

RESEARCH ARTICLE

Open Access

Cestode strobilation: prediction of developmental genes and pathways



Gabriela Prado Paludo^{1,2}, Claudia Elizabeth Thompson^{2,3}, Kendi Nishino Miyamoto², Rafael Lucas Muniz Guedes^{4,5}, Arnaldo Zaha², Ana Tereza Ribeiro de Vasconcelos⁴, Martin Cancela^{1,2} and Henrique Bunselmeyer Ferreira^{1,2*} 

Abstract

Background: Cestoda is a class of endoparasitic worms in the flatworm phylum (Platyhelminthes). During the course of their evolution cestodes have evolved some interesting aspects, such as their increased reproductive capacity. In this sense, they have serial repetition of their reproductive organs in the adult stage, which is often associated with external segmentation in a developmental process called strobilation. However, the molecular basis of strobilation is poorly understood. To assess this issue, an evolutionary comparative study among strobilated and non-strobilated flatworm species was conducted to identify genes and proteins related to the strobilation process.

Results: We compared the genomic content of 10 parasitic platyhelminth species; five from cestode species, representing strobilated parasitic platyhelminths, and five from trematode species, representing non-strobilated parasitic platyhelminths. This dataset was used to identify 1813 genes with orthologues that are present in all cestode (strobilated) species, but absent from at least one trematode (non-strobilated) species. Development-related genes, along with genes of unknown function (UF), were then selected based on their transcriptional profiles, resulting in a total of 34 genes that were differentially expressed between the larval (pre-strobilation) and adult (strobilated) stages in at least one cestode species. These 34 genes were then assumed to be strobilation related; they included 12 encoding proteins of known function, with 6 related to the Wnt, TGF- β /BMP, or G-protein coupled receptor signaling pathways; and 22 encoding UF proteins. In order to assign function to at least some of the UF genes/proteins, a global gene co-expression analysis was performed for the cestode species *Echinococcus multilocularis*. This resulted in eight UF genes/proteins being predicted as related to developmental, reproductive, vesicle transport, or signaling processes.

Conclusions: Overall, the described in silico data provided evidence of the involvement of 34 genes/proteins and at least 3 developmental pathways in the cestode strobilation process. These results highlight on the molecular mechanisms and evolution of the cestode strobilation process, and point to several interesting proteins as potential developmental markers and/or targets for the development of novel antihelminthic drugs.

Keywords: Platyhelminthes, Segmentation, Development, Comparative genomics, Co-expression network

* Correspondence: henrique@cbiot.ufrgs.br

¹Laboratório de Genômica Estrutural e Funcional, Centro de Biotecnologia (CBiot), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil

²Programa de Pós-Graduação em Biologia Celular e Molecular, CBiot, UFRGS, Porto Alegre, RS, Brazil

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

The phylum Platyhelminthes comprise a diverse array of species that occur in all continental masses, seas, rivers, and lakes. Platyhelminths (flatworms) are dorsoventrally flattened bilaterian acoelomates that lack an anus, possess a low level of cephalization, and are usually hermaphroditic [1]. They are divided into four main classes: Turbellaria (free-living planarians), Monogenea (mostly aquatic ectoparasites), Trematoda (flukes), and Cestoda (tapeworms) [2].

Parasitic flatworms (monogeneans, flukes, and tapeworms) form a monophyletic group known as Neodermata [3], which constitutes one of the three largest groups of metazoan parasites that infect vertebrates (the others being nematodes and arthropods). Flukes and tapeworms form a derivate monophyletic group of endoparasitic species, some of which are of great medical and veterinary importance. For instance, the World Health Organization's list of neglected tropical diseases (http://www.who.int/neglected_diseases/diseases/en/) includes several caused by endoparasitic flatworms, namely food-borne trematodiasis (caused by *Clonorchis* spp. and *Fasciola* spp., among other flukes), schistosomiasis (caused by *Schistosoma* spp.), echinococcosis (caused by *Echinococcus* spp.), and taeniasis/cysticercosis (caused by *Taenia* spp.).

Tapeworms are obligate endoparasites of vertebrates that display complex life cycles in which morphologically and physiologically distinct forms alternate, adapted to the survival and development of different intermediate host species [4, 5]. The Cestoda class is divided into two subclasses: Eucestoda and Cestodaria. The Eucestoda subclass ('true' tapeworms) includes *Echinococcus* spp. and *Taenia* spp., which are the main species of medical and veterinary interest. Eucestodes have increased fertility due to metamerism (the serial repetition of body structures, or metameres) [6]. In eucestodes, metamerism is represented by the internal serial repetition of their reproductive organs, called proglottization [4, 6]. In the majority of eucestode evolutionary lineages, proglottization is associated with the external delimitation (segmentation) of proglottids, in a developmental process called strobilation [4, 5]. Eucestode strobilation occurs during the transition from the larval to adult stage and involves the repetitive generation of new proglottids in the base of the head (scolex), in the so-called neck region. Strobilation persists during the whole life of the adult worm, with each proglottid moving toward the posterior end as a new one is generated in the neck region. Strobilation allows adult cestodes to have larger numbers of hermaphroditic sexual organ sets, promoting both cross and self-fertilization and the frequency of egg-release by the progressive excision of gravid proglottids [4, 6].

In cestode evolution, the eucestodes constitute the most recently evolved subclass, and includes orders with different degrees of proglottization/strobilation [4]. The most ancestral eucestodes (those from the order Caryophyllidea) do not undergo proglottization and are non-segmented, similar to the other ancestral cestodes (i.e. those from the Cestodaria subclass). On the other hand, eucestodes of the order Spathebothriidea have proglottization without body segmentation, whereas others from more recently evolved orders (e.g. members of the order Cyclophyllidea, including the most relevant cestodes from an epidemiological point of view) undergo full strobilation (i.e. proglottization along external segmentation).

Thus, the evolution of strobilated tapeworms from non-segmented ancestors could be explained by two hypotheses: the initial loss of segmentation from an ancestral segmented lophotrochozoan, followed by the re-emergence of this process in more recent eucestode lineages in the form of proglottization/strobilation; or the independent evolution of proglottization/strobilation within the eucestode lineage, which gave rise to the present proglottized or fully strobilated orders. Therefore, cestodes are interesting subjects for evolutionary developmental studies [7], especially for those aiming to elucidate the evolutionary origins of developmental novelties related to strobilation. To achieve this, it is important to comparatively analyze cestode genomes and identify candidate genes related to these developmental processes.

So far little is known about strobilation and other developmental processes in cestodes at the molecular level, despite the relevance of such knowledge for both basic evolutionary developmental studies [7] and the identification of targets for new alternative drugs against cestode parasites [8–11]. To better understand cestode molecular biology, several species have been targeted by "omics" studies, including genomic, transcriptomic, and proteomic surveys. *Echinococcus granulosus*, *Echinococcus multilocularis*, *Hymenolepis microstoma*, and *Taenia solium* were the first cestodes to have their genomes completely sequenced [12]. Later, the 50 Helminth Genomes Project provided the draft genome sequence of another 14 cestode species, along with those of many other helminths [13]. Furthermore, transcriptomic and proteomic data are available for different life-cycle stages of cestode species, such as *E. granulosus*, *E. multilocularis*, *H. microstoma*, and *Mesocestoides corti* [12, 14–18]. In some of these studies, the differential expression of transcripts and proteins between larval (non-strobilated) and adult (strobilated) forms has been assessed. For instance, tetrathyridia (larvae) and adult segmented worms of the *M. corti* species were compared with regard to their miRNA [14], mRNA [15, 16] and protein

[17] repertoires, providing some clues about the gene products differentially expressed during the transition between these stages. The initial steps of strobilation were addressed by two proteomic studies; one comparing *M. corti* bona fide tetrathyridia with tetrathyridia after 24 h of strobilation induction [18] and the other identifying proteins newly synthesized in *E. granulosus* pre-adult forms (protoscoleces) upon strobilation induction [19].

Here, a data mining approach, integrating genomic and transcriptomic data, was carried out to identify cestode developmental genes and pathways, and to demonstrate the molecular mechanisms involved in the strobilation process. Eighteen species from the Protostomia clade of Bilateria metazoans were assessed; 10 strobilated and non-strobilated species of flatworms and 8 outgroup species. Their genome sequences and transcriptional profiles were compared to identify tapeworm developmental genes associated with strobilation and, among these genes, those that had differential expression in the cestode pre-strobilated and strobilated stages. Genes associated with the strobilation process had their evolutionary histories investigated through phylogenetic and positive selection analyses. Moreover, functional enrichment provided further information on annotated

gene products, and co-expression network analyses provided further information on the products of hypothetical genes. Overall, 34 proteins associated with strobilation were identified, providing evidence of the involvement of both conserved and novel cellular pathways in cestode strobilation.

Results

Phylogenomic analyses of strobilated and non-strobilated Platyhelminthes cestode species

A phylogenomic analysis was carried out with 10 neodermatan genomes, comprising 5 tapeworm species to represent strobilated platyhelminths (*E. granulosus*, *E. multilocularis*, *H. microstoma*, *M. corti*, and *T. solium*); and by 5 fluke species to represent non-strobilated platyhelminths (*Clonorchis sinensis*, *Schistosoma haematobium*, *Schistosoma japonicum*, *Schistosoma mansoni*, and *Opisthorchis viverrini*). An outgroup set of 8 genomes was used in the analysis, comprising 6 genomes from nematodes to represent non-segmented helminths (*Caenorhabditis elegans*, *Globodera pallida*, *Haemonchus contortus*, *Onchocerca volvulus*, *Strongyloides ratti*, and *Trichuris muris*); 1 genome from an annelid to represent a segmented protostome (*Helobdella robusta*); and 1

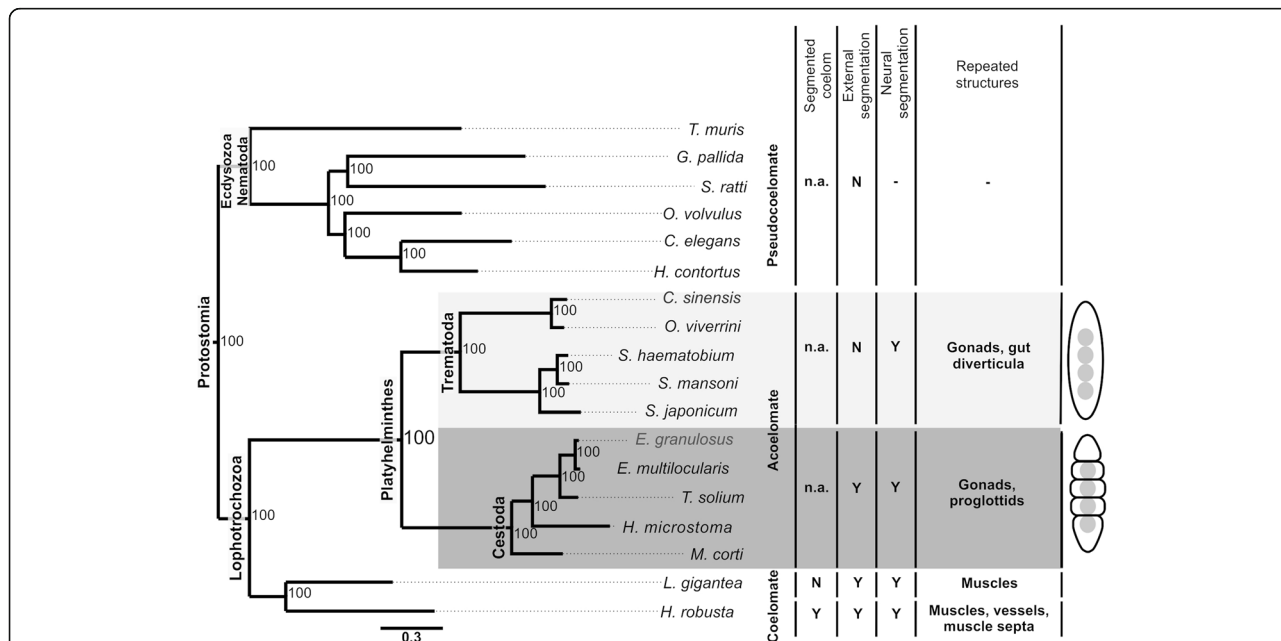


Fig. 1 Platyhelminth evolutionary relationships and segmentation features. The phylogenomic tree (left) was built by MrBayes software with the VT + I + G evolutive model, for 1,688,000 generations, and with a set of 285 orthologs shared by all species. Platyhelminth species are highlighted, with the trematodes (flukes) shaded in light gray and the cestodes (tapeworms) shaded in dark gray. The numbers at the branches are Bayesian posterior probability values. Acelomated (platyhelminths), pseudocoelomated (nematodes), and coelomated (mollusk and annelid) species and corresponding segmentation features are indicated: external segmentation refers to segmented external structures derived from the epidermis (e.g. proglottids in cestodes); neural segmentation refers to ganglia repetition along the longitudinal axis (e.g. the “ladder-like” nervous system of cestodes); segmented structures refer to repeated organs or other anatomical features derived from the mesoderm (e.g. the repeated gonads in cestodes). Cartoons (right) illustrate the metamerism in flukes and full strobilation in tapeworms. Y = yes; N = no; n.a. = not applicable

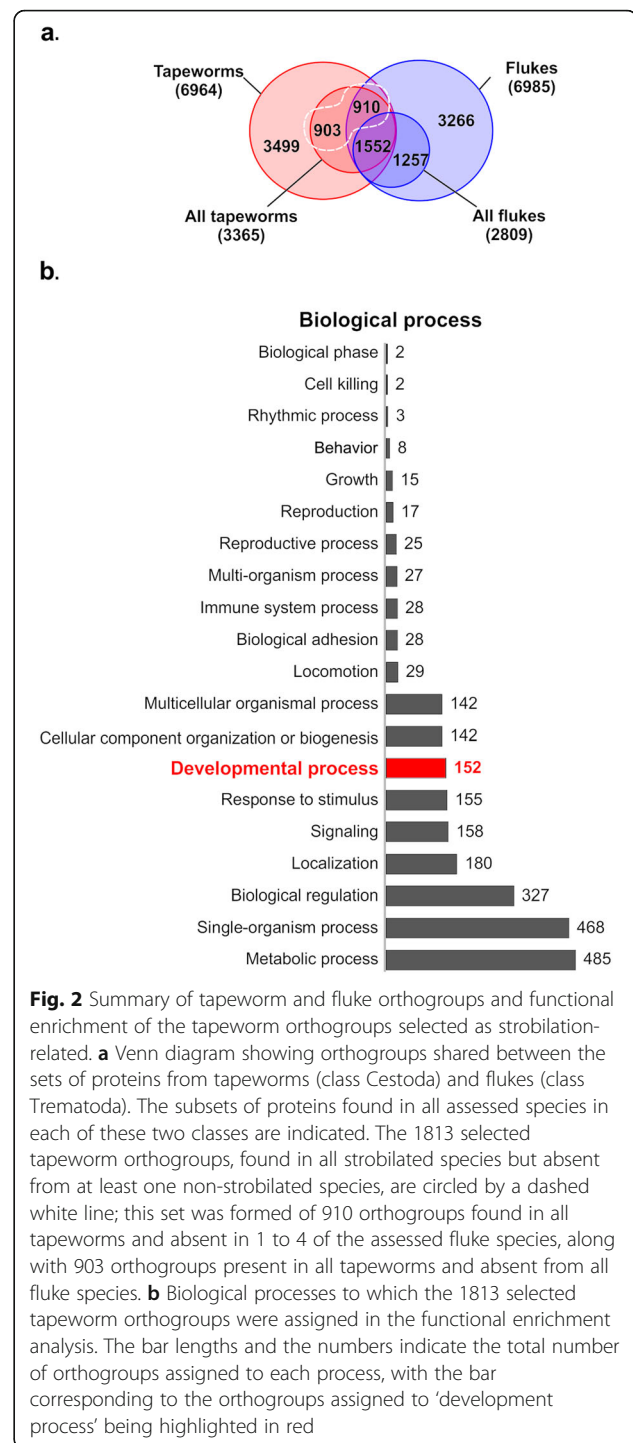
genome from a mollusk to represent a non-segmented protostome (*Lottia gigantea*).

Overall, 11,300 orthogroups of deduced protein sequences were identified, of which 285 have orthologous genes in all 18 analyzed species. These 285 protein sets were aligned, and the alignments were concatenated in a supermatrix for phylogenomic inference via Bayesian analysis. In the resulting tree (Fig. 1), two endoparasitic flatworm monophyletic groups were highly supported, with posterior probability of 100, one corresponding to flukes (trematodes) and the other to tapeworms (cestodes). The platyhelminthes were clearly divided into two clades, one with external body segmentation (with full strobilation) and the other with only internal segmentation (proglottization). Regarding protostome relationships, the tree supports the monophyly of the Ecdysozoa, Lophotrochozoa, Platyhelminthes, Cestoda, and Trematoda clades.

Identification of strobilation-related proteins

To identify proteins related to the strobilation process, orthogroups were selected based on their presence in tapeworm species or absence from fluke species (Fig. 2). From the 11,300 identified orthogroups of deduced protein sequences, 6964 (61.63%) were found in at least one tapeworm species, whereas 6985 (61.81%) were found in at least one fluke species (Fig. 2a). From this subtotal, a set of 3365 orthogroups were shared by all 5 tapeworm species, and a set of 2809 orthogroups were shared by all 5 fluke species (supplementary Figure S1). It was assumed that proteins essential for tapeworm development would be found in all strobilated species but may be absent from non-strobilated species. Based on this criterion, 1813 tapeworm strobilation-related orthogroups were initially selected (Fig. 2a; supplementary Table S1). From the 1813 selected orthogroups, 910 were found in all tapeworms that were absent from 1 to 4 fluke species, whereas 903 orthogroups were found in all tapeworms that were absent from all 5 fluke species.

As tapeworm strobilation is a developmental process, we performed a functional enrichment of the initial set of 1813 tapeworm strobilation-related orthogroups. The functional assignment of these orthogroups is shown in supplementary Table S1; the biological process assignment is summarized in Fig. 2b, and the assigned molecular functions and cellular components are summarized in supplementary Figure S2. Overall, 152 orthogroups were assigned to developmental processes (highlighted in yellow in supplementary Table S1) and then selected for further analyses. A total of 304 orthogroups of UF proteins (highlighted in blue in supplementary Table S1) were also selected, as at least some of these may also be related to developmental processes.



Furthermore, considering that the strobilation process occurs only in the adult tapeworm and not in the larval stage(s), we used available transcriptomic data of three tapeworm species (*E. multilocularis*, *H. microstoma*, and *M. corti*) to identify proteins from the selected 326 orthogroups (152 development-related orthogroups and 304 UF proteins) whose genes have differential

Table 1 Summary of the tapeworm proteins selected as being strobilation-related. The presence of orthologues in species of other taxa is indicated by an 'x'. Differential expression of the corresponding genes in larval versus adult stages is indicated by an arrow, with up (↑) and down (↓) arrows indicating up and down-regulation, respectively, in the adult (strobilated) stage; a dot (●) indicates that there is no differential expression between these stages. Abbreviations of protein names are as follows: bone morphogenetic protein 2 (BMP-2); Cyclin-G-associated kinase (GAK); Groucho protein (Groucho); Homeobox protein B4a (HoxB4a); Lim homeobox protein Ihx1 (LHX1); membrane-associated guanylate kinase protein 2 (MAGI2); serine/threonine protein kinase Mark2 (Mark2); atrial natriuretic peptide receptor 1 (NPR1); RNA binding motif single stranded interacting (RBMS); serine/threonine protein kinase (Ser:Thr kinase); mothers against decapentaplegic homolog 4-like (SMAD4); and Pangolin J (TCF/LCF). Proteins of unknown function (UF) are identified by numbers according to the corresponding orthogroups (Continued)

Protein name	Platyhelminthes			Cestoda							Annelida				Mollusca					Nematoda							
	C.	O.	S.	S.	S.	S.	E.	E.	E.	H.	H.	M.	M.	T.	H.	H.	L.	L.	C.	G.	H.	O.	S.	T.	S.		
	<i>sinensis</i>	<i>viverrini</i>	<i>haematobium</i> ^a	<i>japonicum</i>	<i>manson</i> ^b	<i>granulosus</i>	<i>multilocularis</i> ^c	<i>microstoma</i> ^d	<i>corti</i> ^e	<i>solium</i>	<i>robusta</i>	<i>gigantea</i>	<i>elegans</i>	<i>pallida</i>	<i>contortus</i>	<i>volvulus</i>	<i>ratti</i>										
UF14					x	↑	↑	↑	●	x																	
UF15	x				●	x	↑	●	●	x	x																
UF16			●			x	↓	●	●	x																	
UF17						x	↓	●	●	x																	
UF18	x			↓	●	x	●	●	↑	x																	
UF19	x					x	↑	●	●	x																	
UF20	x				●	x	↑	●	↑	x																	
UF21						x	↑	●	●	x																	
UF22						x	↓	●	●	x																	

^aS. *haematobium* expressed sequence tag libraries analysis (Young et al. 2012) [20]

^bS. *mansonii* RNA-seq analysis (Protasio et al. 2012) [21]

^cE. *multilocularis* RNA-seq analysis (Tsai et al. 2013) [12]

^dH. *microstoma* RNA-seq analysis (Tsai et al. 2013) [12]

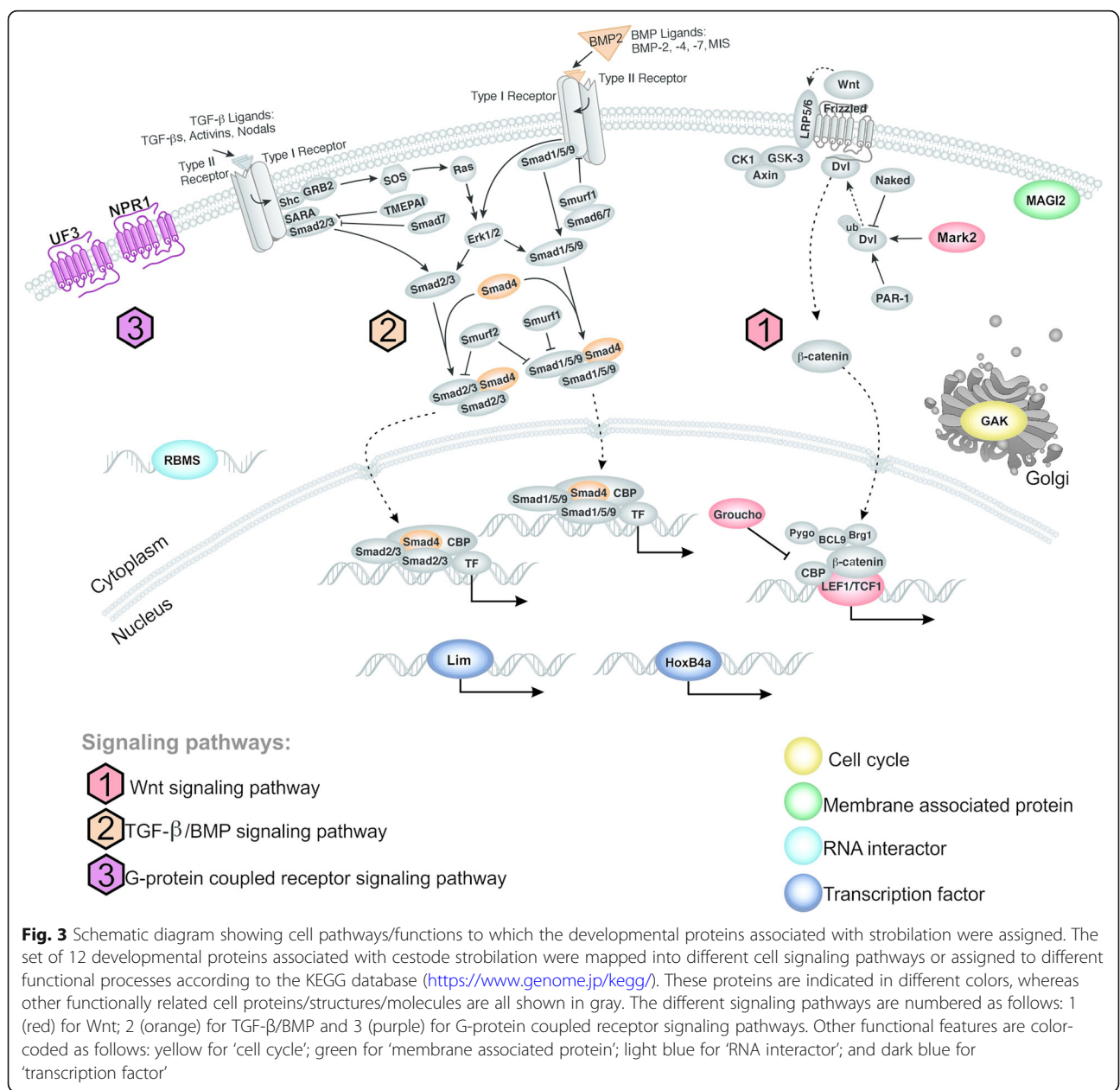
^eM. *corti* RNA-seq data analysis (Basika et al. 2019) [16]

expression in the larval (pre-strobilation) and adult (strobilated) stages. The differentially expressed orthologous genes in the larval and adult stages of the non strobilated flukes *S. haematobium* [20] and *S. mansoni* [21] were excluded to avoid genes not related to strobilation. This resulted in 12 development-related and 22 UF proteins (from now on identified as UF 1–22) being selected as strobilation-related proteins (Table 1, supplementary Table S2).

The selected set of 12 proteins previously associated with development in other organisms was mapped into cellular pathways based on KEGG data (Fig. 3). Among these proteins, Groucho, MARK2, and TFC/LCF were

mapped as components of the Wnt pathway; BMP2 and Smad4 were mapped as components of the TGF-β/BMP pathway; and NPR1 was mapped as component of the G-protein coupled receptor signaling pathway. This provided evidence of the involvement of some well-known developmental pathways in cestode strobilation.

To evaluate and confirm the homology of the proteins in the selected orthogroups, we performed functional domain predictions and comparisons (Fig. 4, supplementary Table S3). In all cases, the proteins within the orthogroup showed the same profile of predicted domains, further confirming their orthologies. In the set of 12 developmental proteins, several functional domains



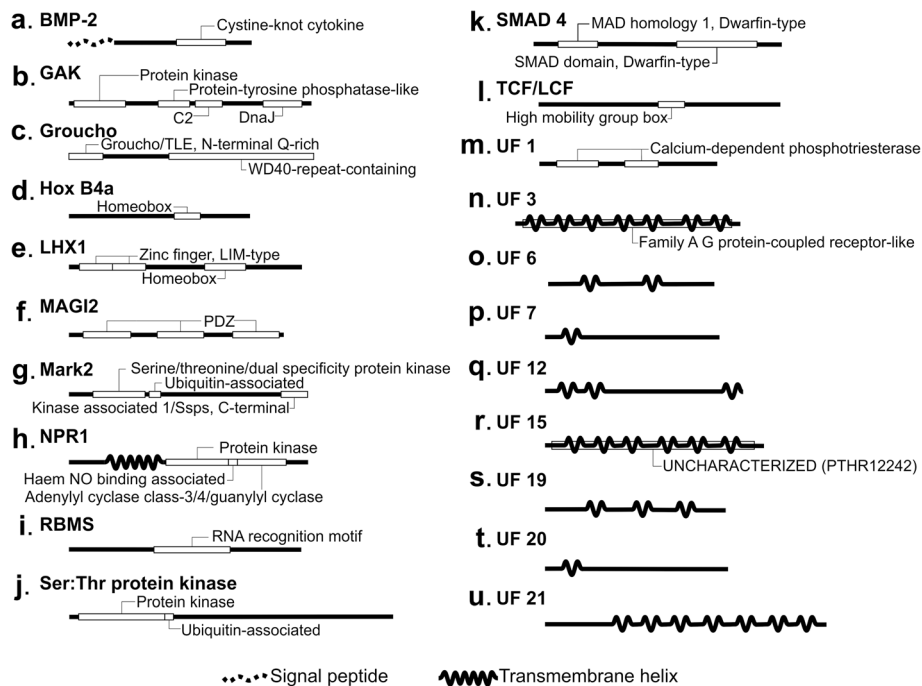


Fig. 4 Domain profiles of the putative strobilation-related proteins. Schematics showing the domains shared by all tapeworm orthologues of (a) BMP2, (b) GAK, (c) Groucho, (d) HoxB4a, (e) LHX1, (f) MAGI2, (g) Mark2, (h) NPR1, (i) RBMS, (j) Ser:Thr kinase, (k) SMAD4, (l) TCF/LCF, (m) UF1, (n) UF3, (o) UF6, (p) UF7, (q) UF12, (r) UF15, (s) UF19, (t) UF20, and (u) UF21. Transmembrane helices are shown as a continuous wavy line, and signal peptides are shown as a dotted wavy line. Other domains are shown as white boxes with their names are indicated

were identified, including a signal peptide and homeobox (Hox B4a and LHX1), RNA recognition motif, and PDZ domains. In the set of 22 UF proteins, 8 had transmembrane domains, two of them in association with another functional domain (a family A G protein-coupled receptor-like in UF 3 and an uncharacterized PTHR12242 domain in UF 15). Two calcium dependent phosphotriesterase domains were also identified in another UF protein (UF 1). No functional domains were found in the remaining 13 UF proteins.

Evolutionary analyses of strobilation-related proteins

A comprehensive search for orthologs was carried out for all the identified strobilation-related proteins, defining 34 orthogroups. The resulting data is summarized in the supplementary Figure S3. It was verified that half (17) of these proteins (MAGI2, Mark2, RBMS, Ser:Thr kinase, TCF/LCF, UF 3, UF 5–14, and UF 22) had orthologs present only among cestodes. Only eight of the strobilation-related proteins had orthologs in non-flatworm lophotrocozoans, namely BMP-2, GAK, Groucho, Hox B4a, LHX1, NPR1, SMAD4, and UF 15. The remaining nine assessed targets (UF 1–2, UF 4, and UF 16–21) showed orthologs only among flatworms. It is interesting to note that considering this last set of 17 proteins were not cestode-exclusive, homologies among cestode orthologues (47.01–81.24%) were considerably

higher than their homologies to orthologs from the other assessed taxa (15.82–48.45%) (supplementary Table S4).

Phylogenetic analyses were then performed to describe the evolutionary history of the 34 putative strobilation-related orthogroups (supplementary Table S5). The resulting trees (supplementary Figure S4–S37) agreed with the monophyly of tapeworms previously established by the phylogenomic analysis. The cyclophyllidean species from the Hymenolepididae and Mesocostoididae families (*H. microstoma* and *M. corti*, respectively) alternated as the most basal, whereas members of the Taeniidae family were more derived.

To address whether the evolution of these putative strobilation-related orthogroups could be under selective pressure, we performed searches to detect codons under positive selection. All codon sequences were aligned and the best phylogenetic tree for each orthogroup was selected (supplementary Table S6) for use in the positive selection analysis. The results suggested that the assessed orthogroups have not been under positive selection, except for a single site in the UF 16 orthogroup (supplementary Table S7).

Functional analyses of strobilation-related UF proteins

Most of the proteins selected as being strobilation-related do not have any previously described biological

function (UF 1–22). Therefore, to help determine their functional roles in tapeworm development, we performed a gene co-expression network analysis using the available transcriptomic data for the *E. multilocularis* pre-adult stage. Data from a pre-adult stage was used because of the lack of RNA-seq data from adult cestodes, in the amount and quality, required to allow the generation of a network based on gene co-expression.

Network construction was based on eight *E. multilocularis* RNA-seq samples, for which no sample outliers were identified (supplementary Figure S38). The resulting network consisted of 2957 nodes and 2,140,516 connections. Gene co-expressions were detected for 15 proteins in the set of 34 selected as strobilation-related, namely Hox B4a, LHX1, MAGI2, and NPR1 and UFs 2, 3, 5, 6, 10, 11, 13, 15, 16, 19, and 20. Only eight of them, namely UFs 2, 6, 10, 11, 15, 16, 19, and 20, were considered hubs (supplementary Table S8) and selected for further analyses. In Fig. 5, the modules to which the corresponding genes of these hub proteins belong to are indicated. With the aim of enhancing the information robustness of these selected network modules, we used the hub genes from each selected module as input for protein-protein interaction (PPI) analyses. Assuming that interacting proteins of a given module are usually involved in the same biological functions [22], we performed a functional enrichment for each of the eight selected modules (supplementary Figure S39 and supplementary Tables S9–15) to evaluate possible functions or pathways involving the strobilation-related UF proteins (Fig. 5). Based on the functional enrichment, the UFs 2, 6, 10, 11, 15, 16, 19, and 20 hub proteins were assigned to Wnt or G protein-coupled signaling pathways, and/or to biological functions such as vesicle-mediated transport, inductive cell migration, cell adhesion, apoptotic processes, and cellular response to interleukin-1, linking these pathways and functions to tapeworm strobilation.

Discussion

Cestode developmental processes, including those involved in the transition from larvae to adult segmented worms (strobilation), are still poorly understood and a truly relevant theme in flatworm biology. Some previous proteomic and transcriptomic studies have started to unravel at least some of the molecular events underpinning cestode strobilation [12, 14–19], but we are still far from understanding all the mechanisms involved in this complex developmental process. In this study, we generated novel and complementary information regarding cestode strobilation using an integrative data mining approach that compared the genomic data of five tapeworm species and five fluke species, and compared the

transcriptomic data of different developmental stages of three tapeworms and one fluke.

The draft genomes selected for analyses were chosen because of their advanced stages of sequencing, assembly, and annotation, and allowed the correct identification of genes/proteins and a precise definition of orthogroups. With that, a dependable assessment of the presence/absence of orthologues in the genomes of closely related tapeworm and fluke species was achieved. Overall, 34 proteins (12 developmental proteins and 22 UF proteins) were selected based on the following criteria (i) presence in all fully strobilated species; (ii) absence in at least one non-strobilated species; (iii) annotation as a developmental gene/protein or with unknown function; and (iv) differential expression in strobilated developmental stages.

The use of such an approach and criteria resulted in the identification of a set of strobilation-related genes that can be regarded as complementary to other sets of genes previously associated with this developmental process based solely on their differential expression at the transcriptional level [12, 14–16] and/or at the protein level [17–19]. As each of these in silico or wet lab approaches rely on different technologies and present different advantages and limitations, they complement each other in the assessment of complex biological processes, such as those involved in development or disease [23–25]. Therefore, the ongoing and progressive assessment of cestode strobilation by different in silico and wet lab approaches is expected to generate a more comprehensive picture of the molecules, mechanisms, and developmental pathways underlying this complex biological process.

An interesting feature of the novel set of 34 proteins identified as being related to strobilation was that half (17) of them were exclusive from tapeworms. Moreover, the remaining 17 proteins, although not exclusive, were considerably more conserved among cestodes (with identities 20.2–44.6% higher) than with their orthologs from other taxa, suggesting some degree of specialization of these proteins within the Cestoda class. As expected, these “non-exclusive” proteins are monophyletic for tapeworms, and nine of them had orthologs identified only for flatworm species, including tapeworms and flukes. Therefore, within the set of 34 proteins selected as being strobilation-related, 17 are assumed to be related to cestode-exclusive functions, whereas 9 are assumed to be associated with functions exclusive to the phylum Platyhelminthes. Further studies are necessary to determine the possible specialized functions of these proteins in tapeworm or flatworm biology.

In this set of 34 strobilation-related proteins, only 8 had orthologs found in non-Platyhelminth lophotrochozoans. The other 26 proteins were conserved only among platyhelminths (with 17 of them being conserved

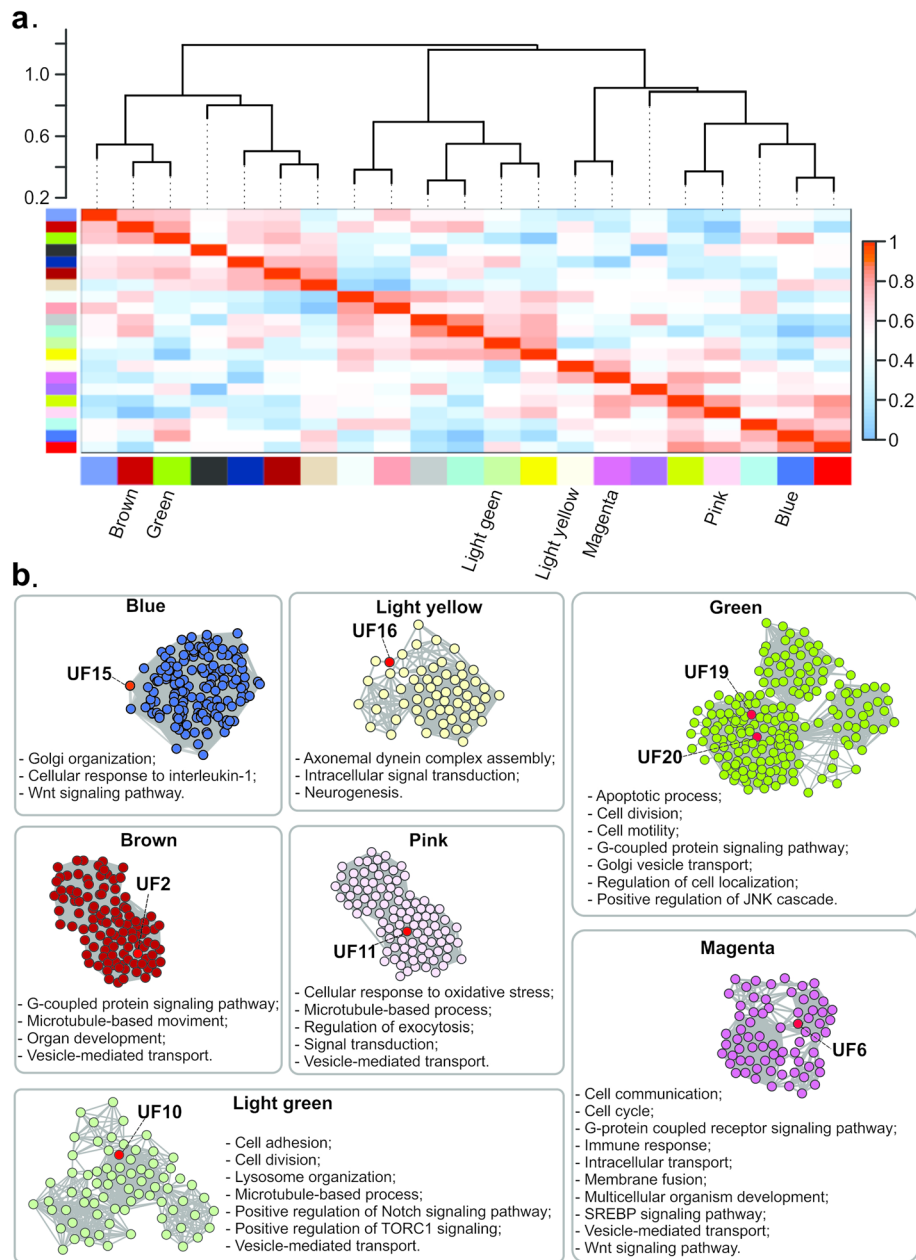


Fig. 5 Analysis of gene co-expression networks and functional predictions for strobilation-related proteins of unknown function (UF). **(a)** Module relationships are represented by the dendrogram and the module Eigengene values of all module comparisons, as represented in the heatmap. Each identified module is represented by a different color and the 7 modules containing hub UF proteins are color-named. **(b)** Each box shows the results of the selected module of the corresponding color. UF proteins are named and highlighted in red in the module PPI networks. Functions predicted for each module are listed

only among cestodes). This suggests that any ancestral lophotrochozoan segmentation mechanism [3] may have changed or gained complexity with the addition of gene/proteins during platyhelminth or cestode evolution. The phylogenetic analyses performed for these 34 strobilation-related proteins showed that their evolutionary history agrees with the currently accepted monophyletic origin of tapeworms [26].

It was interesting to evaluate the synonymous versus non-synonymous substitution rates regarding the set of strobilation-related proteins, as low variability is assumed for developmental genes due to the strong functional constraints they are submitted to [27]. According to the positive selection analyses, the assessed sequences did not show statistically significant high ω values, indicating that they are probably under functional

constraints. This is consistent with what is expected for genes involved in development and other vital processes [28, 29] (supplementary Table S7).

We also investigated whether any of the strobilation-related proteins were under positive selection, considering that at least some of them could be potential targets for the development of novel drugs against flatworm parasites. A stronger positive selection may determine amino acid variations in ortholog proteins, even among closely related species. Therefore, this type of selection would be undesirable in proteins that are potential drug targets [29]. Variations in orthologs of related pathogenic species (for example, among tapeworm or fluke species) could indicate that some of these species would be responsive to the drug, whereas others would be less, or even not at all, responsive [30]. Among the analyzed proteins, there was only one site under positive selection in one protein (UF16). Therefore, almost all the proteins analyzed would be interesting potential targets for broad spectrum drugs in cestodiasis.

As most of these strobilation-related proteins had no functional annotation (22 UF proteins), we performed a systems biology analysis to identify functional modules with the aim of providing additional evidence of their involvement in developmental processes. For cestodes, an important limitation for this type of approach is the scarcity of adequate RNA-seq data, considering the need for large numbers of replicates with low variation among them. Fortunately, RNA-seq data available for the *E. multilocularis* pre-adult (protoscolex) stage matched all technical requirements related to the replicate number and data quality. This allowed us to allocate eight of the UF proteins, which are so far not characterized, as central nodes (hubs) from seven functional network modules. These seven modules were associated with biological functions such as cell signaling, apoptosis, cell adhesion, and transcriptional regulation. Unfortunately, functional assignments were not possible for the remaining 14 UF proteins with the available transcriptomic data. A possible explanation for this could be the absence, or low expression, of these strobilation-related genes/proteins in bona fide protoscolexes, a non-strobilated stage.

Domain analyses of UF proteins allowed the identification of the UF 3 protein as a putative G protein-coupled receptor and the identification of putative transmembrane regions for eight UF proteins, suggesting that these proteins may have strobilation-related functions on the cell surface. Interestingly, transmembrane domains were found in UF 6 (related to the G-protein coupled receptor, SREBP and Wnt signaling pathways), UF 15 (related to the Wnt signaling pathway), UF 19 and UF 20 (related to the G-protein coupled signaling pathway). Additionally, co-expression and PPI analyses

allowed the association of several UF proteins with important cell signaling pathways, namely Wnt (UF 6 and UF 15), G-protein coupled receptor signaling (UF 2, UF 6, UF 19, and UF 20), Notch (UF 10), SREBP (UF 6), and TORC1 (UF 10). These data allow us to infer that these UF proteins play strobilation-related functions in cell-cell interactions mediated by one of these signaling pathways. The involvement of selected UF proteins with strobilation is suggested based on evidence from a more recent proteomic survey carried out by our group with *M. corti* (Camargo de Lima, J., Floriani, M.A., Debarba, J.A., Monteiro, K.M., Moura, H., Barr, J.R., Zaha, A., Ferreira, H.B., personal communication). This study showed that UF 16 is among the proteins that are newly synthesized within 24 h of strobilation induction.

Among the strobilation-related proteins with functional annotation, Groucho, Mark2, and TCF/LCF are components of the Wnt signaling pathway. As discussed above, two UF proteins, UF 6 and UF 15, were also assigned to the same PPI network that includes this pathway. The Wnt pathway is well conserved among metazoans [31] and has already been associated with several different developmental events in diverse organisms. Among flatworms, Wnt signaling was initially involved in planarian anterior-posterior axis (head/tail) specification during regeneration [32]. Moreover, in *E. multilocularis*, Wnt signaling inhibition has been related to protoscolex generation and specification of the primary (antero-posterior) body axis during larval metamorphosis [5, 33]. Our data now provide evidence that this pathway is also required for cestode strobilation.

Evidence for the association of the TGF- β /BMP signaling pathway with strobilation arose from the identification of two of its components, BMP2 and SMAD4, among the selected set of strobilation-related proteins. In metazoans, TGF- β /BMP signaling, including the BMP protein signal and the Smad family of transcription factors, regulates a wide variety of cellular processes, such as proliferation, differentiation, adhesion, migration, and apoptosis [34] and developmental events, such as body axis formation and regeneration [35]. Moreover, in *E. granulosus*, Smad4 transcription factor is expressed in the metacestode (cystic larval stage) and in protoscolexes in a tissue-specific manner, with the highest transcript levels being found in activated (strobilation-induced) protoscolexes [36]. Taken together, all this evidence corroborates the contribution of the TGF- β /BMP signaling pathway in cestode strobilation.

The G-protein coupled signaling pathway also seems to be related to strobilation, based on the initial selection of the NPR1 receptor. Moreover, the UF 3 protein also exhibited the classical domains found in receptors of this pathway. In line with this, the G-protein coupled receptor signaling pathway has been shown to be involved in

invertebrate developmental processes [37], including movement, development, reproduction, and plasticity of the whole worm and neuronal development in *E. granulosus* [38, 39]. It is also notable that serotonergic G-protein coupled receptors have great high potential as drug targets for antiparasitic intervention [39]. However, in *E. granulosus* and other cestodes, many proteins of the G-protein coupled pathway remain unknown because of the considerable sequence divergence among the invertebrate orthologs [38]. Interestingly, based on the transcript co-expression and PPI analyses, we were able to relate UF 2, UF 6, UF 19, and UF 20 proteins to this pathway. Also, according to our differential expression analyses, NPR1, UF 2, UF 6, and UF19 are upregulated in the larval stage. In contrast, UF 3 and UF 20 are up-regulated in the adult stage, suggesting the involvement of the G-protein coupled pathway in different processes related to these developmental stages of cestodes.

Finally, transcript co-expression and systems biology analyses associated UF 2, UF 6, UF10, UF 11, UF 15, UF 19, and UF 20 proteins with vesicle-mediated, or Golgi vesicle, transport. In addition, the UF 15 protein seems to be part of the Golgi organization and has already been identified in a proteomic survey of *E. granulosus* extracellular vesicles (da Silva, E.D., Battistella, M.E., dos Santos, G.B., Monteiro, K.M., Cancela, M., Ferreira, H.B., Zaha, A., personal communication). Additionally, the selected GAK protein is involved in clathrin-dependent trafficking from the trans Golgi network [40]. Vesicle transport has been related to cell signaling during development, as it would facilitate the long-distance trafficking of signaling molecules [41]. Overall, these results suggest that vesicle transport may also be important in signaling pathways involved in cestode strobilation.

Since the vital importance of developmental processes for parasite biology and reproduction, the identified proteins can be considered as targets for the development of novel therapeutic strategies in cestodiasis control. For instance, proteins such as enolase, histone deacetylases, pyruvate kinase, and serotonergic G-protein coupled receptors have been proposed as targets for cestodiasis therapeutic interventions [9, 38, 42, 43]. Here, we have provided details of at least 12 developmental proteins with the potential to be targets for drug repositioning and the design of novel anthelmintic drugs. For instance, Mark2 and MAGI2 already have orthologs with available structural data that could be used for structural modeling, binding site prediction, and drug-protein molecular docking studies [44]. Likewise, the strobilation-related proteins identified as being cestode-exclusive could also constitute an interesting set of proteins for functional studies and for design-specific therapeutic approaches to cestodiasis. Moreover, all proteins identified here as being related to strobilation have

the potential to be molecular markers of cestode development, provided they are further characterized regarding their spatial and temporal expression patterns in a suitable model system, such as *M. corti* [45].

Conclusion

In summary, we performed comprehensive evolutionary and functional analyses using plathyhelminth genomic and transcriptomic data to determine the relationships between segmented and non-segmented species. Such analyses allowed the successful identification of a set of 34 evolutionary conserved cestode proteins, with known (12 proteins) or unknown (22 proteins) function, as possible components of developmental pathways required for strobilation. Several of these proteins could be functionally assigned to cell signaling pathways like Wnt, TGF- β /BMP, and G-protein coupled receptor pathways, linking them to strobilation. Moreover, most of the strobilation-related UF proteins were exclusive of plathyhelminths or cestodes, and they may act in specialized segmentation mechanisms that evolved and operate in these organisms. It was also found that virtually none of the 34 strobilation-related genes are under positive selection. These results provide further information on the molecular mechanisms and evolution of the cestode strobilation process. Our data also highlighted several proteins of interest for future functional studies as potential developmental markers and/or targets for the development of novel antihelminthic drugs.

Methods

Orthogroup identification

Genome data used for the identification of orthologous genes were obtained from the following public databases: National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>), Sanger Institute (<http://www.sanger.ac.uk/>), SchistoDB (<http://schistodb.net/schisto/>), WormBase (<http://www.wormbase.org/#012-34-5>), and WormBase ParaSite (<https://parasite.wormbase.org/species.html>). The species used in this study were the tapeworms *E. granulosus*, *E. multilocularis*, *H. microstoma*, *M. corti*, and *T. solium*; the flukes *C. sinensis*, *S. haematobium*, *S. japonicum*, *S. mansoni*, and *O. viverrini*; the nematodes *C. elegans*, *G. pallida*, *H. contortus*, *O. volvulus*, *S. ratti*, and *T. muris*; the annelid *H. robusta*; and the mollusk *L. gigantea*. Database accession numbers and references of the respective genome sequences are available in the supplementary Table S16. The OrthoMCL algorithm v2.0.8 [46] was used with default parameters to identify groups of orthologs and paralogs, hereinafter called orthogroups, among the whole deduced proteomes of all 18 organisms listed above.

Phylogenomic analysis

A Python (<https://www.python.org/>) script was developed to select, from the OrthoMCL output (section above), all orthogroups with at least a representative sequence for each of the 18 assessed organisms (supplementary data 1). If any given organism was found to have paralogous sequences, only the longest sequence was used for further analysis, keeping only one orthologous sequence for any given protein for each of the 18 organisms. The resulting multi-FASTA ortholog files for each protein sequence were used as the input for multiple alignments, performed using the CLUSTAL Omega algorithm [47] with default parameters. Subsequently, the SCAfos software [48] was used to concatenate the aligned files of amino acid sequences. The selection of supermatrix best-fit model of protein evolution was performed using ProtTest 3 [49]. The MrBayes v3.2.2 software [50] was used to construct the phylogenomic tree, with two runs, four chains in parallel, 25,000,000 generations, sampling every 100 generations, with a burn-in of 25%, and a stopval of 0.01 for the control of topological convergence.

Identification of proteins associated with strobilated species or life-cycle stages

The orthogroups that had orthologous proteins in all five of the assessed cestode (strobilated) species, but lacked orthologs in at least one of the assessed trematode (non-strobilated) species, were selected using the Python script shown in supplementary data 2. Next, the resulting set of orthogroups associated with strobilated species were categorized and functionally enriched based on the BLAST sequence homologies and gene ontology (GO) annotations using the Blast2GO software [51]. The orthogroups were separated into two subsets: one with proteins of known functions and one with proteins of unknown functions (UF). Among the proteins in the known functions subset, we selected only the development-associated orthogroups.

Published transcriptomic data for *E. multilocularis*, *H. microstoma* [12], and *M. corti* [15] were used to identify proteins from selected orthogroups that had genes with different expressions between larval (pre-strobilated) and adult (strobilated) stages. Only bona fide larvae and adult worms were compared, as at present there is no RNA-Seq data available for specimens undergoing strobilation. To avoid differentially expressed genes that were not related to strobilation, those orthologs with differential expression between larval and adult stages of the flukes *S. haematobium* [20] and *S. mansoni* [21] were excluded from further analyses.

Considering the hypothesis that the segmentation mechanism underlying strobilation came from an ancestral lophotrochozoan mechanism [4], whether the

selected strobilation-related proteins have orthologs in other species of the Lophotrochozoa superphylum was investigated. To achieve this, the proteins identified as being differentially expressed in the cestode larval and adult stages were used as query sequences in searches for additional orthologous sequences in other lophotrochozoan species. Searches were performed in the non-redundant database of NCBI-Genbank (<https://www.ncbi.nlm.nih.gov/refseq/>) using the blastp suite (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) and in the UniProtKB database (<http://www.uniprot.org/>) using the HMMER phmmer tool (<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>). Only orthologous sequences with identities and coverages greater than 30 and 70%, respectively, were selected. Functional domain annotation of orthologous proteins was performed using InterProScan 5 version 57.0 [52], and only those orthologous proteins in accordance with the functional domain profile of the orthogroup they were assigned to were used in further analyses.

Phylogenetic and positive selection analyses

Multiple coding DNA sequences (CDSs) or deduced protein sequences were aligned using the CLUSTAL Omega algorithm [47] and the PRANK program [53]. CLUSTAL alignments were performed with default parameters, guided by an external hidden Markov model (HMM). PRANK alignments were performed using two algorithm variants, one based on an amino acid model (PRANK_{AA}) and the other based on an empirical codon model (PRANK_C). Nucleotide alignments were obtained using the PAL2NAL program [54]. Low quality regions of the generated sequence alignments were individually inspected, and manually adjusted, when possible, or removed, when necessary. The final nucleotide and amino acid alignments for all orthogroups were used in the phylogenetic analyses. Selection of the best-fit models of nucleotide and amino acid evolution was performed using the MEGA X software suite [55].

Phylogenetic trees were generated for each orthogroup by distance and probabilistic methods using the MEGA X software suite. For distance methods, the Neighbor-Joining algorithm was used, with pairwise deletion of gaps applied using the p-distance and Poisson evolutionary models for the amino acid sequence evaluation and the p-distance and Jukes-Cantor models for the nucleotide sequence evaluation. For the probabilistic method, the maximum likelihood algorithm with pairwise deletion of gaps was applied. The bootstrap tests of distance and probabilistic phylogenies were performed using 2000 repetitions for all analyses. Phylogenetic trees were also generated for each orthogroup by the bayesian method using the MrBayes v3.2.2 software, with two runs, four chains in parallel, 25,000,000 generations,

sampling every 100 generations, with a burn-in of 25%, and a stopval of 0.01 for the control of topological convergence. The TreeView program [56] was used to visualize and edit the results of all generated phylogenies. The best phylogenetic tree estimated for each gene was selected based on its statistical support and on the agreement with the expected evolution for these species, as indicated by the phylogenomic tree.

Positive selection analyses were performed using the codeml program in the PAML 4 software [57]. The site-specific model analysis was implemented using the M0, M1a, M2a, M3, M7, and M8 nested models. For all models, a Bayes empirical Bayes (BEB) approach was employed to detect codons with a posterior probability of positive selection > 99% [58]. Positive selection is detected when $\omega > 1$, which means non-synonymous substitutions rates (dN) are higher than synonymous substitutions rates (dS).

Gene co-expression network analysis

Two RNA-seq datasets of the *E. multilocularis* pre-adult stage were used as inputs for the gene co-expression analyses. The first one was recovered from the ArrayExpress database (<https://www.ebi.ac.uk/arrayexpress/>, accession number E-ERAD-50), and the second one was recovered from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>, accession number GSE59173). Low quality reads were filtered using Trimmomatic v0.36 [59] and mapped on the *E. multilocularis* reference genome version 3 (<ftp://ftp.sanger.ac.uk/pub/project/pathogens/Echinococcus/multilocularis/genome/>), using the STAR 2.5.3a software [60]. The HTSeq package Version 0.8.0 [61], with intersection-nonempty counting parameter, was used to generate the gene count estimations. Expression data were imported into the R environment and normalized by a variance-stabilizing transformation method, available in the DESeq2 v1.18.1 package [62]. Finally, batch effect correction was performed utilizing the ComBat R package [63], implemented by the SampleNetwork R function created by Oldham et al. [64].

Gene co-expression calculation data was performed using the WGCNA R package, with the biweight mid-correlation method [65]. To provide a co-expression network topology close to a scale-free network, typical of many biological networks [66], a soft thresholding measurement was performed by raising co-expression values to a β power. The β power that generated a graph with the closest free-scale topology was selected. Next, a hierarchical clustering method, followed by a dynamical branch cutting algorithm, was chosen to identify modules. On each cluster, gene expression values were summarized into module eigengenes (ME). Intramodular hub genes (named as kME) were calculated by

correlating the gene expression and the corresponding ME. Genes with kME values higher than 0.85 were considered as hub genes and were used as input to generate PPI networks via STRING (<http://string-db.org>) searches using the medium confidence interaction score (0.40) and excluding text mining source. For each module of interest, the PPI network generated by STRING was joined to the WGCNA co-expression network and only connections with a value > 0.4 were selected. The resulting co-expression and PPI networks were visualized using the Cytoscape software (v3.6.0) [67]. The sets of proteins in these networks were then used for the functional enrichment analyses using the Blast2GO software [51].

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-06878-3>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Abbreviations

BEB: Bayes empirical Bayes; CDS: Coding DNA sequence; HMM: Hidden Markov model; ME: Module eigengenes; PPI: Protein-protein interaction; UF: Unknown function

Acknowledgments

The authors are thankful to Dr. Magdalena Zarowiecki and the Wellcome Trust Sanger Institute (UK) for providing the access to *M. corti* genome data. The access to high-performance computing facilities granted by the Laboratório Nacional de Computação Científica (LNCC, Brazil) is also gratefully acknowledged.

Authors' contributions

GPP, CET, HBF, KNM, AZ and ATV designed the experimental approach. GPP, CET, and RLMG performed the identification and selection of orthogroups. GPP and CET performed the positive selection, phylogenetic and phylogenomic analyses. GPP and KNM performed the gene co-expression networks analysis. GPP, CET, MC and HBF analyzed and interpreted the data. GPP, CET and HBF wrote the manuscript. KNM, AZ and MC revised the manuscript. All authors have read and approved the manuscript.

Funding

This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (grant number 428305/2018–0) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES (grant number 611 23038.010043/2013–02). GPP is a recipient of a CAPES Ph.D. fellowship. The funders had no role in the design of the study, in the collection, analysis, and interpretation of data, in the decision to publish and in writing the manuscript.

Availability of data and materials

All data and materials used in this research are publicly available. Genome web links are available in Supplementary Table S16. Transcriptomic data are available from <https://www.ebi.ac.uk/arrayexpress/experiments/E-ERAD-50/> and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE59173>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no conflicts of interests (financial/non-financial) concerning this article.

Author details

¹Laboratório de Genômica Estrutural e Funcional, Centro de Biotecnologia (CBiot), Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil. ²Programa de Pós-Graduação em Biologia Celular e Molecular, CBiot, UFRGS, Porto Alegre, RS, Brazil. ³Departamento de Farmacociências, Universidade Federal de Ciências Médicas de Porto Alegre, Porto Alegre, RS, Brazil. ⁴Laboratório Nacional de Computação Científica, Petrópolis, RJ, Brazil. ⁵Present address: Instituto Hermes Pardini, Vespasiano, MG, Brazil.

Received: 30 August 2019 Accepted: 2 July 2020

Published online: 16 July 2020

References

- Littlewood DTJ. The evolution of parasitism in flatworms. In: Parasitic flatworms: molecular biology, biochemistry, immunology and physiology. Wallingford: CAB; 2006. p. 1–36. <https://doi.org/10.1079/9780851990279.0001>.
- Kenny NJ, Noreña C, Damborenea C, Grande C. Probing recalcitrant problems in polyclad evolution and systematics with novel mitochondrial genome resources. Genomics. 2019;111:343–55. <https://doi.org/10.1016/j.ygeno.2018.02.009>.
- Lockyer AE, Olson PD, Littlewood DTJ. Utility of complete large and small subunit rRNA genes in resolving the phylogeny of the Neodermata (Platyhelminthes): implications and a review of the cercomer theory. Biol J Linn Soc. 2003;78:155–71. <https://doi.org/10.1046/j.1095-8312.2003.00141.x>.
- Olson PD, Timothy D, Littlewood J, Bray RA, Mariaux J. Interrelationships and evolution of the tapeworms (Platyhelminthes: Cestoda). Mol Phylogenet Evol. 2001;19:443–67.
- Kozioł U, Jarero F, Olson P, Brehm K. Comparative analysis of Wnt expression identifies a highly conserved developmental transition in flatworms. BMC biology. 2016;14:10. <https://doi.org/10.1186/s12915-016-0233-x>.
- Couso JP. Segmentation, metamerism and the Cambrian explosion. Int J Dev Biol. 2009;53:8–10.
- Kozioł U. Evolutionary developmental biology (evo-devo) of cestodes. Exp Parasitol. 2017;180:84–100. <https://doi.org/10.1016/j.exppara.2016.12.004>.
- Cheng Z, Liu F, Li X, Dai M, Wu J, Guo X, et al. EGF-mediated EGFR/ERK signaling pathway promotes germinative cell proliferation in *Echinococcus multilocularis* that contributes to larval growth and development. PLoS Negl Trop Dis. 2017;11:e0005418. <https://doi.org/10.1371/journal.pntd.0005418>.
- Vaca HR, Celentano AM, Macchiaroli N, Kamenetzky L, Camicia F, Rosenzvit MC. Histone deacetylase enzymes as potential drug targets of neglected tropical diseases caused by cestodes. Int J Parasitol Drugs Drug Resist. 2019;9:120–32. <https://doi.org/10.1016/j.ijpddr.2019.02.003>.
- Hemphill A, Stadelmann B, Rufener R, Spiliotis M, Boubaker G, Müller J, et al. Treatment of echinococcosis: albendazole and mebendazole – what else? Parasite. 2014;21:70. <https://doi.org/10.1051/parasite/2014073>.
- Kozioł U, Brehm K. Recent advances in *Echinococcus* genomics and stem cell research. Vet Parasitol. 2015;213:92–102. <https://doi.org/10.1016/j.vetpar.2015.07.031>.
- Tsai IJ, Zarowiecki M, Holroyd N, Garcarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. Nature. 2013;496:57–63. <https://doi.org/10.1038/nature12031>.
- Howe KL, Bolt BJ, Cain S, Chan J, Chen WJ, Davis P, et al. WormBase 2016: expanding to enable helminth genomic research. Nucleic Acids Res. 2016;44:D774–80. <https://doi.org/10.1093/nar/gkv1217>.
- Basika T, Macchiaroli N, Cucher M, Espinola S, Kamenetzky L, Zaha A, et al. Identification and profiling of microRNAs in two developmental stages of the model cestode parasite *Mesocestoides corti*. Mol Biochem Parasitol. 2016. <https://doi.org/10.1016/j.molbiopara.2016.08.004>.
- Bizarro CV, Bengtson MH, Ricachenevsky FK, Zaha A, Sogayar MC, Ferreira HB. Differentially expressed sequences from a cestode parasite reveals conserved developmental genes in platyhelminthes. Mol Biochem Parasitol. 2005;144:114–8. <https://doi.org/10.1016/j.molbiopara.2005.07.002>.
- Basika T, Paludo GP, Araujo FM, Salim AC, Pais F, Maldonado L, et al. Transcriptomic profile of two developmental stages of the cestode parasite *Mesocestoides corti*. Mol Biochem Parasitol. 2019;229:35–46. <https://doi.org/10.1016/j.molbiopara.2019.02.006>.
- Camargo de Lima J, Monteiro KM, Basika Cabrera TN, Paludo GP, Moura H, Barr JR, et al. Comparative proteomics of the larval and adult stages of the model cestode parasite *Mesocestoides corti*. J Proteomics. 2018;175:127–35. <https://doi.org/10.1016/j.jprot.2017.12.022>.
- Laschuk A, Monteiro KM, Vidal NM, Pinto PM, Duran R, Cerveñanski C, et al. Proteomic survey of the cestode *Mesocestoides corti* during the first 24 hours of strobilar development. Parasitol Res. 2011;108:645–56. <https://doi.org/10.1007/s00408-5>.
- Debarba JA, Monteiro KM, Moura H, Barr JR, Ferreira HB, Zaha A. Identification of newly synthesized proteins by *Echinococcus granulosus* Protoscolices upon induction of strobilation. PLoS Negl Trop Dis. 2015;9:e0004085. <https://doi.org/10.1371/journal.pntd.0004085>.
- Young ND, Jex AR, Li B, Liu S, Yang L, Xiong Z, et al. Whole-genome sequence of *Schistosoma haematobium*. Nat Genet. 2012;44:221–5. <https://doi.org/10.1038/ng.1065>.
- Protasio AV, Tsai IJ, Babbage A, Nichol S, Hunt M, Aslett MA, et al. A systematically improved high quality genome and Transcriptome of the human blood fluke *Schistosoma mansoni*. PLoS Negl Trop Dis. 2012;6:e1455. <https://doi.org/10.1371/journal.pntd.0001455>.
- van Dam S, Vösa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2018;19:575–92. <https://doi.org/10.1093/bib/bbw139>.
- Manzoni C, Kia DA, Vandrovicova J, Hardy J, Wood NW, Lewis PA, et al. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. Brief Bioinform. 2018;19:286–302. <https://doi.org/10.1093/bib/bbw114>.
- Steinberg J, Ritchie GRS, Roumeliotis TI, Jayasuriya RL, Clark MJ, Brooks RA, et al. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. Sci Rep. 2017;7:8935. <https://doi.org/10.1038/s41598-017-09335-6>.
- Du Y, Clair GC, Al Alam D, Danopoulos S, Schnell D, Kitzmiller JA, et al. Integration of transcriptomic and proteomic data identifies biological functions in cell populations from human infant lung. Am J Physiol Cell Mol Physiol. 2019;317:L347–60. <https://doi.org/10.1152/ajplung.00475.2018>.
- International Helminth Genomes Consortium. Comparative genomics of the major parasitic worms. Nat Genet. 2019;51:163–74. <https://doi.org/10.1038/s41588-018-0262-1>.
- Dall'Olio G, Laayouni H, Luisi P, Sikora M, Montanucci L, Bertranpetit J. Distribution of events of positive selection and population differentiation in a metabolic pathway: the case of asparagine N-glycosylation. BMC Evol Biol. 2012;12:98. <https://doi.org/10.1186/1471-2148-12-98>.
- Georgi B, Voight BF, Bućan M. From mouse to human: evolutionary genomics analysis of human Orthologs of essential genes. PLoS Genet. 2013;9:e1003484. <https://doi.org/10.1371/journal.pgen.1003484>.
- Jordan IK, Rogozin IB, Wolf YI, Koonin EV. Essential genes are more evolutionarily conserved than are nonessential genes in bacteria. Genome Res. 2002;12:962–8. <https://doi.org/10.1101/gr.87702>.
- Vamathevan JJ, Hasan S, Emes RD, Amrine-Madsen H, Rajagopalan D, Topp SD, et al. The role of positive selection in determining the molecular cause of species differences in disease. BMC Evol Biol. 2008;8:273. <https://doi.org/10.1186/1471-2148-8-273>.
- Croce JC, McClay DR. Evolution of the Wnt pathways. Methods Mol Biol. 2008;469:3–18. https://doi.org/10.1007/978-1-60327-469-2_1.
- Lin AYT, Pearson BJ. Planarian yorkie/YAP functions to integrate adult stem cell proliferation, organ homeostasis and maintenance of axial patterning. Development. 2014;141:1197–208. <https://doi.org/10.1242/dev.101915>.
- Montagne J, Preza M, Castillo E, Brehm K, Kozioł U. Divergent Axin and GSK-3 paralogs in the beta-catenin destruction complexes of tapeworms. Dev Genes Evol. 2019;229:89–102. <https://doi.org/10.1007/s00427-019-00632-w>.
- Zavala-Góngora R, Kroner A, Bernthaler P, Knaus P, Brehm K. A member of the transforming growth factor-beta receptor family from *Echinococcus multilocularis* is activated by human bone morphogenetic protein 2. Mol Biochem Parasitol. 2006;146:265–71. <https://doi.org/10.1016/j.molbiopara.2005.12.011>.
- Epping K, Brehm K. *Echinococcus multilocularis*: molecular characterization of EmSmadE, a novel BR-Smad involved in TGF-β and BMP signaling. Exp Parasitol. 2011;129:85–94. <https://doi.org/10.1016/j.exppara.2011.07.013>.
- Zhang C, Wang L, Wang H, Pu H, Yang L, Li J, et al. Identification and characterization of functional Smad8 and Smad4 homologues from *Echinococcus granulosus*. Parasitol Res. 2014;113:3745–57. <https://doi.org/10.1007/s00436-014-4040-4>.

37. Meeusen T, Mertens I, De Loof A, Schoofs L. G protein-coupled receptors in invertebrates: a state of the art. *Int Rev Cytol.* 2003;230:189–261 <http://www.ncbi.nlm.nih.gov/pubmed/14692683>.
38. Camicia F, Herz M, Prada LC, Kamenetzky L, Simonetta SH, Cucher MA, et al. The nervous and pre-nervous roles of serotonin in *Echinococcus* spp. *Int J Parasitol.* 2013;43:647–59. <https://doi.org/10.1016/j.ijpara.2013.03.006>.
39. Camicia F, Celentano AM, Johns ME, Chan JD, Maldonado L, Vaca H, et al. Unique pharmacological properties of serotonergic G-protein coupled receptors from cestodes. *PLoS Negl Trop Dis.* 2018;12:e0006267. <https://doi.org/10.1371/journal.pntd.0006267>.
40. Zhang CX, Engqvist-Goldstein AEY, Carreno S, Owen DJ, Smythe E, Drubin DG. Multiple roles for cyclin G-associated kinase in clathrin-mediated sorting events. *Traffic.* 2005;6:1103–13. <https://doi.org/10.1111/j.1600-0854.2005.00346.x>.
41. Gradilla A-C, Simon E, Aguilar G, Guerrero I. From intra- to extracellular vesicles: extracellular vesicles in developmental signalling. *Essays Biochem.* 2018;62:215–23. <https://doi.org/10.1042/EBC20180001>.
42. Coghlan A, Tyagi R, Cotton JA, Holroyd N, Rosa BA, Tsai IJ, et al. Comparative genomics of the major parasitic worms. *Nat Genet.* 2019;51:163–74.
43. Sulima A, Savijoki K, Bień J, Näreaho A, Salamatin R, Conn DB, et al. Comparative proteomic analysis of *Hymenolepis diminuta* cysticercoid and adult stages. *Front Microbiol.* 2017;8:2672. <https://doi.org/10.3389/fmicb.2017.02672>.
44. Govindaraj RG, Naderi M, Singha M, Lemoine J, Brylinski M. Large-scale computational drug repositioning to find treatments for rare diseases. *NPJ Syst Biol Appl.* 2018;4:13. <https://doi.org/10.1038/s41540-018-0050-7>.
45. Costa CB, Monteiro KM, Teichmann A, da Silva ED, Lorenzatto KR, Cancela M, et al. Expression of the histone chaperone SET/TAF- β during the strobilation process of *Mesocostoides corti* (Platyhelminthes, Cestoda). *Parasitology.* 2015;142:1171–82. <https://doi.org/10.1017/S003118201500030X>.
46. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 2003;13:2178–89. <https://doi.org/10.1101/gr.1224503>.
47. Sievers F, Higgins DG. Clustal omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol.* 2014;1079:105–16. https://doi.org/10.1007/978-1-62703-646-7_6.
48. Roure B, Rodriguez-Ezpeleta N, Philippe H. SCaFoS: a tool for Selection, Concatenation and Fusion of Sequences for phylogenomics. *BMC Evol Biol.* 2007;7 Suppl 1:S2. <https://doi.org/10.1186/1471-2148-7-S1-S2>.
49. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics.* 2011;27:1164–5. <https://doi.org/10.1093/bioinformatics/btr088>.
50. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61:539–42. <https://doi.org/10.1093/sysbio/sys029>.
51. Conesa A, Götz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics.* 2008;2008:619832. <https://doi.org/10.1155/2008/619832>.
52. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–40. <https://doi.org/10.1093/bioinformatics/btu031>.
53. Löytynoja A, Goldman N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics.* 2010;11:579. <https://doi.org/10.1186/1471-2105-11-579>.
54. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 2006;34 Web Server:W609–12. <https://doi.org/10.1093/nar/gkl315>.
55. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547–9. <https://doi.org/10.1093/molbev/msy096>.
56. Page RDM. Visualizing phylogenetic trees using TreeView. In: current protocols in bioinformatics. Hoboken: John Wiley & Sons, Inc.; 2002. <https://doi.org/10.1002/0471250953.bi0602s01>.
57. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007;24:1586–91. <https://doi.org/10.1093/molbev/msm088>.
58. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 2012;8:e1002764. <https://doi.org/10.1371/journal.pgen.1002764>.
59. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114–20. <https://doi.org/10.1093/bioinformatics/btu170>.
60. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
61. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics.* 2015;31:166–9. <https://doi.org/10.1093/bioinformatics/btu638>.
62. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550. <https://doi.org/10.1186/s13059-014-0550-8>.
63. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* 2007;8:118–27. <https://doi.org/10.1093/biostatistics/kxj037>.
64. Oldham MC, Langfelder P, Horvath S. Network methods for describing sample relationships in genomic datasets: application to Huntington's disease. *BMC Syst Biol.* 2012;6:63. <https://doi.org/10.1186/1752-0509-6-63>.
65. Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics.* 2012;13:328. <https://doi.org/10.1186/1471-2105-13-328>.
66. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004;5:101–13. <https://doi.org/10.1038/nrg1272>.
67. Demchak B, Hull T, Reich M, Liefeld T, Smoot M, Ideker T, et al. Cytoscape: the network visualization tool for GenomeSpace workflows. *F1000Res.* 2014;3:151. <https://doi.org/10.12688/f1000research.4492.2>.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

