supplementary Table 3). A clear separation was identified between pseudophyllidea and cyclophilidea clades. In the pseudophyllidea clade, *S. proliferum* occupied the basal position of the *Spirometra* cluster, in which two *S. erinaceiropaei* isolates (Japan and UK isolates) were placed beside each other.

Phylogenetic tree topology based on mitogenomes of the 14 cestodes and all available mitogenome data of *Spirometra* in the GenBank, was similar to that of the nuclear genome (Supplementary Fig. 3 and Supplementary Table 2). Yet in contrast with the nuclear genome tree, the *S. erinaceiropaei* UK isolate was located in a basal position of the *Spirometra* cluster, placing *S. proliferum* in the middle of *Spirometra* species, albeit with a long branch. These inconsistencies between nuclear and mitogenome trees may reflect uncertainties of species classification in the genus *Spirometra*^{19,20}. Moreover, mitochondrial sequences can give poor inferences of species trees²¹. Cumulatively, these results suggest that *S. proliferum* has a close phylogenetic relationship with *Spirometra* but is clearly distinguished by genomic features and gene contents.

Gene family evolution. Protein family (Pfam) analyses revealed highly similar protein domain distributions of *S. proliferum* and *Spirometra* genomes ($r = 0.99$; Fig. 2 and Supplementary Data 2), confirming the close phylogenetic relationship between the two species. Only a few domains differed significantly in abundance between the two species. Among these, the *S. proliferum* genome was underrepresented in zinc-finger families (Zf-C2H2 and Zf-met), reverse transcriptase (RVT_1), galactosyltransferase, and alpha/beta hydrolase (Abhydrolase_6). Overrepresented Pfam domains in *S. proliferum* included a distinct type of zinc-finger domain (zf-3CxxC), fibronectin type III (fn3), trypsin, RNA polymerase III RPC4 (RNA_pol_Rpc4), and an N-terminal region of glycosyl transferase group 7 (Glyco_transf_7N).

We then performed gene family analysis using OrthoFinder²² with the predicted proteomes of *S. proliferum*, *S. erinaceiropaei*,

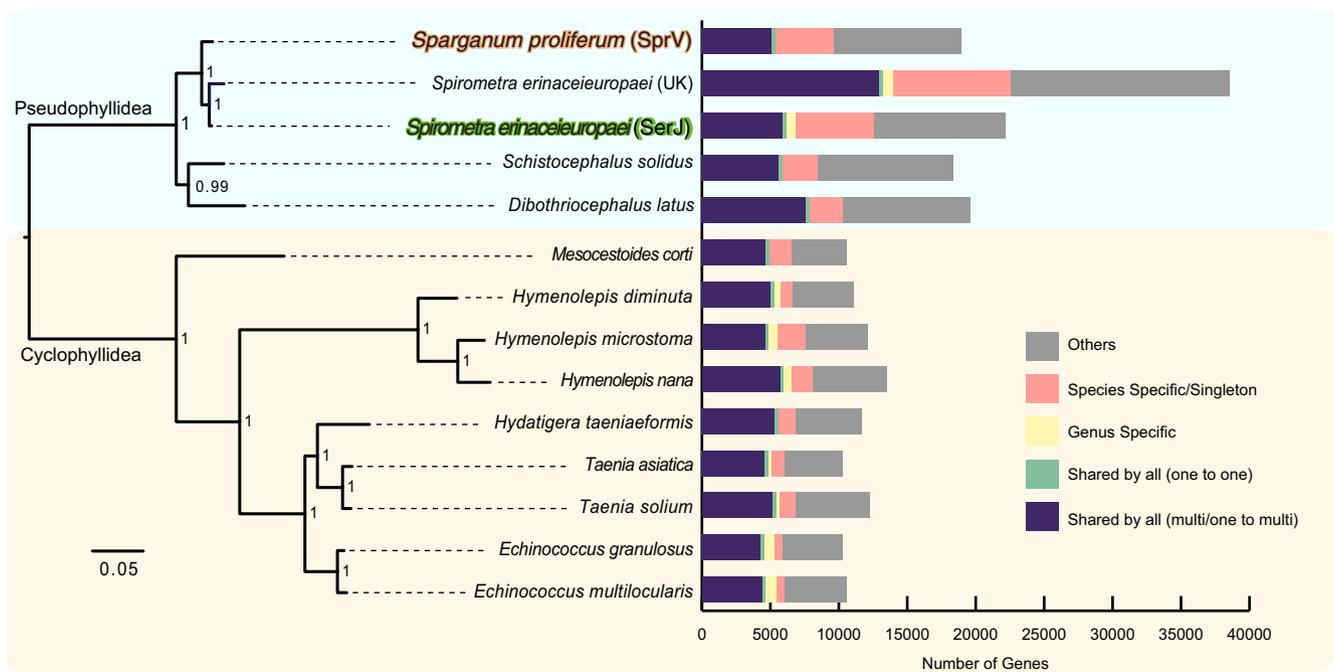


Fig. 1 Phylogeny and gene contents of selected tapeworms. Phylogenetic relationships of *S. proliferum* and other tapeworms were inferred by a maximum-likelihood tree analysis using 205 single-copy orthologous. Bootstrap confidence values via 500 replicates were shown at each node. Genes are categorised in a stack bar, and the length of stack bar is proportional to number of genes.

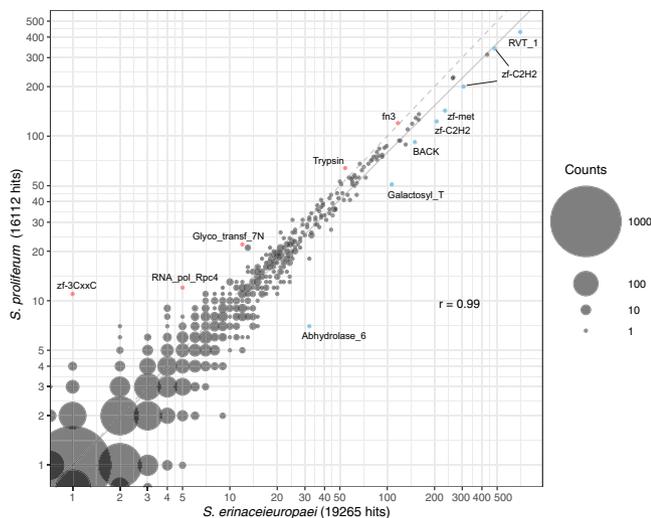


Fig. 2 A scatterplot showing the abundance of Pfam domains in *S. proliferum* and *S. erinaceieuropaei* genomes. The size of a circle represents the number of Pfam domains. Pfam domains that are more enriched in *S. proliferum* than in *S. erinaceieuropaei* are highlighted in red. Those enriched in *S. erinaceieuropaei* relative to *S. proliferum* are highlighted in blue.

and 12 other selected cestode genomes. A total of 219,816 proteins from 14 cestode species were placed into 18,738 gene families and 26,157 unassigned singletons (Fig. 1). The *S. proliferum* proteome (18,919 proteins) was encoded by 9119 gene families, among which 3674 gene families (composed of 5347 proteins) were shared by all 14 cestodes, and 4304 proteins were specific to the species or singletons. The *S. erinaceieuropaei* proteome (22,162 proteins) was clustered into 9,655 gene families. Although the *S. erinaceieuropaei* shared 7639 gene families with the other strain of *S. erinaceieuropaei* (UK), 30.4% and 39.7% of the total proteins for each strain were not shared. One of the

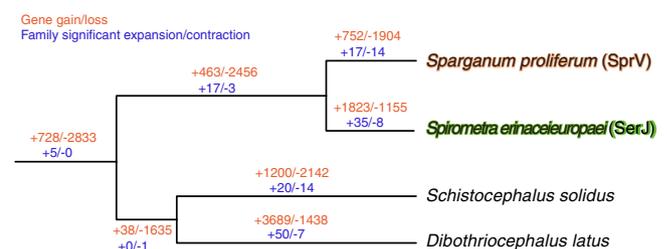


Fig. 3 Gene family evolution of selected cestode species was inferred using computational analysis of gene family evolution (CAFE). Numbers on each branch (or lineage) indicate specific gains/losses of that branch (or lineage).

reasons for this was the fragmented gene models in the UK genomes. To further investigate gene family evolution, we selected eight relatively high-quality genomes and performed computational analysis of gene family evolution (CAFE) to estimate gene family expansion and contraction. The analysis identified gene families with significantly higher than expected rates of gains and losses (Fig. 3 and Supplementary Table 3). Seventeen gene families were significantly expanded in the *S. proliferum* lineage, and these included annotations for fibronectin, trypsin tetraspanin family, Ras, zinc-finger C2H2 type, and core histone (Supplementary Data 3). Significantly contracted gene families (14 families) had annotations relating to signal transduction proteins, such as phosphatases and kinases (Supplementary Data 4). Fibronectin and zinc-finger C2H2 type were present in expanded and contracted families.

In the *S. erinaceieuropaei* lineage, 35 and 8 gene families were significantly expanded or contracted (Supplementary Data 5 and 6), respectively. Among them, highly lineage-specific expansion was found for 12 families (i.e., 6 or more genes in *S. erinaceieuropaei*, whereas one or no genes in *S. proliferum*). Those gene families mostly encode proteins of unknown function, but

Table 3 Homeobox complement in *S. proliferum* and *S. erinaceiropaei* compared with other tapeworms and bilaterians.

Homeobox class	<i>Sparganum proliferum</i> (v2.2)	<i>Spirometra erinaceiropaei</i> (v2.0)	<i>Taenia solium</i> (WBPS13)	<i>Echinococcus multilocularis</i> (WBPS13)	<i>Branchiostoma floridae</i>
ANTP	25	30	36	25	58
PRD	10	8	11	15(18)	21
CUT	3	4	3	3	4
SINE	3	4	2	3	3
TALE	8	10	11	12	10
CERS	2	1	2	2	1
POU	3	4	4	5	6
LIM	6	6	7	7	8
ZF	4	4	3(2)	3(2)	4
Total	64	71	79	75	115

they were likely expanded after speciation from *S. proliferum* and *S. erinaceiropaei* and may have specific roles in the *S. erinaceiropaei* life cycle or parasitism.

Conserved developmental pathway genes. Homeobox transcription factors are involved in patterning of body plans in animals. The homeobox gene numbers are much fewer in parasitic flatworms than in most other bilaterian invertebrates, which have a conserved set of ~100 homeobox genes. Genome sequencing of four cyclophyllid cestodes revealed that out of 96 homeobox gene families that are thought to have existed at the origin of the bilateria, 24 are not present in cestodes²³. We identified and manually curated homeobox genes in the *S. proliferum* and *S. erinaceiropaei* and found that the two pseudophyllid cestodes had similar homeobox class repertoires as those in cyclophyllid cestodes, in which class ANTP was the most abundant, followed by classes PRD and TALE; Table 3). The total numbers of homeobox domains identified in *S. proliferum* and *S. erinaceiropaei* are 64 and 71, respectively, which are the most reduced of any studied bilaterian animal as these were fewer than in the cyclophyllids *Echinococcus multilocularis* and *Taenia solium* (Table 3). The three homeobox families Pou/Pou6, ANTP/Bsx, and ANTP/Meox were not present in *S. proliferum* and *S. erinaceiropaei*, whereas the homeobox family ANTP/Ro was found in *S. proliferum* and *S. erinaceiropaei* but not in *E. multilocularis* and *T. solium* (Supplementary Fig. 4).

Comparisons between *S. proliferum* and *S. erinaceiropaei* showed that the homeobox families TALE/Pknox, ANTP/Hox1, ANTP/Msxlx, and POU/Pou-like are missing in *S. proliferum*, despite being present in the other cestodes. In contrast, the homeobox families ANTP/Dbx and PRD/Alx were found in *S. proliferum* but not in *S. erinaceiropaei*.

Other conserved genes with roles in flatworm developmental pathways, such as Hedgehog and Notch, were largely conserved in *S. proliferum* and *S. erinaceiropaei*. However, in the Wnt pathway, whose complement is much smaller than the ancestral complement in tapeworms²³, two further genes (Axin and LEF1/TCF) were missing in *S. proliferum* and *S. erinaceiropaei* (Supplementary Data 7).

Horizontally transferred genes. To determine whether the present genomes contained horizontally transferred genes (HTGs) from other organisms, we used a genome-wide prediction method based on a lineage probability index using the software Darkhorse²⁴ and identified 19 and 33 putative HTGs in *S. proliferum* and *S. erinaceiropaei*, respectively (Supplementary Data 8 and 9). For these transfers, all possible host organisms were bacteria except for one *Spirometra* gene that has high similarity to a chlorella virus gene. Orthologues of most *S. proliferum* putative HTGs were also detected as horizontally

transferred in *S. erinaceiropaei*. Moreover, possible host bacteria, including *Marinifilum breve*, *Aphanizomenon flos-aquae*, *Alcanivorax* sp., and *Vibrio* sp., were shared by the two cestode species and were aquatic or marine bacteria, indicating that these genes were acquired by a common ancestor of the two tapeworms, which had aquatic phase in the life cycle.

Signatures of selection in the *S. proliferum* lineage. Positive selection is a mechanism by which new advantageous genetic variants sweep through a population and drive adaptive evolution. To investigate the roles of positive selection in the evolution of *S. proliferum* from a non-life-threatening ancestor, we performed dN/dS branch-site model analyses with 755 single-copy orthologous genes from 12 tapeworms (Supplementary Data 10) and identified a total of 35 positively selected genes in the *S. proliferum* lineage (Supplementary Table 4). Evolutionary pressures were identified for some genes that are essential to cellular processes, including transcription/RNA processing/translation genes encoding DNA-directed RNA polymerase II subunit, polypyrimidine tract-binding protein, adenylate kinase, ribosomal protein L21, snu13 NHP2-like protein, and eukaryotic translation initiation factors. Other identified genes were related to transportation (dynein intermediate chain 2) and mitochondrial processes (Rieske). Genes involved in stress and immune responses, such as DNAI/Hsp40, HIKESHI protein, Toll-like receptor, and Ig_3/Ig, were also positively selected in the *S. proliferum* lineage, along with the RAS oncogene *Rab-4A*.

Circumstances where environmental change eliminates or weakens selection pressures formerly important for the maintenance of a particular trait, have been termed “relaxed selection”²⁵. Such circumstances include, for example, eyesight in lightless caves and pathogen resistance after disappearance of pathogens. Genes under relaxed selection can be detected using the ratio of non-synonymous to synonymous substitutions and a comparison in a phylogenetic framework²⁶. In *S. proliferum* we sought to test the relaxed selection that may have acted on genes involved in development or sex reproduction under its contentious asexual reproduction. Using the alignment set used in the positive selection tests, we detected 10 genes that were under “relaxed selection” in the *S. proliferum* lineage, relative to the other tapeworm lineages (Supplementary Table 5). These genes encode proteins with putative roles in developmental regulation and cell differentiation. In particular, the receptor roundabout (ROBO) and secreted molecules of the SLIT family, together, play important roles in guiding axons and proper morphogenesis²⁷. The Rho-GTPase-activating protein is also highly expressed in highly differentiated tissues and affects cell differentiation by negatively regulating Rho-GTPase signalling²⁸. Delta-like protein (DLL) is an inhibitory ligand of the Notch receptor pathway and is expressed during brain development²⁹. Vascular endothelial

growth factor receptor is also known to regulate stem cell homeostasis and repopulation in planarian species³⁰. Hence, these instances of relaxed selection indicate that certain developmental pathways became less important for the worm or, further, possibly the worm has long since used those pathways. We also identified two genes encoding cadherin (protocadherin) that were subject to relaxed selection. Cadherin is a transmembrane protein that mediates cell–cell adhesion in animals and the signatures of relaxed selection suggest cell adhesion processes have diverged between the worm species.

Differential gene expression involved in asexual proliferation and parasitism. We maintained *S. proliferum* via serial infection of mice and found that some plerocercoid worms exhibit a highly branching structure, which was observed frequently in heavily infected mice. In contrast, in mice with low worm burdens, most worms had unadorned non-branching morphology. To investigate physiological characteristics of those worms, we compared morphology, viability and mobility of 81 worms showing variable branching levels with branch numbers ranging from 1 to 79 (Supplementary Table 6). We calculated fractal dimension value for each worm, which represents morphological complexity, and found that the value better represents the worm status than using only the branch numbers. Fractal dimension not only shows a good correlation with the number of branches ($r = 0.59$, p -value = $6e-9$) (Supplementary Fig. 5a), but is also possible to distinguish the unadorned forms representing low branching activity from worms recently split from a highly branching worm, which shows simple but relatively slender forms. Cell viability evaluated with the assays based on cellular NADH/NADPH activity³¹ and extracellular lactate dehydrogenase (LDH) activity³² showed almost no correlation with the fractal dimension values ($r = -0.01$, p -value = 0.88 and $r = 0.08$, p -value = 0.51 , respectively) (Supplementary Figure 5b and S5c). However, higher mobility was observed in the high-complexity worms than the low ones ($r = 0.28$, p -value = 0.04), which possibly results in more frequent fragmentation of branches (Supplementary Fig. 5d). Thus, those high-complexity worms are considered the main sources of new plerocercoid worms in the host via their frequent branching and high mobility. We call this highly branching form “medusa-form” and the unadorned non-branching form “wasabi-form” because of the resemblances of those structures with Medusa’s head” and wasabi rhizomes (Fig. 4a). Further histological observations revealed that the parenchymal cavity, which is characteristic of *S. proliferum*³³, was more developed in the medusa-form and the cavities were often filled with contents stained in pink by hematoxylin and eosin (H-E) (Supplementary Fig. 6). In contrast, most parenchymal cavities in the wasabi-form were empty. Additionally, in the medusa-form, we found loci where nuclei were concentrated. Calcareous bodies, demonstrated by von Kossa staining, were more often observed in the medusa-form than the wasabi-form. These observations suggest that the typical histological characteristics of *S. proliferum* reported previously were of the medusa-form³³.

We performed a transcriptome analysis using nine RNAseq libraries generated from five and four worms showing typical morphological characters of medusa and wasabi forms, respectively, to identify genes involved in the plerocercoid proliferation. Interestingly, RNA expression patterns of the two forms were clearly separated by the first dimension of Multidimensional scaling (MDS) plot based on Spearman correlations (Fig. 4b and Supplementary Table 7). EdgeR RNAseq comparisons identified 357 differentially expressed genes (DE genes) between the medusa and wasabi forms (246 upregulated and 111 downregulated in medusa forms) (Fig. 4c). In this analyses we used a total of 23,043

gene models which included 4124 transposon proteins and found many transposon proteins such as gag-pol polyproteins and reverse transcriptases (36 genes) were upregulated in the medusa form. Other than that, the upregulated set in medusa forms were dominated by genes encoding peptidases and peptidase inhibitors, such as tolloid-like proteins (19 genes), chymotrypsin-like proteins (6 genes) and CAP domain-containing proteins (12 genes) (Supplementary Data 11). This set of DE genes was enriched in the GO categories for metalloendopeptidase activity and proteolysis (Supplementary Table 8). Downregulated genes also encoded a variety of peptidases and peptidase inhibitors, including leucyl aminopeptidase (five genes), chymotrypsin-like proteins (seven genes), and kunitz bovine pancreatic trypsin inhibitor domain protein (three genes), with high representation under the GO terms metalloexopeptidase, aminopeptidase, and manganese ion binding (Supplementary Table 8). Peptidases and peptidase inhibitors are secreted by many types of pathogens, including bacteria, fungi, and parasites, and often play critical roles in survival and virulence^{34–36}. Other genes known to be involved in pathogenicity in other pathogens were also upregulated in the medusa form, including genes encoding multidrug resistance-associated proteins³⁷ and tetraspanins. The latter proteins have four transmembrane domains and not only play roles in a various aspects of cell biology but also are used by several pathogens for infection and regulate cancer progression³⁸.

Genes that are involved in cell-growth and cancer development were also upregulated in the medusa form, including those encoding proteins from wnt (WNT-111 and WNT-5) and ras/rab (RAS-0b, RAS-2 and Rasef) pathways, transcription factors/receptors (SOX1a, fibroblast growth factor receptor) and homeobox proteins (prospero, PAX, orthopedia ALX and ISL2).

It has been shown that expansions of gene families and changes in expression levels have been associated with the evolution of parasitism in previous studies^{39,40}. An upregulation of genes from expanded gene families was also found in *S. proliferum*. For instance, 12 out of 20 genes were identified as upregulated from an expanded gene family encoding CAP domain-containing proteins (orthogroup OG0000040). The orthogroup OG0000044 includes 49 *S. proliferum* genes encoding Chymotrypsin-like (mastin-like) proteins (Supplementary Data 3), and six of these were upregulated and another six were downregulated in the medusa form (Supplementary Data 11). Gene family OG0000044 is Pseudophillidea specific and gene family expansions were observed in *Dibothriocephalus*, *Spirometra* and *S. proliferum* although a phylogenetic analysis of these genes could not define upregulated or downregulated gene cluster in the phylogenetic tree (Supplementary Fig. 7). Phylogenetic analyses of the other expanded gene families containing multiple DE genes indicate that some of these orthogroups are conserved across flatworms, while others are specific to the Pseudophillidea clade of flatworms (Supplementary Fig. 7).

Thirty-six upregulated genes showed fold-changes differences >200 (Supplementary Data 11), including eight genes encoding tolloid-like proteins. In mammals, tolloid-like proteins compose a small group called BMP-1/TLD family with bone morphogenetic proteins (BMP), playing fundamental roles in morphogenesis and extracellular matrix (ECM) assembly via a direct effect on growth factors such as TGF- β and IGFs^{41,42}. The high fold-change of the tolloid-like protein genes in the medusa-form is, therefore, likely reflecting the high-branching activity of the worms and suggests an important role of these proteins in *S. proliferum* asexual proliferation. *Sprfv2_0697700.1* gene encoding acetylcholinesterase also showed a very high fold-change (>1000 folds). Acetylcholinesterases have a well-defined role in cholinergic neurotransmission in both vertebrates and invertebrates⁴³, but they are also known to be secreted by a variety of parasitic nematodes and have been

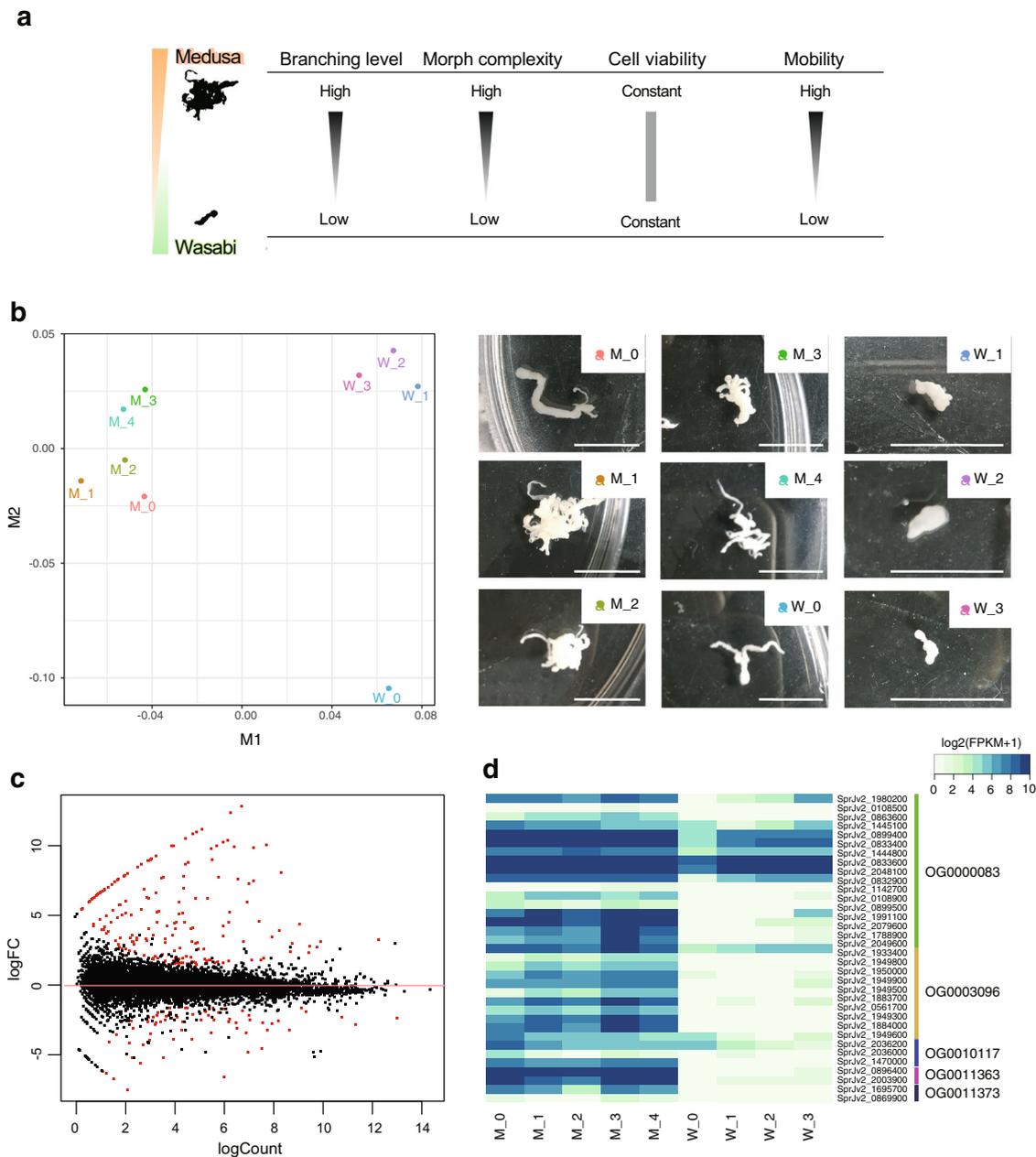


Fig. 4 Comparison of gene expression in highly branching worms (medusa form) relative to static worms (wasabi form) of *S. proliferum*. **a** A schematic illustration to show characteristics of medusa and wasabi forms of plerocercoid. **b** Multidimensional scaling (MDS) analyses of RNA-seq samples clearly separate the two forms by dimension 1. Pictures of used samples are shown on the right. Scale bars indicate 10 mm. **c** Bland-Altman (MA) plot of the two-form comparison; dots represent transcripts and log₂-fold-changes (medusa/wasabi) plotted against average abundance in counts per million. Red dots indicate differentially expressed transcripts with false discovery rates (FDR) of <0.05 and fold-changes of >2. **d** Heatmap of gene families encoding novel secreted proteins; the heat map shows log₂ fragments per kilobase per million reads mapped (FPKM) values for five gene families.

speculated to work in modulating host's immunity, inflammation and physiology^{44,45}. The high expression of the gene in the medusa-form made us conceive a possibility that it plays a role in the parasitism in some manner, though no secretion signal peptides were identified in the protein and little is known about secreted acetylcholinesterases of cestodes⁴⁶.

Notably, 24 of the 36 high fold-change genes were uncharacterised proteins. In total DE genes, 85 out of 246 that were upregulated in medusa forms have no known functions. These included 17, 10, 3, 2, and 2 genes from orthogroups OG0000083, OG0003096, OG0010117, OG0011363, and OG0011373, respectively. These orthogroups were expanded in the *S. proliferum* lineage and most DE genes had extremely high fold-changes

(Fig. 4c). As their products predominantly harbour secretion signal peptides (Supplementary Data 11), they are likely to be secreted by the parasite into the host and play important roles in parasitism, aberrant larval proliferation in the host, and/or modulation of host immunity.

Discussion

S. proliferum is a cryptic parasite with fatal consequences, but its phylogeny and life cycle are poorly understood. In this study, we sequenced the *S. proliferum* genome and performed comparative genomics with other tapeworm species, including the newly sequenced *S. erinaceiropaei* genome. The *S. erinaceiropaei*

genome was sequenced previously¹⁷, with an estimated genome size of more than 1.2 Gb, but because the source material was from a biopsy the assembled sequence was highly fragmented. Hence, the *S. erinaceiropaei* genome presented herein provides a more reliable estimate of the size and contents of this parasite genome although the kmer-based haploid genome size estimate suggests that the new assembly still contains un-collapsed heterozygous haplotypes. The new genome assembly was about two thirds of the size of the previous assembly, but remains the largest genome among sequenced tapeworms. Compared to cyclophyllidean tapeworms, including *Echinococcus* and *Taenia* spp., for which high-quality genome references are available^{23,47,48}, genome information for pseudophyllidean tapeworms is limited⁴⁹. The genomes presented in this study could, therefore, serve as a powerful resource for more comprehensive studies of tapeworm genomics and will facilitate the understanding of pseudophyllidean tapeworm biology and parasitism.

There have been three big knowledge gaps for the cryptic tapeworm *S. proliferum*: (1) its phylogenetic relationship with *Spirometra* species, (2) its life cycle including the definitive and intermediate hosts, and (3) genetic and physiological differences with non-proliferating *Spirometra* species that enable the worm to reproduce asexually in non-definitive hosts, such as humans and mice.

To determine phylogenetic relationships, we confirmed that the genetic sequence of *S. proliferum* is distinct from that of *S. erinaceiropaei*, despite the close relationship between these species. Specifically, the *S. proliferum* genome is about 150-Mb smaller and contains 3000 fewer protein-coding genes than in *S. erinaceiropaei*. Both genomes, nonetheless, showed diploidy. These data suggest that *S. proliferum* is not an aberrant form of *Spirometra* worm by virus infection or by small mutations^{6,7} and not a hybrid origin of multiple *Spirometra* species. In agreement, no virus-like sequences were detected in *S. proliferum* DNA or RNA raw reads.

We were unable to identify definitive or intermediate hosts of *S. proliferum* in the current study. Recent horizontal transfers of genes or mobile elements can indicate ecological relationships, because HGT events occur between closely associated organisms. Well-known examples include HGT from Wolbachia symbionts to their host insect^{50,51} and transfer of BovB retrotransposons between ruminants and snakes via parasitic ticks^{52,53}. We found that RTE/BovB repeats are abundant in the *S. proliferum* genome, but were likely acquired by an ancestral pseudophyllidea, as indicated by their abundance in *D. latus* and *S. solidus*. Moreover, our HGT screening analyses indicate several genes were acquired from bacteria but these HGTs likely have occurred before specification of *S. proliferum* and *S. erinaceiropaei*. The high-quality reference genomes presented herein, however, provide valuable resources for further attempts to identify vestigial *S. proliferum* sequences in other organisms or to perform analyses of protein–protein interactions between hosts and parasites. Recently, a possible adult form of *S. proliferum* was isolated from wild felids and fragments of mitochondrial *cox-1*, *nad-1* and *atp-6* genes were sequenced¹⁰. Although those sequences showed higher similarities to *S. proliferum* than *S. erinaceiropaei* or other tapeworms, the percentages of identities with the *S. proliferum* mitochondrion were relatively low (93% (320/343) in *cox1*, 96% (284/296) in *nad1* and 92% (531/577) in *atp6*), suggesting they are closely related but not identical species. Further sequencing analyses of those worms will help understand the origin, biology and pathogenicity of *S. proliferum*.

Loss of genes that are involved in the development of multicellular organisms and nervous systems, including homeobox genes and genes for zinc-finger domain-containing proteins, as well as relaxed selection of some developmental genes (ROBO,

Slit, RHOGAP, etc.), suggests that *S. proliferum* has lost the ability to undergo proper development and complete the sexual life cycle. Although the precise functions of homeobox genes in tapeworms remain elusive, homeobox families that are missing in *S. proliferum* (TALE/Pknox, ANTP/Hox1, ANTP/Msxlx and POU/Pou-like) appear to have important roles in the development of embryos and adult body plans. For example, Hox1 of the HOX gene family specifies regions of the body plan of embryos and the head–tail axis of animals⁵⁴. Products of the Pknox gene family, also known as the PREP gene family, are implicated as cofactors of Hox proteins⁵⁵. *Msxlx* homeobox gene was highly upregulated in the ovaries, continually expressed in fertilised ova in the uterus in *Hymenolepis microstoma* and suggested to be related to the female reproductive system in this tapeworm⁵⁶. POU class genes are present in all animals playing roles in neurogenesis, pluripotency, and cell-type specification. The POU homologue Oct-4 is one of the four “Yamanaka factors” used to induce pluripotent stem cells in mammals⁵⁷. Specific loss of Pou-like genes and relaxed selection of Pou3 suggest that *S. proliferum* has low dependency on POU genes.

We contend that the loss of sexual maturity of this parasite is related to its fatal pathogenicity in humans, because survival of the parasite is dependent on asexual reproductive traits of budding and branching, which lead to 100% lethality in infected humans. Accordingly, we identified genes that are upregulated in vigorously budding worms using transcriptome analyses and then selected genes that are putatively important for asexual proliferation, such as a variety of peptidase genes and oncogene-like genes. Among them, groups of secreted proteins with unknown functions were of great interest. They were expanded in the *S. proliferum* genome and showed more than 10-fold-changes in expression levels. Recently, an *S. erinaceiropaei* gene belonging to one of those groups (orthogroup OG0000083) was cloned and named plerocercoid-immunosuppressive factor (P-ISF)⁵⁸. P-ISF is a cysteine-rich glycoprotein abundant in plerocercoid excretory/secretory products and likely involved in immunomodulation of its hosts, by suppressing osteoclastogenesis including the gene expression of TNF- α and IL-1 β , and nitric oxide production in macrophages^{59,60}. Upregulation of P-ISF genes in *S. proliferum* proliferating worms is, therefore, reasonable and the expansion of the gene family in *S. proliferum* indicates the considerable contribution to the specific life cycle. The other upregulated gene families of unknown function are also expanded in *S. proliferum*, suggesting possible important roles in the hosts, therefore, future studies of these novel genes are required to fully understand the mechanism underlying the *S. proliferum* parasitism.

Fibronectin is an extracellular matrix (ECM) glycoprotein that controls the deposition of other ECM proteins, including collagens and latent TGF-beta binding protein⁶¹. During branching morphogenesis, accumulations of fibronectin fibrils promote cleft formation by suppressing cadherin localisation, leading to loss of cell–cell adhesion⁶². The present observations of the *S. proliferum* lineage show specific expansions of three gene families containing fibronectin type III domains. *S. proliferum* also had fewer cadherin genes than *S. erinaceiropaei* and three of them are subject to relaxed selection in *S. proliferum*. Furthermore, we found extremely high upregulation of tolloid-like proteins in highly branching medusa worms, which are known to be involved in ECM assembly in mammals via a direct effect on growth factors. These results collectively suggest nonordinal ECM coordination in *S. proliferum*, allowing the formation of highly branching structures and enabling asexual proliferation in the host.

Methods

Biological materials. *S. proliferum* strain Venezuela was used for the genome analyses. The parasite was originally isolated from a Venezuelan patient in 1981

and has been maintained by serial passages using BALB/c mice via intraperitoneal injections of the plerocercoids in National Science Museum as described in Noya and colleagues^{63,64}. *S. erinaceiueuropaei* used for the genome project was a Noya isolate obtained from a Japanese four-striped rat snake (*Elaphe quadrivirgata*) purchased in Yamaguchi prefecture, Japan in 2014.

DNA and RNA extraction and sequencing. *S. proliferum* worms were collected from the abdominal cavity of infected mice and washed thoroughly with 1x PBS. Plerocercoids of *S. erinaceiueuropaei* were isolated from the subcutaneous tissues of the snake. Genomic DNA was extracted from a single worm using Genomic-tip (Qiagen) following the manufacturer's instructions and used for genome sequencing for both species.

Paired-end sequencing libraries (Supplementary Table 1) were prepared using the TruSeq DNA Sample Prep kit (Illumina) according to the manufacturer's instructions. Multiple mate-paired libraries (3, 8, 12 and 16 kb) were also constructed using the Nextera Mate-Paired Library Construction kit (Illumina). Libraries were sequenced on the Illumina HiSeq 2500 sequencer using the Illumina TruSeq PE Cluster kit v3 and TruSeq SBS kit v3 (101, 150 or 250 cycles x 2) or the Illumina MiSeq sequencer with the v3 kit (301 cycles x 2) (Supplementary Table 1). The raw sequence data were analysed using the RTA 1.12.4.2 analysis pipeline and were used for genome assembly after removal of adapter, low quality, and duplicate reads.

RNA was extracted from individual worms using TRI reagent according to the manufacturer's instructions. Total RNA samples were qualified using Bioanalyzer 2100 (Agilent Technology, Inc.). Only samples with an RNA integrity value (RIN) greater than 8.0 were used for library construction. One hundred ng of total RNA was used to construct an Illumina sequencing library using the TruSeq RNA-seq Sample Prep kit according to the manufacturer's recommended protocols (Illumina, San Diego, USA). The libraries were sequenced for 101 or 151 bp paired-ends on an Illumina HiSeq2500 sequencer using the standard protocol (Illumina).

K-mer analysis. A K-mer count analysis was performed using K-mer Counter (KMC)⁶⁵, on the paired-end Illumina data. Only the first read was used to avoid counting overlapping k-mers. Genome size and ploidy estimations were performed using Genomescope⁶⁶ and Smudgeplot, respectively⁶⁷.

Genome assembly. Illumina reads from multiple paired-end and mate-pair libraries (Supplementary Table 1) were assembled using the Platanus assembler⁶⁸ with the default parameter. Haplomerger2¹⁸ was then used to remove remaining haplotypic sequences in the assembly and contigs were further scaffolded using Illumina mate-pair reads using SSPACE⁶⁹, CEGMA v2⁷⁰ and BUSCO⁷¹ were used to assess the completeness of the assemblies.

Mitochondrial genomes (mitogenomes) were reconstructed from Illumina reads with MITObim version 1.6⁷². Mitochondrial fragments in the nuclear genome assembly were identified by BLASTX using *S. mansoni* mitochondrial genes as queries and those fragments were extended by iterative mappings of Illumina short reads using MITObim. Assembled mitogenomes were annotated for protein-coding, tRNA and ribosomal RNA genes using the MITOS web server⁷³. Assemblies and annotations were manually curated using the Artemis genome annotation tool⁷⁴ with based on evidence supports from sequence similarity to other published mitogenomes.

Repeat analysis. Repeats within the genome assemblies were identified using the combined outputs of RepeatModeler (v1.0.8, <http://www.repeatmasker.org/RepeatModeler.html>), MITE-hunter (v11-2011)⁷⁵, LTR-harvest/LTR-digest (v1.6.1)⁷⁶ and TransposonPSI (v08222010) (<http://transposonpsi.sourceforge.net>). For each species, search v7⁷⁷ and RepeatClassifier (v1.0.8, <http://www.repeatmasker.org/RepeatModeler.html>) was used to cluster and classify repeat sequences from multiple tools to generate consensus sequences for a non-redundant repeat library with $\geq 80\%$ identity threshold. RepeatMasker (v3.2.8, <http://www.repeatmasker.org>) was then run with the custom repeat library, to calculate the distribution of each repeat and its abundance in the genome assemblies.

Gene prediction and functional annotation. To predict protein-coding genes, Augustus (v. 3.0.1)⁷⁸ was trained for *S. proliferum* and *S. erinaceiueuropaei*, individually, based on a training set of 500 non-overlapping, manually curated genes. To obtain high-confidence curated genes, a selection of gene models from gene predictions based on Augustus *S. mansoni* parameters, were manually curated in Artemis using aligned RNA-seq data and BLAST matches against the NCBI database. RNA-seq reads were mapped to the genomes using Hisat2 (parameters: --rna-strandness RF --min-intronlen 20 --max-intronlen 10000)⁷⁹. Based on the Hisat2 alignments, the bam2hints programme (part of the Augustus package) was used to create the intron hints, with minimum length set to 20 bp. Augustus were run with trained parameters using all the hints for that species as input. Introns starting with "AT" and ending with "AC" were allowed (--allow_hinted_splicesites=atcac). A weight of 10^5 was given to intron and exonpart hints from RNA-seq. If Augustus predicted multiple, alternatively spliced transcripts for a gene, we only kept the transcript corresponding to the longest predicted protein for further analyses. The

gene prediction sensitivity and specificity at exon level evaluated with ~200 each manually curated genes (not overlapping with the training sets) were 79.0% and 64.2% for *S. proliferum* and 76.2% and 64.0% for *S. erinaceiueuropaei*, respectively.

Functional annotations were performed on the gene models based on multiple pieces of evidence, including BLASTP search against NCBI nr database and the latest version Pfam search (ver. 30.0) with HMMER3⁸⁰. Gene ontology (GO) terms were assigned to genes using Blast2Go (v2)⁸¹ with BLAST search against NCBI nr database and the InterProScan results. TransposonPSI (release 08222010; <http://transposonpsi.sourceforge.net>) was used to identify potential transposon proteins in the protein sets.

Gene family analysis. To estimate branch or lineage-specific gain and loss of orthologous gene families, OrthoFinder²² and CAFE (v3)⁸² under parameters "-p 0.01, -r 1000" were used. The genome datasets of 12 cestode species (*Dibothriocephalus latus*⁴⁹, *Echinococcus granulosus*²³, *Echinococcus multilocularis*²³, *Hymenolepis diminuta*⁴⁹, *Hymenolepis microstoma*⁸³, *Hymenolepis nana*, *Hydatigera taeniaeformis*⁴⁹, *Mesocostoides corti*⁴⁹, *Spirometra erinaceiueuropaei*¹⁷, *Schistocephalus solidus*⁴⁹, *Taenia asiatica*⁸⁴, *Taenia solium*²³) used for the analyses were downloaded from WormBase Parasite WBPS13 (<https://parasite.wormbase.org/>)⁸⁵.

Species tree reconstruction. Amino acid sequences in each single-copy gene family were aligned using MAFFT version v7.22152⁸⁶, poorly aligned regions were trimmed using GBLOCKS v0.91b53⁸⁷, and then the trimmed alignments were concatenated. Genetic distance was calculated with the Poisson model and uniform rates among sites. A maximum-likelihood phylogenetic tree was produced based on the concatenated alignment using RAxML v8.2.754⁸⁸ with 500 bootstrap replicates. The best-fitting substitution model for each protein alignment was identified using the RAxML option (-m PROTGAMMAAUTO). Mitochondrial genome phylogeny was also constructed by the same method using 12 protein-coding genes on mitogenomes.

Identification of developmental pathway genes. Homeobox genes in *S. proliferum* and *S. erinaceiueuropaei* were identified using OrthoFinder results and BlastP analyses using lophotrochozoan genes as queries^{89,90}. Identified genes were manually curated for the gene structures using Artemis⁷⁴ and absence of genes was confirmed by tBlastn analyses using protein queries against the genome sequences. Genes involved in other conserved developmental pathways, including Hedgehog, Notch and Wnt were also identified in the same way²³.

Screening for horizontally transferred genes. To screen potential horizontal gene transfers (HGTs) into the *S. proliferum* and *S. erinaceiueuropaei* lineages, we used DarkHorse v2, which detects phylogenetically atypical proteins based on phylogenetic relatedness of BlastP hits against a taxonomically diverse reference database using a taxonomically weighted distance algorithm²⁴. Options (-n 1 -b 0.5 -f 0.1) were used in DarkHorse HGT screening. Franking genes of the identified candidates were checked for similarity to any cestode genes to exclude possibilities of mis-assembly or contamination.

Positive/relax selection scans (dN/dS). To analyse selection pressures in *S. proliferum* genes, the ETE3 Python package⁹¹ for CODEML⁹² was employed to calculate the non-synonymous (dN) and synonymous (dS) substitutions rates, and the ratio (dN/dS or ω). Nucleotide sequences of 755 single-copy orthologue genes from 12 cestode species (*S. proliferum*, *S. erinaceiueuropaei*, *D. latus*, *S. solidus*, *H. diminuta*, *H. nana*, *H. taeniaeformis*, *T. solium*, *T. asiatica*, *E. multilocularis*, *E. granulosus*, *M. corti*) were aligned based on each corresponding amino acid alignment using Pal2aln v14⁹³ with the parameters (-nomismatch and -nogap). After removing poor alignments manually, dN/dS were estimated using branch-site models with *S. proliferum* as the foreground and other branches in the tree as the background. The non-null model (bsA) were compared with the null model (bsA1) for each tree using a likelihood ratio test (LRT), where log-likelihood ratios were compared to a chi-square distribution with 1 degree of freedom. False discovery rate (FDR) correction was performed over all the *p*-values and genes showing FDR <0.05 were manually curated before obtaining final dN/dS values.

Test for relaxed selection was performed using the RELAX tool²⁶ with aforementioned nucleotide alignments of the single-copy orthologue genes. The relaxation parameter *k* was calculated for each branch and tested by LRT with *S. proliferum* as foreground and the others as background, and genes showing *k*-values <1 and likelihood ratio test values <0.05 were listed as candidate genes under relaxed selection.

Worm viability/mobility assays and histological observation. *S. proliferum* plerocercoid worms isolated from infected mice were washed with PBS and transferred RPMI media (Gibco) both prewarmed at 37 °C. Worm mobility in RPMI at 37 °C was recorded for 15 min using Wormlab instrument (MBF Bioscience) and analysed for the average migration speed using ImageJ and wrMTck plugin^{94,95}. Worm viability and cell damage were evaluated using water-soluble tetrazolium salt (WST)-8 (Cell Counting Kit-8, Dojindo Laboratories) and lactate dehydrogenase (LDH, Cytotoxicity LDH Assay Kit-WST, Dojindo

Laboratories) assays, respectively. Complexities of worm morphs were quantified by calculating fractal dimensions using ImageJ and FracLac plugin⁹⁶. For histological observation, isolated worms were immediately fixed in 4% paraformaldehyde phosphate buffer solution and embedded in paraffin. Tissue slides of 5 µm thick sections were stained with haematoxylin and eosin (H-E) and observed under light microscopy.

RNAseq analysis. For gene expression analyses, *S. proliferum* plerocercoid worms were grouped into two types based on the morphology and proliferation activity; worms vigorously branching to form structure like Medusa's head and worms under static form to form like wasabi rhizomes (Fig. 4a). Five medusa-form and four wasabi-form worms were collected from infected mice on ~50 weeks post inoculation. RNA was extracted from the individual worms and sequenced independently as described above. RNAseq reads were mapped to the *S. proliferum* reference genomes (v2.2) using Hisat2⁷⁹ (parameters: --rna-strandness RF --min-intronlen 20 --max-intronlen 10000). Mapped read count of each gene was calculated using HTSeq⁹⁷ with options (-s no, -a 10, -m union) and differential expression analyses were performed using EdgeR v3.24⁹⁸. A transcript was identified as differentially expressed in a pairwise comparison if the following criteria were met: false discovery rate (FDR) ≤ 0.001 and fold-change ≥ 2.0. Spearman correlation was calculated between samples using R and used to generate for multidimensional scaling (MDS) plots. FPKM values were calculated using Cufflinks packages v2.2.1⁹⁹ and used to generate and gene expression heatmaps.

Statistics and reproducibility. Statistical analyses were performed using independent biological replicates, which is considered to be a large enough sample for the statistical estimations, as described in Methods with available packages including R, edgeR⁹⁸, RAXML⁸⁸, CAFE⁸², DarkHorse²⁴ and ETE3⁹¹. Reproducibility can be accomplished using our own parameters mentioned in Methods.

Ethics approval and consent to participate. All animal experiments in this study were performed under the applicable laws and guidelines for the care and use of laboratory animals, as specified in the Fundamental Guidelines for Proper Conduct of Animal Experiment and Related Activities in Academic Research Institutions under the jurisdiction of the Ministry of Education, Culture, Sports, Science and Technology, Japan, 2006.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All sequence data from the genome projects have been deposited at DDBJ/ENA/GenBank under BioProject accession PRJEB35374 and PRJEB35375. All relevant data are available from the corresponding authors on reasonable request.

Received: 15 September 2020; Accepted: 29 April 2021;

Published online: 31 May 2021

References

- Ijima I. On a New Cestode Larva Parasitic in Man (*Plerocercoides Prolifer*, 1905).
- Stiles, C. W. The occurrence of a proliferating cestode larva (*Sparganum proliferum*) in man in Florida. *Hyg. Lab. Bull.* **40**, 7–18 (1908).
- Meric, R. et al. Disseminated infection caused by *Sparganum proliferum* in an AIDS patient. *Histopathology* **56**, 824–828 (2010).
- Nakamura, T., Hara, M., Matsuoka, M., Kawabata, M. & Tsuji, M. Human proliferative sparganosis: a new Japanese case. *Am. J. Clin. Pathol.* **94**, 224–228 (1990).
- Kikuchi T. & Maruyama H. Human proliferative sparganosis update. *Parasitol. Int.* **75**, 102036 (2019).
- Iwata, S. On the branched plerocercoid (*Sparganum proliferum*) from Japanese snake. *Prog. Med. Parasitol. Jpn* **4**, 587–590 (1972).
- Mueller J. F. & Strano A. J. *Sparganum proliferum*, a sparganum infected with a virus? *J. Parasitol.* 15–19 (1974).
- Kokaze, A. et al. Phylogenetic identification of *Sparganum proliferum* as a pseudophyllidean cestode. *Parasitol. Int.* **46**, 271–279 (1997).
- Miyadera, H. et al. Phylogenetic identification of *Sparganum proliferum* as a pseudophyllidean cestode by the sequence analyses on mitochondrial COI and nuclear sdhB genes. *Parasitol. Int.* **50**, 93–104 (2001).
- Arrabal, J. P., Pérez, M. G., Arce, L. F. & Kamenetzky, L. First identification and molecular phylogeny of *Sparganum proliferum* from endangered felid (*Panthera onca*) and other wild definitive hosts in one of the regions with highest worldwide biodiversity. *Int. J. Parasitology: Parasites Wildl.* **13**, 142–149 (2020).
- Kuchta, R., Scholz, T., Brabec, J., & Narduzzi-Wicht, B. in *Biology of Foodborne Parasites Section III Important Foodborne Parasites* (CRC Press, 2015).
- Reuter, M. & Kreshchenko, N. Flatworm asexual multiplication implicates stem cells and regeneration. *Can. J. Zool.* **82**, 334–356 (2004).
- Thompson, R. A. & Lymbery, A. J. *Echinococcus and Hydatid Disease* (Cab International, 1995).
- Specht, D. & Voge, M. Asexual multiplication of mesocostoides tetrathyridia in laboratory animals. *J. Parasitol.* **51**, 268–272 (1965).
- Yamasaki, H. et al. *Spirometra* species from Asia: genetic diversity and taxonomic challenges. *Parasitol. Int.* **80**, 102181 (2021).
- Kuchta, R. et al. Sparganosis (*Spirometra*) in Europe in the molecular era. *Clin. Infect. Dis.* **72**, 882–890 (2021).
- Bennett, H. M. et al. The genome of the sparganosis tapeworm *Spirometra erinaceieuropaei* isolated from the biopsy of a migrating brain lesion. *Genome Biol.* **15**, 510 (2014).
- Huang, S., Kang, M. & Xu, A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics (Oxf., Engl.)* **33**, 2577–2579 (2017).
- Almeida, G. G., Coscarelli, D., Melo, M. N., Melo, A. L. & Pinto, H. A. Molecular identification of *Spirometra* spp. (Cestoda: Diphyllobothriidae) in some wild animals from Brazil. *Parasitol. Int.* **65**, 428–431 (2016).
- Jeon, H. K. et al. Genetic identification of *Spirometra decipiens* plerocercoids in terrestrial snakes from Korea and China. *Korean J. Parasitol.* **54**, 181–185 (2016).
- Bernt, M. et al. A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. *Mol. Phylogenet. Evol.* **69**, 352–364 (2013).
- Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
- Tsai, I. J. et al. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature* **496**, 57–63 (2013).
- Podell, S. & Gaasterland, T. DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biol.* **8**, R16 (2007).
- Lahti, D. C. et al. Relaxed selection in the wild. *Trends Ecol. Evol.* **24**, 487–496 (2009).
- Wertheim, J. O., Murrell, B., Smith, M. D., Kosakovsky Pond, S. L. & Scheffler, K. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832 (2015).
- Cebria, F. & Newmark, P. A. Morphogenesis defects are associated with abnormal nervous system regeneration following roboA RNAi in planarians. *Development (Camb., Engl.)* **134**, 833–837 (2007).
- Basseres, D. S., Tizzei, E. V., Duarte, A. A., Costa, F. F. & Saad, S. T. ARHGAP10, a novel human gene coding for a potentially cytoskeletal Rho-GTPase activating protein. *Biochem. Biophys. Res. Commun.* **294**, 579–585 (2002).
- Wenemoser, D., Lapan, S. W., Wilkinson, A. W., Bell, G. W. & Reddien, P. W. A molecular wound response program associated with regeneration initiation in planarians. *Genes Dev.* **26**, 988–1002 (2012).
- Lei, K. et al. Egf signaling directs neoblast repopulation by regulating asymmetric cell division in planarians. *Dev. Cell* **38**, 413–429 (2016).
- Ishiyama, M., Miyazono, Y., Sasamoto, K., Ohkura, Y. & Ueno, K. A highly water-soluble disulfonated tetrazolium salt as a chromogenic indicator for NADH as well as cell viability. *Talanta* **44**, 1299–1305 (1997).
- Aslanturk, O. S. *In Vitro Cytotoxicity and Cell Viability Assays: Principles, Advantages, and Disadvantages*. (InTech, 2018).
- Noya, O., de Noya, B., Archedera, H., Torres, J. & Argüello, C. *Sparganum proliferum*: an overview of its structure and ultrastructure. *Int. J. Parasitol.* **22**, 631–640 (1992).
- Klemba, M. & Goldberg, D. E. Biological roles of proteases in parasitic protozoa. *Annu. Rev. Biochem.* **71**, 275–305 (2002).
- Frees, D., Brondsted, L. & Ingmer, H. Bacterial proteases and virulence. *Sub-Cell. Biochem.* **66**, 161–192 (2013).
- Yike, I. Fungal proteases and their pathophysiological effects. *Mycopathologia* **171**, 299–323 (2011).
- Coleman, J. J. & Mylonakis, E. Efflux in fungi: La Pièce de résistance. *PLoS Pathog.* **5**, e1000486 (2009).
- Florin, L. & Lang, T. Tetraspanin Assemblies in Virus Infection. *Front. Immunol.* **9**, 1140 (2018).
- Hunt, V. L., Hino, A., Yoshida, A. & Kikuchi, T. Comparative transcriptomics gives insights into the evolution of parasitism in Strongyloides nematodes at the genus, subclade and species level. *Sci. Rep.* **8**, 5192 (2018).
- Hunt, V. L. et al. The genomic basis of parasitism in the Strongyloides clade of nematodes. *Nat. Genet.* **48**, 299–307 (2016).
- Vadon-Le Goff, S., Hulmes, D. J. S. & Moali, C. BMP-1/tolloid-like proteinases synchronize matrix assembly with growth factor activation to promote morphogenesis and tissue remodeling. *Matrix Biol.* **44–46**, 14–23 (2015).
- Wang, R. N. et al. Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis.* **1**, 87–105 (2014).

43. Taylor, P. The cholinesterases. *J. Biol. Chem.* **266**, 4025–4028 (1991).
44. Sellkirk M., Henson S., Russell W. & Hussein A. in *Parasitic Nematodes: Molecular Biology, Biochemistry and Immunology*, 211–228 (CAB eBooks, 2001).
45. Hunt, V. L., Tsai, I. J., Sellkirk, M. E. & Viney, M. The genome of *Strongyloides* spp. gives insights into protein families with a putative role in nematode parasitism. *Parasitology* **144**, 343 (2017).
46. de Lange, A. et al. *Taenia* larvae possess distinct acetylcholinesterase profiles with implications for host cholinergic signalling. *PLOS Neglect. Tropical Dis.* **14**, e0008966 (2020).
47. Zheng, H. et al. The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nat. Genet.* **45**, 1168–1175 (2013).
48. Li, W. et al. The genome of tapeworm *Taenia multiceps* sheds light on understanding parasitic mechanism and control of coenurosis disease. *DNA Res.* **25**, 499–510 (2018).
49. International Helminth Genomes C. Comparative genomics of the major parasitic worms. *Nat. Genet.* **51**, 163–174 (2019).
50. Kondo, N., Nikoh, N., Ijichi, N., Shimada, M. & Fukatsu, T. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc. Natl Acad. Sci. USA* **99**, 14280–14285 (2002).
51. Aikawa, T. et al. Longicorn beetle that vectors pinewood nematode carries many *Wolbachia* genes on an autosome. *Proc. Biol. Sci.* **276**, 3791–3798 (2009).
52. Walsh, A. M., Kortschak, R. D., Gardner, M. G., Bertozzi, T. & Adelson, D. L. Widespread horizontal transfer of retrotransposons. *Proc. Natl Acad. Sci. USA* **110**, 1012–1016 (2013).
53. Ivancevic, A. M., Kortschak, R. D., Bertozzi, T. & Adelson, D. L. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol.* **19**, 85 (2018).
54. Carroll, S. B. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485 (1995).
55. Moens, C. B. & Selleri, L. Hox cofactors in vertebrate development. *Dev. Biol.* **291**, 193–206 (2006).
56. Olson, P. D. et al. Genome-wide transcriptome profiling and spatial expression analyses identify signals and switches of development in tapeworms. *EvoDevo* **9**, 21 (2018).
57. Gold, D. A., Gates, R. D. & Jacobs, D. K. The early expansion and evolutionary dynamics of POU class genes. *Mol. Biol. Evol.* **31**, 3136–3147 (2014).
58. Kondo, Y. et al. Molecular cloning and characterization of plerocercoid-immunosuppressive factor from *Spirometra erinaceieuropaei*. *Parasitol. Int.* **76**, 102062 (2020).
59. Dirgahayu, P., Fukumoto, S., Tadamoto, S., Kina, Y. & Hirai, K. Excretory/secretory products from plerocercoids of *Spirometra erinaceieuropaei* suppress interleukin-1beta gene expression in murine macrophages. *Int. J. Parasitol.* **34**, 577–584 (2004).
60. Kina, Y. et al. A glycoprotein from *Spirometra erinaceieuropaei* plerocercoids suppresses osteoclastogenesis and proinflammatory cytokine gene expression. *Int. J. Parasitol.* **35**, 1399–1406 (2005).
61. Sakai, T., Larsen, M. & Yamada, K. M. Fibronectin requirement in branching morphogenesis. *Nature* **423**, 876–881 (2003).
62. Wang, S., Sekiguchi, R., Daley, W. P. & Yamada, K. M. Patterned cell and matrix dynamics in branching morphogenesis. *J. Cell Biol.* **216**, 559–570 (2017).
63. Moulinier, R. et al. Human proliferative sparganosis in Venezuela: report of a case. *Am. J. Tropical Med. Hyg.* **31**, 358–363 (1982).
64. Alarcon de Noya, B., Torres, J. R. & Noya, O. Maintenance of *Sparganum proliferum* in vitro and in experimental animals. *Int. J. Parasitol.* **22**, 835–838 (1992).
65. Kokot, M., Dlugosz, M. & Deorowicz, S. KMC 3: counting and manipulating k-mer statistics. *Bioinformatics (Oxf., Engl.)* **33**, 2759–2761 (2017).
66. Vurture, G. W. et al. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics (Oxf., Engl.)* **33**, 2202–2204 (2017).
67. Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M., & Schwander, T. Genomic features of parthenogenetic animals. *Journal of Heredity.* **112**, 19–33 (2021).
68. Kajitani, R. et al. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
69. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics (Oxf., Engl.)* **27**, 578–579 (2010).
70. Parra, G., Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxf., Engl.)* **23**, 1061–1067 (2007).
71. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics (Oxf., Engl.)* **31**, 3210–3212 (2015).
72. Hahn, C., Bachmann, L. & Chevreux, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res.* **41**, e129 (2013).
73. Bernt, M. et al. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol.* **69**, 313–319 (2013).
74. Carver, T., Harris, S. R., Berriman, M., Parkhill, J. & McQuillan, J. A. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics (Oxf., Engl.)* **28**, 464–469 (2012).
75. Han, Y. & Wessler, S. R. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **38**, e199 (2010).
76. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinforma.* **9**, 18 (2008).
77. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics (Oxf., Engl.)* **26**, 2460–2461 (2010).
78. Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics (Oxf., Engl.)* **19**, ii215–ii225 (2003).
79. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360 (2015).
80. El-Gebali, S. et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2018).
81. Gotz, S. et al. High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).
82. De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics (Oxf., Engl.)* **22**, 1269–1271 (2006).
83. Olson, P. D. et al. Complete representation of a tapeworm genome reveals chromosomes capped by centromeres, necessitating a dual role in segregation and protection. *BMC Biol.* **18**, 165 (2020).
84. Wang, S. et al. Comparative genomics reveals adaptive evolution of Asian tapeworm in switching to a new intermediate host. *Nat. Commun.* **7**, 12845 (2016).
85. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite: a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* **215**, 2–10 (2017).
86. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
87. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
88. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxf., Engl.)* **30**, 1312–1313 (2014).
89. Paps, J., Xu, F., Zhang, G. & Holland, P. W. Reinforcing the egg-timer: recruitment of novel lophotrochozoa homeobox genes to early and late development in the Pacific oyster. *Genome Biol. Evol.* **7**, 677–688 (2015).
90. Zhong, Y. F. & Holland, P. W. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. *Evol. Dev.* **13**, 567–568 (2011).
91. Huerta-Cepas, J., Serra, F. & Bork, P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
92. Yang, Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Bioinformatics (Oxf., Engl.)* **13**, 555–556 (1997).
93. Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **34**, W609–W612 (2006).
94. Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
95. Nussbaum-Krammer, C. I., Neto, M. F., Briemann, R. M., Pedersen, J. S., & Morimoto R. I. Investigating the spreading and toxicity of prion-like proteins using the metazoan model organism *C. elegans*. *J. Vis. Exp.* **95**, 52321 (2015).
96. Karperien, A. Fraclac for ImageJ. (Charles Sturt University, 2013).
97. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxf., Engl.)* **31**, 166–169 (2015).
98. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxf., Engl.)* **26**, 139–140 (2010).
99. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562 (2012).

Acknowledgements

Genome data analyses were partly performed using the DDBJ supercomputer system. We thank Ryusei Tanaka, Akemi Yoshida for assistance and comments. This work was supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 26460510 and 19H03212, AMED 18fk0108009h0003 and JST CREST Grant Number JPMJCR18S7.

Author contributions

T. Kikuchi, T. Kuramochi and H.M. conceived the study. T. Kikuchi contributed to study design. V.L.H., H.M. and T. Kikuchi wrote the manuscript with inputs from others. B.A.N., O.N., S.K., Y.K. prepared biological samples. K.I., A.K., S.S., T. Kikuchi and T. Kuramochi conducted experiments. V.L.H., M.D., Y.M., A.T. and T. Kikuchi completed genome assembly and analysed genome data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-021-02160-8>.

Correspondence and requests for materials should be addressed to T.K.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021