Published in final edited form as:
Nature. 2013 April 4; 496(7443): 57–63. doi:10.1038/nature12031.

The genomes of four tapeworm species reveal adaptations to parasitism

A full list of authors and affiliations appears at the end of the article.

[#] These authors contributed equally to this work.

Summary

Tapeworms cause debilitating neglected diseases that can be deadly and often require surgery due to ineffective drugs. Here we present the first 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 115-141 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 ubiquitous in other animals, including 34 homeobox families and several determinants of stem cell fate. Tapeworms have species-specific expansions of non-canonical heat shock proteins and families of known antigens; specialised detoxification pathways, and metabolism finely tuned to rely on nutrients scavenged from their hosts. We identify new potential drug targets, including those on which existing pharmaceuticals may act. The genomes provide a rich resource to underpin the development of urgently needed treatments and control.

Keywords

HSP70; parasitism; Cestoda; cysticercosis; echinococcosis; Platyhelminthes

Corresponding author: Matthew Berriman (mb4@sanger.ac.uk), Klaus Brehm (kbrehm@hygiene.uni-wuerzburg.de) or Juan Pedro Laclette (laclette@unam.mx).

The *Taenia solium* Genome Consortium

Alejandro Garciarrubio[1], Raúl J. Bobes[2], Gladis Fragoso[2], Alejandro Sánchez-Flores[1], Karel Estrada[1], Miguel A. Cevallos[3], Enrique Morett[1], Víctor González[3], Tobias Portillo[1], Adrian Ochoa-Leyva[6], Marco José[2], Edda Sciutto[1], Abraham Landa[4], Lucía Jiménez[4], Víctor Valdés[5], Julio C. Carrero[2], Carlos Larralde[2], Jorge Morales-Montor[2], Jorge Limón-Lason[2], Xavier Soberón[1, 6] and Juan P. Laclette[2]

[1] Institute of Biotechnology, [2]Institute of Biomedical Research, [3]Genomic Sciences Center, [4]School of Medicine and [5]School of Sciences, Universidad Nacional Autónoma de México; 04510 México D.F., [6]National Institute of Genomic Medicine, Ministry of Health, 01900, México D.F., México.

Contributions

Wrote the manuscript: I.J.T., Ma.Z., N.H., M.B. Project conception and design: M.B., K.B., J.P.L., X.S., X.P.C. Project coordination: N.H., M.B, Ma.Z. Parasite material and nucleic acid preparation: R.J.B., P.D., C.F., T.H., J.L.H., K.L., X.N.L., S.N.H., N.M., P.D.O., M.R., E.S., N.P.S., Y.Z., K.B. Genome assembly: I.J.T., Ma.Z., A.G., K.E., A.S.F. Genome assembly improvement: I.J.T., K.L.B., A.T., H.B., S.N., T.H., A.G., K.E. Gene predictions: I.J.T., Ma.Z., A.S.F., X.S., K.L., J.L.H, A.G., K.E. Gene annotation: I.J.T., Ma.Z., K.L.B., A.T., H.B., O.L., S.N., R.C., R.J.B., G.F., E.S., X.S., J.P.L. K.L., J.L.H., A.G., K.E., S.N.H., X.P.C. Data processing, computational and bioinformatics support: N.D.S., M.A., J.A.K., K.E., J.L.H., S.N.H., X.S., Y.Z. Genome structure, comparative genomics & ploidy: I.J.T. Gene structure: A.S.F., I.J.T., Ma.Z. Experimental validation of micro-exons: T.H., H.M.B. Trans-splicing and polycistrons: A.G., K.B., F.K., I.J.T. Metabolism and detoxification: I.J.T., S.S.H., J.P., G.S., Ma.Z. Homeobox gene loss: P.W.H.H., J.Paps, N.R., P.D.O. Stem cell specialisations: K.B., I.J.T, Ma.Z. Domains: Ma.Z., J.D.W. Tapeworm specific genes and expansions: I.J.T., Ma.Z., C.F., G.S. Kinases and proteases: Ma.Z. GPCRs: T.A.D., Mo.Z. Neuropeptides: Mo.Z., T.A.D., U.K., K.B. Neuronal signalling: M.R., L.K., F.C., M.C. Drug targets: Ma.Z. Commented on the manuscript drafts: G.S., M.R., C.F., K.B., P.W.H.H., P.D.O., A.G., R.J.B., G.F., E.S., X.S., J.P.L., J.P.C.

Introduction

Echinococcosis (hydatid disease) and cysticercosis, caused by the proliferation of larval tapeworms in vital organs[1], are amongst the most severe parasitic diseases in humans and account for 2 of the 17 Neglected Tropical Diseases prioritised by the World Health Organization[2]. Larval tapeworms can persist asymptomatically in a human host for decades[3], eventually causing a spectrum of debilitating pathologies and death[1]. When diagnosed, the disease is often at an advanced stage when surgery is no longer an option[4]. Tapeworm infections are highly prevalent worldwide[5], and their human disease burden has been estimated at 1 million disability-adjusted life years, comparable with African trypanosomiasis, river blindness and dengue. Furthermore, cystic echinococcosis in livestock causes an annual loss of $2 billion[6].

Tapeworms (Platyhelminthes, Cestoda) are passively transmitted between hosts and parasitise virtually every vertebrate species[7]. Their morphological adaptations to parasitism include the absence of a gut, head and light sensing organs and a unique surface (tegument) able to withstand host-stomach acid and bile, yet penetrable enough to absorb nutrients[7].

Tapeworms (Cestoda) are the only one of three major groups of human-parasitic worms, the others being flukes (Trematoda) and round worms (Nematoda), for which no genome sequence has been available. Here we present a high quality reference tapeworm genome of a human-infective fox tapeworm *Echinococcus multilocularis.* We also present three comparator genomes: *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 *H. 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 in 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 S2), producing a high quality reference genome where 89% of the sequence is contained in 9 chromosome scaffolds containing only 23 gaps (Supplementary Table S1.2). One chromosome is complete from telomere to telomere and 13 of the expected 18 telomeres are joined to scaffolds (Figure 1A). This quality and completeness is comparable to the first published *C. elegans* and *D. melanogaster* genomes[8,9]. The 115-141 megabase (Mb) nuclear tapeworm genomes were assembled using several high-throughput sequencing technologies (Supplementary Tables S1.1). The tapeworm genomes are approximately one-third of the size of the genome of their distant flatworm relative – the blood fluke *Schistosoma mansoni*[10] – mainly due to fewer repeats (Supplementary Information S3). By sequencing multiple isolates of *E. multilocularis* (Supplementary Table S3.2), we revealed tetraploidy in protoscoleces of one isolate, and a trisomy of chromosome 9 (the smallest chromosome, and possibly the only one for which a trisomy is tolerated) transiently exhibited in protoscoleces and metacestodes from two different isolates (Figure 1C and 1D, Supplementary Figures S3.1 S3.2 and S3.3), consistent with previous observations of karyotype plasticity in flatworms[11].

Aided by deep transcriptome sequencing from multiple lifecycle stages we identified 10,231-12,490 putative genes per genome (Supplementary Table S5.5). Like *S. mansoni*[12] distinct “micro-exon genes” are present in tapeworm genomes, with multiple internal exons that are small (typically <36 bases) and divisible by three (Supplementary Information S5).

To identify gene gain and loss in tapeworms, orthologous relationships were predicted between tapeworms and eight other species (Figure 2). Although gene order has been lost, ancient chromosomal synteny is preserved amongst parasitic flatworms (Figure 1B and Supplementary Table S7.3). Two chromosomes in *E. multilocularis* (Figure 1A and Figure 1B) correspond to the *S. mansoni* Z sex chromosome. Schistosomes are unusual, having distinctive sexual dimorphism but how common ancestors of both tapeworms and flukes evolved into female heterogametic parasites like *S. mansoni* remains to be elucidated.

We report the first genome wide identification of polycistrons in tapeworms, finding 308 putative polycistrons in *E. multilocularis*, the largest containing four genes. Their internal gene order is largely shared with *T. solium* and *H. microstoma* (Supplementary Table S6.5), and even extends to 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[10], further demonstrating that gene order in polycistrons is highly conserved over long evolutionary time[13] (p-value<0.0001, Supplementary Information S6).

Polycistrons are resolved into individual coding transcripts using spliced-leader (SL) *trans*-splicing, but SL *trans*-splicing also occurs in genes outside of polycistrons. Using deep RNA-Seq we found evidence of SL trans-splicing in ~13% of *E. multilocularis* genes (Supplementary Table S6.2), less than the 70% observed in *C. elegans*[14] and 58% in a tunicate[15].

Reduced metabolic versatility and specialised detoxification

The high confidence gene sets reveal extensive reductions in overall metabolic capability compared with other animals, combined with an increased ability to absorb nutrients from the host (Figure 2 and 3, Supplementary Information S9). Their main energy source, carbohydrates, can be catabolised by aerobic respiration or to two complementary anaerobic pathways, lactate fermentation and malate dismutation. 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[16].

Tapeworms, like flukes, lack the ability to synthesise fatty acids and cholesterol *de novo*[17,18]. Instead, they scavenge essential fats from the host using fatty acid transporters and lipid elongation enzymes (Supplementary Table S9.2), along with several tapeworm-specific gene families (Supplementary Information S8) Uptake of fatty acids seems to be crucial in *Echinococcus* spp. metacestodes, where fatty acid binding protein (FABP) and Antigen B are amongst the most highly expressed genes[19] (Supplementary Table S5.7). Tapeworms and flukes have lost many genes associated with the peroxisome (Supplementary Information S8), an organelle in which fatty acid oxidation occurs, and may lack peroxisomes altogether, as seen in several other parasites[20].

Tapeworm capability of amino acid synthesis was even further reduced than in *S. mansoni*[17], with serine and proline biosynthesis enzymes absent from *E. multilocularis* (Figure 3, Supplementary Information S9). 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 utilise 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 S9). We found that, like flukes[23], tapeworms typically have only one cytochrome P450 gene, suggesting their ability

to oxidise many xenobiotics and steroids is substantially reduced compared to 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 (Supplementary Table S9.3). The GST expansion suggests that tapeworms would be able to water-solubilise 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 implicated in the patterning of body plans in animals. Across parasitic flatworms, the homeobox gene numbers are extensively reduced (Supplementary Table S10.1). Most bilaterian invertebrates have a conserved set of ~100 homeobox genes (e.g., *C. elegans* 92, *D. melanogaster* 102, lancelet 133)[25]. Of the 96 homeobox gene families inferred to exist 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]. Amongst the tapeworm-specific gene losses are genes involved in neural development (Mnx, Pax3/7, Gbx, Hbn and Rax), which is somewhat surprising considering that tapeworms possess a well-developed nervous system, albeit with reduced sensory input and cephalisation. Tapeworms also lack the ParaHox genes (Gsx, Pdx, Cdx) ancestrally involved in specification of a through-gut[26,27] although these appear to have been lost before the tapeworm gut was lost. Other conserved genes found in bilaterian developmental pathways such as Hedgehog and Notch were found to be present and intact, although the Wnt complement is greatly reduced as compared to the ancestral (spiralian) complement of 12 Wnt ligands[28] (Supplementary Tables S10.2).

Stem cell specialisations

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 lack the ubiquitous stem-cell marker vasa (Supplementary Information S11). Instead they have two copies of another dead-box helicase (PL10), which we hypothesise may have taken over some of vasa's functions (Supplementary Figure S11.1). Tapeworms and flukes are also missing the piwi sub-family of the argonaute proteins and piwi-interacting tudor-genes. In addition, they have a new sub-family of argonautes (Supplementary Figure S11.2), which may bind a newly discovered potential small RNA precursor[30]. Both piwi and vasa are usually essential in regulating the fate of germline stem cells in animals, and vasa suppression usually leads to infertility or death[31]. These findings suggest that stem cell associated pathways in parasitic flatworms may be highly modified.

Specialisation 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 S8). For instance, several novel domains are found on the ectodomain of cadherins (Supplementary Information S8), and tapeworms have proportionally more tetraspanin copies (30-36) (Supplementary Table S12.1) than the highly expanded repertoires of fruit flies and zebrafish[32]. The acellular carbohydrate-rich laminated layer (LL), coating the outside of *Echinococcus* metacestodes, is a unique genus-specific trait and one of the few morphological traits that differs between the very closely related

species *E. granulosus* and *E. multilocularis*. We identified corresponding species differences in an *Echinococcus*-specific apomucin family (Supplementary Figure 12.1), an important building block of the LL[33]. One particular copy is highly differentiated between the two species (dN/dS ratio > 1) and is the fifth most highly expressed in the metacestode stage of *E. multilocularis* (Supplementary Table S5.7). Similarly diverged are galactosyltransferases that probably decorate the apomucins with galactose residues, the predominant sugar of LL glycans[33] (Supplementary Information S8). Another ~20% of the genes are exclusive to tapeworms and may reveal genes responsible for immuno-modulatory activities within the host. Amongst them we identified many highly expressed antigen families including Antigen B, the GPI-anchored protein GP50[34], and the vaccine target EG95[35] (Supplementary Table S12.4).

One of the most striking gene family expansions in the tapeworm genomes is the heat shock protein (HSP) family. Phylogenetic analysis revealed independent and parallel expansions in both the HSP110 and the cytosolic HSP70 (cHSP70) clades (Figure 4). Several examples of expansions exist at various clades of HSP70 in other systems, including HSP110 expansions to cope with temperature or proteotoxic stress in oysters or cancer cells, respectively[36,37]. *Echinococcus* and *T. solium* have the greatest expansions in the cHSP70 clade with 22 – 32 full copies in each species clustered amongst themselves, compared to 6 in fruit fly and 2 in humans (Figure 4). This expanded clade lack classical cHSP70 features (a conserved EEVD motif for substrate binding and a GGMP repeat unit), and while the canonical cHSP70s are constitutively expressed in different life cycle stages, the non-canonical proteins show almost no expression, suggesting a putative contingency role where individual copies of the expanded family are only highly expressed under certain conditions (Supplementary Figure 12.2). At least 40% of *E. multilocularis* HSP-like genes are found within the sub-telomeric regions of chromosomes, including the extreme case of chromosome 8 where eight copies (including pseudogenes) are located in the sub-telomere (Supplementary Table S12.2). No other genes are over-represented in these regions. Although HSP70 proteins have been found in excretory/secretory products of tapeworms[38] it remains to be determined whether the non-canonical HSPs have a host-interacting role or whether telomere proximity is important for their function or expression.

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[39] that, due to considerable side effects, are administered at parasititstatic rather than parasiticidal concentrations[40]. 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[41]. We identified ~250-300 new protein kinases (Supplementary Table S13.1) covering most major classes (Supplementary Information S13). We also identified 151 proteases and 63 peptidase-like proteins in *E. multilocularis*, a repertoire of similar diversity to *S. mansoni*, but with strongly reduced copy numbers when compared with other animals (Supplementary Table S13.9). Many successful anthelminthic drugs target one of several different forms of neural communication[41], so we mapped the signalling pathways of the serotonin and acetylcholine neurotransmitters, predicted conserved and novel neuropeptides (Supplementary Table S13.6), and classified more than 60 putative GPCRs (Supplementary Table S13.2) and 31 ligand-gated ion channels (Supplementary Table S13.4). A voltage-gated calcium channel subunit[42] - the proposed target of Praziquantel - is not expressed in cysts and thus provides a putative explanation for the drug's low efficacy.

By searching databases for potential attributes for target selection including compounds associated with protein targets and expression in the clinically relevant metacestode life-stage, we assigned weights to rank the entire proteomes (Supplementary Table S13.10). We identified 1,082 *E. multilocularis* proteins with potential druggability. Of these, 150 – 200 with the highest scores have available chemical leads (known drug or approved compounds).

High on the list are acetylcholinesterases, which are inhibited by the anti-malarial Mefloquine that also reduces egg production in *S. mansoni*[43]. 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 casein 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 somewhat similar; both show uncontrolled proliferation, invasion and metastasis, and are difficult to kill without causing damage to the surrounding tissue. Metacestodes could thus be vulnerable to similar strategies as cancer; suppression of mitosis, induction of apoptosis and prevention of DNA replication. In fact, the anthelminthic medicines niclosamide, mebendazole and albendazole have already been shown to inhibit cancer growth[44]

Conclusion

Tapeworms are among the first known parasites of humans, recorded by Hippocrates and Aristotle ~300 B.C.[45], 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 made possible by recent advances in axenic and cell culturing techniques[39,46,47]. Flatworms display an unusually high degree of developmental plasticity. In this study, the high level of sequence completion allowed 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 performed on a combination of platforms. RNA sequencing was performed with Illumina RNA-Seq protocols (*E. multilocularis*, *E. granulosus*, *H. microstoma*) or capillary sequencing of full-length cDNA libraries (*T. solium*). The complete genome annotation is available at www.genedb.org. The tapeworm genome projects were registered under the INSDC project IDs; PRJEB122 (*E. multilocularis*), PRJEB121 (*E. granulosus*), PRJEB124 (*H. microstoma*), PRJNA16816 (*T. solium*, Mexico). Sequence data for *T. solium* isolate (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).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Isheng J. Tsai[#],
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Magdalena Zarowiecki[#],

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Nancy Holroyd[#],
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Alejandro Garciarrubio[#],
Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México

Alejandro Sánchez-Flores,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México

Karen L. Brooks,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Alan Tracey,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Raúl J. Bobes,
Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

Gladis Fragoso,
Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

Edda Sciutto,
Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

Martin Aslett,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Helen Beasley,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Hayley M. Bennett,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Xuepeng Cai,
State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Federico Camicia,
Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Richard Clark,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Marcela Cucher,
Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Nishadi De Silva,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Tim A Day,
Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA

Peter Deplazes,
Institute of Parasitology, Vetsuisse Faculty, University of Zürich, Winterthurerstrasse 266a, CH-8057 Zürich, Switzerland

Karel Estrada,
Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México

Cecilia Fernández,
Cátedra de Inmunologá, Facultad de Quámica, Universidad de la República. Av. Alfredo Navarro 3051, piso 2, Montevideo, CP 11600, Uruguay

Peter W. H. Holland,
Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Junling Hou,
State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Songnian Hu,
Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, PR China

Thomas Huckvale,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Stacy S. Hung,
Department of Biochemistry & Molecular and Medical Genetics, University of Toronto, Program in Molecular Structure & Function, The Hospital for Sick Children, Toronto, Ontario, Canada

Laura Kamenetzky,
Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Europe PMC Funders Author Manuscripts

Jacqueline A. Keane,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Ferenc Kiss,
University of Würzburg, Institute of Hygiene and Microbiology, Würzburg, Germany

Uriel Koziol,
University of Würzburg, Institute of Hygiene and Microbiology, Würzburg, Germany

Olivia Lambert,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Kan Liu,
Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, PR China

Xuenong Luo,
State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Yingfeng Luo,
Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, PR China

Natalia Macchiaroli,
Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Sarah Nichol,
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Jordi Paps,
Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

John Parkinson,
Department of Biochemistry & Molecular and Medical Genetics, University of Toronto, Program in Molecular Structure & Function, The Hospital for Sick Children, Toronto, Ontario, Canada

Natasha Pouchkina-Stantcheva,
Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Nick Riddiford,
Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK; Department of Zoology, School of Natural Sciences and Regenerative Medicine Institute (REMEDI), National University of Ireland Galway, University Road, Galway, Ireland

Mara Rosenzvit,

Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Gustavo Salinas,
Cátedra de Inmunologá, Facultad de Quámica, Universidad de la República. Av. Alfredo Navarro 3051, piso 2, Montevideo, CP 11600, Uruguay

James D. Wasmuth,
Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada

Mostafa Zamanian,
Institute of Parasitology, McGill University, 2111 Lakeshore Road, Ste. Anne de Bellevue, Quebec, H9X 3V9. Canada

Yadong Zheng,
State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

The Taenia solium Genome Consortium, Jianping Cai,
State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Xavier Soberón,
Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México; National Institute of Genomic Medicine, Ministry of Health, 01900, México D.F., México

Peter D. Olson,
Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Juan P. Laclette,
Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

Klaus Brehm, and
University of Würzburg, Institute of Hygiene and Microbiology, Würzburg, Germany

Matthew Berriman
Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Affiliations

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Europe PMC Funders Author Manuscripts

Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Department of Biomedical Sciences, Iowa State University, Ames, Iowa 50011, USA

Institute of Parasitology, Vetsuisse Faculty, University of Zürich, Winterthurerstrasse 266a, CH-8057 Zürich, Switzerland

Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México

Cátedra de Inmunologá, Facultad de Quámica, Universidad de la República. Av. Alfredo Navarro 3051, piso 2, Montevideo, CP 11600, Uruguay

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, PR China

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Department of Biochemistry & Molecular and Medical Genetics, University of Toronto, Program in Molecular Structure & Function, The Hospital for Sick Children, Toronto, Ontario, Canada

Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

University of Würzburg, Institute of Hygiene and Microbiology, Würzburg, Germany

University of Würzburg, Institute of Hygiene and Microbiology, Würzburg, Germany

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, PR China

State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Beijing Institute of Genomics, Chinese Academy of Sciences, No.7 Beitucheng West Road, Chaoyang District, Beijing 100029, PR China

Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Department of Biochemistry & Molecular and Medical Genetics, University of Toronto, Program in Molecular Structure & Function, The Hospital for Sick Children, Toronto, Ontario, Canada

Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK; Department of Zoology, School of Natural Sciences and Regenerative Medicine Institute (REMEDI), National University of Ireland Galway, University Road, Galway, Ireland

Instituto de Microbiología y Parasitología Médica, Universidad de Buenos Aires-Consejo Nacional de Investigaciones Científicas y Tecnológicas (IMPaM, UBA-CONICET). Facultad de Medicina, Paraguay 2155, C1121ABG, Buenos Aires, Argentina

Cátedra de Inmunologá, Facultad de Quámica, Universidad de la República. Av. Alfredo Navarro 3051, piso 2, Montevideo, CP 11600, Uruguay

Department of Ecosystem and Public Health, Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada

Institute of Parasitology, McGill University, 2111 Lakeshore Road, Ste. Anne de Bellevue, Quebec, H9X 3V9. Canada

State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

State Key Laboratory of Veterinary Etiological Biology, Key Laboratory of Veterinary Parasitology of Gansu Province, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, No 1, Xujiaping, Chengguan District, Lanzhou 730046, Gansu Province, China

Institute of Biotechnology, Universidad Nacional Autónoma de México, 04510 México D.F., México; National Institute of Genomic Medicine, Ministry of Health, 01900, México D.F., México

Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Institute of Biomedical Research, Universidad Nacional Autónoma de México, 04510 México D.F., México

University of Würzburg, Institute of Hygiene and Microbiology, Würzburg, Germany

Parasite Genomics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Acknowledgments

We thank Matthew Dunn and OpGen for the *E. multilocularis* optical map; Mike Quail, David Willey, Nathalie Smerdon and Karen Oliver for sequencing libraries at WTSI; Thomas Dan Otto for bioinformatics expertise; Robert Davies and Quan Lin for help with data release; Alex Bateman, Marco Punta, Penny Cogghill and Jaina Minstry of Pfam; Neil Rowlings from MEROPS database, Julian Gough at SUPERFAMILY, and Christophe Dessimoz for OMA; C. Seed and F. Jarero for laboratory assistance; John Overington and Bissan Al-Lazikani; Dr. Brona Brejová for predicting genes with ExonHunter. P.D.O. and N.P.S. were supported in part by a BBSRC grant (BBG0038151) to P.D.O. P.D.O., M.Z. was supported by a SynTax joint-UK Research Council grant. The European Research Council supported P.W.H.H. and J.Paps. The *E. multilocularis*, *E. granulosus* and *H. microstoma* genome projects were funded by the Wellcome Trust through their core support of the Wellcome Trust Sanger Institute (grant 098051). The *Taenia solium* Genome Project (IMPULSA 03) was supported by the Universidad Nacional Autonóma de México. The *Taenia solium* Genome Consortium thanks P. de la Torre, J. Yañez, P. Gaytán, S. Juárez and J. L. Fernández for technical support; L. Herrera-Estrella and LANGEBIO, CINVESTAV-Irapuato and Charles

B. Shoemaker (Tufts University) for sequencing support and other advice; and J. Watanabe (deceased), S. Sugano, Y. Suzuki, (Tokyo University) for the construction and sequencing of the full length cDNA library.

Footnotes

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

The tapeworm genome projects were registered under the INSDC project IDs; PRJEB122 (*E. multilocularis*), PRJEB121 (*E. granulosus*), PRJEB124 (*H. microstoma*), PRJNA16816 (*T. solium*, *Mexico*). Illumina and 454 data are released to the European Nucleotide Archive (http://www.ebi.ac.uk/ena/) under accession numbers ERP000351, ERP000452, 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 is 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 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 EL740221 to EL763490.

Ethical considerations: All experiments involving jirds were carried out in accordance with European and German regulations on the protection of animals. Ethical approval of the study was obtained from the local ethics committee of the government of Lower Franconia (621-2531.01-2/05). Experiments with dogs were conducted according to the Swiss guidelines for animal experimentation and approved by the Cantonal Veterinary Office of Zurich prior to study start. They were carried out with facility-born animals at the experimental units of the Vetsuisse Faculty in Zurich (permission numbers 40/2009, 03/2010). A licenced hunter hunted the fox during the regular hunting season. *Hymenolepis* parasites were reared using laboratory mice in accordance with project license PPL 70/7150, granted to PDO by the U.K. Home Office.

Competing interests

The authors declare no competing financial interests.

Supplementary information

PDF files

1. Supplementary Information **(764k)**

This file contains Supplementary Text 1-13

2. Supplementary Figures **(9.4M)**

File contains Supplementary Figures S3.1-3.3, S7.1, S8.1-8.5, S9.1-9.5, S10.1, S11.1-11.2, S12.1-12.2, S13.1-S13.5

Excel files

1. Supplementary Tables **(6.3M)**

Europe PMC Funders Author Manuscripts

This file contains Supplementary Tables S1.1-1.2, S2.1, S3.1-3.2, S4.1, S5.1-5.9, S6.1-6.5, S7.1-7.3, S8.1-8.7, S9.1-9.4, S10.1-10.2, S11.1, S12.1-12.4 and S13.1-13.9

2. Supplementary Tables **(4.5M)**

This file contains Supplementary Table S13.10.

References

1. Garcia HH, Moro PL, Schantz PM. Zoonotic helminth infections of humans: echinococcosis, cysticercosis and fascioliasis. Current opinion in infectious diseases. 2007; 20:489–494. [PubMed: 17762782]
2. W.H.O.. Neglected Tropical Diseases. 2012. http://www.who.int/neglected_diseases/diseases/en/ http://www.who.int/neglected_diseases/diseases/en/
3. Eckert J, Deplazes P. Biological, epidemiological, and clinical aspects of echinococcosis, a zoonosis of increasing concern. Clinical microbiology reviews. 2004; 17:107–135. [PubMed: 14726458]
4. Brunetti E, Kern P, Vuitton DA. Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. Acta tropica. 2010; 114:1–16. [PubMed: 19931502]
5. Budke CM, White AC Jr. Garcia HH. Zoonotic larval cestode infections: neglected, neglected tropical diseases? PLoS Negl Trop Dis. 2009; 3:e319. [PubMed: 19238190]
6. Torgerson PR, Macpherson CN. The socioeconomic burden of parasitic zoonoses: global trends. Veterinary parasitology. 2011; 182:79–95. [PubMed: 21862222]
7. Burton, J.; Bogitsh, CEC.; Oeltmann, Thomas N. Human parasitology. Fourth edn. Academic Press; 2012.
8. Adams MD, et al. The genome sequence of *Drosophila melanogaster*. Science. 2000; 287:2185–2195. [PubMed: 10731132]
9. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science. 1998; 282:2012–2018. [PubMed: 9851916]
10. Protasio AV, et al. A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*. PLoS Negl Trop Dis. 2012; 6:e1455. [PubMed: 22253936]
11. Spakulova M, Orosova M, Mackiewicz JS. Cytogenetics and chromosomes of tapeworms (Platyhelminthes, Cestoda). Advances in parasitology. 2011; 74:177–230. [PubMed: 21295678]
12. DeMarco R, et al. Protein variation in blood-dwelling schistosome worms generated by differential splicing of micro-exon gene transcripts. Genome research. 2010; 20:1112–1121. [PubMed: 20606017]
13. Qian W, Zhang J. Evolutionary dynamics of nematode operons: easy come, slow go. Genome research. 2008; 18:412–421. [PubMed: 18218978]
14. Allen MA, Hillier LW, Waterston RH, Blumenthal T. A global analysis of *C. elegans* trans-splicing. Genome research. 2011; 21:255–264. [PubMed: 21177958]
15. Matsumoto J, et al. High-throughput sequence analysis of *Ciona intestinalis* SL trans-spliced mRNAs: alternative expression modes and gene function correlates. Genome research. 2010; 20:636–645. [PubMed: 20212022]
16. 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. 2008; 52:164–170. [PubMed: 17954696]
17. Berriman M, et al. The genome of the blood fluke *Schistosoma mansoni*. Nature. 2009; 460:352–U365. [PubMed: 19606141]
18. Frayha GJ. Comparative metabolism of acetate in the taeniid tapeworms Echinococcus granulosus, E. multilocularis and Taenia hydatigena. Comp Biochem Physiol B. 1971; 39:167–170. [PubMed: 5570024]
19. Obal G, et al. Characterisation of the native lipid moiety of *Echinococcus granulosus* antigen B. PLoS Negl Trop Dis. 2012; 6:e1642. [PubMed: 22616019]

Europe PMC Funders Author Manuscripts

20. Kaasch AJ, Joiner KA. Targeting and subcellular localization of *Toxoplasma gondii* catalase. Identification of peroxisomes in an apicomplexan parasite. J Biol Chem. 2000; 275:1112–1118. [PubMed: 10625653]
21. Schwarz G, Mendel RR. Molybdenum cofactor biosynthesis and molybdenum enzymes. Annu Rev Plant Biol. 2006; 57:623–647. [PubMed: 16669776]
22. Zhang Y, Rump S, Gladyshev VN. Comparative Genomics and Evolution of Molybdenum Utilization. Coord Chem Rev. 2011; 255:1206–1217. [PubMed: 22451726]
23. Pakharukova MY, et al. Cytochrome P450 in fluke *Opisthorchis felineus*: identification and characterization. Mol Biochem Parasitol. 2012; 181:190–194. [PubMed: 22115821]
24. Kuntz AN, et al. Thioredoxin glutathione reductase from Schistosoma mansoni: an essential parasite enzyme and a key drug target. PLoS Med. 2007; 4:e206. [PubMed: 17579510]
25. Zhong YF, Holland PW. HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology. Evol Dev. 2011; 13:567–568. [PubMed: 23016940]
26. Holland PW. Beyond the Hox: how widespread is homeobox gene clustering? J Anat. 2001; 199:13–23. [PubMed: 11523814]
27. Brooke NM, Garcia-Fernandez J, Holland PW. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature. 1998; 392:920–922. [PubMed: 9582071]
28. Riddiford N, Olson PD. Wnt gene loss in flatworms. Development genes and evolution. 2011; 221:187–197. [PubMed: 21892738]
29. Brehm K. *Echinococcus multilocularis* as an experimental model in stem cell research and molecular host-parasite interaction. Parasitology. 2010; 137:537–555. [PubMed: 19961652]
30. Parkinson J,WJ, Salinas G, Bizarro CV, Sanford C, Berriman M, Ferreira HB, Zaha A, Blaxter ML, Maizels RM, Fernández C. A transcriptomic analysis of Echinococcus granulosus larval stages: implications for parasite biology and host adaptation. PLoS Negl Trop Dis. 2012
31. Raz E. The function and regulation of vasa-like genes in germ-cell development. Genome Biology. 2000; 1 REVIEWS1017.
32. Garcia-Espana A, et al. Appearance of new tetraspanin genes during vertebrate evolution. Genomics. 2008; 91:326–334. [PubMed: 18291621]
33. Diaz A, et al. Understanding the laminated layer of larval Echinococcus I: structure. Trends Parasitol. 2011; 27:204–213. [PubMed: 21257348]
34. Hancock K, et al. Characterization and cloning of GP50, a *Taenia solium* antigen diagnostic for cysticercosis. Mol Biochem Parasitol. 2004; 133:115–124. [PubMed: 14668018]
35. Heath DD, Jensen O, Lightowlers MW. Progress in control of hydatidosis using vaccination - a review of formulation and delivery of the vaccine and recommendations for practical use in control programmes. Acta tropica. 2003; 85:133–143. [PubMed: 12606090]
36. Zhang G, et al. The oyster genome reveals stress adaptation and complexity of shell formation. Nature. 2012; 490:49–54. [PubMed: 22992520]
37. Subjeck JR, Repasky EA. Heat shock proteins and cancer therapy: the trail grows hotter! Oncotarget. 2011; 2:433–434. [PubMed: 21677361]
38. Vargas-Parada L, Solis CF, Laclette JP. Heat shock and stress response of *Taenia solium* and *T. crassiceps* (Cestoda). Parasitology. 2001; 122:583–588. [PubMed: 11393832]
39. Hemphill A, et al. *Echinococcus* metacestodes as laboratory models for the screening of drugs against cestodes and trematodes. Parasitology. 2010; 137:569–587. [PubMed: 19765346]
40. Brunetti E, White AC Jr. Cestode infestations: hydatid disease and cysticercosis. Infect Dis Clin North Am. 2012; 26:421–435. [PubMed: 22632647]
41. McVeigh P, et al. Parasite neuropeptide biology: Seeding rational drug target selection? International Journal for Parasitology: Drugs and Drug Resistance. 2012; 2:76–91. [PubMed: 24533265]
42. Marks NJ, Maule AG. Neuropeptides in helminths: occurrence and distribution. Adv Exp Med Biol. 2010; 692:49–77. [PubMed: 21189674]

43. Van Nassauw L, Toovey S, Van Op den Bosch J, Timmermans JP, Vercruysse J. Schistosomicidal activity of the antimalarial drug, mefloquine, in Schistosoma mansoni-infected mice. Travel medicine and infectious disease. 2008; 6:253–258. [PubMed: 18760248]

44. Doudican N, Rodriguez A, Osman I, Orlow SJ. Mebendazole induces apoptosis via Bcl-2 inactivation in chemoresistant melanoma cells. Molecular cancer research: MCR. 2008; 6:1308–1315. [PubMed: 18667591]

45. Grove, DI. A history of human helminthology. CAB International; 1990. p. 848

46. Spiliotis M, Brehm K. Axenic in vitro cultivation of *Echinococcus multilocularis* metacestode vesicles and the generation of primary cell cultures. Methods Mol Biol. 2009; 470:245–262. [PubMed: 19089387]

47. Spiliotis M, et al. *Echinococcus multilocularis* primary cells: improved isolation, small-scale cultivation and RNA interference. Mol Biochem Parasitol. 2010; 174:83–87. [PubMed: 20637246]

Europe PMC Funders Author Manuscripts

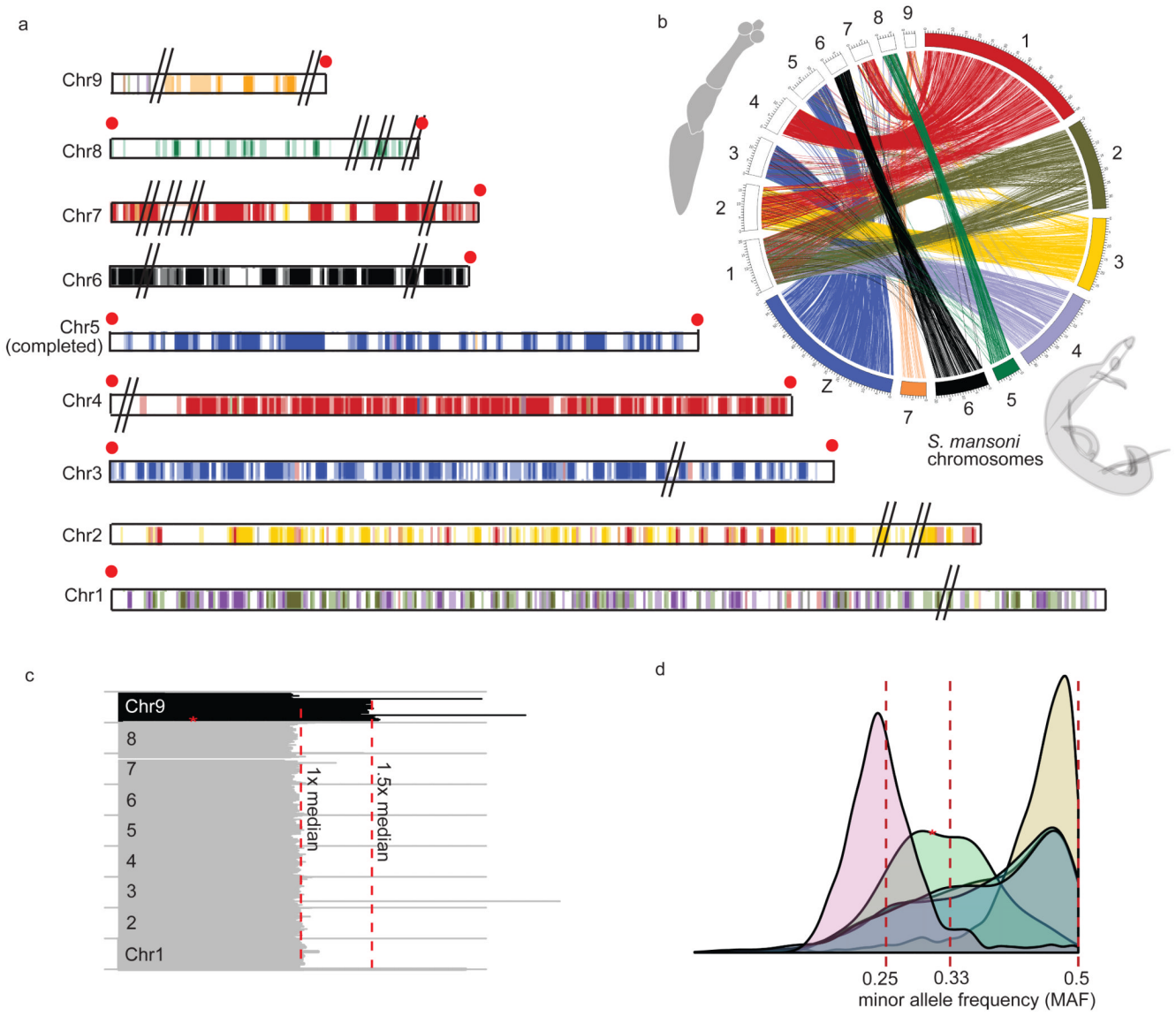

Figure 1. Genome of *E. multilocularis*

A) The nine assembled chromosomes of *E. multilocularis*. Telomeres (red circle) and physical gaps in the sequence assembly (dashed lines), but which an optical map covers, are shown. (B) One-to-one orthologues connecting *E. multilocularis* and *S. mansoni* chromosomes. (C) Distribution of normalised genome coverage on strain GT10/2. Each horizontal line depicts median coverage of 100 kb windows normalised against the mean coverage for the genome (130×). Even coverage was observed across the first eight chromosomes in *E. multilocularis* but 1.5× coverage of chromosome 9 indicates trisomy. Similar plots for other isolates are shown in Supplementary Figure S3.1. D) Distribution of minor allele frequency (MAF) of heterozygous sites in five isolates of *E. multilocularis* (plot for individual isolates in Supplementary Figure S3.1), identified by mapping sequencing reads against the assembled chromosome consensus sequences. At each site, the proportion of bases that disagree with the reference is counted. For four isolates, the MAF peaks at around 0.5, indicative of diploidy, whereas JAVA05/1 peaks at 0.25 suggesting tetraploidy.

*Chr 9 of GT10/2 is plotted separately from Chr1-8 and the MAF display a clear departure of 0.5 and peaks around 0.33, consistent with a trisomy.

Europe PMC Funders Author Manuscripts

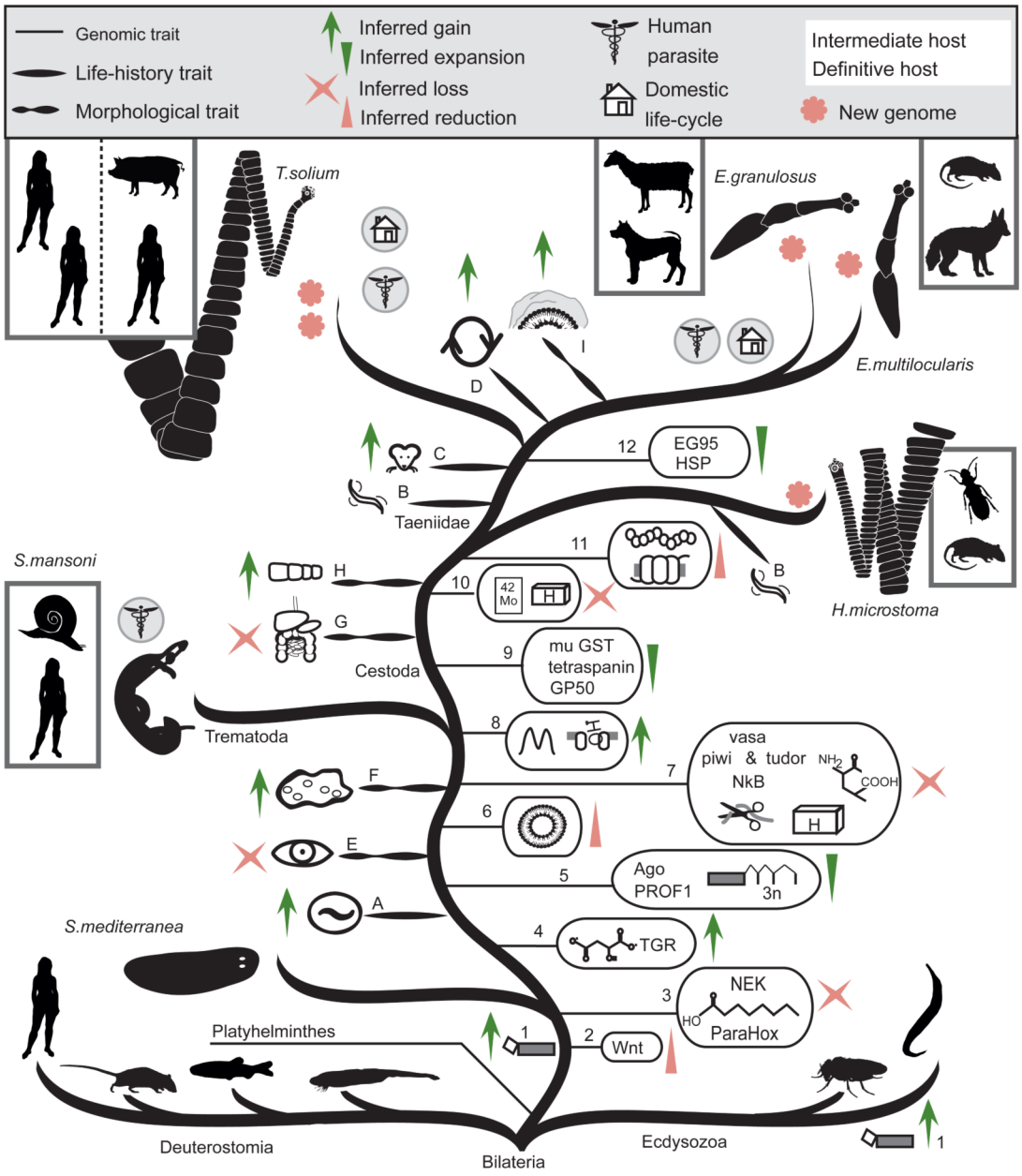

Figure 2. Road to parasitism

Phylogeny of the main branches of Bilateria; Ecdysozoa - including fruit flies and nematodes, Deuterostomia - including lancelet, zebrafish, mice and humans, and Lophotrochozoans, including Platyhelminthes (flatworms), based on phylogeny in Supplementary Figure S7.1. Gains and losses of life cycle traits; A. endoparasitism evolves, B. passively transmitted between hosts, C. acquires vertebrate intermediate host, D. ability for asexual proliferation in intermediate host. Morphological traits that have evolved include E. cup-eyes were lost, F neodermatan syncytial epithelia gained, G. gut was lost, H. segmentation of body plan, I. laminated layer evolved, containing specialised apomucins. Gains and losses of genomic traits: 1. SL-trans-splicing, 2. loss of Wnt genes, 3. loss of NEK kinases, fatty acid biosynthesis and ParaHox genes, 4. anaerobic metabolic ability through the malate dismutation/rodhoquinone pathway, merger of Glutaredoxin (Grx) and

Europe PMC Funders Author Manuscripts

thioredoxin reductase (TR) to thioredoxin glutathione reductase (TGR) 5. evolution of tapeworm and fluke specific Argonaute family, micro exon genes (MEGs) and PROF1 GPCRs, 6. loss of peroxisomal genes 7. complete loss of vasa, tudor and piwi genes, NkB pathway, loss of 24 homeobox gene families, metabolic proteases and amino acid biosynthesis, 8. in tapeworms: innovation of bimodal intron distribution and novel fatty acid transporters 9. expansion of mu glutathione-S-transferases, GP50 antigens and tetraspanins, 10. loss of molybdopterin biosynthesis pathway, loss of 10 homeobox gene families 11. fewer GPCRs and fewer neuropeptides encoded by each protopeptide, 12. expansion of heat shock proteins and species-specific antigens.

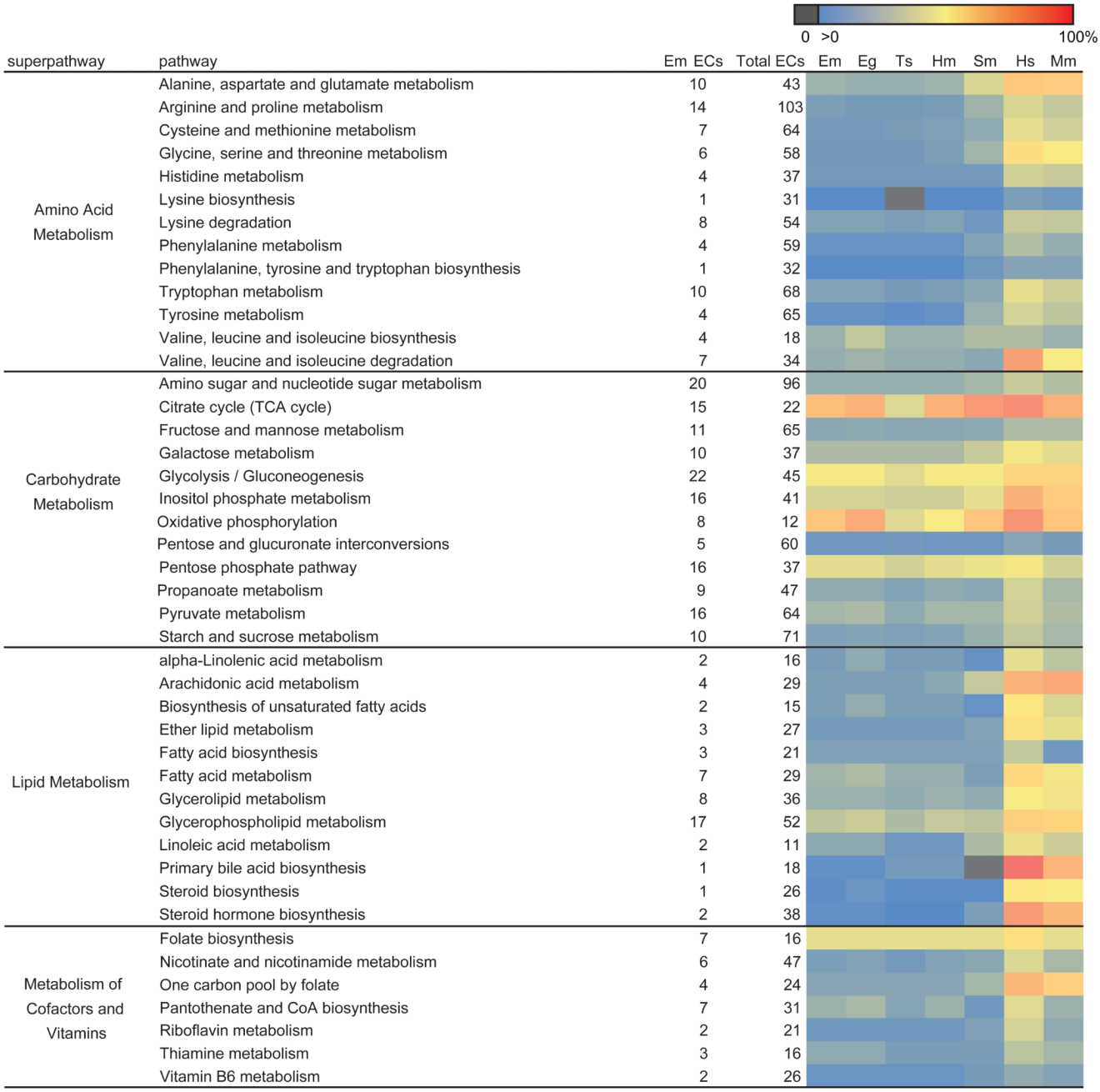

Figure 3. Conservation of individual metabolic pathways

Heatmap showing the conservation of individual metabolic pathways for *E. multilocularis* (Em), *E. granulosus* (Eg), *T. solium* (Ts), *H. microstoma* (Hm) and *S. mansoni* (Sm) compared to those of humans (Hs) and mice (Mm). Each row indicates an individual metabolic pathway grouped by their superclass membership (defined by KEGG). Coloured tiles indicate the level of conservation (percentage of enzymes detected) of each pathway within each species. KEGG pathways with insufficient evidence (i.e. containing only one enzyme) in *E. multilocularis* have been removed.

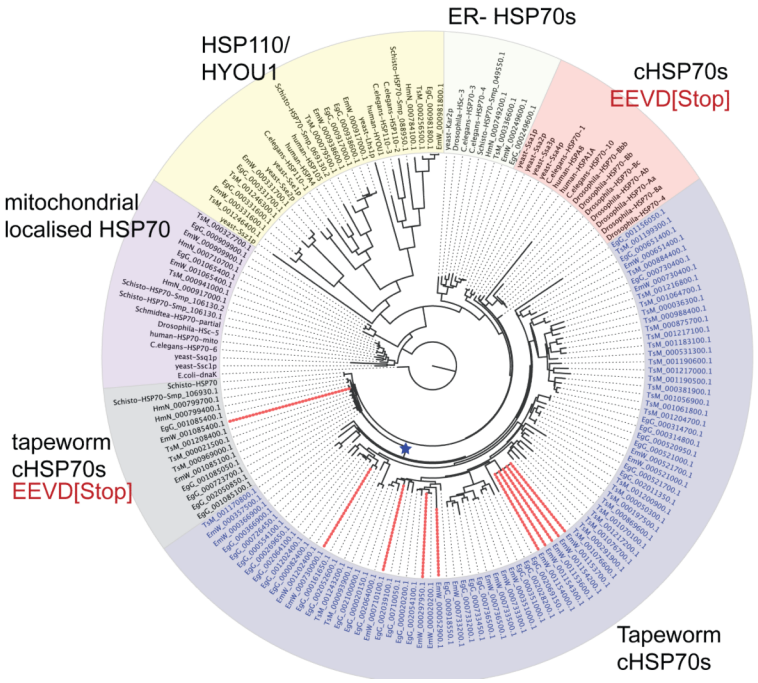

Figure 4. Heat shock protein 70 expansions in the tapeworms

Rooted tree of HSP70 sequences from the eight comparative species used in this study and tapeworms with additional sequences from baker's yeast *Saccharomyces cerevisiae*, and the Pacific oyster *Crassostrea gigas* (a non-flatworm example of a lophotrochozoan) with a recently reported HSP70 expansion. Colour highlights different HSP70 subfamilies. Red stars indicates the *E. multilocularis* cytosolic HSP70 that are located in the sub-telomeres. EEVD denotes the conserved C-terminal residues of a canonical cytosolic HSP70.

Europe PMC Funders Author Manuscripts

Table 1

Top 20 promising targets in *E. multilocularis*

Target	Action	Expression	Drug	Rank
Current targets				
Tubulin beta chain	Cytoskeleton	M,A	Albendazole	406
Voltage dependent calcium channel	Ion transport	-	Praziquantel	277
Potential target				
Thioredoxin glutathione reductase (TGR)	Detoxification	M,A	Experimental compounds	277
Top Predicted targets				
Fatty acid amide hydrolase	Bioactive lipid catabolism	M	Thiopental, Propofol	1
Adenine nucleotide translocator	Mitochondrial ATP export	M	Clodronate	2
Inosine 5′ monophosphate dehydrogenase	Purine biosynthesis	M	Mycophenolic acid, Ribavirin	3
Succinate semialdehyde dehydrogenase	GABA catabolism	M	Chlormerodrin	3
Ribonucleoside diphosphate reductase	Purine biosynthesis	M,A	Motexafin gadolinium	5
Casein kinase II	Cell cycle regulating kinase	M,A	Experimental compounds	6
Hypoxanthine guanine phosphoribosyltransferase	Purine biosynthesis	M,A	Azathioprine	8
Glycogen synthase kinase 3	Multiple signaling pathways	M,A	Lithium	8
Proteasome subunit	Protein degradation	M,A	Bortezomib	16
CalModulin	Transduces calcium signals	M,A	Trifluoperazine	19
FK506 binding protein	Protein folding	M,A	Pimecrolimus	19
UMP:CMP kinase	Phosphorylases ribonucleotides	M	Gemcitabine	39
Na^+/K^+-ATPase	Ion transport	M	Artemether	42
Carbonic anhydrase II	Acidity control	M	Multiple, e.g. Methazolamide	42
NADH dehydrogenase subunit 1	Energy metabolism	M	Multiple, e.g. Methoxyflurane	42
Translocator protein	Multiple functions	M,A	Multiple, e.g. Lorazepam	42
Elongation factor 2	Translation	M,A	Experimental compounds	54
Cathepsin B	Protease	M	Experimental compounds	55
Dual specificity mitogen activated protein	Signaling, activation of p38	M	Experimental compounds	56
Purine nucleoside phosphorylase	Purine metabolism	M,A	Didanosine	63

Expression: M- metacestode, A – adult. Rank is sorted starting from the highest overall score; proteins with tied scores have the same rank. For current targets, the rank is only reported from the highest scoring protein family member. For full scores and information please see Supplementary Table S13.10