The genomes of four tapeworm species reveal adaptations to parasitism

Ishteng J. Tsai1,2, Magdalena Zarowiecki1,4, Nancy Holroyd1, Alejandro Garcia-Rubio1,4, Alejandro Sanchez-Flores1,3, Karen L. Brooks1, Alan Tracey1, Raúl J. Bobes4, Gladis Fragoso4, Edda Sciutto4, Martin Aslett1, Helen Beasley1, Hayley M. Bennett1, Jianping Cai2, Federico Caminca4, Richard Clark4, Marcela Cucher4, Nishadi De Silva1, Tim A. Day3, Peter Deplazes6, Karel Estrada3, Cecilia Fernández4, Peter W. H. Holland10, Junling Hou6, Songnian Hu11, Thomas Huckvale4, Stacy S. Hung4, Laura Kamenetskaya6, Jacqueline A. Keane4, Ferenc Kiss13, Uriel Koziol13, Olivia Lambert4, Kan Liu11, Xuenong Luo5, Yingfeng Luo6, Natalia Macchiarelli3, Sarah Nichol1, Jordi Paps10, John Parkinson12, Natasha Pouchkina-Stantcheva14, Nick Riddiford14,15, Mara Rosenzvit6, Gustavo Salinas6, James D. Wasmuth16, Mostafa Zamanian17, Yadong Zheng5, The Taenia solium Genome Consortium†, Xuepeng Cai2, Xavier Sóberón1,18, Peter D. Olson14, Juan P. Laclette6, Klaus Brehm13 & Matthew Berrian1

Tapeworms (Cestoda) cause neglected diseases that can be fatal and are difficult to treat, owing to inefficient drugs. Here we present an analysis of tapeworm genome sequences using the human-infective species Echinococcus multilocularis, E. granulosus, Taenia solium and the laboratory model Hymenolepis microstoma as examples. The 115– to 141-megabase genomes offer insights into the evolution of parasitism. Synteny is maintained with distantly related blood flukes but we find extreme losses of genes and pathways that are ubiquitous in other animals, including 34 homeobox families and several determinants of stem cell fate. Tapeworms have specialized detoxification pathways, metabolism that is finely tuned to rely on nutrients scavenged from their hosts, and species-specific expansions of non-canonical heat shock proteins and families of known antigens. We identify new potential drug targets, including some on which existing pharmaceuticals may act. The genomes provide a rich resource to underpin the development of urgently needed treatments and control.

Echinococcosis (hydatid disease) and cysticercosis, caused by the proliferation of larval tapeworms in vital organs1, are among the most severe parasitic diseases in humans and account for 2 of the 17 neglected tropical diseases prioritized by the World Health Organization2. Larval tapeworms can persist asymptomatically in a human host for decades3, eventually causing a spectrum of debilitating pathologies and death1. When diagnosed, the disease is often at an advanced stage at which surgery is no longer an option4. Tapeworm infections are highly prevalent worldwide5, and their human disease burden has been estimated at 1 million disability-adjusted life years, comparable with African trypanosomiasis, river blindness and dengue fever. Furthermore, cystic echinococcosis in livestock causes an annual loss of US$2 billion6.

Tapeworms (Platyhelminthes, Cestoda) are passively transmitted between hosts and parasitize virtually every vertebrate species7. Their morphological adaptations to parasitism include the absence of a gut, a head and light-sensing organs, and they possess a unique surface ( tegument) that is able to withstand host-stomach acid and bile but is still penetrable enough to absorb nutrients7.

Tapeworms are the only one of three major groups of worms that parasitize humans, the others being flukes (Trematoda) and round worms (Nematoda), for which no genome sequence has been available so far. Here we present a high-quality reference tapeworm genome of Echinococcus multilocularis. We also present the genomes of three other species, for comparison; E. granulosus (dog tapeworm), Taenia solium (pork tapeworm), both of which infect humans, and Hymenolepis microstoma (a rodent tapeworm and laboratory model for the human parasite Hymenolepis nana). We have mined the genomes to provide a starting point for developing urgently needed therapeutic measures against tapeworms and other parasitic flatworms. Access to the complete genomes of several tapeworms will accelerate the pace at which new tools and treatments to combat tapeworm infections can be discovered.

The genomes and genes of tapeworms

The E. multilocularis genome assembly was finished manually (Supplementary Information, section 2), producing a high-quality reference genome that is a useful starting point for future research.

1Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. 2Division of Parasitology, Department of Infectious Disease, Faculty of Medicine, University of Miyazaki, Miyazaki 889-1692, Japan. 3Institute of Biotechnology, Universidad Nacional Autónoma de México, Cuenavaca, Morelos 62210, México. 4Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 Mexico D.F., Mexico. 5State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of GansuProvince, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No. 1 Kuijaping, Chengguan District, Lanzhou 730046, Gansu Province, China. 6Instituto de Microbiología y Parasitológía Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPA-CONICET), Facultad de Medicina, Paraguay 2155, C1121ABG Buenos Aires, Argentina. 7Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA. 8Institute of Parasitology, Vetsuisse Faculty, University of Zürich, Winterthurerstrasse 266a, CH-8057 Zürich, Switzerland. 9Cátedra de Inmunología, Facultad de Química, Universidad de la República, Avenida Alfredo Navarro 3051, piso 2, Montevideo, CP 11600, Uruguay. 10Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. 11Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Bietucheng West Road, Chaoyang District, Beijing 100029, China. 12Department of Biochemistry & Molecular and Medical Genetics, University of Toronto, Program in Molecular Structure and Function, The Hospital for Sick Children, Toronto, Ontario MSG 1X8, Canada. 13University of Würzburg, Institute of Hygiene and Microbiology, D-97080 Würzburg, Germany. 14Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK. 15Department of Zoology, School of Natural Sciences and Regenerative Medicine (REMEDi), National University of Ireland Galway, University Road, Galway, Ireland. 16Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary T2N 4Z6, Canada. 17Institute of Parasitology, McGill University, 2111 Lakeshore Road, Ste Anne de Bellevue, Quebec H9X 3V9, Canada. 18Instituto Nacional de Medicina Genómica, Periférica Sur No. 4509 Col. Arenal Tepepan, Delegación Tlalpan, 14610 México, D.F., México.

*These authors contributed equally to this work.
†Lists of participants and their affiliations appear at the end of the paper.
genome in which 89% of the sequence is contained in 9 chromosome scaffolds that have only 23 gaps (Supplementary Table 1.2). One chromosome is complete from telomere to telomere, and 13 of the expected 18 telomeres are joined to scaffolds (Fig. 1a). This quality and completeness is comparable to that of the first publications of Caenorhabditis elegans and Drosophila melanogaster genomes. The 115- to 141-megabase (Mb) nuclear tapeworm genomes were sequenced using several high-throughput technologies (Supplementary Table 5.5). Similar to the genome of Schistosoma mansoni of the size of the genome of their distant flatworm relative, the blood fluke Schistosoma mansoni, of the size of the genome of their distant flatworm relative, the blood fluke Schistosoma mansoni, is unusual among flatworms in that it has a sex chromosome. Schistosomes are unusual among flatworms in that they have sexual dimorphism, but how common ancestors of both tapeworms and flukes evolved into female-heterogametic parasites, like S. mansoni, remains to be elucidated.

Genome-wide identification of polycistrons in tapeworms shows that there are 308 putative polycistrons in E. multilocularis, with the largest containing 4 genes. The internal gene order within E. multilocularis polycistrons is largely the same as in T. solium and H. microstoma (Supplementary Table 6.5), and—to some extent—as in flukes; 39% of S. mansoni orthologues of genes within E. multilocularis polycistrons retain colinearity. Of these S. mansoni genes, 40% have transcriptome evidence supporting their polycistronic transcription, demonstrating further that gene order in polycistrons is highly conserved over long evolutionary time (P < 0.0001, Supplementary Information, section 6).

Polycistrons are resolved into individual coding transcripts using spliced-leader trans-splicing, but spliced-leader trans-splicing also occurs in genes outside of polycistrons. Using deep transcriptome sequencing (RNA-seq) we found evidence of spliced-leader trans-splicing in approximately 13% of E. multilocularis genes (Supplementary Table 6.2), less than the 70% observed in C. elegans and 58% in a tunicate.

**Specialized metabolism and detoxification**

The high-confidence gene sets reveal extensive reductions in overall metabolic capability and an increased ability to absorb nutrients, compared to that of other animals (Figs 2 and 3, and Supplementary Information, section 9). Their main energy source, carbohydrates, can be catabolized by aerobic respiration or by two complementary anaerobic pathways; the lactate fermentation and malate dismutation pathways. The parasiticidal effects of mitochondrial fumarate reductase inhibitors have been demonstrated in vitro, suggesting that the malate dismutation pathway would be an effective target for the development of novel therapeutics.

Tapeworms, like flukes, lack the ability to synthesize fatty acids and cholesterol de novo. Instead, they scavenge essential fats from the host using fatty acid transporters and lipid elongation enzymes (Supplementary Table 9.2), as well as several tapeworm-specific gene families, such as fatty acid binding protein (FABP) and the apolipoprotein...
antigen B (Supplementary Information, section 8). Uptake of fatty acids seems to be crucial in *Echinococcus* spp. metacestodes, in which both FABP and antigen B gene families are among the most highly expressed genes\(^1\) (Supplementary Table 5.7). Tapeworms and flukes have lost many genes associated with the peroxisome (Supplementary Information, section 8), an organelle in which fatty acid oxidation occurs, and may lack peroxisomes altogether, as seen in several other parasites\(^2\).

Compared with other animals, *S. mansoni* has a reduced ability to synthesize amino acids\(^7\). In tapeworms, this capacity is reduced further, with serine and proline biosynthesis enzymes absent from *E. multilocularis* (Fig. 3 and Supplementary Information, section 9). Many enzymes in the molybdopterin biosynthesis pathway seemed to be lost in tapeworms, along with enzymes that use molybdopterin as a cofactor. The ability to utilize molybdenum in enzymatic reactions was believed to be present in all animals\(^21\), but has been lost in some eukaryotic parasites\(^22\).

Differences in the detoxification systems between tapeworms and their mammalian hosts may be exploited for drug design (Supplementary Information, section 9). We found that, like flukes\(^5\), tapeworms typically have only one cytochrome P450 gene, suggesting that their ability to oxidize many xenobiotics and steroids is substantially lower than that of their hosts. Uniquely, tapeworms and flukes have merged two key enzymatic functions for redox homeostasis in one single enzyme: thioredoxin glutathione reductase (TGR). TGR is an essential gene and validated drug target in flukes\(^24\). Downstream of TGR we find an unexpected diversity of thioredoxins, glutaredoxins and mu-class glutathione S-transferases (GSTs) (Supplementary Table 9.3). The GST expansion suggests that tapeworms would be able to water-solubilize and excrete a large range of hydrophobic compounds, which may add complexity to the pharmacokinetics of drugs.

**Homeobox gene loss**

Homeobox genes are high-level transcription factors that are implicated in the patterning of body plans in animals. Across parasitic flatworms, the homeobox gene numbers are extensively reduced (Supplementary Table 10.1). Most bilaterian invertebrates have a conserved set of approximately 100 homeobox genes (for example, 92 conserved in *C. elegans*, 102 in *D. melanogaster*, and 133 in the lancelet)\(^9\). Of the 96 homeobox gene families that are thought to have existed at the origin of the Bilateria, 24 are not present in tapeworms and flukes, and a further 10 were lost in tapeworms, making their complement by far the most reduced of any studied bilaterian animal\(^25\). Among the tapeworm-specific gene losses are gene families involved in neural development (indicated by 'H'), metabolic proteases and amino acid biosynthesis (7). In tapeworms, gains and losses of genomic traits include innovation of bimodal intron distribution and novel fatty acid transporters (8), expansion of mu-class glutathione S-transferases, GP50 antigens and tetraspanins (9), loss of the molybdopterin biosynthesis pathway, loss of 10 homeobox gene families (10), fewer GPCRs and fewer neuropeptides encoded by each proteptide (11), and expansion of heat shock proteins (Hsp) and species-specific antigens (12).

![Figure 2](https://example.com/figure2.png)

**Figure 2 | Evolution of tapeworm parasitism.** Phylogeny of the main branches of Bilateria; Ecdysozoa (including fruitflies and nematodes), Deuterostomia (including lancelet, zebrafish, mice and humans), and Lophotrochozoans (including Platychelminthes (flatworms)) (based on phylogeny in Supplementary Fig. 7.1). The gains and losses of life-cycle traits for these parasitic flatworms include the evolution of endoparasitism (a), passive transmission between hosts (b), acquisition of vertebrate intermediate host (c), ability to proliferate asexually in intermediate host (d). Morphological traits that have evolved include the loss of eye cups (e), gain of neodermatan syncretic epithelia (f), loss of gut (g), segmentation of body plan (h), and changes in the laminated layer (to contain specialized aponemotaxis). I). Gains and losses of genomic traits include spliced leader transcript splicing (1), loss of Wnt genes (2), loss of NEK kinases, fatty acid biosynthesis and ParaHox genes (3), anaerobic metabolic ability through the malate dismutation/rodhoquinone pathway, merger of glutaredoxin and thioredoxin reductase to thioredoxin glutathione reductase (TGR) (4), evolution of tapeworm- and fluke-specific Argonaute (Ago) family, micro exon genes (MEGs) and PROF1 GPCRs (5), loss of peroxisomal genes (6), and complete loss of vasa, tudor and piwi genes, NF-κB pathway, loss of 24 homeobox gene families (indicated by ‘H’), metabolic proteases and amino acid biosynthesis (7). In tapeworms, gains and losses of genomic traits include innovation of bimodal intron distribution and novel fatty acid transporters (8), expansion of mu-class glutathione S-transferases, GP50 antigens and tetraspanins (9), loss of the molybdopterin biosynthesis pathway, loss of 10 homeobox gene families (10), fewer GPCRs and fewer neuropeptides encoded by each proteptide (11), and expansion of heat shock proteins (Hsp) and species-specific antigens (12).
Stem cell specializations

Extreme regenerative capability and developmental plasticity, mediated by ever-present somatic stem cells (neoblasts), have made flatworms popular models for stem cell research\(^29\). All multicellular organisms rely on stem cells for proliferation and growth, so it is remarkable that tapeworms and flukes appear to lack the ubiquitous stem cell marker gene \textit{vasa} (Supplementary Information, section 11). Instead tapeworms have two copies of another dead-box helicase (PL10), which we propose may have taken over some of the functions of \textit{vasa} (Supplementary Information, section 11). Tapeworms and flukes are also missing the \textit{vasa} gene subfamily and \textit{piwi}-interacting \textit{tudor}-domain containing proteins. The \textit{piwi} genes belong to a subfamily of genes encoding argonaute proteins, and we also found that tapeworms have a new subfamily of argonaute proteins (Supplementary Fig. 11.2) that may bind a newly discovered potential small RNA precursor\(^{30}\). Both \textit{piwi} and \textit{vasa} are usually essential in regulating the fate of germline stem cells in animals, and \textit{vasa} suppression usually leads to infertility\(^{31}\). These findings suggest that stem-cell-associated pathways in parasitic flatworms may be highly modified.

Specialization of the tapeworm proteome

We sought to identify novel and expanded gene families in tapeworms, and found many frequently occurring novel domains involved in cell–cell adhesion and the formation of the tegument (Supplementary Information, section 8). For example, several novel domains are found on the ectodomain of cadherins (Supplementary Information, section 8). For example, several novel domains are found on the ectodomain of cadherins (Supplementary Information, section 8). Approximately 20% of the genes are exclusive to tapeworms, and these include many highly expressed antigen families, such as antigen B, the glycosylphosphatidylinositol (GPI)-anchored protein GP50 (ref. 34), and the vaccine target EG95 (ref. 35) (Supplementary Table 12.4).

One of the most striking gene family expansions in the tapeworm genomes is the heat shock protein 70 (Hsp70) family. Phylogenetic analysis revealed independent and parallel expansions in both the Hsp110 and the cytosolic Hsp70 clades (Fig. 4). Several examples of expansions exist at various clades of Hsp70 in other systems, including Hsp110 expansions in oysters (to cope with temperature) and in cancer cells (to cope with proteotoxic stress)\(^{36,37}\). \textit{Echinococcus} and \textit{T. solium} have the highest number of gene expansions in the cytosolic Hsp70 clade. These expansions seem to have occurred independently in each species, and have resulted in 22 to 32 full copies in each species (\textit{Echinococcus} and \textit{T. solium}) compared to 6 copies in fruitflies and 2 in humans (Fig. 4). This expanded clade lacks classical cytosolic Hsp70 features (a conserved EEVD motif for substrate binding and a GGMP repeat unit), and whereas the canonical cytosolic hsp70 genes are constitutively expressed in different life-cycle stages, the non-canonical genes show almost no expression, suggesting a putative few morphological traits that differ between the very closely related species \textit{E. granulosus} and \textit{E. multilocularis}. We identified corresponding species differences in an \textit{Echinococcus}-specific apomucin family (Supplementary Fig. 12.1), an important building block of the laminated layer\(^33\). One particular copy is highly differentiated between the two species (non-synonymous to synonymous substitution ratio of >1) and is the fifth most highly expressed in the metacestode stage of \textit{E. multilocularis} (Supplementary Table 5.7). Galactosyltransferases that probably decorate the apomucins with galactose residues, the predominant sugar of laminated layer glycans, are similarly diverged\(^{34}\) (Supplementary Information, section 8). Approximately 20% of the genes are exclusive to tapeworms, and these include many highly expressed antigen families, such as antigen B, the glycosylphosphatidylinositol (GPI)-anchored protein GP50 (ref. 34), and the vaccine target EG95 (ref. 35) (Supplementary Table 12.4).

One of the most striking gene family expansions in the tapeworm genomes is the heat shock protein 70 (Hsp70) family. Phylogenetic analysis revealed independent and parallel expansions in both the Hsp110 and the cytosolic Hsp70 clades (Fig. 4). Several examples of expansions exist at various clades of Hsp70 in other systems, including Hsp110 expansions in oysters (to cope with temperature) and in cancer cells (to cope with proteotoxic stress)\(^{36,37}\). \textit{Echinococcus} and \textit{T. solium} have the highest number of gene expansions in the cytosolic Hsp70 clade. These expansions seem to have occurred independently in each species, and have resulted in 22 to 32 full copies in each species (\textit{Echinococcus} and \textit{T. solium}) compared to 6 copies in fruitflies and 2 in humans (Fig. 4).
Novel drug targets

Tapeworm cysts are treated by chemotherapy or surgical intervention depending on tapeworm species, patient health and the site of the cyst. The only widely used drugs to treat tapeworm cysts are benzimidazoles that, owing to considerable side effects, are administered at parasitic rather than parasiticide concentrations. Novel targets and compound classes are therefore urgently needed.

To identify new potential drug targets, we surveyed common targets of existing pharmaceuticals; kinases, proteases, G-protein-coupled receptors (GPCRs) and ion channels. We identified approximately 250 to 300 new protein kinases (Supplementary Table 13.1), and these cover most major classes (Supplementary Information, section 13). We also identified 151 proteases and 63 peptidase-like proteins in *E. multilocularis*, a repertoire of similar diversity to *S. mansoni*, and found that, like *S. mansoni*, *E. multilocularis* has strongly reduced copy numbers compared to those of other animals (Supplementary Table 13.9). Many successful anthelminthic drugs target one of several different forms of neural communication. We therefore mapped the signalling pathways of the serotonin and acetylcholine neurotransmitters, predicted conserved and novel neuropeptides (Supplementary Table 13.2), and classified more than 60 putative GPCRs (Supplementary Table 13.2) and 31 ligand-gated ion channels (Supplementary Table 13.4). A voltage-gated calcium channel subunit—the proposed target of praziquantel—is not expressed in cysts and thus provides a putative explanation for the drug’s low efficacy.

We searched databases for potential features for target selection, including compounds associated with protein targets and expression in the clinically relevant metacestode life-stage, and using this information we assigned weights to rank the entire proteomes (Supplementary Table 13.10). We identified 1,082 *E. multilocularis* proteins as potential targets, and of these, 150 to 200 with the highest scores have available chemical leads (known drug or approved compounds).

Acetylcholinesterases, which are inhibited by mefloquine (an antimalarial that reduces egg production in *S. mansoni*), are high on the list of potential targets. However, acetylcholinesterase transcription in tapeworm cysts is low, possibly limiting their suitability. After filtering to remove targets with common substrates rather than inhibitors, the top of the list includes several homologues of targets for...
cancer chemotherapy, including cascin kinase II, ribonucleoside reductase, UMP–CMP kinase and proteasome subunits (Table 1). The challenges of inhibiting cancer tumours and metacestodes (particularly those of E. multilocularis) with drugs are in some ways similar; both show uncontrolled proliferation, invasion and metastasis, and are difficult to kill without causing damage to the surrounding tissue. Therefore, metacestodes may be vulnerable to similar strategies as cancer; suppression of mitosis, induction of apoptosis and prevention of DNA replication. In fact, the anthelmintic medicines niclosamide, mebendazole and albendazole have already been shown to inhibit cancer growth.

Conclusion

Tapeworms were among the first known parasites of humans, recorded by Hippocrates and Aristotle in BC. But, a safe and efficient cure to larval tapeworm infection in humans has yet to be found. These genomes provide hundreds of potential drug targets that can be tested using high-throughput drug screenings that were made possible by recent advances in axenic and cell culturing techniques. Flatworms display an unusually high degree of developmental plasticity. In this study, the high level of sequence completion enabled both gene losses and gains to be accurately determined, and has shown how this plasticity has been put to use in the evolution of tapeworms.

METHODS SUMMARY

Genome sequencing was carried out using a combination of platforms. RNA sequencing was performed with illumina RNA-seq protocols for E. multilocularis, E. granulosus and H. microstoma or capillary sequencing of full-length complementary DNA libraries (T. solium). The complete genome annotation is available at http://www.genedb.org. The tapeworm genome projects were registered under the INSDC project IDs PRJEB121 (E. multilocularis), PRJEB121 (E. granulosus), PRJEB124 (H. microstoma) and PRJNA16816 (T. solium). Sequence data for T. solium isolate (from Mexico) were used for all orthologue comparisons, but results relating to gene gains and losses were reconciled against an additional sequenced isolate from China (unpublished). All experiments involving jirds (laboratory host of E. multilocularis) were carried out in accordance with European and German regulations relating to the protection of animals. Ethical approval of the study was obtained from the ethics committee of the government of Lower Franconia (621-2531.01-2/05). Experiments with dogs were conducted under the Swiss guidelines for animal experimentation and approved by the Ethics Committee of the Canton of Thurgau. Experiments described in this article were generated by differential splicing of micro-exon gene transcripts. Genome Res. 20, 1112–1121 (2010).

Qian, W. & Zhang, J. Evolutionary dynamics of nematode operons: easy come, slow go. Genome Res. 18, 412–421 (2008).

Allen, M. A., Hillier, L. W., Waterson, R. H. & Blumenthal, T. A. global analysis of C. elegans trans-splicing. Genome Res. 21, 255–264 (2011).

Matsumoto, J. et al. High-throughput sequence analysis of C. intestinalis SL trans-spliced miRNAs: alternative expression modes and gene function correlates. Genome Res. 20, 636–645 (2010).

Matsumoto, J. et al. Anaerobic NADH-fumarate reductase system is predominant in the respiratory chain of Echinococcus multilocularis, providing a novel target for the chemotherapy of alveolar echinococcosis. Antimicrob. Agents Chemother. 52, 156–179 (2008).

Berriman, M. et al. The genome of the blood fluke Schistosoma mansoni. Nature 460, 352–358 (2009).

Frayha, G. J. Comparative metabolism of acetate in the taeniid tapeworms Echinococcus granulosus, E. multilocularis and Taenia hydatigena. Comp. Biochem. Physiol. B 39, 167–170 (1971).

Obal, G. et al. Characterisation of the native lipid moiety of Echinococcus granulosus antigen B. PLoS Negl. Trop. Dis. 6, e1642 (2012).

Kaasch, A. J. & Joiner, K. A. Targeting and subcellular localization of Toxoplasma gondii catalase. Identification of peroxisomes in an apicomplexan parasite. J. Biol. Chem. 275, 1112–1118 (2000).

Schwarz, G. & Mendel, R. R. Molybdenum cofactor biosynthesis and molybdenum enzymes. Annu. Rev. Plant Biol. 57, 623–647 (2006).

Zhang, Y., Rump, S. & Gladyshev, V. N. Comparative Genomics and Evolution of Molybdenum Utilization. Coord. Chem. Rev. 255, 1206–1217 (2011).

Pakharukova, M. Y. et al. Cytochrome P450 in fluke Opisthorchis felineus: identification and characterization. Mol. Biochem. Parasitol. 181, 190–194 (2012).

Kwon, S. N. & Shin, Y. K. Thioredoxin glutathione reductase from Schistosoma mansoni: an essential parasite enzyme and a key drug target. PLoS Med. 4, e206 (2007).

Zhong, Y. F. & Holland, P. W. HomeODB2: functional expansion of a comparative homeobox genome database for evolutionary developmental biology. Evol. Dev. 13, 557–568 (2011).

Holland, P. W. Beyond the Hox: how widespread is homeobox gene clustering? J. Anat. 199, 13–23 (2001).

Brooke, N. M., Garcia-Fernandez, J. & Holland, P. W. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature 392, 920–922 (1998).

Ruddiford, N. & Olson, P. D. Wnt gene loss in flatworms. Dev. Genes Evol. 221, 187–197 (2011).

Brehm, K. Echinococcus multilocularis as an experimental model in stem cell research and molecular host-parasite interaction. Parasitology 137, 537–555 (2010).

Parkinson, J. et al. A transcriptomic analysis of Echinococcus granulosus larval stages: implications for parasite biology and host adaptation. PLoS Negl Trop. Dis. 6, e1897 (2012).

Raz, E. The function and regulation of vasa-like genes in germ-cell development. Trends Genet. 19, 582–588 (2003).

Diaz, A. et al. Understanding the laminated layer of larval Echinococcus I: structure. Trends Parasitol. 27, 204–213 (2011).

Hancock, K. et al. Characterization and cloning of GP50, a Taenia solium antigen diagnostic for cysticercosis. Mol. Biochem. Parasitol. 133, 115–124 (2004).

Heath, D. D., Jensen, O. & Lightowler, M. W. Progress in control of hydatidosis using vaccination - a review of formulation and delivery of the vaccine and recommendation for practical use in control programmes. Acta Trop. 85, 133–143 (2003).

Zheng, G. et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54 (2012).

Subjecjek, J. & Repasky, E. A. Heat shock proteins and cancer therapy: the trail grows hotter! Oncotarget 2, 433–434 (2011).

Vargas-Parada, L., Solis, C. F. & Laclette, J. P. Heat shock and stress response of Taenia solium and T. crassiceps (Cestoda). Parasitology 122, 583–588 (2001).

Hempfih, A. et al. Echinococcus metacestodes as laboratory models for the screening of drugs against cestodes and trematodes. Parasitology 137, 589–587 (2010).

Brunetti, E. et al. Characterization and cloning of GP50, a Taenia solium antigen diagnostic for cysticercosis. Mol. Biochem. Parasitol. 133, 115–124 (2004).

Heath, D. D., Jensen, O. & Lightowler, M. W. Control of hydatidosis using vaccination - a review of formulation and delivery of the vaccine and recommendation for practical use in control programmes. Acta Trop. 85, 133–143 (2003).

Zheng, G. et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490, 49–54 (2012).

Subjecjek, J. & Repasky, E. A. Heat shock proteins and cancer therapy: the trail grows hotter! Oncotarget 2, 433–434 (2011).

Vargas-Parada, L., Solis, C. F. & Laclette, J. P. Heat shock and stress response of Taenia solium and T. crassiceps (Cestoda). Parasitology 122, 583–588 (2001).

Hempfih, A. et al. Echinococcus metacestodes as laboratory models for the screening of drugs against cestodes and trematodes. Parasitology 137, 589–587 (2010).
Author Contributions I.J.T., Ma.Z., N.H. and M.B. Wrote the manuscript; M.B., K.B., J.P.L., X.S., and X.C. conceived and designed the project; R.J.B., P.D., C.F., T.H., J.H., K.L., X.L., S.H., N.M., P.D.O., M.R., E.S., N.P.S., Y.Z. and X.S. and X.C. examined neuronal signalling; Ma.Z. examined drug targets; and G.S., M.R., C.F., K.B., P.W.H.H., P.D.O., AG, R.J.B., G.F., E.S., X.S., J.P.L. and J.C. commented on the manuscript drafts.

Author Information The tapeworm genome projects were registered under the INSDC project IDs PRJEB122 (E. multilocularis), PRJEB121 (E. granulosus), PRJEB124 (H. microstoma) and PRJNA16816 (T. solium). Illumina and 454 data are released to the European Nucleotide Archive (http://www.ebi.ac.uk/ena/) under accession numbers ERP000331, ERP000452 and PRJNA16816. Capillary data is at http://www.ncbi.nlm.nih.gov/Traces/trace.cgi, SEQ_LIB_ID 98488, 98489 and 101760, CENTER NAME SC (E. multilocularis) and sg1, sg2, sg3, sg4 and sg5 (T. solium, Mexico).

Genome data are available from http://www.sanger.ac.uk/resources/downloads/helminths/ (E. multilocularis, E. granulosus and H. microstoma) and http://www.taeniasolium.unam.mx/taenia/ (T. solium). The complete genome annotation is available at http://www.genedb.org. All RNA-seq data were released to ArrayExpress under accession numbers E-ERAD-50 or E-ERAD-56. T. solium EST sequences were released to http://www.ncbi.nlm.nih.gov/nucest/ under accession numbers ERP000351, ERP000452 and PRJNA16816. Capillary data is at http://www.ncbi.nlm.nih.gov/Traces/trace.cgi, SEQ_LIB_ID 98488, 98489 and 101760, CENTER NAME SC (E. multilocularis) and sg1, sg2, sg3, sg4 and sg5 (T. solium, Mexico).

The Taenia solium Genome Consortium

Alejandro Garcíarrubio1, Raúl J. Bobes2, Gladis Fragoso2, Alejandro Sánchez-Flores1, Karel Estrada1, Miguel A. Cevallos3, Enrique Morett1, Víctor González2, Tobias Portillo1, Adrian Ochoa-Leyva2, Marco V. José2, Edith Sciuoto2, Abraham Landa2, Lucía Jiménez2, Víctor Valdés2, Julio C. Carrero2, Carlos Larralde2, Jorge Morales-Montor2, Jorge Limón-Lason2, Xavier Soberón1,4 & Juan P. Laclette2

1Institute of Biotechnology, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México. 2Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México, D.F. México. 3Genomic Sciences Center, Universidad Nacional Autónoma de México, Cuernavaca, Morelos 62210, México. 4Instituto Nacional de Medicina Genómica, Periférico Sur No. 4809 Col. Arenal Tepepan, Delegación Tlalpan, 14610 México, D.F. México. "School of Medicine, Universidad Nacional Autónoma de México, 04510 México, D.F. México. School of Sciences, Universidad Nacional Autónoma de México, 04510 México, D.F. México.