Universidade Federal do Rio Grande do Sul

Centro de Biotecnologia

Programa de Pós-Graduação em Biologia Celular e Molecular

Estudo filogenômico do desenvolvimento estrobilar em platelmintos da Classe Cestoda

Dissertação de Mestrado

Gabriela Prado Paludo

Porto Alegre, outubro de 2016

Universidade Federal do Rio Grande do Sul

Centro de Biotecnologia

Programa de Pós-Graduação em Biologia Celular e Molecular

Estudo filogenômico do desenvolvimento estrobilar em platelmintos da Classe Cestoda

Dissertação submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para obtenção do grau de Mestre.

Gabriela Prado Paludo

Prof. Dr. Henrique Bunselmeyer Ferreira – Orientador

Dra. Claudia Elizabeth Thompson – Co-orientadora

Porto Alegre, outubro de 2016

Este trabalho foi desenvolvido no Laboratório de Genômica Estrutural e Funcional e na Unidade de Biologia Teórica e Computacional do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul (CBiot/UFRGS), e contou com o apoio financeiro da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

> *“IF IT COULD BE DEMONSTRATED THAT ANY COMPLEX ORGAN EXISTED, WHICH COULD NOT POSSIBLY HAVE BEEN FORMED BY NUMEROUS, SUCCESSIVE, SLIGHT MODIFICATIONS, MY THEORY WOULD ABSOLUTELY BREAK DOWN. BUT I CAN FIND NO SUCH CASE.”*
>
> *— CHARLES DARWIN, THE ORIGIN OF SPECIES*

Sumário

Lista de abreviaturas, símbolos e unidades

BMP-2: proteína morfogenética óssea 2 (de *bone morphogenetic protein 2*)

cAMP: adenosina monofosfatada cíclica

cDNA: DNA complementar

cGTP: guanosina trifosfatada cíclica

CDS: sequência codificante do DNA (de *coding DNA sequence*)

GAK: cinase associada à ciclina G (de *cyclin-g-associated kinase*)

GTP: guanosina trifosfatada

Hox B4a: proteína homeobox Hox B4a (de *homeobox protein Hox B4a*)

LHX1: proteína homeobox Lim 1 (de *Lim homeobox protein 1*)

MAGI2: guanilato-quinase associada à membrana 2 (de *membrane-associated guanilate kinase 2*)

miRNA : microRNA

mRNA: RNA mensageiro

NCBI: *National Center for Biotechnology Information*

NPR1: receptor do peptídeo natriurético atrial 1 (de *atrial natriuretic peptide receptor 1*)

RBMS: proteína com domínio de interação ao RNA de fita simples (de *RNA binding motif single stranded interacting protein*)

SMAD 4: proteína semelhante a “mães contra decapentaplégico homólogo 4” (de *mothers against decapentaplegic homolog 4 like*)

TCF/LCF: proteína pangolin J (de *pangolin J protein*)

TGF-β/BMP: fator de transformação do crescimento beta/ proteína morfogenética óssea (de transforming growth factor-β / bone morphogenetic protein)

Wnt: proteína wingless

Lista de Figuras

Capítulo I

Lista de Tabelas

Capítulo I

Resumo

O Filo Platyhelminthes inclui todos os vermes achatados e contém quatro Classes: Turbellaria, Menogenea, Trematoda e Cestoda. A primeira é composta predominantemente por organismos de vida livre, a segunda por ectoparasitas e as Classes Trematoda e Cestoda são compostas por endoparasitas obrigatórios. Os cestódeos são agentes etiológicos de algumas das principais doenças de seres humanos e animais domésticos, apresentado complexos ciclos de vida que abrangem, pelo menos, dois hospedeiros. Entre as suas adaptações ao parasitismo, alguns cestódeos da Subclasse Eucestoda apresentam repetição seriada dos órgãos reprodutivos (metamerismo) e a segmentação externa destes (proglotização), apresentando, assim, uma enorme capacidade reprodutiva. Porém, pouco se sabe dos aspectos moleculares envolvidos na biologia do desenvolvimento desta estrutura corporal. O presente trabalho descreve as relações evolutivas entre organismos endoparasitas do Filo Platyhelminthes através de análise filogenômica, assim como a interrelação dos platelmintos com demais representantes do Superfilo Lophotrochozoa. Por meio da comparação de dados genômicos, transcritômicos e inferência funcional, este trabalho descreve um total de 34 proteínas associadas ao processo de proglotização, conservadas em platelmintos da Classe Cestoda. Entre estas proteínas, 12 estão relacionadas a processos de desenvolvimento, incluindo vias bem conhecidas como as vias de sinalização da wnt e do TGF-β/BMP. Adicionalmente, a identificação de 22 proteínas hipotéticas conservadas e a descrição de seus domínios, adiciona importantes alvos para o estudo da evolução deste processo de desenvolvimento na Classe Cestoda.

Abstract

The Phylum Platyhelminthes includes all flatworms and contains four classes: Turbellaria, Menogenea, Trematoda, and Cestoda. The first one is predominantly composed of free-living organisms, the second by ectoparasites and the Trematoda and Cestoda Classes are composed of obligatory endoparasites. The cestodes are etiologic agents of some of the major diseases of humans and domestic animals, and present complex life cycles that include at least two hosts. Among its adaptations to parasitism, some cestodes of Eucestoda Subclass have serial repetition of their reproductive organs (metamerism) and external segmentation of these (proglottisation), thus presenting an enormous reproductive capacity. However, little is known about the molecular aspects involved in the biology of development of this kind of body structure. This work describes the evolutionary relationships among endoparasite organisms from Phylum Platyhelminthes through phylogenomic analysis, as well as the interrelationship of flatworms with other species representing the Superphylum Lophotrochozoa. Through genomic data comparison, transcriptomic analysis and functional inference, this work describes a set of 34 proteins associated with the proglottisation process, preserved in flatworms Class Cestoda. Among these proteins, 12 are related to developmental processes, including well described pathways as the Wnt and TGF-β / BMP signaling pathways. Additionally, the identification of 22 conserved hypothetical proteins and the description of its domains adds important targets for the study of the proglottisation evolution in the Class Cestoda.

1. INTRODUÇÃO

1.1. O FILO PLATYHELMINTHES

O Filo Platyhelminthes é composto por uma enorme diversidade de espécies que ocorrem em todos os mares, rios, lagos e em todas as massas continentais. O Filo é constituído pelas Classes Turbellaria, Monogenea, Trematoda e Cestoda (Figura 1). Com a simetria bilateral, ausência de celoma e ânus, estes animais apresentam uma ampla variedade morfológica no que diz respeito ao comprimento, organização e presença de órgãos (Scholz et al., 2009).

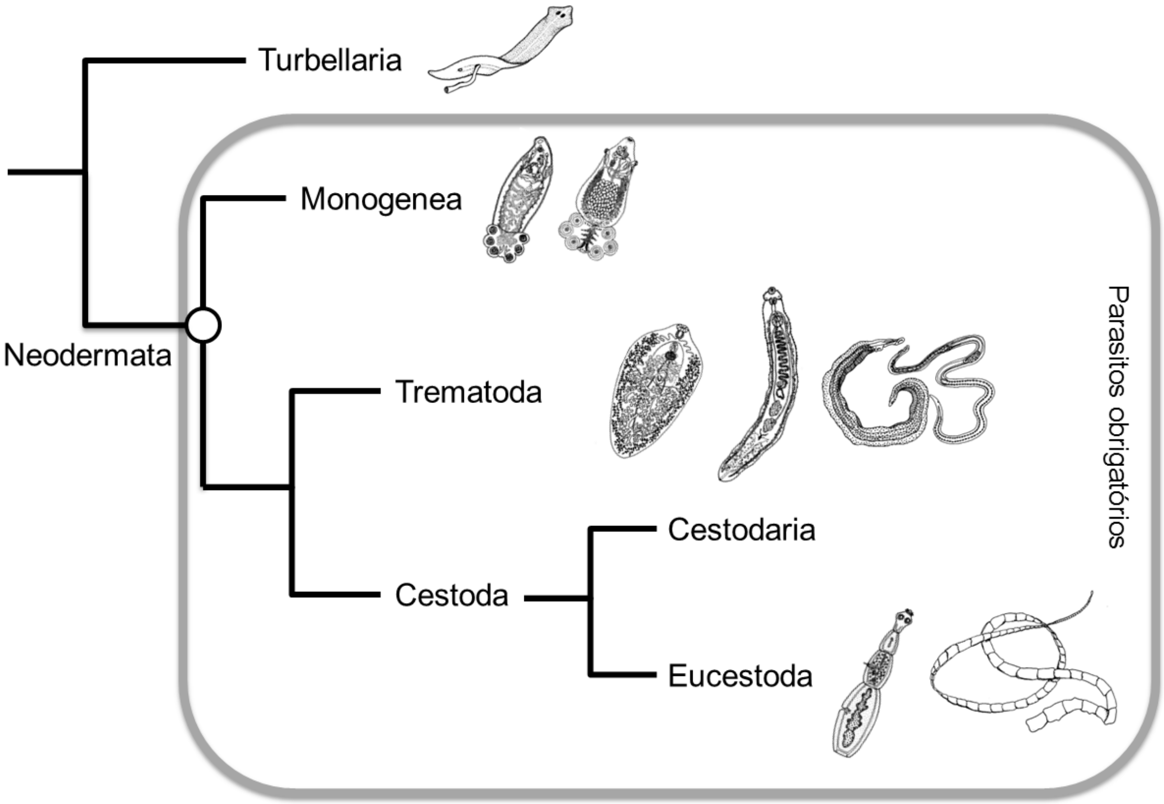


Figura 1. Relações filogenéticas do Filo Phatyhelminthes. Relações filogenéticas das classes do Filo Platyhelminthes e subclasses da classe Cestoda. Classes de parasitos formam um grupo monofilético, o clado Neodermata (Hahn et al., 2014). Figuras provenientes do banco digital ©BIODIDAC (http://biodidac.bio.uottawa.ca/index.htm).

A Classe Turbellaria é composta principalmente por vermes de vida livre, comumente encontrados em ambientes aquáticos. Muitas de suas espécies foram inicialmente descritas como comensais e, posteriormente, algumas passaram a ser descritas como parasitos (Rohde, 1994). Em todo caso, parece evidente a transição entre a vida livre, o comensalismo (com sua estreita associação à outra espécie) e o parasitismo nesta Classe ancestral.

As demais classes de platelmintos são compostas por parasitos obrigatórios de vertebrados, que formam um grupo monofilético chamado Neodermata (Hahn et al., 2014; Lockyer et al., 2003). Este clado é caracterizado por possuir um tegumento secundário em suas formas larvais (Neoderme) que apresenta importantes adaptações ao parasitismo e de defesa contra o hospedeiro, como: o aumento da área de sua superfície (promovendo maior absorção de nutrientes), perda de cílios e sua característica sincicial (permitindo melhor difusão de moléculas) (Dalton et al., 2004).

A classe Monogenea é composta, principalmente, por ectoparasitas de peixes teleósteos. Todos os organismos dessa Classe são dependentes de ambientes aquáticos para o desenvolvimento de seus ovos e a distribuição de suas larvas. Embora predominantemente ectoparasitas, casos de endoparasitismo já foram relatados para organismos da Classe que tendem a se refugiar de ambientes hostis no interior de seus hospedeiros (Kearn, 1994).

As Classes Trematoda e Cestoda são compostas por organismos endoparasitas de vertrebrados (Park et al., 2007). A Classe Trematoda é dividida em duas Subclasses: Aspidograstrea, constituída por aproximadamente 12 gêneros e menos de 100 espécies, e Digenea (Rohde, 2001). A Subclasse Digenea possui o maior número de espécies de trematódeos (com cerca de 18.000 espécies), que apresentam ciclos de vida mais complexos, constituídos por múltiplos hospedeiros (Olson et al., 2003). Neste grupo, estão os agentes etiológicos das principais doenças causadas por trematódeos, como a esquistossomose, causada por espécies do gênero *Schistossoma*, que se estima causar entre 20.000-200.000 mortes de seres humanos por ano de acordo com a *World Health Organization* (http://www.who.int/mediacentre/factsheets/fs115/en/).

A Classe Cestoda inclui endoparasitas de vertebrados e alguns oligoquetos (Heyneman, 1996). Possuem ciclos de vida complexos e são os agentes etiológicos de algumas das principais doenças de seres humanos e animais domésticos. Mais aspectos do parasitismo nessa Classe são discutidos nas próximas sessões.

1.2. PARASITOS CESTÓDEOS E O IMPACTO DAS CESTODÍASES EM SAÚDE HUMANA EM NÍVEL MUNDIAL

Platelmintos parasitas estão entre os agentes infecciosos mais prevalentes no mundo, acometendo, principalmente, seres humanos e animais domésticos de países em desenvolvimento. Há mais de 1.000 espécies conhecidas de platelmintos, a maioria parasita, e praticamente todas as espécies de vertebrados são suscetíveis à infecção por pelo menos uma delas (http://www.earthlife.net/inverts/cestoda.html; Olson et al. 2012).

As doenças causadas por parasitos da Classe Cestoda, cestodíases, estão entre as helmintíases mais prevalentes em todo o mundo. Em seres humanos, apenas os casos relatados e estimados das cestodíases mais comuns ultrapassam os 200 milhões (Tabela 1).

Tabela 1. Prevalência mundial de cestódeos na população humana.

Espécie	Casos	Referência
Diphyllobothrium spp.	20 milhões	(Scholz et al., 2009)
Echinococcus spp.	4 milhões	(Zhang et al., 2016)
Hymenolepis nana	75 milhões	(Muehlenbachs et al., 2015)
Taenia saginata	77 milhões	(Teklemariam & Debash, 2015)
Taenia solium	50 milhões	(Almeida et al., 2009)

Estima-se que as perdas globais determinadas pela hidatidose cística, causada pela forma larval da espécie *Echinococcus granulosus*, e pela cisticercose, causada pela forma larval da *T. solium* em humanos, em termos de *disability-adjusted life years* (DALYs), equivalem às das doenças tropicais negligenciadas mais conhecidas, como a doença de Chagas, a dengue e a tripanossomíase (Budke et al., 2009).

Recentemente, a severidade e os danos causados por cestodíases, levou a *World Helth Oganization* (http://www.who.int/en/) a incluir equinococoses e cisticercose à lista de Doenças tropicais negligenciadas (*Neglected tropical diseases*: http://www.who.int/neglected_diseases/diseases/en/). Essa lista de doenças foi criada visando buscar apoio de organizações de todo o mundo para a busca de tratamentos, controle e formas de erradicação destas cestodíases. Assim, estudos relacionados ao combate destas doenças, assim como elucidação de aspectos biológicos e de relação parasito-hospedeiro dos agentes etiológicos têm sido amplamente realizados (Gabriël et al., 2016; Lorenzatto et al., 2015; Sharma et al., 2016).

1.3. ESTROBILIZAÇÃO COMO UMA ADAPTAÇÃO AO PARASITISMO

Os cestódeos são endoparasitas obrigatórios e, portanto, apresentam características que confirmam sua dependência dos hospedeiros para se desenvolverem. Um exemplo disso é a completa perda de órgãos do sistema digestivo, de forma que o parasito obtém seus nutrientes através da absorção destes do hospedeiro. Todos os cestódeos possuem ao menos dois hospedeiros, embora *Archigetes* possam, ocasionalmente, se desenvolver completamente em seu primeiro hospedeiro, adicionando considerável complexidade a seus ciclos de vida (Figura 2) (Littlewood, 2006). Para completarem seu ciclo, os cestódeos que, frequentemente, sobrevivem a longos períodos de infecção, desenvolveram a capacidade de aumentar seu potencial de reprodução através da repetição seriada dos seus órgãos reprotudivos e, em alguns casos, através de reprodução assexuada com a produção de cistos (Littlewood, 2006).

A Subclasse Cestodaria é formada pelas Ordens Amphilinidea e Gyrocotylidea. Após serem ingeridos por crustáceos, os anfilinídeos atingem sua fase larval e o desenvolvimento para a forma adulta se dá somente através da ingestão do crustáceo por um hospedeiro definitivo adequado (Littlewood, 2006). Em contrapartida, as relações com hospedeiros dos estágios do ciclo de vida dos girocotilídeos ainda não estão elucidadas. Acredita-se que possuam um ciclo de vida direto tendo um peixe como seu hospedeiro (Filo Chordata, Classe Chondrichthyes, Subclasse Holocephali), apesar de haver relatos do seu desenvolvimento no molusco *Mulinia edulis* (Littlewood, 2006).

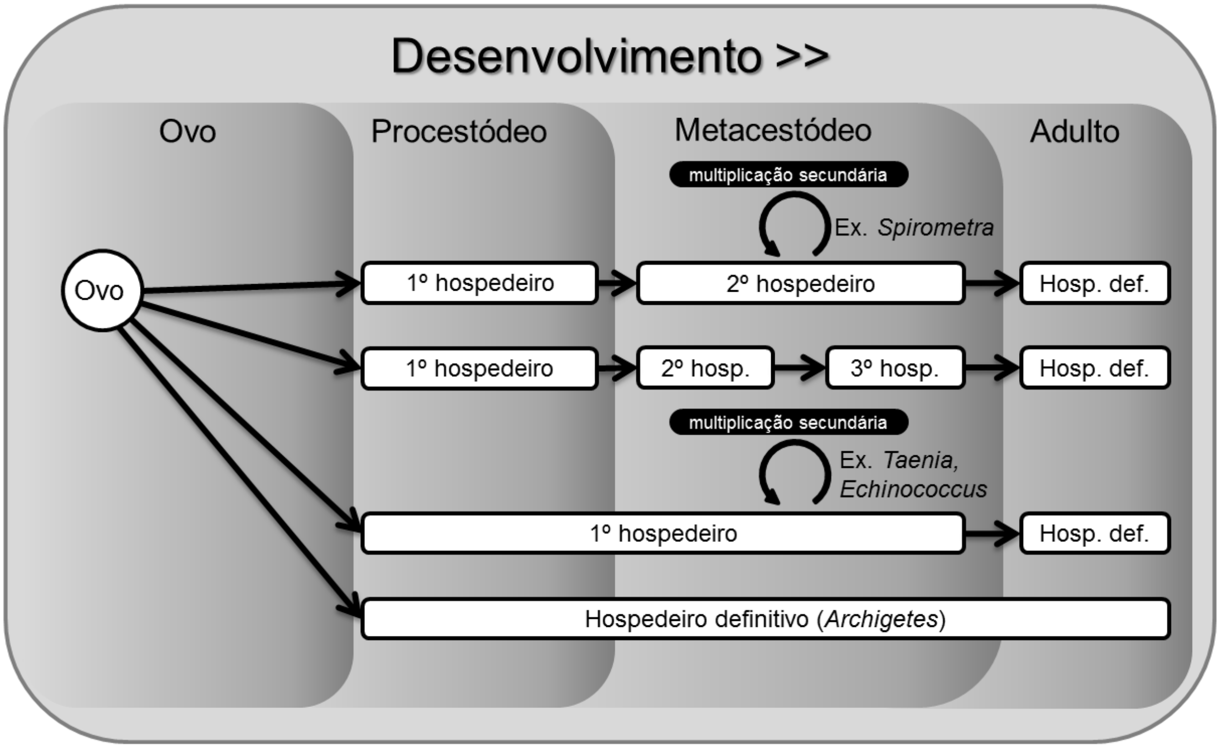


Figura 2. Representação esquemática dos diferentes tipos de ciclo de vida dos cestódeos. Estão indicadas as posições onde desenvolvem-se os principais estágios de vida em relação ao seu hospedeiro. A multiplicação secundária refere-se à multiplicação assexual ocorrida na proliferação do metacestóide. O ciclo de vida do *Archigetes iowensis* (Caryophyllaeidae) pode ocorrer completamente em um único hospedeiro, um anelídeo Oligochaeta, o mesmo ocorre em outras espécies do gênero *Archigetes*. Figura modificada de Littlewood 2006.

Na Subclasse Eucestoda, o ovo é um embrião hexacanto (oncosfera) protegido por envoltórios ovulares (embrióforo) e, para eclodir, o embrióforo precisa ser ingerido e digerido pelas enzimas do primeiro hospedeiro (Chervy, 2002). A oncosfera deve romper o envoltório interno e penetrar na mucosa do hospedeiro pela a ação dos três pares de ganchos (Chervy, 2002). A forma juvenil (metacestóide) se desenvolve no(s) hospedeiro(s) intermediério(s), onde se mantém até que seja ingerida pelo hospedeiro definitivo e atinja a forma adulta.

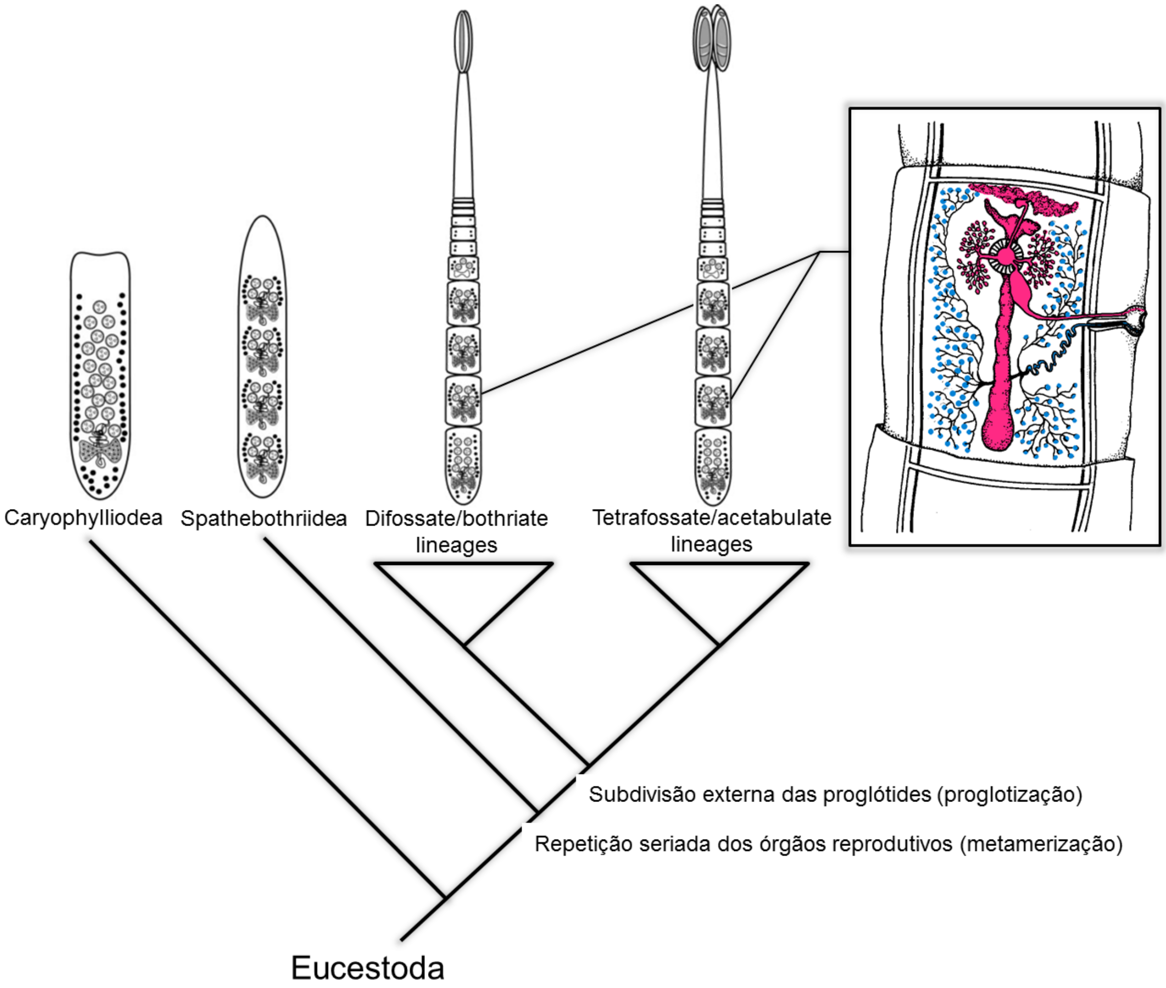


Figura 3. Representação dos passos evolutivos que resultaram na proglotização. O diagrama descreve os passos de segmentação interna (metamerização) dos órgãos reprodutivos e externa (proglotização) da Subclasse Eucestoda (modificado de Olson et al. 2001). À direita está a representação de uma proglótide madura (adquirida do banco digital ©BIODIDAC), mostrando os órgãos do sistema reprodutor feminino (magenta), masculino (azul) e o átrio genital.

Com exceção das Ordens Caryiophylliodea e Spatheobothriidea, os cestódeos em seu estágio adulto possuem uma região anterior (escólex), a partir da qual crescem serialmente as proglótides (Figura 3)(Littlewood, 2006). Desse modo, quanto mais distante do escólex, mais antiga é a proglótide. As proglótides da maioria dos cestódeos são hermafroditas com um ou mais conjuntos de órgãos reprodutivos masculinos e femininos (Figura 3)(Littlewood, 2006).

Como descrita na Figura 3, a evolução da proglotização na Subclasse Eucestoda foi decorrente da condição plesiomórfica na Ordem Caryophylliodea e derivou de forma que a metamerização, repetição seriada dos ógãos reprodutivos, e a segmentação externa, gerando a proglotização, foram eventos evolutivos independentes. Organismos que apresentam ambos os processos de metamerismo e proglotização são chamados estrobilizados. Ambos os processos apresentam potenciais vantagens adaptativas, tal como o aumento da fertilidadeidade gerado pela metamerização (Olson et al., 2001). Já a proglotização promove um aumento da fecundidade, podendo permitir que a fertilização ocorra em diferentes regiões do ambiente em que o parasito se encontra (como o intestino), através de fertilização cruzada.

Do ponto de vista da diversidade de espécies conhecidas, poucas linhagens são conhecidas para as Ordens Caryophylliodea e Spathebothriidea, em contrapartida, mais de 600 gêneros de organismos estrobilizados já foram descritos (Olson et al., 2001). Esses dados sugerem uma forte vantagem adaptativa da proglotização. Além disso, tanto o aumento da complexidade dos ciclos de vida quanto o número de hospedeiros, considerando a simplicidade observada para a Ordem Caryophylliodea (Figura 2), pode nos levar a presumir que a metamerização e proglotização estão intimamente ligadas à evolução dos ciclos de vida dos vermes, envolvendo diferentes números de hospedeiros vertrebrados.

1.4. GENOMAS DE CESTÓDEOS E EVOLUÇÃO MOLECULAR

As relações filogenéticas considerando diferentes genes marcadores do desenvolvimento do Filo Platyhelminthes têm sido amplamente discutidas por décadas (Littlewood, 1999; Olson & Tkach, 2005; Thompson, 2008; Zarowiecki & Berriman, 2015). Por meio destas análises estabeleceram-se as principais relações do Filo Platyhelminthes, como a definição de quatro Classes e a monofilia das Classes platelmintos parasitas formando o Clado Neodermata. Porém, mantiveram-se dúvidas com relação às interrelações do Clado Neodermata, variando conforme o conjunto de marcadores utilizados para a análise filogenética (Littlewood et al., 2001).

Apenas nos últimos anos, os dados genômicos começaram a ser considerados para as análises evolutivas destes organismos. De fato, poucos dados genômicos de platelmintos estão atualmente disponíveis. No entanto, somente com a utilização de dados genômicos em larga escala, algumas questões referentes à biologia molecular destes parasitos passaram a ser elucidadas. Estudos nesse sentido incluem a análise genômica comparativa considerando quatro genomas de espécies pertencentes à Classe Cestoda que foca em adaptações ao parasitismo exclusivas dessa Classe (Tsai et al., 2013). Neste estudo, são descritas perdas de genes e vias metabólicas ubíquas em outros animais, e a associação dessa simplificação metabólica ao parasitismo, além da identificação de possíveis alvos para o desenvolvimento de fármacos anti-helmínticos.

Em sequência, um estudo considerando todas as Classes do Filo Platyhelminthes, incluindo três genomas de cestódeos, três genomas de trematódeos e um genoma de cada uma das outras Classes (dados genômicos não disponibilizados) esclareceu as relações evolutivas do Clado Neodermata através de uma análise filogenômica (Hahn et al., 2014). Esse trabalho descreve a Classe Monogenea como basal aos Trematóteos e Cestódeos, sendo, assim, o ectoparasitismo plesiomórfico dentro do Clado Neodermata. Além disso, a perda de vias de biossíntese de ácidos graxos funcionais e a ausência de peroxissomos foram sugeridas.

Adicionalmente, uma análise de genômica comparativa das Classes Cestoda (cinco genomas), Trematoda (quatro genomas) e Turbellaria (um genoma não publicado) descreve a perda dos sistemas de variação antigênica de superfície dos helmintos parasitas e o desenvolvimento de conjuntos de proteínas imunorregulatórias capazes de suprimir a resposta imunológica do hospedeiro durante os longos períodos de infecção (Zarowiecki & Berriman, 2015). Além disso, reforça-se e amplia-se a descrição de perdas de vias metabólicas em platelmintos endoparasitas.

Assim, muitas descrições do metabolismo e adaptações dos platelmintos ao parasitismo têm sido realizadas através da utilização de dados genômicos. Uma das conclusões a que chegaram indica que a regressão morfológica e a simplificação, em alguns aspectos, do metabolismo nas classes Trematoda e Cestoda, se deve, possivelmente, à redução dos genomas dessas espécies parasitas (Zarowiecki &

Berriman, 2015). A simplificação do metabolismo nestas espécies, porém, também foi descrita como uma adaptação importante para a evolução de organismos metaméricos e segmentados (Couso, 2009).

1.5. JUSTIFICATIVAS

Cestódeos que infectam o homem e animais domésticos são alvos de investigação científica em todo o mundo, com ênfase na busca de formas mais eficientes de prevenção, diagnóstico e tratamento das enfermidades causadas por suas formas larvais (Lorenzatto et al., 2015; Gabriël et al., 2016; Sharma et al., 2016). Porém, apesar dos resultados das pesquisas e da implementação de programas de controle epidemiológico em diversos países, os esforços visando à erradicação de cestodíases, como a cisticercose e a hidatidose cística, têm apresentado resultados bastante limitados (Coral-Almeida et al., 2015; Cucher et al., 2016). O insucesso de programas de prevenção, controle e erradicação de cestodíases deve-se em grande parte à escassez de conhecimentos sobre a biologia do desenvolvimento destes parasitos, sobre aspectos moleculares das interações parasito-hospedeiro e sobre a influência destes fatores sobre a proliferação e, por consequência, a capacidade reprodutiva e dinâmica de transmissão do parasito entre seus hospedeiros. Nesse contexto, este trabalho visa investigar as relações evolutivas entre organismos do Filo Platyhelminthes e identificar diferenças entre os genomas de platelmintos endoparasitas que apresentam ou não o processo de proglotização, gerando uma estrutura corporal intimamente ligada ao aumento da capacidade reprodutiva. Do ponto de vista de pesquisa básica, este estudo se propõe a identificar genes relacionados a esse processo de desenvolvimento, verificando se há eventos de seleção positiva atuando sobre estes genes.

Do ponto de vista de potenciais aplicações, os resultados a serem gerados disponibilizarão novos genes-alvo para estudos funcionais, na tentativa de melhor elucidar a biologia do desenvolvimento destes parasitos, que poderão ser utilizados para o desenvolvimento de drogas anti-helmínticas mais eficientes.

2. OBJETIVOS

2.1. OBJETIVO GERAL

O objetivo geral deste trabalho foi realizar a descrição das relações evolutivas entre organismos do Filo Platyhelminthes e a identificação de genes associados ao processo de estrobilização de platelmintos da Classe Cestoda.

2.2. OBJETIVOS ESPECÍFICOS

- Realizar uma análise evolutiva utilizando dados genômicos (filogenômica) para o estabelecimento das relações filogenéticas no Filo Platyhelminthes.

- Identificar genes associados ao processo de proglotização, através da comparação de dados genômicos, enriquecimento funcional e dados de transcrição.

- Avaliar os processos evolutivos atuantes nas proteínas relacionadas à proglotização.

3. CAPÍTULO I – *PHYLOGENOMIC ANALYSIS OF FLATWORM ENDOPARASITES AND SEARCH FOR DEVELOPMENT-RELATED AND EVOLUTIONARILY CONSERVED PROTEINS IN CESTODES*

3.1. APRESENTAÇÃO

O manuscrito que constitui esta seção foi elaborado conforme o formato exigido para submissão à revista *Development Genes and Evolution* (http://link.springer.com/journal/427). Todos os experimentos descritos no manuscrito, assim como a sua redação, foram realizados pela aluna Gabriela Prado Paludo, sendo os demais autores responsáveis pela sua orientação. Os *scripts* utilizados neste trabalho e todo o material suplementar (*Supplementary Files*) associado a ele estão disponíveis nos Apêndices de 1 a 22.

PHYLOGENOMIC ANALYSIS OF FLATWORM ENDOPARASITES AND SEARCH FOR DEVELOPMENT-RELATED AND EVOLUTIONARILY CONSERVED PROTEINS IN CESTODES

Gabriela Prado Paludo [1,2,3], Claudia Elizabeth Thompson [2,3], and Henrique Bunselmeyer Ferreira [1,3,4]

[1] Laboratório de Genômica Estrutural e Funcional, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS – Brazil.
[2] Unidade de biologia Teórica e Computacional, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS – Brazil.
[3] Programa de Pós-Graduação em Biologia Celular e Molecular, Centro de Biotecnologia, UFRGS, Porto Alegre, RS – Brazil.
[4] Departamento de Biologia Molecular e Biotecnologia, Instituto de Biociências, UFRGS, Porto Alegre, RS – Brazil.

*** To whom correspondence should be sent:**
Dr. Henrique Bunselmeyer Ferreira
Laboratório de Genômica Estrutural e Funcional
Centro de Biotecnologia
Universidade Federal do Rio Grande do Sul
Av. Bento Gonçalves, 9500 - Prédio 43-421, Sala 210
Caixa Postal 15005
91501-970 Porto Alegre, RS
BRAZIL

Phone: (+55 51) 3308-7768

Contract/grant sponsor: CAPES.

ABSTRACT

The Phylum Platyhelminthes includes all flatworms and comprehends four Classes: Turbellaria, Monogenea, Trematoda, and Cestoda. Among flatworms, monogeneans, trematodes and cestodes are exclusively parasites, while most turbelarians are free-living organisms. Some interesting aspects are evident in the evolution of parasitic plathyhelminths, as the increase of their progeny number through an enormous reproductive capacity. The Eucestoda Subclass has increased fecundity through serial repetition of their reproductive organs (proglottisation). However, the development mechanism leading to this body organization is still unknown. The main objective of this work was to understand the evolutionary relationships among segmented and non-segmented species from the Phylum Platyhelminthes and identify proteins related to the proglottisation process. The 10 sequenced and annotated genomes from parasitic platyhelminth species available in public databanks were included in this study, being 5 of them from segmented species and 5 from non-segmented ones. A phylogenomic analysis was performed in order to establish their evolutionary relationships, also including genomes from 6 nematodes (non-segmented helminths), one annelid (segmented deuterostome), and one mollusk (non-segmented deuterostome) as outgroups. Comparative genomics associated with expression data were used to select 12 developmental proteins conserved in proglottised species. The rates of synonymous and nonsynonymous substitutions were used to investigate the molecular evolution of each protein in lophotrocozoans. Thus, this work presents a study of the evolutionary relationships among species of flatworms and highlights a set of evolutionary conserved proteins of cestodes as possible regulators of this adaptive morphologic process, describing a set of targets for further researches.

Key words: Cestode development, Developmental proteins, Proglottisation, Segmentation, Phylogenomics

INTRODUCTION

The Phylum Platyhelminthes (flatworms) is comprised of an enourmous diversity of species, most of them parasites (Scholz et al. 2009). This Phylum comprehends four Classes, namely Turbellaria (planarians), Trematoda (the flukes), Monogea, and Cestoda (tapeworm).

Parasitic flatworms form a monophyletic group known as Neodermata, including the tapeworms, flukes and Monogea, which share a common ancestor (Lockyer et al. 2003). Neodermata clade constitutes one of the three largest groups of metazoan parasites of vertebrates (the others being the nematodes and arthropods) and includes many species of medical and veterinary importance (Koziol et al. 2016). Tapeworms are obligate internal parasites of vertebrates that display a wide range of body forms, life histories, and host associations (Olson et al. 2001). Some species (e.g., those from genus *Echinococcus*, *Taenia* and *Diphyllobothrium*) are etiological agents of major diseases in human beings and domesticated animals and cause morbidity and mortality in humans and domestic livestock, with significant economic and public health impacts (Gabriël et al. 2016; Kinkar et al. 2016). These parasites are now receiving considerable attention from biologists in a variety of fields, from molecular aspects of host-parasite association to epidemiology and distribution (Scholz et al. 2009; Lorenzatto et al. 2012; Hahn et al. 2014; Coral-Almeida et al. 2015).

Among the adaptations to parasitism, each Class possesses its own fitness strategy. The tapeworms' adaptations to their complex life cycles involve the increase of their progeny number through an enormous reproductive capacity.

The Eucestoda Subclass has increased fecundity through serial repetition of their reproductive organs (proglottisation) (Olson et al. 2001). The proglottisation is a kind of segmentation that leads to excision of zooids and its high number of repetitions is one way to promote cross-fertilization, further increasing the adaptive success of this body structure (Olson et al. 2001; Couso 2009). However, the development mechanism leading to this body organization is still unknown.

Here, we have investigated the evolutionary relationships among species of flatworms and identified genes potential related to the proglottisation development. Furthermore, domain and molecular evolution analysis of target proteins link them as possible regulators of this adaptive morphologic process.

RESULTS

Putative proglottisation-related proteins identification

Considering all sequenced and annotated genomes available in the databanks, five species belonging to the flukes (not proglottised neodermatan) and five species of tapeworms (proglottised neodermatan) were included in this study. Additionally, genomes of six nematodes (not segmented helminths), one annelid (segmented deuterostome), and one mollusk (not segmented deuterostome) were included as outgroups. The search for orthologous shared by these organisms generates 11,300 orthologous groups.

In order to find proteins possibly related to the proglottisation process, orthologous sequences were grouped according to the representation of flukes or by the representation of tapeworms, see Fig 1 A-B. Thus, the number of orthologous groups represented by all flukes was 2,809 and by all tapeworms was 3,365. Whereas essential proteins for proglottisation process have orthologues in all proglottised organisms, but may lack in not proglottised, orthologous groups were selected to be present in all tapeworms and absent in at last one fluke, resulting in 910 tapeworms conserved orthologous groups (Fig 1C).

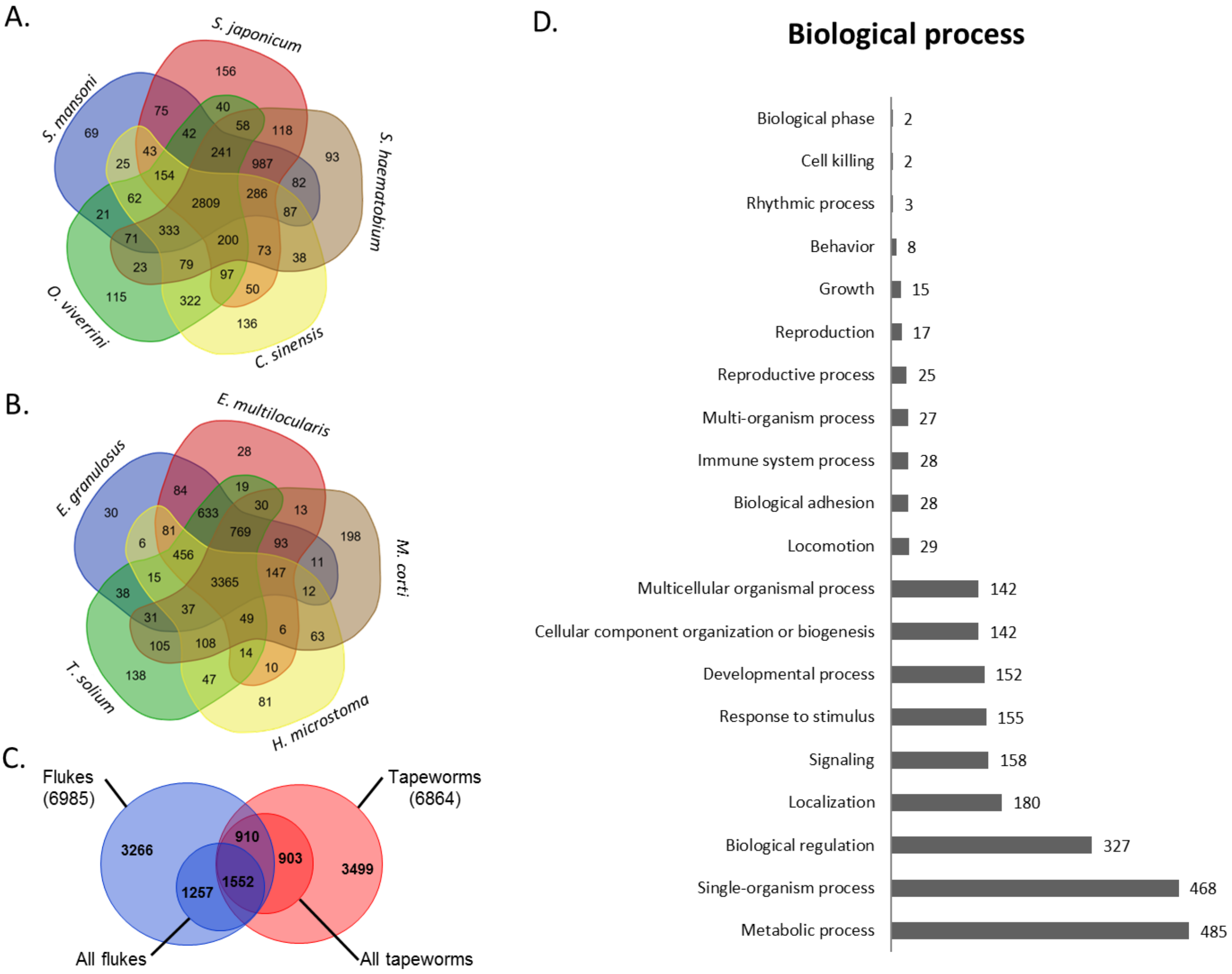


Fig 1. Venn diagrams of flatworm orthologous and functional enrichment. (A) Venn diagram showing orthologous groups shared among the five fluke species: *Clonorchis sinensis*, *Opisthorchis viverrini*, *Schistosoma haemmatobium*, *Schistosoma japonicum*, and *Schistosoma mansoni.* (B) Venn diagram showing orthologous groups shared among the five tapeworm species: *Echinococcus granulosus*, *Echinococcus multilocularis*, *Hymenolepis microstoma*, *Mesocestoides corti*, and *Taenia solium*. (C) Venn diagram showing orthologous groups shared between the sets of proteins from flukes and tapeworms, including their subsets of proteins present in all species of each Class. (D) Biological processes performed by the 910 proteins present in all tapeworms and absent in at least one fluke.

As the proglottisation is a developmental process, we performed a functional enrichment of the tapeworms conserved orthologous groups. Among biological processes mediated by these orthologous (Fig 1D) were selected 152 orthologous groups related to the developmental process. Their molecular functions and the cellular components are showed in Supplementary File 1. Furthermore, considering that the proglottisation is a process that occurs only in the adult stage of tapeworms life, we select only proteins up or down regulated in adult in relation to the larval stage of tapeworms (Table 1), resulting in 12 selected proteins.

Table 1. Putative proglottisation-related proteins. The orthologous presence in each species is highlighted (gray). Protein regulation analysis in larva X adult stages is represented by: UP for up-regulated protein, DOWN for down-regulated protein or ND for non-difference of regulation. Orthologous without expression analysis are represented by 'x'.

	Platyhelminthes										Annelida	Mollusca	Nematoda					
	Trematoda					Cestoda												
	C. sinensis	*O. viverrini*	*S. haematobium*[1]	*S. japonicum*	*S. mansoni*[2]	*E. granulosus*	*E. multilocularis*[3]	*H. microstoma*[4]	*M. corti*[5]	*T. solium*	*H. robusta*	*L. gigantea*	*C. elegans*	*G. pallida*	*H. contortus*	*O. volvulus*	*S. ratti*	*T. muris*
Bone morphogenetic protein 2						x	**DOWN**	**ND**	**ND**	x								
Cyclin-g-associated kinase		x	x	x	**ND**	x	**UP**	**ND**	**ND**	x	x							
Groucho protein	x		**ND**	x		x	**UP**	**ND**	**ND**	x	x	x	x			x		x
Homeobox protein Hox B4a						x	**DOWN**	**ND**	**ND**	x								
Lim homeobox protein lhx1						x	**UP**	**ND**	**ND**	x								
Membrane-associated guanylate kinase protein 2						x	**UP**	**ND**	**ND**	x								
Serine:threonine protein kinase mark2						x	**UP**	**ND**	**ND**	x	x	x	x			x	x	x
Atrial natriuretic peptide receptor 1			x		**ND**	x	**DOWN**	**ND**	**ND**	x		x						
RNA binding motif single stranded interacting						x	**UP**	**ND**	**ND**	x	x							
Serine:threonine protein kinase						x	**ND**	**UP**	**ND**	x	x	x						
Mothers against decapentaplegic homolog 4-like			**ND**			x	**UP**	**ND**	**ND**	x	x	x		x		x	x	x
Pangolin j						x	**DOWN**	**ND**	**ND**	x	x							

[1] *S. haematobium* expressed sequence tag libraries, ftp://ftp.sanger.ac.uk/pub4/pathogens/Schistosoma/mansoni;
[2] *S. mansoni* RNA-seq data from ArrayExpress under accession number E-MTAB-451;
[3] *E. multilocularis* RNA-seq data from ArrayExpress under accession number E-ERAD-50;
[4] *H.microstoma* RNA-seq data from ArrayExpress under accession number E-ERAD-56.
[5] *M. corti* RNA-seq data (Basika et al. unpublished data)

To evaluate the orthology of the selected groups, a domain analysis was performed (Fig 2). All the proteins in each orthologous group showed the same domains' profile. The BMP-2 (bone morphogenetic protein 2) proteins have the transforming growth factor-beta C-terminal domain (IPR001839); the GAK (cyclin-g-associated kinase) proteins have the ser/thr protein kinase (IPR002290), C2 domain (IPR000008), tensin phosphatase (IPR029023), and DnaJ (IPR001623) domains; the groucho proteins have the groucho/TLE N-terminal Q-rich (IPR005617), WD40-repeat-containing (IPR017986), and WD40 repeat (IPR001680) domains; Hox B4a (homeobox protein Hox B4a) proteins have the homeobox (IPR020479) protozoans domain; LHX1 (lim homeobox protein lhx1) proteins have the LIM-type zinc finger (IPR001781) and homeobox (IPR001356) domains; MAGI2 (membrane-associated guanilate kinase 2) proteins have the PDZ (IPR001478) domain; Mark2 proteins have the ser/thr protein kinase (IPR002290), ubiquitin-associated (IPR015940), and C-terminal KA1/Ssp2

(IPR028375) domains; NPR1 (atrial natriuretic peptide receptor 1) proteins have the ser/thr protein kinase (IPR001245), Haem NO binding associated (IPR011645), and adenylyl cyclase class-3/4/guanylyl cyclase (IPR001054) domains; RBMS (RNA binding motif single stranded interacting) proteins have the RNA recognition motif (IPR000504) domain; Ser:Thr protein kinase (serine:threonine protein kinase) proteins have the catalytic ser/thr/dual specificity protein kinase (IPR002290) and ubiquitin-associated (IPR015940) domains; SMAD4 (mothers against decapentaplegic homolog 4 like) proteins have the Dw arfin-type MAD homology (IPR003619) and SMAD/FHA (IPR008984) domains; and TCF/LCF (pangolin J) proteins have the high mobility group box (IPR009071) domain.

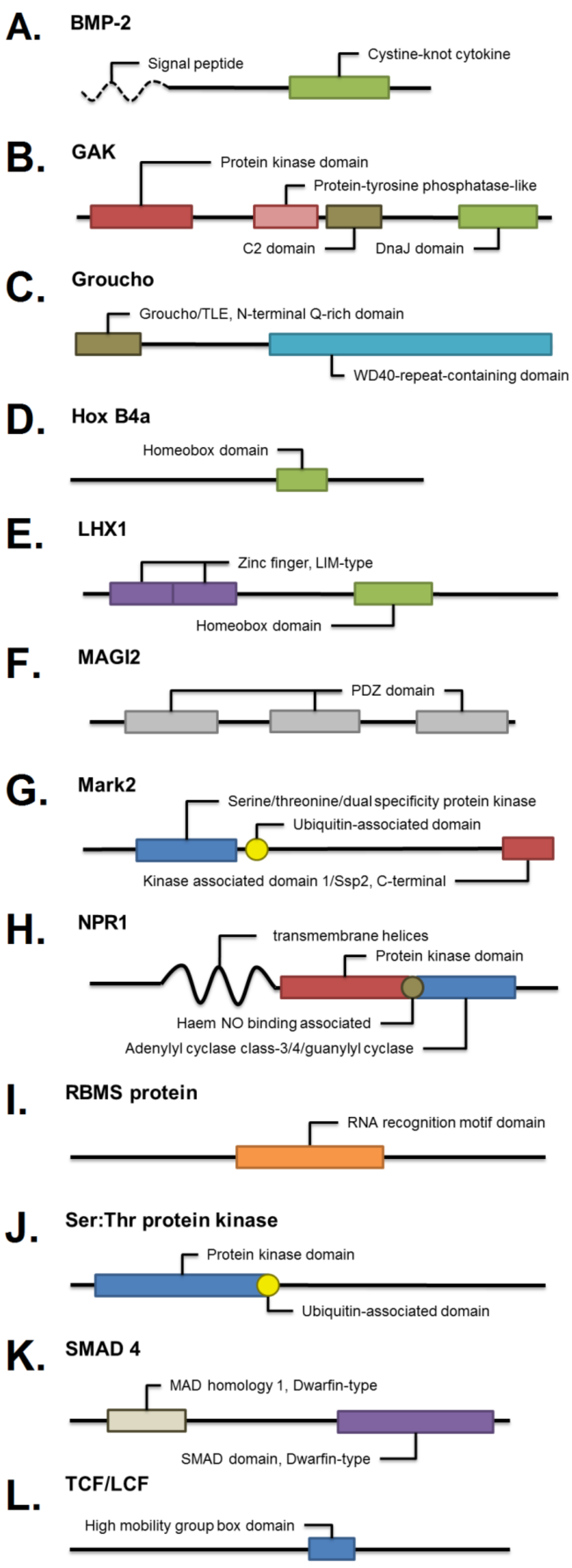


Fig 2. Domain profiles of putative proglottisation-related proteins. Representation of domains shared by all tapeworms orthologous of (A) bone morphogenetic protein 2, (B) cyclin-g-associated kinase, (C) groucho protein, (D) homeobox protein Hox B4a, (E) lim homeobox protein lhx1, (F) membrane-associated guanilate kinase, (G) Mark2 protein, (H) atrial natriuretic peptide receptor 1, (I) RNA binding motif single stranded interacting protein, (J) serine:threonine protein kinase, (K) mothers against decapentaplegic homolog 4 like, and (L) pangolin J protein.

Phylogenomic and phylogenetic analyses

Using the 18 selected genomes (protostome) of this study, we investigated the evolutionary relationships among species of flatworms through phylogenomic analysis. The orthology search for the protostome data set identified 11,300 orthologous groups, out of which 285 passed the selection criteria (see Materials and Methods section). The individual alignments for each selected gene were concatenated in a supermatrix for the subsequent phylogenomic analysis. Within the flatworms, two monophyletic groups of the endoparasitic flukes and tapeworms were highly supported in the analysis (Fig 3). With respect to protostome relationships, the phylogenomic tree obtained is in agreement with previously published results and recovers the monophyly of Protostome, Lophotrochozoa, Platyhelminthes, Cestoda and Trematoda with high statistical support (Bernt et al. 2013; Hahn et al. 2014).

The phylogenetic analysis of the orthologous groups of the putative proglottisation-related proteins was performed in order to identify the evolutive history of each protein (Supplementary files 2-13). In all analyzes, the cestodes are grouped into a monophyletic branch. As observed in the phylogenomic analysis, the species from *Echinococccus* genus form a monophyletic group and are most closely related to *Taenia solium* in all proteins analized, with the exception of SMAD 4 where the branches of these three species are low supported. For the other two tapeworms species, *Hymenolepis microstoma* and *Mesocestoides corti*, was observed a variation of their positions in relation to the species already mentioned in the pyhlogenetic trees. The *H. microstoma* is closer to *Echinococcus* sp. and *T. solium* in Groucho, Hox B4a, MAGI2, Mark2, RBMS protein and TCF/LCF analyses, and the *M. corti* is the closest one in BMP-2, GAK, LHX1, NPR1 and Ser:Thr protein kinase analyses.

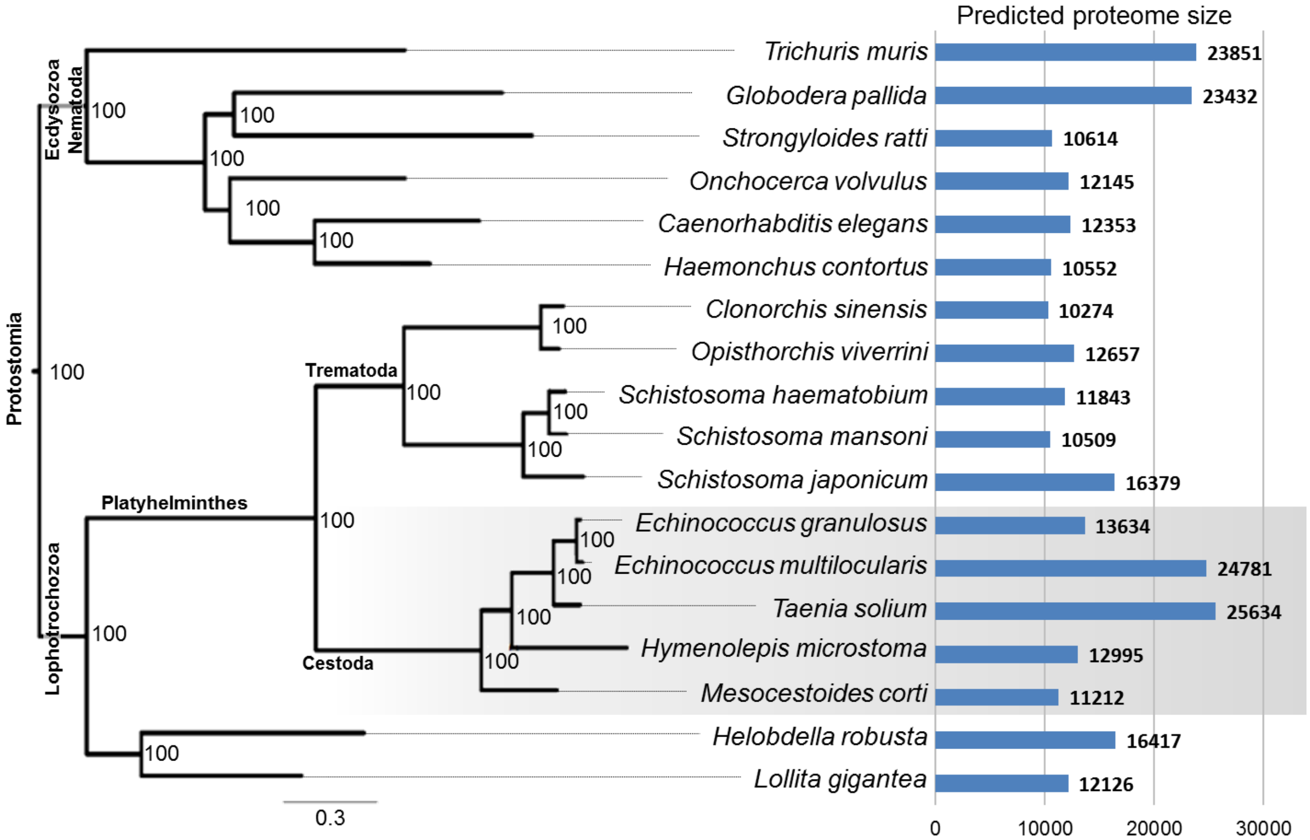


Fig 3. Platyhelminthes evolutionary relationships.
The phylogenomic tree (left) was built by MrBayes software with VT+I+G evolutive model for 1,688,000 generations with a set of 285 orthologous shared by all species. The numbers at the branches stand for Bayesian posterior probability values. The total numbers of predicted proteins for each species genome are showed (right) and the tapeworms data are highlighted by grey.

Analysis of positive selection in proglottisation-related genes

Through the analysis of the rates of nonsynonymous versus synonymous substitutions, we were able to identify if positive selection was acting on the proglottisation-related genes. When submitted to positive selection, there is an increase in the amino acid variability that provides adaptative advantage. Thereby, we used the CODEML package of PAML to detect positive selection acting on the proglottisation-related proteins previously identified. All codon sequences were aligned and for each data set was selected the best phylogenetic tree previously estimated. Thus, the results revealed that none of the proteins is under positive selection (Supplementary file 14).

It has been described that the presence of signatures of positive selection in evolutionarily new proteins may be responsible for the phenotypic diversity of specific developmental processes, such as brain development, sexual development and the tooth development of mammals (Zhang et al. 2011; Bohne et al. 2013; Machado et al. 2016). In contrast, proteins related to constitutional processes, as the proglottisation for these species of tapeworm, tend to have less positive selection that other proteins (Dall'Olio et al. 2012). Our results showed that these proteins are not suffering pressure that favors higher variation in its sequence in the domains regions.

DISCUSSION

Tapeworms are obligatory parasitic flatworms and, therefore, present a wide range of morphological and functional adaptations to their life style. A strategy to improve their fitness is the repetition of a multi-segmented body resulting in a huge capacity of reproduction. To better understand the developmental process that lead these organisms to segment their bodies in proglottides, we conducted comprehensive evolutionary and comparative analyses of organisms with proglottisation and others without this kind of segmentation.

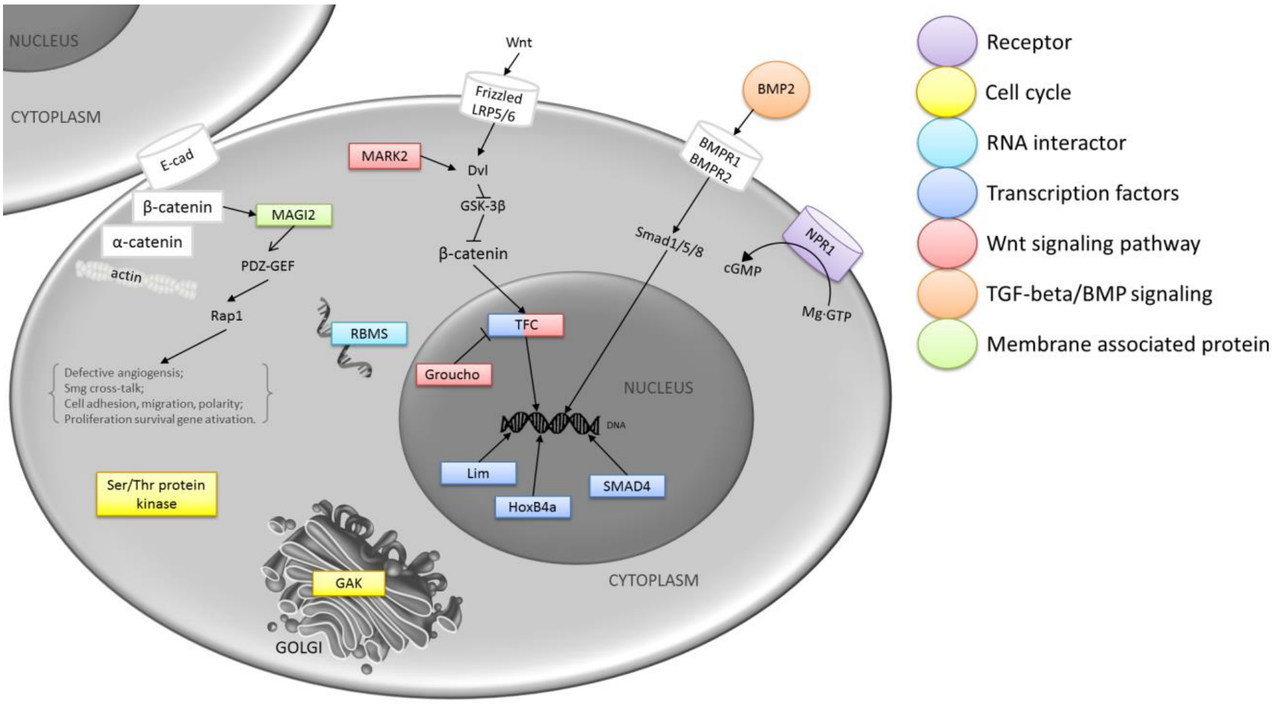


Fig 4. Simplified metabolic scheme of predicted pathways performed by the putative proglottisation-related proteins. Proteins functions/metabolic pathways are showed in colors, white boxes represent physical interaction of proteins.

In this work, we have performed the most extensive phylogenomic analysis of the Neodermata clade up to date, when considering the number of endoparasitic species included (Hahn et al. 2014; Egger et al. 2015). Evolutionary analysis (Fig 3) indicated that Cestoda and Trematoda Classes are sister groups.

Additionally, there was a separation between flatworms and the other Lophotrochozoa species, including the annelid *Helobdella robusta*, which shows external kind of segmentation. Thus, phylogenomic results, in association with the phylogenetic analysis of proglottisation-related proteins retake the idea that the proglottisation and external segmentation were independent evolutionary events (Olson et al. 2001).

Through functional analysis of the putative proglottisattion-related proteins, we could establish a link among them and their metabolic pathways (Fig 4). Among the identified metabolic pathways/functions, we mentioned some of the main pathways of developmental biology studies.

Wnt signaling pathway

Wnt pathway ligands are secreted glycoproteins containing a conserved sequence of cysteine residues. Wnt signalling is involved in a diverse range of cellular interactions throughout development, including regeneration (Broun 2005; Bastakoty and Young 2016), embryo segmentation (Dunty et al. 2007; Bolognesi et al. 2008), and axial patterning (Lin and Pearson 2014; Wei et al. 2016).

The discovery that canonical Wnt/β-catenin signalling is responsible for regulating head/tail specification in planarian regeneration highlighted their importance in flatworm (Phylum Platyhelminthes) development (Lin and Pearson 2014). A recent study showed that, although flatworms have a highly reduced and dispersed complement that includes orthologous of only five subfamilies (Wnt1, Wnt2, Wnt4, Wnt5 and Wnt11) and fewer paralogs in parasitic flatworms (5–6) than in planarians (9), all major signalling components are identified, including antagonists and receptors, and key binding domains are intact, indicating that the canonical (Wnt/β-catenin) and non-canonical (planar cell polarity and Wnt/Ca2+) pathways are functional (Riddiford and Olson 2011).

In fact, it was demonstrated posterior expression of specific Wnt factors during larval metamorphosis and showed that scolex formation is preceded by localized expression of Wnt inhibitors (Koziol et al. 2016). In this way, the identification of 3 signalling componentes (Groucho, Mark2 and PangolinJ) in this work suggests that the Wnt signaling is regulating the cestodes proglottisation and, therefore, is active during adult metamorphosis.

Transforming growth factor-β / bone morphogenetic protein signaling

The transforming growth factor-β (TGF-β) ligands are composed of a carboxy-terminal signaling domain and an amino-terminal propeptide domain that is cleaved before ligand release (Constam 2014). Two major clades of ligands are generally recognized: the TGF-β sensu stricto/TGF-β related (e.g., Activins, Leftys, and GDF8s) and bone morphogenetic protein (BMP) related (e.g., BMPs and Nodals) (Matus et al. 2008).

The TGF-β family of polypeptide growth factors regulates a wide variety of biological processes such as cell division, differentiation, adhesion, migration, and apoptosis in metazoan organisms (Zavala-Góngora et al. 2006). Signaling is initiated by binding of the cytokines to cell surface associated TGF-β receptors, which consist of two transmembrane serine/threonine kinases called the type I and the type II receptor (Richards and Degnan 2009). Once complexed with its ligand, the type II receptor phosphorylates and activates the type I receptor at the GS domain, which is located in the type I receptor's intracellular region. The type I receptors activated recruit and phosphorylate the receptor-regulated Smads (R-Smads; Smad1/5, Smad2/3) that form multisubunit complexes with common partner Smads (Co-Smads; Smad4) before entering the nucleus to regulate gene activity.

Smad family proteins are central components of TGF-β/BMP signaling pathways in metazoans, and regulate key developmental processes, such as body axis formation or regeneration (Epping and Brehm 2011). In this way, studies with the Smad4 from *E. granulosus* showed that the protein is expressed in the larval stages and exhibited the highest transcript levels in activated protoscoleces (pre-adult). The Smad4 and some receptor-regulated Smads proteins were co-localized in the sub-tegumental and tegumental layer of the parasite, suggesting that Smad4 may take part in critical biological processes, including echinococcal growth, development, and parasite-host interaction (Zhang et al. 2014).

Transcription factors

The LIM domain is a cysteine-histidine rich, zinc-coordinating domain, consisting of two tandemly repeat zinc fingers. The LIM homeodomain genes present two tandemly repeat LIM domain fused to a conserved homeodomain, as the LHX1 (Bach 2000). Considering its importance in developmental pathways, it was demonstrated that the LHX1 expression is dependent on the presence of Smad4 in the mouse epiblast and marks the entire definitive endoderm lineage, the anterior mesendoderm, and midline progenitors (Costello et al. 2015). Furthermore, the same work uses transcriptional profiling and ChIP-seq (chromatin immunoprecipitation followed by high-throughput sequencing) experiments to identify Lhx1 target genes, including numerous anterior definitive endoderm markers and components of the Wnt signaling pathway.

Homeobox genes are high-level transcription factors implicated in the patterning of body plans in animals. Across parasitic flatworms, the number of homeobox genes is extensively reduced and most of their functions are still unknown. Thus, the LHX1 identification as a putative proglottisation-related protein is important information about the Homeobox Transcription Factors acting on parasitic flatworms.

MATERIALS AND METHODS

Orthologous groups identification

Considering all the sequenced and annotated genomes available in the databanks, the endoparasitic flatworms were represented by 10 species, five genomes from Cestoda Class: *Echinococcus granulosus* (Tsai et al. 2013), *Echinococcus multilocularis* (Tsai et al. 2013), *Hymenolepis microstoma* (Tsai et al. 2013), *Mesocestoides corti*, and *Taenia solium* (Tsai et al. 2013); and five genomes from Trematoda Class: *Clonorchis sinensis* (Wang et al. 2011), *Schistosoma h*aematobium (Young et al. 2012), *Schistosoma japonicum* (Zhou et al. 2009), *Schistosoma mansoni* (Protasio et al. 2012), and *Opisthorchis viverrini* (Young et al. 2014). Additionally, the genomes of six nematodes were included as outgroups: *Caenorhabditis elegans* (C. elegans Sequencing Consortium 1998), *Globodera pallida* (Cotton et al. 2014), *Haemonchus contortus*, *Onchocerca volvulus*, *Strongyloides ratti* (Hunt et al. 2016), and *Trichuris muris* (Hunt et al. 2016); one annelid: *Helobdella robusta* (Simakov et al. 2012); and one mollusk: *Lollita gigantean* (Simakov et al. 2012). Detailed information about these genomes is described in Supplementary File 15. The OrthoMCL v2.0.8 (Li et al. 2003) was used with the default parameters to identify the orthologs and paralogs among the complete proteomes of all 18 studied organisms.

Search for target proteins

The first search step was performed using Python scripts, in which orthologous were grouped according to organisms to which they belong. Firstly, the selected orthologous groups were functionally enriched and categorized based on the BLAST sequence homologies and gene ontology (GO) annotations using the Blast2GO software (Conesa and Götz 2008) (Supplementary File 16). The protein regulation in different life stages of the organisms was analyzed using available data for: *E. multilocularis* (E-ERAD-50 ArrayExpress accession number), *H. microstoma* (E-ERAD-56 ArrayExpress accession number), *M. corti* (Basika et al. unpublished data), *S. haematobium* (expressed sequence tag libraries, ftp://ftp.sanger.ac.uk/pub4/pathogens/Schistosoma/mansoni), and *S. mansoni* (E-MTAB-451 ArrayExpress accession number).

Phylogenomic analyses

A Python script was developed to select from OrthoMCL output only orthologous group sequences represented by all 18 organisms and, if necessary, select only the longest sequence for each organism. The multi-FASTA ortholog files of each protein sequence were used as input for the multiple alignments using CLUSTAL omega algorithm (Sievers and Higgins 2014) with the default parameters. Subsequently, the SCaFos software (Roure et al. 2007) was used to allow the gene concatenation of 285 alignment files. The selection of supermatrix best-fit model of protein evolution was performed by ProtTest 3 (Darriba et al. 2011). A bayesian tree was constructed using MrBayes v3.2.2 (Ronquist et al. 2012). MrBayes was run using VT+I+G evolutive model for 1,688,000 generations with two runs and four chains in parallel, sampled every 100 generations and with a burn-in of 25%.

Putative proglottisation-related protein analysis

To improve the number of orthologous sequences, we performed searches using blastp in the non-redundant database of NCBI-Genbank, and phmmer tool of HMMER in the UniProtKB database (Supplementary File 17). Only sequences with identity and coverage above 30% and 70%, respectively, were selected. For functional domain annotation of all orthologous proteins, we employed InterProScan 5 version 57.0 (Jones et al. 2014), which uses a consortium of eleven protein domain databases (PROSITE, HAMP, Pfam, PRINTS, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, CATH-Gene3D, and PANTHER). Only proteins containing the same functional domains profile were considered orthologous. The multiple alignments of proteins and CDSs were peformed by CLUSTAL Omega guided by external HMM (hidden Markov model), and two variants of PRANK (Löytynoja and Goldman 2010) based on an amino acid model (PRANK_{AA}) or an empirical codon model (PRANK_{C}). The nucleotide alignments were obtained using PAL2NAL (Suyama et al. 2006) tool. Finally, we performed manual edition and removal of low aligned regions (Supplementary File 17).

The best orthologous alignments for the proteins and nucleotides were subsequently submitted to the phylogenetic analysis (Supplementary File 18). The selection of best-fit model of protein and nucleotide evolution was performed by MEGA 7 (Kumar et al. 2016) software. The orthologous files were submitted to phylogenetic analysis using distance and probabilistic methods implemented by MEGA 7 and bayesian method implemented by MrBayes. In relation to the distance methods, the neighbor-joining with pairwise deletion of gaps were applied to the datasets. The p-distance and poisson models were used for the proteins sequences, and p-distance and Jukes-Cantor models for the nucleotides sequences. The probabilistic method was applied using maximum likelihood with pairwise deletion of gaps. The bootstrap test of phylogeny was performed using 2,000 repetitions for all analyses. Bayesian method was sampled every 100 generations, with two runs and four chains in parallel and a burn-in of 25%. The TreeView program (Page 2002) was used to visualize and edit the resulting phylogenies. Furthermore, to detect orthologous codons under selective pressure, the site-specific model analysis using nested models M0, M1a, M2a, M3, M7 and M8 was implemented in the codeml program in PAML software. For all models, a Bayes empirical Bayes (BEB) approach was employed to detect codons with a posterior probability >90% of being under selection (Murrell et al. 2012).

ACKNOWLEDGMENTS

The authors are thankful to Dr. Magdalena Zarowiecki and the Wellcome Trust Sanger Institute for providing the access to *M. corti* genome data. Access to high-performance computing facilities granted by Laboratório Nacional de Computação Científica (LNCC) is gratefully acknowledged. This work was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

REFERENCES

Bach I (2000) The LIM domain: Regulation by association. Mech Dev 91:5–17. doi: 10.1016/S0925-4773(99)00314-7

Bastakoty D, Young PP (2016) Wnt/ -catenin pathway in tissue injury: roles in pathology and therapeutic opportunities for regeneration. FASEB J. doi: 10.1096/fj.201600502R

Bernt M, Bleidorn C, Braband A, et al (2013) A comprehensive analysis of bilaterian mitochondrial genomes and phylogeny. Mol Phylogenet Evol 69:352–364. doi: 10.1016/j.ympev.2013.05.002

Bohne A, Heule C, Boileau N, Salzburger W (2013) Expression and Sequence Evolution of Aromatase cyp19a1 and Other Sexual Development Genes in East African Cichlid Fishes. Mol Biol Evol 30:2268–2285. doi: 10.1093/molbev/mst124

Bolognesi R, Farzana L, Fischer TD, Brown SJ (2008) Multiple Wnt Genes Are Required for Segmentation in the Short-Germ Embryo of Tribolium castaneum. Curr Biol 18:1624–1629. doi: 10.1016/j.cub.2008.09.057

Broun M (2005) Formation of the head organizer in hydra involves the canonical Wnt pathway. Development 132:2907–2916. doi: 10.1242/dev.01848

C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science (80-) 282:2012–2018. doi: 10.1126/science.282.5396.2012

Conesa A, Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:619832. doi: 10.1155/2008/619832

Constam DB (2014) Regulation of TGFβ and related signals by precursor processing. Semin Cell Dev Biol 32:85–97. doi: 10.1016/j.semcdb.2014.01.008

Coral-Almeida M, Gabriël S, Abatih EN, et al (2015) Taenia solium Human Cysticercosis: A Systematic Review of Sero-epidemiological Data from Endemic Zones around the World. PLoS Negl Trop Dis 9:e0003919. doi: 10.1371/journal.pntd.0003919

Costello I, Nowotschin S, Sun X, et al (2015) Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. Genes Dev 29:2108–2122. doi: 10.1101/gad.268979.115

Cotton JA, Lilley CJ, Jones LM, et al (2014) The genome and life-stage specific transcriptomes of Globodera pallida elucidate key aspects of plant parasitism by a cyst nematode. Genome Biol 15:R43. doi: 10.1186/gb-2014-15-3-r43

Couso JP (2009) Segmentation, metamerism and the Cambrian explosion. Int J Dev Biol 53:8–10. doi: 10.1387/ijdb.072425jc

Dall'Olio G, Laayouni H, Luisi P, et al (2012) Distribution of events of positive selection and population differentiation in a metabolic pathway: the case of asparagine N-glycosylation. BMC Evol Biol 12:98. doi: 10.1186/1471-2148-12-98

Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165. doi: 10.1093/bioinformatics/btr088

Dunty WC, Biris KK, Chalamalasetty RB, et al (2007) Wnt3a/ -catenin signaling controls posterior body development by coordinating mesoderm formation and segmentation. Development 135:85–94. doi: 10.1242/dev.009266

Egger B, Lapraz F, Tomiczek B, et al (2015) A Transcriptomic-Phylogenomic Analysis of the Evolutionary Relationships of Flatworms. Curr Biol 25:1347–1353. doi: 10.1016/j.cub.2015.03.034

Epping K, Brehm K (2011) Echinococcus multilocularis: Molecular characterization of EmSmadE, a novel BR-

Smad involved in TGF-β and BMP signaling. Exp Parasitol 129:85–94. doi: 10.1016/j.exppara.2011.07.013

Gabriël S, Dorny P, Mwape KE, et al (2016) Control of Taenia solium taeniasis/cysticercosis: The best way forward for sub-Saharan Africa? Acta Trop. doi: 10.1016/j.actatropica.2016.04.010

Hahn C, Fromm B, Bachmann L (2014) Comparative Genomics of Flatworms (Platyhelminthes) Reveals Shared Genomic Features of Ecto-and Endoparastic Neodermata. Genome Biol Evol 6:1105–1117. doi: 10.1093/gbe/evu078

Hunt VL, Tsai IJ, Coghlan A, et al (2016) The genomic basis of parasitism in the Strongyloides clade of nematodes. Nat Genet 48:299–307. doi: 10.1038/ng.3495

Jones P, Binns D, Chang H-Y, et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. doi: 10.1093/bioinformatics/btu031

Kinkar L, Laurimäe T, Simsek S, et al (2016) High-resolution phylogeography of zoonotic tapeworm Echinococcus granulosus sensu stricto genotype G1 with an emphasis on its distribution in Turkey, Italy and Spain. Parasitology 1–12. doi: 10.1017/S0031182016001530

Koziol U, Jarero F, Olson P, Brehm K (2016) Comparative analysis of Wnt expression identifies a highly conserved developmental transition in flatworms. BMC bilogy 14:10. doi: 10.1186/s12915-016-0233-x

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870–1874. doi: 10.1093/molbev/msw054

Li L, Stoeckert CJ, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13:2178–89. doi: 10.1101/gr.1224503

Lin AYT, Pearson BJ (2014) Planarian yorkie/YAP functions to integrate adult stem cell proliferation, organ homeostasis and maintenance of axial patterning. Development 141:1197–1208. doi: 10.1242/dev.101915

Lockyer AE, Olson PD, Littlewood DTJ (2003) 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 78:155–171. doi: 10.1046/j.1095-8312.2003.00141.x

Lorenzatto KR, Monteiro KM, Paredes R, et al (2012) Fructose-bisphosphate aldolase and enolase from Echinococcus granulosus: Genes, expression patterns and protein interactions of two potential moonlighting proteins. Gene. doi: 10.1016/j.gene.2012.06.046

Löytynoja A, Goldman N (2010) webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics 11:579. doi: 10.1186/1471-2105-11-579

Machado JP, Philip S, Maldonado E, et al (2016) Positive Selection Linked with Generation of Novel Mammalian Dentition Patterns. Genome Biol Evol 8:2748–2759. doi: 10.1093/gbe/evw200

Matus DQ, Magie CR, Pang K, et al (2008) The Hedgehog gene family of the cnidarian, Nematostella vectensis, and implications for understanding metazoan Hedgehog pathway evolution. Dev Biol 313:501–518. doi: 10.1016/j.ydbio.2007.09.032

Murrell B, Wertheim JO, Moola S, et al (2012) Detecting Individual Sites Subject to Episodic Diversifying Selection. PLoS Genet 8:e1002764. doi: 10.1371/journal.pgen.1002764

Olson PD, Timothy D, Littlewood J, et al (2001) Interrelationships and Evolution of the Tapeworms (Platyhelminthes: Cestoda). Mol Phylogenet Evol 19:443–467. doi: 10.1006/mpev.2001.0930

Page RDM (2002) Visualizing Phylogenetic Trees Using TreeView. In: Current Protocols in Bioinformatics. John Wiley & Sons, Inc., Hoboken, NJ, USA,

Protasio A V., Tsai IJ, Babbage A, et al (2012) A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke Schistosoma mansoni. PLoS Negl Trop Dis 6:e1455. doi: 10.1371/journal.pntd.0001455

Richards GS, Degnan BM (2009) The dawn of developmental signaling in the metazoa. Cold Spring Harb Symp Quant Biol 74:81–90. doi: 10.1101/sqb.2009.74.028

Riddiford N, Olson PD (2011) Wnt gene loss in flatworms. Dev Genes Evol 221:187–197. doi: 10.1007/s00427-011-0370-8

Ronquist F, Teslenko M, van der Mark P, et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–42. doi: 10.1093/sysbio/sys029

Roure B, Rodriguez-Ezpeleta N, Philippe H (2007) SCaFoS: a tool for Selection, Concatenation and Fusion of Sequences for phylogenomics. BMC Evol Biol 7:S2. doi: 10.1186/1471-2148-7-S1-S2

Scholz T, Garcia HH, Kuchta R, Wicht B (2009) Update on the Human Broad Tapeworm (Genus Diphyllobothrium), Including Clinical Relevance. Clin Microbiol Rev 22:146–160. doi: 10.1128/CMR.00033-08

Sievers F, Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. Methods Mol Biol 1079:105–16. doi: 10.1007/978-1-62703-646-7_6

Simakov O, Marletaz F, Cho S-J, et al (2012) Insights into bilaterian evolution from three spiralian genomes. Nature 493:526–531. doi: 10.1038/nature11696

Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612. doi: 10.1093/nar/gkl315

Tsai IJ, Zarowiecki M, Holroyd N, et al (2013) The genomes of four tapeworm species reveal adaptations to parasitism. Nature 496:57–63. doi: 10.1038/nature12031

Wang X, Chen W, Huang Y, et al (2011) The draft genome of the carcinogenic human liver fluke Clonorchis sinensis. Genome Biol 12:R107. doi: 10.1186/gb-2011-12-10-r107

Wei S, Shang H, Cao Y, Wang Q (2016) The coiled-coil domain containing protein Ccdc136b antagonizes maternal Wnt/β-catenin activity during zebrafish dorsoventral axial patterning. J Genet Genomics 43:431–438. doi: 10.1016/j.jgg.2016.05.003

Young ND, Jex AR, Li B, et al (2012) Whole-genome sequence of Schistosoma haematobium. Nat Genet 44:221–225. doi: 10.1038/ng.1065

Young ND, Nagarajan N, Lin SJ, et al (2014) The Opisthorchis viverrini genome provides insights into life in the bile duct. Nat Commun. doi: 10.1038/ncomms5378

Zavala-Góngora R, Kroner A, Bernthaler P, et al (2006) A member of the transforming growth factor-beta receptor family from Echinococcus multilocularis is activated by human bone morphogenetic protein 2. Mol Biochem Parasitol 146:265–71. doi: 10.1016/j.molbiopara.2005.12.011

Zhang C, Wang L, Wang H, et al (2014) Identification and characterization of functional Smad8 and Smad4 homologues from Echinococcus granulosus. Parasitol Res 113:3745–3757. doi: 10.1007/s00436-014-4040-4

Zhang YE, Landback P, Vibranovski MD, Long M (2011) Accelerated Recruitment of New Brain Development Genes into the Human Genome. PLoS Biol 9:e1001179. doi: 10.1371/journal.pbio.1001179

Zhou Y, Zheng H, Chen Y, et al (2009) The Schistosoma japonicum genome reveals features of host–parasite interplay. Nature 460:345–351. doi: 10.1038/nature08140

4. CAPÍTILO II – IDENTIFICAÇÃO DE PROTEÍNAS HIPOTÉTICAS POSSIVELMENTE RELACIONADAS AO PROCESSO DE PROGLOTIZAÇÃO

4.1. APRESENTAÇÃO

O Capítulo II tem como objetivo relacionar proteínas hipotéticas ao processo de proglotização, através da comparação de dados genômicos, enriquecimento funcional e dados de transcrição. O presente capítulo está estruturado em sessões de “Materiais e Métodos” e “Resultados”, e apresenta a identificação de 22 proteínas hipotéticas conservadas em cestódeos possivelmente relacionadas à proglotização. Os *scripts* utilizados neste trabalho estão disponíveis nos Apêndices 1 e 2.

4.2. MATERIAIS E MÉTODOS

4.2.1. Identificação dos grupos de proteínas ortólogas

Os genomas utilizados neste estudo estão descritos no Apêndice 18. A identificação dos grupos de ortólogos foi realizada através da utilização do software OrthoMCL v2.0.8, conforme descrito na sessão “Orthologous groups identification” dos “Materials and methods” do manuscrito apresentado no Capítulo I.

4.2.2. Associação de proteínas ao processo de proglotização

Com o intuito de relacionar proteínas evolutivamente conservadas em cestódeos ao processo de proglotização, foram utilizados *scripts* em Python (Apêndice 2) para selecionar proteínas ortólogas presentes em todas as espécies de cestódeos analisadas e ausentes em, pelo menos, uma das espécies de trematódeos, conforme descrito na sessão “Search for target proteins” dos “Materials and methods” do manuscrito apresentado no Capítulo I.

Em seguida, foram selecionadas as proteínas identificadas como hipotéticas na descrição de produtos gênicos diponibilizada para os genomas de *E. granulosus*, *E. multilocularis* e *H. microstoma.* Por fim, foram selecionadas apenas as proteínas com genes diferencialmente expressos entre os estágios larval e adulto de cestódeos, com base nos dados de transcrição dos genes correspondentes descritos na sessão “Search for target proteins” dos “Materials and methods” do manuscrito apresentado no Capítulo I.

4.2.3. Identificação de domínios funcionais

Para avaliar a ortologia das proteínas identificadas, realizou-se uma busca por domínios funcionais através da ferramenta InterProScan 5 versão 57.0, conforme descrito na sessão “Putative proglottisation-related protein analysis” dos “Materials and methods” do manuscrito apresentado no Capítulo I. Apenas proteínas com o mesmo perfil de domínios foram consideradas ortólogas.

4.2.4. Busca por proteínas ortólogas

A busca por proteínas ortólogas foi realizada conforme descrito na sessão “Putative proglottisation-related protein analysis” dos “Materials and methods” do manuscrito apresentado no Capítulo I.

4.3. RESULTADOS

4.3.1. Identificação de proteínas hipotéticas possivelmente relacionadas ao processo de proglotização

Considerando as espécies estudadas, a proglotização é um processo de devenvolvimento presente nas cinco espécies de cestódeos e ausente em todas as demais espécies. Dessa forma, utilizou-se uma comparação entre o repertório de proteínas presentes nos cinco proteomas preditos de cestódeos em relação aos cinco proteomas dos seus organismos mais próximos evolutivamente, os trematódeos (ver Fig 3 do Capítulo I). A análise foi iniciada com um grupo de 910 proteínas (ver Fig 1 do Capítulo I) que, nas espécies estudadas, possuem ortólogos em todos os cestódeos e que estão ausentes em, pelo menos, um trematódeo. Posteriormente, foram selecionadas apenas as proteínas anotadas como hipotéticas, definindo um total de 174 grupos de proteínas hipotéticas ortólogas.

Considerando que apenas cestódeos adultos podem ser proglotizados, foram selecionadas as proteínas que têm seus genes transcritos diferencialmente na comparação entre as fases larval e adulta de cestódeos. Com base neste critério, foram selecionadas 22 proteínas hipotéticas, descritas na Tabela 2, as quais serão identificadas por numeração sequencial, de 1 a 22. Considerando o conjunto amostral, as proteínas hipotéticas selecionadas não apresentam ortólogos para as espécies de nematódeos, estando estes restritos a animais do Filo Platyhelminthes, com excessão da proteína 15 que apresenta ortólogos em lofotrocozoários. Adicionalmente, entre os dados de transcrição analisados, apenas a proteína 18 é diferencialmente expressa em uma espécie de trematódeo (*S. haematobium*), porém, essa proteína foi mantida por seu transcrito estar com expressão diminuída em trematódeos adultos enquanto os transcritos de seus ortólogos possuem expressão aumentada em cestódeos adultos.

Tabela 2. Proteínas hipotéticas possivelmente relacionadas ao processo de proglotização. A presença de ortólogo em cada espécie está destacada em cinza. Resultados de expressão gênica comparativa dos estágios Larval X Adulto estão representados pelos símbolos: seta para cima para expressão aumentada, seta para baixo para expressão diminuída e círculo preenchido para quando não há diferença significativa da expressão. Ortólogos para os quais não foi analisada a expressão gênica, estão marcadas por 'x'.

	Platyhelminthes											
	Trematoda					Cestoda					Annelida	Mollusca
Proteína hipotética	*C. sinensis*	*O. viverrini*	*S. haematobium*[1]	*S. japonicum*	*S. mansoni*[2]	*E. granulosus*	*E. multilocularis*[3]	*H. microstoma*[4]	*M. corti*[5]	*T. solium*	*H. robusta*	*L. gigantea*
1		x				x	●	↓	●	x		
2						x	●	●	↓	x		
3						x	●	●	↑	x		
4						x	↑	●	●	x		
5						x	↓	●	●	x		
6						x	↓	●	↓	x		
7						x	↓	●	●	x		
8						x	↓	●	●	x		
9						x	●	●	↓	x		
10						x	↑	●	●	x		
11						x	↑	●	●	x		
12						x	●	●	↑	x		
13						x	●	●	↑	x		
14						x	↓	↓	●	x		
15	x	x			●	x	●	↑	●	x	x	x
16			●			x	●	↓	●	x		
17						x	↑	●	●	x		
18	x	x	↓		●	x	●	●	↑	x		
19	x	x				x	↓	●	●	x		
20	x				●	x	●	●	↑	x		
21		x				x	↓	●	●	x		
22						x	↑	●	●	x		

[1] *S. haematobium*, ftp://ftp.sanger.ac.uk/pub4/pathogens/Schistosoma/mansoni;

[2] *S. mansoni*, número de acesso E-MTAB-451 do ArrayExpress;

[3] *E. multilocularis*, número de acesso E-ERAD-50 do ArrayExpress;

[4] *H. microstoma,* número de acesso E-ERAD-56 do ArrayExpress;

[5] *M. corti,* Basika et al. dados não publicados.

A ortologia das proteínas identificadas foi avaliada através da comparação entre seus perfis de domínios (Figura 4). Das 22 proteínas, 13 não apresentam resultado algum na análise de domínios, 6 apresentaram apenas resultados de domínios transmembranas, duas apresentaram domínios transmembranas e a sua associação com algum domínio: proteína 3 apresenta o domínio “family A G protein-coupled receptor-like superfamily” (SSF81321) e a proteína 15 apresenta um domínio não carcterizado (PTHR12242); para a proteína 1 foram identificados dois domínios “calcium-dependent phosphotriesterase” (SSF63829). Como esperado, pouca informação é obtida através da análise de domínios das proteínas hipotéticas e, através destes resultados, não foi possível realizar inferência funcional para nenhuma das proteínas. Porém, os domínios identificados estão presentes em todas as proteínas de cada grupo, de forma que todas as ortólogas apresentam o mesmo perfil de domínios. Assim, essa análise valida a identificação dos grupos de ortólogas realizada com base em sua sequência.

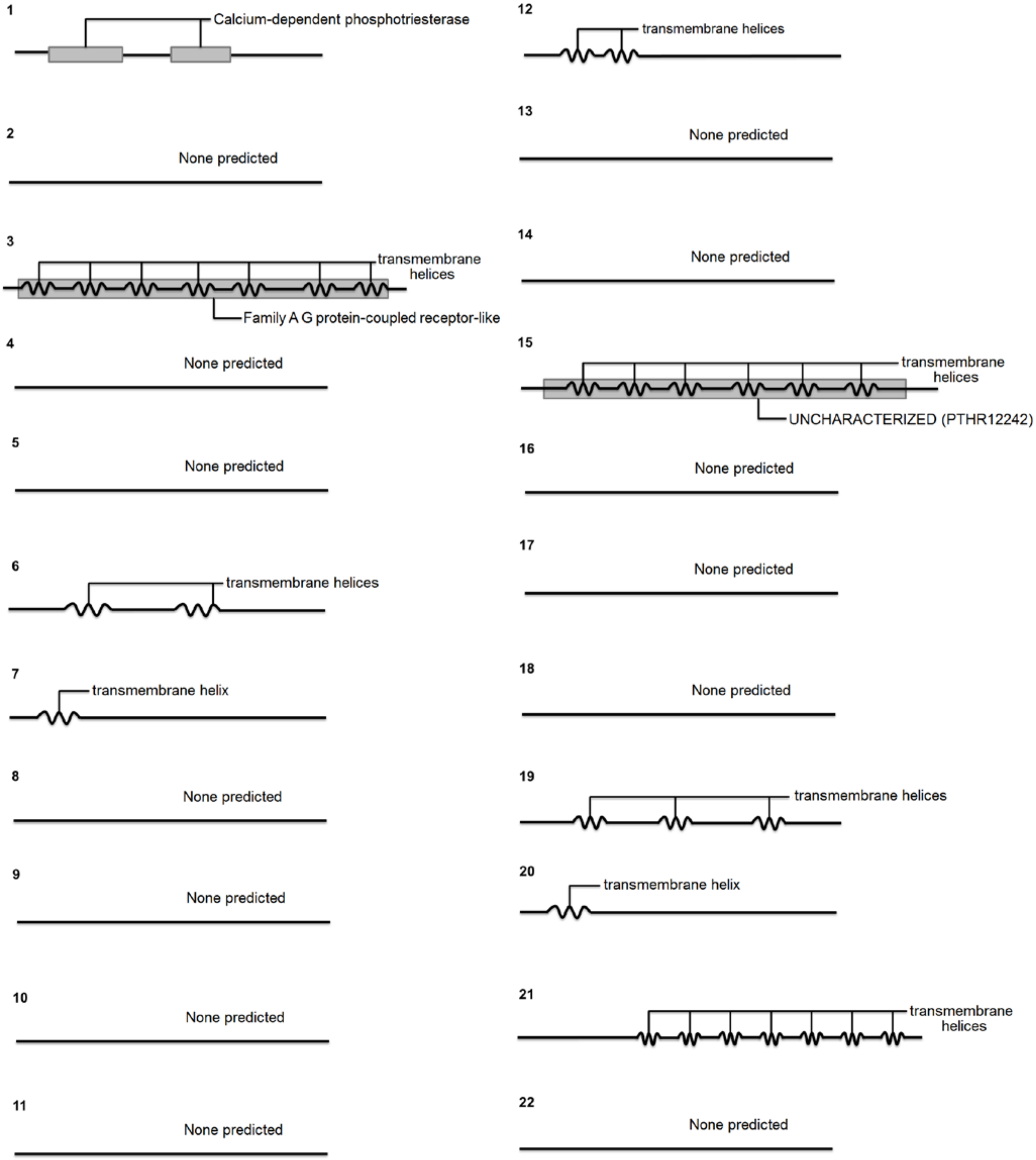


Figura 4. Domínios identificados para as proteínas hipotéticas conservadas. Descrição dos domínios conservados em todos os ortólogos das proteínas hipotéticas de cestódeos.

4.3.2. Ampliação do conjunto amostral das proteínas ortólogas

Considerando os resultados obtidos na sessão anterior, observou-se que as ortólogas das proteínas hipotéticas selecionadas estão restritas aos lofotrocozoários. Como a análise anterior se restringiu às 18 espécies estudadas (Apêndice 18), realizou-se uma busca por ortólogos para avaliar a presença destes em outras espécies.

De forma análoga ao observado para as proteínas do Capítulo I, poucos ortólogos foram identificados para as proteínas hipotéticas. A Tabela 3 descreve os resultados finais dos grupos de ortólogas, abrangendo as proteínas obtidas na análise inicial (Tabela 2) e a nova busca. Nesta última etapa, podemos observar que não ocorreu grande ampliação do número de espécies para cada grupo de ortólogos, porém, muitos parálogos foram adicionados. Novamente, apenas a proteína 15 apresenta ortólogas de espécies de moluscos e anelídeos, estando, portanto, restritas ao Superfilo Lophotrochozoa. Os demais grupos de ortólogos estão restritos apenas a espécies do Filo Platyhelminthes e, mais especificamente, doze grupos de ortólogos (3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 e 22), restritos à cestódeos.

Tabela 3. Resultados da busca por ortólogos das proteínas hipotéticas. Os táxons ao qual cada espécie está vinculada estão representados por diferentes cores: vermelho, para cestódeos; azul, para trematódeos; verde, para moluscos; e amarelo, para anelídeos.

Proteína hipotética	Táxon	Espécie	Identificação no NCBI[1]
1	🟥	*Echinococcus granulosus*	gi|674568676|emb|CDS17794.1| hypothetical protein EgrG_001056100
	🟥	*Echinococcus multilocularis*	gi|674572416|emb|CDS42841.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674594877|emb|CDS26379.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000802201-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684396902|ref|XP_009171675.1| hypothetical protein T265_14384, partial
	🟥	*Taenia asiatica*	gi|1046523282|gb|OCK26927.1| hypothetical protein TAS_TASs00013g02848
	🟥	*Taenia saginata*	gi|1046539392|gb|OCK37496.1| hypothetical protein TSA_TSAs00029g04632
	🟥	*Taenia solium**	TsM_001128600
2	🟦	*Clonorchis sinensis*	gi|358333364|dbj|GAA51882.1| hypothetical protein CLF_106961
	🟥	*Echinococcus granulosus*	gi|674568014|emb|CDS17128.1| hypothetical protein EgrG_000985800
	🟥	*Echinococcus multilocularis*	gi|674571737|emb|CDS42155.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595904|emb|CDS25473.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000991601-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684385662|ref|XP_009168232.1| hypothetical protein T265_05057
	🟦	*Schistosoma haematobium*	gi|844839738|ref|XP_012792983.1| hypothetical protein MS3_01373, partial
	🟦	*Schistosoma mansoni*	gi|353231386|emb|CCD77804.1| hypothetical protein Smp_023830
	🟥	*Taenia asiatica*	gi|1046524317|gb|OCK27898.1| hypothetical protein TAS_TASs00007g01841
	🟥	*Taenia solium**	TsM_000497700
3	🟥	*Echinococcus granulosus*	gi|674563883|emb|CDS21567.1| hypothetical protein EgrG_000105500
	🟥	*Echinococcus granulosus*	gi|674562419|emb|CDS23129.1| hypothetical protein EgrG_001089200
	🟥	*Echinococcus granulosus*	gi|576692638|gb|EUB56277.1| hypothetical protein EGR_08822
	🟥	*Echinococcus multilocularis*	gi|674266900|emb|CDI97288.1| conserved hypothetical protein
	🟥	*Echinococcus multilocularis*	gi|674571166|emb|CDS43160.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674594154|emb|CDS27120.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674592582|emb|CDS28604.1| amine GPCR
	🟥	*Hymenolepis microstoma*	gi|674592581|emb|CDS28603.1| amine GPCR
	🟥	*Hymenolepis microstoma*	gi|961499149|emb|CUU98304.1| centrin 3
	🟥	*Mesocestoides corti**	MCOS_0000773401-mRNA-1
	🟥	*Taenia asiatica*	gi|1046521111|gb|OCK24942.1| hypothetical protein TAS_TASs00042g05037
	🟥	*Taenia asiatica*	gi|1046517272|gb|OCK21598.1| hypothetical protein TAS_TASs00162g08633
	🟥	*Taenia saginata*	gi|1046537631|gb|OCK35758.1| hypothetical protein TSA_TSAs00052g06544
	🟥	*Taenia saginata*	gi|1046536077|gb|OCK34239.1| hypothetical protein TSA_TSAs00087g08235
	🟥	*Taenia solium**	TsM_000622300
4	🟦	*Clonorchis sinensis*	gi|358340976|dbj|GAA48759.1| hypothetical protein CLF_102001
	🟥	*Echinococcus granulosus*	gi|674564018|emb|CDS21702.1| hypothetical protein EgrG_000120000
	🟥	*Echinococcus granulosus*	gi|674569849|emb|CDS15917.1| hypothetical protein EgrG_000832200
	🟥	*Echinococcus multilocularis*	gi|674267035|emb|CDI97423.1| conserved hypothetical protein
	🟥	*Echinococcus multilocularis*	gi|674573805|emb|CDS40728.1| hypothetical transcript
	🟥	*Hymenolepis microstoma*	gi|674588949|emb|CDS32060.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000667601-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684379696|ref|XP_009166407.1| hypothetical protein T265_03608
	🟦	*Schistosoma haematobium*	gi|844856703|ref|XP_012797005.1| hypothetical protein MS3_05576
	🟦	*Schistosoma japonicum*	gi|56757137|gb|AAW26740.1| SJCHGC09165 protein
	🟦	*Schistosoma mansoni*	gi|353231296|emb|CCD77714.1| hypothetical protein Smp_065370.2
	🟥	*Taenia asiatica*	gi|1046524634|gb|OCK28192.1| hypothetical protein TAS_TASs00005g01325
	🟥	*Taenia asiatica*	gi|1046525943|gb|OCK29436.1| hypothetical protein TAS_TASs00001g00250
	🟥	*Taenia saginata*	gi|1046539835|gb|OCK37935.1| hypothetical protein TSA_TSAs00025g04276
	🟥	*Taenia saginata*	gi|1046538582|gb|OCK36695.1| hypothetical protein TSA_TSAs00038g05438
	🟥	*Taenia solium**	TsM_000367100
5	🟥	*Echinococcus granulosus*	gi|674564264|emb|CDS21264.1| hypothetical protein EgrG_000165400
	🟥	*Echinococcus multilocularis*	gi|674266400|emb|CDI97849.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595432|emb|CDS25834.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000886801-mRNA-1
	🟥	*Taenia asiatica*	gi|1046519163|gb|OCK23187.1| hypothetical protein TAS_TASs00084g06930
	🟥	*Taenia saginata*	gi|1046536420|gb|OCK34571.1| hypothetical protein TSA_TSAs00076g07850
	🟥	*Taenia solium**	TsM_001053500
6	🟥	*Echinococcus granulosus*	gi|674561323|emb|CDS24345.1| Shisa domain containing protein

	🟥	*Echinococcus multilocularis*	gi|674578243|emb|CDS36181.1| hypothetical transcript
	🟥	*Hymenolepis microstoma*	gi|674590297|emb|CDS30793.1| hypothetical protein HmN_000314600
	🟥	*Mesocestoides corti**	MCOS_0000192801-mRNA-1
	🟥	*Taenia asiatica*	gi|1046519749|gb|OCK23710.1| hypothetical protein TAS_TASs00070g06388
	🟥	*Taenia saginata*	gi|1046529841|gb|OCK29794.1| hypothetical protein TSA_TSAs01884g12961
	🟥	*Taenia solium**	TsM_000764000
7	🟥	*Echinococcus granulosus*	gi|576696995|gb|EUB60542.1| hypothetical protein EGR_04561
	🟥	*Echinococcus multilocularis*	gi|961439464|emb|CUT98960.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595985|emb|CDS25297.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000657801-mRNA-1
	🟥	*Taenia asiatica*	gi|1046520367|gb|OCK24266.1| hypothetical protein TAS_TASs00055g05808
	🟥	*Taenia saginata*	gi|1046537702|gb|OCK35828.1| hypothetical protein TSA_TSAs00051g06485
	🟥	*Taenia solium**	TsM_000941800
8	🟥	*Echinococcus granulosus*	gi|674568962|emb|CDS15019.1| hypothetical protein EgrG_000742100
	🟥	*Echinococcus multilocularis*	gi|674572964|emb|CDS39872.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674593925|emb|CDS27298.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000902901-mRNA-1
	🟥	*Taenia asiatica*	gi|1046518442|gb|OCK22565.1| zinc finger C2H2 type
	🟥	*Taenia saginata*	gi|1046535786|gb|OCK33958.1| zinc finger C2H2 type
	🟥	*Taenia solium**	TsM_000992500
9	🟥	*Echinococcus granulosus*	gi|674568982|emb|CDS15040.1| hypothetical protein EgrG_000744200
	🟥	*Echinococcus multilocularis*	gi|674572983|emb|CDS39892.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595269|emb|CDS26053.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000440601-mRNA-1
	🟥	*Taenia asiatica*	gi|1046517934|gb|OCK22135.1| hypothetical protein TAS_TASs00128g08059
	🟥	*Taenia saginata*	gi|1046535372|gb|OCK33564.1| hypothetical protein TSA_TSAs00117g08925
	🟥	*Taenia solium**	TsM_000207900
10	🟥	*Echinococcus granulosus*	gi|674569295|emb|CDS15358.1| hypothetical protein EgrG_000775300
	🟥	*Echinococcus multilocularis*	gi|674573272|emb|CDS40186.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674590959|emb|CDS30258.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000970201-mRNA-1
	🟥	*Taenia asiatica*	gi|1046525587|gb|OCK29100.1| hypothetical protein TAS_TASs00002g00637
	🟥	*Taenia saginata*	gi|1046542078|gb|OCK40159.1| hypothetical protein TSA_TSAs00006g01611
	🟥	*Taenia solium**	TsM_000232000
11	🟥	*Echinococcus granulosus*	gi|674569942|emb|CDS16010.1| hypothetical protein EgrG_000842400
	🟥	*Echinococcus multilocularis*	gi|674573900|emb|CDS40823.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674588787|emb|CDS32269.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000259301-mRNA-1
	🟥	*Taenia asiatica*	gi|1046526019|gb|OCK29512.1| hypothetical protein TAS_TASs00001g00330
	🟥	*Taenia saginata*	gi|1046541817|gb|OCK39900.1| hypothetical protein TSA_TSAs00008g02141
	🟥	*Taenia solium**	TsM_000189400
12	🟥	*Echinococcus granulosus*	gi|674561049|emb|CDS24598.1| hypothetical protein EgrG_000934900
	🟥	*Echinococcus multilocularis*	gi|674572720|emb|CDS41678.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595230|emb|CDS26086.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000552701-mRNA-1
	🟥	*Taenia asiatica*	gi|1046520378|gb|OCK24272.1| hypothetical protein TAS_TASs00054g05731
	🟥	*Taenia saginata*	gi|1046538573|gb|OCK36687.1| hypothetical protein TSA_TSAs00039g05608
	🟥	*Taenia solium**	TsM_000499600
13	🟥	*Echinococcus granulosus*	gi|674568682|emb|CDS17800.1| hypothetical protein EgrG_001056700
	🟥	*Echinococcus multilocularis*	gi|674572422|emb|CDS42847.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674586073|emb|CDS34689.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000802601-mRNA-1
	🟥	*Taenia asiatica*	gi|1046523286|gb|OCK26931.1| hypothetical protein TAS_TASs00013g02852
	🟥	*Taenia asiatica*	gi|1046513210|gb|OCK19242.1| hypothetical protein TAS_TASs00691g11178
	🟥	*Taenia asiatica*	gi|1046513218|gb|OCK19247.1| hypothetical protein TAS_TASs00690g11177
	🟥	*Taenia solium**	TsM_000588300
14	🟥	*Echinococcus granulosus*	gi|674560738|emb|CDS24912.1| Pfam-B_2037 domain containing protein
	🟥	*Echinococcus multilocularis*	gi|674570679|emb|CDS43750.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595483|emb|CDS25885.1| hypothetical protein HmN_000131700
	🟥	*Mesocestoides corti**	MCOS_0000832501-mRNA-1
	🟥	*Taenia asiatica*	gi|1046520025|gb|OCK23958.1| expressed conserved protein
	🟥	*Taenia saginata*	gi|1046533151|gb|OCK31616.1| expressed conserved protein
	🟥	*Taenia solium**	TsM_000515500
15	🟦	*Clonorchis sinensis*	gi|358342287|dbj|GAA49786.1| hypothetical protein CLF_103597
	🟩	*Crassostrea gigas*	gi|405964788|gb|EKC30234.1| Protein rolling stone
	🟥	*Echinococcus granulosus*	gi|674565835|emb|CDS20385.1| expressed protein
	🟥	*Echinococcus granulosus*	gi|674561716|emb|CDS24031.1| hypothetical protein EgrG_000146900

	🟥	*Echinococcus granulosus*	gi|576696242|gb|EUB59798.1| hypothetical protein EGR_05274
	🟥	*Echinococcus multilocularis*	gi|674266228|emb|CDI98735.1| expressed protein
	🟥	*Echinococcus multilocularis*	gi|674266707|emb|CDI97586.1| conserved hypothetical protein
	🟨	*Helobdella robusta*	gi|675872564|ref|XP_009021842.1| hypothetical protein HELRODRAFT_176378
	🟥	*Hymenolepis microstoma*	gi|674592844|emb|CDS28382.1| hypothetical protein HmN_000810600
	🟥	*Hymenolepis microstoma*	gi|674588718|emb|CDS32315.1| expressed conserved protein
	🟥	*Hymenolepis microstoma*	gi|961496169|emb|CDS35323.2| hypothetical transcript
	🟥	*Hymenolepis microstoma*	gi|674587493|emb|CDS33452.1| hypothetical protein HmN_000519000
	🟥	*Hymenolepis microstoma*	gi|961387800|emb|CUU99937.1| hypothetical transcript
	🟥	*Hymenolepis microstoma*	gi|961390005|emb|CUU98388.1| hypothetical transcript
	🟥	*Hymenolepis microstoma*	gi|674584895|emb|CDS35369.1| protein rolling stone
	🟥	*Hymenolepis microstoma*	gi|674588989|emb|CDS32040.1| expressed protein
	🟥	*Hymenolepis microstoma*	gi|674589218|emb|CDS31821.1| expressed protein
	🟩	*Lollita gigantea*	gi|676437423|ref|XP_009048199.1| hypothetical protein LOTGIDRAFT_205169
	🟥	*Mesocestoides corti**	MCOS_0000928701-mRNA-1
	🟥	*Mesocestoides corti**	MCOS_0000669301-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684379816|ref|XP_009166443.1| hypothetical protein T265_03644
	🟦	*Schistosoma haematobium*	gi|844876470|ref|XP_012801761.1| Protein rolling stone, partial
	🟦	*Schistosoma mansoni*	gi|353230088|emb|CCD76259.1| hypothetical protein Smp_059820
	🟥	*Taenia asiatica*	gi|1046515954|gb|OCK20680.1| hypothetical protein TAS_TASs00282g09656
	🟥	*Taenia saginata*	gi|1046543251|gb|OCK41327.1| hypothetical protein TSA_TSAs00001g00133
	🟥	*Taenia saginata*	gi|1046537844|gb|OCK35968.1| hypothetical protein TSA_TSAs00049g06369
	🟥	*Taenia solium**	TsM_000360800
	🟥	*Taenia solium**	TsM_000164000
16	🟥	*Echinococcus granulosus*	gi|674562746|emb|CDS23002.1| expressed conserved protein
	🟥	*Echinococcus granulosus*	gi|674562747|emb|CDS23003.1| hypothetical protein EgrG_000701600
	🟥	*Echinococcus multilocularis*	gi|674574984|emb|CDS39486.1| expressed conserved protein
	🟥	*Echinococcus multilocularis*	gi|674574985|emb|CDS39487.1| hypothetical transcript
	🟥	*Hymenolepis microstoma*	gi|961497798|emb|CDS27390.2| expressed conserved protein
	🟥	*Hymenolepis microstoma*	gi|961497799|emb|CDS27391.2| expressed protein
	🟥	*Mesocestoides corti**	MCOS_0000895701-mRNA-1
	🟥	*Mesocestoides corti**	MCOS_0000951901-mRNA-1
	🟥	*Mesocestoides corti**	MCOS_0001007101-mRNA-1
	🟥	*Mesocestoides corti**	MCOS_0000382601-mRNA-1
	🟦	*Schistosoma haematobium*	gi|844863585|ref|XP_012798606.1| hypothetical protein MS3_07259, partial
	🟥	*Taenia asiatica*	gi|1046523441|gb|OCK27075.1| expressed conserved protein
	🟥	*Taenia asiatica*	gi|1046523440|gb|OCK27074.1| hypothetical protein TAS_TASs00012g02742
	🟥	*Taenia saginata*	gi|1046542501|gb|OCK40580.1| expressed conserved protein
	🟥	*Taenia saginata*	gi|1046542500|gb|OCK40579.1| hypothetical protein TSA_TSAs00004g01267
	🟥	*Taenia solium**	TsM_001234000
	🟥	*Taenia solium**	TsM_001245100
	🟥	*Taenia solium**	TsM_000507900
	🟥	*Taenia solium**	TsM_001233900
17	🟥	*Echinococcus granulosus*	gi|674566918|emb|CDS18265.1| hypothetical protein EgrG_000602500
	🟥	*Echinococcus granulosus*	gi|576697007|gb|EUB60554.1| hypothetical protein EGR_04573
	🟥	*Echinococcus multilocularis*	gi|961439472|emb|CUT98968.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595997|emb|CDS25309.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000870601-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684390686|ref|XP_009169754.1| hypothetical protein T265_06270
	🟥	*Taenia asiatica*	*gi|1046520357|gb|OCK24256.1| hypothetical protein TAS_TASs00055g05797*
	🟥	*Taenia saginata*	gi|1046539543|gb|OCK37646.1| hypothetical protein TSA_TSAs00028g04590
	🟥	*Taenia solium**	TsM_000941200
18	🟦	*Clonorchis sinensis*	gi|358342778|dbj|GAA50229.1| hypothetical protein CLF_104262
	🟥	*Echinococcus granulosus*	gi|674566917|emb|CDS18264.1| hypothetical protein EgrG_000602400
	🟥	*Echinococcus granulosus*	gi|576697006|gb|EUB60553.1| hypothetical protein EGR_04572
	🟥	*Echinococcus multilocularis*	gi|961439471|emb|CUT98967.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674595996|emb|CDS25308.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000870701-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684377312|ref|XP_009165679.1| hypothetical protein T265_03026
	🟦	*Schistosoma haematobium*	gi|844873123|ref|XP_012800968.1| hypothetical protein MS3_09709
	🟦	*Schistosoma mansoni*	gi|360044317|emb|CCD81864.1| hypothetical protein Smp_015760
	🟥	*Taenia asiatica*	gi|1046520359|gb|OCK24258.1| hypothetical protein TAS_TASs00055g05799
	🟥	*Taenia saginata*	gi|1046539545|gb|OCK37648.1| hypothetical protein TSA_TSAs00028g04592
	🟥	*Taenia solium**	TsM_000941300
19	🟦	*Clonorchis sinensis*	gi|358254857|dbj|GAA56484.1| hypothetical protein CLF_110980
	🟥	*Echinococcus granulosus*	gi|674564898|emb|CDS20445.1| hypothetical protein EgrG_001110400
	🟥	*Echinococcus multilocularis*	gi|674570730|emb|CDS43351.1| conserved hypothetical protein

	🟥	*Hymenolepis microstoma*	gi|674590644|emb|CDS30534.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000580801-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684406449|ref|XP_009174784.1| hypothetical protein T265_10213
	🟥	*Taenia asiatica*	gi|1046519383|gb|OCK23386.1| hypothetical protein TAS_TASs00079g06727
	🟥	*Taenia saginata*	gi|1046536344|gb|OCK34497.1| hypothetical protein TSA_TSAs00078g07920
	🟥	*Taenia solium**	TsM_001060200
	🟥	*Taenia solium**	TsM_000069400
20	🟦	*Clonorchis sinensis*	gi|358336271|dbj|GAA54817.1| hypothetical protein CLF_105500
	🟥	*Echinococcus granulosus*	gi|576698627|gb|EUB62159.1| hypothetical protein EGR_02911
	🟥	*Echinococcus multilocularis*	gi|674572190|emb|CDS42615.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674594064|emb|CDS27186.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674594039|emb|CDS27260.1| fructose 26 bisphosphatase TIGAR
	🟥	*Mesocestoides corti**	MCOS_0000237701-mRNA-1
	🟦	*Schistosoma mansoni*	gi|360043561|emb|CCD78974.1| hypothetical protein Smp_015100
	🟥	*Taenia saginata*	gi|1046542981|gb|OCK41058.1| hypothetical protein TSA_TSAs00002g00817
	🟥	*Taenia solium**	TsM_001099900
21	🟥	*Echinococcus granulosus*	gi|674564525|emb|CDS20841.1| hypothetical protein EgrG_000518800
	🟥	*Echinococcus granulosus*	gi|576694239|gb|EUB57831.1| hypothetical protein EGR_07302
	🟥	*Echinococcus multilocularis*	gi|674576199|emb|CDS37897.1| hypothetical protein EmuJ_000518800
	🟥	*Hymenolepis microstoma*	gi|674590032|emb|CDS31159.1| hypothetical protein HmN_000058200
	🟥	*Mesocestoides corti**	MCOS_0000072301-mRNA-1
	🟦	*Opisthorchis viverrini*	gi|684388333|ref|XP_009169039.1| hypothetical protein T265_13834
	🟥	*Taenia asiatica*	gi|1046518042|gb|OCK22226.1| regulator of G protein signaling 3
	🟥	*Taenia saginata*	gi|1046539173|gb|OCK37280.1| regulator of G protein signaling 3
	🟥	*Taenia solium**	TsM_001120700
	🟥	*Taenia solium**	TsM_000568300
22	🟥	*Echinococcus granulosus*	gi|674567674|emb|CDS16784.1| hypothetical protein EgrG_000949700
	🟥	*Echinococcus granulosus*	gi|576692312|gb|EUB55965.1| hypothetical protein EGR_09169
	🟥	*Echinococcus multilocularis*	gi|674571402|emb|CDS41816.1| conserved hypothetical protein
	🟥	*Hymenolepis microstoma*	gi|674592177|emb|CDS29001.1| conserved hypothetical protein
	🟥	*Mesocestoides corti**	MCOS_0000969401-mRNA-1
	🟥	*Mesocestoides corti**	MCOS_0000375801-mRNA-1
	🟥	*Taenia asiatica*	gi|1046522587|gb|OCK26289.1| hypothetical protein TAS_TASs00021g03578
	🟥	*Taenia saginata*	gi|1046539994|gb|OCK38092.1| hypothetical protein TSA_TSAs00023g04009
	🟥	*Taenia solium**	TsM_000994400
	🟥	*Taenia solium**	TsM_000431000

---

[1] https://www.ncbi.nlm.nih.gov/

[2] Identificação retirada do genoma de referência (Apêndice 18)

5. DISCUSSÃO

A proglotização tem sido considerada uma anomalia pela maioria dos biólogos do desenvolvimento, já que esse tipo de segmentação corporal está presente apenas na Subclasse Eucestoda, não ocorrendo em nenhum outro tipo de platelminto ou animal (Blair, 2008). Além disso, ao contrário de outros metazoários, esse tipo de segmentação evoluiu como uma adaptação ao parasitismo (aumentando a fecundidade), e não à locomoção (Riddiford & Olson, 2011).

Nesse sentido, estudos envolvendo construção de bibliotecas de cDNA (Bizarro et al., 2005), análises proteômicas (Laschuk et al., 2011; Cui et al., 2013; Debarba et al., 2015) e estudos transcritômicos de mRNAs e miRNAs (Tsai et al., 2013; Basika et al., 2016) entre as fases larval e adulta de cestódeos proglotizados, descrevem conjuntos de transcritos/proteínas enriquecidos e exclusivos de cada estágio de desenvolvimento. Porém, todas as abordagens descritas até o momento focaram-se em descrever o processo de desenvolvimento utilizando apenas uma espécie em cada estudo. Apenas recentemente tornou-se possível realizar análises de genômica comparativa para avaliar características evolutivamente conservadas em grupos de cestódeos. Portanto, esse trabalho aborda o tema de uma forma alternativa às anteriormente descritas obtendo, assim, resultados complementares de proteínas que podem ter passado despercebidas em análises realizadas com apenas uma espécie.

Considerando as espécies estudadas, os resultados deste trabalho descrevem proteínas com alto nível de conservação em cestódeos e que estão ausentes em, pelo menos, um trematódeo, considerando os grupos de proteínas ortólogas identificadas com base na conservação das sequências aminoacídicas. Sabendo-se que as Classes Cestoda e Trematoda são altamente relacionadas evolutivamente, essa análise pode ser considerada bastante estringente e, de fato, muitas proteínas identificadas possuem sequências muito diferenciadas das de outros animais e conservadas exclusivamente em cestódeos. Devido a este fato, poderia ser esperado que se identificassem poucos ortólogos para as proteínas analisadas, pois poucas regiões são conservadas em espécies de outros táxons, e poucas espécies de cestódeos possuem as sequências de seus proteomas disponíveis nos bancos de dados.

No Capítulo I estão descritos os resultados da busca por proteínas relacionadas ao desenvolvimento. A escolha destas proteínas se deve ao fato de que, em geral, um pequeno número de diferentes sistemas de sinalização são compartilhados em todos os animais e responsáveis pelo desenvolvimento destes (Pires-daSilva & Sommer, 2003). Levando em consideração esse fato, associado à simplificação genômica descrita para cestódeos, seria esperado que vias de sinalização conhecidas e bem descritas, relacionadas ao desenvolvimento, estivessem envolvidas na proglotização.

Considerando as proteínas relacionadas ao desenvolvimento, podemos dividir as 12 proteínas identificadas em dois conjuntos principais. O primeiro conjunto é referente às proteínas que podem ser associadas a vias de sinalização conhecidas. Entre estas, a identificação de proteínas da via de sinalização da Wnt não é surpreendente, visto que trabalhos anteriores já sugeriram o envolvimento desta via com a segmentação de platelmintos (Riddiford & Olson, 2011). A proteína TFC é um fator de transcrição desta via metabólica e tem sua atividade reprimida pela interação com seu repressor Groucho. Interessantemente, a análise transcritômica em *E. multilocularis* (Table 1) identifica o fator de transcrição com expressão aumentada na fase adulta, e seu repressor com expressão diminuída, o que sugere o envolvimento deste fator de transcrição no desenvolvimento deste estágio de vida. Sabendo-se que a via de sinalização da Wnt está envolvida na especificação dos eixos ântero-posterior na regeneração de alguns platelmintos não proglotizados (planárias) (Lin & Pearson, 2014), facilmente levantam-se questionamentos quanto ao local em que atuam estas proteínas na metamorfose de cestódeos adultos. O padrão da proglotização sugere que essa via de sinalização apresente a sua atividade serialmente na extensão do corpo destes organismos, de forma a regular as extremidades ântero-posterior do desenvolvimento de cada proglótide. Esse tipo de informação poderia ser avaliada através de estudos de localização espacial de proteínas marcadoras desta via como, por exemplo, por imuno-histoquímica.

O segundo conjunto é composto pelas proteínas com anotação funcional relacionada ao desenvolvimento, e sem vinculação à nenhuma via metabólica descrita. Quanto a estas, pode-se especular quanto às suas funções através da descrição destas em proteínas com o mesmo perfil de domínios descritos para outros animais. Em ratos, a proteína GAK atua na regulação da via de transporte vesicular dependente de clatrina no complexo de Golgi (Zhang et al., 2005) e é essecial para o desenvolvimento de órgãos como o cérebro, fígado e pele (Lee et al., 2008). Assim, as informações da alta conservação de sequência da proteína GAK de cestódeos juntamente com os dados de transcrição, expressão aumentada no estágio adulto de *E. multilocularis* (Table 1), sugerem seu envolvimento no desenvolvimento de cestódeos através da regulação do transporte vesicular.

No entanto, pouco se pode inferir sobre as possíveis funções que algumas proteínas identificadas possam estar desempenhando na proglotização. Um exemplo é a proteína MAGI, um importante regulador da plasticidade e adesão de junções celulares e envolvida na neurogênese (Wright, 2004; Funke et al., 2005). A proteína NPR1 pertencente à família das guanilil-cinases, catalisa a transformação de GTP em cGTP, e pode estar envolvida em muitos processos regulados pela via de sinalização mediada por cAMP (Johnston et al., 2001).

Finalmente, são descritas proteínas às quais não foi possível atribuir um processo biológico específico. Esse é o caso das proteínas serina/treonina-cinase e RBMS, para as quais serão necessários estudos complementares para a atribuição de possíveis funções. Uma forma de identificá-las seria através, por exemplo, da análise da correlação de transcritos, que permitiria a identificação de outras proteínas relacionadas a estas (Langfelder & Horvath, 2008). Além disso, a análise dos agrupamentos de transcritos com altos valores de correlação da expressão permitiria inferir possíveis genes regulados pelos fatores de transcrição identificados nesse trabalho.

Além de estudar proteínas previamente relacionadas ao desenvolvimento, é necessário levar em consideração que a proglotização é um processo que ocorre unicamente em uma subclasse de cestódeos. Dessa forma, não pode ser ignorado o possível envolvimento de proteínas desconhecidas neste processo. Assim, o Capítulo II descreve o estudo realizado na busca de proteínas hipotéticas relacionadas à proglotização, complementando os resultados obtidos no Capítulo I.

No Capítulo II é apresentado o terceiro conjunto de proteínas, que não possuem nem anotação funcional, nem vinculação a vias metabólicas. Estes resultados, ainda em caráter preliminar, dão margens a futuros estudos *in silico* e *in vitro* do envolvimento destas no processo de proglotização. As proteínas hipotéticas avaliadas apresentam altos níveis de conservação entre os cestódeos e são comprovadamente expressas em, pelo menos, uma espécie (Tabela 2). A conservação destas proteínas apenas em cestódeos, em sua maioria, torna possível inferir que estejam relacionadas a processos específicos dessa Classe. Análises futuras incluem avaliar as regiões de similaridade das proteínas ortólogas já identificadas neste estudo, além de dar seguimento às análises de evolução molecular para melhor elucidar a história evolutiva individual de cada proteína.

6. PERSPECTIVAS

- Realização das análises de evolução molecular das proteínas hipotéticas;

- Realização de estudos *in silico* da coexpressão de transcritos para identificação das possíveis interações e funções desempenhadas pelas proteínas associadas ao processo de proglotização;

- Análise dos padrões espaço-temporais de expressão de genes e proteínas de interesse durante a estrobilização de *M. corti*, de modo a evidenciar o envolvimento destes genes/proteínas neste processo de desenvolvimento;

- Caracterização funcional de proteínas envolvidas no processo de estrobilização de platelmintos da Classe Cestoda, como base para a identificação de genes/proteínas marcadoras de desenvolvimento.

REFERÊNCIAS BIBLIOGRÁFICAS

Almeida, C. R., Stoco, P. H., Wagner, G., Sincero, T. C., Rotava, G., Bayer-Santos, E., Rodrigues, J. B., Sperandio, M. M., Maia, A. A., Ojopi, E. P., Zaha, A., Ferreira, H. B., Tyler, K. M., Dávila, A. M., Grisard, E. C., & Dias-Neto, E. (2009) Transcriptome analysis of Taenia solium cysticerci using Open Reading Frame ESTs (ORESTES). *Parasites & Vectors*, *2*(1), 35.

Basika, T., Macchiaroli, N., Cucher, M., Espínola, S., Kamenetzky, L., Zaha, A., Rosenzvit, M., & Ferreira, H. B. (2016) Identification and profiling of microRNAs in two developmental stages of the model cestode parasite Mesocestoides corti. *Molecular and Biochemical Parasitology*.

Bizarro, C. V., Bengtson, M. H., Ricachenevsky, F. K., Zaha, A., Sogayar, M. C., & Ferreira, H. B. (2005) Differentially expressed sequences from a cestode parasite reveals conserved developmental genes in platyhelminthes. *Molecular and Biochemical Parasitology*, *144*(1), 114–118.

Blair, S. S. (2008) Segmentation in animals. *Current Biology*, *18*(21), R991–R995.

Budke, C. M., White, A. C., & Garcia, H. H. (2009) Zoonotic larval cestode infections: neglected, neglected tropical diseases? *PLoS neglected tropical diseases*, *3*(2), e319.

Chervy, L. (2002) The terminology of larval cestodes or metacestodes. *Systematic Parasitology*, *52*(1), 1-33.

Coral-Almeida, M., Gabriël, S., Abatih, E. N., Praet, N., Benitez, W., & Dorny, P. (2015) Taenia solium Human Cysticercosis: A Systematic Review of Sero-epidemiological Data from Endemic Zones around the World. *PLOS Neglected Tropical Diseases*, *9*(7), e0003919.

Couso, J. P. (2009) Segmentation, metamerism and the Cambrian explosion. *The International Journal of Developmental Biology*, *53*, 8–10.

Cucher, M. A., Macchiaroli, N., Baldi, G., Camicia, F., Prada, L., Maldonado, L., Avila, H. G., Fox, A., Gutiérrez, A., Negro, P., López, R., Jensen, O., Rosenzvit, M., & Kamenetzky, L. (2016) Cystic echinococcosis in South America: systematic review of species and genotypes of Echinococcus granulosus sensu lato in humans and natural domestic hosts. *Tropical Medicine & International Health*, *21*(2), 166–175.

Cui, S. J., Xu, L. L., Zhang, T., Xu, M., Yao, J., Fang, C. Y., Feng, Z., Yang, P. Y., Hu, W., & Liu, F. (2013) Proteomic characterization of larval and adult developmental stages in Echinococcus granulosus reveals novel insight into host–parasite interactions. *Journal of Proteomics*, *84*, 158–175.

Dalton, J. P., Skelly, P., & Halton, D. W. (2004) Role of the tegument and gut in nutrient uptake by parasitic platyhelminths. *Canadian Journal of Zoology*, *82*(2), 211–232.

Debarba, J. A., Monteiro, K. M., Moura, H., Barr, J. R., Ferreira, H. B., & Zaha, A. (2015) Identification of Newly Synthesized Proteins by Echinococcus granulosus

Protoscoleces upon Induction of Strobilation. *PLOS Neglected Tropical Diseases*, *9*(9), e0004085.

Funke, L., Dakoji, S., & Bredt, D. S. (2005) Membrane-associated guanylate kinases regulate adhesion and platicity at cell junctions. *Annual Review of Biochemistry*, *74*(1), 219–245.

Gabriël, S., Dorny, P., Mwape, K. E., Trevisan, C., Braae, U. C., Magnussen, P., Thys, S., Bulaya, C., Phiri, I. K., Sikasunge, C. S., Makungu, C., Afonso, S., Nicolau, Q., & Johansen, M. V. (2016) Control of Taenia solium taeniasis/cysticercosis: The best way forward for sub-Saharan Africa? *Acta Tropica*.

Hahn, C., Fromm, B., & Bachmann, L. (2014) Comparative Genomics of Flatworms (Platyhelminthes) Reveals Shared Genomic Features of Ecto-and Endoparastic Neodermata. *Genome Biology and Evolution*, *6*(5), 1105–1117.

Heyneman, D. (1996) Cestodes. S. Baro (Ed), *Medical Microbiology* (4th ed). University of Texas Medical Branch at Galveston, Galveston (TX).

Johnston, S. D., Enomoto, S., Schneper, L., McClellan, M. C., Twu, F., Montgomery, N. D., Haney, S. A., Broach, J. R., & Berman, J. (2001) CAC3(MSI1) suppression of RAS2(G19V) is independent of chromatin assembly factor I and mediated by NPR1. *Molecular and cellular biology*, *21*(5), 1784–94.

Kearn, G. C. (1994) Evolutionary expansion of the Monogenea. *International journal* *for parasitology*, *24*(8), 1227–71.

Langfelder, P., & Horvath, S. (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*(1), 559.

Laschuk, A., Monteiro, K. M., Vidal, N. M., Pinto, P. M., Duran, R., Cerveñanski, C., Zaha, A., & Ferreira, H. B. (2011) Proteomic survey of the cestode Mesocestoides corti during the first 24 hours of strobilar development. *Parasitology Research*, *108*(3), 645–656.

Lee, D. W., Zhao, X., Yim, Y. I., Eisenberg, E., & Greene, L. E. (2008) Essential Role of Cyclin-G-associated Kinase (Auxilin-2) in Developing and Mature Mice. *Molecular Biology of the Cell*, *19*(7), 2766–2776.

Lin, A. Y. T., & Pearson, B. J. (2014) Planarian yorkie/YAP functions to integrate adult stem cell proliferation, organ homeostasis and maintenance of axial patterning. *Development*, *141*(6), 1197–1208.

Littlewood, D. T. J. (1999) Phylogeny of the Platyhelminthes and the evolution of parasitism. *Biological Journal of the Linnean Society*, *68*(1–2), 257–287.

Littlewood, D. T. J., Cribb, T. H., Olson, P. D., e Bray, R. A. (2001) Platyhelminth phylogenetics – a key to understanding parasitism? *Belgian Journal of Zoology*, *131*(1), 35–46.

Littlewood, D. T. J. (2006) Evolution of Parasitism in Flatworms. A. G. Maule & N. J. Marks (Eds), *Parasitic Flatworms: Molecular Biology, Biochemistry, Immunology* *and Physiology* (p. 1–36). CABI.

Lockyer, A. E., Olson, P. D., & Littlewood, D. T. J. (2003) 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. *Biological Journal of the Linnean Society*, *78*(2), 155–171.

Lorenzatto, K. R., Kim, K., Ntai, I., Paludo, G. P., Camargo de Lima, J., Thomas, P. M., Kelleher, N. L., & Ferreira, H. B. (2015) Top Down Proteomics Reveals Mature Proteoforms Expressed in Subcellular Fractions of the Echinococcus granulosus Preadult Stage. *Journal of Proteome Research*, *14*(11), 4805–4814.

Muehlenbachs, A., Bhatnagar, J., Agudelo, C. A., Hidron, A., Eberhard, M. L., Mathison, B. A., Frace, M. A., Ito, A., Metcalfe, M. G., Rollin, D. C., Visvesvara, G. S., Pham, C. D., Jones, T. L., Greer, P. W., Vélez Hoyos, A., Olson, P. D., Diazgranados, L. R., & Zaki, S. R. (2015) Malignant Transformation of Hymenolepis nana in a Human Host. *New England Journal of Medicine*, *373*(19), 1845–1852.

Olson, P. D., Timothy, D., Littlewood, J., Bray, R. A., & Mariaux, J. (2001) Interrelationships and Evolution of the Tapeworms (Platyhelminthes: Cestoda). *Molecular Phylogenetics and Evolution*, *19*(3), 443–467.

Olson, P. D., Cribb, T. H., Tkach, V. V., Bray, R. A., & Littlewood, D. T. J. (2003) Phylogeny and classification of the Digenea (Platyhelminthes: Trematoda). *International Journal for Parasitology*, *33*(7), 733–755.

Olson, P. D., & Tkach, V. V. (2005) Advances and Trends in the Molecular Systematics of the Parasitic Platyhelminthes. (p. 165–243).

Olson, P. D., Zarowiecki, M., Kiss, F., & Brehm, K. (2012) Cestode genomics - progress and prospects for advancing basic and applied aspects of flatworm biology. *Parasite immunology*, *34*(2–3), 130–50.

Park, J.-K., Kim, K.-H., Kang, S., Kim, W., Eom, K. S., & Littlewood, D. (2007) A common origin of complex life cycles in parasitic flatworms: evidence from the complete mitochondrial genome of Microcotyle sebastis (Monogenea: Platyhelminthes). *BMC Evolutionary Biology*, *7*(1), 11.

Pires-daSilva, A., & Sommer, R. J. (2003) The evolution of signalling pathways in animal development. *Nature Reviews Genetics*, *4*(1), 39–49.

Riddiford, N., & Olson, P. D. (2011) Wnt gene loss in flatworms. *Development Genes and Evolution*, *221*(4), 187–197.

Rohde, K. (1994) The origins of parasitism in the platyhelminthes. *International journal for parasitology*, *24*(8), 1099–115.

Rohde, K. (2001) The Aspidogastrea: an archaic group of Platyhelminthes. D. T. J. Littlewood & R. A. Bray (Eds), *Interrelationships of the Platyhelminthes* (1st ed, p. 159–167). Taylor & Francis, London and New York.

Scholz, T., Garcia, H. H., Kuchta, R., & Wicht, B. (2009) Update on the Human Broad Tapeworm (Genus Diphyllobothrium), Including Clinical Relevance. *Clinical*

Microbiology Reviews, *22*(1), 146–160.

Sharma, S., Lyngdoh, D., Roy, B., & Tandon, V. (2016) Differential diagnosis and molecular characterization of Hymenolepis nana and Hymenolepis diminuta (Cestoda: Cyclophyllidea: Hymenolepididae) based on nuclear rDNA ITS2 gene marker. *Parasitology Research*.

Teklemariam, A. D., & Debash, W. (2015) Prevalence of Taenia Saginata/Cysticercosis and Community Knowledge about Zoonotic Cestodes in and Around Batu, Ethiopia. *Journal of Veterinary Science & Technology*, *6*(6).

Thompson, R. C. A. (2008) The taxonomy, phylogeny and transmission of Echinococcus. *Experimental Parasitology*, *119*(4), 439–446.

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garciarrubio, A., Sanchez-Flores, A., Brooks, K. L., Tracey, A., Bobes, R. J., Fragoso, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H. M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T. A., Deplazes, P., Estrada, K., Fernández, C., Holland, P. W. H., Hou, J., Hu, S., Huckvale, T., Hung, S. S., Kamenetzky, L., Keane, J. A., Kiss, F., Koziol, U., Lambert, O., Liu, K., Luo, X., Luo, Y., Macchiaroli, N., Nichol, S., Paps, J., Parkinson, J., Pouchkina-Stantcheva, N., Riddiford, N., Rosenzvit, M., Salinas, G., Wasmuth, J. D., Zamanian, M., Zheng, Y., Taenia solium Genome Consortium, Cai, X., Soberón, X., Olson, P. D., Laclette, J. P., Brehm, K., & Berriman, M. (2013) The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, *496*(7443), 57–63.

Wright, G. J. (2004) Delta proteins and MAGI proteins: an interaction of Notch ligands with intracellular scaffolding molecules and its significance for zebrafish development. *Development*, *131*(22), 5659–5669.

Zarowiecki, M., & Berriman, M. (2015) What helminth genomes have taught us about parasite evolution. *Parasitology*, *142 Suppl*, S85-97.

Zhang, C. X., Engqvist-Goldstein, Å. E. Y., Carreno, S., Owen, D. J., Smythe, E., & Drubin, D. G. (2005) Multiple Roles for Cyclin G-Associated Kinase in Clathrin-Mediated Sorting Events. *Traffic*, *6*(12), 1103–1113.

Zhang, C., Wang, L., Ali, T., Li, L., Bi, X., Wang, J., Lü, G., Shao, Y., Vuitton, D. A., Wen, H., & Lin, R. (2016) Hydatid cyst fluid promotes peri-cystic fibrosis in cystic echinococcosis by suppressing miR-19 expression. *Parasites & Vectors*, *9*(1), 278.

CURRICULUM VITAE RESUMIDO

PALUDO, GABRIELA PRADO; PALUDO, G.P.

1. DADOS PESSOAIS

Nome:
Gabriela Prado Paludo

Local e Data de Nascimento:
Porto Alegre, Rio Grande do Sul, Brasil, 28/07/1990

Endereço Profissional:
Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia
Avenida Bento Gonçalves, 9500 Prédio 43421 salas 210/223
91501-970 Porto Alegre, RS, Brasil
Telefone: (051) 33087769

E-mail:
gabrielappaludo@gmail.com
gabriela.paludo@ufrgs.br

2. FORMAÇÃO

2015 - Atual

Mestrado em Biologia Celular e Molecular
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Orientador: Henrique Bunselmeyer Ferreira
Co-orientadora: Claudia Elizabeth Thompson
Bolsista da: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

2010 – 2014

Graduação em Biotecnologia (Bioinformática)
Universidade Federal do Rio Grande do Sul, UFRGS, Porto Alegre, Brasil
Bolsista do: Conselho Nacional de Desenvolvimento Científico e Tecnológico

3. ESTÁGIOS

2014 – 2014

Estágio Curricular
Enquadramento Funcional: Estagiário
Carga horária: 20h
Unidade de Biologia Teórica e Computacional (Centro de Biotecnologia/ UFRGS)
Supervisor: Dr. Augusto Schrank e Claudia E. Thompson.

2010 – 2014

Bolsista
Enquadramento Funcional: Estagiário – Iniciação Científica
Carga horária: 20h
Laboratório de Genômica Estrutural e Funcional (Centro de Biotecnologia/UFRGS)
Orientador: Dr. Henrique Bunselmeyer Ferreira

4. PRÊMIOS E DISTINÇÕES

2014 Destaque – Salão de Iniciação Científica UFRGS

5. PROJETOS DE PESQUISA

2011 - 2015

ESTUDO DE ASPECTOS MOLECULARES DA BIOLOGIA DE PLATELMINTOS PARASITAS DA CLASSE CESTODA

NATUREZA: Pesquisa.
ALUNOS ENVOLVIDOS: Graduação: (1) / Mestrado acadêmico: (2) / Doutorado: (2) .
INTEGRANTES: **Gabriela Prado Paludo** - Integrante / Henrique Bunselmeyer Ferreira - Coordenador / Karina Mariante Monteiro - Integrante / Aline Teichmann - Integrante / Caroline Borges Costa - Integrante / Daiani Machado de Vargas - Integrante / Karina Rodrigues Lorenzatto - Integrante / Arnaldo Zaha - Integrante.

2010 - 2013

PROSPECÇÃO E ESTUDOS FUNCIONAIS DE PROTEÍNAS RELEVANTES PARA A RELAÇÃO PARASITO-HOSPEDEIRO NA HIDATIDOSE CÍSTICA E NA HIDATIDOSE ALVEOLAR

DESCRIÇÃO: Projeto de pesquisa aprovado no Edital Pesquisador Gaúcho FAPERGS..
NATUREZA: Pesquisa.
ALUNOS ENVOLVIDOS: Graduação: (1) / Mestrado acadêmico: (1) / Doutorado: (2) .
INTEGRANTES: **Gabriela Prado Paludo** - Integrante / Henrique Bunselmeyer Ferreira - Coordenador / Karina Mariante Monteiro - Integrante / Aline Teichmann - Integrante / Daiani Machado de Vargas - Integrante / Karina Rodrigues Lorenzatto - Integrante / Arnaldo Zaha - Integrante.
FINANCIADOR(ES): Universidade Federal do Rio Grande do Sul.

2010 - 2012

ESTUDO DE PROTEÍNAS POTENCIALMENTE ENVOLVIDAS NA INTERAÇÃO PARASITO-HOSPEDEIRO DURANTE A INFECÇÃO PELO METACESTÓDEO DE ECHINOCOCCUS GRANULOSUS (PLATYHELMINTHES, CESTODA)

DESCRIÇÃO: Projeto financiado através do Edital Universal CNPq.
NATUREZA: Pesquisa.
ALUNOS ENVOLVIDOS: Graduação: (1) / Mestrado acadêmico: (1) / Doutorado: (2) .
INTEGRANTES: **Gabriela Prado Paludo** - Integrante / Henrique Bunselmeyer Ferreira - Integrante / Karina Mariante Monteiro - Integrante / Aline Teichmann - Integrante / Daiani Machado de Vargas - Integrante / Karina Rodrigues Lorenzatto - Integrante / Arnaldo Zaha - Coordenador.
FINANCIADOR(ES): Universidade Federal do Rio Grande do Sul.

6. ARTIGOS COMPLETOS PUBLICADOS

6.1. Lorenzatto, Karina R.; Kim, Kyunggon; Ntail, Ioanna; **Paludo, Gabriela P**. ; Camargo de Lima, Jeferson; Thomas, Paul M. ; Kelleher, Neil L. ; Ferreira, Henrique B.. Top Down Proteomics Reveals Mature Proteoforms Expressed in Subcellular Fractions of the *Echinococcus granulosus* Preadult Stage. Journal of Proteome Research, v. 14, p. 4805–4814, 2015. Citações:1

6.2. **Paludo, Gabriela Prado**; Lorenzatto, Karina Rodrigues; Bonatto, Diego; Ferreira, Henrique Bunselmeyer. Systems biology approach reveals possible evolutionarily conserved moonlighting functions for enolase. Computational Biology and Chemistry, v. 58, p. 1-8, 2015. Citações:5

6.3. Lorenzatto, Karina Rodrigues; Monteiro, Karina Mariante; Paredes, Rodolfo; **Paludo, Gabriela Prado**; da Fonsêca, Marbella Maria; Galanti, Norbel; Zaha, Arnaldo ; Ferreira, Henrique Bunselmeyer. Fructose-bisphosphate aldolase and enolase from Echinococcus granulosus: Genes, expression patterns and protein interactions of two potential moonlighting proteins. Gene, v. 506, p. 76-84, 2012. Citações:16

7. RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS

7.1. **Paludo, Gabriela Prado**; Thompson, Claudia Elizabeth ; Ferreira, Henrique Bunselmeyer. Phylogenomic study of the segmentation process in flatworm species. 2015. (Apresentação de Trabalho/Congresso).

7.2. **Paludo, Gabriela Prado**; Lorenzatto, K. R.; Bonatto, D.; Ferreira, Henrique Bunselmeyer . Investigation of possible moonlighting functions of an Echinococcus granulosus enolase. 2012. (Apresentação de Trabalho/Congresso).

7.3. **Paludo, Gabriela Prado**; Lorenzatto, Karina Rodrigues ; Bonatto, D. ; Ferreira, H. B. . Investigação de possíveis funções moonlighting da enzima glicolítica enolase de Echinococcus granulosus. 2012. (Apresentação de Trabalho/Outra).

7.4. Lorenzatto, K. R.; Paredes, R.; **Paludo, Gabriela Prado**; Monteiro, K. M.; Zaha, A.; Ferreira, H. B.. Eatudo de duas enzimas da via glicolítica de *Echinococcus granulosus* com possíveis funções moonlighting na interação da forma larval com o hospedeiro intermediário. 2011. (Apresentação de Trabalho/Congresso).

7.5. **Paludo, Gabriela Prado**; Lorenzatto, K. R.; Zaha, A.; Ferreira, H. B.. Investigação das funções das proteínas aldolase e enolase de Echinococcus granulosus na interação da forma larval do parasito com o hospedeiro intermediário. 2011. (Apresentação de Trabalho/Outra).

Apêndices

APÊNDICE 1: ALGORITMOS EM LINGUAGEM PYTHON PARA SELEÇÃO DE ORTÓLOGOS 1:1

Os dados utilizados para a análise filogenômica foram filtrados de acordo com os filtros 1 e 2 descritos abaixo:

Filtro 1: Seleciona os arquivos de ortólogos que possuem representantes de todas as espécies do estudo.

- Recebe os arquivos em formato fasta, salvos na pasta Platyhelminthes;
- Salva apenas os arquivos que possuem pelo menos um ortólogo para cada espécie do estudo em uma nova pasta (NewPlatyhelminthes).

```
from numpy import *
import os, sys
from os.path import join as pjoin

def read_FASTA (filename):
        with open (filename) as file:
                return file.read()[0:]

#Os arquivos com as listas de ortólogos estão salvos na pasta Platyhelminthes
file_names = os.listdir('/home/Platyhelminthes')
for f in range (file_names): #lê cada arquivo da pasta Platyhelminthes individualmente
        data = read_FASTA(f)
        data_names = [] #cria uma lista contend os nomes das espécies presentes no arquivo
        for x in (data):
                if (x != ’’): #ignora linhas em branco
                        if (x[0] == ’>’):
                                data_names = data_names + [x]
#Inicia um teste para avaliar se todas as espécies estão presentes na lista criada
        if ‘>Sma’ in data_names:
                if ‘>Sja’ in data_names:
                        if ‘>Csi’ in data_names:
                                if ‘>Egr’ in data_names:

                                        if ‘>Emu’ in data_names:
                                                if ‘>Tso’ in data_names:
                                                        if ‘>Hmi’ in data_names:
                                                                if ‘>Mco’ in data_names:
                                                                        if ‘>Cel’ in data_names:
                                                                                if     ‘>Gpa’     in
data_names:
```

```
                                                                                  if ‘>Hco’ in
data_names:
                                                                                        if
‘>Ovo’ in data_names:

      if ‘>Sra’ in data_names:

            if ‘>Tmu’ in data_names:

                  if ‘>Ovi’ in data_names:

                        if ‘>Sha’ in data_names:

                              if ‘>Hro’ in data_names:

                                    if ‘>Lgi’ in data_names:
#Se o arquivo passer possui todas as espécies, escreve o arquivo na pasta nova
                                          new = raw_input(f)

                                          filepath = '/home/NewPlatyhelminthes'

                                          file = open(filepath, “w”)

                                          file.write(data)

                                          file.close
```

Filtro 2: Garante que cada espécie esteja representada apenas uma vez por arquivo.

- Recebe os arquivos salvos em formato fasta salvos na pasta NewPlatyhelminthes;
- Caso exista mais de uma sequência para uma mesma espécie, remove as sequências de menor tamanho;
- Escreve os arquivos na pasta FinalPlatyhelminthes.

```
from numpy import *
import os, sys
from os.path import join as pjoin

def read_FASTA (filename):
      with open (filename) as file:
            return file.read().split(‘\n’)[0:]

#Função ‘filtro’ recebe uma lista das sequências e o nome da espécie a ser avaliada
#A função excui as sequências repetidas, mandendo apenas a mais longa da espécie ‘name’
def filtro(Data,name):
      seqs = []
      newData = []
      for x in range (len(Data):
            if (Data[x] != name):
                  newData = newData + [Data[x]
      else:
            x = x+1
            seqs = seqs + [Data[x]]
      new = seqs[0]
      for x in range (len(seqs) – 1):
```

```
                if (len(new) < len(seqs[x+1])):
                        new = seqs[x+1]
        newData = newData + [name] + [new]
        return newData


#Os arquivos com as listas de ortólogos estão salvos na pasta NewPlatyhelminthes
file_names = os.listdir('/home/NewPlatyhelminthes')

for f in range (file_names): #lê cada arquivo da pasta Platyhelminthes individualmente
        data = read_FASTA(f)
        data_names = [] #cria uma lista contend os nomes das espécies presentes no arquivo
        for x in (data):
                if (x != ’’): #ignora linhas em branco
                        if (x[0] == ’>’):
                                data_names = data_names + [x]

#Remove as quebras de linha entre as sequências e salva em uma nova lista ‘newData’
        newData = []
        string1 = ‘’
        string2 = ‘’
        for x in range (len(data)):
                if (data[x][:1] == ‘>’):
                        string1 = data[x] #Salva o nome de cada proteína
                        cont = 1
                        string2 = ‘’
                        #Salva a sequência de cada proteína em uma única palavra
                        while ((data[x+cont] != ‘end’) and (data[x+cont] != ‘>’)):
                                string2 = string2 + data[x+cont]
                                cont = cont + 1
                        newData = newData + [string1] + [string2]

#Submete todas as esécies à função ‘filtro’
        if (data_names.cound(‘>Sma’)!= ‘):
                newData = filtro(newData,’>Sma’)
        if (data_names.cound(‘>Sja’)!= ‘):
                newData = filtro(newData,’>Sja’)
        if (data_names.cound(‘>Csi’)!= ‘):
                newData = filtro(newData,’>Csi’)
        if (data_names.cound(‘>Egr’)!= ‘):
                newData = filtro(newData,’>Egr’)
        if (data_names.cound(‘>Emu’)!= ‘):
                newData = filtro(newData,’>Emu’)
        if (data_names.cound(‘>Tso’)!= ‘):
                newData = filtro(newData,’>Tso’)
        if (data_names.cound(‘>Hmi’)!= ‘):
                newData = filtro(newData,’>Hmi’)
        if (data_names.cound(‘>Mco’)!= ‘):
                newData = filtro(newData,’>Mco’)
        if (data_names.cound(‘>Cel’)!= ‘):
                newData = filtro(newData,’>Cel’)
        if (data_names.cound(‘>Gpa’)!= ‘):
                newData = filtro(newData,’>Gpa’)
        if (data_names.cound(‘>Hco’)!= ‘):
                newData = filtro(newData,’>Hco’)
        if (data_names.cound(‘>Ovo’)!= ‘):
                newData = filtro(newData,’>Ovo’)
        if (data_names.cound(‘>Sra’)!= ‘):
                newData = filtro(newData,’>Sra’)
        if (data_names.cound(‘>Tmu’)!= ‘):
                newData = filtro(newData,’>Tmu’)
        if (data_names.cound(‘>Ovi’)!= ‘):
                newData = filtro(newData,’>Ovi’)
        if (data_names.cound(‘>Sha’)!= ‘):
                newData = filtro(newData,’>Sha’)
        if (data_names.cound(‘>Hro’)!= ‘):
                newData = filtro(newData,’>Hro’)
        if (data_names.cound(‘>Lgi’)!= ‘):
```

```
               newData = filtro(newData,’>Lgi’)

#Escreve o arquivo em uma pasta nova
       new = raw_input(f)
       filepath = '/home/FinalPlatyhelminthes'
       file = open(filepath, “w”)
       file.write(data)
       file.close
```

APÊNDICE 2: ALGORITMOS EM LINGUAGEM PYTHON PARA IDENTIFICAÇÃO DE ORTÓLOGOS CONSERVADAS EM CESTÓDEOS

Os dados utilizados para a seleção dos grupos de ortólogos compartilhadas entre todas as espécies de cestódeos estudadas e ausentes em, pelo menos, uma das espécies de trematódeos estudadas, foram filtrados de acordo com o filtros 3 descrito abaixo:

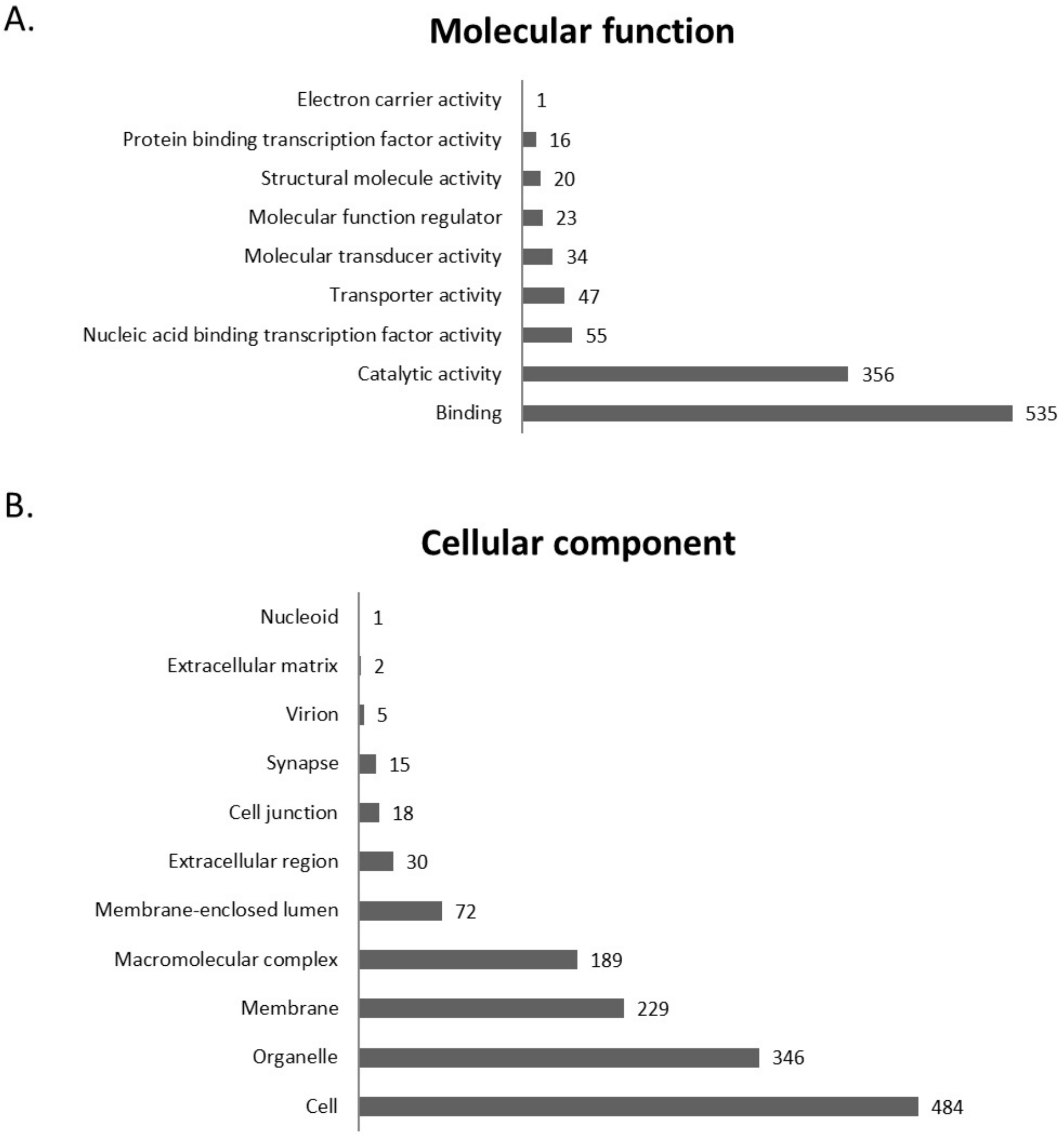
Filtro 3: Seleciona os arquivos de ortólogos que possuem representantes de todas as espécies de cestódeos mas não em todas as espécies de trematódeos.

- Recebe os arquivos em formato fasta, salvos na pasta Platyhelminthes;
- Salva em uma noma pasta (AllCestodes) apenas os arquivos que passarem pela análise.

```
from numpy import *
import os, sys
from os.path import join as pjoin

def read_FASTA (filename):
        with open (filename) as file:
                return file.read()[0:]
#A função ‘TremTest’ retorna o resultado lógico ‘True’ caso alguma espécie de trematódeo
esteja ausente na lista ‘names’; e retorna o resultado lógico ‘False’ caso todas as espécies
de trematódeos estejam presentes na lista ‘names.
def TremTest(names)
        resp = True
        if ‘>Csi’ in names:
                if ‘>Ovi’ in names:
                        if ‘>Sha’ in names:
                                if ‘>Sma’ in names:
                                        if ‘>Sja’ in names:
                                                resp = False
        return resp


#Os arquivos com as listas de ortólogos estão salvos na pasta Platyhelminthes
file_names = os.listdir('/home/Platyhelminthes')
for f in range (file_names): #lê cada arquivo da pasta Platyhelminthes individualmente
        data = read_FASTA(f)
        data_names = [] #cria uma lista contend os nomes das espécies presentes no arquivo
        for x in (data):
                if (x != ’’): #ignora linhas em branco
                        if (x[0] == ’>’):
                                data_names = data_names + [x]
#Inicia um teste para avaliar se todas as espécies de cestódeos estão presentes na lista
criada
```

```
        if ‘>Egr’ in data_names:
                if ‘>Emu’ in data_names:
                        if ‘>Tso’ in data_names:
                                if ‘>Hmi’ in data_names:
                                        if ‘>Mco’ in data_names:
#Inicia o teste para selecionar arquivos que não possuem alguma espécie de trematódeo
                                                if (TremTest(data_names)):
#Se o arquivo passar pelos critérios, escreve o arquivo na pasta nova
                                                        new = raw_input(f)
                                                        filepath = '/home/AllCestodes'
                                                        file = open(filepath, “w”)
                                                                        file.write(data)
                                                        file.close
```

APÊNDICE 3: SUPPLEMENTARY FILE 1



Supplementary File 1. Functional enrichment of orthologous groups present in all tapeworms and absent in at last one fluke. (A) Molecular function and (B) cellular component related to the 910 orthologus groups selected.

APÊNDICE 4: DIAGNÓSTICOS DE CONVERGÊNCIA DO MRBAYES

Apêndice 4.1: Phylogenomic analysis

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -4112821.41
      |                            1    2                     2    |
      |                                               2            |
      |                              1 2 2    2                    |
      |    2        1    1 211  1           1    1  11     11      |
      |1           1      2       1    1  2             1 12  1111 |
      |    12           1  1  1  1  1      1         2   12        |
      |2              1            2  2 1   21 11       2   2     1|
      |  2   *    2      2  2 2  2       1     2   *   2     1     |
      | 212    21 12   1  1    1      1   12 2  2     11     2 2 22|
      |     1                                 1   2                |
      |        121                2  2                   2         |
      | 1        2     22      2                 2                 |
      |       1       2         2   2             1 2           2  |
      |   1   2      1       2                                     |
      |             22                                             |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -4112824.80
      ^                                                            ^
      422000                                                       1688000

         Model parameter summaries over the runs sampled in files
         "phylogenomic.nxs.run1.p" and "phylogenomic.nxs.run2.p":
         Summaries are based on a total of 25322 samples from 2 runs.
         Each run produced 16881 samples of which 12661 samples were included.
         Parameter summaries saved to file "phylogenomic.nxs.pstat".
         Appending to file "phylogenomic.nxs.pstat"
```

```
                                     95% HPD Interval
                                   --------------------
   Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
   --------------------------------------------------------------------------------------------------
   TL         10.761010    0.001378   10.687080   10.832960   10.761260   5665.42   6030.09    1.000
   alpha       0.912277    0.000014    0.905040    0.919761    0.912272   3091.38   3343.87    1.000
   pinvar      0.000004    0.000000    0.000000    0.000010    0.000003     31.37     63.79    1.009
   --------------------------------------------------------------------------------------------------
    * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
      correspond to minimal and average ESS among runs.
      ESS value below 100 may indicate that the parameter is undersampled.
    + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
      and Rubin, 1992) should approach 1.0 as runs converge.
```

```
      Summary statistics for informative taxon bipartitions
      (saved to file "phylogenomic.nxs.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   19  25322    1.000000    0.000000    1.000000    1.000000    2
   20  25322    1.000000    0.000000    1.000000    1.000000    2
```

```
21  25322    1.000000    0.000000    1.000000    1.000000    2
22  25322    1.000000    0.000000    1.000000    1.000000    2
23  25322    1.000000    0.000000    1.000000    1.000000    2
24  25322    1.000000    0.000000    1.000000    1.000000    2
25  25322    1.000000    0.000000    1.000000    1.000000    2
26  25322    1.000000    0.000000    1.000000    1.000000    2
27  25322    1.000000    0.000000    1.000000    1.000000    2
28  25322    1.000000    0.000000    1.000000    1.000000    2
29  25322    1.000000    0.000000    1.000000    1.000000    2
30  25322    1.000000    0.000000    1.000000    1.000000    2
31  25322    1.000000    0.000000    1.000000    1.000000    2
32  25322    1.000000    0.000000    1.000000    1.000000    2
33  25322    1.000000    0.000000    1.000000    1.000000    2
------------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "phylogenomic.nxs.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
---------------------------------------------------------------------------------------
length[1]     0.529785    0.000019    0.521169    0.538016    0.529743    1.000    2
length[2]     0.071937    0.000001    0.069811    0.074100    0.071942    1.000    2
length[3]     0.012429    0.000000    0.011736    0.013103    0.012426    1.000    2
length[4]     0.019624    0.000000    0.018786    0.020439    0.019624    1.000    2
length[5]     0.860489    0.000046    0.847438    0.873866    0.860413    1.000    2
length[6]     0.370554    0.000014    0.363212    0.377855    0.370548    1.000    2
length[7]     0.369596    0.000007    0.364403    0.375024    0.369591    1.000    2
length[8]     0.716282    0.000033    0.704759    0.727570    0.716235    1.000    2
length[9]     0.514257    0.000025    0.504751    0.524170    0.514265    1.000    2
length[10]    0.243140    0.000007    0.238160    0.248400    0.243123    1.000    2
length[11]    0.057003    0.000001    0.055064    0.059126    0.056991    1.000    2
length[12]    0.562559    0.000021    0.553369    0.571065    0.562475    1.000    2
length[13]    0.051779    0.000001    0.050080    0.053469    0.051780    1.000    2
length[14]    0.191591    0.000005    0.187495    0.195910    0.191582    1.000    2
length[15]    0.054513    0.000001    0.052833    0.056249    0.054503    1.000    2
length[16]    0.957802    0.000048    0.944437    0.971592    0.957790    1.000    2
length[17]    1.022474    0.000056    1.008252    1.037485    1.022504    1.000    2
length[18]    0.086159    0.000001    0.083940    0.088330    0.086159    1.000    2
length[19]    0.380582    0.000025    0.371297    0.390753    0.380538    1.000    2
length[20]    0.303819    0.000023    0.294361    0.313103    0.303778    1.001    2
length[21]    0.532706    0.000026    0.522657    0.542550    0.532713    1.000    2
length[22]    0.074321    0.000001    0.072330    0.076485    0.074319    1.000    2
length[23]    0.097737    0.000005    0.093632    0.102175    0.097735    1.000    2
length[24]    0.133004    0.000003    0.129475    0.136570    0.133007    1.000    2
length[25]    0.079849    0.000007    0.074391    0.084858    0.079850    1.000    2
length[26]    0.383366    0.000015    0.375833    0.390995    0.383355    1.000    2
length[27]    0.082716    0.000003    0.079294    0.086133    0.082714    1.000    2
length[28]    0.736419    0.000044    0.723158    0.749156    0.736386    1.000    2
length[29]    0.283996    0.000018    0.275491    0.292160    0.283967    1.000    2
length[30]    0.439523    0.000015    0.431968    0.447177    0.439522    1.000    2
length[31]    0.271989    0.000013    0.264842    0.278954    0.271933    1.000    2
length[32]    0.175791    0.000017    0.167655    0.183680    0.175798    1.000    2
length[33]    0.093219    0.000011    0.086452    0.099704    0.093215    1.000    2
---------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
```

```
Average PSRF for parameter values (excluding NA and >10.0) = 1.000
Maximum PSRF for parameter values = 1.001
```

Apêndice 4.2: Bone morphogenetic protein 2 – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

+------------------------------------------------------------+ -3190.98
|                                2            1              |
|                 2                                2         |
|                           1               1            1   |
| 2 2             1   2                   2  1      1     11 |
|  11  1       1         1       1  2 2 21   2   *       2 2 |
|* 2      11 2   *  21   221    2 21   2  1*2  11 1  2    2 2|
|      2 * 22 1    *   21 1       12                  1      |
| 1  *  2 2  1  *       2   2*12                2   2  21    |
|           1       1211   2        121 1         21    2   1|
|     1 1     22               11      1       2      2      |
|     2                              1        2      1 1     |
|                             2                              |
|                                                            |
|                                                            |
|                                        2                   |
+------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -3199.76
      ^                                                            ^
      13200                                                        53000

 Model parameter summaries over the runs sampled in files
         "BMP2_CDS.nexus.run1.p" and "BMP2_CDS.nexus.run2.p":
         Summaries are based on a total of 798 samples from 2 runs.
         Each run produced 531 samples of which 399 samples were included.
         Parameter summaries saved to file "BMP2_CDS.nexus.pstat".

                                95% HPD Interval
                              --------------------
   Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
   ------------------------------------------------------------------------------------------------
   TL          4.239867    0.299277    3.342848    5.371899    4.168797    153.31    195.99    1.001
   kappa       3.601397    0.166024    2.874864    4.485213    3.559607    135.82    190.53    1.000
   pi(A)       0.320110    0.000205    0.289098    0.344591    0.319837    146.05    159.93    1.000
   pi(C)       0.263238    0.000154    0.237431    0.286109    0.263104    150.61    160.12    1.001
   pi(G)       0.202294    0.000109    0.179901    0.219737    0.202744    107.88    125.06    0.999
   pi(T)       0.214359    0.000139    0.190648    0.237147    0.214527    117.05    130.44    0.999
   alpha       1.375789    0.236210    0.645351    2.305631    1.274457     54.45     68.22    0.999
   pinvar      0.123593    0.002312    0.022641    0.208701    0.127764     61.62     82.78    1.001
   ------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.

      Summary statistics for informative taxon bipartitions
      (saved to file "BMP2_CDS.nexus.tstat"):

   ID   #obs    Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   10   798    1.000000    0.000000    1.000000    1.000000    2
   11   798    1.000000    0.000000    1.000000    1.000000    2
   12   784    0.982456    0.003544    0.979950    0.984962    2
   13   736    0.922306    0.000000    0.922306    0.922306    2
   14   549    0.687970    0.001772    0.686717    0.689223    2
```

```
15   439    0.550125    0.040761    0.521303    0.578947    2
16   181    0.226817    0.019494    0.213033    0.240602    2
17   156    0.195489    0.000000    0.195489    0.195489    2
18    88    0.110276    0.010633    0.102757    0.117794    2
----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "BMP2_CDS.nexus.vstat"):

                                   95% HPD Interval
                                 --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
---------------------------------------------------------------------------------------
length[1]     0.005408    0.000016    0.000020    0.012722    0.004664    1.006    2
length[2]     0.008257    0.000023    0.000866    0.018162    0.007526    1.006    2
length[3]     0.078193    0.000607    0.038633    0.128020    0.076206    0.999    2
length[4]     0.316206    0.004809    0.195553    0.470278    0.313447    1.000    2
length[5]     0.264895    0.013720    0.035537    0.456838    0.259528    0.999    2
length[6]     0.310602    0.005645    0.170469    0.467351    0.309839    0.999    2
length[7]     0.133041    0.002455    0.034378    0.229634    0.131237    1.000    2
length[8]     0.317745    0.018061    0.084547    0.599355    0.296801    0.999    2
length[9]     0.537814    0.029001    0.212985    0.878916    0.519294    1.000    2
length[10]    0.767433    0.049287    0.377112    1.230906    0.743968    1.004    2
length[11]    0.965846    0.067840    0.479025    1.451521    0.943814    0.999    2
length[12]    0.043274    0.000349    0.007715    0.077459    0.041747    0.999    2
length[13]    0.286880    0.012779    0.047216    0.498171    0.281911    1.005    2
length[14]    0.199615    0.008444    0.046516    0.396308    0.188600    0.998    2
length[15]    0.065485    0.001173    0.000630    0.124679    0.063879    1.004    2
length[16]    0.040732    0.000458    0.001979    0.078246    0.038954    1.000    2
length[17]    0.072683    0.001141    0.004197    0.126769    0.072174    0.994    2
length[18]    0.108227    0.005200    0.000377    0.259137    0.090608    0.989    2
---------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.008467
    Maximum standard deviation of split frequencies = 0.040761
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.006
```

Apêndice 4.3: Bone morphogenetic protein 2 – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

   +------------------------------------------------------------+ -1697.71
   |  2                                         1    2          |
   |   1            1              2         2          2       |
   |                   2                            2           |
   |               2            2                    1    2 1  2|
   |     1 1 1*        1          1 1     122 2    1            |
   |        2    21  22  22             2 2    2             121|
   | *       2    2   1  11 1 1   2   1 11  1  1  22       2221 |
   |1     1 1  11  12   2  2  2  2 12    2    1     1 221 1     |
   |   22 22               1     1         1           1        |
   |    1        1           1 21    2 1         *         1    |
   |  1  2     22           22               1           1      |
   |                 1  1             22        2     1  2      |
   |2                                             1             |
   |                                 1                          |
   |                           1                                |
   +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -1698.75
   ^                                                            ^
   625000                                                       2500000


      Model parameter summaries over the runs sampled in files
       "BMP2_Prot.nexus.run1.p" and "BMP2_Prot.nexus.run2.p":
       Summaries are based on a total of 37502 samples from 2 runs.
       Each run produced 25001 samples of of which 18751 samples were included.
       Parameter summaries saved to file "BMP2_Prot.nexus.pstat".

                                       95% HPD Interval
                                     --------------------
   Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
   --------------------------------------------------------------------------------------------------
   TL         2.819544    0.086833    2.280275    3.423562    2.799334  16144.26  16379.99    1.000
   alpha      1.367554    0.115483    0.814122    2.066669    1.315613  11685.67  11970.94    1.000
   --------------------------------------------------------------------------------------------------
    * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
      correspond to minimal and average ESS among runs.
      ESS value below 100 may indicate that the parameter is undersampled.
    + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
      and Rubin, 1992) should approach 1.0 as runs converge.


    Summary statistics for informative taxon bipartitions
    (saved to file "BMP2_Prot.nexus.tstat"):

 ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
 -----------------------------------------------------------------
 10  37502    1.000000    0.000000    1.000000    1.000000    2
 11  37502    1.000000    0.000000    1.000000    1.000000    2
 12  37402    0.997333    0.000453    0.997013    0.997653    2
 13  37336    0.995574    0.000453    0.995254    0.995894    2
 14  37138    0.990294    0.001207    0.989441    0.991147    2
 15  36563    0.974961    0.000641    0.974508    0.975415    2
 -----------------------------------------------------------------
 + Convergence diagnostic (standard deviation of split frequencies)
```

```
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "BMP2_Prot.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.014168    0.000104    0.000159    0.033995    0.011868    1.000    2
length[2]     0.019327    0.000141    0.000933    0.042573    0.017037    1.000    2
length[3]     0.093678    0.000890    0.039076    0.152502    0.090542    1.000    2
length[4]     0.191060    0.002149    0.102545    0.280694    0.187143    1.000    2
length[5]     0.075936    0.001963    0.000057    0.159197    0.069147    1.000    2
length[6]     0.171723    0.002491    0.079915    0.271818    0.167392    1.000    2
length[7]     0.119883    0.001847    0.039489    0.204520    0.115899    1.000    2
length[8]     0.250056    0.006070    0.108585    0.407777    0.243523    1.000    2
length[9]     0.440742    0.012312    0.237930    0.663902    0.430918    1.000    2
length[10]    0.636718    0.018734    0.389292    0.913663    0.624539    1.000    2
length[11]    0.397990    0.011958    0.193931    0.613944    0.388573    1.000    2
length[12]    0.046481    0.000482    0.008983    0.090445    0.043557    1.000    2
length[13]    0.078310    0.000899    0.023615    0.137542    0.075065    1.000    2
length[14]    0.169060    0.004714    0.041914    0.303744    0.162681    1.000    2
length[15]    0.117432    0.002212    0.028654    0.209993    0.113695    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000459
    Maximum standard deviation of split frequencies = 0.001207
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.4: Cyclin-g-associated kinase – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -29479.85
      |                    1                                       |
      |                                                            |
      |                                            2               |
      |                               1                            |
      |                    2                                       |
      |                   2  212   *                   2   1       |
      |1    1  1 12                               1             2  |
      | 2   2     1  1*          2         222     1   1      1    |
      |  212  2 1   22 1  1 2  1  2 11  1 2 1  *           2*  2   |
      | 1    *  2  1    12  11       2222       *     2  *1   2  12|
      |2 12   1    2    2       2   2    2    2       1      2  12 |
      |    1     2  1            1        11 11   2 1   2 2  1 1  1|
      |        2       2 1    2 1        1           1  1          |
      |                           1              *  22             |
      |                                1                           |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -29482.09
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "GAK_CDS.nexus.run1.p" and "GAK_CDS.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "GAK_CDS.nexus.pstat".

                                         95% HPD Interval
                                       --------------------
   Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
   ------------------------------------------------------------------------------------------------------
   TL          9.720695    0.310271    8.668933   10.828200    9.687636   6498.10   6920.64    1.000
   kappa       3.125313    0.014867    2.886397    3.360811    3.123325   8342.36   8383.94    1.000
   pi(A)       0.293580    0.000018    0.285325    0.301989    0.293546   4948.99   4988.17    1.000
   pi(C)       0.239993    0.000015    0.232320    0.247599    0.239977   5080.18   5349.12    1.000
   pi(G)       0.211457    0.000014    0.204042    0.218493    0.211460   5663.32   5745.95    1.000
   pi(T)       0.254970    0.000016    0.247297    0.262972    0.254875   4825.70   5243.18    1.000
   alpha       0.971393    0.007043    0.802124    1.131295    0.971179   2855.48   2925.62    1.000
   pinvar      0.060780    0.000239    0.029658    0.089765    0.062099   2996.12   3003.73    1.000
   ------------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "GAK_CDS.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   15  37502    1.000000    0.000000    1.000000    1.000000    2
   16  37502    1.000000    0.000000    1.000000    1.000000    2
   17  37502    1.000000    0.000000    1.000000    1.000000    2
   18  37502    1.000000    0.000000    1.000000    1.000000    2
   19  37502    1.000000    0.000000    1.000000    1.000000    2
```

```
20  37502    1.000000    0.000000    1.000000    1.000000    2
21  37502    1.000000    0.000000    1.000000    1.000000    2
22  37485    0.999547    0.000490    0.999200    0.999893    2
23  37457    0.998800    0.000113    0.998720    0.998880    2
24  37410    0.997547    0.001810    0.996267    0.998827    2
25  37403    0.997360    0.001999    0.995947    0.998773    2
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "GAK_CDS.nexus.vstat"):

                                      95% HPD Interval
                                    --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]     0.743171    0.003754    0.625652    0.864375    0.740204    1.000    2
length[2]     0.346548    0.001205    0.280222    0.415978    0.345392    1.000    2
length[3]     0.419523    0.001413    0.348059    0.494150    0.418185    1.000    2
length[4]     0.023478    0.000042    0.010882    0.036097    0.023247    1.000    2
length[5]     0.022176    0.000041    0.009968    0.035054    0.022074    1.000    2
length[6]     0.236881    0.000976    0.174957    0.297135    0.236112    1.000    2
length[7]     0.085452    0.000095    0.066719    0.104874    0.085118    1.000    2
length[8]     0.007381    0.000004    0.003456    0.011404    0.007205    1.000    2
length[9]     0.010822    0.000006    0.006428    0.015708    0.010644    1.000    2
length[10]    0.186178    0.000486    0.144457    0.230657    0.185673    1.000    2
length[11]    0.283368    0.000411    0.244321    0.323136    0.282553    1.000    2
length[12]    1.156920    0.011804    0.955243    1.377736    1.151876    1.000    2
length[13]    2.750349    0.123172    2.115290    3.464807    2.722922    1.000    2
length[14]    0.748166    0.027843    0.420106    1.065958    0.742550    1.000    2
length[15]    0.124524    0.000219    0.095393    0.153012    0.124045    1.000    2
length[16]    0.104581    0.000406    0.065125    0.143618    0.103978    1.000    2
length[17]    0.324744    0.001418    0.251785    0.398480    0.323754    1.000    2
length[18]    0.155827    0.001213    0.090067    0.226647    0.154681    1.000    2
length[19]    0.053494    0.000074    0.037220    0.070645    0.053172    1.000    2
length[20]    0.409263    0.001311    0.338894    0.481104    0.408116    1.000    2
length[21]    0.237241    0.001237    0.170499    0.307533    0.235829    1.000    2
length[22]    0.400730    0.002652    0.303627    0.501521    0.398988    1.000    2
length[23]    0.108004    0.001290    0.039806    0.178526    0.105944    1.000    2
length[24]    0.217828    0.004145    0.095228    0.346590    0.215630    1.000    2
length[25]    0.565263    0.026457    0.252412    0.887646    0.560852    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000401
    Maximum standard deviation of split frequencies = 0.001999
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.5: Cyclin-g-associated kinase - Proteína

Below are rough plots of the generation (x-axis) versus the log probability of observing the data (y-axis). You can use these graphs to determine what the burn in for your analysis should be. When the log probability starts to plateau you may be at stationarity. Sample trees and parameters after the log probability plateaus. Of course, this is not a guarantee that you are at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt' commands for all the parameters in your model. Remember that the burn in is the number of samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

```
      +------------------------------------------------------------+ -15507.95
      |    1    1                             2                  2 |
      |                                          1                 |
      | 1        1            2             2                 1    |
      |    2 2    1    2* 1  *     2      2     1   2              |
      |2        2 2 2    2  2 1      1   2         2   1    2   1 2|
      | 2 2      2 212     2               1        1 1   2  1 2   |
      |  2         1 1    21      1 12 1*11211    1  2  22 * 2     |
      |     1  2      1     1  * 2  2 12       122   1   1     1  1|
      |  1                      1     2       1                    |
      |1    2 21         1       1           2    21  2   1 1      |
      |      1        21        2 2            2        1       21 |
      |       1                    1                               |
      |   1                                                        |
      |                                                       2    |
      |                                                2           |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -15509.44
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
          "GAK_Prot.nexus.run1.p" and "GAK_Prot.nexus.run2.p":
          Summaries are based on a total of 37502 samples from 2 runs.
          Each run produced 25001 samples of which 18751 samples were included.
          Parameter summaries saved to file "GAK_Prot.nexus.pstat".

                                        95% HPD Interval
                                      --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ---------------------------------------------------------------------------------------------------
     TL          6.809590    0.094227    6.226508    7.420869    6.799374  10450.16  10570.75    1.000
     alpha       1.408259    0.011435    1.206091    1.622193    1.402493  10885.23  11071.78    1.000
     pinvar      0.000858    0.000001    0.000000    0.002575    0.000596   8013.62   8277.45    1.000
     ---------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "GAK_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   15  37502    1.000000    0.000000    1.000000    1.000000    2
   16  37502    1.000000    0.000000    1.000000    1.000000    2
   17  37502    1.000000    0.000000    1.000000    1.000000    2
   18  37502    1.000000    0.000000    1.000000    1.000000    2
   19  37502    1.000000    0.000000    1.000000    1.000000    2
   20  37502    1.000000    0.000000    1.000000    1.000000    2
   21  37502    1.000000    0.000000    1.000000    1.000000    2
   22  37502    1.000000    0.000000    1.000000    1.000000    2
   23  37502    1.000000    0.000000    1.000000    1.000000    2
```

```
24  37502    1.000000    0.000000    1.000000    1.000000    2
25  36303    0.968028    0.000490    0.967682    0.968375    2
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "GAK_Prot.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.498910    0.001981    0.415828    0.589505    0.497327    1.000    2
length[2]     0.243032    0.000752    0.189724    0.296440    0.242127    1.000    2
length[3]     0.329918    0.001003    0.268771    0.391949    0.328845    1.000    2
length[4]     0.014700    0.000027    0.005150    0.025113    0.014232    1.000    2
length[5]     0.017436    0.000030    0.007544    0.028736    0.017009    1.000    2
length[6]     0.181520    0.000644    0.132489    0.231806    0.180423    1.000    2
length[7]     0.058114    0.000101    0.038152    0.077229    0.057556    1.000    2
length[8]     0.005937    0.000007    0.001497    0.011401    0.005559    1.000    2
length[9]     0.004873    0.000006    0.000867    0.009749    0.004489    1.000    2
length[10]    0.152593    0.000281    0.120384    0.185533    0.151922    1.000    2
length[11]    0.133640    0.000356    0.096742    0.170643    0.132839    1.000    2
length[12]    0.816222    0.005135    0.672342    0.951940    0.813967    1.000    2
length[13]    1.782829    0.035616    1.430279    2.162368    1.773713    1.000    2
length[14]    0.408553    0.010093    0.216650    0.608914    0.405827    1.000    2
length[15]    0.065960    0.000250    0.036033    0.097236    0.065066    1.000    2
length[16]    0.112526    0.000678    0.063744    0.165026    0.111506    1.000    2
length[17]    0.208048    0.000655    0.159215    0.259194    0.207303    1.000    2
length[18]    0.034667    0.000064    0.019701    0.050656    0.034135    1.000    2
length[19]    0.210665    0.000933    0.151870    0.271085    0.209468    1.000    2
length[20]    0.054622    0.000119    0.033892    0.076511    0.054015    1.000    2
length[21]    0.595077    0.011608    0.389100    0.811753    0.591507    1.000    2
length[22]    0.162597    0.002210    0.071888    0.254219    0.160512    1.000    2
length[23]    0.276247    0.001033    0.214882    0.340646    0.275253    1.000    2
length[24]    0.365205    0.001841    0.281357    0.449089    0.363674    1.000    2
length[25]    0.076845    0.000804    0.024310    0.134362    0.075232    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000045
    Maximum standard deviation of split frequencies = 0.000490
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.6: Groucho protein – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -17869.17
      |                            1                               |
      |                                                            |
      |                               1       1        1           |
      |                                     1                   *  |
      | *             1  1                                       1 |
      |2           2 2 1  1              1       1 1        1     *|
      |       1          2   211    1 2 1     2              22    |
      |  1     1 2 1    2 2       2     2       1    2* 1  *       |
      |1  11 2 2     1     12     1  2    2    12  21  2 2         |
      |    22     2   2      1   1 2     2 1   2    2   2 2 2  2 2 |
      |   2 112  1  2      2  2 12  2  1  1  2   21      1   1 1   |
      |         * 1 1  2       2           2         1             |
      |  2              1   1   2      2    21    2       1   1    |
      |                                                            |
      |                              1                             |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -17870.79
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "Groucho_CDS.nexus.run1.p" and " Groucho_CDS.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "Groucho_CDS.nexus.pstat".

                                        95% HPD Interval
                                      --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ---------------------------------------------------------------------------------------------------
     TL         6.785397    0.231347    5.888357    7.755557    6.760788    6946.98   7249.20    1.000
     kappa      3.455907    0.032112    3.106412    3.808353    3.451372    7544.32   8081.89    1.000
     pi(A)      0.254634    0.000026    0.244172    0.264435    0.254644    5550.58   5982.27    1.000
     pi(C)      0.270589    0.000027    0.260388    0.280866    0.270601    5998.63   6166.50    1.000
     pi(G)      0.215172    0.000022    0.206079    0.224685    0.215197    6689.32   6730.41    1.000
     pi(T)      0.259605    0.000026    0.249630    0.269581    0.259662    5732.10   6046.39    1.000
     alpha      0.492797    0.000666    0.443385    0.543214    0.492116    6413.76   6635.29    1.000
     ---------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "Groucho_CDS.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   14  37502    1.000000    0.000000    1.000000    1.000000    2
   15  37502    1.000000    0.000000    1.000000    1.000000    2
   16  37502    1.000000    0.000000    1.000000    1.000000    2
   17  37502    1.000000    0.000000    1.000000    1.000000    2
   18  37502    1.000000    0.000000    1.000000    1.000000    2
```

```
19  37502    1.000000    0.000000    1.000000    1.000000    2
20  37502    1.000000    0.000000    1.000000    1.000000    2
21  37502    1.000000    0.000000    1.000000    1.000000    2
22  37499    0.999920    0.000038    0.999893    0.999947    2
23  34340    0.915684    0.004978    0.912165    0.919204    2
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "Groucho_CDS.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     1.199699    0.036068    0.856729    1.587131    1.184724    1.000    2
length[2]     0.327854    0.001640    0.252309    0.410157    0.327476    1.000    2
length[3]     0.290900    0.001391    0.221125    0.366779    0.289574    1.000    2
length[4]     0.300763    0.003068    0.198240    0.413763    0.297103    1.000    2
length[5]     0.004765    0.000004    0.001031    0.008846    0.004549    1.000    2
length[6]     0.006557    0.000005    0.002352    0.011027    0.006364    1.000    2
length[7]     0.091447    0.000172    0.066064    0.117414    0.090954    1.000    2
length[8]     0.124528    0.000815    0.069470    0.180807    0.123668    1.000    2
length[9]     0.453548    0.001533    0.379210    0.531060    0.451659    1.000    2
length[10]    0.023996    0.000066    0.007528    0.038954    0.024529    1.000    2
length[11]    0.011471    0.000059    0.000002    0.026035    0.010251    1.000    2
length[12]    0.029081    0.000075    0.012762    0.046097    0.028637    1.000    2
length[13]    0.024182    0.000072    0.008215    0.041044    0.023947    1.000    2
length[14]    0.100864    0.000455    0.058449    0.142063    0.100122    1.000    2
length[15]    0.055601    0.000133    0.033227    0.078290    0.055249    1.000    2
length[16]    0.541497    0.015945    0.308814    0.801069    0.534713    1.000    2
length[17]    0.171484    0.001014    0.109684    0.235040    0.170346    1.000    2
length[18]    1.388428    0.039076    1.009994    1.775426    1.375917    1.000    2
length[19]    0.455598    0.003069    0.348974    0.565912    0.453233    1.000    2
length[20]    0.390508    0.006984    0.230402    0.556639    0.386977    1.001    2
length[21]    0.402628    0.002776    0.301622    0.506794    0.400696    1.000    2
length[22]    0.301276    0.006102    0.151509    0.455780    0.298215    1.000    2
length[23]    0.091892    0.001665    0.016965    0.173317    0.088956    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000502
    Maximum standard deviation of split frequencies = 0.004978
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.001
```

Apêndice 4.7: Groucho protein – Proteína

Below are rough plots of the generation (x-axis) versus the log probability of observing the data (y-axis). You can use these graphs to determine what the burn in for your analysis should be. When the log probability starts to plateau you may be at stationarity. Sample trees and parameters after the log probability plateaus. Of course, this is not a guarantee that you are at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt' commands for all the parameters in your model. Remember that the burn in is the number of samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

```
      +------------------------------------------------------------+ -8062.66
      |                                     2      1   1           |
      |                                                            |
      |    2      2  2        2                                  1 |
      | 11  2          2                 2              1          |
      |   111      22       12  2                             1   1|
      |      112       1   1         2*   11    2    1  2  221 1   |
      |       2  21      212        2  1      11 2    2   1  2     |
      |2         1 1 1  21         1 1       12 1  2  1     1 2  2 |
      |             1       2   1       212  2   1  2  2       22 2|
      |1 22  2 12       1      * 1 21   1  2      * 1    22        |
      | 2                    1    1         1  2     2          1  |
      |         1     1   2   1  2                         1       |
      |               2                2                 1         |
      |                                                            |
      |                           2                                |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -8064.16
      ^                                                            ^
      625000                                                       2500000
```

```
         Model parameter summaries over the runs sampled in files
         "Groucho_Prot.nexus.run1.p" and Groucho_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file Groucho_Prot.nexus.pstat".

                                         95% HPD Interval
                                       --------------------
   Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
   ----------------------------------------------------------------------------------------------------
   TL          2.939532    0.032488    2.597479    3.299828    2.931159    9764.90  10050.17    1.000
   alpha       1.068708    0.013614    0.856232    1.309533    1.060679    8928.86   9281.25    1.000
   ----------------------------------------------------------------------------------------------------
   * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
     correspond to minimal and average ESS among runs.
     ESS value below 100 may indicate that the parameter is undersampled.
   + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
     and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "Groucho_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   14  37502    1.000000    0.000000    1.000000    1.000000    2
   15  37502    1.000000    0.000000    1.000000    1.000000    2
   16  37502    1.000000    0.000000    1.000000    1.000000    2
   17  37502    1.000000    0.000000    1.000000    1.000000    2
   18  37502    1.000000    0.000000    1.000000    1.000000    2
   19  37502    1.000000    0.000000    1.000000    1.000000    2
   20  37497    0.999867    0.000038    0.999840    0.999893    2
   21  37483    0.999493    0.000113    0.999413    0.999573    2
   22  37351    0.995974    0.000038    0.995947    0.996000    2
   23  34743    0.926431    0.002149    0.924911    0.927951    2
```

```
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file “Groucho_Prot.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.666934    0.008940    0.485448    0.851193    0.661393    1.000    2
length[2]     0.079023    0.000253    0.049221    0.110863    0.078194    1.000    2
length[3]     0.104092    0.000314    0.069906    0.138679    0.103058    1.000    2
length[4]     0.091702    0.000459    0.052479    0.135558    0.090236    1.000    2
length[5]     0.003240    0.000005    0.000080    0.007518    0.002760    1.000    2
length[6]     0.002213    0.000003    0.000000    0.005862    0.001726    1.000    2
length[7]     0.037029    0.000080    0.020389    0.054847    0.036376    1.000    2
length[8]     0.083640    0.000252    0.053521    0.115245    0.082845    1.000    2
length[9]     0.221644    0.000533    0.177129    0.267107    0.220755    1.000    2
length[10]    0.004956    0.000009    0.000013    0.010618    0.004454    1.000    2
length[11]    0.002568    0.000006    0.000000    0.007279    0.001897    1.000    2
length[12]    0.019627    0.000042    0.008211    0.032684    0.018978    1.000    2
length[13]    0.009632    0.000025    0.001008    0.019370    0.008955    1.000    2
length[14]    0.153951    0.001026    0.090559    0.215648    0.152498    1.000    2
length[15]    0.172241    0.000562    0.126519    0.218647    0.171317    1.000    2
length[16]    0.140878    0.000969    0.080852    0.202516    0.139743    1.000    2
length[17]    0.190072    0.000598    0.143381    0.239139    0.189208    1.000    2
length[18]    0.623160    0.006431    0.470070    0.781639    0.619062    1.000    2
length[19]    0.066719    0.000227    0.039069    0.097272    0.065952    1.000    2
length[20]    0.035672    0.000135    0.013808    0.058513    0.034879    1.000    2
length[21]    0.189642    0.002798    0.089362    0.295108    0.187217    1.000    2
length[22]    0.009993    0.000026    0.001153    0.019874    0.009276    1.000    2
length[23]    0.032399    0.000248    0.002535    0.061869    0.031131    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000234
    Maximum standard deviation of split frequencies = 0.002149
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.8: Homeobox protein HoxB4a – CDS

Below are rough plots of the generation (x-axis) versus the log probability of observing the data (y-axis). You can use these graphs to determine what the burn in for your analysis should be. When the log probability starts to plateau you may be at stationarity. Sample trees and parameters after the log probability plateaus. Of course, this is not a guarantee that you are at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt' commands for all the parameters in your model. Remember that the burn in is the number of samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

```
   +------------------------------------------------------------+ -8014.21
   |                                                         2  |
   |                                                            |
   |                                2    2       2           1  |
   |       2             11             11              2       |
   |      2          1  2   1 12     2  2 2 2  2                |
   |2        2  21 *   11       1 22   2        1 1   1  1      |
   |11                1  2222             11       22 2 1 2     |
   |   2 2  2   121  2             1  1       1    1 2 * 21    *|
   | 2  2 11   1  2    2     1   1    2    21 2 2 2  1     22   |
   |  2      1      1      1                 * 1              1 |
   |  1 1     *       2      221 21 1                       1   |
   |           2                2    1                     1  2 |
   |   1 1  1                                                   |
   |                                                1           |
   |                2                  1         1              |
   +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -8015.47
   ^                                                            ^
   625000                                                       2500000
```

```
     Model parameter summaries over the runs sampled in files
      “HoxB4a_CDS.nexus.run1.p" and "HoxB4a_CDS.nexus.run2.p":
      Summaries are based on a total of 37502 samples from 2 runs.
      Each run produced 25001 samples of which 18751 samples were included.
      Parameter summaries saved to file "HoxB4a_CDS.nexus.pstat".

                                         95% HPD Interval
                                       --------------------
   Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
   ------------------------------------------------------------------------------------------------------
   TL          2.694830    0.030002    2.364079    3.037263    2.686743  12734.96  12870.77    1.000
   kappa       2.990905    0.041840    2.592151    3.390170    2.984709   9990.83  10213.51    1.000
   pi(A)       0.306055    0.000060    0.290609    0.320868    0.306021   6442.38   6673.46    1.000
   pi(C)       0.256759    0.000051    0.242708    0.270688    0.256675   6505.74   6756.44    1.000
   pi(G)       0.198290    0.000041    0.185680    0.210801    0.198292   6739.32   7157.08    1.000
   pi(T)       0.238896    0.000048    0.225103    0.252146    0.238826   6911.65   7093.28    1.001
   pinvar      0.212454    0.000322    0.178366    0.248379    0.212518  10337.82  10611.52    1.000
   ------------------------------------------------------------------------------------------------------
    * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
      correspond to minimal and average ESS among runs.
      ESS value below 100 may indicate that the parameter is undersampled.
    + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
      and Rubin, 1992) should approach 1.0 as runs converge.
```

```
   Summary statistics for informative taxon bipartitions
   (saved to file “HoxB4a_CDS.nexus.tstat"):

ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
---------------------------------------------------------------
 9  37502    1.000000    0.000000    1.000000    1.000000    2
10  37502    1.000000    0.000000    1.000000    1.000000    2
11  37502    1.000000    0.000000    1.000000    1.000000    2
12  37300    0.994614    0.001735    0.993387    0.995840    2
13  22553    0.601381    0.002828    0.599381    0.603381    2
```

```
14  12526    0.334009    0.003545    0.331502    0.336515    2
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file “HoxB4a_CDS.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.091409    0.002708    0.000005    0.184839    0.086719    1.000    2
length[2]     0.241029    0.005582    0.097274    0.389224    0.237783    1.000    2
length[3]     0.670247    0.013416    0.454409    0.901884    0.661584    1.000    2
length[4]     0.006409    0.000007    0.001419    0.011821    0.006188    1.000    2
length[5]     0.018198    0.000013    0.011314    0.025431    0.017979    1.000    2
length[6]     0.175407    0.000267    0.143979    0.207359    0.174881    1.000    2
length[7]     0.248622    0.000460    0.208013    0.290015    0.248317    1.000    2
length[8]     0.096270    0.003278    0.000021    0.197940    0.090834    1.000    2
length[9]     0.690748    0.015518    0.451875    0.934036    0.680590    1.000    2
length[10]    0.092138    0.000301    0.058474    0.126589    0.091785    1.000    2
length[11]    0.054464    0.000128    0.032983    0.077034    0.054084    1.000    2
length[12]    0.200896    0.003432    0.086233    0.309088    0.203665    1.000    2
length[13]    0.123070    0.005317    0.000004    0.254777    0.114948    1.000    2
length[14]    0.099068    0.004205    0.000034    0.219089    0.090412    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.001351
    Maximum standard deviation of split frequencies = 0.003545
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.9: Homeobox protein HoxB4a - Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -4767.46
      |         1                                2                 |
      |                                                            |
      |                             1                2             |
      |          1  2  2               2           2      1     2  |
      |   1             *     1 122  *       *        2          2 |
      |1    1       1       12 1 1    2   22  * 2 1 2  1   2*2 * 12|
      |221      2  2 *     *212    *2  1 1     *1    11    1    1  |
      |    1222    1     2*     2       1   2    1211  21*    2   1|
      |  2   11   1                     2  1              2  1     |
      | 1  2   1  2   *1 1     2      1  21 1                      |
      |   2                       1                     2     1    |
      |          2                                                 |
      |                                                            |
      |                                                            |
      |        2                                                   |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -4768.61
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "HoxB4a_Prot.nexus.run1.p" and "HoxB4aP_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "HoxB4a_Prot.nexus.pstat".

                                       95% HPD Interval
                                     --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ------------------------------------------------------------------------------------------------------
     TL          1.950341    0.012250    1.733727    2.166488    1.947872  15140.81  15733.87    1.000
     pinvar      0.001328    0.000002    0.000000    0.003987    0.000915   9978.16   9994.86    1.000
     ------------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file “HoxB4a_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
    9  37502    1.000000    0.000000    1.000000    1.000000    2
   10  37502    1.000000    0.000000    1.000000    1.000000    2
   11  37500    0.999947    0.000000    0.999947    0.999947    2
   12  36734    0.979521    0.000754    0.978988    0.980054    2
   13  33695    0.898485    0.000415    0.898192    0.898779    2
   ----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.
```

```
   Summary statistics for branch and node parameters
   (saved to file “HoxB4a_Prot.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.049687    0.000800    0.000240    0.101811    0.045570    1.000    2
length[2]     0.061403    0.000956    0.009390    0.122229    0.056655    1.000    2
length[3]     0.381434    0.007189    0.220693    0.549396    0.377143    1.000    2
length[4]     0.012865    0.000026    0.003800    0.022938    0.012338    1.000    2
length[5]     0.033104    0.000056    0.019050    0.047745    0.032578    1.000    2
length[6]     0.197283    0.000631    0.150057    0.247945    0.196254    1.000    2
length[7]     0.180230    0.000619    0.134329    0.230265    0.179818    1.000    2
length[8]     0.206900    0.001949    0.122305    0.295552    0.204415    1.000    2
length[9]     0.511610    0.008657    0.331050    0.692179    0.508983    1.000    2
length[10]    0.070527    0.000249    0.040502    0.101716    0.069702    1.000    2
length[11]    0.086778    0.000420    0.047734    0.126965    0.085958    1.000    2
length[12]    0.098224    0.001432    0.026884    0.173339    0.096428    1.000    2
length[13]    0.063175    0.001315    0.004213    0.133418    0.057290    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000234
    Maximum standard deviation of split frequencies = 0.000754
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.10: Lim homeobox protein lhx1 – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -6644.49
      |            1      1           2      2                     |
      |    2             1       1                          1      |
      |                       1 12          2      2               |
      |      *     2   22 2 1 222 1       2   22  1      21        |
      | 2         1      2 1   1  22  12   21        2            1|
      |*   1   1122   2 1          1*           12 1  1*1   2    1 |
      | 12  2        2       2         1  1    1212 1 2      1   22|
      |     1 1     2      22        2  11 1            2  2   2*  |
      |   *                  1           2                 1 22    |
      |       222   1111                2    1      21    2    1   |
      |  1                           1                        1    |
      |          1                            1                    |
      |                                                            |
      |                                                            |
      |                                                  1         |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -6656.15
      ^                                                            ^
      14500                                                        58000

         Model parameter summaries over the runs sampled in files
         "LHX1_CDS.nexus.run1.p" and “LHX1_CDS.nexus.run2.p":
         Summaries are based on a total of 872 samples from 2 runs.
         Each run produced 581 samples of which 436 samples were included.
         Parameter summaries saved to file "LHX1_CDS.nexus.pstat".

                                      95% HPD Interval
                                    --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ---------------------------------------------------------------------------------------------------
     TL          8.488856    0.819350    6.843690   10.284510    8.397763    204.41    215.12    0.999
     kappa       3.447326    0.073957    2.948521    3.963319    3.474155     78.93    110.19    0.999
     pi(A)       0.287864    0.000081    0.270663    0.305319    0.287980     88.16    102.87    0.999
     pi(C)       0.243623    0.000067    0.227868    0.259184    0.242788    111.55    137.98    0.999
     pi(G)       0.222686    0.000064    0.207866    0.237483    0.222928     90.14    111.62    1.000
     pi(T)       0.245827    0.000070    0.229619    0.261716    0.245932    145.67    148.41    0.999
     alpha       0.507680    0.001726    0.430428    0.598573    0.504999    195.44    208.96    1.001
     ---------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "LHX1_CDS.nexus.tstat"):

   ID   #obs    Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   16   872    1.000000    0.000000    1.000000    1.000000    2
   17   872    1.000000    0.000000    1.000000    1.000000    2
   18   872    1.000000    0.000000    1.000000    1.000000    2
   19   872    1.000000    0.000000    1.000000    1.000000    2
   20   863    0.989679    0.001622    0.988532    0.990826    2
```

```
21   851    0.975917    0.001622    0.974771    0.977064    2
22   816    0.935780    0.009731    0.928899    0.942661    2
23   803    0.920872    0.021083    0.905963    0.935780    2
24   787    0.902523    0.011353    0.894495    0.910550    2
25   778    0.892202    0.016218    0.880734    0.903670    2
26   739    0.847477    0.017840    0.834862    0.860092    2
27   593    0.680046    0.017840    0.667431    0.692661    2
28   142    0.162844    0.016218    0.151376    0.174312    2
29    99    0.113532    0.001622    0.112385    0.114679    2
----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "LHX1_CDS.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]     0.461453    0.007754    0.284826    0.624833    0.457311    1.000    2
length[2]     0.174859    0.004762    0.041613    0.306979    0.168731    1.001    2
length[3]     0.134870    0.004788    0.008152    0.263901    0.130669    0.999    2
length[4]     0.410042    0.008630    0.221080    0.605662    0.414420    0.999    2
length[5]     0.479865    0.007864    0.314365    0.655197    0.471299    0.999    2
length[6]     1.157551    0.093641    0.585903    1.728442    1.120830    1.000    2
length[7]     0.004550    0.000011    0.000061    0.010194    0.003742    0.999    2
length[8]     0.012336    0.000024    0.004195    0.021989    0.011749    1.001    2
length[9]     0.639920    0.010074    0.450776    0.833515    0.629911    1.001    2
length[10]    0.301530    0.005927    0.154326    0.449205    0.297643    1.003    2
length[11]    0.065460    0.000417    0.028547    0.104979    0.064242    0.999    2
length[12]    0.646171    0.060326    0.198988    1.076544    0.615420    0.999    2
length[13]    0.016111    0.000074    0.001246    0.030672    0.015675    1.001    2
length[14]    0.017671    0.000082    0.002443    0.034922    0.016893    1.004    2
length[15]    0.129539    0.004138    0.009202    0.244756    0.123466    1.001    2
length[16]    0.253596    0.005044    0.110598    0.390992    0.249811    1.006    2
length[17]    1.012407    0.043766    0.649432    1.445448    0.991403    1.000    2
length[18]    0.465770    0.014591    0.209630    0.670926    0.464490    0.999    2
length[19]    0.394181    0.010889    0.213483    0.605395    0.388781    0.999    2
length[20]    0.041615    0.000347    0.009055    0.078942    0.041645    1.007    2
length[21]    0.188048    0.006187    0.048409    0.339265    0.181311    1.000    2
length[22]    0.520519    0.053759    0.101770    0.974284    0.503556    1.000    2
length[23]    0.099761    0.001823    0.017903    0.179256    0.098685    1.006    2
length[24]    0.332328    0.013143    0.100441    0.534944    0.325078    0.999    2
length[25]    0.185815    0.004675    0.066057    0.318176    0.181706    0.999    2
length[26]    0.199285    0.009422    0.029329    0.378103    0.189183    0.999    2
length[27]    0.196511    0.010685    0.046103    0.431150    0.181083    1.000    2
length[28]    0.160329    0.009946    0.006257    0.353167    0.145095    0.995    2
length[29]    0.221855    0.019984    0.005059    0.445052    0.223841    0.997    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.008225
    Maximum standard deviation of split frequencies = 0.021083
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.007
```

Apêndice 4.11: Lim homeobox protein lhx1 – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -3006.62
      |               2       1           1                 12    2|
      |         2   12 11    2 1    1                   2          |
      |                     1  2  1        1   21      1         2 |
      |  21      12       *       21   1                 1 1       |
      |   2       1  1  21  21                      1           2  |
      | 2    2 1           1     2     22     1   1     1     *    |
      |      1 2   *          2 2   2             21      1        |
      |    *    1   2  2             22    2 2 1      *2   2 1   1 |
      |1              1          1 2    1*2 11  2    1             |
      |     2 2          2           1           1   2         2   |
      |211                 2                     2        2     1  |
      |                                       2                1   |
      |       1                                     2       2      |
      |          2                    1            2               |
      |     1                   1           2            2        1|
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -3012.80
      ^                                                            ^
      20700                                                        83000

         Model parameter summaries over the runs sampled in files
         "LHX1_Prot.nexus.run1.p" and "LHX1_Prot.nexus.run2.p":
         Summaries are based on a total of 1248 samples from 2 runs.
         Each run produced 831 samples of which 624 samples were included.
         Parameter summaries saved to file "LHX1_Prot.nexus.pstat".

                                       95% HPD Interval
                                     --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ------------------------------------------------------------------------------------------------
     TL          5.337037    0.349374    4.266499    6.528099    5.306429    276.12    402.53    1.001
     alpha       0.579447    0.004930    0.451319    0.724886    0.572935    306.24    383.82    1.000
     ------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "LHX1_Prot.nexus.tstat"):

   ID   #obs    Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   16  1248    1.000000    0.000000    1.000000    1.000000    2
   17  1235    0.989583    0.003400    0.987179    0.991987    2
   18  1208    0.967949    0.018131    0.955128    0.980769    2
   19  1203    0.963942    0.016998    0.951923    0.975962    2
   20  1160    0.929487    0.027196    0.910256    0.948718    2
   21  1132    0.907051    0.013598    0.897436    0.916667    2
   22  1014    0.812500    0.011332    0.804487    0.820513    2
   23   932    0.746795    0.004533    0.743590    0.750000    2
   24   894    0.716346    0.004533    0.713141    0.719551    2
   25   687    0.550481    0.003400    0.548077    0.552885    2
```

```
26   460    0.368590    0.018131    0.355769    0.381410    2
27   449    0.359776    0.028330    0.339744    0.379808    2
28   380    0.304487    0.006799    0.299679    0.309295    2
29   368    0.294872    0.002266    0.293269    0.296474    2
30   343    0.274840    0.001133    0.274038    0.275641    2
31   316    0.253205    0.002266    0.251603    0.254808    2
32   284    0.227564    0.004533    0.224359    0.230769    2
33   253    0.202724    0.007932    0.197115    0.208333    2
34   216    0.173077    0.018131    0.160256    0.185897    2
35   203    0.162660    0.007932    0.157051    0.168269    2
36   157    0.125801    0.007932    0.120192    0.131410    2
37   152    0.121795    0.009065    0.115385    0.128205    2
----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "LHX1_Prot.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.147413    0.001949    0.060764    0.225494    0.143138    0.999    2
length[2]     0.061604    0.001041    0.005200    0.123087    0.059070    1.002    2
length[3]     0.081838    0.001501    0.004028    0.152865    0.079233    0.999    2
length[4]     0.271220    0.005919    0.111191    0.414000    0.273149    1.000    2
length[5]     0.283420    0.003659    0.182826    0.406745    0.277376    1.000    2
length[6]     1.003538    0.034017    0.662331    1.379594    0.999999    0.999    2
length[7]     0.009513    0.000046    0.000041    0.022765    0.007805    0.999    2
length[8]     0.010381    0.000051    0.000195    0.024145    0.008860    1.003    2
length[9]     0.381737    0.006072    0.232379    0.528674    0.376125    1.000    2
length[10]    0.111851    0.001876    0.030448    0.194734    0.109649    1.000    2
length[11]    0.025889    0.000161    0.004837    0.050369    0.023644    1.000    2
length[12]    0.651854    0.043499    0.302355    1.058900    0.631150    1.000    2
length[13]    0.019988    0.000125    0.001233    0.041498    0.018742    1.003    2
length[14]    0.018839    0.000124    0.001418    0.040863    0.016828    1.000    2
length[15]    0.036338    0.000700    0.000131    0.083538    0.031745    0.999    2
length[16]    0.827884    0.031227    0.536249    1.223547    0.812095    0.999    2
length[17]    0.018103    0.000119    0.000781    0.039597    0.016370    0.999    2
length[18]    0.100246    0.001974    0.022763    0.187047    0.096790    1.000    2
length[19]    0.229447    0.008780    0.062419    0.414045    0.226266    1.003    2
length[20]    0.286054    0.009105    0.074582    0.465953    0.289964    1.000    2
length[21]    0.231634    0.010706    0.027627    0.419032    0.223281    1.007    2
length[22]    0.041266    0.000587    0.000645    0.086941    0.037582    1.000    2
length[23]    0.152407    0.005144    0.019708    0.290636    0.145752    1.001    2
length[24]    0.069391    0.001749    0.000077    0.149504    0.062479    0.999    2
length[25]    0.056760    0.001234    0.000114    0.121770    0.051068    1.002    2
length[26]    0.136352    0.005848    0.000183    0.273180    0.127406    0.999    2
length[27]    0.054029    0.001141    0.000871    0.116706    0.049102    0.998    2
length[28]    0.116748    0.006944    0.000132    0.280059    0.102355    1.014    2
length[29]    0.124127    0.005806    0.000110    0.259505    0.112382    0.998    2
length[30]    0.132151    0.009945    0.000029    0.324391    0.106361    0.998    2
length[31]    0.184989    0.012992    0.000194    0.380719    0.189514    1.033    2
length[32]    0.128707    0.005149    0.001929    0.252444    0.118653    0.997    2
length[33]    0.061211    0.001096    0.012038    0.133109    0.058458    0.997    2
length[34]    0.121854    0.006467    0.000378    0.261742    0.109591    0.997    2
length[35]    0.100629    0.008096    0.000132    0.280284    0.077796    1.011    2
length[36]    0.014345    0.000086    0.001181    0.029941    0.012282    1.015    2
length[37]    0.093397    0.003626    0.005080    0.209149    0.085052    0.994    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.
```

```
       Summary statistics for partitions with frequency >= 0.10 in at least one run:
       Average standard deviation of split frequencies = 0.009890
       Maximum standard deviation of split frequencies = 0.028330
       Average PSRF for parameter values (excluding NA and >10.0) = 1.002
       Maximum PSRF for parameter values = 1.033
```

Apêndice 4.12: Membrane-associated guanylate kinase protein 2 – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -6053.72
      |  1             2                                        1  |
      |     1          1                2              2           |
      |        1  2        2     2                *                |
      |              2     12 2           1   2*      *           2|
      | 1  2 2   2      2          2     22         1      12  2   |
      |   2   *  11 *    11    1     2  11              22   2     |
      |12    1  2       1     1 1  1 1 2   1       1 1    2  1   2 |
      |              1       1  2 2 1         1 **     1  1   *12 1|
      |  2 1   21  *  1  2   2      2 1     21     2     1 2     1 |
      |               2           1        21           1          |
      |2  1               2 1    1           2      2       1      |
      |                                1             2             |
      |     2                                                      |
      |                                                            |
      |                        2      2                            |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -6054.65
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "MAGUK2_CDS.nexus.run1.p" and "MAGUK2_CDS.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "MAGUK2_CDS.nexus.pstat".

                                          95% HPD Interval
                                        --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ------------------------------------------------------------------------------------------------
     TL         1.443225    0.012248    1.237414    1.662680    1.432499   7064.82   7684.91    1.000
     kappa      3.453216    0.112199    2.818083    4.120849    3.430545   6429.67   7050.09    1.000
     pi(A)      0.285893    0.000072    0.269527    0.302852    0.285897   6555.83   6580.55    1.000
     pi(C)      0.254210    0.000065    0.238603    0.270314    0.254183   6479.15   6615.95    1.000
     pi(G)      0.219870    0.000058    0.204652    0.234474    0.219740   6335.73   6924.74    1.000
     pi(T)      0.240027    0.000061    0.224816    0.255291    0.240004   6870.28   7053.71    1.000
     pinvar     0.254008    0.000764    0.198090    0.306243    0.254961   7291.39   8012.13    1.000
     ------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.

Summary statistics for informative taxon bipartitions
      (saved to file "MAGUK2_CDS.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
```

```
-----------------------------------------------------------------
 6  37496    0.999840    0.000075    0.999787    0.999893    2
 7  37286    0.994240    0.000754    0.993707    0.994774    2
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "MAGUK2_CDS.nexus.vstat"):

                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.601499    0.005619    0.466461    0.754206    0.593202    1.000    2
length[2]    0.013674    0.000013    0.006715    0.020679    0.013419    1.000    2
length[3]    0.007347    0.000009    0.002034    0.013130    0.007101    1.000    2
length[4]    0.141914    0.000374    0.105736    0.181635    0.141189    1.000    2
length[5]    0.485825    0.003865    0.372880    0.610647    0.479937    1.000    2
length[6]    0.062110    0.000263    0.030877    0.094192    0.061632    1.000    2
length[7]    0.131478    0.001466    0.054284    0.205523    0.131964    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000415
    Maximum standard deviation of split frequencies = 0.000754
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.13: Membrane-associated guanylate kinase protein 2 – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -3607.92
      |         2      * 2                                         |
      |          1   2             2                               |
      |           2      1        2    21            1 2  2       2|
      |*   2 21               2      2  2  1 1                  11 |
      |        2   12 *        2                         21* *     |
      | 1   1              2 2  1 111 2        1       1    1      |
      |  2          1   2      1       1 11 2   1*       1  2      |
      |       2         1   2    2       2    *         1      2   |
      |           1  1       1             2 2    *2          2 2  |
      |     2      2      1   1     2          2   11              |
      |   11   11                     1         2     2 2         1|
      |                          1                    1        1   |
      |  12  1   2        211             2 1       22           2 |
      | 2                       2                                  |
      |                              1                        1    |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -3608.41
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "MAGUK2_Prot.nexus.run1.p" and "MAGUK2_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "MAGUK2_Prot.nexus.pstat".

                                        95% HPD Interval
                                      --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ---------------------------------------------------------------------------------------------------
     TL         1.395510    0.009856    1.203252    1.589831    1.390445  15788.07  15893.15    1.000
     alpha      1.542839    0.140839    0.914687    2.271613    1.481454  10319.84  10772.24    1.000
     ---------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "MAGUK2_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37502    1.000000    0.000000    1.000000    1.000000    2
   ----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.


      Summary statistics for branch and node parameters
      (saved to file "MAGUK2_Prot.nexus.vstat"):
```

```
                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.607746    0.004620    0.480577    0.745491    0.603816    1.000    2
length[2]    0.031894    0.000077    0.015759    0.049403    0.031110    1.000    2
length[3]    0.014291    0.000039    0.002953    0.026500    0.013557    1.000    2
length[4]    0.123751    0.000584    0.078936    0.172301    0.122595    1.000    2
length[5]    0.361837    0.002186    0.271393    0.453385    0.359488    1.000    2
length[6]    0.158393    0.001353    0.087455    0.231288    0.156603    1.000    2
length[7]    0.097598    0.000496    0.056036    0.142674    0.096455    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.14: Serine:threonine protein kinase Mark2 – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -15431.90
      |           2                   2                            |
      |      1                                                     |
      |                                                            |
      |2            1 2     2                            2   2  22 |
      | 2              211     2     1     1 1 2  2                |
      |   12           1           2     *      2                12|
      |   2 2   1  2                    1    2    1    21 12      1|
      |  2    21    2     1 1             1   2 1*  2    1    *1   |
      |    1 2  2  1 2    2       *     2                 2 11  1  |
      |              1  2  1 2 1*1    1    2  1    1 *1 2  12      |
      |          1    1  2    1    1*       2  1       1           |
      | 11    12 2           1   2     *                       2   |
      |     1              2         2    2        21              |
      |1          1           2                                    |
      |                                     1         2            |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -15432.91
      ^                                                            ^
      625000                                                       2500000

        Model parameter summaries over the runs sampled in files
         "Mark2_CDS.nexus.run1.p" and "Mark2_CDS.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "Mark2_CDS.nexus.pstat".

                                           95% HPD Interval
                                         --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     --------------------------------------------------------------------------------------------------
     TL          1.214717    0.008457    1.045404    1.399924    1.206528   3890.97   4038.81    1.000
     kappa       4.066306    0.095529    3.490958    4.698132    4.048351   3986.98   4215.15    1.000
     pi(A)       0.251241    0.000025    0.241431    0.261108    0.251262   5879.42   6132.92    1.000
     pi(C)       0.287956    0.000028    0.277448    0.298194    0.287913   5392.50   5906.21    1.000
     pi(G)       0.232509    0.000024    0.222880    0.241982    0.232463   5870.14   6652.94    1.000
     pi(T)       0.228294    0.000023    0.218857    0.237622    0.228299   6666.94   6743.03    1.000
     alpha       0.667060    0.007204    0.508395    0.835653    0.660219   3467.09   3671.83    1.000
     --------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "Mark2_CDS.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   -----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37502    1.000000    0.000000    1.000000    1.000000    2
   -----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.
```

```
   Summary statistics for branch and node parameters
   (saved to file "Mark2_CDS.nexus.vstat"):

                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.387189    0.001634    0.311758    0.469205    0.384227    1.000    2
length[2]    0.413440    0.001885    0.333166    0.500049    0.409718    1.000    2
length[3]    0.006607    0.000004    0.003105    0.010463    0.006500    1.000    2
length[4]    0.017476    0.000006    0.012608    0.022346    0.017383    1.000    2
length[5]    0.072035    0.000103    0.052172    0.091711    0.071746    1.000    2
length[6]    0.265756    0.000844    0.210327    0.322786    0.263841    1.000    2
length[7]    0.052213    0.000088    0.033928    0.070457    0.052012    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.15: Serine:threonine protein kinase Mark2 – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -8260.84
      |                             1             2                |
      |1          1                                                |
      |   2           2    1    1   2       2                2     |
      |             1    12        1                      1    2   |
      |221    2        1      21      1  1    1 1 1    2 1        2|
      |   1 211  2  2   1 1          1    1          *  1  2   112 |
      |    1   1 12  2   2  22   2     2     1 2 2 2  *12   11     |
      | 12      1  2 1  2   1 1   2   2  2 * 2 1    2       2   21 |
      |               1        2       12        1            *   1|
      |         2            1            2     2          1       |
      |     1          2             2        2     1              |
      |    2 2 2   1            21      1                          |
      |                           12        1            22        |
      |                                                            |
      |                    2                       1               |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -8261.47
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "Mark2_Prot.nexus.run1.p" and "Mark2_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "Mark2_Prot.nexus.pstat".

                                                   95% HPD Interval
                                                --------------------
      Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
      ---------------------------------------------------------------------------------------------------
      TL          0.661319    0.001005    0.600304    0.723523    0.660408  15169.20  15974.82    1.000
      alpha       1.347632    0.095008    0.830478    1.950329    1.295864  11678.34  12075.99    1.000
      ---------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "Mark2_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   -----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37502    1.000000    0.000000    1.000000    1.000000    2
   -----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.


      Summary statistics for branch and node parameters
      (saved to file "Mark2_Prot.nexus.vstat"):
```

```
                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.208861    0.000369    0.171954    0.247200    0.208177    1.000    2
length[2]    0.171531    0.000263    0.140314    0.203961    0.171061    1.000    2
length[3]    0.009761    0.000007    0.004722    0.015162    0.009525    1.000    2
length[4]    0.016909    0.000012    0.010510    0.023827    0.016683    1.000    2
length[5]    0.067691    0.000081    0.050583    0.085446    0.067321    1.000    2
length[6]    0.032433    0.000045    0.019676    0.045762    0.032040    1.000    2
length[7]    0.154133    0.000253    0.123023    0.185350    0.153679    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.16: Atrial natriuretic peptide receptor 1 – CDS

Below are rough plots of the generation (x-axis) versus the log probability of observing the data (y-axis). You can use these graphs to determine what the burn in for your analysis should be. When the log probability starts to plateau you may be at stationarity. Sample trees and parameters after the log probability plateaus. Of course, this is not a guarantee that you are at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt' commands for all the parameters in your model. Remember that the burn in is the number of samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC analysis.

```
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -14005.06
      |                                          1                 |
      |        1       2                                           |
      |1            1             1                                |
      |   11       1     1     1       11   1                      |
      |   2       2 2            1       1        *   1    *2111 22|
      | 11 222  1 1  1       11       2    2    12 221 1       2   |
      |       12 *   211  2 2      2      2 2211      22 2   2  1  |
      |                   1   22*  11  2 2 1  2 2  11     *   2  1 |
      |2    1              11        2         2     2   1  1      |
      | 2    12    2  2 2  2         11   1             2       2  |
      |                 12   2      2   2    1                     |
      |  2                       22                     1          |
      |                                                            |
      |         2                                                  |
      |                                                           1|
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -14018.49
      ^                                                            ^
      11200                                                        45000

         Model parameter summaries over the runs sampled in files
         "NPR1_CDS.nexus.run1.p" and "NPR1_CDS.nexus.run2.p":
         Summaries are based on a total of 678 samples from 2 runs.
         Each run produced 451 samples of which 339 samples were included.
         Parameter summaries saved to file "NPR1_CDS.nexus.pstat".

                                         95% HPD Interval
                                       --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ------------------------------------------------------------------------------------------------------
     TL          4.881069    0.048088    4.463393    5.308056    4.863719     99.47    125.88    1.000
     r(A<->C)    0.148031    0.000153    0.124755    0.171196    0.147391     60.63     65.54    1.015
     r(A<->G)    0.206207    0.000181    0.183385    0.234976    0.205686    102.42    130.23    1.039
     r(A<->T)    0.090999    0.000086    0.072292    0.108074    0.091198    128.04    140.92    1.004
     r(C<->G)    0.128492    0.000137    0.108617    0.151562    0.127114     57.10     87.68    1.001
     r(C<->T)    0.369683    0.000336    0.334821    0.401847    0.369075     84.81    101.45    1.049
     r(G<->T)    0.056588    0.000070    0.040779    0.072467    0.056290    125.32    153.12    0.999
     pi(A)       0.307515    0.000066    0.294412    0.324040    0.307776     65.34     74.27    1.006
     pi(C)       0.214402    0.000040    0.202968    0.225866    0.215018     61.24     83.74    1.002
     pi(G)       0.230695    0.000045    0.217964    0.242028    0.230442     57.40     76.26    1.006
     pi(T)       0.247388    0.000046    0.234211    0.260698    0.247629    105.66    109.82    1.063
     alpha       1.342886    0.035718    1.005537    1.717966    1.327450    107.88    179.54    1.001
     pinvar      0.125514    0.000458    0.090357    0.171953    0.127170    125.07    179.83    1.007
     ------------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


    Summary statistics for informative taxon bipartitions
      (saved to file "NPR1_CDS.nexus.tstat"):

   ID   #obs    Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
```

```
13   678    1.000000    0.000000    1.000000    1.000000    2
14   678    1.000000    0.000000    1.000000    1.000000    2
15   678    1.000000    0.000000    1.000000    1.000000    2
16   678    1.000000    0.000000    1.000000    1.000000    2
17   678    1.000000    0.000000    1.000000    1.000000    2
18   661    0.974926    0.006258    0.970501    0.979351    2
19   631    0.930678    0.010429    0.923304    0.938053    2
20   391    0.576696    0.002086    0.575221    0.578171    2
21   374    0.551622    0.012515    0.542773    0.560472    2
22   204    0.300885    0.025030    0.283186    0.318584    2
23   158    0.233038    0.025030    0.215339    0.250737    2
24   138    0.203540    0.020859    0.188791    0.218289    2
----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "NPR1_CDS.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.005974    0.000006    0.002160    0.011177    0.005703    0.999    2
length[2]     0.007111    0.000007    0.001639    0.011748    0.006725    0.999    2
length[3]     0.065997    0.000108    0.048252    0.087856    0.065138    0.999    2
length[4]     0.255467    0.000577    0.211724    0.301906    0.253803    1.001    2
length[5]     0.175358    0.000731    0.123538    0.224997    0.174890    1.008    2
length[6]     0.067585    0.000165    0.041018    0.090618    0.068541    1.061    2
length[7]     0.019621    0.000117    0.001018    0.038579    0.018354    1.056    2
length[8]     0.397338    0.001685    0.312162    0.469417    0.396345    0.999    2
length[9]     0.508140    0.002307    0.425297    0.605976    0.503847    1.010    2
length[10]    0.423746    0.002299    0.331910    0.512645    0.421708    0.999    2
length[11]    0.406757    0.001991    0.325233    0.498178    0.404084    0.999    2
length[12]    0.964878    0.007669    0.805316    1.128855    0.958724    1.000    2
length[13]    0.074553    0.000249    0.045234    0.106864    0.074296    1.000    2
length[14]    0.627146    0.003579    0.509534    0.734688    0.623689    1.000    2
length[15]    0.056789    0.000096    0.038333    0.074992    0.056947    1.001    2
length[16]    0.286335    0.002129    0.189192    0.370234    0.285555    0.999    2
length[17]    0.247402    0.001740    0.177629    0.340778    0.243892    1.003    2
length[18]    0.110743    0.001613    0.035419    0.187420    0.110900    0.998    2
length[19]    0.063250    0.000524    0.026081    0.108235    0.062336    1.001    2
length[20]    0.079546    0.000981    0.027062    0.145938    0.077392    0.998    2
length[21]    0.057532    0.000935    0.000077    0.114430    0.054603    0.997    2
length[22]    0.067455    0.000935    0.005479    0.126134    0.064357    1.009    2
length[23]    0.040331    0.001068    0.000078    0.110668    0.033815    1.030    2
length[24]    0.046672    0.000813    0.000259    0.100884    0.042716    1.005    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.008517
    Maximum standard deviation of split frequencies = 0.025030
    Average PSRF for parameter values ( excluding NA and >10.0 ) = 1.007
    Maximum PSRF for parameter values = 1.061
```

Apêndice 4.17: Atrial natriuretic peptide receptor 1 – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -6991.53
      |                                                         1  |
      |      2                                               1     |
      |      1 1          21                2      1        1      |
      | 1       2 1     2         11   1              1          22|
      |  1             2   2          1       *1        2 1   2    |
      |     2 2  *    1      1  22   2       1                     |
      |  2     2   21    21 1 1         2  1 2    1    1       2   |
      |12 *2        2  111     11  21    11      12 *2221 21    2  |
      |2    1     2              1        2        2       2 2 1 11|
      |       1    1         22     212 12  1  2 2          2 1    |
      |    1    1              2  2             1        1         |
      |              *2                    2    2        2         |
      |                     2                                      |
      |                                              1             |
      |                                2                           |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -6992.94
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "NPR1_Prot.nexus.run1.p" and "NPR1_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "NPR1_Prot.nexus.pstat".

                                      95% HPD Interval
                                    --------------------
      Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
      ---------------------------------------------------------------------------------------------------
      TL          4.603227    0.085599    4.048086    5.185627    4.587339  14977.12  15054.49    1.000
      alpha       0.799899    0.004879    0.664391    0.936643    0.796387  10664.15  10794.91    1.000
      ---------------------------------------------------------------------------------------------------
       * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
         correspond to minimal and average ESS among runs.
         ESS value below 100 may indicate that the parameter is undersampled.
       + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
         and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "NPR1_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   13  37502    1.000000    0.000000    1.000000    1.000000    2
   14  37502    1.000000    0.000000    1.000000    1.000000    2
   15  37502    1.000000    0.000000    1.000000    1.000000    2
   16  37502    1.000000    0.000000    1.000000    1.000000    2
   17  37500    0.999947    0.000000    0.999947    0.999947    2
   18  37406    0.997440    0.000453    0.997120    0.997760    2
   19  36313    0.968295    0.003582    0.965762    0.970828    2
   20  26748    0.713242    0.002413    0.711535    0.714949    2
   21  25315    0.675031    0.004261    0.672017    0.678044    2
   22   8921    0.237881    0.002979    0.235774    0.239987    2
```

```
23   6235    0.166258    0.000566    0.165858    0.166658    2
24   5702    0.152045    0.003319    0.149699    0.154392    2
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "NPR1_Prot.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     0.003432    0.000008    0.000000    0.008964    0.002713    1.000    2
length[2]     0.002677    0.000007    0.000000    0.007770    0.001920    1.000    2
length[3]     0.015299    0.000046    0.003200    0.028605    0.014487    1.000    2
length[4]     0.163531    0.000611    0.116667    0.213226    0.162147    1.000    2
length[5]     0.068655    0.000329    0.035050    0.105500    0.067751    1.000    2
length[6]     0.072088    0.000262    0.041582    0.104674    0.071317    1.000    2
length[7]     0.014020    0.000108    0.000001    0.034135    0.011931    1.000    2
length[8]     0.354220    0.002683    0.256365    0.457031    0.351640    1.000    2
length[9]     0.497585    0.004277    0.368953    0.624341    0.495733    1.000    2
length[10]    0.388418    0.003008    0.288205    0.499636    0.385232    1.000    2
length[11]    0.388056    0.004089    0.265614    0.513378    0.385037    1.000    2
length[12]    1.029434    0.013829    0.804592    1.262789    1.022406    1.000    2
length[13]    0.626173    0.005716    0.483562    0.775872    0.622637    1.000    2
length[14]    0.326827    0.003709    0.208580    0.444602    0.323748    1.000    2
length[15]    0.029681    0.000075    0.013561    0.046721    0.028913    1.000    2
length[16]    0.266622    0.002563    0.170614    0.367491    0.263758    1.000    2
length[17]    0.035384    0.000149    0.012780    0.059492    0.034335    1.000    2
length[18]    0.123063    0.001754    0.041209    0.204455    0.120389    1.000    2
length[19]    0.031568    0.000237    0.004095    0.061265    0.029855    1.000    2
length[20]    0.078583    0.001493    0.007553    0.152128    0.074897    1.000    2
length[21]    0.097219    0.001445    0.025304    0.172975    0.095061    1.000    2
length[22]    0.074684    0.001442    0.004397    0.144529    0.072216    1.000    2
length[23]    0.083003    0.001504    0.006662    0.154227    0.079524    1.000    2
length[24]    0.078081    0.001002    0.021081    0.143550    0.076117    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.001464
    Maximum standard deviation of split frequencies = 0.004261
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.18: RNA binding motif single stranded interacting – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -6493.89
      |                       22    2   1                        2 |
      | 2    1                    2                   2    1 1     |
      |       1 1         2         1   2      1 1                 |
      |                      1                  2   1          2   |
      |                                1     22    2  1     1     1|
      |          2 1   *     2  2         2    2  2 21 2 1 2   1   |
      |2      22  1 2       2      1          1      2      2    12|
      |             1 1     1 1    2 2    1 2           2 1   * 1  |
      |  11*2      2     *11   1                                   |
      |     1     2  *2 2       1 1  112         2        2        |
      |1 2      2       1             2  1   1                  2  |
      | 1    2 1                 2          1     1                |
      |   2      1               1       2 2    1  1     2         |
      |                    2                            1    2     |
      |                                    1           1           |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -6494.78
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "RBM_CDS.nexus.run1.p" and "RBM_CDS.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "RBM_CDS.nexus.pstat".

                                               95% HPD Interval
                                             --------------------
      Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
      ---------------------------------------------------------------------------------------------------
      TL          1.302312    0.020287    1.051092    1.591867    1.285205   5191.87   5210.62    1.000
      kappa       4.361826    0.267597    3.421587    5.419221    4.316801   5084.42   5239.04    1.000
      pi(A)       0.292834    0.000067    0.276303    0.308306    0.292781   6449.55   6724.73    1.000
      pi(C)       0.254189    0.000059    0.239037    0.269233    0.254139   7477.55   7584.23    1.000
      pi(G)       0.207835    0.000051    0.193899    0.221754    0.207806   6863.97   6921.00    1.000
      pi(T)       0.245141    0.000057    0.230027    0.259532    0.245123   6912.08   7166.93    1.000
      pinvar      0.411901    0.000641    0.358697    0.457803    0.413208   5231.33   5786.49    1.000
      ---------------------------------------------------------------------------------------------------
       * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
         correspond to minimal and average ESS among runs.
         ESS value below 100 may indicate that the parameter is undersampled.
       + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
         and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "RBM_CDS.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
    6  37088    0.988961    0.000528    0.988587    0.989334    2
    7  36345    0.969148    0.002376    0.967468    0.970828    2
   ----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.
```

```
   Summary statistics for branch and node parameters
   (saved to file "RBM_CDS.nexus.vstat"):

                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.454697    0.006862    0.309630    0.625129    0.444706    1.000    2
length[2]    0.558667    0.008328    0.396182    0.741897    0.548001    1.000    2
length[3]    0.007523    0.000007    0.002686    0.012909    0.007290    1.000    2
length[4]    0.014134    0.000011    0.007966    0.020935    0.013927    1.000    2
length[5]    0.108271    0.000314    0.074378    0.143566    0.107132    1.000    2
length[6]    0.044894    0.000223    0.015658    0.074211    0.045001    1.000    2
length[7]    0.117511    0.002147    0.024708    0.207228    0.117045    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.001452
    Maximum standard deviation of split frequencies = 0.002376
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.19: RNA binding motif single stranded interacting – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -3801.09
      |                2       1                                   |
      |         2  2                                               |
      |             1                1       1   1    2      2     |
      |              2               2                    2        |
      |2  1                2               *   12 1      2      1  |
      |               1  2       2                 1         11    |
      |1 2  *1  1      12   1       2         *  22   1*           |
      |       12 1    2 11    2   111 *  12     1       2     2   *|
      | 212  2 1    2     1 2  22       2   2       12   1121  *2  |
      |       2   2  1     1 *1   22     2           1  1  12    2 |
      | 1  1       1                        1       2              |
      |          2              11     *1    2 2   2               |
      |    2                                                     1 |
      |                   2                                        |
      |           1                       1                        |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -3801.76
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "NPR1_Prot.nexus.run1.p" and "NPR1_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "NPR1_Prot.nexus.pstat".

                                           95% HPD Interval
                                         --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ---------------------------------------------------------------------------------------------------
     TL          0.760194    0.002054    0.674687    0.851804    0.758902  17663.37  17906.02    1.000
     pinvar      0.001366    0.000002    0.000000    0.004084    0.000956   9896.71  10202.25    1.000
     ---------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "NPR1_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   -----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37502    1.000000    0.000000    1.000000    1.000000    2
   -----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.


      Summary statistics for branch and node parameters
      (saved to file "NPR1_Prot.nexus.vstat"):
```

```
                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.257819    0.001180    0.193125    0.327253    0.256051    1.000    2
length[2]    0.222190    0.000796    0.169797    0.280239    0.221445    1.000    2
length[3]    0.013618    0.000024    0.004876    0.023354    0.013042    1.000    2
length[4]    0.017831    0.000031    0.007696    0.028983    0.017271    1.000    2
length[5]    0.088104    0.000189    0.061632    0.114922    0.087411    1.000    2
length[6]    0.086791    0.000549    0.042971    0.133925    0.085584    1.000    2
length[7]    0.073839    0.000163    0.049594    0.099291    0.073266    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.20: Serine:threonine protein kinase – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -11137.70
      |                                                        1   |
      |                                                            |
      |       1                                                    |
      |  2                2   2   1       1      2            2  1 |
      | * 2    1      2 1     11 2     21         1  2 1  1       1|
      |      * 2* 221               2  1 *          11  21  1     2|
      |  1  2           2 1222       22 2       2       1  *2 1    |
      |2   1  2    1 * 1    11  *1 2       1 *2          2      12 |
      |1         2  2  2 2            1   2   121  22  2           |
      |   1       1      1     2                 121  2   2     2  |
      |                    1      2111     21  1      1      2 2   |
      |    21         1                     2                1     |
      |                                                            |
      |                                                            |
      |          1                                                 |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -11138.82
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "Ser_Thr_kinase_CDS.nexus.run1.p" and "Ser_Thr_kinase_CDS.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "Ser_Thr_kinase_CDS.nexus.pstat".

                                       95% HPD Interval
                                     --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ------------------------------------------------------------------------------------------------------
     TL         1.456818    0.022252    1.188378    1.762343    1.442044    4322.32    4330.48    1.000
     kappa      4.590682    0.185178    3.790609    5.452715    4.561520    4297.60    4559.49    1.000
     pi(A)      0.225717    0.000032    0.214944    0.237163    0.225756    6877.59    7000.86    1.000
     pi(C)      0.315682    0.000041    0.302956    0.327976    0.315657    6593.19    6682.74    1.000
     pi(G)      0.239247    0.000034    0.227853    0.250807    0.239212    6946.50    7039.94    1.000
     pi(T)      0.219355    0.000031    0.208306    0.230127    0.219302    7016.76    7264.43    1.000
     alpha      0.533529    0.004612    0.405770    0.668175    0.527983    3832.71    3940.68    1.000
     ------------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "Ser_Thr_kinase_CDS.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37502    1.000000    0.000000    1.000000    1.000000    2
   ----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.
```

```
   Summary statistics for branch and node parameters
   (saved to file "Ser_Thr_kinase_CDS.nexus.vstat"):

                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.640293    0.007786    0.475398    0.813483    0.630752    1.000    2
length[2]    0.005384    0.000004    0.001452    0.009505    0.005239    1.000    2
length[3]    0.013545    0.000008    0.008369    0.018996    0.013375    1.000    2
length[4]    0.077405    0.000181    0.051073    0.103800    0.077077    1.000    2
length[5]    0.452395    0.003882    0.338991    0.577647    0.446401    1.000    2
length[6]    0.086572    0.000189    0.059996    0.113618    0.086160    1.000    2
length[7]    0.181224    0.001097    0.120217    0.250235    0.179504    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.21: Serine:threonine protein kinase – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -6086.24
      |                 2                                          |
      |                                                      1     |
      |                1                                           |
      |    2 21           1                2      2   12 1    2 2  |
      |          2      1       1       2  1   2        *          |
      |   112   11        22     2  12       21 21 221   2         |
      |1    1 2       22        2        2   1  1 1 1        2     |
      | 1         2      1 1 21   2   2 1   1 2    1  21  22   1 * |
      |2212    12  2        *    11      1  2  1 2   2             |
      |                             2                       1  2  *|
      |      1 2    22        22       2                    2      |
      |            11 1  2     1   2      2               11    1  |
      |  2        1          1     1                               |
      |                              11   1                   1    |
      |              1                 1                           |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -6086.83
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "Ser_Thr_kinase_Prot.nexus.run1.p" and "Ser_Thr_kinase_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "Ser_Thr_kinase_Prot.nexus.pstat".

                                        95% HPD Interval
                                      --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     ------------------------------------------------------------------------------------------------------
     TL          0.742024    0.001816    0.659613    0.826043    0.740307  13946.92  14958.10    1.000
     alpha       1.387214    0.118981    0.836127    2.083979    1.324828  10252.25  11257.49    1.000
     ------------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "Ser_Thr_kinase_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37502    1.000000    0.000000    1.000000    1.000000    2
   ----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.


      Summary statistics for branch and node parameters
      (saved to file "Ser_Thr_kinase_Prot.nexus.vstat"):
```

```
                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.291827    0.000725    0.240624    0.345168    0.290636    1.000    2
length[2]    0.005763    0.000007    0.001259    0.010894    0.005398    1.000    2
length[3]    0.012396    0.000014    0.005756    0.019921    0.012032    1.000    2
length[4]    0.063084    0.000108    0.043663    0.084055    0.062499    1.000    2
length[5]    0.236992    0.000535    0.193799    0.283497    0.236081    1.000    2
length[6]    0.077950    0.000127    0.056864    0.100671    0.077381    1.000    2
length[7]    0.054012    0.000176    0.028717    0.080495    0.053338    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

Apêndice 4.22: Mothers against decapentaplegic homolog 4-like – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -10752.54
      |              1                                       1     |
      |     2           2  2                                       |
      |      2         2                        2                  |
      |    1        1     2                 1     1           1    |
      |  112  1  2  22  1 1             1  1 2           2   2  1* |
      |      1         1 1  1  22       2  2  *2 221   121 1   22  |
      |2*2     12111     2   1 1 22 *2   12     11  2*2       2   2|
      |1      2   22  1    12 *  1        1  1        1 1         1|
      |   2 1   1     2         1  2           1   2   2  * 1      |
      |                            1 1 1                   2   1   |
      |                      2    1      2  2               2      |
      |        2                      1             1              |
      |                               2                            |
      |                                                            |
      |                                2                           |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -10764.73
      ^                                                            ^
      15200                                                        61000

         Model parameter summaries over the runs sampled in files
         "SMAD4_CDS.nexus.run1.p" and "SMAD4_CDS.nexus.run2.p":
         Summaries are based on a total of 918 samples from 2 runs.
         Each run produced 611 samples of which 459 samples were included.
         Parameter summaries saved to file "SMAD4_CDS.nexus.pstat".

                                        95% HPD Interval
                                      --------------------
      Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
      -------------------------------------------------------------------------------------------------
      TL          7.531327    0.398569    6.366890    8.802810    7.504724    177.17    261.09    0.999
      kappa       3.722810    0.054589    3.278888    4.157000    3.720967    130.29    144.12    0.999
      pi(A)       0.278579    0.000046    0.263974    0.290675    0.278090     80.81     97.31    1.000
      pi(C)       0.240468    0.000040    0.227710    0.252282    0.240417    124.73    161.71    0.999
      pi(G)       0.196993    0.000035    0.186894    0.206870    0.196784     93.96     95.59    1.002
      pi(T)       0.283960    0.000051    0.272025    0.299233    0.283510    112.63    132.57    1.002
      alpha       0.485303    0.001045    0.427538    0.555070    0.482550    172.99    223.45    0.999
      -------------------------------------------------------------------------------------------------
       * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
         correspond to minimal and average ESS among runs.
         ESS value below 100 may indicate that the parameter is undersampled.
       + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
         and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "SMAD4_CDS.nexus.tstat"):

   ID   #obs    Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   16   918    1.000000    0.000000    1.000000    1.000000    2
   17   918    1.000000    0.000000    1.000000    1.000000    2
   18   918    1.000000    0.000000    1.000000    1.000000    2
   19   918    1.000000    0.000000    1.000000    1.000000    2
   20   918    1.000000    0.000000    1.000000    1.000000    2
```

```
21   917    0.998911    0.001541    0.997821    1.000000    2
22   908    0.989107    0.000000    0.989107    0.989107    2
23   848    0.923747    0.012324    0.915033    0.932462    2
24   797    0.868192    0.016946    0.856209    0.880174    2
25   768    0.836601    0.033892    0.812636    0.860566    2
26   600    0.653595    0.015405    0.642702    0.664488    2
27   476    0.518519    0.000000    0.518519    0.518519    2
28   421    0.458606    0.001541    0.457516    0.459695    2
29   217    0.236383    0.016946    0.224401    0.248366    2
30   149    0.162309    0.032351    0.139434    0.185185    2
31   100    0.108932    0.003081    0.106754    0.111111    2
32    94    0.102397    0.021568    0.087146    0.117647    2
----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "SMAD4_CDS.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
--------------------------------------------------------------------------------------
length[1]     1.052340    0.034111    0.733679    1.432258    1.033172    1.001    2
length[2]     1.482984    0.050195    1.092958    1.943845    1.463823    0.999    2
length[3]     0.345848    0.007951    0.179702    0.530323    0.340972    0.999    2
length[4]     0.086347    0.000436    0.045131    0.125229    0.084490    1.000    2
length[5]     0.104169    0.001070    0.039080    0.158903    0.103851    1.000    2
length[6]     0.186638    0.000784    0.134786    0.238529    0.184725    0.999    2
length[7]     0.136239    0.000967    0.073308    0.195377    0.134155    1.000    2
length[8]     0.231376    0.002920    0.139311    0.351924    0.227697    0.999    2
length[9]     0.550492    0.005440    0.426352    0.707098    0.546169    1.001    2
length[10]    0.339921    0.002960    0.236470    0.443365    0.336192    0.999    2
length[11]    0.004926    0.000007    0.000393    0.010129    0.004540    1.000    2
length[12]    0.012814    0.000016    0.006401    0.021643    0.012341    0.999    2
length[13]    0.035708    0.000205    0.007068    0.063329    0.036531    1.001    2
length[14]    0.332109    0.002560    0.247643    0.431191    0.331597    1.003    2
length[15]    0.152170    0.001412    0.087851    0.229088    0.149997    1.002    2
length[16]    0.138539    0.001960    0.055060    0.219420    0.137636    0.999    2
length[17]    0.282297    0.003708    0.173245    0.411963    0.277383    0.999    2
length[18]    0.577531    0.013022    0.357531    0.797049    0.567016    1.000    2
length[19]    0.344265    0.003806    0.224863    0.455347    0.337539    0.999    2
length[20]    0.346275    0.002954    0.243179    0.448948    0.342045    1.004    2
length[21]    0.179105    0.002051    0.080026    0.259447    0.180552    0.999    2
length[22]    0.333345    0.013996    0.109677    0.557433    0.321963    0.999    2
length[23]    0.032935    0.000197    0.006924    0.062560    0.031535    0.999    2
length[24]    0.081908    0.001145    0.023218    0.153292    0.080070    0.999    2
length[25]    0.077627    0.000697    0.028036    0.127047    0.077893    0.999    2
length[26]    0.039528    0.000563    0.000514    0.080430    0.036220    0.999    2
length[27]    0.058448    0.000816    0.009504    0.112554    0.055968    1.004    2
length[28]    0.060841    0.000884    0.011813    0.119880    0.057082    1.009    2
length[29]    0.029377    0.000334    0.001133    0.062590    0.028081    0.996    2
length[30]    0.069702    0.000961    0.005877    0.119700    0.065225    0.994    2
length[31]    0.019906    0.000199    0.000402    0.045592    0.017007    0.993    2
length[32]    0.057235    0.001028    0.006017    0.113846    0.055576    1.000    2
--------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.009153
    Maximum standard deviation of split frequencies = 0.033892
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.009
```

Apêndice 4.23:Mothers against decapentaplegic homolog 4-like – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -4310.48
      |                2    1  2         2                         |
      | 2            1  221   1                    1               |
      |  1  2      1  2 1                                     21 1 |
      |  2    1*2   1    1       2    2          1           1     |
      |*              11           2    2   2 1   1 11   2    1 121|
      | 1 1  2   2             1        1         2         2     2|
      |   221                1  2 21 1 2  121 2 2      1  11   2   |
      |       2     2         2  11   1   2    2   2222 2       2  |
      |    1 1            2*    1    2       * 11       1          |
      |              2       2           1               12 12     |
      |         1 *         2       1                      2       |
      |            2                2      1           2           |
      |                                          2                 |
      |          1                     1                           |
      |                                               1            |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -4322.14
      ^                                                            ^
      9200                                                         37000

          Model parameter summaries over the runs sampled in files
         "SMAD4_Prot.nexus.run1.p" and "SMAD4_Prot.nexus.run2.p":
         Summaries are based on a total of 558 samples from 2 runs.
         Each run produced 371 samples of which 279 samples were included.
         Parameter summaries saved to file "SMAD4_Prot.nexus.pstat".

                                          95% HPD Interval
                                        --------------------
     Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
     -------------------------------------------------------------------------------------------------
     TL         3.047942    0.040366    2.688812    3.461710    3.042203    225.89    233.89    1.000
     alpha      1.080686    0.018101    0.792305    1.313652    1.068747    145.99    154.14    1.002
     -------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "SMAD4_Prot.nexus.tstat"):

   ID   #obs    Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
   16   558    1.000000    0.000000    1.000000    1.000000    2
   17   558    1.000000    0.000000    1.000000    1.000000    2
   18   558    1.000000    0.000000    1.000000    1.000000    2
   19   558    1.000000    0.000000    1.000000    1.000000    2
   20   558    1.000000    0.000000    1.000000    1.000000    2
   21   558    1.000000    0.000000    1.000000    1.000000    2
   22   558    1.000000    0.000000    1.000000    1.000000    2
   23   558    1.000000    0.000000    1.000000    1.000000    2
   24   484    0.867384    0.015207    0.856631    0.878136    2
   25   396    0.709677    0.030413    0.688172    0.731183    2
```

```
26   239    0.428315    0.002534    0.426523    0.430108    2
27   142    0.254480    0.000000    0.254480    0.254480    2
28   137    0.245520    0.002534    0.243728    0.247312    2
29   136    0.243728    0.045620    0.211470    0.275986    2
30   122    0.218638    0.015207    0.207885    0.229391    2
31    98    0.175627    0.000000    0.175627    0.175627    2
32    82    0.146953    0.015207    0.136201    0.157706    2
33    78    0.139785    0.020275    0.125448    0.154122    2
34    76    0.136201    0.010138    0.129032    0.143369    2
35    73    0.130824    0.012672    0.121864    0.139785    2
36    69    0.123656    0.027879    0.103943    0.143369    2
-----------------------------------------------------------------
+ Convergence diagnostic (standard deviation of split frequencies)
  should approach 0.0 as runs converge.


   Summary statistics for branch and node parameters
   (saved to file "SMAD4_Prot.nexus.vstat"):

                                        95% HPD Interval
                                      --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
---------------------------------------------------------------------------------------
length[1]     0.498053    0.006035    0.352159    0.640517    0.491884    1.008    2
length[2]     0.617802    0.007296    0.446470    0.769312    0.616365    0.999    2
length[3]     0.141044    0.001427    0.067388    0.210573    0.140125    1.014    2
length[4]     0.010771    0.000042    0.000439    0.022373    0.009681    0.998    2
length[5]     0.038016    0.000124    0.018119    0.059748    0.037487    1.006    2
length[6]     0.022550    0.000085    0.007640    0.042127    0.021123    1.003    2
length[7]     0.020369    0.000079    0.004526    0.036617    0.019855    0.998    2
length[8]     0.068202    0.000471    0.027749    0.106774    0.067223    0.999    2
length[9]     0.270290    0.001423    0.207816    0.342093    0.267821    1.000    2
length[10]    0.101807    0.000593    0.058301    0.148228    0.099501    1.000    2
length[11]    0.005205    0.000018    0.000030    0.013747    0.004041    1.000    2
length[12]    0.002698    0.000008    0.000001    0.008098    0.001795    0.999    2
length[13]    0.003022    0.000012    0.000000    0.009672    0.002139    0.999    2
length[14]    0.131034    0.001061    0.078317    0.199278    0.127423    1.000    2
length[15]    0.052681    0.000516    0.013481    0.096293    0.049147    0.998    2
length[16]    0.148698    0.000776    0.093666    0.199692    0.145527    1.000    2
length[17]    0.054913    0.000590    0.015341    0.103914    0.050384    0.998    2
length[18]    0.023045    0.000089    0.006771    0.040038    0.022498    0.998    2
length[19]    0.265827    0.002356    0.174859    0.356075    0.260324    0.999    2
length[20]    0.207196    0.002902    0.114487    0.314254    0.203518    1.007    2
length[21]    0.124588    0.000705    0.076931    0.174965    0.123393    0.998    2
length[22]    0.160006    0.000997    0.093506    0.221609    0.159827    1.009    2
length[23]    0.026660    0.000110    0.010132    0.047976    0.025361    0.999    2
length[24]    0.021562    0.000133    0.002139    0.043355    0.019888    0.998    2
length[25]    0.029058    0.000274    0.003519    0.061854    0.026986    0.998    2
length[26]    0.005382    0.000017    0.000030    0.013922    0.004549    1.003    2
length[27]    0.003257    0.000009    0.000045    0.009269    0.002294    0.997    2
length[28]    0.003579    0.000011    0.000023    0.010125    0.002490    0.993    2
length[29]    0.019928    0.000188    0.001631    0.048879    0.016737    0.993    2
length[30]    0.002735    0.000008    0.000033    0.007003    0.001899    1.022    2
length[31]    0.004883    0.000024    0.000100    0.014944    0.003550    0.991    2
length[32]    0.002919    0.000007    0.000059    0.008729    0.002109    1.011    2
length[33]    0.003576    0.000011    0.000130    0.009771    0.002555    0.992    2
length[34]    0.002956    0.000010    0.000008    0.009448    0.001792    0.987    2
length[35]    0.003066    0.000011    0.000023    0.009821    0.001940    0.992    2
length[36]    0.002896    0.000008    0.000015    0.008399    0.001992    0.986    2
---------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.009414
```

```
Maximum standard deviation of split frequencies = 0.045620
Average PSRF for parameter values (excluding NA and >10.0) = 1.000
Maximum PSRF for parameter values = 1.022
```

Apêndice 4.24: Pangolin J – CDS

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -11500.35
      |           1                                   1            |
      |                                                   1        |
      |       2         1   1  1            1         21       1 1 |
      |  2      2          1 1  1     222      1        1      21  |
      |       1 12  1         1 2        2        2 1     2        |
      |   121       2    1           2  1     22   2   2          2|
      | *1 12        1           2  2 11   1    2    1  22       2 |
      |              2*1  12      11 1    12     *1  2       1     |
      |2  2  1           2   222   21        11            1  1   1|
      |      2 * 1 1    2 2 2    12         22  1                  |
      |1          22   2                           1        * 2 2  |
      |                                             2    1 2       |
      |                                  1                         |
      |                                                            |
      |                                   2                  2     |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -11501.33
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "PangolinJ_CDS.nexus.run1.p" and "PangolinJ_CDS.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "PangolinJ_CDS.nexus.pstat".

                                             95% HPD Interval
                                           --------------------
      Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
      -------------------------------------------------------------------------------------------------
      TL          1.295310    0.009399    1.115839    1.488964    1.287087   6861.43   7111.22    1.000
      kappa       3.589085    0.102490    2.994870    4.238404    3.561646   5942.43   6261.46    1.000
      pi(A)       0.222455    0.000030    0.211709    0.233389    0.222316   7045.02   7126.99    1.000
      pi(C)       0.315047    0.000040    0.302997    0.327575    0.314993   6139.15   6282.04    1.000
      pi(G)       0.246412    0.000034    0.235366    0.258124    0.246277   6779.59   6870.49    1.000
      pi(T)       0.216086    0.000029    0.205705    0.226627    0.216061   7070.62   7123.28    1.000
      pinvar      0.439898    0.000248    0.409600    0.471428    0.440247   6798.38   7406.53    1.000
      -------------------------------------------------------------------------------------------------
      * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
        correspond to minimal and average ESS among runs.
        ESS value below 100 may indicate that the parameter is undersampled.
      + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
        and Rubin, 1992) should approach 1.0 as runs converge.


      Summary statistics for informative taxon bipartitions
      (saved to file "PangolinJ_CDS.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37465    0.999013    0.000113    0.998933    0.999093    2
   ----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.
```

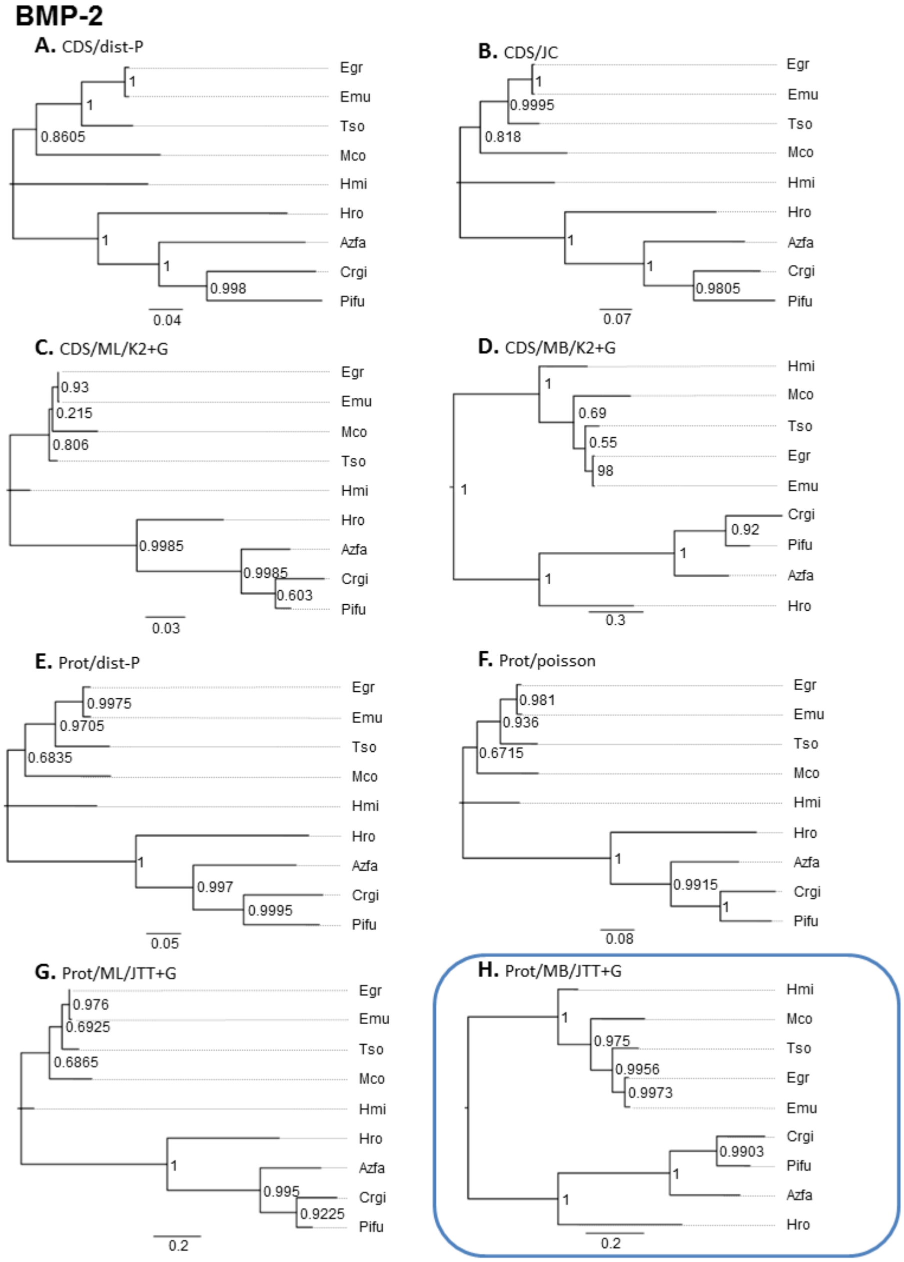
```
   Summary statistics for branch and node parameters
   (saved to file "PangolinJ_CDS.nexus.vstat"):

                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.545416    0.003805    0.433820    0.672662    0.540148    1.000    2
length[2]    0.009659    0.000005    0.005511    0.014052    0.009516    1.000    2
length[3]    0.011488    0.000006    0.007011    0.016133    0.011346    1.000    2
length[4]    0.073162    0.000150    0.048989    0.096863    0.072873    1.000    2
length[5]    0.458836    0.002769    0.363783    0.565557    0.454237    1.000    2
length[6]    0.063595    0.000141    0.040451    0.087120    0.063432    1.000    2
length[7]    0.133263    0.001018    0.070408    0.195683    0.133532    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000057
    Maximum standard deviation of split frequencies = 0.000113
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
```

Apêndice 4.25: Pangolin J – Proteína

```
Below are rough plots of the generation (x-axis) versus the log probability of observing the
data (y-axis). You can use these graphs to determine what the burn in for your analysis should
be. When the log probability starts to plateau you may be at stationarity. Sample trees and
parameters after the log probability plateaus. Of course, this is not a guarantee that you are
at stationarity. Also examine the convergence diagnostics provided by the 'sump' and 'sumt'
commands for all the parameters in your model. Remember that the burn in is the number of
samples to discard. There are a total of ngen / samplefreq samples taken during a MCMC
analysis.
Overlay plot for both runs:
(1 = Run number 1; 2 = Run number 2; * = Both runs)

      +------------------------------------------------------------+ -5744.97
      |                           22     1                         |
      |            2     1    1  1   1         2        1      *   |
      |       2            2                     1     *     1     |
      |                                               1     2      |
      |   1  1  2             2 1   2  *  22            22      2  |
      |    2  12          21    2  1    12        1               2|
      |    1   112                            1 2 2                |
      |22   1         1      2 1             1     2  2       1    |
      |11   2          2 2  21              12       1    1      2 |
      |   2        1 *2 2 1    2 2  1 1 2 11    1   12    2     1 1|
      |      2    2 *   1   1         2     2  1           1     1 |
      |                1             2           2       1         |
      |           1               1           2               2    |
      |  2       1                                  2       12     |
      |  1                                         1       2       |
      +------+-----+-----+-----+-----+-----+-----+-----+-----+-----+ -5745.49
      ^                                                            ^
      625000                                                       2500000

         Model parameter summaries over the runs sampled in files
         "PangolinJ_Prot.nexus.run1.p" and "PangolinJ_Prot.nexus.run2.p":
         Summaries are based on a total of 37502 samples from 2 runs.
         Each run produced 25001 samples of which 18751 samples were included.
         Parameter summaries saved to file "PangolinJ_Prot.nexus.pstat".

                                      95% HPD Interval
                                    --------------------
   Parameter      Mean      Variance     Lower       Upper       Median    min ESS*  avg ESS    PSRF+
   ---------------------------------------------------------------------------------------------------
   TL          0.610806    0.001549    0.534726    0.688004    0.609313  14501.89  15631.07    1.000
   alpha       0.684464    0.016975    0.457731    0.947398    0.667333  11115.74  11214.91    1.000
   ---------------------------------------------------------------------------------------------------
   * Convergence diagnostic (ESS = Estimated Sample Size); min and avg values
     correspond to minimal and average ESS among runs.
     ESS value below 100 may indicate that the parameter is undersampled.
   + Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
     and Rubin, 1992) should approach 1.0 as runs converge.

     Maximum PSRF for parameter values = 1.000


      Summary statistics for informative taxon bipartitions
      (saved to file "PangolinJ_Prot.nexus.tstat"):

   ID   #obs     Probab.     Sd(s)+      Min(s)      Max(s)   Nruns
   ----------------------------------------------------------------
    6  37502    1.000000    0.000000    1.000000    1.000000    2
    7  37502    1.000000    0.000000    1.000000    1.000000    2
   ----------------------------------------------------------------
   + Convergence diagnostic (standard deviation of split frequencies)
     should approach 0.0 as runs converge.
```

```
   Summary statistics for branch and node parameters
   (saved to file "PangolinJ_Prot.nexus.vstat"):

                                       95% HPD Interval
                                     --------------------
Parameter      Mean       Variance     Lower       Upper       Median     PSRF+  Nruns
-------------------------------------------------------------------------------------
length[1]    0.264692    0.000690    0.214181    0.316014    0.263665    1.000    2
length[2]    0.004658    0.000005    0.000998    0.008877    0.004341    1.000    2
length[3]    0.007971    0.000007    0.003118    0.013428    0.007667    1.000    2
length[4]    0.032243    0.000057    0.018071    0.047160    0.031773    1.000    2
length[5]    0.175390    0.000389    0.137423    0.213921    0.174405    1.000    2
length[6]    0.090798    0.000237    0.061412    0.121181    0.090211    1.000    2
length[7]    0.035054    0.000059    0.020848    0.050531    0.034635    1.000    2
-------------------------------------------------------------------------------------
+ Convergence diagnostic (PSRF = Potential Scale Reduction Factor; Gelman
  and Rubin, 1992) should approach 1.0 as runs converge. NA is reported when
  deviation of parameter values within all runs is 0 or when a parameter
  value (a branch length, for instance) is not sampled in all runs.


    Summary statistics for partitions with frequency >= 0.10 in at least one run:
    Average standard deviation of split frequencies = 0.000000
    Maximum standard deviation of split frequencies = 0.000000
    Average PSRF for parameter values (excluding NA and >10.0) = 1.000
    Maximum PSRF for parameter values = 1.000
```

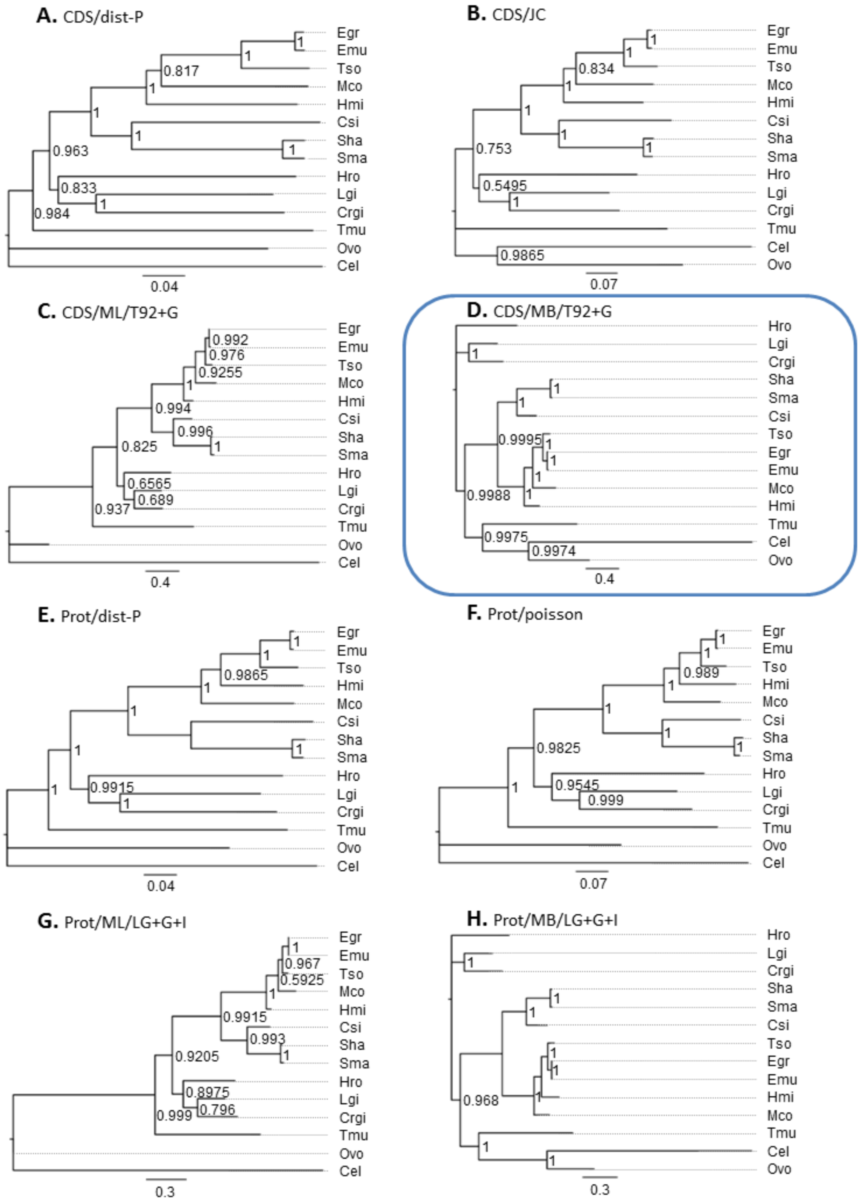
APÊNDICE 5: SUPPLEMENTARY FILE 2



Suppementary File 1. Bone morphogenetic protein 2 (BMP-2) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by K2 with gamma distribution and (D) bayesian by K2 with gamma distribution models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with gamma distribution and (H) bayesian by JTT with gamma distribution models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Azfa: *Azumapecten farreri*; Crgi: *Crassostrea gigas*; Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hro: *Helobdella robusta*; Hmi: *Hymenolepis microstoma*; Mco: *Mesocestoides corti*; Pifu: *Pinctada fucata* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

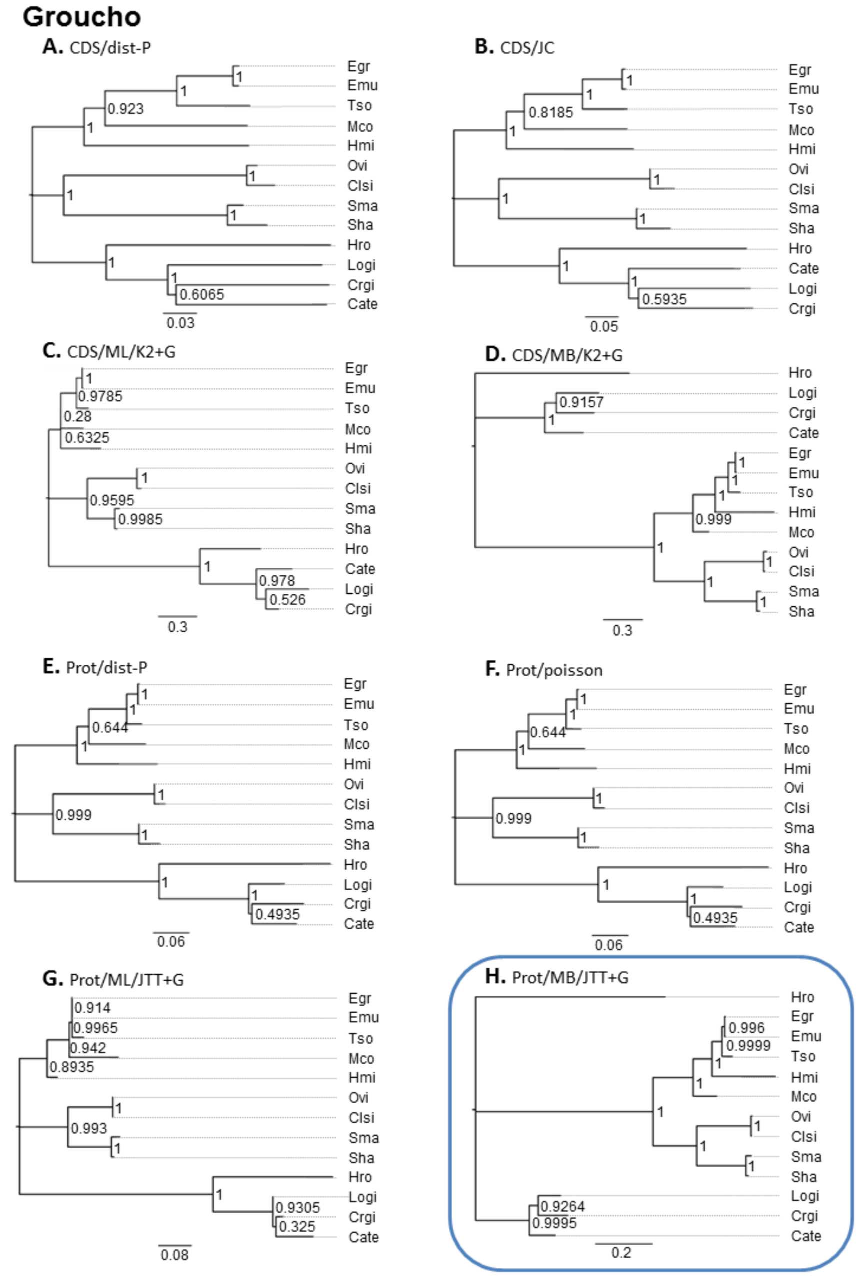
APÊNDICE 6: SUPPLEMENTARY FILE 3

GAK



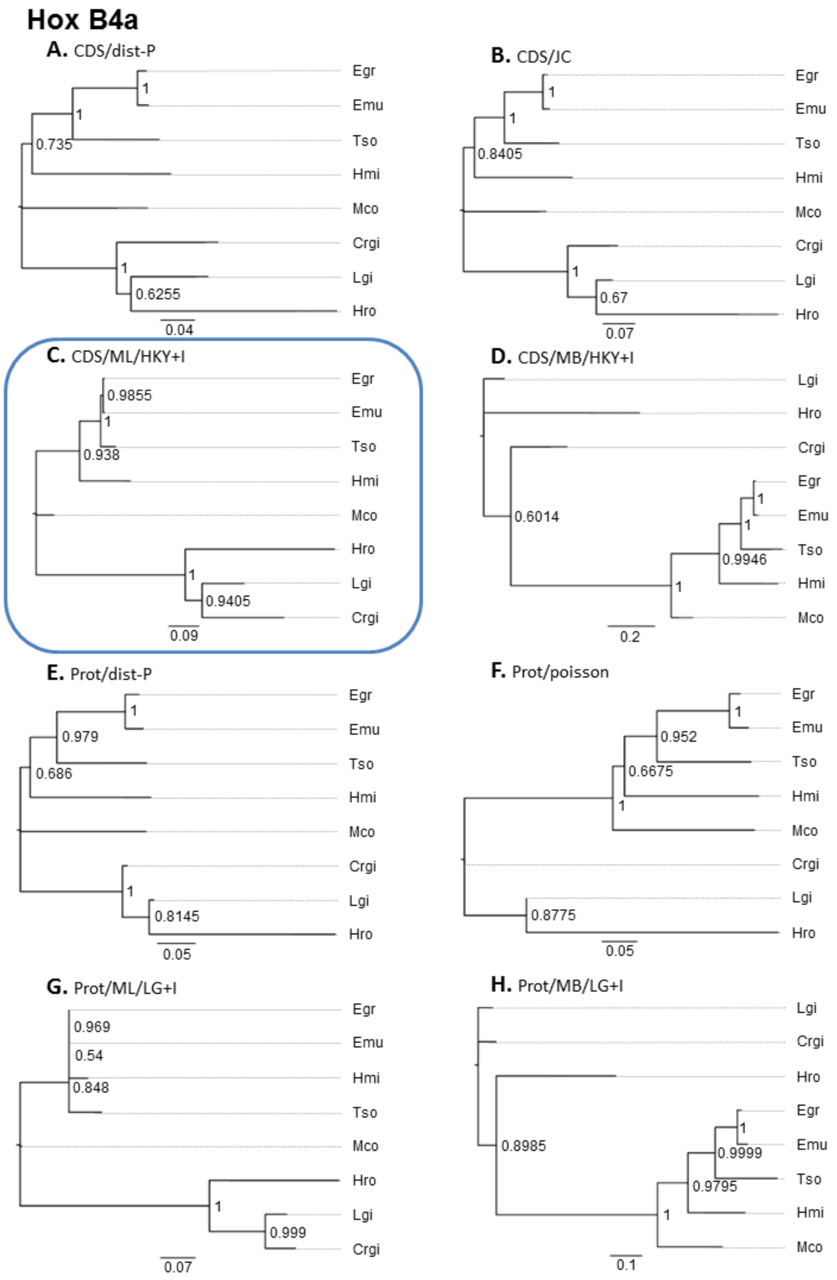
Suppementary File 2. Cyclin-g-associated kinase (GAK) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by T92 with gamma distribution and (D) bayesian by T92 with gamma distribution models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by LG with gamma distribution and proportion of invariable sites and (H) bayesian by LG with gamma distribution and proportion of invariable sites models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Cel: *Caenorhabditis elegans*; Csi: *Clonorchis sinensis*; Crgi: *Crassostrea gigas*; Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hro: *Helobdella robusta*; Hmi: *Hymenolepis microstoma*; Lgi: *Lollita gigantea*; Mco: *Mesocestoides corti*; Ovo: *Onchocerca volvulus*; Sha: *Schistosoma haematobium*; Sma: *Schistosoma mansoni*; Tso: *Taenia solium* and Tmu: *Trichuris muris*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 7: SUPPLEMENTARY FILE 4



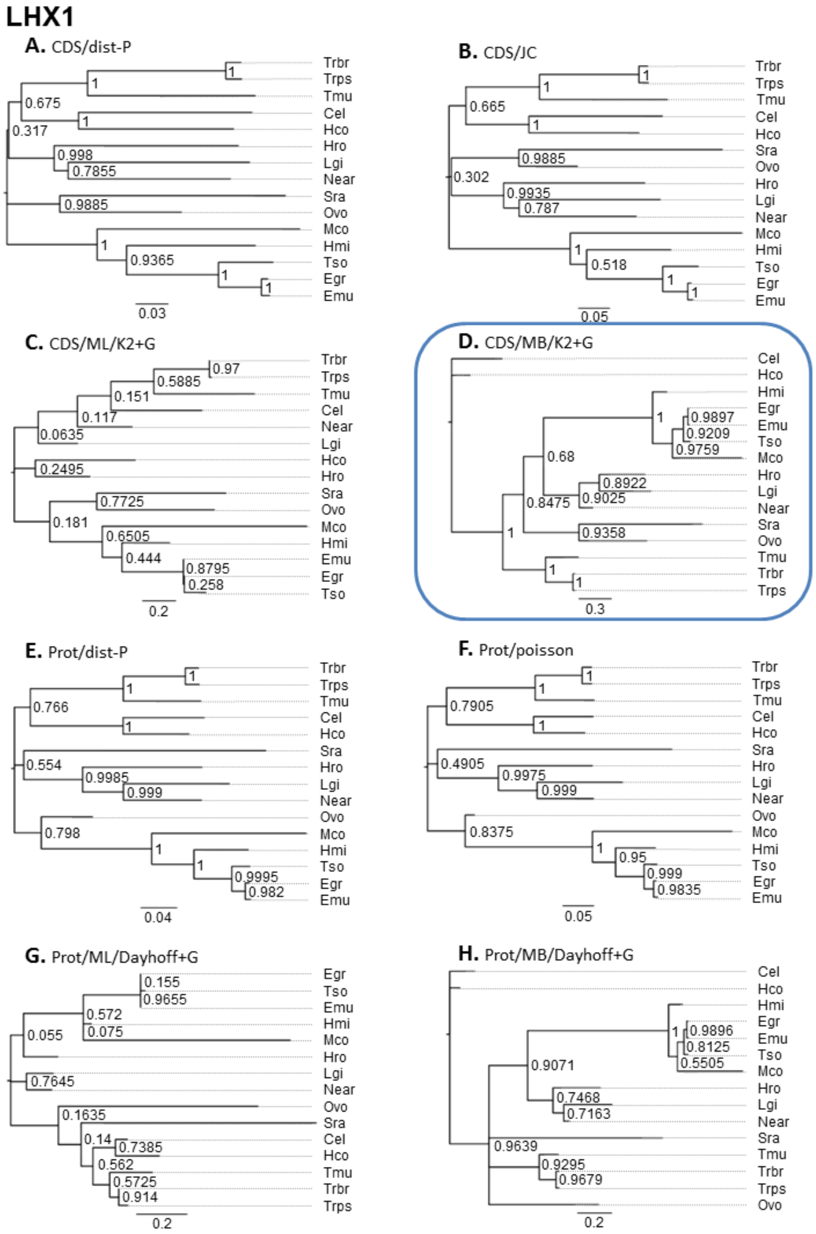
Suppementary File 3. Groucho protein phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by K2 with gamma distribution and (D) bayesian by K2 with gamma distribution models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with gamma distribution and (H) bayesian by JTT with gamma distribution models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Cate: *Capitella teleta*; Csi: *Clonorchis sinensis*; Crgi: *Crassostrea gigas*; Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hro: *Helobdella robusta*; Hmi: *Hymenolepis micróstoma*; Lgi: *Lottia gigantea*; Mco: *Mesocestoides corti*; Ovi: *Opisthorchis viverrini*; Sha: *Schistosoma haematobium*; Sma: *Schistosoma mansoni* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 8: SUPPLEMENTARY FILE 5



Suppementary File 4. Homeobox protein HoxB4a (Hox B4a) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by HKY with proportion of invariable sites and (D) bayesian by HKY with proportion of invariable sites models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by LG with proportion of invariable sites and (H) bayesian by LG with proportion of invariable sites models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Crgi: *Crassostrea gigas*; Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hro: *Helobdella robusta*; Hmi: *Hymenolepis microstomal*; Lgi: *Lollita gigantea*; Mco: *Mesocestoides corti* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 9: SUPPLEMENTARY FILE 6


LHX1

A. CDS/dist-P

B. CDS/JC

C. CDS/ML/K2+G

D. CDS/MB/K2+G

E. Prot/dist-P

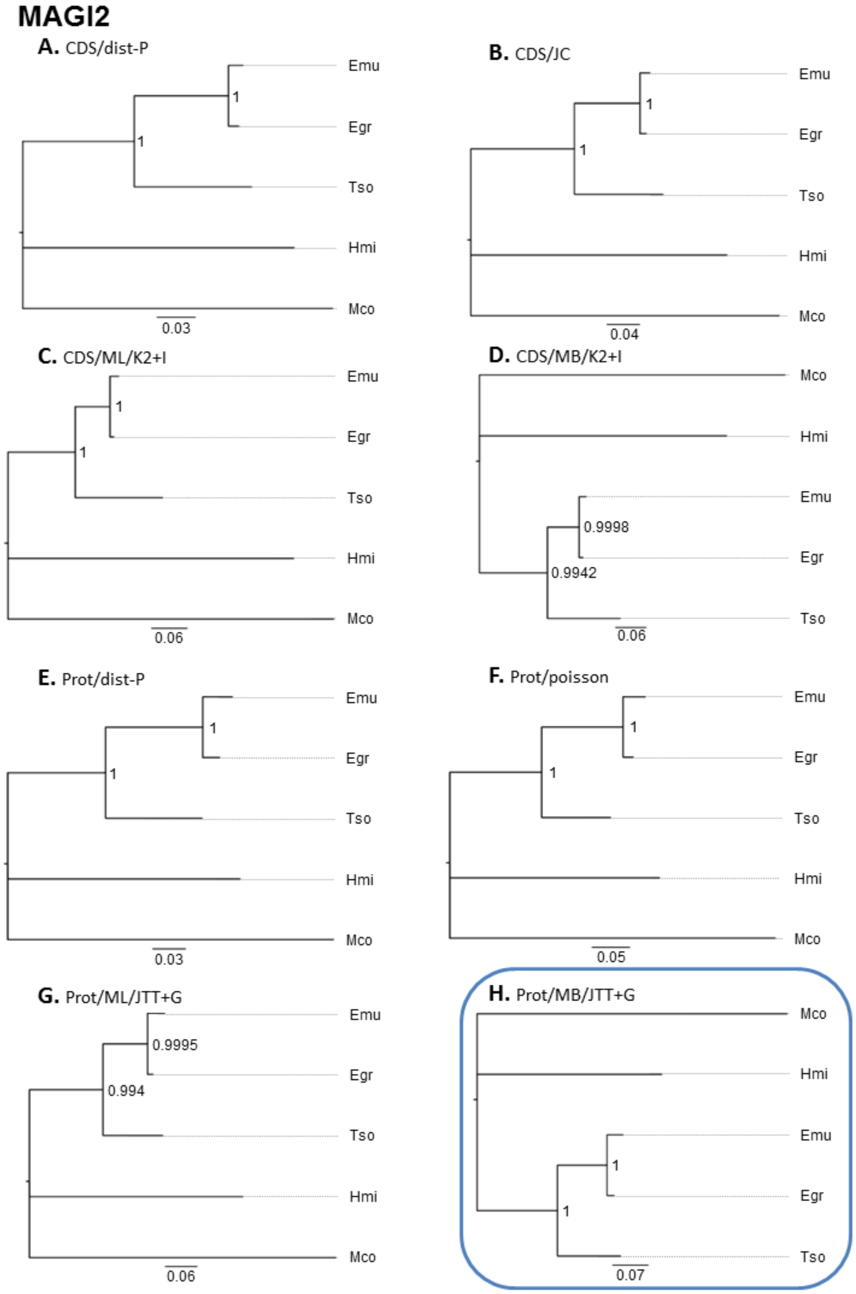
F. Prot/poisson

G. Prot/ML/Dayhoff+G

H. Prot/MB/Dayhoff+G

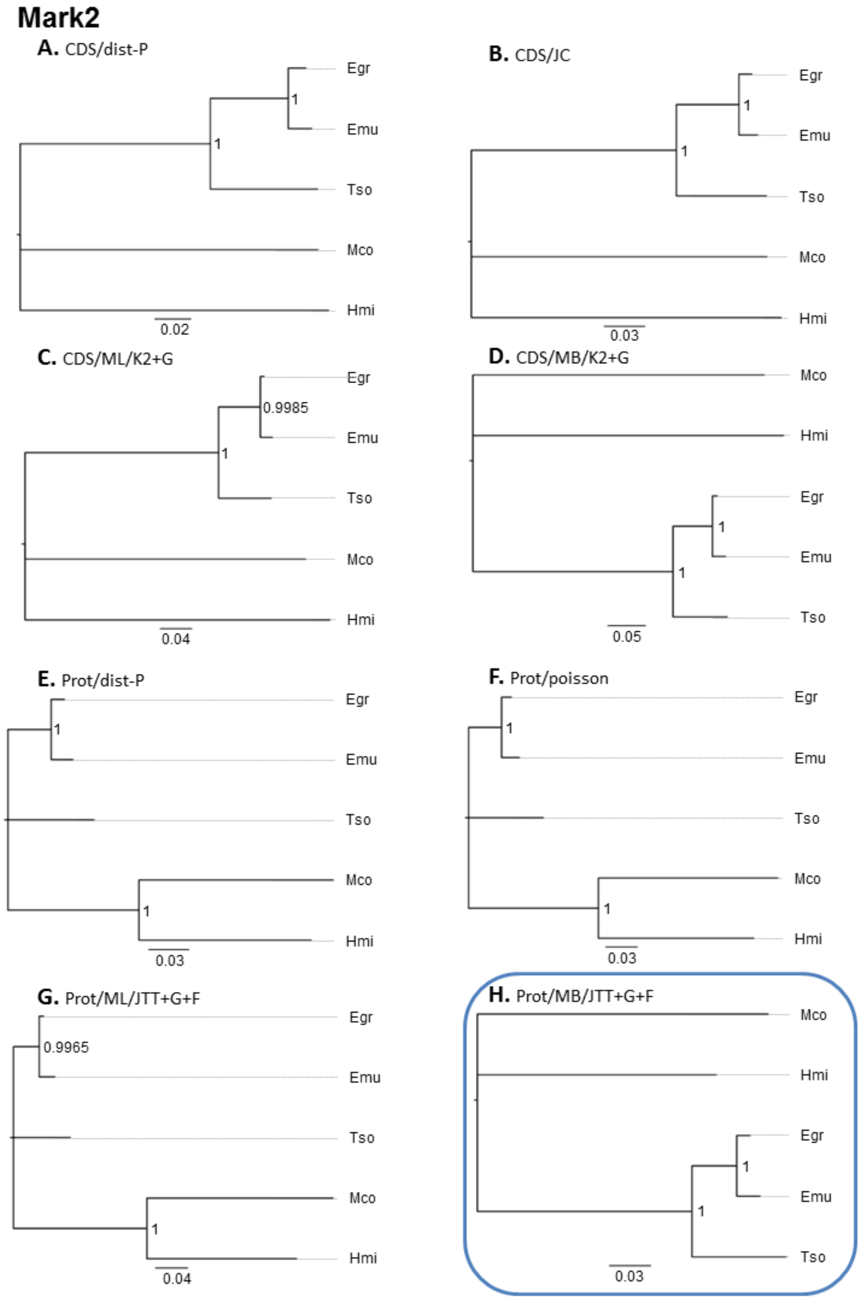
Suppementary File 5. Lim homeobox protein lhx1 (LHX1) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by K2 with gamma distribution and (D) bayesian by K2 with gamma distribution models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by Dayhoff with gamma distribution and (H) bayesian by Dayhoff with gamma distribution models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Cel: *Caenorhabditis elegans*; Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hco: *Haemonchus contortus*; Hro: *Helobdella robusta*; Hmi: *Hymenolepis micróstoma*; Lgi: *Lollita gigantea*; Mco: *Mesocestoides corti*: Ovo: *Onchocerca volvulus*; Sra: *Strongyloides ratti*; Tso: *Taenia solium*; Trbr: *Trichinella britovi*; Trps: *Trichinella pseudospiralis*; Tmu: *Trichuris muris* and Near: *Neanthes arenaceodentata*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 10: SUPPLEMENTARY FILE 7



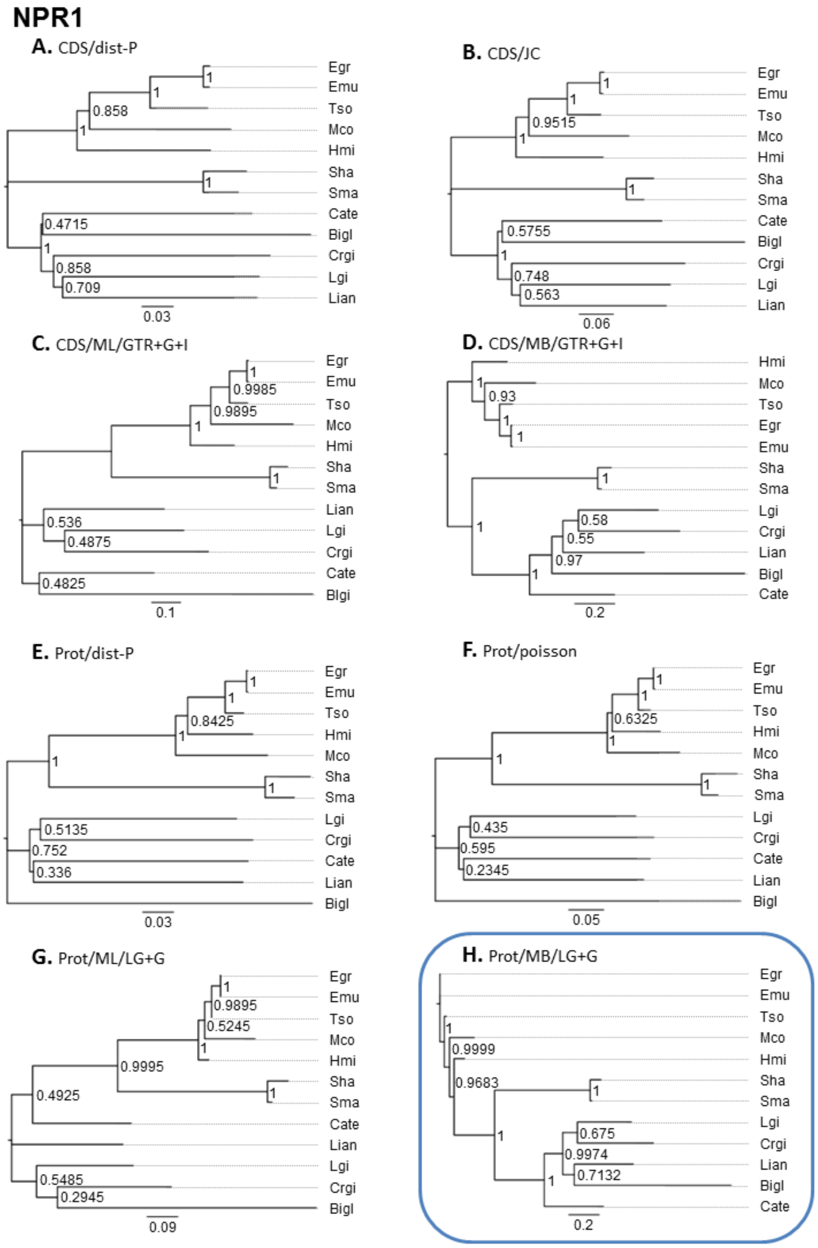
Suppementary File 6. Membrane-associated guanylate kinase protein 2 (MAGI2) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by K2 with proportion of invariable sites and (D) bayesian by K2 with proportion of invariable sites models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with gamma distribution and (H) bayesian by JTT with gamma distribution models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hmi: *Hymenolepis microstoma*; Mco: *Mesocestoides corti* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 11: SUPPLEMENTARY FILE 8



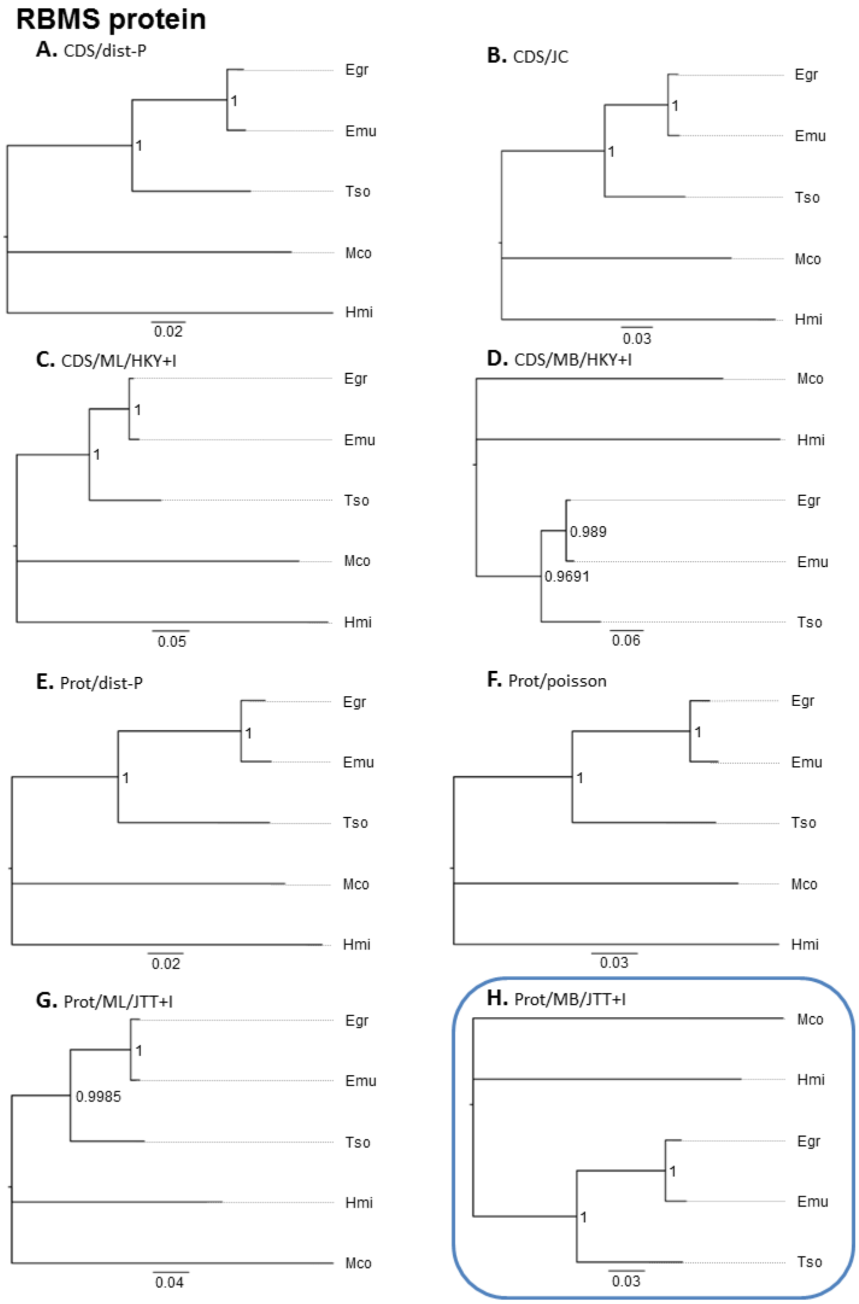
Suppementary File 7. Serine:threonine protein kinase Mark2 (Mark2) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by K2 with gamma distribution and (D) bayesian by K2 with gamma distribution models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with gamma distribution and observed amino acid frequencies and (H) bayesian by JTT with gamma distribution and observed amino acid frequencies models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hmi: *Hymenolepis micróstoma*; Mco: *Mesocestoides corti* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 12: SUPPLEMENTARY FILE 9



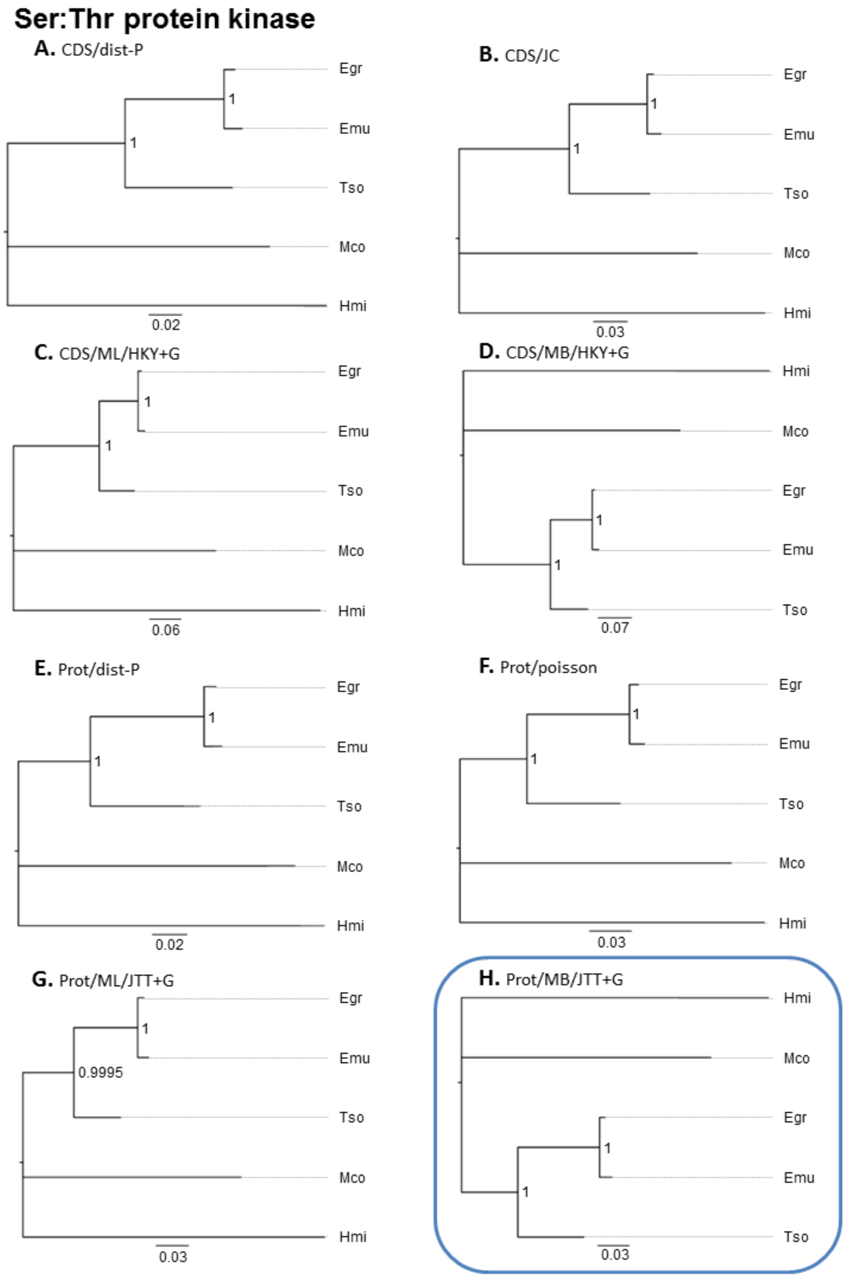
Suppementary File 8. Atrial natriuretic peptide receptor 1 (NPR1) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by GTR with gamma distribution and proportion of invariable sites (D) bayesian by GTR with gamma distribution and proportion of invariable sites models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by LG with gamma distribution and (H) bayesian by LG with gamma distribution models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Bigl: *Biomphalaria glabrata*; Cate: *Capitella teleta*; Crgi: *Crassostrea gigas*; Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hmi: *Hymenolepis microstoma*; Lian: *Lingula anatine*; Lgi: *Lollita gigantea*; Mco: *Mesocestoides corti*; Sha: *Schistosoma haematobium*; Sma: *Schistosoma mansoni* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 13: SUPPLEMENTARY FILE 10



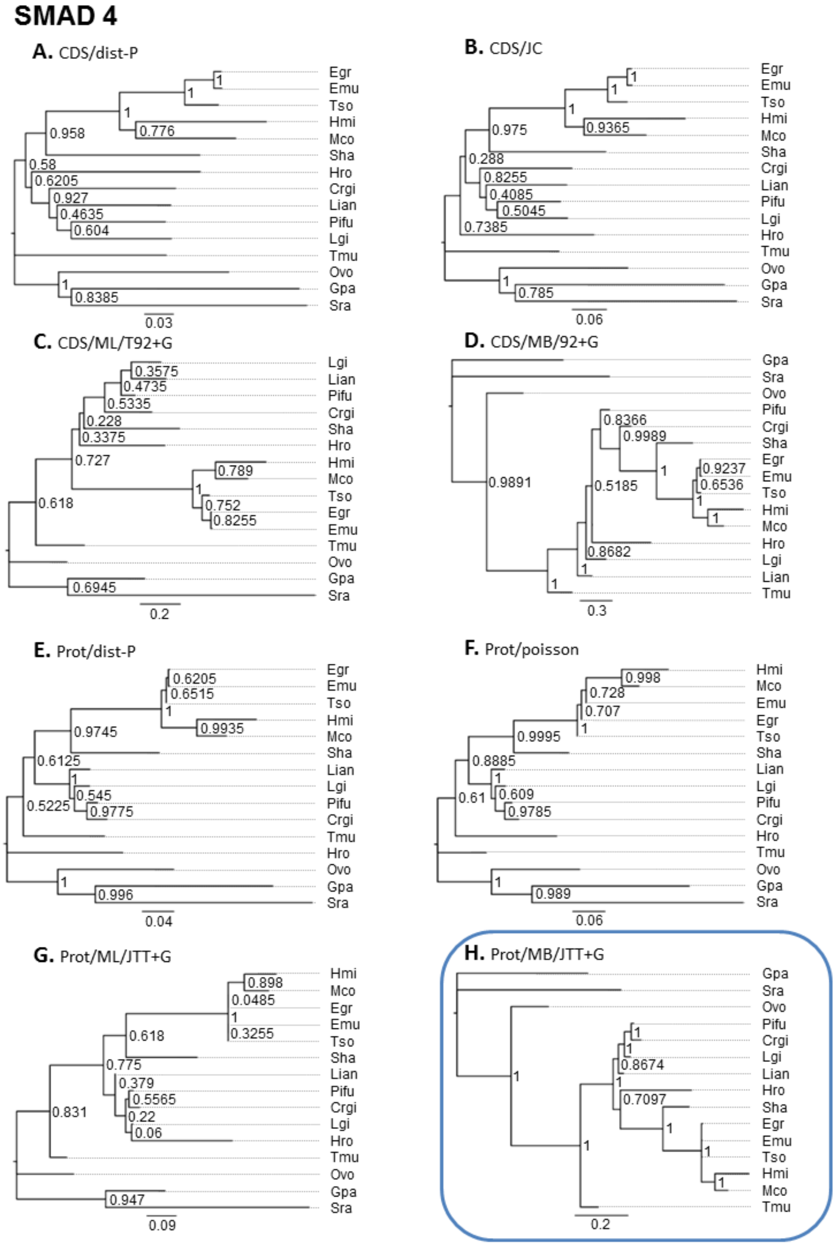
Suppementary File 9. RNA binding motif single stranded interacting (RBMS protein) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by HKY with proportion of invariable sites and (D) bayesian by HKY with proportion of invariable sites models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with proportion of invariable sites and (H) bayesian by JTT with proportion of invariable sites models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hmi: *Hymenolepis microstoma*; Mco: *Mesocestoides corti* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 14: SUPPLEMENTARY FILE 11



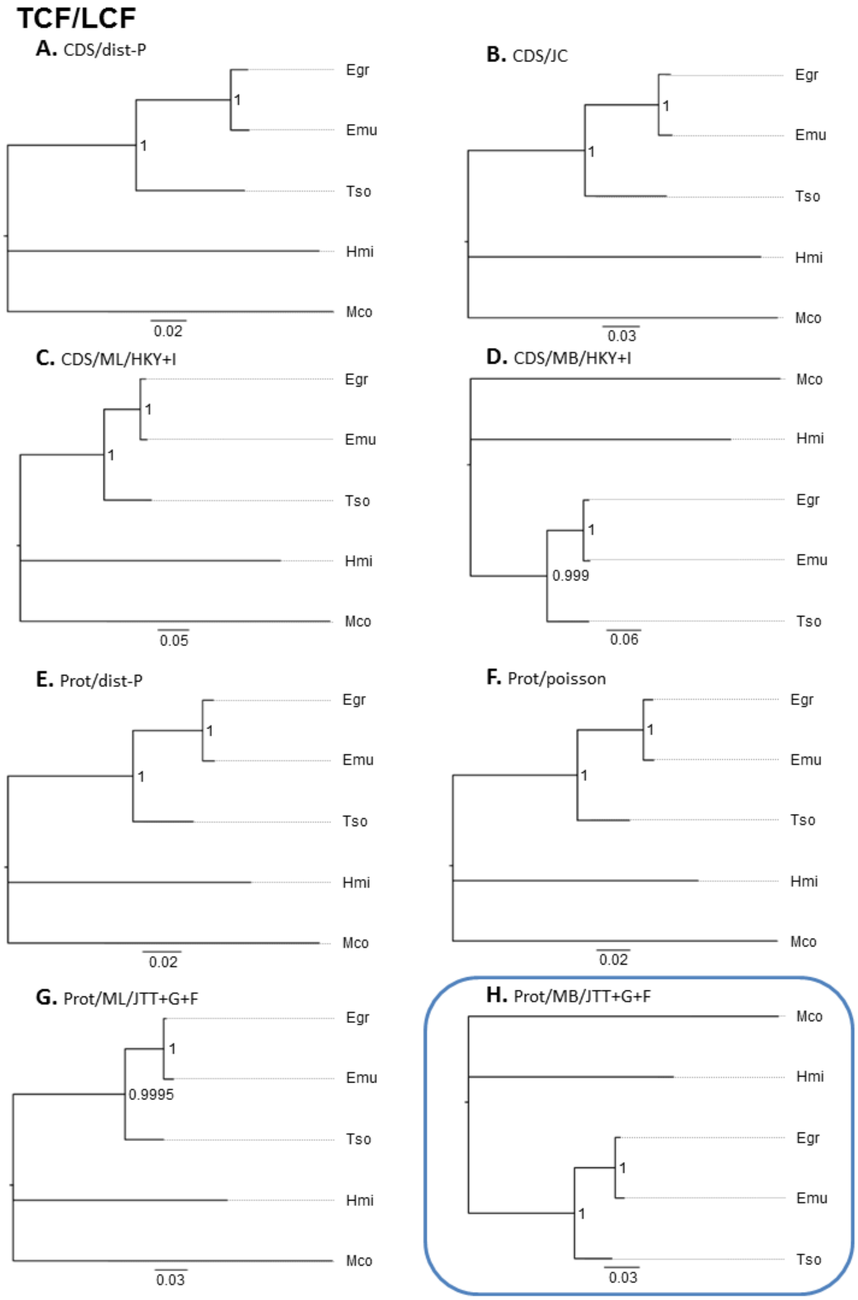
Suppementary File 10. Serine:threonine protein kinase (Ser:Thr protein kinase) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by HKY with gamma distribution and (D) bayesian by HKY with gamma distribution models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with gamma distribution and (H) bayesian by JTT with gamma distribution models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hmi: *Hymenolepis microstoma*; Mco: *Mesocestoides corti* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 15: SUPPLEMENTARY FILE 12



Suppementary File 11. Mothers against decapentaplegic homolog 4-like (SMAD 4) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by T92 with gamma distribution and (D) bayesian by T92 with gamma distribution models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with gamma distribution and (H) bayesian by JTT with gamma distribution models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Crgi: *Crassostrea gigas*; Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Gpa: *Globodera pallida*; Hro: *Helobdella robusta*; Hmi: *Hymenolepis micróstoma*; Lian: *Lingula anatina*; Lgi: *Lollita gigantea*; Mco: *Mesocestoides corti*; Ovo: *Onchocerca volvulus*; Pifu: *Pinctada fucata*; Sha: *Schistosoma haematobium*; Sra: *Strongyloides ratti*; Tso: *Taenia solium* and Tmu: *Trichuris muris*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 16: SUPPLEMENTARY FILE 13



Suppementary File 12. Pangolin J (TCF/LCF) phylogenetic analysis. Phylogenetic trees from CDS alignment build using (A) p-distance, (B) Jukes-Cantor, (C) maximum likelihood by HKY with proportion of invariable sites and (D) bayesian by HKY with proportion of invariable sites models. Phylogenetic trees from protein alignment build using (E) p-distance, (F) poisson, (G) maximum likelihood by JTT with gamma distribution and observed amino acid frequencies and (H) bayesian by JTT with gamma distribution and observed amino acid frequencies models. The best phylogenetic tree is highlighted by a blue box. The species analyzed are Egr: *Echinococcus granulosus*; Emu: *Echinococcus multilocularis*; Hmi: *Hymenolepis microstoma*; Mco: *Mesocestoides corti* and Tso: *Taenia solium*. CDS and protein alignments were described in **Supplementary File 17**.

APÊNDICE 17: SUPPLEMENTARY FILE 14

Supplementary File 14. Analysis of positive selection of the putative proglottisation-related genes.

Protein	Model[1]	Estimates of parameters[2]	-lnL	BEB[3]	NEB[4]
Bone morphogenetic protein 2	M1a: nearly neutral (2)	p0= 0.94844; p1= 0.05156; ω0= 0.05957 ; ω1= 1.00000	2955.986426	NA	NA
	M2a: positive selection (4)	p0= 0.94844 ; p1= 0.05156; p2= 0.00000; ω0= 0.05957; ω1= 1.00000; ω2= 32.95918	2955.986426	NA	NA
	M7: β (2)	p= 0.67448; q= 7.98175	2912.834240	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99999; p1= 0.00001; p= 0.67449; q= 7.98227; ω= 1.00000	2912.834516	0	0
Cyclin-g-associated kinase	M1a: nearly neutral (2)	p0= 0.88087; p1= 0.11913; ω0= 0.04927; ω1= 1.00000	27819.232636	NA	NA
	M2a: positive selection (4)	p0= 0.88087; p1= 0.11913 ; p2= 0.00000; ω0= 0.04927; ω1= 1.00000; ω2= 31.92412	27819.232638	NA	NA
	M7: β (2)	p= 0.94400; q=12.82596	27486.155497	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99506; p1= 0.00494; p= 0.95841; q= 13.33317; ω= 1.00000	27485.924922	0	0
Groucho protein	M1a: nearly neutral (2)	p0= 0.92751; p1= 0.07294; ω0= 0.03053; ω1= 1.00000	16350.199285	NA	NA
	M2a: positive selection (4)	p0= 0.92751; p1= 0.07249; p2= 0.00000; ω0= 0.03053; ω1= 1.00000; ω2= 30.14746	16350.199288	NA	NA
	M7: β (2)	0.24149	16111.693054	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99999; p1= 0.00001; p= 0.54152; q= 10.22404; ω= 5.20234	16111.696723	0	0
Homeobox protein HoxB4a	M1a: nearly neutral (2)	p0= 0.77492; p1= 0.22508; ω0= 0.07604; ω1= 1.00000	7543.879071	NA	NA
	M2a: positive selection (4)	p0= 0.77492; p1= 0.18191; p2= 0.04317; ω0= 0.07604; ω1= 1.00000; ω2= 1.00000	7543.879071	NA	NA
	M7: β (2)	p= 0.48046; q= 1.98811	7495.987477	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99999; p1= 0.00001; p= 0.48047; q= 1.98827; ω= 2.80638	7495.987534	0	0
Lim homeobox protein lhx1	M1a: nearly neutral (2)	p0= 0.90183; p1= 0.09817; ω0= 0.01899; ω1= 1.00000	6071.326703	NA	NA
	M2a: positive selection (4)	p0= 0.90183; p1= 0.09817; p2= 0.00000; ω0= 0.01899; ω1= 1.00000; ω2= 11.71650	6071.326703	NA	NA
	M7: β (2)	p= 0.42728; q= 9.64316	5911.689538	NA	NA
	M8: β & ω > 1 (4)	p0= 0.98573; p1= 0.01427; p= 0.44357; q= 12.11042; ω= 1.99512	5909.497162	0	1 (128/N/0.958)
Membrane-associated guanylate kinase protein 2	M1a: nearly neutral (2)	p0= 0.80943; p1= 0.19057; ω0= 0.11908; ω1= 1.00000	5778.909283	NA	NA
	M2a: positive selection (4)	p0= 0.8094; p1= 0.13801; p2= 0.05255; ω0= 0.11908; ω1= 1.00000; ω2= 1.00000	5778.909283	NA	NA
	M7: β (2)	p= 0.68139; q= 2.38910	5764.532287	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99567; p1= 0.00433; p= 0.69670; q= 2.50687; ω= 5.43515	5764.462324	0	0
Serine:threonine protein kinase Mark2	M1a: nearly neutral (2)	p0= 0.82257; p1= 0.17743; ω0= 0.05050; ω1= 1.00000	14266.657574	NA	NA
	M2a: positive selection (4)	p0= 0.82257; p1= 0.13025; p2= 0.04718; ω0= 0.05050; ω1= 1.00000; ω2= 1.00000	14266.657574	NA	NA
	M7: β (2)	p= 0.30914; q= 1.72381	14252.057177	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99981; p1= 0.00019; p= 0.30976; q= 1.73128; ω= 6.41769	14252.052920	0	0
Atrial natriuretic peptide receptor 1	M1a: nearly neutral (2)	p0= 0.88405; p1= 0.11595; ω0= 0.03759; ω1= 1.00000	13114.022891	NA	NA
	M2a: positive selection (4)	p0= 0.88405; p1= 0.11595; p2= 0.00000; ω0= 0.03759; ω1= 1.00000; ω2= 37.15439	13114.022891	NA	NA
	M7: β (2)	p= 0.66587; q= 11.25087	12843.228735	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99585; p1= 0.00415; p= 0.67595; q= 11.94233; ω= 6.09965	12841.573244	0	0
RNA binding motif single stranded interacting	M1a: nearly neutral (2)	p0= 0.72904; p1= 0.27096; ω0= 0.04888; ω1= 1.00000	6113.503358	NA	NA
	M2a: positive selection (4)	p0= 0.72904; p1= 0.22384; p2= 0.04712; ω0= 0.04888; ω1= 1.00000; ω2= 1.00000	6113.503358	NA	NA
	M7: β (2)	p= 0.25136; q= 0.89618	6101.638216	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99999; p1= 0.00001; p= 0.25136; q= 0.89623; ω= 1.00000	6101.638276	0	0
Serine:threonine protein kinase	M1a: nearly neutral (2)	p0= 0.83951; p1= 0.16049; ω0= 0.05909; ω1= 1.00000	10367.954359	NA	NA
	M2a: positive selection (4)	p0= 0.83951; p1= 0.10874; p2= 0.05175; ω0= 0.05909; ω1= 1.00000; ω2= 1.00000	10367.954359	NA	NA
	M7: β (2)	p= 0.39313; q= 2.25980	10358.846846	NA	NA
	M8: β & ω > 1 (4)	p0= 0.93181; p1= 0.06819; p= 0.55389; q= 4.83056; ω= 1.00000	10356.580326	0	0

Mothers against decapentaplegic homolog 4-like	M1a: nearly neutral (2)	p0= 0.98297; p1= 0.01703; ω0= 0.01139; ω1= 1.00000	9649.215110	NA	NA
	M2a: positive selection (4)	p0= 0.98297; p1= 0.00146; p2= 0.01557; ω0= 0.01139; ω1= 1.00000; ω2= 1.00000	9649.215110	NA	NA
	M7: β (2)	p= 0.67763; q= 45.65467	9454.473046	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99999; p1= 0.00001; p= 0.67770; q= 45.67450; ω= 1.00000	9454.476171	0	0
Pangolin J	M1a: nearly neutral (2)	p0= 0.83239; p1= 0.16761; ω0= 0.02792; ω1= 1.00000	10452.671165	NA	NA
	M2a: positive selection (4)	p0= 0.83239; p1= 0.16735; p2= 0.00026; ω0= 0.02792; ω1= 1.00000; ω2= 1.00000	10452.671165	NA	NA
	M7: β (2)	p= 0.21504; q= 2.04292	10424.734738	NA	NA
	M8: β & ω > 1 (4)	p0= 0.99488; p1= 0.00512; p= 0.21853; q= 2.16274; ω= 1.00000	10424.725950	0	0

[1] Parentheses: number of free parameters of the mode

[2] Model estimates of parameters generated by CodeML analysis

[3] Number of posivitaly selected sited by Bayes Empirical Bayes analysis. Parentheses: alignment syte position/aminiacid/posterior probability. N/A: Not allowed

[4] Number of posivitaly selected sited by Naive Empirical Bayes analysis. Parentheses: alignment syte position/aminiacid/posterior probability. N/A: Not allowed

APÊNDICE 18: SUPPLEMENTARY FILE 15

Supplementary File 15. Taxonomic information of the 18 studied species and genomes source.

Organism	Philum	Class	Source	Reference
Echinococcus granulosus	Platyhelminthes	Cestoda	Sanger Institute[1]	Tsai et al. 2013
Echinococcus multilocularis	Platyhelminthes	Cestoda	Sanger Institute[1]	Tsai et al. 2013
Hymenolepis microstoma	Platyhelminthes	Cestoda	Sanger Institute[1]	Tsai et al. 2013
Mesocestoides corti	Platyhelminthes	Cestoda	Sanger Institute[1]	UNPUBLISHED
Taenia solium	Platyhelminthes	Cestoda	National University of Mexico[2]	Tsai et al. 2013
Clonorchis sinensis	Platyhelminthes	Trematoda	National Center for Biotechnology Information[3]	Wang et al. 2011
Schistosoma haematobium	Platyhelminthes	Trematoda	SchistoDB[4]	Young et al. 2012
Schistosoma japonicum	Platyhelminthes	Trematoda	Shanghai Center for Life Science & Biotechnology Information[5]	Zhou et al. 2009
Schistosoma mansoni	Platyhelminthes	Trematoda	Sanger Institute[1]	Protasio et al. 2012
Opisthorchis viverrini	Platyhelminthes	Trematoda	National Center for Biotechnology Information[3]	Young et al. 2014
Caenorhabditis elegans	Nematoda	Secernentea	WormBase[6]	C. elegans Sequencing Consortium 1998
Globodera pallida	Nematoda	Secernentea	Sanger Institute[1]	Cotton et al. 2014
Haemonchus contortus	Nematoda	Secernentea	Sanger Institute[1]	UNPUBLISHED
Onchocerca volvulus	Nematoda	Secernentea	Sanger Institute[1]	UNPUBLISHED
Strongyloides ratti	Nematoda	Secernentea	Sanger Institute[1]	Hunt et al. 2016
Trichuris muris	Nematoda	Adenophorea	Sanger Institute[1]	Hunt et al. 2016
Helobdella robusta	Annelida	Clitellata	National Center for Biotechnology Information[3]	Simakov et al. 2012
Lollita gigantea	Mollusca	Gastropoda	National Center for Biotechnology Information[3]	Simakov et al. 2012

[1] Sanger Institute database access: http://www.sanger.ac.uk/
[2] National University of Mexico database access: https://www.unam.mx/
[3] National Center for Biotechnology Information database access: http://www.ncbi.nlm.nih.gov/
[4] SchistoDB database access: http://schistodb.net/schisto/
[5] Shanghai Center for Life Science & Biotechnology Information database access: http://lifecenter.sgst.cn/schistosoma/en/schistosomaCnIndexPage.do
[6] WormBase database access: http://www.wormbase.org/#012-34-5

APÊNDICE 19: SUPPLEMENTARY FILE 16

Supplementary File 16. Gene ontology annotations and domains inferred for the selected proteins.

FO[1]	Protein	Blast E-Value Min	Blast similarity mean	Blast GO Number	Top-Hit Species	GO Accession	InterPro IDs	Number of Blast Hits
0	baculoviral iap repeat-containing protein 5	7.0E-62	63.75	34	*Echinococcus granulosus*	GO:0000228,GO:0000777,GO:0005814,GO:0005829, GO:0005876,GO:0005881,GO:0030496,GO:0031021, GO:0032133,GO:0008017,GO:0008270,GO:0008536, GO:0042803,GO:0043027,GO:0046982,GO:0048037, GO:0051087,GO:0000086,GO:0000087,GO:0000226, GO:0000910,GO:0006468,GO:0007067,GO:0009790, GO:0031503,GO:0031536,GO:0031577,GO:0043154, GO:0043524,GO:0045892,GO:0051303,GO:0061178, GO:0061469, GO:0072358	IPR001370 (SMART),IPR001370 (G3DSA:1.10.1170.GENE3D),IPR001370 (PFAM),PTHR10044 (PANTHER),IPR001370 (PROSITE_PROFILES), SSF57924 (SUPERFAMILY)	20
0	bcl-2 homologous antagonist killer	5.1E-30	51.1	27	*Echinococcus multilocularis*	GO:0005739,GO:0016020,GO:0031967,GO:0005515, GO:0002376,GO:0006950,GO:0007548,GO:0008637, GO:0009653,GO:0009967,GO:0031323,GO:0032504, GO:0043065,GO:0044702,GO:0044765,GO:0044802, GO:0048468,GO:0048513,GO:0048523,GO:0048872, GO:0051246,GO:0065009,GO:0097191,GO:0097193, GO:0097285,GO:1902589, GO:2001233	SM00337 (SMART),IPR026298 (PFAM),G3DSA:1.10.437.10 (GENE3D),IPR026308 (PTHR11256:PANTHER),IPR026298 (PANTHER),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),IPR002475 (PROSITE_PROFILES),SSF56854 (SUPERFAMILY),TMhelix (TMHMM), TMhelix (TMHMM)	20

0	bone morphogenetic protein 2	2.2E-106	70.5	9	*Echinococcus multilocularis*	GO:0005615,GO:0005125,GO:0005160,GO:0008083,GO:0010862,GO:0042981,GO:0043408,GO:0048468,GO:0060395	IPR002405 (PRINTS),IPR001839 (SMART),IPR029034 (G3DSA:2.10.90.GENE3D),IPR001839 (PFAM),IPR015615 (PANTHER),IPR017948 (PROSITE_PATTERNS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),SIGNAL_PEPTIDE (PHOBIUS),SIGNAL_PEPTIDE_C_REGION (PHOBIUS),IPR001839 (PROSITE_PROFILES),SignalP-noTM (SIGNALP_EUK),IPR029034 (SUPERFAMILY), TMhelix (TMHMM)	20
0	calmodulin	2.1E-60	73.95	57	*Echinococcus multilocularis*	GO:0000922,GO:0005654,GO:0005813,GO:0005829,GO:0005876,GO:0005886,GO:0030017,GO:0034704,GO:0043005,GO:0070062,GO:0005509,GO:0019901,GO:0019904,GO:0031432,GO:0031996,GO:0031997,GO:0043274,GO:0043539,GO:0044325,GO:0072542,GO:0002027,GO:0002223,GO:0002576,GO:0005513,GO:0005980,GO:0006006,GO:0006936,GO:0007173,GO:0007202,GO:0007264,GO:0007268,GO:0008543,GO:0010800,GO:0010801,GO:0010881,GO:0016056,GO:0021762,GO:0022400,GO:0030168,GO:0030801,GO:0031954,GO:0032465,GO:0032516,GO:0035307,GO:0038095,GO:0043647,GO:0045087,GO:0046209,GO:0048010,GO:0048011,GO:0050999,GO:0051343,GO:0060315,GO:0060316,GO:0061024,GO:0071902,GO:1901844	PR00450 (PRINTS),IPR002048 (SMART),IPR011992 (PFAM),IPR011992 (G3DSA:1.10.238.GENE3D),PTHR23050 (PANTHER),IPR018247 (PROSITE_PATTERNS),IPR002048 (PROSITE_PROFILES), SSF47473 (SUPERFAMILY)	20
0	dynein light chain cytoplasmic	9.4E-56	85.05	13	*Echinococcus granulosus*	GO:0000776,GO:0005813,GO:0005875,GO:0008180,GO:0016020,GO:0043186,GO:0070062,GO:0072686,GO:0003777,GO:0007017,GO:0008152,GO:0021762,GO:0042326	IPR001372 (PFAM),IPR001372 (G3DSA:3.30.740.GENE3D),IPR001372 (PANTHER),IPR019763 (PROSITE_PATTERNS), SSF54648 (SUPERFAMILY)	20

0	early growth response protein 3	0.0E0	66.65	12	*Echinococcus granulosus*	GO:0005634,GO:0005737,GO:0003676,GO:0003700,GO:0046872,GO:0006355,GO:0007274,GO:0007422,GO:0033089,GO:0045586,GO:0071310, GO:0071495	IPR015880 (SMART),IPR013087 (G3DSA:3.30.160.GENE3D),IPR013087 (G3DSA:3.30.160.GENE3D),PF13465 (PFAM),IPR013087 (G3DSA:3.30.160.GENE3D),PTHR24409 (PANTHER),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PROFILES), SSF57667 (SUPERFAMILY)	20
0	f-actin-capping protein subunit beta	4.3E-77	49.7	8	*Echinococcus granulosus*	GO:0016043,GO:0032501,GO:0044087,GO:0048523,GO:0048856,GO:0048869,GO:0051128, GO:0065008	IPR001698 (PRINTS),IPR001698 (PFAM),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),SSF90096 (SUPERFAMILY), TMhelix (TMHMM)	20
0	fibroblast growth factor receptor 4	2.8E-178	61.4	38	*Echinococcus granulosus*	GO:0005794,GO:0005887,GO:0030133,GO:0043235,GO:0000166,GO:0005007,GO:0008201,GO:0017134,GO:0042803,GO:0001503,GO:0006950,GO:0007409,GO:0008283,GO:0008284,GO:0010468,GO:0010863,GO:0016477,GO:0018108,GO:0030900,GO:0035295,GO:0035556,GO:0042511,GO:0042517,GO:0043009,GO:0043406,GO:0043552,GO:0045666,GO:0046777,GO:0048523,GO:0048598,GO:0048839,GO:0051093,GO:0051216,GO:0060350,GO:0060429,GO:0070374,GO:0090596, GO:1902178	IPR001245 (PRINTS),IPR020635 (SMART),G3DSA:1.10.510.10 (GENE3D),IPR001245 (PFAM),G3DSA:3.30.200.20 (GENE3D),PTHR24416 (PANTHER),IPR017441 (PROSITE_PATTERNS),IPR008266 (PROSITE_PATTERNS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),TRANSMEMBRANE (PHOBIUS),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),SIGNAL_PEPTIDE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),SIGNAL_PEPTIDE_C_REGION (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),IPR000719 (PROSITE_PROFILES),SignalP-TM (SIGNALP_EUK),IPR011009 (SUPERFAMILY), TMhelix (TMHMM)	20

0	forkhead box protein d1	2.9E-66	86.1	18	*Echinococcus granulosus*	GO:0000790,GO:0000978,GO:0001077,GO:0001227, GO:0003690,GO:0008301,GO:0000122,GO:0001829, GO:0001892,GO:0006366,GO:0007411,GO:0030513, GO:0045944,GO:0060678,GO:0072076,GO:0072210, GO:0072267, GO:0090184	IPR001766 (PRINTS),IPR001766 (SMART),IPR001766 (PFAM),IPR011991 (G3DSA:1.10.10.GENE3D),PTHR11829 (PANTHER),PTHR11829:SF85 (PANTHER),IPR018122 (PROSITE_PATTERNS),PS00658 (PROSITE_PATTERNS),IPR001766 (PROSITE_PROFILES), SSF46785 (SUPERFAMILY)	20
0	forkhead box protein j3	1.8E-78	70.45	6	*Echinococcus multilocularis*	GO:0005634,GO:0003700,GO:0043565,GO:0006351, GO:0006355, GO:0032502	IPR001766 (PRINTS),IPR001766 (SMART),IPR011991 (G3DSA:1.10.10.GENE3D),IPR001766 (PFAM),PTHR11829 (PANTHER),PTHR11829:SF104 (PANTHER),PS00658 (PROSITE_PATTERNS),IPR018122 (PROSITE_PATTERNS),IPR001766 (PROSITE_PROFILES), SSF46785 (SUPERFAMILY)	20

0	frizzled-10	0.0E0	59.25	4	*Echinococcus granulosus*	GO:0016021,GO:0004888,GO:0007275, GO:0016055	IPR000539 (PRINTS),IPR020067 (SMART),IPR020067 (G3DSA:1.10.2000.GENE3D),IPR020067 (PFAM),IPR000539 (PFAM),IPR026554 (PTHR11309:PANTHER),IPR015526 (PANTHER),TRANSMEMBRANE (PHOBIUS),SIGNAL_PEPTIDE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),SIGNAL_PEPTIDE_C_REGION (PHOBIUS),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),IPR020067 (PROSITE_PROFILES),IPR017981 (PROSITE_PROFILES),SignalP-noTM (SIGNALP_GRAM_NEGATIVE),SignalP-noTM (SIGNALP_EUK),SignalP-TM (SIGNALP_GRAM_POSITIVE),IPR020067 (SUPERFAMILY),TMhelix (TMHMM),TMhelix (TMHMM),TMhelix (TMHMM),TMhelix (TMHMM),TMhelix (TMHMM),TMhelix (TMHMM), TMhelix (TMHMM)	20

0	homeobox protein	1.9E-74	57.85	3	*Echinococcus multilocularis*	GO:0003677,GO:0007275, GO:0050794	IPR001356 (SMART),IPR001356 (PFAM),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR24327:SF3 (PANTHER),PTHR24327 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	homeobox protein arx	1.2E-107	71.9	5	*Echinococcus multilocularis*	GO:0005634,GO:0043565,GO:0006355,GO:0022029, GO:0048666	IPR001356 (SMART),IPR009057 (G3DSA:1.10.10.GENE3D),IPR001356 (PFAM),PTHR24329 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	homeobox protein hox-partial	4.8E-137	85.9	8	*Echinococcus granulosus*	GO:0005634,GO:0003700,GO:0043565,GO:0071837, GO:0006355,GO:0009952,GO:0048704, GO:0051216	IPR020479 (PRINTS),IPR001356 (SMART),IPR001356 (PFAM),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR24326 (PANTHER),PTHR24326:SF111 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	homeobox protein meis1-like	5.9E-99	84.9	13	*Echinococcus multilocularis*	GO:0005634,GO:0000978,GO:0001077,GO:0003714, GO:0008134,GO:0000122,GO:0001654,GO:0006366, GO:0009612,GO:0031016,GO:0045638,GO:0045944, GO:0070848	IPR001356 (SMART),IPR009057 (G3DSA:1.10.10.GENE3D),IPR008422 (PFAM),PTHR11850 (PANTHER),PTHR11850:SF24 (PANTHER),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	homeobox protein nkx-	8.0E-66	74.35	19	*Echinococcus multilocularis*	GO:0005634,GO:0043565,GO:0001776,GO:0002317, GO:0003007,GO:0006641,GO:0022612,GO:0030225, GO:0035050,GO:0042127,GO:0042475,GO:0043367, GO:0045944,GO:0048535,GO:0048536,GO:0048541, GO:0048621,GO:0048738, GO:0050900	IPR020479 (PRINTS),IPR001356 (SMART),IPR001356 (PFAM),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR24340 (PANTHER),PTHR24340:SF9 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20

0	homeobox protein orthopedia	6.2E-74	85.65	7	*Echinococcus granulosus*	GO:0005634,GO:0043565,GO:0002052,GO:0006355, GO:0021879,GO:0021979, GO:0021985	IPR001356 (SMART),IPR009057 (G3DSA:1.10.10.GENE3D),IPR001356 (PFAM),PTHR24329:SF274 (PANTHER),PTHR24329 (PANTHER),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	homeobox protein partial	2.8E-83	78.95	12	*Echinococcus granulosus*	GO:0005634,GO:0005515,GO:0043565,GO:0006355, GO:0009952,GO:0021542,GO:0021796,GO:0021846, GO:0021885,GO:0030182,GO:0042493, GO:0072001	IPR000047 (PRINTS),IPR020479 (PRINTS),IPR001356 (SMART),IPR001356 (PFAM),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR24339:SF24 (PANTHER),PTHR24339 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	inhibitor of growth protein	6.7E-105	79.15	20	*Echinococcus granulosus*	GO:0005794,GO:0016580,GO:0016602,GO:0003677, GO:0003682,GO:0008270,GO:0032403,GO:0035064, GO:0035091,GO:0007141,GO:0007286,GO:0008285, GO:0016568,GO:0030317,GO:0030511,GO:0043065, GO:0045893,GO:0048133,GO:0072520, GO:2001234	IPR001965 (SMART),IPR019787 (PFAM),IPR013083 (G3DSA:3.30.40.GENE3D),IPR028651 (PANTHER),IPR019786 (PROSITE_PATTERNS),IPR019787 (PROSITE_PROFILES), IPR011011 (SUPERFAMILY)	20
0	krueppel-like factor 5	0.0E0	76.1	8	*Echinococcus granulosus*	GO:0043231,GO:0003676,GO:0046872,GO:0006357, GO:0040025,GO:0044763,GO:0045087, GO:0048522	IPR015880 (SMART),IPR013087 (G3DSA:3.30.160.GENE3D),PF13465 (PFAM),IPR013087 (G3DSA:3.30.160.GENE3D),PTHR23223 (PANTHER),PTHR23223:SF143 (PANTHER),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PROFILES), SSF57667 (SUPERFAMILY)	20
0	lim homeobox protein lhx1	9.2E-81	54.0	4	*Echinococcus multilocularis*	GO:0005634,GO:0003677,GO:0006355, GO:0048513	IPR001781 (SMART),IPR001356 (SMART),IPR001356 (PFAM),IPR001781 (PFAM),IPR001781 (G3DSA:2.10.110.GENE3D),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR24208:SF81 (PANTHER),PTHR24208 (PANTHER),IPR001781 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES),IPR001781 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20

0	lim homeobox protein lhx2 isoform x2	3.1E-84	47.8	6	*Echinococcus granulosus*	GO:0005488,GO:0006355,GO:0009653,GO:0009888, GO:0021537, GO:0030182	IPR001781 (SMART),IPR001356 (SMART),IPR001781 (PFAM),IPR001356 (PFAM),IPR009057 (G3DSA:1.10.10.GENE3D),IPR001781 (G3DSA:2.10.110.GENE3D),PTHR24208:SF8 (PANTHER),PTHR24208 (PANTHER),IPR001781 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES),IPR001781 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	mediator of rna polymerase ii transcription	0.0E0	52.58	9	*Echinococcus granulosus*	GO:0031981,GO:0003712,GO:0005102,GO:0006357, GO:0006366,GO:0030154,GO:0048513,GO:0048522, GO:0051716	IPR019680 (PFAM),PTHR12881 (PANTHER),PTHR12881:SF6 (PANTHER),PTHR12881 (PANTHER),PTHR12881:SF6 (PANTHER),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),SIGNAL_PEPTIDE (PHOBIUS), SIGNAL_PEPTIDE_C_REGION (PHOBIUS)	12
0	membrane-associated guanylate ww and pdz domain-containing protein 2	3.9E-124	51.75	9	*Echinococcus multilocularis*	GO:0005737,GO:0005515,GO:0007165,GO:0007399, GO:0016043,GO:0032879,GO:0048523,GO:0051130, GO:0051641	IPR001478 (SMART),IPR001478 (PFAM),IPR001478 (G3DSA:2.30.42.GENE3D),IPR001478 (G3DSA:2.30.42.GENE3D),PTHR10316 (PANTHER),PTHR10316:SF40 (PANTHER),IPR001478 (PROSITE_PROFILES), IPR001478 (SUPERFAMILY)	20
0	metastasis suppressor protein 1	5.9E-108	59.95	7	*Echinococcus granulosus*	GO:0005856,GO:0003779,GO:0007009,GO:0009888, GO:0030036,GO:0050794, GO:0072001	Coil (COILS),Coil (COILS),IPR013606 (PFAM),G3DSA:1.20.1270.80 (GENE3D),PTHR15708:SF4 (PANTHER),IPR030127 (PANTHER), SSF103657 (SUPERFAMILY)	20

0	mitogen-activated protein kinase kinase kinase 9	0.0E0	42.35	7	*Echinococcus granulosus*	GO:0004672,GO:0002009,GO:0006468,GO:0007275, GO:0044763,GO:0050794, GO:0051716	Coil (COILS),G3DSA:1.10.510.10 (GENE3D),IPR000719 (PFAM),PTHR23257 (PANTHER),IPR015785 (PTHR23257:PANTHER),IPR000719 (PROSITE_PROFILES), IPR011009 (SUPERFAMILY)	20
0	paired box protein pax 1	6.7E-63	84.45	5	*Echinococcus granulosus*	GO:0005634,GO:0043565,GO:0006351,GO:0006355, GO:0007275	IPR001523 (PRINTS),IPR001523 (SMART),IPR011991 (G3DSA:1.10.10.GENE3D),IPR001523 (PFAM),PTHR24329 (PANTHER),IPR001523 (PROSITE_PATTERNS),IPR001523 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	pancreas transcription factor 1 subunit alpha	1.8E-100	74.0	8	*Echinococcus multilocularis*	GO:0043229,GO:0003700,GO:0043565,GO:0046983, GO:0006355,GO:0009887,GO:0030154, GO:0031016	IPR011598 (SMART),IPR011598 (G3DSA:4.10.280.GENE3D),IPR011598 (PFAM),PTHR23349 (PANTHER),PTHR23349:SF44 (PANTHER),IPR011598 (PROSITE_PROFILES), IPR011598 (SUPERFAMILY)	20

0	pleckstrin y	0.0E0	51.6	15	*Echinococcus granulosus*	GO:0005634,GO:0030864,GO:0044463,GO:0097458, GO:0098590,GO:0005515,GO:0006605,GO:0007178, GO:0016482,GO:0032880,GO:0044707,GO:0044767, GO:0051128,GO:0060341, GO:0090002	Coil (COILS),Coil (COILS),Coil (COILS),Coil (COILS),Coil (COILS),IPR001849 (SMART),IPR001715 (SMART),IPR018159 (SMART),G3DSA:1.20.58.60 (GENE3D),G3DSA:1.20.58.60 (GENE3D),G3DSA:1.20.58.60 (GENE3D),IPR001715 (PFAM),G3DSA:1.20.58.60 (GENE3D),IPR002017 (PFAM),IPR001715 (G3DSA:1.10.418.GENE3D),IPR011993 (G3DSA:2.30.29.GENE3D),IPR001715 (G3DSA:1.10.418.GENE3D),G3DSA:1.20.58.60 (GENE3D),PTHR11915 (PANTHER),PTHR11915:SF65 (PANTHER),PTHR11915:SF65 (PANTHER),PTHR11915:SF65 (PANTHER),PTHR11915:SF65 (PANTHER),PTHR11915:SF65 (PANTHER),PTHR11915 (PANTHER),PTHR11915 (PANTHER),PTHR11915 (PANTHER),PTHR11915 (PANTHER),IPR001715 (PROSITE_PROFILES),IPR001849 (PROSITE_PROFILES),SSF46966 (SUPERFAMILY),SSF46966 (SUPERFAMILY),SSF46966 (SUPERFAMILY),SSF50729 (SUPERFAMILY),IPR001715 (SUPERFAMILY),SSF46966 (SUPERFAMILY), SSF46966 (SUPERFAMILY)	20
0	pou class transcription factor 1 isoform x4	3.7E-54	54.85	3	*Echinococcus multilocularis*	GO:0005634,GO:0003677, GO:0048513	IPR013847 (PRINTS),IPR001356 (SMART),IPR009057 (G3DSA:1.10.10.GENE3D),IPR001356 (PFAM),PTHR11636:SF5 (PANTHER),PTHR11636 (PANTHER),IPR001356 (PROSITE_PROFILES),IPR000327 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	pre-b-cell leukemia	0.0E0	80.65	9	*Echinococcus granulosus*	GO:0005634,GO:0005667,GO:0003700,GO:0043565, GO:0000060,GO:0001654,GO:0006351,GO:0006357, GO:0007422	IPR001356 (SMART),IPR001356 (PFAM),IPR005542 (PFAM),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR11850 (PANTHER),PTHR11850:SF61 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20

0	protein pangolin j	0.0E0	82.1	77	*Echinococcus multilocularis*	GO:0005634,GO:0005667,GO:0005737,GO:0032993, GO:0000978,GO:0001077,GO:0003682,GO:0003705, GO:0008013,GO:0008134,GO:0008301,GO:0030284, GO:0030331,GO:0035326,GO:0042393,GO:0043027, GO:0045295,GO:0070016,GO:0070742,GO:0000122, GO:0001649,GO:0001755,GO:0001756,GO:0001837, GO:0002040,GO:0006366,GO:0010718,GO:0021542, GO:0021854,GO:0021861,GO:0021873,GO:0021879, GO:0021943,GO:0022408,GO:0022409,GO:0030223, GO:0030307,GO:0030326,GO:0030335,GO:0030509, GO:0030854,GO:0030879,GO:0032696,GO:0032713, GO:0032714,GO:0033153,GO:0042100,GO:0042475, GO:0043154,GO:0043392,GO:0043401,GO:0043586, GO:0043923,GO:0043966,GO:0043967,GO:0045063, GO:0045843,GO:0045944,GO:0048069,GO:0048341, GO:0048747,GO:0050909,GO:0060021,GO:0060033, GO:0060070,GO:0060325,GO:0060326,GO:0060710, GO:0061153,GO:0071353,GO:0071864,GO:0071866, GO:0071895,GO:0071899,GO:0090068,GO:0090090, GO:1902262	IPR009071 (SMART),IPR009071 (PFAM),IPR009071 (G3DSA:1.10.30.GENE3D),IPR028782 (PTHR10373:PANTHER),IPR024940 (PANTHER),IPR009071 (PROSITE_PROFILES),IPR009071 (SUPERFAMILY), TMhelix (TMHMM)	20
0	protein sox-15	2.1E-158	71.65	15	*s multiloculari s*	GO:0005737,GO:0044798,GO:0000981,GO:0003677, GO:0003682,GO:0046982,GO:0000122,GO:0006366, GO:0014718,GO:0043403,GO:0045843,GO:0045944, GO:0048627,GO:0070318, GO:2000288	PR00886 (PRINTS),IPR009071 (SMART),IPR009071 (G3DSA:1.10.30.GENE3D),IPR009071 (PFAM),PTHR10270 (PANTHER),IPR009071 (PROSITE_PROFILES), IPR009071 (SUPERFAMILY)	20
0	protein tiptop	0.0E0	68.9	4	*Echinococcus multilocularis*	GO:0046872,GO:0010468,GO:0044767, GO:0048856	IPR015880 (SMART),PF12756 (PFAM),IPR027008 (PANTHER),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PROFILES), SignalP-TM (SIGNALP_GRAM_POSITIVE)	20

0	protein wnt-11b-1-like	0.0E0	54.9	14	*Echinococcus multilocularis*	GO:0005578,GO:0005102,GO:0007507,GO:0009888, GO:0009966,GO:0010604,GO:0016055,GO:0030154, GO:0031323,GO:0048522,GO:0048523,GO:0048598, GO:0051179, GO:0080090	IPR005817 (PRINTS),IPR005817 (SMART),IPR005817 (PFAM),PTHR12027:SF37 (PANTHER),IPR005817 (PANTHER),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS), CYTOPLASMIC_DOMAIN (PHOBIUS)	20
0	protein yippee-like 1	1.8E-127	76.45	1	*Echinococcus granulosus*	GO:0007420	IPR004910 (PFAM),PTHR13847 (PANTHER), PTHR13847:SF179 (PANTHER)	20
0	ras responsive element binding protein 1	0.0E0	59.7	16	*Echinococcus multilocularis*	GO:0005730,GO:0005737,GO:0016604,GO:0070062, GO:0000979,GO:0046872,GO:0000122,GO:0006366, GO:0007265,GO:0007275,GO:0010634,GO:0033601, GO:0045893,GO:1900026,GO:1903691, GO:2000394	IPR015880 (SMART),IPR013087 (G3DSA:3.30.160.GENE3D),IPR013087 (G3DSA:3.30.160.GENE3D),IPR013087 (G3DSA:3.30.160.GENE3D),IPR013087 (G3DSA:3.30.160.GENE3D),IPR013087 (G3DSA:3.30.160.GENE3D),PF13465 (PFAM),IPR007087 (PFAM),PTHR24409 (PANTHER),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PROFILES), SSF57667 (SUPERFAMILY)	20
0	rna binding motif single stranded interacting	7.3E-116	58.05	3	*Echinococcus granulosus*	GO:0003676,GO:0016301, GO:0006796	IPR000504 (SMART),IPR012677 (G3DSA:3.30.70.GENE3D),IPR012677 (G3DSA:3.30.70.GENE3D),IPR000504 (PFAM),IPR031096 (PTHR24011:PANTHER),PTHR24011 (PANTHER),IPR000504 (PROSITE_PROFILES), SSF54928 (SUPERFAMILY)	20
0	sal protein 3	9.0E-88	69.6	3	*Echinococcus granulosus*	GO:0003676,GO:0046872, GO:0009790	IPR015880 (SMART),IPR013087 (G3DSA:3.30.160.GENE3D),PF13465 (PFAM),IPR013087 (G3DSA:3.30.160.GENE3D),PTHR23233 (PANTHER),PTHR23233:SF49 (PANTHER),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PROFILES), SSF57667 (SUPERFAMILY)	20

0	serine threonine-protein kinase pak partial	0.0E0	73.25	14	*Echinococcus granulosus*	GO:0005737,GO:0000166,GO:0004674,GO:0048365,GO:0007266,GO:0007346,GO:0016477,GO:0023014,GO:0030036,GO:0031098,GO:0031175,GO:0032147,GO:0042981, GO:0043408	IPR002290 (SMART),G3DSA:1.10.510.10 (GENE3D),IPR000719 (PFAM),G3DSA:3.30.200.20 (GENE3D),PTHR24361:SF25 (PANTHER),PTHR24361 (PANTHER),PTHR24361 (PANTHER),IPR017441 (PROSITE_PATTERNS),IPR008271 (PROSITE_PATTERNS),IPR000719 (PROSITE_PROFILES), IPR011009 (SUPERFAMILY)	20
0	serine:threonine protein kinase	0.0E0	81.05	21	*Echinococcus granulosus*	GO:0005634,GO:0005884,GO:0016328,GO:0045180,GO:0097427,GO:0000287,GO:0004674,GO:0005515,GO:0005524,GO:0008289,GO:0044822,GO:0050321,GO:0001764,GO:0010976,GO:0016055,GO:0030010,GO:0035556,GO:0045197,GO:0046777,GO:0050770,GO:0051493	IPR015940 (SMART),IPR002290 (SMART),G3DSA:1.10.8.10 (GENE3D),G3DSA:3.30.200.20 (GENE3D),G3DSA:1.10.510.10 (GENE3D),IPR000719 (PFAM),PTHR24346 (PANTHER),PTHR24346:SF23 (PANTHER),IPR017441 (PROSITE_PATTERNS),IPR008271 (PROSITE_PATTERNS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),IPR015940 (PROSITE_PROFILES),IPR000719 (PROSITE_PROFILES), IPR011009 (SUPERFAMILY)	20
0	serine:threonine protein kinase mark2	0.0E0	72.25	27	*Echinococcus granulosus*	GO:0005634,GO:0005739,GO:0005884,GO:0016328,GO:0045180,GO:0097427,GO:0000287,GO:0004674,GO:0005515,GO:0005524,GO:0008289,GO:0030295,GO:0044822,GO:0050321,GO:0000422,GO:0001764,GO:0010976,GO:0016055,GO:0018107,GO:0030010,GO:0032147,GO:0035556,GO:0045197,GO:0046777,GO:0050770,GO:0051493, GO:0051646	IPR015940 (SMART),IPR002290 (SMART),G3DSA:1.10.8.10 (GENE3D),IPR028375 (G3DSA:3.30.310.GENE3D),IPR001772 (PFAM),G3DSA:1.10.510.10 (GENE3D),IPR000719 (PFAM),PTHR24346 (PANTHER),IPR008271 (PROSITE_PATTERNS),IPR015940 (PROSITE_PROFILES),IPR000719 (PROSITE_PROFILES),IPR001772 (PROSITE_PROFILES),IPR028375 (SUPERFAMILY), IPR011009 (SUPERFAMILY)	20

0	single-stranded dna-binding protein 3	6.1E-91	80.95	14	*Echinococcus granulosus*	GO:0005634,GO:0005737,GO:0043234,GO:0003697, GO:0005515,GO:0002244,GO:0006351,GO:0006461, GO:0008284,GO:0021501,GO:0021547,GO:0045944, GO:0060323, GO:2000744	IPR008116 (PRINTS),PTHR12610 (PANTHER),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TMhelix (TMHMM), TMhelix (TMHMM)	20
0	single-stranded dna-binding protein 3	4.3E-153	72.3	5	*Echinococcus granulosus*	GO:0005634,GO:0003697,GO:0006355,GO:0035220, GO:0048812	IPR008116 (PRINTS), IPR006594 (PROSITE_PROFILES)	20
0	six1	7.7E-130	75.2	28	*Hymenolepis microstoma*	GO:0031981,GO:0003700,GO:0043565,GO:0044212, GO:0001657,GO:0003156,GO:0006357,GO:0008284, GO:0021545,GO:0042472,GO:0043066,GO:0043586, GO:0045165,GO:0045595,GO:0045893,GO:0048638, GO:0048646,GO:0048699,GO:0048704,GO:0048732, GO:0051179,GO:0051960,GO:0060537,GO:0061061, GO:0061213,GO:0072171,GO:0072358, GO:0090189	IPR001356 (SMART),IPR001356 (PFAM),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR10390:SF29 (PANTHER),PTHR10390 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES), IPR009057 (SUPERFAMILY)	20
0	t cell transcription factor 4 long c terminal	1.7E-78	77.25	19	*Echinococcus multilocularis*	GO:0005634,GO:0005667,GO:0003682,GO:0003700, GO:0008013,GO:0043565,GO:0044212,GO:0008595, GO:0030217,GO:0035019,GO:0043588,GO:0045892, GO:0046022,GO:0048319,GO:0048546,GO:0048562, GO:0048699,GO:0060070, GO:2000036	IPR009071 (SMART),IPR009071 (PFAM),IPR009071 (G3DSA:1.10.30.GENE3D),IPR024940 (PANTHER),IPR028782 (PTHR10373:PANTHER),IPR009071 (PROSITE_PROFILES), IPR009071 (SUPERFAMILY)	20

0	t-box transcription factor tbx2	0.0E0	77.85	26	*Echinococcus multilocularis*	GO:0005634,GO:0005667,GO:0000978,GO:0001078,GO:0005515,GO:0000122,GO:0003148,GO:0003203,GO:0003256,GO:0006351,GO:0007521,GO:0035050,GO:0035909,GO:0036302,GO:0042733,GO:0048596,GO:0048738,GO:0060021,GO:0060037,GO:0060045,GO:0060465,GO:0060560,GO:0060596,GO:0090398,GO:1901208, GO:1901211	IPR001699 (PRINTS),IPR001699 (SMART),IPR001699 (PFAM),IPR001699 (G3DSA:2.60.40.GENE3D),PTHR11267:SF23 (PANTHER),IPR001699 (PANTHER),IPR018186 (PROSITE_PATTERNS),IPR018186 (PROSITE_PATTERNS),SIGNAL_PEPTIDE_C_REGION (PHOBIUS),SIGNAL_PEPTIDE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),IPR001699 (PROSITE_PROFILES),SignalP-TM (SIGNALP_GRAM_POSITIVE), IPR008967 (SUPERFAMILY)	20
0	transcription factor gata-6	8.4E-41	74.85	30	*Echinococcus multilocularis*	GO:0005634,GO:0005667,GO:0000979,GO:0001103,GO:0003682,GO:0003705,GO:0008270,GO:0001701,GO:0001889,GO:0003309,GO:0003310,GO:0006366,GO:0006644,GO:0007493,GO:0014898,GO:0035239,GO:0042981,GO:0043627,GO:0045892,GO:0045944,GO:0048645,GO:0051145,GO:0055007,GO:0060045,GO:0060430,GO:0060486,GO:0060510,GO:0071371,GO:0071635, GO:0071773	IPR000679 (PRINTS),IPR000679 (SMART),IPR000679 (PFAM),IPR013088 (G3DSA:3.30.50.GENE3D),PTHR10071 (PANTHER),PTHR10071:SF165 (PANTHER),IPR000679 (PROSITE_PATTERNS),IPR000679 (PROSITE_PROFILES), SSF57716 (SUPERFAMILY)	20
0	transcription factor sox-14	4.5E-134	72.35	3	*Echinococcus granulosus*	GO:0003677,GO:0006355, GO:0007275	IPR009071 (PFAM),IPR009071 (G3DSA:1.10.30.GENE3D),IPR009071 (PROSITE_PROFILES), IPR009071 (SUPERFAMILY)	20
0	transcription factor sum-1	1.3E-180	86.3	5	*Echinococcus multilocularis*	GO:0005634,GO:0003677,GO:0046983,GO:0006355, GO:0007517	IPR002546 (SMART),IPR011598 (SMART),IPR011598 (G3DSA:4.10.280.GENE3D),IPR011598 (PFAM),IPR002546 (PFAM),PTHR11534 (PANTHER),PTHR11534:SF9 (PANTHER),IPR011598 (PROSITE_PROFILES), IPR011598 (SUPERFAMILY)	20

0	tyrosine protein kinase fes:fps	0.0E0	49.75	34	*Echinococcus multilocularis*	GO:0016020,GO:0043231,GO:0044444,GO:0071944, GO:0004713,GO:0005515,GO:0006468,GO:0007420, GO:0008284,GO:0008543,GO:0010468,GO:0016043, GO:0022612,GO:0030324,GO:0030850,GO:0035239, GO:0035272,GO:0043009,GO:0043085,GO:0043410, GO:0043583,GO:0045595,GO:0048523,GO:0048562, GO:0048589,GO:0048646,GO:0048666,GO:0048705, GO:0051094,GO:0051240,GO:0060485,GO:0061138, GO:0072358, GO:2000027	G3DSA:3.30.200.20 (GENE3D),IPR001245 (PFAM),G3DSA:1.10.510.10 (GENE3D),PTHR24418 (PANTHER),SIGNAL_PEPTIDE (PHOBIUS),SIGNAL_PEPTIDE_C_REGION (PHOBIUS),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),IPR000719 (PROSITE_PROFILES), IPR011009 (SUPERFAMILY)	20
0	vang-like protein 2	0.0E0	56.9	2	*Hymenolepis microstoma*	GO:0016021, GO:0007275	IPR009539 (PFAM),IPR009539 (PANTHER),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),TRANSMEMBRANE (PHOBIUS),TMhelix (TMHMM),TMhelix (TMHMM), TMhelix (TMHMM)	20
0	zinc finger c4h2 domain containing protein	2.6E-97	76.85	5	*s multilocularis*	GO:0016607,GO:0010172,GO:0016337,GO:0040025, GO:0048730	Coil (COILS),IPR018482 (PFAM),PTHR31058 (PANTHER),PTHR31058 (PANTHER), PTHR31058:SF2 (PANTHER)	20

0	zinc finger protein	3.1E-112	60.25	7	*Echinococcus multilocularis*	GO:0005634,GO:0005737,GO:0097159,GO:1901363, GO:0002119,GO:0006352, GO:0048665	IPR015880 (SMART),PF13465 (PFAM),IPR013087 (G3DSA:3.30.160.GENE3D),IPR013087 (G3DSA:3.30.160.GENE3D),PF13894 (PFAM),IPR007087 (PFAM),PTHR24409 (PANTHER),PTHR24409:SF14 (PANTHER),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PATTERNS),IPR007087 (PROSITE_PROFILES),IPR007087 (PROSITE_PROFILES),IPR007087 (PROSITE_PROFILES),IPR007087 (PROSITE_PROFILES),SSF57667 (SUPERFAMILY), SSF57667 (SUPERFAMILY)	20
0	zinc finger protein basonuclin-2	0.0E0	60.7	5	s *multiloculari* s	GO:0005654,GO:0005737,GO:0005886,GO:0046872, GO:0048066	IPR015880 (SMART),PTHR15021 (PANTHER),PTHR15021:SF0 (PANTHER),IPR007087 (PROSITE_PATTERNS), IPR007087 (PROSITE_PROFILES)	20

1	frizzled-partial	2.7E-65	61.35	4	*Echinococcus granulosus*	GO:0016021,GO:0042813,GO:0007275, GO:0016055	IPR000539 (PRINTS),IPR020067 (SMART),IPR020067 (PFAM),IPR000539 (PFAM),IPR020067 (G3DSA:1.10.2000.GENE3D),IPR015526 (PANTHER),IPR026554 (PTHR11309:PANTHER),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),IPR017981 (PROSITE_PROFILES),IPR020067 (PROSITE_PROFILES),IPR020067 (SUPERFAMILY),TMhelix (TMHMM),TMhelix (TMHMM),TMhelix (TMHMM),TMhelix (TMHMM), TMhelix (TMHMM)	20

1	mothers against decapentaplegic homolog 4-like	3.0E-138	83.05	64	*Echinococcus multilocularis*	GO:0000790,GO:0005654,GO:0005737,GO:0005813, GO:0032444,GO:0071141,GO:0000978,GO:0001076, GO:0001077,GO:0001085,GO:0003682,GO:0005518, GO:0030616,GO:0031005,GO:0042803,GO:0046872, GO:0046982,GO:0070411,GO:0070412,GO:0001658, GO:0001666,GO:0001701,GO:0001702,GO:0003190, GO:0003198,GO:0003251,GO:0003279,GO:0003360, GO:0006366,GO:0007179,GO:0007183,GO:0007411, GO:0007492,GO:0007498,GO:0008283,GO:0008285, GO:0010718,GO:0010862,GO:0014033,GO:0030308, GO:0030509,GO:0030511,GO:0030513,GO:0032525, GO:0032909,GO:0035556,GO:0036302,GO:0042118, GO:0042177,GO:0045892,GO:0045944,GO:0048589, GO:0048663,GO:0048733,GO:0048859,GO:0051098, GO:0051797,GO:0060021,GO:0060391,GO:0060395, GO:0060548,GO:0060956,GO:0072133, GO:0072134	Coil (COILS),IPR003619 (SMART),IPR003619 (PFAM),IPR013019 (G3DSA:3.90.520.GENE3D),IPR013790 (PANTHER),PTHR13703:SF19 (PANTHER),IPR013019 (PROSITE_PROFILES), IPR013019 (SUPERFAMILY)	20

1	protein jagged-1	5.3E-118	47.45	5	*Echinococcus granulosus*	GO:0007507,GO:0009653,GO:0009888,GO:0044763, GO:0045595	IPR001881 (SMART),IPR000742 (SMART),IPR000742 (PFAM),G3DSA:2.10.25.10 (GENE3D),G3DSA:2.10.25.10 (GENE3D),G3DSA:2.10.25.10 (GENE3D),G3DSA:2.10.25.10 (GENE3D),PTHR24033 (PANTHER),PTHR24033:SF0 (PANTHER),IPR013032 (PROSITE_PATTERNS),IPR013032 (PROSITE_PATTERNS),IPR013032 (PROSITE_PATTERNS),IPR013032 (PROSITE_PATTERNS),IPR013032 (PROSITE_PATTERNS),IPR013032 (PROSITE_PATTERNS),IPR013032 (PROSITE_PATTERNS),IPR000742 (PROSITE_PROFILES),IPR000742 (PROSITE_PROFILES),IPR000742 (PROSITE_PROFILES),IPR000742 (PROSITE_PROFILES),SSF57196 (SUPERFAMILY),SSF57196 (SUPERFAMILY), SSF57196 (SUPERFAMILY)	20
1	sh3 domain-containing kinase-binding protein 1	9.3E-161	56.35	9	*Echinococcus granulosus*	GO:0044464,GO:0005515,GO:0016301,GO:0016310, GO:0044707,GO:0044763,GO:0044767,GO:0048856, GO:0050794	Coil (COILS),IPR001452 (PRINTS),IPR001452 (SMART),IPR001452 (PFAM),G3DSA:2.30.30.40 (GENE3D),PTHR14167 (PANTHER),PTHR14167:SF19 (PANTHER),IPR001452 (PROSITE_PROFILES),IPR001452 (PROSITE_PROFILES),IPR001452 (SUPERFAMILY), IPR001452 (SUPERFAMILY)	20
2	atrial natriuretic peptide receptor 1	1.4E-159	80.8	21	*Echinococcus granulosus*	GO:0005622,GO:0005886,GO:0016021,GO:0004383, GO:0004672,GO:0004888,GO:0005524,GO:0005525, GO:0016941,GO:0017046,GO:0042802,GO:0006182, GO:0006468,GO:0007168,GO:0008217,GO:0035556, GO:0044702,GO:0051447,GO:0060348,GO:0097011, GO:1900194	IPR001054 (SMART),IPR001054 (PFAM),IPR001054 (G3DSA:3.30.70.GENE3D),PTHR11920 (PANTHER),PTHR11920:SF281 (PANTHER),IPR018297 (PROSITE_PATTERNS),IPR001054 (PROSITE_PROFILES), IPR029787 (SUPERFAMILY)	20

2	coiled-coil domain-containing protein partial	6.8E-73	53.65	3	*Echinococcus multilocularis*	GO:0044464,GO:0007368, GO:0048731	Coil (COILS),Coil (COILS),Coil (COILS),Coil (COILS),Coil (COILS),PTHR18962:SF0 (PANTHER), PTHR18962 (PANTHER)	20
2	dna-dependent protein kinase catalytic subunit	0.0E0	46.05	11	*Echinococcus multilocularis*	GO:0044424,GO:0016772,GO:0097159,GO:1901363, GO:0002520,GO:0006259,GO:0006974,GO:0030154, GO:0044710,GO:0048513, GO:0065007	IPR000403 (SMART),IPR000403 (G3DSA:1.10.1070.GENE3D),G3DSA:3.30.1010.10 (GENE3D),IPR000403 (PFAM),IPR012582 (PFAM),PTHR11139:SF54 (PANTHER),PTHR11139 (PANTHER),PTHR11139 (PANTHER),PTHR11139:SF54 (PANTHER),PTHR11139 (PANTHER),PTHR11139:SF54 (PANTHER),PTHR11139 (PANTHER),IPR017900 (PROSITE_PATTERNS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),IPR000403 (PROSITE_PROFILES),IPR017896 (PROSITE_PROFILES),IPR016024 (SUPERFAMILY), IPR011009 (SUPERFAMILY)	20
2	egfp:bcl2 fusion protein	3.4E-103	55.45	22	*Echinococcus multilocularis*	GO:0005815,GO:0031966,GO:0044428,GO:0070013, GO:0098588,GO:0042802,GO:0006810,GO:0008637, GO:0010033,GO:0022402,GO:0044702,GO:0044707, GO:0044767,GO:0048522,GO:0048856,GO:0051129, GO:0051704,GO:0051726,GO:0065008,GO:0097190, GO:1902589, GO:2001243	IPR026298 (PRINTS),SM00337 (SMART),IPR026298 (PFAM),G3DSA:1.10.437.10 (GENE3D),IPR002475 (PROSITE_PROFILES), SSF56854 (SUPERFAMILY)	20

2	mitogen-activated protein kinase 3 isoform x2	6.6E-141	72.55	61	*Hymenolepis microstoma*	GO:0005635,GO:0005654,GO:0005739,GO:0005769,GO:0005770,GO:0005794,GO:0005829,GO:0005901,GO:0005925,GO:0015630,GO:0070062,GO:0004707,GO:0005524,GO:0019902,GO:0000186,GO:0000187,GO:0002755,GO:0006361,GO:0006975,GO:0007173,GO:0007265,GO:0007411,GO:0008286,GO:0008543,GO:0030168,GO:0030509,GO:0031281,GO:0032872,GO:0033129,GO:0034134,GO:0034138,GO:0034142,GO:0034146,GO:0034162,GO:0034166,GO:0034605,GO:0035066,GO:0035666,GO:0038083,GO:0038095,GO:0038096,GO:0038123,GO:0038124,GO:0045087,GO:0045944,GO:0048010,GO:0048011,GO:0048513,GO:0051090,GO:0051403,GO:0051493,GO:0051704,GO:0060397,GO:0070374,GO:0070498,GO:0070849,GO:0071260,GO:0072584,GO:0090170,GO:1900034,GO:2000641	IPR002290 (SMART),G3DSA:3.30.200.20 (GENE3D),IPR000719 (PFAM),G3DSA:1.10.510.10 (GENE3D),PTHR24055 (PANTHER),PTHR24055:SF111 (PANTHER),IPR017441 (PROSITE_PATTERNS),IPR008271 (PROSITE_PATTERNS),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),IPR000719 (PROSITE_PROFILES), IPR011009 (SUPERFAMILY)	20
2	pou class transcription factor partial	0.0E0	84.2	19	*Echinococcus multilocularis*	GO:0005634,GO:0005667,GO:0001105,GO:0003700,GO:0043565,GO:0071837,GO:0006351,GO:0008284,GO:0021799,GO:0021869,GO:0022011,GO:0030216,GO:0043066,GO:0045892,GO:0045944,GO:0072218,GO:0072227,GO:0072233, GO:0072240	IPR013847 (PRINTS),IPR000327 (SMART),IPR001356 (SMART),IPR001356 (PFAM),IPR000327 (PFAM),IPR010982 (G3DSA:1.10.260.GENE3D),IPR009057 (G3DSA:1.10.10.GENE3D),PTHR11636 (PANTHER),PTHR11636:SF77 (PANTHER),IPR017970 (PROSITE_PATTERNS),IPR000327 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES),IPR000327 (PROSITE_PROFILES),IPR010982 (SUPERFAMILY), IPR009057 (SUPERFAMILY)	20

2	protocadherin fat 3	0.0E0	49.05	6	*Echinococcus multilocularis*	GO:0005886,GO:0016021,GO:0005509,GO:0007156, GO:0009653, GO:0048513	IPR002126 (SMART),IPR002126 (PFAM),IPR002126 (G3DSA:2.60.40.GENE3D),IPR002126 (G3DSA:2.60.40.GENE3D),PTHR24028 (PANTHER),PTHR24028:SF47 (PANTHER),IPR020894 (PROSITE_PATTERNS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),IPR002126 (PROSITE_PROFILES),IPR002126 (PROSITE_PROFILES),IPR015919 (SUPERFAMILY), TMhelix (TMHMM)	20
2	ring and yy1 binding protein	7.6E-51	71.0	7	s *multilocularis*	GO:0016604,GO:0003677,GO:0008270,GO:0010623, GO:0031146,GO:0043066, GO:0050777	IPR001876 (SMART),IPR001876 (PFAM),PTHR12920:SF1 (PANTHER),PTHR12920 (PANTHER),IPR001876 (PROSITE_PATTERNS),IPR001876 (PROSITE_PROFILES), SSF90209 (SUPERFAMILY)	20
2	tau tubulin kinase 1	0.0E0	67.95	6	*Hymenolepis microstoma*	GO:0005737,GO:0000166,GO:0004674,GO:0008360, GO:0018105, GO:0021762	IPR002290 (SMART),G3DSA:3.30.200.20 (GENE3D),IPR000719 (PFAM),G3DSA:1.10.510.10 (GENE3D),PTHR11909 (PANTHER),PTHR11909:SF19 (PANTHER),PTHR11909 (PANTHER),PTHR11909:SF19 (PANTHER),IPR008271 (PROSITE_PATTERNS),IPR000719 (PROSITE_PROFILES), IPR011009 (SUPERFAMILY)	20

2	transcription factor ap 2 gamma	2.2E-165	69.75	56	*Hymenolepis microstoma*	GO:0005654,GO:0005667,GO:0005794,GO:0005813,GO:0005829,GO:0000978,GO:0000979,GO:0000980,GO:0001077,GO:0001078,GO:0001105,GO:0001106,GO:0003682,GO:0008134,GO:0042803,GO:0000122,GO:0001822,GO:0003151,GO:0003334,GO:0003404,GO:0003409,GO:0006366,GO:0007605,GO:0008285,GO:0009880,GO:0010172,GO:0010842,GO:0010944,GO:0014032,GO:0021506,GO:0021559,GO:0021623,GO:0021884,GO:0030335,GO:0030501,GO:0032496,GO:0035115,GO:0042059,GO:0042472,GO:0043524,GO:0043525,GO:0045664,GO:0045944,GO:0048485,GO:0048701,GO:0048730,GO:0060021,GO:0060235,GO:0060325,GO:0060349,GO:0061029,GO:0061303,GO:0070172,GO:0071281,GO:0071711, GO:2000378	IPR013854 (PRINTS),IPR013854 (PFAM), IPR004979 (PANTHER)	20
3	histone deacetylase 7	0.0E0	67.25	46	*Echinococcus granulosus*	GO:0000118,GO:0017053,GO:0043232,GO:0044444,GO:0001025,GO:0001047,GO:0003682,GO:0003714,GO:0005080,GO:0008270,GO:0030955,GO:0032041,GO:0033613,GO:0042826,GO:0043565,GO:0046969,GO:0046970,GO:0070491,GO:0071889,GO:0097372,GO:0000122,GO:0001570,GO:0006338,GO:0006954,GO:0007043,GO:0007165,GO:0007399,GO:0008284,GO:0010832,GO:0014898,GO:0016202,GO:0030183,GO:0032703,GO:0033235,GO:0034983,GO:0040029,GO:0043393,GO:0043433,GO:0045668,GO:0045944,GO:0051091,GO:0061647,GO:0070555,GO:0070933,GO:0090050, GO:1990619	IPR000286 (PRINTS),IPR023801 (G3DSA:3.40.800.GENE3D),IPR023801 (PFAM),IPR000286 (PANTHER),PTHR10625:SF107 (PANTHER),PTHR10625:SF107 (PANTHER),IPR000286 (PANTHER),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS), SSF52768 (SUPERFAMILY)	20

3	low quality protein: x-box-binding protein 1	1.7E-33	60.0	33	*Echinococcus granulosus*	GO:0005634,GO:0044444,GO:0000981,GO:0019899,GO:0043565,GO:0044212,GO:0001889,GO:0002699,GO:0006357,GO:0006366,GO:0009605,GO:0009725,GO:0009888,GO:0010557,GO:0030968,GO:0031329,GO:0031401,GO:0035188,GO:0042493,GO:0045621,GO:0048468,GO:0048584,GO:0048785,GO:0051047,GO:0051222,GO:0051602,GO:0055088,GO:0060096,GO:0060341,GO:0071236,GO:0071417,GO:1901701,GO:1902236	Coil (COILS),IPR004827 (SMART),G3DSA:1.20.5.170 (GENE3D),IPR004827 (PFAM),PTHR13301 (PANTHER),IPR004827 (PROSITE_PROFILES), SSF57959 (SUPERFAMILY)	20
3	lysine-specific histone demethylase partial	2.1E-174	54.9	17	*Echinococcus granulosus*	GO:0031981,GO:0032991,GO:0008134,GO:0032453,GO:0097159,GO:1901363,GO:0006357,GO:0044707,GO:0044767,GO:0045892,GO:0045893,GO:0048856,GO:0051090,GO:0051100,GO:0070076,GO:1901797,GO:2001021	IPR002937 (PFAM),G3DSA:1.10.287.80 (GENE3D),G3DSA:3.90.660.10 (GENE3D),G3DSA:3.90.660.10 (GENE3D),IPR011991 (G3DSA:1.10.10.GENE3D),G3DSA:3.50.50.60 (GENE3D),PTHR10742 (PANTHER),PTHR10742 (PANTHER),PTHR10742:SF245 (PANTHER),PTHR10742:SF245 (PANTHER),PTHR10742 (PANTHER),PTHR10742 (PANTHER),PTHR10742:SF245 (PANTHER),PTHR10742 (PANTHER),PTHR10742:SF245 (PANTHER),TRANSMEMBRANE (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),IPR009057 (SUPERFAMILY),SSF54373 (SUPERFAMILY),SSF51905 (SUPERFAMILY), SSF51905 (SUPERFAMILY)	20
3	nadh dehydrogenase	0.0E0	63.05	1	*Echinococcus granulosus*	GO:0043009	IPR003788 (PFAM),PTHR12049:SF5 (PANTHER),IPR003788 (PANTHER), IPR029063 (SUPERFAMILY)	20

3	serum response factor	1.2E-73	79.3	70	*Echinococcus granulosus*	GO:0000790,GO:0005654,GO:0005737,GO:0000978,GO:0000983,GO:0001076,GO:0001077,GO:0003705,GO:0008134,GO:0010736,GO:0031490,GO:0042803,GO:0001569,GO:0001666,GO:0001707,GO:0001764,GO:0001829,GO:0001947,GO:0002011,GO:0002042,GO:0003257,GO:0007160,GO:0007264,GO:0007616,GO:0008285,GO:0008306,GO:0009636,GO:0009725,GO:0010669,GO:0010735,GO:0021766,GO:0022028,GO:0030038,GO:0030155,GO:0030168,GO:0030220,GO:0030336,GO:0031175,GO:0033561,GO:0034097,GO:0035855,GO:0035912,GO:0042789,GO:0043149,GO:0043589,GO:0045059,GO:0045214,GO:0045773,GO:0045987,GO:0046016,GO:0046716,GO:0048589,GO:0048821,GO:0051091,GO:0051150,GO:0051491,GO:0055003,GO:0060055,GO:0060218,GO:0060261,GO:0060292,GO:0060347,GO:0060947,GO:0061029,GO:0070830,GO:0071333,GO:0090009,GO:0090136,GO:0090398, GO:1900222	IPR002100 (PRINTS),IPR002100 (SMART),IPR002100 (PFAM),PTHR11945 (PANTHER),PTHR11945:SF32 (PANTHER),IPR002100 (PROSITE_PROFILES), IPR002100 (SUPERFAMILY)	20
3	transcription factor coe4 isoform x1	0.0E0	85.7	9	*Hymenolepis microstoma*	GO:0005634,GO:0000977,GO:0001228,GO:0008134,GO:0046872,GO:0046983,GO:0006366,GO:0007275,GO:0045944	Coil (COILS),IPR002909 (SMART),IPR013783 (G3DSA:2.60.40.GENE3D),IPR002909 (PFAM),IPR003523 (PANTHER),IPR018350 (PROSITE_PATTERNS), IPR014756 (SUPERFAMILY)	20
4	acidic fibroblast growth factor intracellular-binding protein	3.0E-169	63.3	6	*Echinococcus granulosus*	GO:0005634,GO:0017134,GO:0007368,GO:0060026,GO:0060271, GO:0070121	Coil (COILS),IPR008614 (PFAM),PTHR13223:SF2 (PANTHER),IPR008614 (PANTHER),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),TMhelix (TMHMM), TMhelix (TMHMM)	20

4	activin receptor type-1	0.0E0	58.15	22	*Echinococcus granulosus*	GO:0005887,GO:0000166,GO:0004675,GO:0019838,GO:0043167,GO:0003007,GO:0006357,GO:0006468,GO:0007178,GO:0009968,GO:0042692,GO:0045778,GO:0045893,GO:0048598,GO:0051094,GO:0060485,GO:0060562,GO:0060911,GO:0071310,GO:0071495,GO:0090092, GO:2000026	IPR002290 (SMART),IPR003605 (SMART),IPR000719 (PFAM),G3DSA:3.30.200.20 (GENE3D),G3DSA:1.10.510.10 (GENE3D),IPR003605 (PFAM),PTHR23255:SF14 (PANTHER),IPR000333 (PANTHER),IPR017441 (PROSITE_PATTERNS),IPR008271 (PROSITE_PATTERNS),CYTOPLASMIC_DOMAIN (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),IPR003605 (PROSITE_PROFILES),IPR000719 (PROSITE_PROFILES),SignalP-TM (SIGNALP_GRAM_POSITIVE),IPR011009 (SUPERFAMILY), TMhelix (TMHMM)	20
4	calmodulin	4.1E-94	97.4	73	*Trichuris trichiura*	GO:0000922,GO:0005654,GO:0005813,GO:0005829,GO:0005876,GO:0005886,GO:0030017,GO:0030426,GO:0034704,GO:0070062,GO:0004689,GO:0005246,GO:0005509,GO:0008179,GO:0017022,GO:0019901,GO:0019904,GO:0030235,GO:0031432,GO:0031800,GO:0031996,GO:0031997,GO:0043274,GO:0043539,GO:0043548,GO:0044325,GO:0048306,GO:0050998,GO:0072542,GO:0000086,GO:0001975,GO:0002027,GO:0002223,GO:0002576,GO:0005513,GO:0005980,GO:0006006,GO:0006468,GO:0006936,GO:0007173,GO:0007190,GO:0007202,GO:0007264,GO:0007268,GO:0008543,GO:0010800,GO:0010801,GO:0010881,GO:0016056,GO:0021762,GO:0022400,GO:0030168,GO:0031954,GO:0032465,GO:0032516,GO:0035307,GO:0038095,GO:0043388,GO:0043647,GO:0045087,GO:0046209,GO:0048010,GO:0048011,GO:0051000,GO:0051343,GO:0051412,GO:0060315,GO:0060316,GO:0061024,GO:0071902,GO:1901339,GO:1901841,GO:1901844	IPR002048 (SMART),IPR011992 (G3DSA:1.10.238.GENE3D),IPR011992 (PFAM),IPR011992 (G3DSA:1.10.238.GENE3D),PTHR23050:SF155 (PANTHER),PTHR23050 (PANTHER),IPR018247 (PROSITE_PATTERNS),IPR002048 (PROSITE_PROFILES), SSF47473 (SUPERFAMILY)	20

4	cyclin-g-associated kinase	0.0E0	68.2	10	*Echinococcus multilocularis*	GO:0005794,GO:0016020,GO:0000166,GO:0004674, GO:0005515,GO:0006464,GO:0016043,GO:0016310, GO:0044763, GO:0048513	Coil (COILS),IPR002290 (SMART),IPR001623 (G3DSA:1.10.287.GENE3D),G3DSA:1.10.510.10 (GENE3D),G3DSA:3.30.200.20 (GENE3D),G3DSA:2.60.40.1110 (GENE3D),IPR014020 (PFAM),IPR029021 (G3DSA:3.90.190.GENE3D),IPR000719 (PFAM),PTHR23172 (PANTHER),PTHR23172:SF19 (PANTHER),IPR008271 (PROSITE_PATTERNS),IPR029023 (PROSITE_PROFILES),IPR000719 (PROSITE_PROFILES),IPR014020 (PROSITE_PROFILES),IPR011009 (SUPERFAMILY),IPR029021 (SUPERFAMILY),IPR000008 (SUPERFAMILY), IPR001623 (SUPERFAMILY)	20
4	forkhead box protein d2	1.9E-74	86.2	13	*Echinococcus multilocularis*	GO:0000790,GO:0000977,GO:0001227,GO:0000122, GO:0001755,GO:0001829,GO:0001892,GO:0006366, GO:0007398,GO:0007498,GO:0030336,GO:0030900, GO:0045944	IPR001766 (PRINTS),IPR001766 (SMART),IPR001766 (PFAM),IPR011991 (G3DSA:1.10.10.GENE3D),PTHR11829 (PANTHER),PTHR11829:SF85 (PANTHER),IPR018122 (PROSITE_PATTERNS),PS00658 (PROSITE_PATTERNS),IPR001766 (PROSITE_PROFILES), SSF46785 (SUPERFAMILY)	20
4	groucho protein	0.0E0	62.7	11	*Echinococcus granulosus*	GO:0005730,GO:0005829,GO:0001106,GO:0008134, GO:0042802,GO:0007166,GO:0007275,GO:0010628, GO:0043124,GO:0045892, GO:2000811	IPR001680 (SMART),IPR005617 (PFAM),IPR015943 (G3DSA:2.130.10.GENE3D),IPR001680 (PFAM),PTHR10814:SF22 (PANTHER),PTHR10814 (PANTHER),IPR019775 (PROSITE_PATTERNS),IPR017986 (PROSITE_PROFILES),IPR001680 (PROSITE_PROFILES), IPR017986 (SUPERFAMILY)	20

4	intraflagellar transport protein 52	0.0E0	77.2	11	*Echinococcus granulosus*	GO:0005814,GO:0030992,GO:0031514,GO:0044441, GO:0007368,GO:0042073,GO:0048598,GO:0048731, GO:0050794,GO:0060271, GO:0060562	Coil (COILS),IPR019196 (PFAM),PTHR12969:SF6 (PANTHER),PTHR12969 (PANTHER),SIGNAL_PEPTIDE_C_REGION (PHOBIUS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS), SIGNAL_PEPTIDE (PHOBIUS)	20
4	jumonji domain containing 1a	0.0E0	50.7	15	*Echinococcus granulosus*	GO:0005488,GO:0016706,GO:0006325,GO:0010628, GO:0031323,GO:0032259,GO:0032501,GO:0044237, GO:0044702,GO:0044710,GO:0044763,GO:0044767, GO:0050896,GO:0071704, GO:0080090	IPR003347 (SMART),IPR003347 (PFAM),PTHR12549 (PANTHER), SSF51197 (SUPERFAMILY)	20
4	kinesin heavy chain	0.0E0	67.65	30	*Echinococcus granulosus*	GO:0005739,GO:0005871,GO:0035371,GO:0005524, GO:0008017,GO:0008574,GO:0001754,GO:0006886, GO:0007154,GO:0007303,GO:0007310,GO:0007315, GO:0007317,GO:0007411,GO:0008088,GO:0008103, GO:0008152,GO:0008345,GO:0030011,GO:0040023, GO:0044700,GO:0046785,GO:0046843,GO:0047497, GO:0048312,GO:0048741,GO:0048813,GO:0051012, GO:0051299, GO:0061572	Coil (COILS),Coil (COILS),Coil (COILS),Coil (COILS),Coil (COILS),Coil (COILS),IPR001752 (PRINTS),IPR001752 (SMART),IPR001752 (G3DSA:3.40.850.GENE3D),IPR001752 (PFAM),IPR027640 (PANTHER),PTHR24115:SF380 (PANTHER),IPR027640 (PANTHER),PTHR24115:SF380 (PANTHER),IPR019821 (PROSITE_PATTERNS),IPR001752 (PROSITE_PROFILES), IPR027417 (SUPERFAMILY)	20
4	methyltransferase-like protein 8	6.6E-98	61.9	11	*s multilocularis*	GO:0005634,GO:0005737,GO:0004402,GO:0008170, GO:0008171,GO:0008175,GO:0008276,GO:0008649, GO:0007519,GO:0016573, GO:0045444	PF13489 (PFAM),IPR029063 (G3DSA:3.40.50.GENE3D),IPR026113 (PANTHER), IPR029063 (SUPERFAMILY)	20

4	mothers against decapentaplegic homolog 3 isoform x3	0.0E0	69.6	95	*Echinococcus granulosus*	GO:0000790,GO:0005637,GO:0005654,GO:0005829,GO:0005886,GO:0043235,GO:0071144,GO:0000978,GO:0000983,GO:0000988,GO:0001102,GO:0003690,GO:0005160,GO:0005518,GO:0008013,GO:0008270,GO:0019901,GO:0019902,GO:0030618,GO:0031490,GO:0031625,GO:0035326,GO:0042803,GO:0043130,GO:0043425,GO:0046982,GO:0070410,GO:0070412,GO:0000122,GO:0001666,GO:0001701,GO:0001707,GO:0001756,GO:0001889,GO:0001933,GO:0001947,GO:0002076,GO:0002520,GO:0006367,GO:0006810,GO:0006955,GO:0007050,GO:0007179,GO:0007183,GO:0007492,GO:0009880,GO:0010694,GO:0010718,GO:0016202,GO:0019049,GO:0023019,GO:0030308,GO:0030335,GO:0030501,GO:0030512,GO:0030878,GO:0031053,GO:0032332,GO:0032731,GO:0032909,GO:0032916,GO:0033689,GO:0035413,GO:0035556,GO:0038092,GO:0042060,GO:0042110,GO:0042177,GO:0042993,GO:0043066,GO:0045216,GO:0045668,GO:0045930,GO:0045944,GO:0048340,GO:0048589,GO:0048617,GO:0048701,GO:0050678,GO:0050728,GO:0050776,GO:0050821,GO:0050927,GO:0051098,GO:0051496,GO:0051894,GO:0060039,GO:0060290,GO:0060395,GO:0061045,GO:0070306,GO:0090263,GO:0097191,GO:0097296, GO:1901203	IPR003619 (SMART),IPR001132 (SMART),IPR001132 (PFAM),IPR013019 (G3DSA:3.90.520.GENE3D),IPR003619 (PFAM),IPR017855 (G3DSA:2.60.200.GENE3D),IPR013790 (PANTHER),PTHR13703:SF30 (PANTHER),IPR013790 (PANTHER),PTHR13703:SF30 (PANTHER),PTHR13703:SF30 (PANTHER),IPR013019 (PROSITE_PROFILES),IPR001132 (PROSITE_PROFILES),IPR013019 (SUPERFAMILY), IPR008984 (SUPERFAMILY)	20
4	nanos protein	3.1E-56	72.65	3	s *multiloculari* s	GO:0003723,GO:0008270, GO:0042462	IPR024161 (PFAM),PTHR12887:SF2 (PANTHER),IPR008705 (PANTHER), IPR024161 (PROSITE_PROFILES)	20

4	nhp2 non-histone chromosome protein 2-like 1	3.6E-83	87.85	7	*Echinococcus multilocularis*	GO:0005730,GO:0030529,GO:0003723,GO:0005525, GO:0007264,GO:0042254, GO:0061053	IPR018492 (PRINTS),IPR002415 (PRINTS),IPR029064 (G3DSA:3.30.1330.GENE3D),IPR004038 (PFAM),PTHR23105 (PANTHER),IPR004037 (PROSITE_PATTERNS), IPR029064 (SUPERFAMILY)	20
4	nucleoside diphosphate kinase homolog 5	5.9E-93	59.1	7	*Echinococcus granulosus*	GO:0004550,GO:0005524,GO:0006165,GO:0006183, GO:0006228,GO:0006241, GO:0044767	IPR001564 (SMART),IPR001564 (G3DSA:3.30.70.GENE3D),IPR007858 (PFAM),IPR001564 (PFAM),PTHR11349 (PANTHER),PTHR11349:SF52 (PANTHER), IPR001564 (SUPERFAMILY)	20
4	peptidyl-prolyl cis-trans isomerase fkbp8	1.5E-91	57.0	6	*Echinococcus granulosus*	GO:0043231,GO:0044444,GO:0044267,GO:0044707, GO:0044767, GO:0050789	IPR001179 (PFAM),IPR011990 (G3DSA:1.25.40.GENE3D),G3DSA:3.10.50.40 (GENE3D),IPR023566 (PANTHER),PTHR10516:SF290 (PANTHER),IPR001179 (PROSITE_PROFILES),SSF54534 (SUPERFAMILY), SSF48452 (SUPERFAMILY)	20
4	pou class transcription factor 2	0.0E0	83.7	21	*Echinococcus granulosus*	GO:0016607,GO:0000978,GO:0001077,GO:0001078, GO:0003682,GO:0046872,GO:0000122,GO:0000165, GO:0006351,GO:0007605,GO:0021562,GO:0030520, GO:0031290,GO:0042472,GO:0042491,GO:0045597, GO:0045944,GO:0048675,GO:0050885,GO:0051402, GO:0060041	IPR013847 (PRINTS),IPR000327 (SMART),IPR001356 (SMART),IPR009057 (G3DSA:1.10.10.GENE3D),IPR000327 (PFAM),IPR010982 (G3DSA:1.10.260.GENE3D),IPR001356 (PFAM),PTHR11636 (PANTHER),PTHR11636:SF42 (PANTHER),IPR000327 (PROSITE_PATTERNS),IPR000327 (PROSITE_PATTERNS),IPR017970 (PROSITE_PATTERNS),IPR001356 (PROSITE_PROFILES),IPR000327 (PROSITE_PROFILES),IPR010982 (SUPERFAMILY), IPR009057 (SUPERFAMILY)	20
4	pre-mrna-splicing regulator female-lethal d	4.6E-125	64.5	7	*Echinococcus granulosus*	GO:0005634,GO:0005515,GO:0000375,GO:0000381, GO:0007530,GO:0046331, GO:0048749	Coil (COILS),Coil (COILS),PTHR15217 (PANTHER), IPR029732 (PTHR15217:PANTHER)	20

4	protein prenyltransferase alpha subunit	2.1E-99	55.05	4	*Echinococcus granulosus*	GO:0005968,GO:0004663,GO:0007423, GO:0018344	G3DSA:1.25.40.120 (GENE3D),IPR002088 (PFAM),PTHR11129 (PANTHER),PTHR11129:SF3 (PANTHER), SSF48439 (SUPERFAMILY)	20
4	protein tilb homolog	9.9E-121	59.7	10	*Echinococcus granulosus*	GO:0005737,GO:0005515,GO:0001947,GO:0003146, GO:0003351,GO:0009953,GO:0044458,GO:0048793, GO:0060027, GO:0060294	IPR003603 (SMART),SM00365 (SMART),G3DSA:3.80.10.10 (GENE3D),PTHR10588:SF114 (PANTHER),PTHR10588 (PANTHER),IPR001611 (PROSITE_PROFILES),IPR001611 (PROSITE_PROFILES),IPR001611 (PROSITE_PROFILES),IPR001611 (PROSITE_PROFILES), SSF52058 (SUPERFAMILY)	20
4	ras-related protein rab-11a	7.0E-136	81.35	12	*Echinococcus multilocularis*	GO:0016020,GO:0055037,GO:0003924,GO:0005525, GO:0001525,GO:0006886,GO:0006887,GO:0006913, GO:0008152,GO:0032402,GO:0032482, GO:0070121	IPR001806 (PRINTS),IPR002041 (SMART),IPR003578 (SMART),IPR003579 (SMART),IPR020849 (SMART),IPR027417 (G3DSA:3.40.50.GENE3D),IPR005225 (TIGRFAM),IPR001806 (PFAM),PTHR24073 (PANTHER),PTHR24073:SF335 (PANTHER),PS51419 (PROSITE_PROFILES), IPR027417 (SUPERFAMILY)	20
4	receptor-type tyrosine-protein phosphatase c	0.0E0	53.3	32	*Echinococcus granulosus*	GO:0005622,GO:0009986,GO:0044459,GO:0004725, GO:0005515,GO:0043168,GO:0097367,GO:1901681, GO:0001914,GO:0001933,GO:0002711,GO:0002891, GO:0002923,GO:0006470,GO:0006928,GO:0009605, GO:0032879,GO:0040011,GO:0042102,GO:0042113, GO:0043410,GO:0045061,GO:0045582,GO:0045860, GO:0046651,GO:0048468,GO:0048585,GO:0050851, GO:0050854,GO:0050871,GO:0051179, GO:2001236	IPR003961 (SMART),IPR000242 (PFAM),IPR029021 (G3DSA:3.90.190.GENE3D),IPR013783 (G3DSA:2.60.40.GENE3D),PTHR19134:SF81 (PANTHER),PTHR19134:SF81 (PANTHER),PTHR19134 (PANTHER),PTHR19134:SF81 (PANTHER),PTHR19134 (PANTHER),CYTOPLASMIC_DOMAIN (PHOBIUS),TRANSMEMBRANE (PHOBIUS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),IPR000242 (PROSITE_PROFILES),IPR029021 (SUPERFAMILY),IPR003961 (SUPERFAMILY), TMhelix (TMHMM)	20

4	rna binding protein musashi rbp6	2.5E-114	73.75	7	*Schistosoma japonicum*	GO:0005737,GO:0005844,GO:0000166,GO:0005515, GO:0008266,GO:0044822, GO:0048864	IPR000504 (SMART),IPR000504 (PFAM),IPR012677 (G3DSA:3.30.70.GENE3D),PTHR24012:SF324 (PANTHER),PTHR24012 (PANTHER),IPR000504 (PROSITE_PROFILES),IPR000504 (PROSITE_PROFILES),SSF54928 (SUPERFAMILY), SSF54928 (SUPERFAMILY)	20
4	sphingosine-1-phosphate lyase 1	5.2E-107	76.5	13	*Echinococcus granulosus*	GO:0008117,GO:0016831,GO:0030170,GO:0006807, GO:0007165,GO:0008202,GO:0008585,GO:0009653, GO:0019752,GO:0030154,GO:0034754,GO:0044255, GO:0048609	IPR015422 (G3DSA:3.90.1150.GENE3D),PTHR11999 (PANTHER),PTHR11999:SF63 (PANTHER), IPR015424 (SUPERFAMILY)	20
4	succinate dehydrogenase	0.0E0	87.8	8	*ubiquinone*	GO:0005749,GO:0005515,GO:0008177,GO:0050660, GO:0006099,GO:0006105,GO:0007399, GO:0022904	PIRSF000171 (PIRSF),IPR014006 (TIGRFAM),G3DSA:3.50.50.60 (GENE3D),IPR011281 (TIGRFAM),G3DSA:4.10.80.40 (GENE3D),IPR003953 (PFAM),IPR015939 (PFAM),IPR027477 (G3DSA:3.90.700.GENE3D),IPR015939 (G3DSA:1.20.58.GENE3D),PTHR11632:SF37 (PANTHER),PTHR11632 (PANTHER),IPR003952 (PROSITE_PATTERNS),NON_CYTOPLASMIC_DOMAIN (PHOBIUS),SIGNAL_PEPTIDE_H_REGION (PHOBIUS),SIGNAL_PEPTIDE_N_REGION (PHOBIUS),SIGNAL_PEPTIDE (PHOBIUS),SIGNAL_PEPTIDE_C_REGION (PHOBIUS),SignalP-TM (SIGNALP_GRAM_POSITIVE),SignalP-noTM (SIGNALP_GRAM_NEGATIVE),SSF51905 (SUPERFAMILY),IPR027477 (SUPERFAMILY), IPR015939 (SUPERFAMILY)	20

4	t-box transcription factor tbx2	5.7E-172	84.75	49	*Echinococcus granulosus*	GO:0005634,GO:0005667,GO:0000978,GO:0001078, GO:0001102,GO:0000122,GO:0001501,GO:0001701, GO:0001947,GO:0003148,GO:0003167,GO:0003203, GO:0003256,GO:0006366,GO:0007219,GO:0007521, GO:0008595,GO:0010159,GO:0019827,GO:0021761, GO:0030539,GO:0030540,GO:0030857,GO:0032275, GO:0035115,GO:0035909,GO:0036302,GO:0042733, GO:0043066,GO:0045662,GO:0045787,GO:0045893, GO:0046884,GO:0048332,GO:0048596,GO:0060021, GO:0060037,GO:0060045,GO:0060412,GO:0060444, GO:0060465,GO:0060560,GO:0060596,GO:0060923, GO:0060931,GO:0090398,GO:1901208,GO:1901211, GO:2000648	IPR001699 (PRINTS),IPR002070 (PRINTS),IPR001699 (SMART),IPR001699 (G3DSA:2.60.40.GENE3D),IPR001699 (PFAM),IPR001699 (PANTHER),PTHR11267:SF23 (PANTHER),IPR018186 (PROSITE_PATTERNS),IPR018186 (PROSITE_PATTERNS),IPR001699 (PROSITE_PROFILES), IPR008967 (SUPERFAMILY)	20
4	transcription factor 12	7.7E-68	85.3	56	*Echinococcus multilocularis*	GO:0000788,GO:0005654,GO:0005737,GO:0090575, GO:0000978,GO:0000980,GO:0001077,GO:0001078, GO:0001205,GO:0003682,GO:0003713,GO:0008013, GO:0030165,GO:0031435,GO:0035497,GO:0042803, GO:0043425,GO:0046332,GO:0046982,GO:0070491, GO:0070644,GO:0070888,GO:0071837,GO:0000122, GO:0001779,GO:0002088,GO:0002326,GO:0006366, GO:0006955,GO:0007369,GO:0007517,GO:0008595, GO:0021986,GO:0030890,GO:0032496,GO:0033077, GO:0033152,GO:0035019,GO:0035462,GO:0042493, GO:0043588,GO:0043966,GO:0043967,GO:0044333, GO:0045666,GO:0045787,GO:0046022,GO:0048319, GO:0048541,GO:0050821,GO:0051091,GO:0060070, GO:0060729,GO:2000036,GO:2000045, GO:2001237	IPR011598 (SMART),IPR011598 (G3DSA:4.10.280.GENE3D),IPR011598 (PFAM),PTHR11793 (PANTHER),IPR011598 (PROSITE_PROFILES), IPR011598 (SUPERFAMILY)	20
4	ubiquinone biosynthesis protein coq7 homolog	2.5E-58	74.45	12	*Strongylocentrotus purpuratus*	GO:0005634,GO:0005743,GO:0004497,GO:0001306, GO:0001701,GO:0001841,GO:0006744,GO:0008340, GO:0022008,GO:0034599,GO:0042775, GO:0070584	IPR012347 (G3DSA:1.20.1260.GENE3D),IPR011566 (PFAM),PTHR11237:SF1 (PANTHER),IPR011566 (PANTHER), IPR009078 (SUPERFAMILY)	20

4	ubiquitin-like modifier-activating enzyme atg7	0.0E0	59.9	15	*s multilocularis*	GO:0005737,GO:0008641,GO:0006464,GO:0006914, GO:0006996,GO:0007417,GO:0009267,GO:0009893, GO:0010506,GO:0044712,GO:0048468,GO:0048513, GO:0048522,GO:0048523, GO:0051246	IPR000594 (PFAM),IPR016040 (G3DSA:3.40.50.GENE3D),PTHR10953 (PANTHER),IPR006285 (PTHR10953:PANTHER), IPR009036 (SUPERFAMILY)	20
4	voltage-dependent l-type calcium channel subunit beta-2	0.0E0	80.25	6	*Taenia solium*	GO:0005891,GO:0008331,GO:0007268,GO:0007528, GO:0070588, GO:1901385	Coil (COILS),IPR000584 (PRINTS),IPR008145 (SMART),IPR008145 (PFAM),G3DSA:2.30.30.40 (GENE3D),IPR027417 (G3DSA:3.40.50.GENE3D),IPR000584 (PANTHER),PTHR11824:SF5 (PANTHER),IPR001452 (SUPERFAMILY), IPR027417 (SUPERFAMILY)	20
4	zinc finger protein zpr1	0.0E0	62.95	8	*Echinococcus granulosus*	GO:0031981,GO:0097458,GO:0008270,GO:0007275, GO:0044763,GO:0048518,GO:0048856, GO:0050794	IPR004457 (SMART),IPR004457 (PFAM),IPR004457 (TIGRFAM),PTHR10876 (PANTHER), PTHR10876:SF0 (PANTHER)	20

APÊNDICE 20: SUPPLEMENTARY FILE 17

Supplementary File 17. Orthologous identification and alignments.

Protein	Organism	Identification	Codon alignment	Protein alignment
Bone morphogenetic protein 2	*Azumapecten farreri*	gi|451782459|gb|AGF68558.1| bone morphogenetic protein 2 [Azumapecten farreri]	AGAAGTAGT------------------------GGTAGA---------AAGTCTAAACGGAAGAGACCA---CGCAACCGGAAGAGGCGG- ---CGACATAAGAAGTAC--------- TGTAGGAGGAAACCGTTGTATGTAGACTTTACAGCCGTAGGCTGGACAGACTGGATTTTTGCGCCTCCGGGC TATCAGGCATATTATTGTCAGGGCGAATGTGAGTTTCCGTTTTCGGAGCACATGAACGCTACAAATCATGCAAT AGTGCAGGACTTGGTGAATTCGATAGAC------ TCCAAGTCCGTACCCAAACCGTGCTGTGTACCTACAGAACTGAGTCCTATTTCACTTCTCTATGTGGACGAGTA TGAAAAAGTGATACTGAAGAGCTACCAAAACATGACCGTCGTGAGCTGCGGCTGTCGG----------------------	RSS--------GR---KSKRKRP-RNRKRR--------- -------RHKKY--- CRRKPLYVDFTAVGWTDWIFAPPGYQAYY CQGECEFPFSEHMNATNHAIVQDLVNSID- - SKSVPKPCCVPTELSPISLLYVDEYEKVILK SYQNMTVVSCGCR-------
	Crassostrea gigas	gi|405950786|gb|EKC18750.1| Bone morphogenetic protein 2 [Crassostrea gigas]	CGAGCAACT------------------------AGTGAT---------AAAAAAGTGAAAAAGAATAAG--- AAAAATAGAAAAAACAAAAATAAGCGAAGGAAAAATAGG--- AAGAAAAATAGAAAAAATAAAACTAAGAGGAAAAAGTATAACAACCAGTGTCGTAGGAAGGAATTAAATGTAGA CTTCAAAGCCGTGGGTTGGAACGATTGGATATTCGCGCCACCCGGTTATAATGCGTATTATTGCGATGGTTCG TGTCATTGGCCGTACGATGACCACATGAATGTGACCAATCACGCAATAGTCCAAGACTTAGTGAACTCCATTG AC------ CCTAGGGCAGCCCCAAAGCCCTGCTGTGTACCCACAGAACTCAGTTCTTTGTCCTTGTTATATACTGACGAAC ACGGCGCGGTGGTGCTGAAAGTTTATCAAGACATGGTAGTAGAAGGCTGTGGTTGCCGG---------------------	RAT--------SD---KKVKKNK- KNRKNKNKRRKNR- KKNRKNKTKRKKYNNQCRRKELNVDFKAV GWNDWIFAPPGYNAYYCDGSCHWPYDD HMNVTNHAIVQDLVNSID-- PRAAPKPCCVPTELSSLSLLYTDEHGAVVL KVYQDMVVEGCGCR-------
	Echinococcus granulosus	gi|674566209|emb|CDS19307.1| bone morphogenetic protein 2 [Echinococcus granulosus]	AAACCAATT------------------------GGTGAA---------AATCGCAAAAAGCGTCGACGAACACGT--- CTTAAAAGCAAAAGTTGG--- ACAAATAATCGGGAAAAGCGCAATTCGGGATATCTCATGAACCAGCGATACGTCGCCTCAACATGCCAACGGC GTGACCTCGTGGTTAACTTTAATGCAGTTGGTTGGTCGCGTTGGGTGATTGCTCCGCCTGCCTACAACGCTGG CTACTGCTACGGCTACTGTCCCTTCCCCCTTTCAGCCCATTTCAACACTACCAACCACGCGATCATCATCCAC CTCATGTACAACTTGGGCGTGGCCCCACCCCAAGTCAAACCGCCCTGTTGCACCCCTGTCACCTTCAGTCCC CAGTCTATTCTCTTCTTCGACAGCGACGAG--- GTCGTTCTGCAAGTCTACGAGGACATGATTGTCGAGACTTGTGGCTGTCGG---------------------	KPI--------GE---NRKKRRRTR-LKSKSW- TNNREKRNSGYLMNQRYVASTCQRRDLV VNFNAVGWSRWVIAPPAYNAGYCYGYCP FPLSAHFNTTNHAIIIHLMYNLGVAPPQVKP PCCTPVTFSPQSILFFDSDE- VVLQVYEDMIVETCGCR-------

	Echinococcus multilocularis	gi|674576919| emb|CDS37359.1| bone morpholocus tagtic protein 2 [Echinococcus multilocularis]	AAACCAATC------------------------GGTGAA---------AATCGCAAAAAGCGTCGACGAACACGT--- CCTAAAAGCAAAAGTTGG--- ACAAATAATCGGGAAAAGCGCAATTCGCGATATCTCATGAACCAGCGATACGTCGCCTCAACATGCCAACGGC GTGACCTCGTGGTTAACTTTAATGCAGTTGGTTGGTCGCGTTGGGTGATTGCTCCGCCTGCCTACAACGCTGG CTACTGCTACGGCTACTGTCCCTTCCCCCTTTCAGCCCATTTCAACACTACCAACCACGCGATCATCATCCAC CTCATGTACAACTTGGGCGTGGCCCCACCCCAAGTCAAACCGCCCTGTTGCACCCCTGTCACCTTCAGTCCC CAGTCTATTCTCTTCTTCGACAGCGACGAG--- GTCGTTCTGCAAGTCTACGAGGACATGGTTGTCGAGACTTGTGGCTGTCGG---------------------	KPI--------GE---NRKKRRRTR-PKSKSW- TNNREKRNSRYLMNQRYVASTCQRRDLV VNFNAVGWSRWVIAPPAYNAGYCYGYCP FPLSAHFNTTNHAIIIHLMYNLGVAPPQVKP PCCTPVTFSPQSILFFDSDE- VVLQVYEDMVVETCGCR-------
	Helobdella robusta	gi|675874980| ref|XP_009023050.1| hypothetical protein HELRODRAFT_107188 [Helobdella robusta]	AGAAACCTC------------------------GGAGAA---------TTCAAAACCAGGAAAGTGCGC---AAGGAT------ AGTCGGAACAAG---AGAAATTCGTAC---AAGAAGAAACCAAACTATCTTACCAGACTC------------------ TGCCAACGTCGAAAGTTATACGTGGACTTCGGCGATCTAGGCTGGGAGGATTGGGTGATAGCTCCAGTGGGC TACACGGCCAACTACTGTTACGGCGAATGCACCTACCCGATGAATTCATACATGAATGCCACGAACCATGCCA TCATACAAACTCTAGCTCACTCCCTCAAC------ TCCTCATATGTGCCCAAGCCATGCTGTGCCCCCATCAAGTTGTCCACGCAATCTGTCCTCTACATCGACGACA ACAGCAACATCGTTCTTAAGTTCTACAAGAACATGGTAGTGAGGGCTTGTGGTTGTTTA--------------------	RNL--------GE---FKTRKVR-KD--SRNK- RNSY-KKKPNYLTRL------ CQRRKLYVDFGDLGWEDWVIAPVGYTAN YCYGECTYPMNSYMNATNHAIIQTLAHSLN -- SSYVPKPCCAPIKLSTQSVLYIDDNSNIVLK FYKNMVVRACGCL-------
	Hymenolepis microstoma	gi|674594923| emb|CDS26370.1| bone morphogenetic protein 2 [Hymenolepis microstoma]	AGAAATTCT------------------------GGGGAA---------AAGAGAAATAAACGCCGAAGA--- AAGAATCGTAAAGGTAAAAGCGAT--- ACTAGTAACCGAGAAAAACGAAGTCCTCAATACTTAATGAATAAGCGGTACATTGAGTCAACCTGCCAACGTC GCGATTTAATGGTCAATTTCAATGCAATCGGTTGGTCAAAGTGGGTTATAGCTCCGATGGCTTACAACGCAGG ATACTGCTATGGCAACTGCCCATTTCCCCTATCTGCCCATTTCAACACCACAAATCATGCAATAATACTTCATTT GATGCACAATTTAGGAGTGGCTCATTCTCAAATCAATTCGCCCTGTTGTACACCTGTGACGTTTGGTCCACAGT CTATACTCTTCTTTGATGGTGACGAC--- GTTGTGTTGCAAGTTTATGAAGATATGATTGTAGAGTCCTGTGGATGTCGG---------------------	RNS--------GE---KRNKRRR-KNRKGKSD- TSNREKRSPQYLMNKRYIESTCQRRDLMV NFNAIGWSKWVIAPMAYNAGYCYGNCPFP LSAHFNTTNHAIILHLMHNLGVAHSQINSPC CTPVTFGPQSILFFDGDD- VVLQVYEDMIVESCGCR-------

	*Mesocestoides corti**	MCOS_0001004001-mRNA-1	AAACCAACA-----------------------AACGCAGCCACAAAACAGCGCAAAGAGAGACGCGCA---CGC--- CGCAAAGGCCGAGGTGCT--- GCGTACTTTCGGGAGAAACACAACTCCCAATACCTCATGAACCAGCGCTACATCGCCTCCACATGCCAGAGAC GCGACCTGATGGTTAACTTCGACGCAGTTGGCTGGTCGCGGTGGGTGATCGCACCAATGGCCTACGACGCC GGCTACTGCTACGGCCACTGTCCCTTCCCCCTGTCCTCCCACTTCAACACCACCAACCACGCGATCATAATTC ACCTCATGTACAACCTGCAAGTCGCGCCATCACACATTCCACCGCCCTGTTGCACACCACTCACCTTCAGCCC TCAATCGCTATTGTTCTTTGAAGGCGACGAA--- GTGGTCCTGCAGGTGTATGAGAACATGGTCGTGGAGACGTGCGGTTGCCGG--------------------	KPT--------NAATKQRKERRA-R-RKGRGA- AYFREKHNSQYLMNQRYIASTCQRRDLMV NFDAVGWSRWVIAPMAYDAGYCYGHCPF PLSSHFNTTNHAIIIHLMYNLQVAPSHIPPP CCTPLTFSPQSLLFFEGDE- VVLQVYENMVVETCGCR-------
	Pinctada fucata	gi|46518286|dbj|BAD16731.1| bone morphogenetic protein-2 [Pinctada fucata]	AGACAAACAATTAACGATAAAGACGATAAAAGAAGAAAT---------AGAAAAAGGAGGAGGCGGAGG--- AAAAATAGACGACGAAAAAAATAAG---AGGAAAAAATAAG--- AAGAATAGAAAAAATAATAAAACAAAAAGAAAGAAGTACACCGATGCGTGTAAAAGAAAACCATTATATGTTGA CTTCAAAGCAGTGGGGTGGAATGACTGGATTTTTGCACCTCCTGGATACGAAGCTTATTATTGTCATGGGTCA TGTAACTGGCCGTATGACGATCATATGAACGTCACAAACCATGCAATAGTGCAGGACTTAGTGAATTCTATAAA C------ CCAGGGTCAGTACCCAAACCTTGCTGTGTACCCACAGAACTTAGCTCTCTTTCATTGTTATATACCGACGAACA TGAAGTTGTCGTCTTAAAAGTTTATCCCGATATGGTTGTTGAAGGATGCGGATGTCGG--------------------	RQTINDKDDKRRN---RKRRRRR- KNRRRKNK-RKNK- KNRKNNKTKRKKYTDACKRKPLYVDFKAV GWNDWIFAPPGYEAYYCHGSCNWPYDDH MNVTNHAIVQDLVNSIN-- PGSVPKPCCVPTELSSLSLLYTDEHEVVVL KVYPDMVVEGCGCR-------
	*Taenia solium**	TsM_000538200	AAACCATTT-----------------------CGTGGA---------AATCGTAAAAAGCGCCGAGGA---CGT--- CTAAAAAGCAAAAGCTGG--- ACAAGTGATCAGGAAAAGCACAATTCGCGATATTTTATGAACCAGCGATACATTGCCTCAACATGCCAACGAC GTGATCTAATGGTTAACTTTAATGCAGTTGGTTGGTCGCGTTGGGTGATTGCTCCTACTGCCTACAACGCCGG CTACTGCTTCGGCTACTGTCCCTTTCCCCTCTCAGCCCACTTTAACACTACAAATCACGCGATCATCATCCACC TCATGTACAATTTACGCGTGGCCCCACCCCAAGTCAAACCGCCCTGCTGTACCCCGCTCACCTTCAGTCCCCA ATCTATTCTCTTCTTCGACGGCGACGAG--- GTTGTCCTCCAAGTCTTTGAGGATATGATTGTCGAGACTTGCGGTTGTCGTTCTGTCCTTGGTCAGTTGATA	KPF--------RG---NRKKRRG-R-LKSKSW- TSDQEKHNSRYFMNQRYIASTCQRRDLMV NFNAVGWSRWVIAPTAYNAGYCFGYCPF PLSAHFNTTNHAIIIHLMYNLRVAPPQVKPP CCTPLTFSPQSILFFDGDE- VVLQVFEDMIVETCGCRSVLGQLI

Cyclin-g-associated kinase	*Caenorhabditis elegans*	gi|17567783|ref|NP_508971.1| Uncharacterized protein CELE_F46G11.3 [Caenorhabditis elegans]	CAGAGTGAGCTCTACGATCGT--- GGTCAGACATTCAGCATTAACGGAAACAATTATCGAGTGGAAAAGGTA--------- ATAGCCAAAGGCGGTTTTGGAACGGTGTTTCTTGCGACCAAC--- ACCAAGGGAAAACAAGTCGCCGTGAAGATCATGCTGAGCCACGATGCAGCCGCGACGAAGGATATTGATAAT GAAATTGATATGATGAAGAAGCTCCAA---CACGAAAATATTATTCAACTGTTTGATGCGTCAGCTGAAAGC--- AGAAGTTCAAATCGGTCCGTGAAAGAGTACAAAATATCAATGGAATATTGC-----------------AAATTTTCAATT------ --- GCGGATGTGCTTCTCAAGTACAAAGAAGTCTCAATTGACTTTGTCGTTCGCATAATCTACTTTACAACAAGAGC CCTAGTATATTTGCATTCCGTCGGCGCC------ ATCCACCGAGATATCAAAGCAGAAAATTTGCTTATAAACGGAAATGGAAAACTAAAACTATGCGATTTTGGAAG TGCTACAACAAAGTCCATCGAGATG--- GCACCACTATCAAATTCCGAAAGATTAGCGGTTCAGGAAGAGATGTTC--- AAATACACAACACCTATCACCCGATCCCCAGAAGTTTGTGATGTCTACTCAAATTGGCCTATCGGAAAACAACA AGACAACTGGGCAATGGGGTGTTTGATCTATTTTGTTGCGTTTGGAGAGCATCCATTTGATGGATCAGCG--- CTGGCAATTATCAACGGAAAGTACAAGAAGCCA---CCACCG------------ GTTCAGCAAAACCAGTTATCAGCGTTTGCAGATTTAATTGCAAAGTGTCTGACACCCAATCCGGAT--------- GAACGAATTACTGCTGCAAAAATTGAGGAATACATGAAACTA----------------------------- TTCACGGATATACTCGAT--- --CTGATGAATGTTCAACCG----------- GTACAAGCGGAGCAAAGTATCGAATCGCAA-----------------------------------GCTGCAAAAGGATTC--------------------- --TTTACA--- -- -- -- -- ATGCAGGACAAACTATTTTCCAATTTGACATCACTCAAAAATACAGTTGTACAGCAGACGAATAAAATGGGATG GGGAATGGAGCCAACAAAT--- ----------------------------------ACT-----------------------------ACACCACGCCCCGGG------CACCCTTCA-------------- ACATCTCCGAAGCTTGTG-- -------------------TCACGAGACTTGTTTGAC-- ---TTTGACGACCTGATGCTC-----------------CGACAC--------- ACGACACCTTCCGCCGAATCT------ TCTCAAGCTGTTTTACAACCAATACGCCAAGTGGAGAACAAGAACTTGACCAAAGTCGATGTT--------------------- -- -- TCAAAAAACGGAATTGGTTCGTCATCCTCTTCTGCCAGCCTGGATGATATGGTCAGCGATATGATGAAAATGTC AACCAAGAAA-- --	QSELYDR--------------------- GQTFSINGNNYRVEKV--- IAKGGFGTVFLATN- TKGKQVAVKIMLSHDAAATKDIDNEIDMMK KLQ-HENIIQLFDASAES- RSSNRSVKEYKISMEYC------KFSI--- ADVLLKYKEVSIDFVVRIIYFTTRALVYLHSV GA-- IHRDIKAENLLINGNGKLKLCDFGSATTKSI EM-APLSNSERLAVQEEMF- KYTTPITRSPEVCDVYSNWPIGKQQDNWA MGCLIYFVAFGEHPFDGSA- LAIINGKYKKP-PP---- VQQNQLSAFADLIAKCLTPNPD--- ERITAAKIEEYMKL----------FTDILD------------ -- LMNVQP----VQAEQSIESQ------------AAKGF- -------FT--- -- -- -- MQDKLFSNLTSLKNTVVQQTNKMGWGME PTN--T-- --------TPRPG--HPS-----TSPKLV---------------- -----------------------------SRDLFD------------------ -------------------------------FDDLML------RH--- TTPSAES--SQAVLQPIRQVENKNLTKVDV-- -- --- SKNGIGSSSSSASLDDMVSDMMKMSTKK-- -- ----------------------

Clonorchis sinensis	gi|358336029| dbj|GAA31493 .2| cyclin G- associated kinase [Clonorchis sinensis]	CAGAGTGAGCTCTACGATCGT--- GGTCAGACATTCAGCATTAACGGAAACAATTATCGAGTGGAAAAGGTA--------- ATAGCCAAAGGCGGTTTTGGAACGGTGTTTCTTGCGACCAAC--- ACCAAGGGAAAACAAGTCGCCGTGAAGATCATGCTGAGCCACGATGCAGCCGCGACGAAGGATATTGATAAT GAAATTGATATGATGAAGAAGCTCCAA---CACGAAAATATTATTCAACTGTTTGATGCGTCAGCTGAAAGC--- AGAAGTTCAAATCGGTCCGTGAAAGAGTACAAAATATCAATGGAATATTGC------------------AAATTTTCAATT------ --- GCGGATGTGCTTCTCAAGTACAAAGAAGTCTCAATTGACTTTGTCGTTCGCATAATCTACTTTACAACAAGAGC CCTAGTATATTTGCATTCCGTCGGCGCC------ ATCCACCGAGATATCAAAGCAGAAAATTTGCTTATAAACGGAAATGGAAAACTAAAACTATGCGATTTTGGAAG TGCTACAACAAAGTCCATCGAGATG--- GCACCACTATCAAATTCCGAAAGATTAGCGGTTCAGGAAGAGATGTTC--- AAATACACAACACCTATCACCCGATCCCCAGAAGTTTGTGATGTCTACTCAAATTGGCCTATCGGAAAACAACA AGACAACTGGGCAATGGGGTGTTTGATCTATTTTGTTGCGTTTGGAGAGCATCCATTTGATGGATCAGCG--- CTGGCAATTATCAACGGAAAGTACAAGAAGCCA---CCACCG------------ GTTCAGCAAAACCAGTTATCAGCGTTTGCAGATTTAATTGCAAAGTGTCTGACACCCAATCCGGAT--------- GAACGAATTACTGCTGCAAAAATTGAGGAATACATGAAACTA------------------------------ TTCACGGATATACTCGAT-- ---CTGATGAATGTTCAACCG------------ GTACAAGCGGAGCAAAGTATCGAATCGCAA-------------------------------------GCTGCAAAAGGATTC---------------------- --TTTACA--- -- -- -- --- ATGCAGGACAAACTATTTTCCAATTTGACATCACTCAAAAATACAGTTGTACAGCAGACGAATAAAATGGGATG GGGAATGGAGCCAACAAAT-- ------------------------------------ACT------------------------------ACACCACGCCCCGGG------CACCCTTCA--------------- ACATCTCCGAAGCTTGTG--- --------------------TCACGAGACTTGTTTGAC-- ---TTTGACGACCTGATGCTC------------------CGACAC--------- ACGACACCTTCCGCCGAATCT------ TCTCAAGCTGTTTTACAACCAATACGCCAAGTGGAGAACAAGAACTTGACCAAAGTCGATGTT---------------------- -- -- TCAAAAAACGGAATTGGTTCGTCATCCTCTTCTGCCAGCCTGGATGATATGGTCAGCGATATGATGAAAATGTC AACCAAGAAA-- ---	--- ---------------------------MVL--------------------------------- --- --LE- KFTTPMYRAPEMLDLYQNYPIGTAADIWAL GCILFYLSCTYHPFEDAAKLAILNAKYTLPT PTSREETMFHSLIRQMLLVDPRQRPDINDV LREVSEVAAVLEIRVGLLKGGAGNLLRNIR D--ASNKVLESV-- -SSTLS-SDLDFQLITSRIAVMS----YP----A---- --ESGLE- SIGTGNSMEEVQNMLNSRYPNAYAVYNLS PRPYRSDQWFDGRVSHRVFDAHRAPSLR SLIELCLNARLWLSQKPGNLCVVHCMDGR AASAMLVCSLLCFCHLFDNVSPALQLFSSK RGNPRLNASQTRYIDYVAQLV--------H---- NRVPMPHHRPLKLISLTVTPIPTFNKSKNG CRPYVEVYEGKTQVLSTYTDYDSLRSYVL- EDRKIEFLL- NGISVLGDLTVIVYHCRSSFAGR-- GKVAAVKIAQFQLYTGFVEPDQSELIYFKS DLDHLDTSSGFGN----------Y---- TSRYADSFNLTLEFMVSPNERPRQGKNL- VYPW--ETLPP--PEA----- LRPELCVSDSEELRSLLSD--------------------- G----AATT-GS-----AY-------------------- PPNTPNRT--------- GTRPHVGKDAFSDLLGDF--S-GAS-- NASGQGQPKTVNEMR-- REKLAKTVDPEQL----------------------------- KV--- ------QDWAHGKDRNLRALLCSLPGILWDG- VKWKPVGITDLMTAEQVKRQYRNAARVIH PDKWM----------NTEHEQL--------------- ARMIFVELNDAMAE-------------------FE

Crassostrea gigas	gi|405976458|gb|EKC40964.1| Cyclin G-associated kinase [Crassostrea gigas]	ATGACAGAATTCTTCAAATCAGCGTTTAGTGCCCTGACTGGAAATATATCA---------------GGGATAGAAAATGATTTCGTCGGTCAGATCGTTGAGTTGGGACAGCAGAAACTTCGTGTACGCCGTGTC---------ATAGCGGAAGGTGGATTCGCATTTGTGTATGTTGCACAAGATGTTACGACTGGAAAAGATTATGCCTTAAAAAGGTTGCTTGCTCATGATAAGGAAAAGAATGAGATGGTTATGAATGAAATCAAATATCTTAAAAAACTATCAGGTCATCCCAACATTGTCGAGTTTATAGCTGCTGCCTCG------------GACTCCGACAAAGGACAGTGCGAGTACCTCTTGCTGACTGAGCTTTGTACAGGC---CAACTTATTTCTGTTCTAAAT---------GGTGCTGGA------------TCACCTCTACCGTGCAGTGATGTTATTCAGATTTTTTTCCAAGCATCTCTTGCAATACAGCACATGCACAGACAAAATCCTCCAATTATTCACAGAGATTTAAAGGTTGAGAACCTACTGGTAAGTTCCAAAGGAATGATCAAGCTGTGTGACTTTGGCAGTGCCACAACAGAAACACACTTTCCAGATAGGTCTTGGTCAGCCATTCAGAGAAGCCTGGTTGAAGATGAGATTACC---AAAAATACAACCCCAATGTATCGAGCACCTGAAATGTTAGACCTTTATCAGAACTCTCCCATATGTGAAGCAAGTGATATATGGGCTCTAGGATGTTTACTTTACATGCTTTGCTTTGGCAATCATCCGTTTGAGGATTCAGCCAAATTGAGAATCATCAATGCAAACTATAGCATTCCA---AGCACTGATACCCAGTATAAGGTGTTGCACAATTTAATCAAGTCAATGTTGCAAGTTAATCCAAATGATAGACCAACCATCAATGACATTATAGACAGGATTAAGGAAATAGCAACTGCCAAAAATGTAAACCTGAGCTCCCTGAGGGGTGGAGCAGGAAATCTAATGAAAAATATCAAAGAT------GCTTCAGCAAAAGTCATGGAAACTGTG---TCTGCGACAATTGGAAAAACTGACTTGGATATAAACTACATTACATCAAAAATTGCAGTTATGTCA------------TTTCCT------------GCG------------------GAG---GGAGTGGAG---TCA---GCTTTTAAGAACCACATAGATGAAGTTCGGAGTTACTTAGAGAGTCGTCACAAGGATTGTTATGCAGTGTACAACCTATCCCAGAGGTCCTACAGAGCAACAAAG---TTTGAGAACCGGGTTTCTGAGTGTGGTTGGCCAGCCAAAAAAGCTCCAATGCTTTCAAGCCTCTTTGCAATTTGCCAGAACATGCATCTGTGGCTTCGACAGAATCCCCAAAATATCTGTGTGATTCACTGCATAGATGGAAAGGCATCGTCAGCTACTGTAGTAGGGGCATTCTTTTCTTTTCTTCACCTGTTTGACTCCCCAGAGCAGAGTATGCACATGTTTTCTGTAAAACGAGGACCACCTGGAGTCACTCCTTCCCAAAAACGATACATTGGATATATATCTGAGATAGTT-------------------------GCT------------GATCCTCCATATTCACCCCACAGTGCAGCGGTTCTTCTGAAATCGATGACTTTGTCACCTGTTCCCTTGTTTAACAAAATGAGAAATGGTTGCAGACCATTTGTTGAAGTGTATGTCAATGAAGAGAGAATTCTTACCACATCTCAGGAATATGACAAAATGAGAGGTTATCAGACG---GAGGATGGATTAGCTATACTTCCTGTA---AAC---ACTGCAGTGCAAGGGGATGTCACTGTGTATGTCTACCATGCCAGGTCAACATTTGGGGGAAAGGTGCAAGGAAAGATAACATCTATGAAAATGTTTCAAGTTCAGTTTCACACTGGGATGATTAGACCAGGAACTACATCACTCAAGTTTACACAGTTTGATTTAGATCAGTTAGATACTGCAGAG--AAGTACCCGGAACAATTCAGTGTGATGCTAGACATTTCAGTGTCTAAAAATGAAAGACCA------ACC-------------------------------CAGAAGACTCCATCA------TGGAAAGCATATGATGCATCCAAAGTGTCTCCAAAGGTTCTATTTTCCAACAAAGAAGAAATGATGGAAACATTGCAGGCA--CAG------------AATAAACCAAAT---TATAAT----------------	MTEFFKSAFSALTGNIS-----GIENDFVGQIVELGQQKLRVRRV---IAEGGFAFVYVAQDVTTGKDYALKRLLAHDKEKNEMVMNEIKYLKKLSGHPNIVEFIAAAS----DSDKGQCEYLLLTELCTG-QLISVLN---GAG----SPLPCSDVIQIFFQASLAIQHMHRQNPPIIHRDLKVENLLVSSKGMIKLCDFGSATTETHFPDRSWSAIQRSLVEDEIT-KNTTPMYRAPEMLDLYQNSPICEASDIWALGCLLYMLCFGNHPFEDSAKLRIINANYSIP-STDTQYKVLHNLIKSMLQVNPNDRPTINDIIDRIKEIATAKNVNLSSLRGGAGNLMKNIKD--ASAKVMETV-------------------------------------SATIGKTDLDINYITSKIAVMS----FP----A------E-GVE-S-AFKNHIDEVRSYLESRHKDCYAVYNLSQRSYRATK-FENRVSECGWPAKKAPMLSSLFAICQNMHLWLRQNPQNICVIHCIDGKASSATVVGAFFSFLHLFDSPEQSMHMFSVKRGPPGVTPSQKRYIGYISEIV--------A----DPPYSPHSAAVLLKSMTLSPVPLFNKMRNGCRPFVEVYVNEERILTTSQEYDKMRGYQT-EDGLAILPV-N-TAVQGDVTVYVYHARSTFGGKVQGKITSMKMFQVQFHTGMIRPGTTSLKFTQFDLDQLDTAE--------------------KYPEQFSVMLDISVSKNERP--T----------QKTPS--WKAYDASKVSPKVLFSNKEEMMETLQA--------------------Q---NKPN-YN-----VG-------------------GFAT-GAKNDRSVKNPYGVKPQVNVNAFEDLLG------NHQFTSSKQNNAPKTIGDMK--KLQMAEEMDPEKL----------------------------KV--

Echinococcus granulosus	gi|674563543| emb|CDS21892.1| cyclin g associated kinase [Echinococcus granulosus]	---ATG---------TTACTTTTAGGCGGTTATGCTGTCGTATTTGAGGGCTATGACCCATCTCAAGGAAAGTCGTTTGCTATAAAG---------------------GAAGCCGTTACTGATATTATGAACGAAATTGATATACTGAAACGGCTCTCTGGACATCCAAATATTATGAAGTTCTTTGGGGTTGCATGCGCTGGTAAGGAAAGGGGAAAACAAGTCGGGAACGAGTTTCTCATCGTTACGGAACTATGTTCGGGAGGTCCTCTAAAAGACTATCTG---CCTCCTTCTCATCAAGGC------------AAGCACCTACCTTTGAATATTGTCTTACAAATCCTTGCGCAGACCAGCTGCGCCATACAGCACATGCACAAGCAGTCACCTCCGATAATCCATCGTGATCTGAAAATTGAAAATATTCTCCTCTCAGAATCGTTCACCATAAAATTATGTGATTTCGGCAGTGCCACGACGGAAACTTTTGCTCCTAGTGTGACGTGGTCGGCTGTTGAACGTGGTCGAGTGCAGGAAAGTCTGGAG---AAGGTTACAACGCCTATGTATAGAGCTCCAGAAATGTTGGATTTATATCTCAACTATCCCATCAACGAAGCTCTGGACATTTGGGCTTTCGGATGCATAATGTTCTACCTAGTCTGTGGATATCATCCCTTCGAAGATTCGGCAAAACTGGCCATCTTGAACGCAAATTTCAATCTCCCT---CCCTGTGATACCGGCTTTGAACCCTTTCATAACCTGATTCGCCAAATGTTATTGGTTGATCCGACCCAACGGCCCAGCATAAACTCCATCTACGGTGAGCTCTCCGACCTTGCCACTACCATAAACGTATCTGCTGGTTTCCTCAAGGGTAGTGCAGGTCACTTGATCAAAAATATTCGAGAG------ACGTCATCCAAGGTTATGGAAGCGGTG--ACTGCCGCTATCCCA---AACGACTTGGATCTGCAGTATATTACCTCACGAATAGCGGTAATGTCC------------TACCCT------------GCT------------------GAGAGCGGTCTCGAAACCGTTGTCGGTTCCCGCAATTCCATGGAGGAGGCCCAACTCTACCTAGACCGTCACCACCCTAATAGCTACGCCGTCTATAACCTCAGCTCACGAGCTTATCGTTTTCCCCATTGGTTTGGTGGCCGTGTTTCCTTCCGCCCCTTCGAGTCCAACCGAGCGCCAACGTTGTGGTCCCTCGTCGAGTTGTGCCAAAACGCGCGCCTCTGGTTGTCGCAGAAACCGGACAATGTGTGCATCATTCACTGCACTGATGGACGACAGCTGTCCGCAGTGTTAGTCTGTTGTCTCCTCTGCTTTTGCCGTGTGTTTACGGAGGCGTCTACGGCTGTAAAGTTTTTCGCCTCGAAGCGAGGACCTGTGCGACTTACGCCGTCTCAAATGCGCTACATCGATTACGTGGCAAAGCTTGTG------------------------TCT------------GTACCGCCCCAAGTTCCGCATAATCATCCAGTGGAATTACTGAGCATCATCATCGCCCCTATCCCATCTTTCAACCGACTCAAAAACGGTTGTCGTCCCTATGTCGATGTTTACCAAGGCAACAAGAAGGTCTTTTCGTCAATGACTGACTACGAGTCTTTGCGGACCTACGAGTTC---GAGGATCGTAAAATTGAGATCACTTTC---CTTGACCTCTCAGTTTACGGGGATGTTACAATCAGTGTCTATCATGCTCGCTCCTTCTTTGCTGGAAAG------GGGAAGGCGGCGTCGGTCAAGATATGCCAATTGCAATTTCACACCGGTTTCATTAATCTTGATGCTGGGAAGGTTACCTTCTTGAAATCGGAACTGGACCACCTGCATTCA---TCCAACGCTGGCTCC---------CAAGGT---ATCGGCACGCCATCG------------ACTAGTCGGTATGCTGAGAGTTTCCACACTGTGCTCCACCTCTCCGTCTCACTTAACGAACGTCCACGGCAGGGTGCAGCCCTC---GTCTTCCCTTGG------GAGACCTTACCGCCC------GAGAGCTCA---------------CGAAAACCCATGCTTTGCTTCTCCGATGCCTTCGAGATGCGTTCGGTCATGACGCCG---CCACCACCCCAACAACAACAGTACCAA---TCGTCG---------------GCTTCCTACACGTCGCCGAGGCACAACCCTGCACCGAACACC---------------GCCTCACCTAGT---------------	---M---LLLGGYAVVFEGYDPSQGKSFAIK-------EAVTDIMNEIDILKRLSGHPNIMKFFGVACAGKERGKQVGNEFLIVTELCSGGPLKDYL-PPSHQG----KHLPLNIVLQILAQTSCAIQHMHKQSPPIIHRDLKIENILLSESFTIKLCDFGSATTETFAPSVTWSAVERGRVQESLE-KVTTPMYRAPEMLDLYLNYPINEALDIWAFGCIMFYLVCGYHPFEDSAKLAILNANFNLP-PCDTGFEPFHNLIRQMLLVDPTQRPSINSIYGELSDLATTINVSAGFLKGSAGHLIKNIRE--TSSKVMEAV-------------------------------------TAAIP-NDLDLQYITSRIAVMS----YP----A------ESGLETVVGSRNSMEEAQLYLDRHHPNSYAVYNLSSRAYRFPHWFGGRVSFRPFESNRAPTLWSLVELCQNARLWLSQKPDNVCIIHCTDGRQLSAVLVCCLLCFCRVFTEASTAVKFFASKRGPVRLTPSQMRYIDYVAKLV--------S----VPPQVPHNHPVELLSIIIAPIPSFNRLKNGCRPYVDVYQGNKKVFSSMTDYESLRTYEF-EDRKIEITF-LDLSVYGDVTISVYHARSFFAGK--GKAASVKICQLQFHTGFINLDAGKVTFLKSELDHLHS-SNAGS---QG-IGTPS----TSRYAESFHTVLHLSVSLNERPRQGAAL-VFPW--ETLPP--ESS-----RKPMLCFSDAFEMRSVMTP--------------------PPPQQQQYQ-SS-----ASYTSPRHNPAPNT-----ASPS-----FH---------GARPRVTEDTFEDLLGGFSSSSGEAFGSSRRNQQPKTLAEIR--HQKLAQTEDREAL----------------------------KV---IEWLEGKEKNVRALLCTLNSVLWEG-VRWEQISMADVMTVKQVKQQYRKAARAV

Echinococcus multilocularis	gi|674267522|emb|CDI97027.1| cyclin g associated kinase [Echinococcus multilocularis]	--ATG---------TTACTTTTAGGCGGTTATGCTGTCGTATTTGAGGGCTATGACCCATCTCAAGGAAAGTCGTTTGCTATAAAG---------------------GAAGCTGTTACCGATATTATGAACGAAATTGATATACTGAAACGGCTCTCTGGACATCCAAATATTATGAAGTTCTTTGGGGTTGCATGTGCTGGTAAGGAAAGGGGAAAACAAGTCGGGAACGAGTTTCTCATCGTTACGGAACTATGTTCGGGAGGTCCCCTAAAAGACTATCTG---CCTCCTTCTCATCAAGGC------------AAGCATCTACCTCCGAATATTGTCTTACAGATCCTTGCGCAGACCAGCTGCGCCATACAGCACATGCACAAGCAGTCACCTCCGATAATCCATCGTGATCTGAAAATTGAAAATATTCTCCTCTCAGAATCGTTCACCATAAAATTATGTGATTTCGGCAGTGCCACGACGGAAACTTTTGCTCCTAGTGTGACGTGGTCGGCTGTTGAACGTGGTCGAGTGCAGGAAAGTCTGGAG---AAGGTCACGACGCCTATGTATAGAGCTCCAGAAATGTTGGATTTATATCTCAACTATCCCATCAACGAAGCTCTGGACATTTGGGCTTTCGGATGCATAATGTTCTGCCTAGTCTGTGGATATCATCCCTTCGAAGATTCTGCAAAACTGGCCATCTTGAACGCAAATTTCAATCTCCCT---CCCTGTGATACCGGCTTTGAACCCTTTCATAACCTGATTCGCCAAATGTTATTGGTTGATCCGACCCAACGGCCCAACATAAACTCCATCTACGGTGAGCTCTCCGACCTTGCTACTACTATAAACGTATCTGCTGGTTTCCTCAAGGGTAGTGCAGGTCACTTGATCAAAAATATTCGAGAG------ACGTCATCCAAGGTTATGGAAGCGGTG---ACTGCCGCTATCCCA---AACGACTTGGATCTGCAGTATATTACCTCACGAATAGCGGTAATGTCC------------TACCCT------------GCT------------------GAGAGCGGTCTCGAAACCGTTGTCGGTTCCCGCAATTCCATGGAGGAGGCCCAACTCTACCTAGACCGTCACCACCCCAATAGCTACGCCGTCTATAACCTCAGCTCACGAGCTTATCGTTTTCCCCATTGGTTCGGTGGCCGTGTTTCCTTCCGCCCCTTCGAGTCCAACCGAGCGCCAACGTTGTGGTCCCTCGTCGAGTTGTGCCAAAACGCGCGCCTCTGGTTATCGCAGAAACCGGGCAATGTGTGTATCATTCACTGCACTGATGGACGACAGCTGTCCGCAGTGTTAGTCTGTTGTCTTCTCTGCTTCTGCCGTGTGTTTACGGAGGCGTCTACGGCTGTAAAGTTTTTCGCCTCGAAACGAGGACCTGTGCGACTCACACCGTCTCAAATGCGCTACATCGATTACGTGGCAAAGCTTGTG------------------------TCT------------GTACCGCCTCAAGTTCCGCATAATCATCCAGTGGAATTACTGAGCCTCATCATCGCCCCTATCCCATCTTTCAATCGACTCAAAAACGGTTGTCGTCCCTATGTCGATGTTTACCAAGGCAACAAGAAGGTTTTTTCGTCAATGACTGACTACGAGTCTTTGCGGACCTATGAGTTC---GAGGATCGTAAAATTGAGATCACTTTC---CTTGGCCTCTCAGTTTACGGGGATGTTACAATCAGTGTCTATCACGCACGCTCCTTCTTTGCTGGAAAG------GGGAAGGCGGCGTCGGTCAAGATATGCCAATTGCAATTTCACACCGGTTTCATTAATCTTGATGCTGGGAAGGTTACCTTCTTGAAATCGGAACTGGACCACCTGCATTCA---TCCAACGCTGGCTCC---------CAAGGT---ATCGGCACGCCATCG------------ACTAGTCGATATGCTGAGAGTTTCCACACTGTGCTCCACCTCTCCGTCTCACCCAACGAACGTCCACGGCAGGGTGCAGCCCTC---GTCTTCCCTTGG------GAGACCTTACCGCCC------GAGAGCTCA---------------CGAAAACCCATGCTTTGCTTCTCCGATGCCTTCGAGATGCGTTCGGTCATGACGCCG---CCACCACCCCAACAACAACAGTACCAA---TCGTCG---------------GCCTCCTACACGTCGCCGAGGCACAACCCCGCACCGAACACC---------------GCCTCACCTAGT---------------	--M---LLLGGYAVVFEGYDPSQGKSFAIK-------EAVTDIMNEIDILKRLSGHPNIMKFFGVACAGKERGKQVGNEFLIVTELCSGGPLKDYL-PPSHQG----KHLPPNIVLQILAQTSCAIQHMHKQSPPIIHRDLKIENILLSESFTIKLCDFGSATTETFAPSVTWSAVERGRVQESLE-KVTTPMYRAPEMLDLYLNYPINEALDIWAFGCIMFCLVCGYHPFEDSAKLAILNANFNLP-PCDTGFEPFHNLIRQMLLVDPTQRPNINSIYGELSDLATTINVSAGFLKGSAGHLIKNIRE--TSSKVMEAV--TAAIP-NDLDLQYITSRIAVMS----YP----A------ESGLETVVGSRNSMEEAQLYLDRHHPNSYAVYNLSSRAYRFPHWFGGRVSFRPFESNRAPTLWSLVELCQNARLWLSQKPGNVCIIHCTDGRQLSAVLVCCLLCFCRVFTEASTAVKFFASKRGPVRLTPSQMRYIDYVAKLV--------S----VPPQVPHNHPVELLSLIIAPIPSFNRLKNGCRPYVDVYQGNKKVFSSMTDYESLRTYEF-EDRKIEITF-LGLSVYGDVTISVYHARSFFAGK--GKAASVKICQLQFHTGFINLDAGKVTFLKSELDHLHS-SNAGS---QG-IGTPS----TSRYAESFHTVLHLSVSPNERPRQGAAL-VFPW--ETLPP--ESS-----RKPMLCFSDAFEMRSVMTP---------------------PPPQQQQYQ-SS-----ASYTSPRHNPAPNT-----ASPS-----FH---------GARPRVTEDTFEDLLGGFSSSSGEAFGSSRRNQQPKTLAEIR--HQKLAQTEDREAL-----------------------------KV---IEWLEGKEKNVRALLCTLNSVLWEG-VRWEQISMADVMTVKQVKQQYRKAARAV

Helobdella robusta	gi|675886148| ref|XP_009028576.1| hypothetical protein HELRODRAFT_194118 [Helobdella robusta]	ATGGCCGAATTCTTCAAATCGGCTTTAGGATATTTAGGAAATACGGTTTCTTCA------------AATAGTGAAAATGATTTTGTCGGACAGAATGTCGAGATGGGGAGGCAAAAACTACGCGTCAAAAGGATG---------ATAGCAGAAGGTGGGTTTGCGTTTGTTTTTGTGGCCCAAGATGTAAAATCAGGGAAAGAATATGCATTGAAGAGACTTCTATCAAGTGACTCTGAGAAGACCAAATCAATAATCCAAGAAATAACCATTCTGAAAAAGGTTTCAGGACATAAAAACATTATAGAGTTTGTAGCAGCTGCCAGTGTAGGCAAGGAAGAATCTGGTCATGGGCAGGCTGAGTTTCTCCTCTTAACTGAATTTTGTTCAGGTGGAGAGCTGATCAAAGCAATG------------AGAAACCAG------------ACA---CTGACAAGAAATTTGGTCCTGCAAATATTTTACCAGACCTGCTCAGCTGTCTTACACATGCACAAATTAAAACCACCAATCATACATAGAGATCTTAAGTTGGAAAATCTTCTCCTAAGCAGTGACAACAGGGTCAAGTTATGTGATTTTGGTAGTAGCACGACATTAGCTCAATATCCAGATGACACCTGGACAACTGCTAGGAGGAATACTGTGGAAGATGAGATGGCC---CAAAACACAACCCCCATGTATAGGGCACCTGAGATGTTGGATCTGTATCAAAATTATCCAATCAATGAACAGGCTGAT---GACTCTGCAAAGTTGAGGATTATAAATGCCAACTTCACAATACCT---CCTACTGACACACAATTCACTGAATTCCATAGTCTCATTAAAATGATGTTGAAGCCACACCCTCTTGAAAGACCAAACATCGAGGACATAGTTGAACAGCTCGAACTTCTAGCAACCACATTCAACGTCAACCCAGGCCAGCTGAAGGAGAGTGCGGCTGGCCTGATGAAAAATGTTAGGGAGAGACATGCATGTGCTGACACACTCACATGTGTATGTATGATGATTGCATTTAAATCCGATCGAAGACAAAACAGACAACTGAGACTACCAGCCATCTTCTCATTCATCAAAACAATCAACAGTCAATCCAGTATGCCTTCATCATCGTCGTCGTCGAACAACAACTCAACAATGGACTTGAGCTACCTGACTAGCAGGTTGATCGTGACGTCA------------TTTCCG------------CAC------------------GAT---GGCGTCGAC---TCG---AGCACGCGCAACCATATTGATGAC--AACATGTACAAGTGGCTGTCGCAATCGCACCACAACGTCGTCGTGCTGCATTGCTTGAACGGCGGGAACATCTCGGCGACGGTCACGTGCTCCTTTCTTTGTTACTGTCGTCTGGTCGATGACATCATGCAGGCCGTCTCATTGTTCTCTAGCAGAAGGTTCAATCCTAAACTCACCGCCTCTCAGTTGAGATACATAGGCTACATAAACCAGATAGCTAATAGGATGTCATCATCATCGTCGTCG------------TCGTCCTCACAACTACCCAACAACAACTTTGTAACTCTGCGGCGAATCACAATACAGCCTGTCCCTCTATTTAATAAGATGAAGACAGGCTGTAGACCTTACATCGAAGTGTATGTTGGCGATGAGAAGATATTTGTCAGTAGCGAGGACTATGACTCCATGAAGTCATACGTCATA---GGGGACGGCAGGGTGCGTCTGGAACTA---GAC---ACCATGGTGACCGATGACGTCACCATCATCCTCTACCACGCTAGGTCAACTATCGGCTCCAAAGTTCAAGGAAAGATGACGTCAGTGAAGATATTTCAAATTCAGTTCAATGCTGGCTTCATAGAGGACTCTGTTAAAAATATCAACTTTCAAATAACTGACCTCGATAATCTTGACCCACAACAAGAT---GCAAAGTACCCTGAAGAGTTCGAGGTCGACCTTGAACTTTCTGTGACGTCGCAGGAAGTTCCGCCCAACAAT------------------------------CAGATGAAGGAGATG------CTTGCAGCTATGCATAGGAGCAGGCCGTTTTATACAGCCCTGATAACCAGCGAAAACGAGTACATCAACCTGGCC--AGT---------------GCTTTT--TCAAGCGTTATC---	MAEFFKSALGYLGNTVSS----NSENDFVGQNVEMGRQKLRVKRM---IAEGGFAFVFVAQDVKSGKEYALKRLLSSDSEKTKSIIQEITILKKVSGHKNIIEFVAAASVGKEESGHGQAEFLLLTEFCSGGELIKAM----RNQ----T-LTRNLVLQIFYQTCSAVLHMHKLKPPIIHRDLKLENLLLSSDNRVKLCDFGSSTTLAQYPDDTWTTARRNTVEDEMA-QNTTPMYRAPEMLDLYQNYPINEQAD-------------------DSAKLRIINANFTIP-PTDTQFTEFHSLIKMMLKPHPLERPNIEDIVEQLELLATTFNVNPGQLKESAAGLMKNVRERHACADTLTCVCMMIAFKSDRRQNRQLRLPAIFSFIKTINSQSSMPSSSSSSNNNSTMDLSYLTSRLIVTS----FP----H------D-GVD-S-STRNHIDD---NMYKWLSQSHHNVVVLHCLNGGNISATVTCSFLCYCRLVDDIMQAVSLFSSRRFNPKLTASQLRYIGYINQIANRMSSSSSS----SSSQLPNNNFVTLRRITIQPVPLFNKMKTGCRPYIEVYVGDEKIFVSSEDYDSMKSYVI-GDGRVRLEL-D-TMVTDDVTIILYHARSTIGSKVQGKMTSVKIFQIQFNAGFIEDSVKNINFQITDLDNLDPQQD-------------------AKYPEEFEVDLELSVTSQEVPPNN----------QMKEM--LAAMHRSRPFYTALITSENEYINLA-----------------------------------S-----AF---------------------SSVI-GGREDRGLKKGFGVKPAVPQDAFEDLLG------GHSFTSSSAKKEPKTIKEMR--RELDAKDIDPERL-------------------------------KV---REWIEGKERNIRALLCSLHTVLWDGEDRW

Hymenolepis microstoma	gi|674591869|emb|CDS29298.1| cyclin g associated kinase [Hymenolepis microstoma]	ATGGGAGATTTCATAAAGTCGGCCTTTGGCTATTTTGGAGGACCAAATCAAGGA------------GTAAAAGACCATGAATTGGTGGGAGAGTCCGTGGAAGTTGGCAAAGCTCACTTTCGTATTCGTCGAGTA---------ATTGCCGATGGTGGTTATGGTGTGGTATTTGAGGGATATGATAGTTCTCTGGGACGATCTTTCGCTATAAAGCGACTTTTTGCTCCGGATCAAGATGAAGTCAATGTTATTATGAATGAGATTAATGTTCTGAAACAGCTTTCGGGACATCCAAATATAATGACATTCTATGGTTGCGCTTGTGCTGATAAAGAGCGGGGCAAGCTTGCTGGGAATGAATTTCTCATTGTCACCGAATTGTGTTCAGGAGGTCAACTTAAAGATTATCTG---CCTGCTCCACATCAGGGA------------AAACACCTTCCAGCAGACATTGTACTCCAGGTTCTTTCTCAAACCAGTCGAGCCATTCAACACATGCACAAACAATCACCCCCAATTATTCATCGCGACTTAAAGATTGAAAATATTCTTCTTTCTGAGTCATTTACTATCAAATTATGCGATTTCGGTAGTGCCACTACAGAAACACTCGCTCCAAATATATCCTGGTCTGCCATTGAGCGTGGCAGAGCTCAGGAAAGTTTAGAA---AAGGTTACGACTCCCATGTACAGATCTCCCGAAATGTTGGATTTATATCTCAATTATCCCATTAACGAGGCTCTAGACATTTGGGCCCTTGGATGTATTATGTTCTACCTGGTCTGTGGCTTTCATCCCTTTGAAGACTCGGCCAAATTGGCCATTCTCAATGCAAATTACAACCTTCCA---CCCTCCGATAGTGACTTTGAGCCTTTTCACAATCTCATCCGTCAAATGCTACTGGTAGATCCCACCCAAAGGCCGAATATAAACTCCGTTTATGGAGAATTATCGGATTTAGCTACCACTCGCAATGTTCCCGCAGGGTTTCTCAAAGGCAGTGCAGGCCACTTGATTAAAAATATCAGGGAA------ACTTCGTCTAAGGTGATGGAAGCTGTC---ACGGCCGCAATGCCC---GGTGATTTGGATCTGCAATATATCACTACACGCATTGCAGTTATGTCC------------TACCCA------------GCT------------------GAAAGTGGACTAGAAGCTGTAGTCGGCTCTCGCAACTCAATGGAAGAAGCTCAAGCCTACCTTGATCGTCACCATCCAAACAGTTATGCTGTCTACAACCTCAGCTCACGACAATATCGATCAACCTGCTGGTTTGGTGGTCGGGTCTCTTTCCGACGTTTGGAGCTCAATCGGGCCCCTTCTCTAGCCTCCCTCATTGAGTTGTGTCAAAATGCCCGACTCTGGTTGTCTCAAAAACCTGATAATATATGTGTTATCCACTGTGGTGATGGCAGACAGCTGTCTGCTGTGTTTGTTTCCAGTCTCCTCTGTTTTTGTGGTGTGTTTTCTGATGCTAGTTCGGCTGCAAGATTCTTCAATTCTAAAAGGAGTTCTTCAAGACTTACTCCTGCACAGTTGAGGTACATTGACTATGTGGCAAAACTGGTC------------------------GCT------------GTACCTCCTCATATTCCTCATAACAAACCAATCAAATTGTTGAGTGTTATGATTGCTCCTATCCCCTCCTTCAATCGACTCAAGAACGGTTGTCGACCCTACGTTGAAGTCTATCAAGGTGATAAGAGGGTTTTCACCTCAATGACGGATTATGAATCGCTACGGACCTATGAATTC---GAGGACCGAAAAATCGAAGTCTTTCTA---AACAACCTCATGGTTTACGGAGATGTAACCATCAGCGTCTACCATGCTCGGTCATTCTTCGCTGGCAAA------GGGAAAGTATCAGCCGTGAAAATTTGTCAGTTTCAATTTCACACCGGTTTCATTGATCCTAGCATCACAAAGATCACATTCCTTAAATCTGAATTGGATCACCTCTACTCACTCCATACAAGTGGCACT---------CAGAAT---GCGAGTATTCCCAGTGGTGCTTCAGTAACAAGTCGGTATGCGGAGAGCTTTCATACTGTTCTTCATCTTGCTGTCAGTCCAACAGAACGTCCACGCCAGGGCTCCACTCTT---GTTTTCCCCTGG------GAAACACTCCCCCCT------GAATCTGCT---------------CGGAAACCCATTCTCTGTTTCACCGACGTCTATGAGATGAGGTCAATCATGCCACCT--CCC------------ACCAAACCACCC---GGCACA------------	MGDFIKSAFGYFGGPNQG----VKDHELVGESVEVGKAHFRIRRV---IADGGYGVVFEGYDSSLGRSFAIKRLFAPDQDEVNVIMNEINVLKQLSGHPNIMTFYGCACADKERGKLAGNEFLIVTELCSGGQLKDYL-PAPHQG----KHLPADIVLQVLSQTSRAIQHMHKQSPPIIHRDLKIENILLSESFTIKLCDFGSATTETLAPNISWSAIERGRAQESLE-KVTTPMYRSPEMLDLYLNYPINEALDIWALGCIMFYLVCGFHPFEDSAKLAILNANYNLP-PSDSDFEPFHNLIRQMLLVDPTQRPNINSVYGELSDLATTRNVPAGFLKGSAGHLIKNIRE--TSSKVMEAV--------------------------------------TAAMP-GDLDLQYITTRIAVMS----YP----A------ESGLEAVVGSRNSMEEAQAYLDRHHPNSYAVYNLSSRQYRSTCWFGGRVSFRRLELNRAPSLASLIELCQNARLWLSQKPDNICVIHCGDGRQLSAVFVSSLLCFCGVFSDASSAARFFNSKRSSSRLTPAQLRYIDYVAKLV--------A----VPPHIPHNKPIKLLSVMIAPIPSFNRLKNGCRPYVEVYQGDKRVFTSMTDYESLRTYEF-EDRKIEVFL-NNLMVYGDVTISVYHARSFFAGK--GKVSAVKICQFQFHTGFIDPSITKITFLKSELDHLYSLHTSGT---QN-ASIPSGASVTSRYAESFHTVLHLAVSPTERPRQGSTL-VFPW--ETLPP--ESA-----RKPILCFTDVYEMRSIMPP---------------------P----TKPP-GT-----VPFTSPKHNPPPQTN----AAPTPGS-GFS---------GARPRVTENAFEDLLGGFNSS-GAAFGSSRKNQQPKTLAEIR--HQKLAQTEDPEVL-----------------------------KV---IEWLDGKEKNVRALLCTLGTVLWEG-

Lollita gigantea	gi|676449849| ref|XP_009052187.1| hypothetical protein LOTGIDRAFT_174596 [Lottia gigantea]	ATGGCCGATTTTTTCAAGTCAGCTTTTGGATATTTGAGTGGGGGGCAGAAC---------------AAAGAAGATAATGACTTCGTAGGTCAATTGGTGGAGTTGGGAAACCAGAAACTTCGTGTTAAAAAAGTC---------ATTGCCGAAGGTGGATTTGCTTTTGTTTTTATTGCCCAAGATGTAAGCAGTGGTAAAGAATATGCATTAAAGAGGCTATTAGCCAATGACGAAGAGAAGAACAATGCTGTTATGAATGAAATTCGATTTTTAAAAAAATTAACAGGTCATCCAAATATCATTCAGTTTATATCAGCAGCAGCTATTAATAAAGACCAGTCGGATCACGGCCAGTCTGAATATCTTATACTTACTGAACTCTGTCCTGGTGGAGAGCTAGTGTCATATTTAAAT---------CGTAATGAG------------TTACAATTAGCTTGTAATCAGATCTTACTAGTATTTCATCAAACATGTCGAGCTGTACAACATATGCATAAACAGAAACCTCCTATTATACATAGAGATTTAAAAGTAGAGAATTTATTAATAAGTGCCCAGGGAACTATAAAATTGTGTGATTTTGGTAGTTCAACAACTAAACCTCAATATCCAGATACATCGTGGACAGCTATACAGCGTAGTTTAGTTGAAGATGAGATTGCT---AAAAATACTACCCCTATGTACAGAGCACCTGAAATGTTAGATTTATATCAGAATTATCCTATAGATGAACGTGGAGATATATGGGCTTTAGGTTGTATATTATTTACACTGTGTTTTAGAGAGCATCCATTTGAAGATTCAGCCAAATTACGTATTATTAATGCAAACTATACCATTCCA---GAAAGTGATAGAAAATATAGCATATTTCATGATCTTATACGTTGTATGTTAAGAGTGGACCCCACTACTAGACCAAATATAAATGATATATTAGATAGATTAAATGATATAGCCAATGCTAGAAATGTGAATATAAATTCTCTACGAGGAGGGGCAGGAAACTTGATGAAGAATATTAAAGAT------GCTTCTAGTAAAGTTATGGAAACAGTT---TCTGCGACTATGAATAAAGGTGATTTAGATATATCATATATAACGTCTAGGCTTATAGTTATGTCA------------TTCCCT------------GCT------------------GAA---GGCGTAGAA---TCA---GCCGTCAAAAACCACATAGAAGATGTACGAACCTTTCTAGAAGCTAAACATAAAAACTGTTACGCTGTTTATAATTTATCACAAAGATCATATCGTGTCGGTTGT---TTTGAGAACAGGGTATCTGAGTGTGGCTGGCCTCAAAGAAAAGCGCCAACCTTAGCTAGTTTATTTGCTATCTGTAAAAATATGCATCTATGGTTACGGCAAAATCCTAAAAATATATGTGTTGTTCATTGCACAGATGGGAGATCTAACTCTGCTACTGTGGTAGGAGCTTTCTTGGTATTTTGTCGTCTGTTTGAAAAACATACGTCTGCTATGCACATGTTTACATCCAGACGCTCACCACCAGGGGTCTCTCCAGCACAAAAAAGATATATAGGTTATATTAGTGAAATGGTG------------------------TTA------------GAAAATCCAGTTATACCCCATTCTTATCCCATAGTATTAGATAGTATAGTCATGACTCCGGTACCGTTGTTTAATAAAATGAAGAATGGTGTGACACCGTTTGCTGAAGTGTTTGTGGGTGAAGAAAGAGTTATGTCTTCCTCTCAAGAATATGAAAGAATGAAACAATTTGTAATA---GAAGATGGTAGAGCCAAGTTAAATTTA---GAT---GTTACTGTTAGTGGTGATGTTACTATTGTAGTTTATCATGCTAGATCAACTTTTGGTGGGAAAATACAGGGCAAGATTACATCTATGAAAATGTTTCAAATCCAGTTTCATACTGGTTTTATTAAACCAGAAACAGAAATTATCAAATATACATTGTATGATTTAGACCAGTTAGATACTCCTGAT---AAATATCCTGACATGTTTCATGTAAAACTAAATGTTCAAGTATTAACTAATAAAAAGACATTAACAAGT------------------------------GAATCCAGACCGCCT------TGGGAAAAATTTGATCGTGAAAAATTAGGACCTAAAATTCTCTTCTCCAGTAAAGAAGAAGTACATGAAGCAATAAGTCCT---CAA------------AGACAGCCATCT---AAATCT---------------AATTAT--AGTAGTGTTATA---GGTGGTAGAGAGGAACGTGGTACCAGATCTCATTTCGGTCCCAAAAAGAAAGTAGATGAAAATGCATTTGATG	MADFFKSAFGYLSGGQN-----KEDNDFVGQLVELGNQKLRVKKV---IAEGGFAFVFIAQDVSSGKEYALKRLLANDEEKNNAVMNEIRFLKKLTGHPNIIQFISAAAINKDQSDHGQSEYLILTELCPGGELVSYLN---RNE----LQLACNQILLVFHQTCRAVQHMHKQKPPIIHRDLKVENLLISAQGTIKLCDFGSSTTKPQYPDTSWTAIQRSLVEDEIA-KNTTPMYRAPEMLDLYQNYPIDERGDIWALGCILFTLCFREHPFEDSAKLRIINANYTIP-ESDRKYSIFHDLIRCMLRVDPTTRPNINDILDRLNDIANARNVNINSLRGGAGNLMKNIKD--ASSKVMETV---------------------------------------SATMNKGDLDISYITSRLIVMS----FP----A------E-GVE-S-AVKNHIEDVRTFLEAKHKNCYAVYNLSQRSYRVGC-FENRVSECGWPQRKAPTLASLFAICKNMHLWLRQNPKNICVVHCTDGRSNSATVVGAFLVFCRLFEKHTSAMHMFTSRRSPPGVSPAQKRYIGYISEMV--------L----ENPVIPHSYPIVLDSIVMTPVPLFNKMKNGVTPFAEVFVGEERVMSSSQEYERMKQFVI-EDGRAKLNL-D-VTVSGDVTIVVYHARSTFGGKIQGKITSMKMFQIQFHTGFIKPETEIIKYTLYDLDQLDTPD---------------------KYPDMFHVKLNVQVLTNKKTLTS----------ESRPP--WEKFDREKLGPKILFSSKEEVHEAISP---------------------Q----RQPS-KS-----NY--------------------SSVI-GGREERGTRSHFGPKKKVDENAFDDLLG------GVNFT---KKSEPKTIAEMR--REQEVEEMDPDKL-----------------------------KV---

*Mesocestoides corti**	MCOS_0000093901-mRNA-1	ATGGGGGATTTCATAAGGTCAGCCTTTGGTTATTTCGGAGGGCCAAGTCAAAGT------------ GTCAAGGAAAATGAGTTGGTGGGCCAACTAGTTGATGTCGGGAAAACACAATTCCGAATCCGTCGAGTT-------- - ATCGCTGAGGGTGGTTTCGCCGTGGTGTTCGAGGGATATGAACATTCCAAGGGGAAATCCTTTGCTATAAAGC GGCTTCTTGCACAAGATAAAGAAATGTGTGATGCTGTAATGCGCGAAGTCGACATTATCAAGCGATTGTCTGG GCATCCAAACATAATCGGCTTTTTTGGAGCTGCCTGCGTTGGGAAGGAAAAATCTAGACACGTAGGGACTGAG TTCCTAATTGTCACGGAATTGTGCTCAGGAGGAGCTCTCGCCGACTTTCTT---CCTACACCTCATCAACAA------- ----- AAGCCTCTGCCCTTGAACATTGTTCTTCAAATTCTCGCACAAACTAGTCGAGCCGTGCAGCACATGCACAAGC AGTCACCTCCAATAATCCATCGGGACCTAAAGGTCGAAAATCTCCTGCTTTCAGAAAAATTCACCATTAAGTTG TGTGATTTTGGTAGCGCCTCGACAGAAACATATGCTCCTACTTTACTTTGGTCTGCCACAGAACGTGGACGAA CTCAGGAGGCCCTGGAA--- AAAGTGACAACACCCATGTACAGATCGCCAGAAATGTTGGATCTTTATCTCAACTATCCCATCAACGAAGCCGT GGATATTTGGGCCCTCGGTTGCATAATCTTCTACCTCGTCTGTGGTTACCACCCCTTTGAGGACTCCGCCAAA TTGGCCATTCTGAACGCCAATTTCAATCTTCCC---CCTTGCGATCCCGATTTCGAGACCTTTCATAATTTAATT--- -- CACGCGTCTGCTGGCTTTCTCAAGGGCAGCGCCGGGCATTTGATCAAAAACATCCGTGAA------ ACCTCGGTCAGGGTTATGGAAGCTGTC-- -----------------ACTGCTGCAATCCCG---ACCGATCTGGACCTACAATACATAACTTCGAGAATCGCAGTAATGTCG- -----------TATCCG------------GCC-----------------GAAAGTGGCTTAGAG--- GTCTTCGGGAGCCGAAATTCGATGGAGGAGGCGCAGTCCTACCTCGATCGACACCACCCCGGCAGTTACGCT GTCTACAATTTGAGCCCCCGCTCGTATCGGTCGAGCCACTGGTTTGGCGGCCGCGTCTCCTTCCGCCCCTTC GAGGCCCATCGCGCGCCCACCCTCTGGTCCCTGGTCGAGTTGTGCCAGAACGCGCGACTCTGGCTCTCCCA GAAGCCCAACAATGTCTGTGTTGTCCACTGCACTGACGGACGCCAGTTGTCCGCTGTTCTCGTCTGCTGTCTG CTCTGCTTCTGTCGAGTCTTCTCCGAGGCGTCCTCGGCGGTCAAGTTTTTCATTTCAAAACGCGGACCTGTTC ATCTCACGCCCGCGCAAATGCGTTACATCGACTACGTGGCAAAATTGGTA-----------------------GCG----------- GTACCTCCACACCTTCCGCACAACCACCCCATTGGTCTGCTCAGCATTATACTCAGTCCTATTCCCACCTTCAA TCGTCTCAAGAATGGCTGCCGCCCCTTTGTCGAAATTTACCAAGGTGGCAAGAAGGTTTTCTCGTCAATGACT GATTATGAGTCCCTGCGAACTTACGAGCTA---GAGGATCGGAAGATGGAGATTATCCTC--- AATGGTTTATCCGTGTACGGAGACGTTACGATCGCCGTCTATCACGCTCGGTCCTTCTTTGCCGGAAAA------ GGAAAGGCCTCGTCTGTCAAAATTTGTCAAATTCAGTTCCACACTGGCTTCATCGGCCAGGAGGTTGAAAAGG TTTCCTTCATGAAGTGTGAGCTGGACCACCTGCACAGCGCGGGCCCGAGTGGCGCCAACGCGCCTCAGGCC- --ACCGCTGCGCCACCC------------ ACCAGTCGGTACGCCGAGAGTTTCCACGCCACCCTCCACCTCACGGTTTCGCCCAACGAGCGTCCGCGAGG CGCCTCCTCCCTG---GTGTTCCCCTGG------GAGACCCTGCCCTCC------GAGGGCGCT-------------- CGCAAGCCCATGCTGTGCTTCTCCGACGCCTTCGAGATGCGCGGTGTCATGACGCCA------------------------------ ---ACCCAC---GGTAGC-------------- AACTCGTGCGCATCTCCCAGGCACAGTCAACCACCACAGTCA---	MGDFIRSAFGYFGGPSQS---- VKENELVGQLVDVGKTQFRIRRV--- IAEGGFAVVFEGYEHSKGKSFAIKRLLAQD KEMCDAVMREVDIIKRLSGHPNIIGFFGAA CVGKEKSRHVGTEFLIVTELCSGGALADFL -PTPHQQ---- KPLPLNIVLQILAQTSRAVQHMHKQSPPIIH RDLKVENLLLSEKFTIKLCDFGSASTETYAP TLLWSATERGRTQEALE- KVTTPMYRSPEMLDLYLNYPINEAVDIWAL GCIIFYLVCGYHPFEDSAKLAILNANFNLP- PCDPDFETFHNLI---------------------------- HASAGFLKGSAGHLIKNIRE-- TSVRVMEAV-------------------------------------- TAAIP-TDLDLQYITSRIAVMS----YP----A------ ESGLE- VFGSRNSMEEAQSYLDRHHPGSYAVYNL SPRSYRSSHWFGGRVSFRPFEAHRAPTL WSLVELCQNARLWLSQKPNNVCVVHCTD GRQLSAVLVCCLLCFCRVFSEASSAVKFFI SKRGPVHLTPAQMRYIDYVAKLV--------A--- - VPPHLPHNHPIGLLSIILSPIPTFNRLKNGCR PFVEIYQGGKKVFSSMTDYESLRTYEL- EDRKMEIIL- NGLSVYGDVTIAVYHARSFFAGK-- GKASSVKICQIQFHTGFIGQEVEKVSFMKC ELDHLHSAGPSGANAPQA-TAAPP---- TSRYAESFHATLHLTVSPNERPRGASSL- VFPW--ETLPS--EGA----- RKPMLCFSDAFEMRGVMTP------------------- --------TH-GS-----NSCASPRHSQPPQS- QPPPPPPPPPS-ASQ--------- GARPWLNEDAFEDLLGDFGAP-GQSFG- SKQNRTPRTLAEIR--HQKLAETEDPEAM--- ------------------------- KIVAVPRGLENEFMLSSGVLRAATMRLHG RASVLWTVITVIEKRWTPCALEGGVAESTS

*Onchocerca volvulus**	OVOC2655 OVP08570 WBGene00239464	ATGACAGAGTTATTTCGGTCAGCGTTCAGCTATTTGTCGCAAACAACACCATCAAATGTTATCGGTAAATTGGATCATCCCCTTGTGGGTACAAACATTGAGATTGATGGACTGAGACTGAAAATTCGTTCTCTG---------ATTGCTGAAGGAGGTTATGCGTTAGTATTTTCGGCCCAGGAC---ACACAAGGTAATTGGTTTGCTCTGAAACGACAATTAGCAGCAGATGGAAAAGCTGCTGAAGCAATATTAAAGGAAATTCGTTTTTTGCGAGAGCTAACTGGTTTTCAATCCATTTTGCGTTTTGTGCAGGCGGCACAGCTGAGCCCACAGGAGAGTGGGCATGGTCGAGCTGAATTTCTTCTGTTGACCGAACTTTGTCCGGGCAAATCTTTGTTTACTTTCTTACTC---------GTGATCGAGCTAATTCAAAAAGGTCCACTGCCCGTTAAACAAGTGACGAAGATTTTTTATGCGGCTTGCAGTGCCATTAGACAAATGCATACAAGAAACCCGTCCATAACACATAGAGATATAAAGATTGAAAATTTGCTTTTTGATGCATCTGGATATGTGAAACTATGCGACTTTGGAAGTGCAACAACAGAAATCGTAACACCGGATGAAACCTGGTCCGCTTTACAGCGCGCACAACTCGAAGAGGAAGTTTTAGCACGGCATACGACGCCGATGTACAGAGCACCAGAATCACTGGATATGTATTCGAATTTCCCAGTGGGTCCAGCCCAAGATATTTGGGCTCTTGGATGTGTACTTTATTATCTTTGCTACAGAAAACATCCGTTTGAAGACAGTGCAAAACTTCGCATTATTAACGCGAAATATTCACTTCCT---GATGCAGAATCAGAATATGCCCAATTGAATCAACTCATTCAAGCAACATTACAAGTAGATCCAAGACTACGTCCAAATGCGGAAGATTTATGTGAACGAGTTGAAGCGCTTGCTGCGGCATTATACATAGATCTTGGTGCTATTAAGGAGCAAAGCATGTCTTGGTTCAAAAATATCAAAGAC------AAAACAGCCGCTGTTGCGCAAACAGTA--CAACTGACTTATGGTAATCGAGGTCCAGATGTAACATTTGTTACATCGCGTCTTGTAATAGCGCCACTTGCTGATTGTATACCTGAAGCACTGATTGCACAGACTGAAGAAGCAATGCGGGCCCGTATT--------------------------TTGGACCAGGCTCGAGGACAA------------------------TTTGCAATTTATAACTTATCGAATCGGTGGGTTTTTGTATTGAAA---GAATTGATTGGGCAAAGAAATATAGGTCTTCATAGTAATGGAATAAAATGTTTGCGGTGTGATTACAGTCATTTGATAGAAACTCCTCTGCCAGTTGTCGAAAGTGCTCAATGTATA--------------------------------------CTGGTAGCAGCAACTCTGTTATTATATGCTCGACTTGTACCAAGGCCACTGGCAGCTCTAGAATTAATTTGTACCAAAAGACAACCACCAAATCTTCCACCATCTTATCATCGTCAGCTTGATATAATAAAAACAATTGTT-----------------------TCCATGGATTCATCTGATTTATGTTTAATGGTTCACAATAGACGGGTAGTGCTTCAATCATTACTGATATCTCCGGTGCCATTATTTAATCGAGCTCGAAGTGGTTGTCGACCATTTGCTGATATTTACGCTGGTGGTGTGAAGGTGTGGTCTACATGCAAAAACTATGAAGAATTAAAGTCTTTTGAAGTTCCCGAATCAGTTTTAATAGATCTTCCTTTGGGAGAT---ATTCCAGTTGGTGACGATGTTCAGATAGTTGTTTATCATGCACGACTGATG---------AAGATGCAAAATCGTATGCAGCAATATCTAATGTTTTCACTGAGCTTTCATGCTAATTTCATTGATCCAAATATCCATATGTTGGAATTCGGTCTTGGTGATCTTGATCGTCAG------TTTGATGAT--TTGAGATTTGGATTGTCGCTAAGGGTTGGCTTACAGTTGAAAATTGATGAACATGATAGAGGATTTAGTACG----------------------------AGAGAGCCACCAGCA------ATTCTTAGT--------------TACGAACCAAAAACTATT-------------------------------AATAGT---CAG-----------CAACAACCTAATTACAGTCGAGCACATTTTGATATTCCC--AGTTCAGTTAAG---GGT---	MTELFRSAFSYLSQTTPSNVIGKLDHPLVGTNIEIDGLRLKIRSL---IAEGGYALVFSAQD-TQGNWFALKRQLAADGKAAEAILKEIRFLRELTGFQSILRFVQAAQLSPQESGHGRAEFLLLTELCPGKSLFTFLL---VIELIQKGPLPVKQVTKIFYAACSAIRQMHTRNPSITHRDIKIENLLFDASGYVKLCDFGSATTEIVTPDETWSALQRAQLEEEVLARHTTPMYRAPESLDMYSNFPVGPAQDIWALGCVLYYLCYRKHPFEDSAKLRIINAKYSLP-DAESEYAQLNQLIQATLQVDPRLRPNAEDLCERVEALAAALYIDLGAIKEQSMSWFKNIKD--KTAAVAQTV---------------------------------------QLTYGNRGPDVTFVTSRLVIAPLADCIPEALIAQTEEAMRARI---------LDQARGQ--------FAIYNLSNRWVFVLK-ELIGQRNIGLHSNGIKCLRCDYSHLIETPLPVVESAQCI-------------LVAATLLLYARLVPRPLAALELICTKRQPPNLPPSYHRQLDIIKTIV--------SMDSSDLCLMVHNRRVVLQSLLISPVPLFNRARSGCRPFADIYAGGVKVWSTCKNYEELKSFEVPESVLIDLPLGD-IPVGDDVQIVVYHARLM---KMQNRMQQYLMFSLSFHANFIDPNIHMLEFGLGDLDRQ--FDD------------------LRFGLSLRVGLQLKIDEHDRGFST----------REPPA--ILS-----YEPKTI-----------NS---------------------Q----QQPN-YSRAHFDIP----------------------SSVK-G-----------GKSKNFENAFGDLLT------SQGFQTSPKTMK--SLGEMK--RQDEIKELDPVKV---------------------------KI---KEWTNGKERNIRALLGSMNNILWPNAENWIQPSIGDLLTAQQVKKYYRKACLVIHPDKQV----------GTENEAL--------------ARAIFTELNDAWTA------------------YE

Schistosoma haematobium	gi|844871592|ref|XP_012800557.1| Cyclin-G-associated kinase [Schistosoma haematobium]	ATGGCTGACTATCTGAAGTCTGCATTAAGCTATTTCGCCACCTCAAACACTACT------------AAAAATGAAAATGAATTTCTTGGCAGTAGTATATCTGTCGGGCAACTAAGTTTAAAAGTTAAGAGAATC---------ATTGCTGAAGGTGGCTATGGAATAGTTTATGAAGCTCAAGATGTCAATGAAAATACATCATATGCGTTAAAGCGTATGCTCGCTCACGACAAACCTTCAATGGATTTAATTCTTCATGAAGTTCGTTTACTGAAGCAATTAAACGGACATCCAAACATCCTTAAGTTTTTTAGTGCAGCATCTGTTGGTAAAGAGAAAATGAAGGTTATTGGTACTGAGTTCTTAATAGTCACAGAGTTTTGTAAAGGGGGTCAGTTGGATAAATATTTA---CCGGCATCAAAATGTGAA------------AATCCCCTGCCATCGAACGTTATCTTACAAATATTTCATCAATGTTGTCGTGGTGTACAACATATGCATAGTCAATGTCCACCTGTGATACATCGAGATTTGAAAATTGAAAATCTATTACTTACTGATAATTTTATCATAAAGTTATGCGATTTTGGCAGTGCAACTACAATTACATATAGTCCAGATCAATCATGGACTGCTTTAAAACGTGGAAGTGTTCAAGAAGAACTGGAG---CGTTTCACAACACCAATGTATCGGGCTCCAGAAATGCTTGATCTTTACCAAAATTATCTTATTGGTACTCCATCGGATATTTGGGCTTTAGGTTGTATATTATTTTATTTAACATGTACTTATCATCCATTTGAGGATTCATCGAAATTGGCTATTCTCAATGCTAATTATAGTATACCGGCATCAATAAGTTCAGTGAATGCACCATTTTGCAGTTTAATTCGACAATTACTTCTCATAAATCCATCTCAACGTCCTAACATAAATGAAATTCTTGGTGAGTTATCTGAATTGGCGTCTATGAGAGAAGTTCGAGTAGGTATGATTAAAGGAGGCGCTGGCAATTTAATTCGTAATATACGAGAT------GCATCAACTAAAGTTATTGGGACTGTT---TCTACCACGTTAAAT---ACAGAGCTTGATTTCCAGTTTATCACTTCACGTATAGCAGTTACATCT------------TTCCCC------------ATT------------------GAAAGTGGATTTGAA---GTGCTTGCCAATGTGAATTCCATGGAAGAAATGCAAAATATGCTAGATGCACGTTTTTCAGATGCTTATGCTGTTTATAATCTTAGTAATCGTCCATATCGTTCAGATCATTGGTTTCATGGTAGAGTATCACATCGTGGTTTCGAAGCACACCGTGCTCCAGCGCTAAAATCATTAATTGAATTATGCTTAAATGCAAGGTTATGGTTAGCTCAAAAATCGAACAATATTTGTGTTATACATTGTACCGATGGTAAAACATTGTCCGCAGTACTAGCATGTTCCTTATTATGCTTTTGTCGGGTATTTGATAACGCGTCACCTGCAATTCAATTGTTTGCATCGAAACGTGGCAATCCTGGCTTGAATGCATCACAGATTAGATATATCAACTATGTAGCTCAGTTAAGT-----------------------CAT------------AATCGTGTATTTACACCTCATCATTTCCCATTGAATCTTATCAGTTTGATAATTGCACCGGTACCAACATTTAATAAATCCAAAAATGTATGTCGACCGTATGTGGAAATATTTGAAGGAAAGACCAAAGTATACTCAACTTACGCAGAATACGATAATTTAAGATCCTATGTTCTT---GAAGATGGAAAGATCCAAATACTTTTG---AATGGATTAACTGTCTTAGGTGATCTTACAGTTATCATTTATCATTGTCGATCTTCATTTGCTGGACGT------GGCAAGATCAGTTCAGTGAAGATCGCTCAGTTTCAATTACACACTGGATTTGTAGAACATAATTTAACAGAACTAATTTATTTTAAATCCGATTTAGATCATCTAGATAATAGTAGCAGCTTTGCCGGT-----------------------------TTC------------ACTAGTCGTTATGCTGAAAGTTTTCATGTTACATTGGAATTTATGGTTTCACCAACTGAAAGACCAAGACAAAGTGACAATCTT---GTTTATCCGTGG------GAAACTTTGCCATCA------AATGATCTA---------------CTTAGTCCTAAACTCTGTGTATCAAATCATGATGAACTGAAAAATATCTTAAGTCCTGATAATTTATTCCATGGTTATGAAAAATCTAGTCAATCGTATTCTTCCTACTCCAATTTTCCTAAT------------ACTACTACTTCT---GGATCT---------------ACATCA---ATACCAACTAGT---GGAGGAACT--------------------------GGTGCTAGACCTCATGTTAATAAAGATGCATTTTCTGATTTACTCGGTGGATTTGGTTCTTCT---CGAACAAAC------AACACAGATGATAATAAACAACCGAAAACAGTAAATCAGATTCGT------	MADYLKSALSYFATSNTT----KNENEFLGSSISVGQLSLKVKRI---IAEGGYGIVYEAQDVNENTSYALKRMLAHDKPSMDLILHEVRLLKQLNGHPNILKFFSAASVGKEKMKVIGTEFLIVTEFCKGGQLDKYL-PASKCE----NPLPSNVILQIFHQCCRGVQHMHSQCPPVIHRDLKIENLLLTDNFIIKLCDFGSATTITYSPDQSWTALKRGSVQEELE-RFTTPMYRAPEMLDLYQNYLIGTPSDIWALGCILFYLTCTYHPFEDSSKLAILNANYSIPASISSVNAPFCSLIRQLLLINPSQRPNINEILGELSELASMREVRVGMIKGGAGNLIRNIRD--ASTKVIGTV-------------------------------------STTLN-TELDFQFITSRIAVTS----FP----I------ESGFE-VLANVNSMEEMQNMLDARFSDAYAVYNLSNRPYRSDHWFHGRVSHRGFEAHRAPALKSLIELCLNARLWLAQKSNNICVIHCTDGKTLSAVLACSLLCFCRVFDNASPAIQLFASKRGNPGLNASQIRYINYVAQLS--------H----NRVFTPHHFPLNLISLIIAPVPTFNKSKNVCRPYVEIFEGKTKVYSTYAEYDNLRSYVL-EDGKIQILL-NGLTVLGDLTVIIYHCRSSFAGR--GKISSVKIAQFQLHTGFVEHNLTELIYFKSDLDHLDNSSSFAG----------F----TSRYAESFHVTLEFMVSPTERPRQSDNL-VYPW--ETLPS--NDL-----LSPKLCVSNHDELKNILSPDNLFHGYEKSSQSYSSYSNFPN----TTTS-GS-----TS--------------------IPTS-GGT---------GARPHVNKDAFSDLLGGFGSS-RTN--NTDDNKQPKTVNQIR--REKMAKTVDPEQEKIKNCRFIHTTVTVNFPSVKSSNTISGHYKV--CDWAEGKDRNLRALLCSLPAILWDG-

Schistosoma mansoni	gi|360044570| emb|CCD8211 8.1| serine/threonin e kinase [Schistosoma mansoni]	ATGGCTGACTATCTGAAGTCTGCATTGAGCTATTTCTCCACCTCAAACACTACT------------ AAAAATGAAAATGAATTTCTTGGCAGTAGTATATCTGTCGGGCAACTAAGTTTAAAAGTTAGAAGAGTT--------- ATTGCTGAAGGCGGCTACGGAATAGTTTATGAAGCTCAAGATGTCAATGAAAATATATTATATGCGTTGAAGCG TATGCTCGCTCACGACAAACCTTCAGCAGATTTAATTCTTCATGAAGTTCGTTTGCTGAAACAGTTGAACGGAC ATCCAAACATCCTTAAGTTTTTTAGTGCAGCATCTGTTGGTAAAGAGAAAATGAAGGTTATTGGTACTGAGTTC CTGATAGTCACAGAGTTTTGTAAAGGGGGTCAGTTGGATAAATATTTA---CCGGCATCAAAATGTGAA------------ AATCCCCTGCCGTCGAACATTATCTTACAAATATTTCATCAATGTTGTCGTGGTGTACAACATATGCATAGTCAA TGTCCACCTGTGATACATCGAGATTTGAAAATTGAGAATTTATTACTTACTGATAATTTTATCATAAAGTTATGC GATTTTGGTAGTGCAACCACAATTACATATAGTCCAGATCAATCATGGACAGCTTTAAAACGTGGAAGTGTTCA AGAAGAGTTGGAA--- CGTTTCACAACACCGATGTATCGGGCCCCAGAAATGCTTGATCTTTACCAAAATTATCCTATTGGTACACCATC GGATATTTGGGCTTTGGGTTGTATATTATTTTATTTAACATGTACTTATCATCCATTTGAGGATTCATCGAAGTT GGCTATTCTCAATGCTAATTATAGTATACCAGCGTCAATAAGTTCAGAGAATGCACCATTTTGCAGTTTAATTCG ACAATTACTCCTCATAAATCCATCTCAACGTCCTAACATAAACGAAATTCTTGGTGAGTTATCTGAATTAGCGTC TATGAGAGAAGTTCAAGTAGGTATGATTAAAGGAGGCGCTGGCAATTTAATTCGTAATATACGAGAT------ GCATCAACTAAAGTTATTGGAACTGTT--- ----------------TCTACCACGTTGAAT---ACAGAGCTTGATTTCCAGTTTATCACTTCACGTATAGCAGTTACATCT----- -------TTCCCC------------ATT------------------GAAAGTGGATTTGAA--- GTGCTTGCTAACGTGAATTCTATTGAAGAGATGCAAAATATGCTAGATACACGTTTTCAAGATGCTTATGCCGT GTATAATCTTAGTAATCGTCCATATCGTTCAGATCATTGGTTTCATGGTAGAGTATCACATCGTGGTTTCGAAG CACACCGTGCTCCTACGCTAAAATCATTAATTGAATTATGCTTGAATGCAAGGTTATGGTTAGCTCAAAAGTCG AATAATATTTGTGTTATACATTGTACTGATGGTAAAACATTATCCGCTGTATTAGCATGTTCCTTATTATGCTTTT GTCGGGTATTTGATAACGCATCACCTGCAATTCAATTGTTTGCATCGAAACGTGGCAATCCTGGTTTGAATGCA TCACAAATTAGATATATTAACTATGTAGCTCAGTTAAGT------------------------CAT------------ AATCGTGCATTTACACCTCATCATTTTCCATTGAATCTTATCAGTTTGATAATTGCACCGGTACCGACATTTAAT AAATCCAAAAATGGATGTCGACCGTATGTGGAAATATTTGAAGGGAAGACCAAAGTGTACTCAACTTACGCAG ATTACGATAATTTAAGATCCTATGTTCTT---GAAGATGGAAAGATCCAAATACTTTTG--- AATGGATTAACTGTTTTAGGTGATCTCACAGTTATCATTTATCATTGTCGATCTGCATTTGCTGGACGT------ GGCAAGATCAGTTCAGTAAAGATTGCTCAGTTTCAATTACACACTGGATTTGTAGAACATAATTTAACTGAACTA ATTTATTTCAAATCCGATTTAGATCATCTGGATAATAGTAGCAGCTTTGCCGGT-----------------------------TTC------- ----- ACTAGTCGTTATGCTGAAAGTTTTCATGTTACATTGGAATTTATGGTTTCACCAACTGAAAGACCAAGACAAAGT GACAATCTT---GTTTATCCATGG------GAAACTCTGCCATCA------AATGATCTC--------------- CTTAGTCCTAAACTCTGTGTATCAAATCATGATGAACTAAAAGATATCTTAAGTTCTGATAATTTATTCCACGGT TATGAGAAATCTAGTCAATCATATTCTTCCTACTCCAATTTTCCTAAT------------ACTACCACTACTACTGGATCA-- -------------ACATCA---ATACCAAATAGT---GGAGGAACT--------------------- ------GGTGCTAGACCTCATGTCAATAAAGATGCATTTTCTGATTTACTTGGTGGGTTTGGTTCTTCT--- CGAACAAAC------AACACAGATGATAATAAACAACCAAAAACAGTTAATCAGATTCGT------	MADYLKSALSYFSTSNTT---- KNENEFLGSSISVGQLSLKVRRV--- IAEGGYGIVYEAQDVNENILYALKRMLAHD KPSADLILHEVRLLKQLNGHPNILKFFSAAS VGKEKMKVIGTEFLIVTEFCKGGQLDKYL- PASKCE---- NPLPSNIILQIFHQCCRGVQHMHSQCPPVI HRDLKIENLLLTDNFIIKLCDFGSATTITYSP DQSWTALKRGSVQEELE- RFTTPMYRAPEMLDLYQNYPIGTPSDIWAL GCILFYLTCTYHPFEDSSKLAILNANYSIPA SISSENAPFCSLIRQLLLINPSQRPNINEILG ELSELASMREVQVGMIKGGAGNLIRNIRD-- ASTKVIGTV------------------------------------- STTLN-TELDFQFITSRIAVTS----FP----I------ ESGFE- VLANVNSIEEMQNMLDTRFQDAYAVYNLS NRPYRSDHWFHGRVSHRGFEAHRAPTLK SLIELCLNARLWLAQKSNNICVIHCTDGKTL SAVLACSLLCFCRVFDNASPAIQLFASKRG NPGLNASQIRYINYVAQLS--------H---- NRAFTPHHFPLNLISLIIAPVPTFNKSKNGC RPYVEIFEGKTKVYSTYADYDNLRSYVL- EDGKIQILL- NGLTVLGDLTVIIYHCRSAFAGR-- GKISSVKIAQFQLHTGFVEHNLTELIYFKSD LDHLDNSSSFAG----------F---- TSRYAESFHVTLEFMVSPTERPRQSDNL- VYPW--ETLPS--NDL----- LSPKLCVSNHDELKDILSSDNLFHGYEKSS QSYSSYSNFPN----TTTTTGS-----TS---------- ----------IPNS-GGT--------- GARPHVNKDAFSDLLGGFGSS-RTN-- NTDDNKQPKTVNQIR--REKMAKTIDPEQL-- --------------------------KV--------------------------- ------------------------------- FDWAEGKDRNLRALLCSLPAILWDG- AQWNHVGMADLITRDQVKRQYRKAARVV

*Taenia solium**	TsM_001047700	--ATGTATAAATATATAATTTTAGGTGGTTACGCTGTCGTATTTGAGGGCTACGATTCATCTCAAGGGAAGTCGTTTGCTATAAAGCGTCTTTTTGCGCAAGATCAGGAAACTGTTTCTGTTGTT	--MYKYIILGGYAVVFEGYDSSQGKSFAIKRLFAQDQETVSVVMNEIDVLKRLSGHPNIMKFFGVACAGKERGKLAGNEFLIVTELCSGGPLKDYL-PPSHQG----KHLPPNLVLQILAQTSRAIQHMHKQSPPIIHRDLKIENILLSESFTIKLCDFGSATTETFAPTVAWSAVERGRVQESLE-KVTTPMYRPPEMLDLYLNYPINEALD

*Trichuris muris**	TMUE_s0017000900|cyclin_G_associated_kinase	ATGGTAGAGCTGTTTCGCAATGCATTCAACTACTTGACGTCGAGCAACGAG---------------AAGCAGGACAATGAATTTGTCGGCAGAGTGATAGTGCTCGGCTCGAAACGTCTGTTGGTCCGTCGATTG---------TTAGCGGAAGGCGGTTTCGGTTACGTGTTTGTCGTCACCGAC---AGCGATGGCAACTCATACGCGTTGAAGCGTTTGATTGCGGGCGATAAGGATTCAGCAGATGCAATTCGTCGGGAGATCTTTATTCTGAAGGAAGCCTCTGGACATCCGAACATACTTCATTTCTGCCAAGCTGCTTGC---------GAGCAGCAGCCAAACGGGCGAACGGAGTTCCTCATATTGACGGAATTGTGCTCAGGCGGACCTCTTTTGGATCGGCTCAGG---------GCTCGCAGG------------ACGCCCTTGGAATTTTGTGAGGTGCTTCCGCTGTTCTATCAAATATGCTGCGCCGTCGATCACTTGCACGGCCACAAACCGCCGATAATCCATCGCGATCTGAAGATGGAAAATTTGCTGTTGGATAGCATGGAGCGAGTCAAGTTATGCGACTTTGGCAGCGCTACCGAAAAGTCCTACAAGCCGGACGACACGTGGACCGCGCAAAGGCGATCCATGCTGGAGGAGGAGTTGAAC---AAGTGCACTACTCCAATGTACCGGGCACCGGAAATGCTCGACTCGTACCAGAATCTACCTATCGACCAAAGAATAGACGTTTGGGCTCTTGGCTGCATACTGTACTACCTTTGCTACATGGTTCATCCGTTCGAGGATAGCGCGAAGCTGCGAATCTTGAACGCCAACTACACGCTGCCC---GAGAAGGACGATCGCTGCGGCGTTCTGCATTCCTTGATTGGGAAAATTTTGCAAGTGACTCCCGACGATCGTCCGTCGGTGAAAGAGATTTTGCGCTTGCTGGAGGACCTGGCGGCGTGCTACGGGGCGGACCTGGAGTCGCTGAAAGGCCAGGCCGGCTCGATATTCAAAAGTCTGATTGAA------ACGTCCTCAAAGATAGTGCAGCCAACC---TCCGCAATGCCAAGCAGGAATTCCCTT------ACTTACTTAACAAGCCGTTTGGTCCTTCTGTCG------------TCTCCA------------GTG------------------GATGTTGGTTGT---------------------------TTGGAAGATTCGGTGAAGCAGTTGACCGCTCGCCACGGCAAACGCTTCTTTGTTTACGACTTATGCTGTTGGTACAATGGCGACTTGCCC---TTAGGGGAAAGGGTGATGCGCTGTCGCTTTCCCGCTGGTTCGGCACCTACGTTGAAGTCTATGTTCTCGCTTTGTAAAAATGTGTACCTATGGCTCCGTCAGGACGTTGCAAATGTCGCCGTCTTCTTT---TCCGACAACGAAGGGAACAGCGCGTGCGTTGCGTGCAGTTTTCTTGTCTTTTCCAAAAATCTCTCGCATTCGGCTCACTGTCGCCAGCTGTTAACGGACAGAGGATGCACCGCTCCCTTGACGCCTTCTCAAGCGAGATATGTAGATTACGTTGCCAGCGTGCTG------------------------CGG------------CATCCGACTTTCTATCCGCAGGCCAGCAAAGTCGTAGCCTCCAAGTTATCCGTTAGTCCAGTGCCAATTTTCAATAGGATGCGGAACGGATGTCGACCGTTCGTTGAGGTATTCAGCGGAGACAATAAATTGTTGTCCACTTGCCAGGAGTACGAGAAGATTCCGTTCATTGACGTG---TCTGAGGCGGGCTTCACCATTCCGTTG---AAT---ATGAAGTTCGACGGCGACGCTACCTTCGTTGTAAGCCATGCTCGGTCGACGCTGGGCACGAAAATGCAAAGGAAGCTCAACATTGTAAAGATGTTCCAATTTCAAGTGCACGCCGGATTTTTGGACCCTTCAAGGAGCGCGTTGACCCTTGAGCTAAACGACCTGTGCTGCATC------GGTGAGGAG---TGGAAGTATCCCAGACCTTTTGTGGTTCGCCTAGCCTACGTGGCAACCGAGGATGAGCGGTGCAACGATAGC------------------------TTGGGTCAGTTCTTGCCAAACTTCAATCCGGATGCC---------------ATTGTCACACATTTGCCGTTTTCTTCTCCCGAGGAGAGCGCCAACTTCCAGCCTGGA---GAA------------TGCAAGGCAAGG---GCGGAT---------------ATGTGG--TCGTCAGCTACT---TCG------------------------------------	MVELFRNAFNYLTSSNE-----KQDNEFVGRVIVLGSKRLLVRRL---LAEGGFGYVFVVTD-SDGNSYALKRLIAGDKDSADAIRREIFILKEASGHPNILHFCQAAC---EQQPNGRTEFLILTELCSGGPLLDRLR---ARR----TPLEFCEVLPLFYQICCAVDHLHGHKPPIIHRDLKMENLLLDSMERVKLCDFGSATEKSYKPDDTWTAQRRSMLEEELN-KCTTPMYRAPEMLDSYQNLPIDQRIDVWALGCILYYLCYMVHPFEDSAKLRILNANYTLP-EKDDRCGVLHSLIGKILQVTPDDRPSVKEILRLLEDLAACYGADLESLKGQAGSIFKSLIE--TSSKIVQPT--------------------------------------SAMPSRNSL--TYLTSRLVLLS----SP----V------DVGC---------LEDSVKQLTARHGKRFFVYDLCCWYNGDLP-LGERVMRCRFPAGSAPTLKSMFSLCKNVYLWLRQDVANVAVFF-SDNEGNSACVACSFLVFSKNLSHSAHCRQLLTDRGCTAPLTPSQARYVDYVASVL--------R----HPTFYPQASKVVASKLSVSPVPIFNRMRNGCRPFVEVFSGDNKLLSTCQEYEKIPFIDV-SEAGFTIPL-N-MKFDGDATFVVSHARSTLGTKMQRKLNIVKMFQFQVHAGFLDPSRSALTLELNDLCCI--GEE-------------------WKYPRPFVVRLAYVATEDERCNDS--------LGQFLPNFNPDA-----IVTHLPFSSPEESANFQPG---------------------E----CKAR-AD-----MW-------------------SSAT-S------------ARAKLDSTAFEDLLS------SHGFSGSASSKQ--SLASMK--QEAECQGLTEEEA-----------------------------

Groucho protein	*Capitella teleta*	gi|443689787|gb|ELT92095.1| hypothetical protein CAPTEDRAFT_168379 [Capitella teleta]	ATGTACCCC------------AACAGACACCCGGGCCCCCACACCCCTGGCCAGCCATTCAAATTCACTGTGGCTGAATCATGTGACCGAATCAAGGAAGAGTTCAGCTTTTTGCAAGCACAATACCACAGTTTGAAGATGGAATGTGAGAAGCTAGCTCAGGAGAAGACGGAAATGCAGAGACATTACGTCATGTACTACGAGATGTCCTATGGTTTGAATGTCGAAATGCACAAACAGACTGAAATTGCCAAGCGGCTCAATGCCATATGCGCACAGATCATTCCTTTTCTCTCCCAAGAG------------------------------------CATCAGCAACAAGTTGCGGCTGCTGTCGAAAGAGCAAAACAGGTCACCATGCAAGAACTCAACGCTATAATTGGGAACGCTGATTTGAAGCAGATGTACGCTCAGAGCGAATGGTTCAACCAAGTTAAA--CAGATGCAGGCTCAGCAAATG---GGTGGTCCG---CATGGCCAT---GGT---------CCCCCT---------------------------ATGCCGTTACCCCCGCACCCCGGAGCGGGTCTTCCCGGTCCAGTACCCCCAGGACTTCCCCCGAGTTCCACGGCC--AGTTTGATC---GCCAGTCTTGGCAGCGCGGGA---ATACCGGGTTCCGCTTCACATCTGCTCAGTGGCGCCTCCTCTGGCACGTCTCCAGATCGCGTTCTC---------------------GAACGCAGC------------------------------------AAAATGGGCAGAAGTCGCAGTCGTTCGCCCGACATAAACAGT---------------GAAGCGCTGAAACGCCTAAAAACAGAGGACAAA------GTGCAGATG---CCAGGGATGCCGGGGTTTGATCCGCAC------------------------------CCCCACATGCGGGGGCCTCTGACT--AGTATTCCC---GGGGGTAAACCTGCGTACTCGTTCCACGTCAGCCATGACGGCCAA---ATGCAGCCGGTGCCCTTT---------CCG------CATGATGCCCTTATTGGGCCTGGAATTCCACGACACGCCCGCCAGATCAACACCCTGAATCACGGTGAGGTCGTGTGTGCTGTGACCATC------------------------------------CCGTGTCGACACGTCTACACCGGAGGCAAGGGATGCGTCAAGGTCTGGGACATCAGCCAACCG---------------GGCAAC---------AAG---AGT---------------CCCGTTTCACAGTTGGACTGCTTACCCCACGACAACTACATTCGCTCAGTGCGTTTGGCCAATGAGGGAACGACCCTGATCGTTGGCGGAGAAGCCTCGTTCCTCTGCATGTGGGATCTAGCTGCGCCAACGCCTCGTATCAAAGCTGAGCTGACCTCGAGCGCCCCTGCCTGCTATGCTCTCGCTATGAGCCCAGACTCCAAGGTTTGCTTCAGTTGCTGCAGCGACGGAAATATTGCGGTTTGGGATCTGCATAATCAGACACTAGTCAGACAATTCCAAGGCCATACTGATGGCGCCAGTTGCATTGATATATCTCCCGATGGCACTAAGTTATGGACCGGAGGTCTGGACAACACAGTGCGCTCGTGGGATCTGCGG--GAAGGCAGG------------------CAGCTGCAGCAGCATGACTTCACTTCGCAGATCTTCTCCCTCGGGTACTGTCCGACC---GGAGAG---------------------TGGTTGGCTGTAGGAATGGAGTCGAGTAACGTTGAAGTT--CTTCAT---------------CACTCCAAACCA------------------------GACAAGTACCAGCTGCATTTGCACGAAAGCTGCGTACTTTCACTCAAGTTTGCGTACTGTGGGAAATGGTTTGTGAGTACTGGCAAAGATAACCTACTGAATGCGTGGAGAACACCCTATGGAGCCAGCATCTTCCAGTCTAAGGAATCGTCGTCAGTGTTGTGTTGTGATATATCAACGGACGATAAATACATTGTGACTGGATCGGGAGACAAGAAA	MYP----NRHPGPHTPGQPFKFTVAESCDRIKEEFSFLQAQYHSLKMECEKLAQEKTEMQRHYVMYYEMSYGLNVEMHKQTEIAKRLNAICAQIIPFLSQE------------HQQQVAAAVERAKQVTMQELNAIIGNADLKQMYAQSEWFNQVK----------------------------QMQAQQM-------------------GGP-HGH-G---PP---------MPLPPHPGAGLPGPVPPGLPPSSTA------------------SLI-ASLGSAG-----------------IPGSASHLLSGASSGTSPDRVL-------ERS------------KMGRSRSRSPDINS-----EALKRLKTEDK--VQM-PGMPGFDPH----------PHMRGPLT----------------------------SIP-GGKPAYSFHVSHDGQ-MQPVPF---P--HDALIGPGIPRHARQINTLNHGEVVCAVTI------------PCRHVYTGGKGCVKVWDISQP-----GN---K-S-----PVSQLDCLPHDNYIRSVRLANEGTTLIVGGEASFLCMWDLAAPTPRIKAELTSSAPACYALAMSPDSKVCFSCCSDGNIAVWDLHNQTLVRQFQGHTDGASCIDISPDGTKLWTGGLDNTVRSWDLR------------------EGR------QLQQHDFTSQIFSLGYCPT--GE-------WLAVGMESSNVEV---LH-----HSKP--------DKYQLHLHESCVLSLKFAYCGKWFVSTGKDNLLNAWRTPYGASIFQSKESSSVLCCDISTDDKYIVTGSGDKKATLYEVIF-----?---------

Clonorchis sinensis	gi|358334530| dbj|GAA53008 .1| protein groucho [Clonorchis sinensis]	-- -- -- --- ATGTCTGCTTTATCGGGCACTCCTCCCTCAAGCCTCCAGACGGGATTACATGGTCTGAATATGTCTCACGGTG GCAGTGGCCCA---AGCTACAGC---GCC---------GCCACC--------------------------AATCCGTCTAATTCCCTTCCT----- -GGTGTCCTCCCTTCT------------ATCTCCCCACCAAGCAACCCA-- GCTATGTCCGCAGCCCTTTTAAAT------GGA--- TTAATGTCAGCTGCGGGAGCC--AATCCACTTGGTCCTAGTGGCTTT----------------- ---------------------------ATGGTACCAACGACGGGTGCAACAAGTTTGAATTTTGGACGG-------------- GAATCAGAAAAA--- AACTCTGCGGACGACAAACAACGTGCTCAACTGAAGTCCGAAAGTAGTTCCTCGTACCCAACA-------------------- ---------ACCTCAAGTGCCCCTGCTGTCGGTTCAGGCGGGAGTGGTGGT--- GGCCGGTCGTCGCAGGTATCTGAATTTTCGGCTGGT---ACTCGTTGGCGGGGTTCAAGCAAAAGC---TCC--- TCCGGACCCGGTGCTTATTCTTTCGTAGTCCTTCCTGGTGGTCAA---CTCAGACCTTGTGCGCTA--------- GCATCTGCTCCAGGTGCTGCAACTGCCTCAGGTTTACCCAAGCGGCTTCAGCACCTCGCCAGTTTGCCACAT GGAGATGTGGTTTGTGCAGTCACGATCGGAGCATGCCCGGTCGGTCGT--------- GCCTCCGCTGCTCACTTCGCTTACACCGGTGCTCGGGGCAGCGTCAAGCTATGGGACCTGTCAGCGATCAGT GCAACAAACTTGTCCGGG---------ATG--- TCAACGAGAGATATTGCTCCTCTAATGAATTTCGACTGCCTTTGCCCAGACAGTTATGTTCGGTCGATCAAGTT GTTCCCGGATGCGAGTCACCTCATCATAGGCGGCGAATCAAACGCGTTAACCTTATGGGATTTGAACGGTCC CGGGCGTCGT--- AAAGCAGAATTAACGTTTGAAGCCCCTGCGTGCTACGCTTTGGCCTTATCACCCAATGGAAAACTTTGTTACAG CTGTTGCTCCGATGGTTATGTTTCTGTGTGGGATGTGCACAATCAGAGTGTCGTCCATCAGTTCCACGGTCAT ACAGACGGAACCTCTTGCGTTGAACTGTGTCCTGATGGAAACCGTCTATGGACCGGTGGTTTGGATCACAAAG TCTGCTGTTGGGATATTCGTGCTCCC-------------------------------------ACGTCC---------CGC----------------- TCTCTCGCCTTCATGGAATTCAAATCCCAAGTTTTCTCCTTGGGGCTGACCACTAATTACGGTTCTCCACATGG GCATGGTTTCGCTGGTCGACGGCAG------ ACTGCTTCGGTTTCCCCTCGATCAGGCGATGGGGACTCAGCCTCTAGCACCGGCTTG-------------- ATGCGTGACTCATCG------AGTCCCCAGTCT--- GTCAGTGGCGGCGGGGGTGTCGGAAGCAGGGGTGGTTCGGGTGTTGGCAGTTGGTTAGCTGTGGGTCTTGA GTCATCCGAGGTCGAAGTGGTTGCAATCGGTCCTGATGGGTTCGCCGTCAGCAGT----------------------------------- ------TCGAATTCCGCTGCTGGATCGCGTGATGAATCTAGTCCCGCGTTG-------------------- TCTCCCATGTCCTATTCTCAACCCCCG---------------ACACCGCAACCT----------------------- CAACACTTCCGATTAACACACCACGAGAGTTGTGTTTTAGCCCTCCGATTCGCCCACCATGGTGATTGGTTCC TAACTACAGGCAAAGATCATCAGGTGAACGCTTGGCGTACACCGTATGGTGCTTGTCTCCTAGAGACCAAAGA AGCTGCATCGGTCCTCACGTGTGACATTTCACCGGACGACAAATTCGTTGTGACTGGGTCTGGTGATAAACGC GCCAACCTGTACGAGGTGATTTTT--------------GCCTCCTCCTCTTCTAATAACTCGTCTNNN	--- --- --- -------- MSALSGTPPSSLQTGLHGLNMSHGGSGP- SYS-A---AT---------NPSNSLP--GVLPS---- ISPPSNP------------------AMSAALLN--G------- ---------LMSAAGA-------------NPLGPSGF----- ----------MVPTTGATSLNFGR-----ESEK- NSADDKQRAQLKSESSSSYPT---------- TSSAPAVGSGGSGG-GRSSQVSEFSAG- TRWRGSSKS-S-SGPGAYSFVVLPGGQ- LRPCAL--- ASAPGAATASGLPKRLQHLASLPHGDVVC AVTIGACPVGR--- ASAAHFAYTGARGSVKLWDLSAISATNLS G---M- STRDIAPLMNFDCLCPDSYVRSIKLFPDAS HLIIGGESNALTLWDLNGPGRR- KAELTFEAPACYALALSPNGKLCYSCCSD GYVSVWDVHNQSVVHQFHGHTDGTSCVE LCPDGNRLWTGGLDHKVCCWDIRAP------- ------TS---R------ SLAFMEFKSQVFSLGLTTNYGSPHGHGFA GRRQ--TASVSPRSGDGDSASSTGL----- MRDSS--SPQS- VSGGGGVGSRGGSGVGSWLAVGLESSEV EVVAIGPDGFAVSS-------------- SNSAAGSRDESSPAL-------SPMSYSQPP--- --TPQP-------- QHFRLTHHESCVLALRFAHHGDWFLTTGK DHQVNAWRTPYGACLLETKEAASVLTCDI SPDDKFVVTGSGDKRANLYEVIF----- ASSSSNNSS?

	Crassostrea gigas	gi|405962653|gb|EKC28310.1| Transducin-like enhancer protein 4 [Crassostrea gigas]	ATGTATCCA------------AACAGGCACCCGGCTCCACATCAGCCAAGCCAGCCTTTTAAGTTCACTGTTGCTGAATCGTGTGATAGAATCAAAGAAGAATTCAGTTTTCTTCAAGCACAGTACCACACTTTAAAAATGGAGTGTGAAAAATTAGCCCAAGAGAAAACAGAAATGCAAAGACACTATGTCATGTACTATGAGATGTCCTACGGTCTCAATGTTGAAATGCACAAACAG---CATCAGCAGCAGGTTGCAGCAGCAGTGGAGCGTGCAAAACAAGTCACCATGACAGAGCTCAACCAGATCATTGGG---ATGCAGGCAGGGCAGCACTTA--GCGGGTCATGGACACGGTGCC---------CCACCC---------------------------TTCCCCATGCCACCCCACCCC---CCAGGACTACAACCCCCA------------CTACCTGTCTCCAGCGCTGCT---AGTCTTCTG---GCTCTCCAGGGG------GGCGCCCTAGGGCCCTCCCATCTCCTCCCCAAGGAAGATAAAGATGACAAACACAACAGATCTTCAGCTTCACCC---AGC---------------------GAAAGGGAAAGAGAGAGGGAAAAGTTTGAACACAAATTCAGTGAAAAATAC---GCCAGTCGTAGCAGAACTCCCGAAGTCAATGAA---------TCA-----------AAGAAGAGAAGAGTGGAGGAGAAG------GAATTTAACCACCGAAATCTATTCCAAGGGGATCCTCAT------------GCCAACCCCGCCCACGCAGCTCATATACGACCAGCCCTGGGG---AATTCTGCA---GGAGGAAAACCTGCCTACTCGTTCCATGTCAGCGTAGACGGCCAG---ATGCAGCCCGTGCCCTTT---------CCT------CCGGATGCTCTGATTGGACCAGGAATTCCTCGCCATGCGCGGCAGATAAACACCCTGAACCACGGGGAAGTGGTGTGCGCCGTCACCATC-------------------------------------CCCACAAGACACGTCTACACTGGAGGAAAGGGGTGTGTTAAAGTGTGGGACATCAGCCAGCCC---------------GGAAAC---------AAG---AGT---------------CCCATATCTCAGCTAGATTGTCTGCAAAGAGACAACTACATCCGATCAATCAAGTTACTTCAGGATGGGCGGACCCTCATAGTGGGAGGAGAAGCCAGCACTCTATCAATATGGGATCTTGCTGCGCCCACTCCGCGCATTAAAGCAGAACTGACTTCAAGTGCTCCTGCGTGTTACGCTCTGGCGATCAGCCCAGACTCCAAAGTATGCTTCAGCTGTTGCAGTGACGGAAACATCGCCGTCTGGGACCTTCACAATCAGACTCTGGTACGACAATTCCAGGGGCACACAGATGGCGCCAGTTGTATTGACATCTCGCCTGATGGCACCAAATTGTGGACCGGTGGTCTGGACAACACGGTCCGCTCTTGGGATCTCAGA---GAGGGAAGA--------------------------CAGTTGAAGCAACACGACTTCAGTTCTCAGATCTTCTCCCTTGGCTACTGTCCAACC--GGCGAA---------------------TGGCTAGCTGTAGGAATGGAGAGCAGTAATGTGGAGGTA---CTCCAC---------------TGCAGTAAGCCG-------------------------GACAAGTACCAGCTCCACCTCCACGAGAGCTGTGTCCTCTCCCTCAAGTTCGCCTACTGCGGCAAGTGGTTCGTCAGCACTGGCAAGGACAACCTCCTGAATGCCTGGAGGACTCCATACGGAGCCAGCATATTCCAGTCTAAAGAATCATCCTCAGTCTTGAGCTGCGACATTTCAACAGATGACAAATACATTGTCACTGGATCCGGCGACAAGAAAGCCACATTATACGAAGTCATCTTT---------------NNN----------------------------	MYP----NRHPAPHQPSQPFKFTVAESCDRIKEEFSFLQAQYHTLKMECEKLAQEKTEMQRHYVMYYEMSYGLNVEMHKQ---------------------------------HQQQVAAAVERAKQVTMTELNQIIG---MQAGQHL---------------------AGHGHGA---PP---------FPMPPHP-PGLQPP----LPVSSAA------------------SLL-ALQG--GALGPSHLLPKEDKDDKHNRSSASP--------------S-------EREREREKFEHKFSEKY-ASRSRTPEVNE---S----KKRRVEEK--EFNHRNLFQGDPH----ANPAHAAHIRPALG----------------------------NSA-GGKPAYSFHVSVDGQ-MQPVPF---P--PDALIGPGIPRHARQINTLNHGEVVCAVTI------------PTRHVYTGGKGCVKVWDISQP-----GN---K-S-----PISQLDCLQRDNYIRSIKLLQDGRTLIVGGEASTLSIWDLAAPTPRIKAELTSSAPACYALAISPDSKVCFSCCSDGNIAVWDLHNQTLVRQFQGHTDGASCIDISPDGTKLWTGGLDNTVRSWDLR-----------------EGR------QLKQHDFSSQIFSLGYCPT--GE-------WLAVGMESSNVEV---LH-----CSKP--------DKYQLHLHESCVLSLKFAYCGKWFVSTGKDNLLNAWRTPYGASIFQSKESSSVLSCDISTDDKYIVTGSGDKKATLYEVIF-----?---------

Echinococcus granulosus	gi|674567992|emb|CDS17106.1| groucho protein [Echinococcus granulosus]	ATGTATCCT------------AGCAGGCCTCCTGTGCCCTCGGGGCCAGGCCAGTCCTATAAATTTACAGTTGTCGAGACCTGTGAACGCATCAAGGATGAATTTAACTTCGTTCAGCAGCAGTGTCATCAACTTCAAACTGAGCGAGAGAAGCTCCTGAGTGAGCGCGTTGATATGCAGCGCATATGCGTTGTCTACTATGAAATGGCCAATGGATTAAACCTAGAAATGCATCGGCAGATGGAAATCTCAAAACGTCTCAACGCAATTCTCAATCAAGTCATTCCTTTTCTCGCCACTGAG---GCTCAAATGGAAGACTCAAAGTTAGCCAAATTTGCGAATGGTGGAAATATGGGACTGTCCAATGATCAGCTAGAAATTTATAAACAAATCCAAGCCCACCAGATTTCGTCAGCAGGTAGTCCCACGTCCTCCCTTCCCGGAAGTTCA---TCCGCCAATCAGGCCGGTCCG---CACGGGTCA---AACGTCGGCAGTGCTCCT---------------------------GCTACGGGTGCAGCATTTCCT------GGGTCTACCCCAGGT------------CTACTTCCGTCCACTGCACCC--GGACTGCAAAATGCATTCATCAAT------GGG---CTAATGTCTGCTGCTGGTGGT---GCC---ATGATGCCCCCTCCTGGGACTGCTGGTCTTCCCCTTGGTAGCGGTGGTGTT------CGTGACATTAAGGACTATTCAAGTGAAGATAAACAGCGAGTGCAGCTGCCTAGTTCCATGGGAAGCAACTATCCGACA------------------------------ACGAACAGTGCCCCCGCAGTCGGAACT------------------------CCGCACAATAGATCAACTTTCTCCGAAAATCGGTCAAAGTATGGTGGAAAAGGCCACTCTTCTTCCTGTTCTGGTCCGGGGACCTATTCGTACGTGGTACTTCCAGGCGGTGGAGGCACACGCCCTTGCGCGCTA---------GCGTCGGCACCCGAAGCAACATCGAATCCCAACCTACCGCGGCGTCTACAGCATTTAGCCAGTTTACCGCACGGTGAAGTGGTCTGTGCAGTGACTATCGCTCCCGCGCCCTCTGGTCGGGGTTCCTCACCCACCCCA---CATTTCGCCTACACTGGTGGCCGCGGTTGTGTTAAGCTCTGGGACCTTGACGCCATT---------------GCTGGC---------TCCTCTTCCTCTAGAGATGTCGTCGCCCTCGCCTCCTTTGACTGTCTTCGTCCCGAAAGCTATGTTCGCTCTATCAAGCTTTTCCCGGATTGCTCCACCCTTTTGATTGGAGGCGAATCGAGTTGCCTAACGATTTGGGATTTGAACGGCCCCGGTCGACGG---AAAGCGGACCTCACTTTTGACGCGCCCGCCTGTTATGCACTCGCACTCTCCCCCGACTGCAAGCTTTGTTATAGTTGCTGTTCCGATGGTCAAGTTGCGATATGGGATATCCACAACCAGAGCGTCGTCCAACAGTTCCACGCGCACGCTGATGGCGCTTCCTGCATCGAATTAGTGGGACAGGGCACTCGACTTTGGACCGGTGGCCTTGACAACAAGGTTCGCTGCTGGGACATCCGCGGCACT--------------------------------------TCATCC---------AGT-----------------AGGCTCCACCACGTCGAGTTCAAGTCGCAGGTCTTCTCATTGGGTCTCTCCCCCGTCTACGGCATGCAC------------------GGTCGTCGGCCCAGCCTAGCAGCAGCGGCGTCTCCTATGGGAGCAGAA--GACGCACCT------TCCTCCGAATCGCTCTTCTATGGC---------------TCGCAA--------------------TGGCTAGCTGTCGGCCTCGAGTCCTCAGAGGTCGAAGTCGTTGCCATTGGTCCCGACGGTCCGCCATCCTCCGCT--------------------------------CCCCCGTCCACCATCAGTAGTAGTGGTACTCCT--------------------CCCAACTCCACTCTTCTCAAAACACCTGCCAGTCCA---CAAACTACAAGTGCAGCTGTTCCCAGTCCTCAGGCTCCATCCTAC------------GACCAACACTTCCGCCTGACACGCCACGAAAGCTGCGTCCTCGCTCTCCGCTTCGCGCACCATGCCGACTGGTTCATCACCACCGGCAAGGACCACCAGGTGAACGCCTGGAGGACTCCCTATGGAGCTTGTCTCCTCGAGACCAAAGAAGCCGCCTCTGTGCTGACTTGTGACATCTCGCCGGACGACAAGTTTGTGGTGACGGGTTCGGGTGAC	MYP----SRPPVPSGPGQSYKFTVVETCERIKDEFNFVQQQCHQLQTEREKLLSERVDMQRICVVYYEMANGLNLEMHRQMEISKRLNAILNQVIPFLATE---AQMEDSKLAKFANGGNMGLSNDQLEIYKQIQAHQISSAGSPTSSLPGSS-SANQAGP-HGS-NVGSAP---------ATGAAFP--GSTPG----LLPSTAP------------------GLQNAFIN--G-----------------LMSAAGG-------------A---------------------MMPPPGTAGLPLGSGGV--RDIKDYSSEDKQRVQLPSSMGSNYPT----------TNSAPAVGT--------PHNRSTFSENRSKYGGKGHSSSCSGPGTYSYVVLPGGGGTRPCAL---ASAPEATSNPNLPRRLQHLASLPHGEVVCAVTIAPAPSGRGSSPTP-HFAYTGGRGCVKLWDLDAI-----AG---SSSSRDVVALASFDCLRPESYVRSIKLFPDCSTLLIGGESSCLTIWDLNGPGRR-KADLTFDAPACYALALSPDCKLCYSCCSDGQVAIWDIHNQSVVQQFHAHADGASCIELVGQGTRLWTGGLDNKVRCWDIRGT-------------SS---S------RLHHVEFKSQVFSLGLSPVYGMH------GRRPSLAAAASPMGAE----------------DAP--SSESLFYG-----SQ-------WLAVGLESSEVEVVAIGPDGPPSSA-----------PPSTISSSGTP-------PNSTLLKTPASP-QTTSAAVPSPQAPSY----DQHFRLTRHESCVLALRFAHHADWFITTGKDHQVNAWRTPYGACLLETKEAASVLTCDISPDDKFVVTGSGDKRANLYEIIYGAGSVASSASNVSSH

Echinococcus multilocularis	gi|674571713| emb|CDS42131.1| groucho protein [Echinococcus multilocularis]	ATGTATCCT------------ AGCAGGCCTCCTGTGCCCTCGGGGCCAGGCCAGTCCTATAAATTTACAGTTGTCGAGACCTGTGAACGCATC AAGGATGAATTTAACTTCGTCCAGCAGCAGTGTCATCAACTTCAAACTGAGCGAGAGAAGCTCCTGAGTGAGC GCGTTGATATGCAGCGCATATGTGTTGTCTACTATGAAATGGCCAATGGATTAAACCTAGAAATGCATCGGCA GATGGAAATCTCAAAACGTCTCAACGCAATTCTCAATCAAGTTATTCCTTTTCTCGCCACTGAG--------------------- --- GCTCAAATGGAAGACTCAAAGTTAGCCAAATTTGCGAATGGTGGAAATATGGGACTGTCCAATGATCAGCTAG AAATTTACAAACAAATCCAAGCCCACCAGATGTCGTCAGCAGGTAGTCCCACGTCCTCCCTTCCCGGAAGTTC A---TCCGCCAATCAGGCCGGTCCG---CACGGGTCA---AACGTTGGCAGTGCTCCT-------------------------- GCTACGGGTGCAGCATTTCCG------GGGTCTACCCCAGGT------------CTACTTCCGTCCACCGCACCC------------- ---------------------------------------GGACTGCAAAATGCATTCATCAAT------GGA-- TTAATGTCTGCTGCTGGTGGT---GCC-- ------ATGATGCCCCCTCCTGGGACTGCTGGTCTTCCCCTTGGTAGCGGTGGTGTT------ CGTGACATTAAGGACTATTCAAGTGAAGATAAACAGCGAGTGCAGCTGCCTAGTTCCATGGGAAGCAACTATC CGACA-----------------------------ACGAACAGTGCCCCCGCGGTCGGAACT----------------------- CCGCACAATAGATCAACTTTCTCCGAAAATCGGTCAAAATATGGTGGAAAAGGCCACTCTTCTTCCTGTTCTGG TCCGGGGACCTATTCGTACGTGGTACTTCCAGGCGGTGGAGGCACACGCCCTTGCGCGCTA--------- GCGTCGGCACCCGAAGCAACATCGAATCCCAACCTACCGCGGCGTCTACAGCATTTAGCCAGTTTACCGCAC GGTGAAGTGGTCTGTGCAGTGACTATCGCTCCCGCGCCCTCTGGTCGGGGTTCCTCACCCACCCCA--- CATTTCGCCTACACTGGTGGCCGCGGTTGTGTTAAGCTCTGGGACCTTGACGCCATT---------------GCTGGC----- ---- TCCTCTTCCTCTAGAGATGTCGTCGCCCTCGCCTCCTTTGACTGTCTTCGTCCCGAAAGCTATGTTCGCTCTAT CAAGCTTTTCCCAGATTGCTCCACCCTTTTGATTGGAGGCGAATCGAGTTGCCTAACGATTTGGGATTTGAAC GGCCCCGGTCGACGG--- AAAGCGGACCTCACTTTTGACGCGCCCGCCTGTTATGCACTCGCACTCTCCCCCGACTGCAAGCTTTGCTATA GTTGCTGTTCCGATGGTCAAGTTGCGATATGGGATATCCACAACCAGAGCGTCGTCCAACAGTTCCACGCGC ACGCTGATGGCGCTTCCTGCATCGAATTAGTGGGACAGGGCACTCGACTTTGGACCGGTGGCCTTGACAACA AGGTTCGCTGCTGGGACATCCGCGGCACT--------------------------------------TCATCC---------AGT----------------- AGACTCCACCACGTCGAGTTCAAGTCGCAGGTCTTCTCATTGGGTCTCTCCCCCGTCTACGGCATGCAC-------- ----------GGTCGTCGGCCCAGCCTAGCAGCAGCGGCGTCTCCTATGGGAACAGAA--- --------GACGCACCT------TCCTCCGAATCGCTCTTCTATGGC---------------TCGCAA--------------------- TGGCTAGCTGTCGGCCTCGAGTCCTCAGAGGTCGAAGTCGTTGCCATTGGTCCCGACGGTCCGCCATCCTCC GCT--------------------------------CCCCCGTCCACCATCAGTAGTAGTGGTACTCCT--------------------- CCCAACTCCACTCTTCTTAAAACACCTGCCAGTCCA--- CAAACTACAAGTGCAGCTGTTCCCAGTCCTCAGGCTCCATCCTAC------------ GATCAACACTTCCGCCTGACACGCCACGAAAGCTGCGTTCTCGCTCTCCGCTTCGCGCACCATGCCGACTGG TTCATCACCACCGGCAAAGACCACCAGGTGAACGCCTGGAGGACTCCCTATGGAGCCTGTCTCCTCGAGACC AAAGAAGCCGCCTCTGTGCTGACTTGTGACATCTCGCCGGACGACAAGTTTGTGGTGACGGGTTCGGGTGAC	MYP---- SRPPVPSGPGQSYKFTVVETCERIKDEFN FVQQQCHQLQTEREKLLSERVDMQRICVV YYEMANGLNLEMHRQMEISKRLNAILNQVI PFLATE--- ------- AQMEDSKLAKFANGGNMGLSNDQLEIYKQ IQAHQMSSAGSPTSSLPGSS-SANQAGP- HGS-NVGSAP---------ATGAAFP--GSTPG---- LLPSTAP------------------GLQNAFIN--G-------- ---------LMSAAGG-------------A------------------ ---MMPPPGTAGLPLGSGGV-- RDIKDYSSEDKQRVQLPSSMGSNYPT------ ----TNSAPAVGT-------- PHNRSTFSENRSKYGGKGHSSSCSGPGT YSYVVLPGGGGTRPCAL--- ASAPEATSNPNLPRRLQHLASLPHGEVVC AVTIAPAPSGRGSSPTP- HFAYTGGRGCVKLWDLDAI-----AG--- SSSSRDVVALASFDCLRPESYVRSIKLFPD CSTLLIGGESSCLTIWDLNGPGRR- KADLTFDAPACYALALSPDCKLCYSCCSD GQVAIWDIHNQSVVQQFHAHADGASCIEL VGQGTRLWTGGLDNKVRCWDIRGT--------- ----SS---S------ RLHHVEFKSQVFSLGLSPVYGMH------ GRRPSLAAAASPMGTE----------------DAP-- SSESLFYG-----SQ------- WLAVGLESSEVEVVAIGPDGPPSSA--------- --PPSTISSSGTP-------PNSTLLKTPASP- QTTSAAVPSPQAPSY---- DQHFRLTRHESCVLALRFAHHADWFITTG KDHQVNAWRTPYGACLLETKEAASVLTCD ISPDDKFVVTGSGDKRANLYEIIYGAGSVA SSASNVSSH

	Helobdella robusta	gi|675883172| ref|XP_00902 7088.1| hypothetical protein HELRODRAF T_87650, partial [Helobdella robusta]	-- -- -- -- -- -- -- -- --- TACGCGTTTCAAGTCGTTGGGCCCAAAACT---TTACAACCGATTTGGTTC---------CAC------ CCGGACATGTTAGGGGGTCCGGGTATTCCCCAAGAGATAAAACCCCTCGGTATTTTGCCACAGGGTGAGGTG GTCTGCGCCATTGCACTA----------------------------------- CCGGTTCAAAATGTTTTCACTGGTGGCAAGGGTTGTGTCAAAGTTTGGGACATAAATTCCGTC--------------- GGCAAG---------TCA---TCT--------------- CACATCCACCAACTCAACTGTCTGTCTTCTGATAGCTACATAAGATCAGTGAAGTTACTGAACGACGGCGTTAC ACTAATCGTAGGTGGGGAGGCCAATACACTAACTGTCTGGGATCTGGCTGCTCCAACCCCTGTTATAAAAGGA GAGCTGACATCCGGAGCTCAGGCATGTTACGCAATGGCTGTTTCTCACGACTCGAGACTCTGCTACAGCTGC TACAGTGACGGCAACGTGGCCATCTGGGATGTACATAATAATGAAATTGTCAAACAATTCCAAGGGCACAGCG ATGGAGCCAGCTGCATAGACATCTCCCCAGACGGCACCAACATTTGGACCGGCGGCCTAGACAACACGGTTA AACAGTGGGACATAAGAGGAGTG-------------------------------------- GGAGGAGGGGAAGAGAGGGCTGGCTCGGGCGACAACCCCCTACAGAAATACGAGTACATGTCTCAGGTGTT CTCCCTCGGTGTTTCTCCACTC-- ---GGTGAA--------------------- TGGATAGCGATAGGATTGGAATCGGCTGAAATTCGTCTG-- ---ACGAAC--------------ACAATAACCCAG------ ------------------ GACTCCTACCAGGTCATCCTGCACACCAGTTGCATTCTCACTCTCAAGTACTCGCCAGATGGGCTGTGGTTTA TATCGGCCGGCAAAGATAATGGTCTGTTTGGCTGGAAGGCTCCCTACGGCATTAACCTCTTCCAGAATAAAGA GCAAACTTCAATACTCTGCTGCGACATATCAAACGACAACAAGTACATCGTAACTGGCTCTGGGGATAAGAAA GCCACAGTGTATGAAGTCATCTAC--	--- --- --- --- --- --- --- ----------------------------------YAFQVVGPKT- LQPIWF---H-- PDMLGGPGIPQEIKPLGILPQGEVVCAIAL-- ----------PVQNVFTGGKGCVKVWDINSV----- GK---S-S----- HIHQLNCLSSDSYIRSVKLLNDGVTLIVGGE ANTLTVWDLAAPTPVIKGELTSGAQACYA MAVSHDSRLCYSCYSDGNVAIWDVHNNEI VKQFQGHSDGASCIDISPDGTNIWTGGLD NTVKQWDIRGV------------- GGGEERAGSGDNPLQKYEYMSQVFSLGV SPL-- ------GE-------WIAIGLESAEIRL------------------- -----------------------------------TN-----TITQ------- - DSYQVILHTSCILTLKYSPDGLWFISAGKDN GLFGWKAPYGINLFQNKEQTSILCCDISND NKYIVTGSGDKKATVYEVIY---------------

	Hymenolepis microstoma	gi|674593226| emb|CDS2798 5.1| groucho protein [Hymenolepis microstoma]	ATGTATCCT------------ AATCGGCCTCCGGTGCCTTCTGGGCCCGTTCAACCTTATAAGTTTACGGTTCTTGAAACTTGTGATCGTATTAA AGAAGAATTCAATTATGTTCAACAGCAATGCCATCAGCTTCAAGCTGAAAGGGAAAAACTAATGAGCGAGCGT GTTGATATGCAACGGATATGTGTAGTATATTACGAAATGGCAAATGGATTGAACTTGGAAATGCATCGTCAAAT GGAAATTTCAAAACGTCTTAATGCCATTCTCAATCAAGTCATTCCTTTTCTAGCTGCTGAG--------------------------- -------- CATCAGTCGCAAGTCGCGTCTGCAATTGACAGAGCCAAACAGGTTACCATGCAGGAACTGAATTCAGTTCTAA CA--- GCTCAGATGGAAGATACGAAACTTTCAAAATTTGCCAATGGTGGAAATATGGGAATTTCAAATGATCAATTAGA GATTTATAAGCAAATGCAAGCCCATCAG---------GGTGGTAGTCCCACAGCAGGCTTGCCTGGAGCATCG--- GCAGGAAATCAGAGTAGCCCC---CATGGAAAT---AAT---------CCCACT-------------------------- GCATCATCGTCAGCTTTCCCT------GGC------------------------TTACCCGCTAGTATGACAGCG---------------------------- -------------------------GGACTTCAAAGCGCATTTTTAAAT------GGA--- TTAATGTCTGCTGCTGGTGGT--GGT--- -----ATGATGCCTCCTCCTGGTTCAGCCGGTCTCCCGCTTGGGAAC--------------- CGAGATATCAAGGATTATTCAAATGATGACAAGCAGCGT--- CAAATGTCTAACTCAATGGGGAGCAATTATCCCACC------------------------------ ACAAACAGCGCTCCCGCTTTTGGTTCC------------------------ CCGCAAAATCGATCTACTTTCCCTGAAAATCGGTCCAAGTATAAGGGACAG------------CCG--- TCTGGCCCTGGCGCATACTCTTCTGTGGTTTTACCAAACGGAGCAGGAACTCGTTCTTGTGTTCTT--------- GCAGCTGCGCCAGAAGCCATTTCAGACCCCAGTTTACCACGTCGCCTGCAGCATGTTGCTAGTCTACCTCATG GAGAGGTAGTTTGTGCTGTTACTATCGCCCCAGCACCTGCAGGCCAT---------AGTTCACCA--- CACTTTGCTTACACCGGTGGACGCGGATGTGTCAAGCTGTGGGATCTTGATACTATC---------------ACTTCC------ ---TCC--- ACAGCTAAAGATGTCATTGCACTGGCGTCCTTTGATTGTCTCCGTCCGGAGAGCTATGTGCGATCTATCAAAC TTTTCCCGGACTGTTCCACACTTCTCATTGGAGGTGAATCGAGCTGCCTCTCAATCTGGGATTTGAATGGATCT GGAAGGAGA--- AAGAAGGATCTGACTTTTGACGCCCCTGCTTGCTATGCGCTAGCCCTTTCGCCTGACTGCAGATTCTGTTATA GTTGCTGTTCTGATGGACAGGTGGCAATTTGGGATATTCACAATCAGAATATTGTTCATCAGTTCCATGCACAT GAAGACGGTGCTTCTTGCATCCAGTTGGCCGCACAAGGCACTCGACTCTGGACGGGTGGACTTGATAATAAA GTTCGCTGCTGGGATATTCGTGGAACG--------------------------------------TCCTCG---------AGC------------------ GAACTGCATTGTGTAGAATTTAAGTCTCAAGTCTTCTCTCTTGGATTATCTCCGATCTTCGGGTCACAC----------- -------GGTCGGCGGGCCAGTATTGCG---GCCATATCTCCTACTAGGGGTGAT-- -GAAGGTTCC------TCCATAGATGCTGCGTTCCAAGGC---------------ACTCAG--------------------- TGGCTTGCAGTCGGTCTGGAGTCTTCTGAAGTTGAAGTTGTGGCCATTGGTCCTGATGGCCCAGCTTCACCTA CG--------------------------------AATTCCGTAAGCAACAACGGCGCTGGAAACTCG---------------------TCC------------ TTAACCGGAAGTCCATCACAC--------------------------AGCCCCGTTGTTAACACTTAT------------ GATAACCATTTCCGTCTCACTCGACACGACAGTTGTGTTCTAGCTCTGCGCTTCGCCCATCAAGCGGACTGGT	MYP---- NRPPVPSGPVQPYKFTVLETCDRIKEEFNY VQQQCHQLQAEREKLMSERVDMQRICVV YYEMANGLNLEMHRQMEISKRLNAILNQVI PFLAAE------------ HQSQVASAIDRAKQVTMQELNSVLT--------- --------- AQMEDTKLSKFANGGNMGISNDQLEIYKQ MQAHQ---GGSPTAGLPGAS-AGNQSSP- HGN-N---PT---------ASSSAFP--G-------- LPASMTA------------------GLQSAFLN--G------- ----------LMSAAGG--------------G------------------ ----MMPPPGSAGLPLGN----- RDIKDYSNDDKQR-QMSNSMGSNYPT------ ----TNSAPAFGS-------- PQNRSTFPENRSKYKGQ----P- SGPGAYSSVVLPNGAGTRSCVL--- AAAPEAISDPSLPRRLQHVASLPHGEVVCA VTIAPAPAGH---SSP- HFAYTGGRGCVKLWDLDTI-----TS---S- TAKDVIALASFDCLRPESYVRSIKLFPDCST LLIGGESSCLSIWDLNGSGRR- KKDLTFDAPACYALALSPDCRFCYSCCSD GQVAIWDIHNQNIVHQFHAHEDGASCIQLA AQGTRLWTGGLDNKVRCWDIRGT----------- --SS---S------ ELHCVEFKSQVFSLGLSPIFGSH------ GRRASIA-AISPTRGD----------------EGS-- SIDAAFQG-----TQ------- WLAVGLESSEVEVVAIGPDGPASPT--------- --NSVSNNGAGNS-------S----LTGSPSH------- --SPVVNTY---- DNHFRLTRHDSCVLALRFAHQADWFLTTG KDHQVNAWKTPYGANLLDTREAASVLTCD ISPDDKFVVTGSGNRIANLYEIIYGEGSVVS TG---SSH

	Lottia gigantea	gi|556105939| gb|ESO94591. 1| hypothetical protein LOTGIDRAFT _206444 [Lottia gigantea]	ATGTACCCT------------ AACCGTCACCCTGGACCACACCAACCAGGCCAGCCTATAAAATTTACAGTATCAGAATCATGTGATCGTATTAA AGAAGAATTCAGTTTTCTCCAAGCCCAGTATCACAATTTAAAAATGGAGTGTGAAAAACTGGCACAGGAGAAAA CTGAAATGCAGAGACACTACGTCATGTATTATGAAATGTCATATGGATTAAATGTTGAAATGCATAAACAGACT GAAATAGCCAAGAGATTAAATGCCATATGTGCTCAAGTTATCCCATTTTTATCCCAAGAG------------------------------ ------ CATCAGCAACAAGTGGCAGCAGCAGTGGAAAGAGCAAAACAAGTCACGATGCAGGAATTAAATGCTGTTATTG GT--- CAGATGCAAGCACAGCATTTA--CCA---CATGCACAT---GCT--------- CCACCC---------------------------ATACCTATGACACCTCATCCT---GCAGGATTACCTCCCCCT------------ GGTGGTCTCTCGTCAGTACCTGGTGGATCAGGATTATTATCATTGCCTCATGGATCATTTCCAACACATCCTCA CAGTCTCAGT---GCATTGAAAGAT------GGA-- ATTAGAAGATCAGCATCTCCA---GCT---------------------GAAAGAGAA--------------------- ---------------AAATTC---CGTCCTCGAAGTAGATCACCGGATGTCAGAAAC--------- AGTGGTGAACCTCCCAAAAAACGCAGACAAGATGAAAAA------ GTACAAAGACACCCCAGCATGCCAGGAGGTGGGGGATAT------------ GATGGACCCTCCCAGCATGAACACATGAGACCCGCACTTGGC-- ---------------------AACATTCCA---GGGGGCAAACCTGCATATTCTTTTCATGTGAGTGGTGATGGACAA--- ATACAGCCAGTTCCTTTT---------CCA------ CCAGATGCTTTGATTGGTCCAGGAATTCCAAGACATGCTCGTCAAATTAATACTTTAAATCATGGGGAGGTTGT TTGTGCAGTGACAGTT------------------------------------- CCAACAAGACATGTATATACTGGAGGAAAAGGTTGTGTTAAAGTTTGGGATATTAGTCAACCT--------------- GGAAAT---------AAA---AGT--------------- CCTGTTTCTCAACTTGATTGTTTACAAAGAGACAACTATATCAGATCAATAAAATTATTACAAGATGGTCGAACA TTAATCGTTGGTGGAGAAGCCAGTACATTATCAATATGGGATTTAGCAGCTCCAACCCCCAGAATAAAAGCAG AATTAACATCGAGTGCTCCAGCTTGTTATGCTCTAGCTTTAAGTCCTGATAATAAAGTGTGTTTTAGTTGTTGTA GTGATGGTAATATAGCTGTATGGGATTTACACAATCAAACTTTAGTCAGACAATTTCAAGGGCATACAGATGGA GCTAGTTGTATTGATATATCACCTGATGGTAGTAAATTATGGACTGGAGGTTTAGATAATACAGTCAGATCCTG GGATTTAAGA---GAGGGGAGA------------------ CAACTCCAGCAACATGATTTTACTTCTCAGATATTTTCACTAGGATATTGTCCTACT-------------------------------------- -- GGTGAC---------------------TGGCTAGCTGTAGGTATGGAGAGCAGTAATGTGGAAGTT------------------------------------- ---TTACAT------------- --CACGCTAAACCT------------------------ GACAAATATCAACTCCATTTACATGAAAGTTGTGTTCTTTCTCTTAAATTTGCATATTGTGGTAAATGGTTTGTG AGTACAGGAAAAGATAATTTACTCAATGCCTGGAGAACTCCATATGGTGCCAGCATCTTTCAGTCTAAAGAATC ATCGTCAGTGTTAAGTTGTGATATATCAGGAGATGATAAATATATTGTTACTGGATCAGGAGATAAAAAAGCAA CATTATATGAAGTTATATTT---------------NNN---------------------------	MYP---- NRHPGPHQPGQPIKFTVSESCDRIKEEFSF LQAQYHNLKMECEKLAQEKTEMQRHYVM YYEMSYGLNVEMHKQTEIAKRLNAICAQVI PFLSQE------------ HQQQVAAAVERAKQVTMQELNAVIG-------- --------------------------------------QMQAQHL------ ---------------P-HAH-A---PP---------IPMTPHP- AGLPPP---- GGLSSVPGGSGLLSLPHGSFPTHPHSLS- ALKD--G-----------------IRRSASP--------------A- ------ERE------------KF-RPRSRSPDVRN--- SGEPPKKRRQDEK--VQRHPSMPGGGGY-- --DGPSQHEHMRPALG--------------------------- -NIP-GGKPAYSFHVSGDGQ-IQPVPF---P-- PDALIGPGIPRHARQINTLNHGEVVCAVTV- -----------PTRHVYTGGKGCVKVWDISQP----- GN---K-S----- PVSQLDCLQRDNYIRSIKLLQDGRTLIVGG EASTLSIWDLAAPTPRIKAELTSSAPACYAL ALSPDNKVCFSCCSDGNIAVWDLHNQTLV RQFQGHTDGASCIDISPDGSKLWTGGLDN TVRSWDLR------------------EGR------ QLQQHDFTSQIFSLGYCPT---------------------- --------------------------------------GD------- WLAVGMESSNVEV-------------------------------- -----------------------LH-----HAKP-------- DKYQLHLHESCVLSLKFAYCGKWFVSTGK DNLLNAWRTPYGASIFQSKESSSVLSCDIS GDDKYIVTGSGDKKATLYEVIF-----?---------

	*Mesocestoides corti**	MCOS_0000253001-mRNA-1	--------------------------------- GTGCCTTCTGGGCCAGGTCCGTCCTATAAGTTTACTGTTGTTGAGACCTGTGAGCGCATTAAAGATGAATTTAA CTACGTCCAGCAACAATGTCATCAACTTCAAACTGAGCGTGAGAAACTTCTCAGTGAACGCGTTGATATGCAA CGCATTTGTGTTGTTTATTATGAAATGGCTAATGGGCTCAACCTAGAGATGCACCGTCAAATGGAGATTGCAAA ACGCCTCAACGCTATTTTGACGCAAGTTATCCCCTTTCTTGCTCAAGAG----------------------------------- CACCAATCCCAGGTTGCATCAGCGATCGAGCGAGCGAAGCAGGTGACAATGCAAGAATTAAACTCTGTTCTC GCG--- GCACAGATGGAAGACTCAAAGTTATCCAAATTTGCCAATGGTGGGAGCATGGGGTTATCTAATGACCAGCTGG AAATTTATAAGCAAATTCAAGCTCATCAAATGTCGTCTACTGGAAGTCCAGCAACATCACTTCCTGGTGGTTCG- --TCTAGCAACCAAACGGGCTCT---CATGGATCT---AATATCGGCAACACCCCA-------------------------- GTATCAAGTGCTACATTTCCT------GGATCTAATCCAAAC------------ATGATCCCGTCCAATCCATCG----------------- ------------------------------------TCTTTGCCAGCAGCTTTTCTTAAT------GGA-- TTAATGTCGGCTGCAGGTGGT--GTT-- ------ATGATGCCTCCCCCTGGATCAACTGGTATTCCTCTTAGTGGTGGAACTGGA------ CGAGAAATTAAGGATTATTCATCTGATGATAAACAGCGAGTACAACTGGCCAATTCAATGGGTAGCAGCTATCC GACC------------------------------ACAAATAGTGCTCCTGCTGTCGGAACT------------------------ GCACTTAACCGACCTCCCCAAAGTGAAAGCCGTTCGAGGTACGGCGGAAAAAGCAGCCTGTCATCA--- TCAGGTCCGGGTGCTTACTCGTATGTCGTCTTACCGGGTGGTTCGGGCACGCGCCCCTGTGCTCTG--------- GCGTCAGCGCCAGAAGCTACCACGAATCCGAGTCTGCCTCGGCGTCTACAGCACCTAGCTAGTCTCCCGCAC GGTGAAGTCGTCTGTGCTGTGACCATCGGCCCGTCACCAACTGGACGA---------AGCACCCCA--- CACTTTGCTTACACCGGTGGACGGGGTTGTGTCAAACTTTGGGACTTAGACGTGATT---------------GCTGGC------ ---TCC--- TCTTCCAGAGATGTTGTTGCACTTGCTTCTTTCGATTGTCTTCGACCTGACAGCTATGTTCGATCAATTAAGCT CTTCCCTGATGCCTCGGGTCTTGTGATTGGTGGAGAATCAAGTGCTCTTACGATTTGGGATCTGAACGGTCCC GGTCGACGG--- AAAGCTGACCTCAATTTTGACGCTCCCGCTTGCTACGCTCTTGCTCTTTCCCCTGACTGCAAACTTTGTTACAG TTGCTGTTCTGACGGTCACGTGGCGGTGTGGGATATCCATAATCAAAGCGTTGTCCAACAATTTCATGCTCAT GGAGATGGAGCTTCTTGTATTGAATTGGCATCACAAGGGACTAGACTGTGGACTGGTGGGCTCGATAACAAA GTTCGCTGCTGGGACATTCGTGGTTCC--------------------------------------TCCTCT---------CAC----------------- CGACTTCACCACATTGAATTTAAGTCGCAAGTTTTTTCCCTGGGTCTTTCGCCAATTTGCGGCCTTCGA----------- -------GGACGCCGTCCTAGTTTTGCTGTGGCCCCATCTCCCGCTGCTTTAGAG-- ---GACACAGCCAACGGCTCCTCTGACGCCCTCTTCTACGGG---------------ACCCAG-------------------- TGGATAGCCGTCGGCCTCGAGTCCTCCGAGGTCGAAGTGGTTGCTATTGGTTCCGACGGCCCTCCTCCC------ ---CTTATAACTGCGGGCAACCCT---------------------CCT------------ TTAAAACCCTCTTCAAGCCCG---CAGGCCTCC--------------- AGTCCCCAGGGCTCGTCGTCGATAGCGGTCCAGGACCAACATTTCCGGTTGACGCATCATGAGAGCTGCGTC CTGGCCTTGAGGTTCGCGCATCATGCTGATTGGTTCATTACAACTGGAAAGGACCACCAAGTGAACGCCTGG AGAACGCCTTACGGAGCTTGCCTGCTTGAGACAAAAGAAGCGGCCTCAGTGCTGACTTGCGACATCTCGCCG	----------- VPSGPGPSYKFTVVETCERIKDEFNYVQQ QCHQLQTEREKLLSERVDMQRICVVYYEM ANGLNLEMHRQMEIAKRLNAILTQVIPFLA QE------------ HQSQVASAIERAKQVTMQELNSVLA--------- --------- AQMEDSKLSKFANGGSMGLSNDQLEIYKQ IQAHQMSSTGSPATSLPGGS-SSNQTGS- HGS-NIGNTP---------VSSATFP--GSNPN---- MIPSNPS------------------SLPAAFLN--G-------- ---------LMSAAGG--------------V------------------- ---MMPPPGSTGIPLSGGTG-- REIKDYSSDDKQRVQLANSMGSSYPT------ ----TNSAPAVGT-------- ALNRPPQSESRSRYGGKSSLSS- SGPGAYSYVVLPGGSGTRPCAL--- ASAPEATTNPSLPRRLQHLASLPHGEVVC AVTIGPSPTGR---STP- HFAYTGGRGCVKLWDLDVI-----AG---S- SSRDVVALASFDCLRPDSYVRSIKLFPDAS GLVIGGESSALTIWDLNGPGRR- KADLNFDAPACYALALSPDCKLCYSCCSD GHVAVWDIHNQSVVQQFHAHGDGASCIEL ASQGTRLWTGGLDNKVRCWDIRGS--------- ----SS---H------ RLHHIEFKSQVFSLGLSPICGLR------ GRRPSFAVAPSPAALE---------------- DTANGSSDALFYG-----TQ------- WIAVGLESSEVEVVAIGSDGPPP-------------- ---LITAGNP-------P----LKPSSSP-QAS----- SPQGSSSIAVQDQHFRLTHHESCVLALRF AHHADWFITTGKDHQVNAWRTPYGACLLE TKEAASVLTCDISPDDKFVVTGSGDKRANL YEVIYGAGSVASSASNVSSH

Opisthorchis viverrini	gi|684402996|ref|XP_009173655.1| hypothetical protein T265_09350 [Opisthorchis viverrini]	CTCTGGCCTGAGTATGTTTTCAAACAGGCCACCGGGCCGTCG---------------TACAAATTCACAGTTTTGGAGACATGTGATAGAATCAAAGAAGAATTTAATTATTTGCAGCAGCAAAATCATTCATTGCATATTGAACGTGAGAAGCTTGTTAGCGAGAGAACCGATATGCAACGGATTTGCGTGATGTACTATGAGATGGCAAACGGCCTTAATTTAGAGATGCACAGACAGATGGAAATTGCCAAGCGTCTCAGTGCAATCTTGACTCAAGTTGTGCCATTCCTTTCACAAGAAGTGAGTTCATGTTTGATCATTCAATACCGTCAGCAGCACCAAACACAAGTCCTTTCGGCCATCGAGAGAGCAAAGCAGGTGACAATGCAAGAATTGAACTCAGTACTAGCG--GCTCAAATCGAGGATGCTAAGTCTTCGAAATTCTTTAATGGCAGTAACTCAAATCTGGTCGCTGAGCAATTAGAAGCATACAAG---ATGTCTGCTTTATCGGGCACTCCTCCCTCAAGCCTCCAAACGGGATTACATGGTCTGAACATGTCTCATGGTGGCAGTGGCCCA---AGCTACAGC---GCC---------GCCACC--------------------------AATCCATCTAATTCCCTTCCA------GGTGTCCTCCCTTCT------------ATCTCCCCACCAAACAACCCG--GCTATGTCCGCAGCCCTTTTAAAT------GGA--CTAATGTCAGCCGCGGGAGCC--AATCCGCTTGGTCCTAGTGGCTTT---ATGGTACCAACGACGGGTGCGACAAGTCTGAATTTTGGACGG--------------GAATCAGAGAAA---AACTCTGCGGATGAGAAACAACGCGCTCAACTGAAGTCCGAAAGTAGTTCCTCGTACCCAACA-----------------------------ACCTCAAGTGCCCCCGCTGTCGGTTCAGGTGGGGGTGGTGGCGGTGGCCGGTCGTCGCAGGTATCTGAATTTTCGGCTGGT---ACTCGTTGGCGGGGTTCGAGCAAAAGC---TCC---TCCGGACCCGGTGCTTATTCTTTCGTAGTCCTTCCTGGTGGTCAA---CTCAGGCCTTGTGCGCTA---------GCGTCCGCTCCAGGTGCTGCAACTGCCTCAGGTTTACCCAAGCGGCTCCAGCACCTCGCTAGTTTGCCACATGGAGATGTGGTTTGTGCAGTCACGATCGGAGCATGCCCAGTCGGTCGT---------GCCTCCGCTGCTCACTTCGCTTACACCGGTGCTCGGGGTAGCGTCAAGCTATGGGACTTGTCAGCGATCAGTGCAACAAACTTGTCCGGG---------ATG---TCAACGAGAGATATCGCTCCCCTAATGAATTTCGACTGCCTTTGCCCAGACAGTTATGTTCGGTCGATCAAGTTGTTCCCGGATGCGAGTCACCTCATCATAGGCGGCGAATCAAACGCTCTAACCTTATGGGATTTGAACGGCCCTGGGCGTCGT---AAAGCCGAATTAACGTTTGAAGCCCCTGCGTGCTACGCGTTGGCCTTATCACCTAATGGAAAACTTTGTTACAGCTGTTGCTCCGATGGTTATGTTTCTGTGTGGGATGTGCACAATCAAAGTGTCGTTCATCAGTTCCACGGTCATACAGACGGAACCTCTTGCGTTGAACTGTGTCCAGATGGAAACCGTCTATGGACCGGTGGTCTGGATCACAAAGTCTGCTGTTGGGATATTCGTGCTCCC--------------------------------------ACGTCC---------CGC-----------------TCTCTCGCCTTCATGGAATTCAAATCCCAAGTTTTCTCCTTGGGGCTGACCACTAATTACGGTTCTCCACATGGACATGGTTTTGCTGGTCGACGGCAG------ACTGCTTCGGTTTCCCCTCGATCAGGCGATGGGGACTCAGCTTCTAGCACCGGCTTG--------------ATGCGTGACTCATCG------AGTCCTCAGTCT---GTTAGTGGTGGCGGTGGTGTCGGAAGCAGGGGTGGTTCGGGTGTTGGCAGTTGGTTGGCTGTGGGTCTCGAATCATCCGAGGTCGAAGTAGTTGCAATCGGTCCTGATGGGTTTGCCGTCAGTAGT------------------------------------	LWPEYVFKQATGPS-----YKFTVLETCDRIKEEFNYLQQQNHSLHIEREKLVSERTDMQRICVMYYEMANGLNLEMHRQMEIAKRLSAILTQVVPFLSQEVSSCLIIQYRQQHQTQVLSAIERAKQVTMQELNSVLA------------------AQIEDAKSSKFFNGSNSNLVAEQLEAYK-MSALSGTPPSSLQTGLHGLNMSHGGSGP-SYS-A---AT---------NPSNSLP--GVLPS----ISPPNNP------------------AMSAALLN--G------------------LMSAAGA--------------NPLGPSGF---------------MVPTTGATSLNFGR-----ESEK-NSADEKQRAQLKSESSSSYPT----------TSSAPAVGSGGGGGGGRSSQVSEFSAG-TRWRGSSKS-S-SGPGAYSFVVLPGGQ-LRPCAL---ASAPGAATASGLPKRLQHLASLPHGDVVCAVTIGACPVGR---ASAAHFAYTGARGSVKLWDLSAISATNLSG---M-STRDIAPLMNFDCLCPDSYVRSIKLFPDASHLIIGGESNALTLWDLNGPGRR-KAELTFEAPACYALALSPNGKLCYSCCSDGYVSVWDVHNQSVVHQFHGHTDGTSCVELCPDGNRLWTGGLDHKVCCWDIRAP-------------TS---R------SLAFMEFKSQVFSLGLTTNYGSPHGHGFAGRRQ--TASVSPRSGDGDSASSTGL-----MRDSS--SPQS-VSGGGGVGSRGGSGVGSWLAVGLESSEVEVVAIGPDGFAVSS--------------SNSAAGSRDESSPAL-------SPMSYSQPP-----TPQP--------QHFRLTHHESCVLALRFAHHGDWFLTTGKDHQVNAWRTPYGACLLETKEAASVLTCDISPDDKFVVTGSGDKRANLYEVIF-----ASSSSNNSS-

Schistosoma haematobium	gi|844852349| ref|XP_012795902.1| Transducin-like enhancer protein 3 [Schistosoma haematobium]	-- -- ATGGAAATAGCTAAAAGATTAAGCGCTATCTTGACTCAAGTATTACCATTCCTATCACAAGAG------------------------- -- --CAGCTGTCTGTCATTGGGGGTGCTAACCCAGGCGTT---- --TCCGGTTTACATGCTTCTAATTTGACTTCTAACCATTCTGCGGGC---AGTACATCG---TCC--------- AGCACCCCTGGATCTCTCCAGTCAGGGTCAGCTAACGCACAGACGTCC-----------AGTCTGCTTCCATCT-------- ----CTTTCTCCATCTGGTAATCCG---TCAGTTTCAGCGGCACTGTTGAAT-- ----GGC--TTGATGTCAGCAGCCGGTGCA-- GCATCTGTTGGGCCGGGAGCCCTT--- ATGTTACAGTCTTCTGGTTCAAGGATATCTCCACTTGGACGT--------------- GAAGCCGAAAAGTTAATGAGTAATGAAGATAAACAACGCGCTCAGCTTAGTTCCGGTCACAACACTACCTATC CAACT-------------------------------ACTTCAAGTGCACCTGCAGTCGGCACTGCCAACGCGAATTCT--- GGACGTCCATCTCAGGCGTCTGAGGTCCCAGCTTCA---ACCCGGTGGAAACAATCATCTAAAAATGTTCAG--- TCTGGTCCCGGTGCTTATTCATTCATAGTGTTGCCTAATGGACAA---ACTCGACCGTGTGCATTG--------- GCGTCCGCTCCCGGAGCGACTACTGCTCAGGGATTGCCTCGTCGACTTCAACACCTTGCTAGTTTACCTCATG GTGACGTCGTTTGTGCTGTCACTTTAGGCCCATGTCCAGTAGGACGT--------- GCTTCTCCCGCCCACTTTGCCTATACTGGTGCTCGAGGCAGCGTCAAGTTATGGGATCTTGCATCGATTAGTG CTAAC------TCTGGAACTGCCCCACTT--- GCTAATAGAGATATTACTCCCCTTGCTACTTTTGACTGTTTATGCTACGATAGCTATGTTAGATCTATCAAACTT TTCCCGGATGTTTCTGGTCTTATAGTTGGTGGTGAATCTAACGCACTCACTGTTTGGGATTTAAATGGTCCAGG ACGACGT--- AGAGCTGAATTGACATTTGAAGCGCCTGCTTGCTATGCACTTGCACTTTCGCTTGATGGAAAACTGGCCTATA GTTGCTGCTCCGATGGCCAAGTTGCTGTTTGGGATATACATAATCAGAGTATTGTTCACCAGTTTCATGGACAT GTTGATGGGACATCATGTGTTGAAATTACTGGAGATGGGAATCGTTTGTGGACAGGTGGTTTAGATCATAAAG TCCGTTGCTGGGATATTCGTGGAAATCCACATGTATCCACCAGTTATTGTTTTCTGAATTTACAGCCTTCA-------- -CGA------------------GATGTCTGTCACATTGAATTCAAGTCCCAAGTGTTTTCACTTGGCTTATCTTCA----------------- ----CATATGGCCCCAACTCGGCGCCAA------ GCTGAATCAGTCTCTCCAGTTTCGGTTGACGGAGAGTCTACCTCTAGTGGTGGTGCTGGTGCTGGAGGAGGT ACTCGCGACACAGTT------AGTCCACAGTCGTTCGTCAGTGGA--------------- AGCAATAAGTCACTATCAATTGAAAGCTGGTTAGCGGTGGGTTTGGAGTCATCTGAAGTTGAAGTCCTAGCTA TAGGACCAGATGGACCACCTGTTTCAGCATTCCCTGTTAATTATAATGTACCACGTGAAGAA--------- TCTGGAAGTGGGAGTGGTAATCCT--TCCCCAATATCCCAT---CATACTGGA-------- -------TCTAATCAACCT------------------------- CAACATTTTCGTCTTACTCATCATGAAAGTTGTGTGTTAGCTTTAAGGTTTGCACATCATGCTGACTGGTTCCTT ACAACAGGTAAAGATCATCAAGTTAATGCTTGGCGAACTCCATATGGTGCCTGTCTTCTTGAAACGAAAGAAG CGGCGTCAGTTCTAACATGTGACATATCCTTAGATGATAAATTTGTGGTTACCGGTTCCGGGGATAAGCGTGC AAATTTGTATGAGGTTATCTTC---------------ACTTCTCAACGA-------------------	-- ---------------------- MEIAKRLSAILTQVLPFLSQE--------------------- -- ----QLSVIGGANPGV-- SGLHASNLTSNHSAG-STS-S--- STPGSLQSGSANAQTS----SLLPS---- LSPSGNP------------------SVSAALLN--G------- ----------LMSAAGA--------------ASVGPGAL---- -----------MLQSSGSRISPLGR----- EAEKLMSNEDKQRAQLSSGHNTTYPT------ ----TSSAPAVGTANANS-GRPSQASEVPAS- TRWKQSSKNVQ-SGPGAYSFIVLPNGQ- TRPCAL--- ASAPGATTAQGLPRRLQHLASLPHGDVVC AVTLGPCPVGR--- ASPAHFAYTGARGSVKLWDLASISAN-- SGTAPL- ANRDITPLATFDCLCYDSYVRSIKLFPDVS GLIVGGESNALTVWDLNGPGRR- RAELTFEAPACYALALSLDGKLAYSCCSDG QVAVWDIHNQSIVHQFHGHVDGTSCVEIT GDGNRLWTGGLDHKVRCWDIRGNPHVST SYCFLNLQPS---R------ DVCHIEFKSQVFSLGLSS------- HMAPTRRQ-- AESVSPVSVDGESTSSGGAGAGGGTRDT V--SPQSFVSG----- SNKSLSIESWLAVGLESSEVEVLAIGPDGP PVSAFPVNYNVPREE---SGSGSGNP--------- -----SPISH-HTG-----SNQP-------- QHFRLTHHESCVLALRFAHHADWFLTTGK DHQVNAWRTPYGACLLETKEAASVLTCDI SLDDKFVVTGSGDKRANLYEVIF-----TSQR- -----

Schistosoma mansoni	gi|360043238| emb|CCD78651.1| groucho-related [Schistosoma mansoni]	ATGTTTCCA------------AGCAGACCACCTGGGCCCTCTGGGCCGGGGGGTTCATGCAAGTTTACTGTATTAGAAACTTGTGATAGAATAAAAGAAGAGTTCACTTGCATACAGCAGCAAAACCACTCGCTGCAACTGGAAAGAGAGAAGCTCCTCAGTGAGCGGTCAGATATGCAACGAATTTGTGTAATGTATTATGAGATGGCTAATGGCTTAAACCTAGAAATGCACAGACAAATGGAAATAGCTAAAAGATTAAGCGCTATTTTGACTCAAGTATTACCATTCTTATCACAAGAG-----------------------------------CATCAATCTCAAGTCCTTTCAGCTATCGAAAGAGCTAAACAAGTGACAATGCAGGAATTGAACTCAGTATTGGCA---GCACAAATAGAAGATGTAAAGTCCTCGAAGTTCTTCAATGGAAATAATATGGGGCTCCCATTAGAACAGTTCGAAGCATACAAGCAGCTGTCTGTTATCGGGGGTGCTAACCCAGGCGTT------TCCGGTTTACATGCTTCTAATTTGACTTCTAATCATTCTGGGGGT---AGTACATCG---TCC---------AGCACCCCTGGATCTCTCCAGTCAGGGTCAGCTAACGCACAGACGTCC------------AGTCTGCTTCCCTCT------------CTTTCTCCATCTGGTAATCCG--TCAGTTTCGGCGGCACTGTTGAAT------GGT---TTAATGTCAGCAGCCGGTGCA---GCATCTGTTGGACCTGGGGCCCTA---ATGCTACAGTCTTCTGGTTCAAGGATATCTCCACTTGGACGT---------------GAAGCCGAAAAGTTAATGAGTAATGAAGATAAACAACGCGCTCAGCTAAGTTCCGGTCACAACACTACCTACCCAAATTACTTCAAGTGC------------------ACCTGCAGT--CGGCACTGCCAACTCGAATTCTGGGCGTCCGTCTCAGGC--CCCGCTCACTTTGCCTATACCGGTGCTCGAGGCAGCGTCAAGTTATGGGATCTTGCATCGATTAGTGCTAAC------TCTGGAAATACCCCACTT---GCTAATAGAGATATTACTCCCCTTGCTACTTTTGACTGTTTGTGCTACGACAGCTATGTTAGATCTATCAAACTTTTCCCGGATGTTTCTGGTCTCATAGTCGGTGGGGAATCCAACGCACTTACTGTTTGGGATTTAAATGGTCCAGGAAGACGT---AAAGCTGAACTGACATTTGAAGCGCCTGCTTGCTATGCACTTGCACTTTCGCTTGATGGAAAACTGGCCTATAGTTGCTGCTCCGATGGTCAAGTTGCTGTTTGGGATATACATAATCAAAGTATCGTTCACCAGTTCCATGGACATGTTGATGGGACATCATGTGTTGAAATTACTGGAGATGGGAATCGTTTGTGGACAGGTGGTTTAGATCATAAAGTCCGTTGCTGGGATATTCGTGGAAAT---------------------------------------CCTTCA---------CGA------------------GATGTCTATCACATTGAATTCAAGTCTCAAGTGTTTTCACTTGGCTTATCTTCA---------------------CATATGGTCCCAACTCGGCGCCAA------GCTGAATCAGTATCTCCAGTTTCGGTTGATGGAGAGTCTACCTCTAGTGGTGGTGCTGGTGCTGGGGGAGGTACTCGCGACACAGCT------AGTCCACAGTCGTTTGTTAGCGGA---------------AGCAATAAGTCACTATCAATTGAAAGTTGGTTAGCGGTGGGTTTGGAGTCATCTGAGGTTGAAGTTCTAGCTATAGGACCAGATGGACCACCTGCTTCAGCATTCCCTGTTAATTATAATGTACCACGTGAAGAA---------TCTGGAAGTGGAAGTGGTAATCCT---TCCCCAATATCTCAT---CATACTGGG---------------TCTAATCAACCT------------------------	MFP----SRPPGPSGPGGSCKFTVLETCDRIKEEFTCIQQQNHSLQLEREKLLSERSDMQRICVMYYEMANGLNLEMHRQMEIAKRLSAILTQVLPFLSQE------------HQSQVLSAIERAKQVTMQELNSVLA------------------AQIEDVKSSKFFNGNNMGLPLEQFEAYKQLSVIGGANPGV--SGLHASNLTSNHSGG-STS-S---STPGSLQSGSANAQTS----SLLPS----LSPSGNP------------------SVSAALLN--G-----------------LMSAAGA--------------ASVGPGAL---------------MLQSSGSRISPLGR-----EAEKLMSNEDKQRAQLSSGHNTTYPNYFKC------TCS--RHCQLEFWASVSG---------------------------------------PAHFAYTGARGSVKLWDLASISAN--SGNTPL-ANRDITPLATFDCLCYDSYVRSIKLFPDVSGLIVGGESNALTVWDLNGPGRR-KAELTFEAPACYALALSLDGKLAYSCCSDGQVAVWDIHNQSIVHQFHGHVDGTSCVEITGDGNRLWTGGLDHKVRCWDIRGN-------------PS---R------DVYHIEFKSQVFSLGLSS-------HMVPTRRQ--AESVSPVSVDGESTSSGGAGAGGGTRDTA--SPQSFVSG-----SNKSLSIESWLAVGLESSEVEVLAIGPDGPPASAFPVNYNVPREE---SGSGSGNP--------------SPISH-HTG-----SNQP--------QHFRLTHHESCVLALRFAHHADWFLTTGKDHQVNAWRTPYGACLLETKEAASVLTCDISLDDKFVVTGSGDKRANLYEVIF-----TPQR------

	*Taenia solium**	TsM_001113700	ATGTATCCC------------AGCAGGCCTCCTGTGCCCTCGGGGCCAGGTCAGTCCTACAAATTTACGGTTGTCGAGACCTGTGAACGCATCAAGGATGAATTCAACTTCGTTCAGCAGCAGTGCCATCAACTTCAAACTGAGCGGGAAAAGCTCTTGAGTGAGCGCGTTGATATGCAACGCATATGTGTTGTCTATTATGAAATGGCCAATGGATTGAACCTGGAGATGCATCGACAGATGGAAATCTCGAAACGTCTCAACGCCATTCTTAATCAAGTCATTCCTTTTCTCGCTACCGAG-----------------------------------CACCAGTCACAGGTCGCGTCAGCTATTGAACGTGCAAAGCAGGTTACAATGCAAGAGTTAAATTCAGTGCTAACG---GCCCAAATGGAAGACTCAAAACTGGCAAAGTTTGCGAATGGTGGAAACATGGGACTGTCCAGTGATCAACTAGAAATCTATAAGCAAATCCAAGCCCACCAGATGTCGTCAGCAGGCAGTCCTTCGTCTTCTCTCCCCGGGAGCTCA---TCCGCCAATCAGACCGGTCCT---CACGGATCA---AACGTTGGCAGCGCCCCCC--------------------------GCTGCAGGTGCAGCATTCCCC------GGATCTACCCCAGGC------------CTCCTTCCGCCCAGTGCACCC--GGGTTGCAAAATGCATTCATTAAT------GGA--TTAATGTCTGCCGCTGGTGGT---GCC---ATGATGCCTCCTCCCGGGACCGCTGGTCTCCCTCTGAGCAGTGGCGCTGTT------CGTGACATTAAGGACTATTCAAGTGAAGACAAGCAGCGAGTGCAACTGCATAGTTCTATGGGAAGCAACTATCCGACA-----------------------------ACAAACAGTGCTCCTGCGGTCGGAACC-----------------------CCGCACAATCGATCAACTTTCTCCGAAAATCGGTCAAAATATGGTGCAAAAGGTCACTCTGCCTCT---TCCGGTCCTGGAGCCTACTCGTACGTGGTGCTTCCAGGCAGTGGAGGCACACGCCCTTGTGCGCTG---------GCGTCAGCACCCGAAGCAACATCGAATCCCAGTCTACCGCGGCGTCTACAACACCTGGCCAGTTTGCCGCACGGTGAAGTGGTATGTGCAGTGACTATCGCACCTGCGCCCTCTGGCCGAGGCTCTTCACCCACCCCA---CATTTCGCCTACACTGGTGGACGCGGTTGTGTCAAGCTCTGGGACCTCGACGCCATT---------------GCTGGC---------TCCTCTTCCTCCAGGGATGCCGTTGCTCTCGCTTCCTTTGACTGTCTTCGTCCAGAAAGCTATGTCCGCTCCATTAAACTCTTCCCGGACTGCTCTACCCTGCTGATTGGAGGCGAATCGAGTTGCCTAACGATTTGGGATCTGAACGGCCCCGGTCGAAGA---AAGGCTGACCTCACATTTGACGCGCCCGCCTGTTATGCACTCGCTCTCTCCCCCGATTGCAAACTTTGTTACAGTTGCTGTTCTGATGGTCAAGTTGCGATATGGGATATCCACAACCAGAGCGTCGTCCAACAGTTTCACGCACACGCTGATGGAGCCTCCTGCATCGAACTGGTGGGCCAGGGCACCCGACTTTGGACCGGCGGCCTTGACAACAAGGTTCGCTGCTGGGACATTCGTGGCACT--GTAGGTTCTCTTTCCCTAATAATAGTA---AAGAGCCTTTTC---AGTGTACGA	MYP----SRPPVPSGPGQSYKFTVVETCERIKDEFNFVQQQCHQLQTEREKLLSERVDMQRICVVYYEMANGLNLEMHRQMEISKRLNAILNQVIPFLATE------------HQSQVASAIERAKQVTMQELNSVLT------------------AQMEDSKLAKFANGGNMGLSSDQLEIYKQIQAHQMSSAGSPSSSLPGSS-SANQTGP-HGS-NVGSAP---------AAGAAFP--GSTPG----LLPPSAP------------------GLQNAFIN--G-----------------LMSAAGG-------------A---------------------MMPPPGTAGLPLSSGAV--RDIKDYSSEDKQRVQLHSSMGSNYPT----------TNSAPAVGT--------PHNRSTFSENRSKYGAKGHSAS-SGPGAYSYVVLPGSGGTRPCAL---ASAPEATSNPSLPRRLQHLASLPHGEVVCAVTIAPAPSGRGSSPTP-HFAYTGGRGCVKLWDLDAI-----AG---SSSSRDAVALASFDCLRPESYVRSIKLFPDCSTLLIGGESSCLTIWDLNGPGRR-KADLTFDAPACYALALSPDCKLCYSCCSDGQVAIWDIHNQSVVQQFHAHADGASCIELVGQGTRLWTGGLDNKVRCWDIRGT----------------------------------VGSLSLIIV---KSLF--SVR

Homeobox protein HoxB4a	*Crassostrea gigas*	gi|405967565|gb|EKC32713.1| Homeobox protein Hox-B4a [Crassostrea gigas]	-- -- -- -- -- -- -- -- ----- GAATCAAAAAGGAATCGAACCGCCTACACAAGACATCAGATTTTAGAGTTGGAGAAGGAATTCCATTTCAATCG ATACTTGACGCGGAGGCGGCGAATTGAAATCGCCCATACCTTGTGTTTGTCCGAACGACAAATAAAAATCTGG TTTCAGAACCGGAGAATGAAATGGAAAAAAGAACATAAACTACCGAACACAAAAACAAGACTC------------------------- -- -- -- -- ATGGAAGCCGGA---ACTGGGCTTCTT------------------GAT--- CCGACTCACCATTTTACTCATATGTTGGGACATCCGGATTTAACTCAGATCNNN------------------------------	-- -- -- -- -- -- -- ESKRNRTAYTRHQILELEKEFHFNRYLTRR RRIEIAHTLCLSERQIKIWFQNRRMKWKKE HKLPNTKTRL--- -- -- --- MEAG-TGLL------D------------- PTHHFTHMLGHPDLTQI?----------

<table>
<tr><td>Echinococcus granulosus</td><td>gi|674569660| emb|CDS15727.1| deformed [Echinococcus granulosus]</td><td>ATGGACTCAGACAGTGGTGACCTCAAT---CGCCAT------AGTAACAGTGCATCAACCCTTTTCTCATTTCCTCTACAGCCTACTGTTCACAAGGAACTTTCCCTCAATCCCTCTATCCACGCACATCAGCAG---CAACACCAACCTTTAAATAATTTGCAGGAAATGCTTGAGACATCAACAACTCTGGACCACTCCCATCACATCCAATTCGATAGTCAA---GCCTCAACGTACTATCCATCGCAACTCCCATCTACGTACGATGAGGGCTCTCCTGAGAGCCTCCGAAAGGCCTCACCGGTGAATTTCCTCCGCAATTTTCCGCCATCTAAGCGACTAATTGACCTTGAAACGGGAGCCCTGCATGCTCTCGATGCTACAAGTGGTTTTGACAGCGTCAAAGGTGATTACTACAGCAATGCGGAGGGATATGATGCAAATTCCAGGGCA---CAATTCCTGCGAAATCCACCTTCAGACTATTATTCAATGGAAACAGTATTTTCAGTACCTTCATCTCATCCATCTCAAATGACC---TCTATAACGTTCCCAAAGGAGGATGAAGCGCAAGGGGAACTTTCTGGCGATAGTGGGAGAGGGTATGACTACACCTCACCACATTCCAACGCCCAATTGAGACAGCCACAG---TCATTGATGTCCCAAATGCGTTCAGAACCTGGGTACTCCAGTTTAGAGGAAAGCTTCGCGAAAGTCTCCGCGCCA---CCACATGCGACAATTCCATCG---CAGCATCCTCCTTCGACCATACCATCCACGAATAATGCTAACAACAATGGTGGTAAAGCCGTGGTATCCAGACTTCAG---GGTAGCAATGCTGGTACGACAGCGATAATCTATCCGTGGATGAAAAGGGTGCACTCCAAAGGCCTTGTAGCTCAAGGTCCTGTTAAAACGGGCAAGCAAGTG--------------------------------AATTTCAGCTCCCAAAAACCCAGTGAGGGCATCAAAAATCGGAAGCGATTACTCGAAGACGATCCATTGCCTTCGAAGAAAGCCTTTTCTGATTCCGATAAGCTTGAAAATACCTCGGAGTCAGTGGGTTCATTGAATCAAGGTAGTGAAGACAGTGACGCTGCAAGTGGT------------ATGGGTGGAGATGACTTGAGTTTGGATATGGTAGGCTCCTCC---------------TGCGATCCAAAACGAACTCGCACTGCTTACACTAGACAGCAGATACTAGAACTTGAAAAGGAATTTCATTATAACAAGTACCTAACAAGGAAACGGCGTCTCGAAATCGCGCACACACTCAGCCTTTCTGAAAGACAGATAAAGATATGGTTTCAGAATCGCCGAATGAAATGGAAGAAGGAGCATTGTTTGCCAGGAAATAAACAGCGTCTTTCAGAGGCACCGATTCTCACTGTTCCAAATCAAAACTTTCCCATGCGCAACCAGGAGTCGATGCAATTCTGCTCAAATCGTCATGGTTTTGACGTGAGTGCCGTCGGTGGAGATCCGATGATGTCAACCCGGTTTTTACCCTTTTTCCCAAAGTATCTCTCTCCAACCACTGGTGCAGCCAAGGCTACGGGATTGACGCCACTTTCAGCACCACCAAAAAGATTTTACGATGACGAAAGTGGATTAGACTCAGTGTGGTTCAACGACAGATGCTATCAGCAGTCCCCGCCCCAA---CTTAGTTTTAACTATCCAACAACTTCTACGCAATCAACATCTCCTCCACCTTCGCTA--------------CCACCTCTCAATAACCCCTTTTCTAAT---ACCCTTGCTGTAGCCCCAACGGATTTTTACTCGGGCCTCCTAACGAGTGGTTGGTCGTCA---AAAGTTCCTTCTCAGCAGCAACAGCACCAGCAAATGCGGCACATCGTCAATCCATCCTATGACGGGGATTATGTAGACAAC------ACTAATAACAATAGCAGCTCACCATTTTTCATT------GCGACTCATAAGCATCCCCACTTCCCTTCAGATGATGAA---TTAGCA------ATATCAGGCAGCAAT-----------------GCA--------------------------GGAAATGAAATCGGACTTCCT-----------------------------AATATGGTGGGAGGGGGTATGAAAAAGGAGTGTCGAATTGAAAAANNN</td><td>MDSDSGDLN-RH--SNSASTLFSFPLQPTVHKELSLNPSIHAHQQ-QHQPLNNLQEMLETSTTLDHSHHIQFDSQ-ASTYYPSQLPSTYDEGSPESLRKASPVNFLRNFPPSKRLIDLETGALHALDATSGFDSVKGDYYSNAEGYDANSRA-QFLRNPPSDYYSMETVFSVPSSHPSQMT-SITFPKEDEAQGELSGDSGRGYDYTSPHSNAQLRQPQ-SLMSQMRSEPGYSSLEESFAKVSAP-PHATIPS-QHPPSTIPSTNNANNNGGKAVVSRLQ-GSNAGTTAIIYPWMKRVHSKGLVAQGPVKTGKQV-----------NFSSQKPSEGIKNRKRLLEDDPLPSKKAFSDSDKLENTSESVGSLNQGSEDSDAASG----MGGDDLSLDMVGSS-----CDPKRTRTAYTRQQILELEKEFHYNKYLTRKRRLEIAHTLSLSERQIKIWFQNRRMKWKKEHCLPGNKQRLSEAPILTVPNQNFPMRNQESMQFCSNRHGFDVSAVGGDPMMSTRFLPFFPKYLSPTTGAAKATGLTPLSAPPKRFYDDESGLDSVWFNDRCYQQSPPQ-LSFNYPTTSTQSTSPPPSL-----PPLNNPFSN-TLAVAPTDFYSGLLTSGWSS-KVPSQQQQHQQMRHIVNPSYDGDYVDN--TNNNSSSPFFI--ATHKHPHFPSDDE-LA--ISGSN------A---------GNEIGLP----------NMVGGGMKKECRIEK?</td></tr>
</table>

Echinococcus multilocularis	gi|674573625|emb|CDS40544.1| homeobox protein Hox B4a [Echinococcus multilocularis]	ATGGACTCAGACAGTGGTGACCTCAAT---CGCCAT------AGTAACAGTGCGTCAACCCTTTTCGCATTTCCTCTACAGCCTACTGTTCACAAAGAACTTCCCCTCAATCCCTCTATCCACGCACATCAGCAG---CAACACCAACCTTTAAATAATTTGCATGAAATGCTTGAGACATCAACAACTCTGGACCACTCCCATCACATCCAATTCGATGGTCGA---GCCTCAACGTACTATCCGTCGCAACTCCCATCTACGTACGATGAGGGCTCTCCTGGGAGCCTCCGAAAGGCCTCACCGGTGAATTTCCTCCGCAATTTTCCGCCATCTAAGCGACTAATTGACCTTGAAACGGGAGCCCTGCATGCTCTCGATGCTACAAGTGGTTTTGACAGCGCCAAAAGTGATTACTACAGCAATGCGGAGGGATATGATGTAAATTCCAGGGCA---CAATTCCTGCGAAATCCACCTTCAGACTATTATTCAATGGAAACGGTATTTTCAGTACCTTCATCTCATCCATCTCAAATGACC---TCTGTAACGTTCCCAAAGGAGGATGAAGCGCAAGGGGAACTTTCTGGCGATAGTGGGCGAGGGTATGACTACACCTCACCACATTCCACCGCCCAATTGAGACAGCCACAG---TCACTGATGTCCCAAATGCGTTCAGAACCTGGGTACTCCAGTTTAGAGGAAAGCTTTGCGAAAGGCTCCGCGCCA---CCACATGCGACGCTTCCATCG---CAGCATCCTCCTTCGACCATACCATCCACGAATAATACTAACAACAGTGGTGGTAAAGCCGTGGTATCCAGACTTCAG---GGTAGCACTGCTGGTACGACAGCGATAATCTATCCGTGGATGAAAAGGGTGCACTCCAAAGGCCTTGTAGCTCAAGGTCCTGTTAAAACGGGCAAGCAAGTG--------------------------------AATTCCAGCTCTCAAAAACCCAGTGAGGGCATTAAAAATCGGAAGCAATTACTCGAAGACGATCCATTGCCTTCGAAGAAAGCCTTTTCTGATTCCGATAAGCTTGAAAATGCCTCGGAGTCAGTTGGTTCATTGAATCAAGGTAGTGAAGACAGTGACGCTGCAAGTGGT------------ATGGGTGGAGATGACTTGAGTTTGGATATGGTAGGCTCCTCC---------------TGCGATCCAAAGCGAACTCGCACTGCTTACACTAGACAGCAGATACTAGAACTTGAAAAGGAATTTCATTATAACAAGTACCTAACAAGGAAACGGCGTCTTGAAATCGCGCACACACTCAGCCTTTCTGAAAGACAGATAAAGATATGGTTTCAGAATCGCCGAATGAAATGGAAGAAGGAGCATTGTTTGCCAGGAAATAAACAGCGTCTTTCAGAGGCACCGATTCTCACTATTCCAAATCAAAACTTTCCCATGCGCAACCAGGAGTCGATGCAATTCTGCTCAAATCGTCATGGTTTTGACGTAAGTGCCGTCGGTGGAAATTCGACGATGTCAACCCGGTTTTTACCCTTTTTCCCAAAGTATCTCTCTCCAACCACCGGTGCAGCCAAGGCTACGGGATTGACGCCACTTTCAGCACCACCAAAAAGATTTTACGATGACGAAAGTGGATTAGACTCAGTGTGGTTCAACGACAGATGCTATCAGCAGTCCCCGCCCCAA---CTTAGTTTTAACTATCCAACATCTTCTACGCAATCAACATCTCCTCCACCTCCGCCA--------------CCACCTCTCAATAACCCTTTTTCTAAT---ACCCTTGCTGTAGCCCCAACGGATTTTTACTCGGGCCTCCTAACGAGTGGTTGGTCGTCA---AAAGTTCCTTCTCAGCAGGAACAGCACCAGCAAATGCGGCGCATCGTCAACCCATCCTATGATGGGGATTATGTAGACAAC------ACTAATAACAATAGCAGCTCACCATTTTTCATT------GCGACTCATAAGCATCCCCACTTCCCTTCAGATGATGAA---TTAGCA------ATATCAGGCAGCAAT-----------------GCA--------------------------GGAAATGAGATCGGACTTCCT-----------------------------AATATGGTGGGAGGGGGTATGAAAAAGGAGTGTCGAATTGAAAAANNN	MDSDSGDLN-RH--SNSASTLFAFPLQPTVHKELPLNPSIHAHQQ-QHQPLNNLHEMLETSTTLDHSHHIQFDGR-ASTYYPSQLPSTYDEGSPGSLRKASPVNFLRNFPPSKRLIDLETGALHALDATSGFDSAKSDYYSNAEGYDVNSRA-QFLRNPPSDYYSMETVFSVPSSHPSQMT-SVTFPKEDEAQGELSGDSGRGYDYTSPHSTAQLRQPQ-SLMSQMRSEPGYSSLEESFAKGSAP-PHATLPS-QHPPSTIPSTNNTNNSGGKAVVSRLQ-GSTAGTTAIIYPWMKRVHSKGLVAQGPVKTGKQV-----------NSSSQKPSEGIKNRKQLLEDDPLPSKKAFSDSDKLENASESVGSLNQGSEDSDAASG----MGGDDLSLDMVGSS-----CDPKRTRTAYTRQQILELEKEFHYNKYLTRKRRLEIAHTLSLSERQIKIWFQNRRMKWKKEHCLPGNKQRLSEAPILTIPNQNFPMRNQESMQFCSNRHGFDVSAVGGNSTMSTRFLPFFPKYLSPTTGAAKATGLTPLSAPPKRFYDDESGLDSVWFNDRCYQQSPPQ-LSFNYPTSSTQSTSPPPPP-----PPLNNPFSN-TLAVAPTDFYSGLLTSGWSS-KVPSQQEQHQQMRRIVNPSYDGDYVDN--TNNNSSSPFFI--ATHKHPHFPSDDE-LA--ISGSN------A---------GNEIGLP----------NMVGGGMKKECRIEK?

<table>
<tr>
<td>Helobdella robusta</td>
<td>gi|675857970|ref|XP_009014545.1| hypothetical protein HELRODRAFT_76537 [Helobdella robusta]</td>
<td>TCCGATTGCGAGAGCAACAACGAAAATGAACGCATC------ATTCCATCG---

---CCGACGAACTCAGCAACATTC--------------

CAGCAACAACAACAACAGCAGCAGCAACAACAACAGCAACAACTTCCTGTTATATACCCGTGGATGAAGATGG
CTCATAACAACAACAGTAATAAACCGAATCAAATTAATAGGAACACTAGGACATTAGGCTACTCACCAAACCTT
CCATACCAACAC------AGTAACGCTAACAACGTTCCACAACAATCTCATCGCCACGAACCAAAGAAC----------------
----
GATAGCAATGGCATGAACATGCTCTGCTCAACGACAAACTTTTCATACTGCGACAACAAGCGAGCTAGAACAG
CCTACACGAAGCATCAAATTTTAGAACTGGAAAAGGAATTCCACTTCAACAGATACCTAACGAGAAGAAGGAG
AATAGAGATTGCTCACTGCCTATGCCTCAGTGAGAGGCAGATAAAAATTTGGTTTCAAAACCGCCGGATGAAA
TGGAAAAAAGAAAACAAGCTACCGAATACTAAAACTTTGAAAAATAAA---------------------
AACAGGGATGTGAGTATTGATAAGAATGATAGCAATAGTTTTAATAACAATAATAATAAT-------------------------------
----

---AATAAT------
AATAATAATGACGACAACGAT--GATGACGAT---CTAGAC---------
GAGAAATTCAAT------------------GATAGCTACGAGAACGACGATGATGAAAAT---GATGATGACAGCTTAAGT-------
-----------------------AATGAAAAAATGGGANNN------------------------------</td>
<td>SDCESNNENERI--IPS----------------------------

-----------PTNSATF---------------------------------

QQQQQQQQQQQQQLPVIYPWMKMAHN
NNSNKPNQINRNTRTLGYSPNLPYQH--
SNANNVPQQSHRHEPKN-----------------
DSNGMNMLCSTTNFSYCDNKRARTAYTK
HQILELEKEFHFNRYLTRRRRIEIAHCLCLS
ERQIKIWFQNRRMKWKKENKLPNTKTLKN
K-------NRDVSIDKNDSNSFNNNNNNN---------

----------------------NN--NNNDDND---------------
-DDD-LD---EKFN------DSYENDDDEN-
DDDSLS---------NEKMG?----------</td>
</tr>
</table>

Hymenolepis microstoma	gi|674593341|emb|CDS27878.1| deformed [Hymenolepis microstoma]	ATGGACGCAGACAGTGGTGACCTTAAT---CATCATAATAACAGTAACAGCGCGACTACACTCTTTTCATTTCCTCTACAACCGGCTGCTCATAAAGGACTTCCA---------------------CACCAGTCGCTTCAACATCCTCACCTGAACAACTTACAGGAAATGATTGAATCTTCTACCACTTTGGACCACTCCCATTACATCCATTTCGATAATCGAGGTGTTTCAATGTACTACCCTTCTCAGCCCTCCTCAGCTTATGATGAAGGTTCTCCAGAGAGTCTTCGAAAGACTTCTCCAGTTAACTTTAACCGCAATATCCCTCCACAGAAAAGAATAATCGATCTCGAGTCTCGGACCATACATGGTAGTGATGTT---------TTTGAACCAATCAAAAGTGATTACTATAGCAACACAGAGAACTATGGGGTAAATCCTAATGTTGGCCAATTTTTGAGAAGTCCACAAGCTGAATACTATTCTATGGAGGGAGGATTCGCACTACCACCTTCCCATTCTGCACAGATAACGGGTAGTTTGGGATTTTCAAAGGATGACGAAGTTCAAGGAGATATATCTGGAGATAGTGGAAGAGGTTACGACTATACCTCTCATCATTCCAGTGCTCAGTTGGGTCATCCTCCTTTGTCTTTAATGCCAATGATGCGAACGGAACCTGCCTACGCGCACTTGGAGGAGACTTTTGGGAAGGCCTCTGCCACAATTCCACATTCAGCTATCCCATCCCAACAGCATCCCCCTTCATCGACATCTTCAACTAACAGTGCTTCCAATAATGGAAGCAAAGCGGCTGGTAATAGACTGCAAAGTGGTAGCAATGCTGGTTCCAATGCTATTATCTATCCGTGGATGAAACGGGTACACTCCAAAGGTCTAGTTGCACAAGGTCCAGTGAAATCGAGTAAACAAGTG--------------------------------AATTCCAATTCGCAAAAACCCAATGACGGTGCTAAAAGTCGCAAACGTATGCTGGAAGATGAGGCGCTGTCTTCTAAGAAGGCCTTCTCTGATTCAGATAAAATGGAAAATATGTCAGAATCAGTGGATTCGCTGAATCAGGGCAGTGAAGATAGTGATACCGCTAGTGGA------------ACAGGTGGCGATGATGCAAGTTTGGATATGGTGGGTTCTTCT---------------GGCGATCCTAAACGCACTCGTACTGCCTATACTCGTCAACAGATACTTGAACTTGAAAAGGAGTTCCATTATAATAAGTATCTTACGAGGAAGCGGAGGCTGGAGATAGCACATACTCTCAGTTTGTCTGAAAGACAAATCAAAATATGGTTTCAGAATCGTCGAATGAAGTGGAAGAAGGAGCATTGTCTACCAGGAAATAAGCAACGTCTCTCAGAACCGCCACTCATTACACTCTCCAATCAGAACTTCTCTATGCGTAATCAAGACTCTATGCAGTTTGGTCCAAATCGTCACGGATTTGATGTAGGTGGTGTTAGTGCCGATCCAATGATGACTTCTCGGTTTCTACCCTTCTTTCCAAAGTATCTCCCACCCACAAGTGGTGGACCAAAAAGCCCATCTATGACATCACTTCCCTCACAACCTAAGCGATTTTATGATGATGGTGGTAGCTTTGATTCCACGTGGTTCAATGATCGTTGCTACCAGCAGTCATCCCCCCAAGCGCCGAGGTTCAATTACCCCACAAACTCTACTCAATCAACTCCTCCACCACCTCCCCTGCAAAATATTCCTGGACCGCCACCGCTAACTCCTTTTCCAAATCAATCCTCGGCAGTACCACCAACGGATTTCTACTCCGGCCTTTTGACAAGTAGCTGGTCGACCACAAAAGTATCATCTTCTCAACCTATGCCTCAACCCATGCGACAAAGTGTCAACTCATCCTATGATAGGGAGTATGCAAATAATAATGCCAGCAACAACAATACCAGCTCTCCATTCTTTATCAACGCAGCGGCTCAGAAACATCCCCAGTTTCCCACTGCAGATGAAGACTTAGCC------TTATCAGGCAGCAATATAGTGGGAGGAGTGGGGGGA--------------------------GGAAATGACATGGATATTCCT-----------------------------AACATGGTGGGAGCACGGATCAAGAAAGAGTGTNNN------------	MDADSGDLN-HHNNSNSATTLFSFPLQPAAHKGLP-------HQSLQHPHLNNLQEMIESSTTLDHSHYIHFDNRGVSMYYPSQPSSAYDEGSPESLRKTSPVNFNRNIPPQKRIIDLESRTIHGSDV---FEPIKSDYYSNTENYGVNPNVGQFLRSPQAEYYSMEGGFALPPSHSAQITGSLGFSKDDEVQGDISGDSGRGYDYTSHHSSAQLGHPPLSLMPMMRTEPAYAHLEETFGKASATIPHSAIPSQQHPPSSTSSTNSASNNGSKAAGNRLQSGSNAGSNAIIYPWMKRVHSKGLVAQGPVKSSKQV-----------NSNSQKPNDGAKSRKRMLEDEALSSKKAFSDSDKMENMSESVDSLNQGSEDSDTASG----TGGDDASLDMVGSS-----GDPKRTRTAYTRQQILELEKEFHYNKYLTRKRRLEIAHTLSLSERQIKIWFQNRRMKWKKEHCLPGNKQRLSEPPLITLSNQNFSMRNQDSMQFGPNRHGFDVGGVSADPMMTSRFLPFFPKYLPPTSGGPKSPSMTSLPSQPKRFYDDGGSFDSTWFNDRCYQQSSPQAPRFNYPTNSTQSTPPPPPLQNIPGPPPLTPFPNQSSAVPPTDFYSGLLTSSWSTTKVSSSQPMPQPMRQSVNSSYDREYANNNASNNNTSSPFFINAAAQKHPQFPTADEDLA--LSGSNIVGGVGG---------GNDMDIP----------NMVGARIKKEC?----

Lollita gigantea	gi|676432025| ref|XP_009046440.1| hypothetical protein LOTGIDRAFT_110972, partial [Lottia gigantea]	---GATTCCAAAAGAAACCGAACAGCTTATACCAGACATCAAGTTCTCGAATTAGAAAAGGAATTCCATTTTAATCGTTATTTGACCAGAAGACGCCGAATAGAGATTGCACACACCTTATGTCTGTCAGAGAGACAAATTAAAATCTGGTTTCAAAATCGCAGAATGAAATGGAAAAAAGAACATAAATTACCGAACACTAAGAGTCGAATG---CTAGAA---------ACGGGGTTAGAT------------------GAT------------------------------GACGATATGAGTCCGACA------------------------------GATCTGACAGTTATTNNN------------------------------	---DSKRNRTAYTRHQVLELEKEFHFNRYLTRRRRIEIAHTLCLSERQIKIWFQNRRMKWKKEHKLPNTKSRM---LE---TGLD------D----------DDMSPT----------DLTVI?----------

*Mesocestoides corti**	MCOS_0000529001-mRNA-1	---------------------ATGAAT---CGCTTT------TCCAGTTGT---CTTGAGACACGTTCAATCCAA---AGAGGCATTGTGACTCAAGGTCCCGTAAAATCTGGAAAACAACCG---------------------------------AATTCTGGCTCCCAGAAACTCAATGAGGGGTCAAAGGTTCGCAAACGACCGATTGACGATGACCAAATC---TCGAAGAAATCCTTCCCTGACTCCGAAAAATTAGAGCACGTGCCTGAATCATTTGATTCTCTTAACCAGGGAAGTGAAGATAGCGATGCTGCAAGTGGTACAGACGGTGGAACAGGGGGAGATGACTTAGGTTTGGACATGGTTGGTTCCTCC---------------TGCGATCCAAAGCGAACCCGGACGGCCTACACTAGACAACAGATACTGGAGCTTGAAAAGGAATTCCATTACAACAAGTATCTGACACGAAAACGGCGCCTAGAAATTGCTCATACGCTAAGTCTGTCAGAGAGGCAAATCAAGATCTGGTTTCAGAATCGCCGAATGAAGTGGAAGAAGGAGCACTGCTTGCCTGGTAATAAGCAGCGTCTCTCAGAACCTCCTCTTCTGAACCTTCCAAACCAAAACTTCCACATACGTAATCCTGATCCTATGCAGTTCTGTGTGAGCCGTCACAATTTCGAAATAAGTAGTGGAACTACGGATCCAATGGCGCCAAGCCGATTTCTGCCTTTTTTTCCAAAGTACCTTTCGTCAGCTGCTGGAGTAACCAAGCCTATGGGAGTGATGTCACTTTCAGCACCACCGAAAAGATTTTACGGCGATGGAGGTGGGTTCGATTCAGCTTGGCTCAATGACAGATGTTACCAGCAACCATCACCTCAA---TCTAGGTTTACCTACCCTGCAACATCTACACAGTCCTCAACCCCT------TCCCTA---------------CCTACATCTAGTATCCCTTTTTCAAAC---CCTACAGCTGTTGCGACGCCGGATTTCTATTCAGGCCTACTAACAGGTGGGTGGCAGTCA---AAAGTCTCCTCACAA------------CACCAAATTCGGCTACACACGAGTTCTTCATATGATGCGGGGTATATTAACAAC------ACCAATAGCAACAGCAACTCTACCTTTTTTGCC------TCTGCCCAGAAGAACCCGAATTTCTCCTTGGACGATGAG---ATAGCA------ATGGCAGGCAGCAAT------------------AAC---------------------------GGAAACGAAATGGACCTCTCG------------------------------GATATGGAGCGAGGTGGAATTAAGAAAGAGGGTCAAATCGACAAGNNN	-------MN-RF--SSC---LETRSIQ---RGIVTQGPVKSGKQP-----------NSGSQKLNEGSKVRKRPIDDDQI-SKKSFPDSEKLEHVPESFDSLNQGSEDSDAASGTDGGTGGDDLGLDMVGSS-----CDPKRTRTAYTRQQILELEKEFHYNKYLTRKRRLEIAHTLSLSERQIKIWFQNRRMKWKKEHCLPGNKQRLSEPPLLNLPNQNFHIRNPDPMQFCVSRHNFEISSGTTDPMAPSRFLPFFPKYLSSAAGVTKPMGVMSLSAPPKRFYGDGGGFDSAWLNDRCYQQPSPQ-SRFTYPATSTQSSTP--SL-----PTSSIPFSN-PTAVATPDFYSGLLTGGWQS-KVSSQ----HQIRLHTSSSYDAGYINN--TNSNSNSTFFA--SAQKNPNFSLDDE-IA--MAGSN------N---------GNEMDLS----------DMERGGIKKEGQIDK?

*Taenia solium**	TsM_000864600	--- ATGCACCCACATATAACGCGCGTGCTAGCAGCCGCGAGCACACGGCATGCG--- --CGG---TATTCTGCTTTTTGTTTCTGCAACCGACCC--------------------------------- -----------------------AACTGGCTG--- --CCCACCGTTTCGTGGTCCGCCACCACCACCTTTACCACCCCCTCT------------ --- -------------------CTCCTCCACTACTTTGAAGCTTCCGCCGATTGT---AAGCCATGTACACTC------------------------------------- -----------------------AACGGTGGG--- ATCGACCACTTAGGACTTGTAGCTCAGGGTCCTATTAAAACGGGCAAACAGGTG------------------------------------- AATTTCAACTCCCCAAAACCCACTGACGGTATGAAATATCGGAAACGATTGCTCGAAGACGAGCCGCTGCCTT CAAAGAAAACCTTTTCTGATTCCGATAAGCTTGAAAATACCTCCGAATCAGTGGATTCATTGAATCAAGGTAGT GAAGACAGTGATGCTGCAAGTGGT-------------ACAGGAGGAGATGACTTGAGTTTGGACATGGTAGGCTCCTCG- --------------- TGTGATCCAAAACGAACTCGCACTGCTTACACCAGACAGCAGATACTAGAACTTGAAAAGGAATTTCATTATAA CAAGTACCTAACAAGGAAACGGCGTCTTGAAATCGCGCACACGCTTAACCTTTCTGAAAGACAGATTAAGATA TGGTTTCAGAATCGCCGAATGAAATGGAAGAAGGAGCATTGTTTGCCAGGAAACAAACAGCGTCTCTCAGAGC CACCAATTCTCAATATTTCAAATCAAAACTTCCCTATGCGCAATCAGGAGTCTATGCAATTCTGCTCAAATCGTC ATGGTTTTGAC--- TATCTCTCTACGACCACTGGTGCAGTCAAGACCACGGGGGTGACTCCACTTTCAGCACCACCGAAAAGATTTT ACGATGAGGAAGGTGGATTCGACTCAGTGTGGTTCAACGATAGATGCTATCAGCAGTCCTCACCCCAA--- CCTAATTTTAACTATCCAACCACTTCTACGGAATCAACCCCCCCTCCACCTCCGCTG----------------- CCACCTTCAAATAACCTCTTCTCTAAC--- CCTCTTGCTGTGGCCCCAACGGATTTCTATTCGGGCCTCCTGACGAGTGGTTGGTCGTCA--- AAAGTTCCTTCCCAACAGCAACAG------ CAAATGCGGCACACTGTCAACCCATCCTACGACGGAGAGTATGTAAACAAC------ ACTAATGACAATAGCAGCTCACCGTTTTGCATT------ ACAGCTAATAAACATCCTCACTTCCTCTCAGGTGATGAA---TTAGCA------ATATCAGGCAGGAAT-------------------- GCA------------------------------GGAAATGAGATTGAAATTCAA--------------------------------- AATATGGTGGCAGGGGGGATGAAAAAGGAGTGTCGAATTGAAAAANNN	----------------------------MHPHITRVLAAASTRHA- ----------------------------------R-YSAFCFCNRP----- -------------NWL--- -------------PTVSWSATTTFTTPS-------------------- -- LLHYFEASADC-KPCTL--------------------NGG-- -------------------------IDHLGLVAQGPIKTGKQV- ---------- NFNSPKPTDGMKYRKRLLEDEPLPSKKTF SDSDKLENTSESVDSLNQGSEDSDAASG-- --TGGDDLSLDMVGSS----- CDPKRTRTAYTRQQILELEKEFHYNKYLTR KRRLEIAHTLNLSERQIKIWFQNRRMKWKK EHCLPGNKQRLSEPPILNISNQNFPMRNQ ESMQFCSNRHGFD---------------------- YLSTTTGAVKTTGVTPLSAPPKRFYDEEG GFDSVWFNDRCYQQSSPQ- PNFNYPTTSTESTPPPPPL----- PPSNNLFSN- PLAVAPTDFYSGLLTSGWSS-KVPSQQQQ- -QMRHTVNPSYDGEYVNN-- TNDNSSSPFCI--TANKHPHFLSGDE-LA-- ISGRN------A---------GNEIEIQ---------- NMVAGGMKKECRIEK?

Lim homeobox protein lhx1	*Caenorhabditis elegans*	gi|17508255|ref|NP_492696.1| Protein lin-11 [Caenorhabditis elegans]	GGAAATGAGTGTGCCGCGTGTGCACAGCCTATTCTTGACAGATATGTATTCACTGTGCTCGGAAAATGCTGGCATCAGTCTTGTCTTCGATGTTGTGACTGTCGAGCGCCCATGTCAATGACTTGTTTTAGTAGAGATGGTCTTATTCTGTGTAAAACTGATTTTTCAAGGCGATATAGTCAACGGTGTGCCGGATGCGATGGGAAATTGGAAAAAGAGGATCTGGTGAGACGAGCAAGAGAT---AAAGTTTTTCATATTCGATGTTTTCAATGCTCCGTTTGTCAAAGGTTATTGGACACTGGTGACCAGCTTTATATCATG---GAAGGCAATCGATTCGTGTGTCAAAGTGATGGGAAGGACAATTCAGATGACTCGAATTCTGCGAAAAGGCGTGGCCCTCGAACGACAATTAAAGCAAAACAACTTGAAACCTTAAAAAACGCGTTCGCCGCAACACCCAAACCAACTCGACATATCCGTGAACAACTTGCCGCCGAGACAGGACTCAACATGAGAGTCATTCAGGTGTGGTTTCAAAACCGCCGAAGCAAGGAACGAAGAATGAAACAACTTCGTTTTGGAGGATATCGTCAATCCCGA	GNECAACAQPILDRYVFTVLGKCWHQSCLRCCDCRAPMSMTCFSRDGLILCKTDFSRRYSQRCAGCDGKLEKEDLVRRARD-KVFHIRCFQCSVCQRLLDTGDQLYIM-EGNRFVCQSDGKDNSDDSNSAKRRGPRTTIKAKQLETLKNAFAATPKPTRHIREQLAAETGLNMRVIQVWFQNRRSKERRMKQLRFGGYRQSR
	Echinococcus granulosus	gi|674568672|emb|CDS17790.1| Lim1 [Echinococcus granulosus]	GCCAACTGCTGCGTTGGATGCGAGCGACCAATTACGGACAAGTACTATCCATGTATCGATGATCAGATTTGGCACCAGGACTGTCTTCGCTGTGTGGTCTGCCGCGTTCAATTAGTGGGCCGGTGTTTCCTCCGAAATGGACAGTACTTTTGTAGAAACGACTTTATACGTCTCTTCAGTCCGCGATGTCCCACCTGCATGGAGACGATCCTGTCGACAGACATGGTCCGACTGCTAGGATCT---GTTGCCTATCACGCTGACTGCTTCCGCTGCGTCCTCTGCGCGCGCTGTCTCTCTACGGGGGATGAGTGTCGATCCCTGGGCGATGGCGTACGATTTGTCTGCATGGAGCACTCATTGGAAGGCAGTGGAAACACCCTCGTCACAAAGCGACGCGGACCTCGTACCACTATCAAGGCTAAGCAGTTGGACACCCTAAAACAGGCCTTCGCCACCACACCAAAACCCACCAGACATATTCGTGAACAATTGGCTCAAGAGACCGGTCTCTCTATGCGCGTCATTCAGGTATGGTTCCAAAATCGTCGCAGTAAGGAGCGTCGGATGAAACAGCTATCGGCGCTGGGAGTACGAAGGTCCTTC	ANCCVGCERPITDKYYPCIDDQIWHQDCLRCVVCRVQLVGRCFLRNGQYFCRNDFIRLFSPRCPTCMETILSTDMVRLLGS-VAYHADCFRCVLCARCLSTGDECRSLGDGVRFVCMEHSLEGSGNTLVTKRRGPRTTIKAKQLDTLKQAFATTPKPTRHIREQLAQETGLSMRVIQVWFQNRRSKERRMKQLSALGVRRSF
	Echinococcus multilocularis	gi|674572412|emb|CDS42837.1| Lim1 [Echinococcus multilocularis]	GCCAACTGCTGCGTTGGATGCGAGCGTCCAATTACGGACAAGTACTATCCATGTATCGATGATCAGATTTGGCACCAGGACTGTCTTCGCTGTGTGGTCTGCCGTGTTCAATTAGTGGGCCGGTGTTTCGTCCGAAATGGACAGTACTTTTGTAGAAACGACTTTATACGTCTCTTCAGTCCGCGATGTCCCACCTGCATGGAGACGATCCTGTCTACAGACATGGTCCGACTGCTGGGATCT---GTTGCCTATCACGCTGGCTGCTTCCGCTGCGTCCTCTGCGCGCGCTGTCTCTCTACGGGGGATGAGTGTCGATCCCTGGGCGATGGCGTACGATTTGTCTGCATGGAGCACTCATTGGAAGGCAGTGGAAACACCCTCGTCACAAAGCGACGCGGACCTCGTACCACTATCAAGGCTAAGCAGTTGGACACCCTAAAACAGGCCTTTGCCACCACACCAAAACCCACCAGACATATTCGTGAACAATTGGCTCAAGAGACCGGTCTCTCTATGCGCGTCATTCAGGTATGGTTCCAAAATCGTCGCAGTAAGGAGCGTCGGATGAAACAGCTATCGGCGCTGGGAGTACGAAGGTCATTC	ANCCVGCERPITDKYYPCIDDQIWHQDCLRCVVCRVQLVGRCFVRNGQYFCRNDFIRLFSPRCPTCMETILSTDMVRLLGS-VAYHAGCFRCVLCARCLSTGDECRSLGDGVRFVCMEHSLEGSGNTLVTKRRGPRTTIKAKQLDTLKQAFATTPKPTRHIREQLAQETGLSMRVIQVWFQNRRSKERRMKQLSALGVRRSF

	Haemonchus contortus	gi|560139938|emb|CDJ81631.1| Zinc finger and Homeobox domain containing protein [Haemonchus contortus]	ATGAACGAGTGCGCCGGCTGCGCTCAACCAATTCTTGACAGGTATGTATTCAATGTGGTCGGTAAGTCGTGGCACCAAGCATGTTTACGCTGCTCTGACTGTCTATCGCCGATGACCGAAACATGCTTCAGTCGCGACGGGTTGATTCTGTGTAAAAGCGATTTCGCCAGACGTTATGGTCAGCGCTGTGCAGGTTGCGATGGTGCGTTGGAGAAGGAAGACTTAGTTCGGAAGGCCCGTGAT---AAAGTGTTTCACATTCAATGCTTCCAGTGTTCTGTCTGTCAGAGGCGGTTGGACACTGGGGAACAGTTGTACATTCTG---GAGGGCAACCGATTCGTCTGTCAGCAGGACGGCGAGGGA---AAGGACGATGCGGCGGCTGCGAAGCGACGAGGACCGCGGACGACCATCAAAGCCAAGCAGTTAGAGACATTGAAGAATGCGTTTGCTTCGACGCCAAAACCGACACGGCACATTCGAGAACAGTTAGCACAGGAAACTGGACTCAATATGAGGGTCATTCAGGTCTGGTTCCAAAACCGTCGAAGTAAAGAACGGCGCATGAAACAGTTACGATTCGGGGGTTTTCGTCCCACGCGG	MNECAGCAQPILDRYVFNVVGKSWHQACLRCSDCLSPMTETCFSRDGLILCKSDFARRYGQRCAGCDGALEKEDLVRKARD-KVFHIQCFQCSVCQRRLDTGEQLYIL-EGNRFVCQQDGEG-KDDAAAAKRRGPRTTIKAKQLETLKNAFASTPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLRFGGFRPTR
	Helobdella robusta	gi|675889286|ref|XP_009030145.1| hypothetical protein HELRODRAFT_70524 [Helobdella robusta]	GTGGTGTTGTGTGCTGGCTGTGATGCGCCCATCCTGGACAAGTTTCTCCTGAATGTTCTAGATCGCACATGGCATACTGACTGCGTGCAGTGTTACGATTGCAAAACTGTTCTCACCGAAAAGTGCTTCTCCAGAGATGGCAAACTTTACTGCAAGATGGATTTTCATAGGTCTGTTGCTGTTAAGTGCGGCGGTTGTGGCCAGGACATCTCAGCCACAGAACTCGTGAGGAGGGCGCGCGAT---CGCGCCTACCACCTCAAATGTTTCACTTGCATCGCCTGTGGCAAGCAGCTATCAACTGGAGAGGAGTTGTACATGTTG---GACGAAGCTCGCTTCCTCTGTAAAGATGACGACGATGATGATGATGGAAGTGATATGACAAGCAAACGCCGAGGGCCCCGCACAACCATCAAAGCAAAGCAGCTGGAAACGTTGAAAGCTGCATTTGCGGCCACACCGAAACCAACGCGCCACATCAGAGAACAGCTGGCACAAGAAACTGGACTCAACATGAGGGTCATTCAGGTCTGGTTCCAGAACAGGCGATCAAAAGAACGACGGATGAAACAACTGAGTGCTCTTGGTGTTCGTCGTCAGTTC	VVLCAGCDAPILDKFLLNVLDRTWHTDCVQCYDCKTVLTEKCFSRDGKLYCKMDFHRSVAVKCGGCGQDISATELVRRARD-RAYHLKCFTCIACGKQLSTGEELYML-DEARFLCKDDDDDDDGSDMTSKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLSALGVRRQF
	Hymenolepis microstoma	gi|674594811|emb|CDS26481.1| LIM/homeobox protein Lhx1 [Hymenolepis microstoma]	GTCAACTGCTGTGTTGGCTGTGAGAAACCAATTATGGACAAATACTACCCCTGCATTGATGATCAGATATGGCATCAGGATTGTCTTCGGTGTGTTGTCTGTCGTATTCAACTGATCGATCGGTGCTTCTTTCGAAATGGGCAGTACTTTTGCAGAAATGACTTCATCAGACTCTTCAGCCCTAGATGTCCAACCTGTCTCGAAACCATTCATCCCACAGACATGGTTCGAATCCTTGGTTCA---GTAGCGTATCATGCTGACTGCTTGCGTTGTGTACTTTGTGCTCGATGTTTATCCACTGGCGATGAATGCAGACCTCTTGGTGACGGTGTTCGATTTGTCTGTTTGGAACATGGCGACAAGAGTGAAGGTGGCAGTATTATTACAAAACGACGAGGACCAAGGACTACGATCAAGGCCAAACAACTCGACACTTTAAAACAAGCATTTGCGACTACACCAAAACCCACTAGACACATTCGCGAACAATTGGCTCAAGAAACAGGTCTATCTATGCGAGTTATTCAGGTTTGGTTTCAAAATCGTCGTAGCAAGGAGCGCCGAATGAAACAGCTCTCCGCACTTGGTGTTCGACGGAACTTC	VNCCVGCEKPIMDKYYPCIDDQIWHQDCLRCVVCRIQLIDRCFFRNGQYFCRNDFIRLFSPRCPTCLETIHPTDMVRILGS-VAYHADCLRCVLCARCLSTGDECRPLGDGVRFVCLEHGDKSEGGSIITKRRGPRTTIKAKQLDTLKQAFATTPKPTRHIREQLAQETGLSMRVIQVWFQNRRSKERRMKQLSALGVRRNF

Lollita gigantea	gi|676436215|ref|XP_009047809.1| hypothetical protein LOTGIDRAFT_200317, partial [Lottia gigantea]	ATGGTTCACTGTGCTGGATGTGAACGCCCTATATCCGACAGGTTTCTTCTCAATGTTTTAGATCGTGCCTGGCA TGCCAAGTGTGTCCAGTGCTTTGACTGTAAAAATAATTTGACGGAAAAGTGTTTTTATCGAGAGGGAAAATTAT ATTGTCGACTCGATTTTTTCAGCCTTGTCACACCG-------------------------------- TCTTCATTAAAACATACTGGACGTTACACTCTGTTAAATTTCGTTTCATGTTTCGCTTGTATGGTGTGTAGAAAA CAGCTATCAACAGGCGAGGAACTTTATATTTTA--- GACGAAAATAAATTTATTTGTAAAGAAGATGGGAAAGACTGTGGTGGCGGTACTGTCGGAGCTAAAAGGCGCG GACCACGAACGACGATAAAAGCCAAGCAGTTAGAAGTGTTAAAAGCAGCTTTTAATGCCACACCAAAACCTAC ACGTCATATCCGTGAACAACTTGCTCAAGAAACCGGACTCAATATGAGAGTGATTCAGGTTTGGTTCCAAAACA GAAGATCGAAAGAGAGAAGAATGAAGCAGTTAAGTGCGCTGGGAGCAAGGCGGCACTTC	MVHCAGCERPISDRFLLNVLDRAWHAKCV QCFDCKNNLTEKCFYREGKLYCRLDFFSL VTP----------- SSLKHTGRYTLLNFVSCFACMVCRKQLST GEELYIL- DENKFICKEDGKDCGGGTVGAKRRGPRTT IKAKQLEVLKAAFNATPKPTRHIREQLAQET GLNMRVIQVWFQNRRSKERRMKQLSALG ARRHF
*Mesocestoides corti**	MCOS_0000652101-mRNA-1	-------------------------------- ATGGACAGGTATTACCCCTGTGTTGATGACAGGACTTGGCACCAGGACTGCCTTCGATGTGTAGTCTGTCGAG CACAGCTCTTTGGTCGGTGCTACGCACGAAATGGACGGTACTTTTGCCGAAATGACTTTATCCGCCAGTCCAG CCCACGATGCCCAACATGTACGGAGGCCATCCTACCGACTGATATGGTGCGTCTGTTGGGTTCC--- ATAGCCTACCATGCCGACTGCTTCCTCTGCATCCTCTGTTCACGTCGCCTAGCAACCGGAGACGAGTGTCGA CTCATCGGCGACGGCGTGAGATTTATTTGCCTCGAGCACAATCAGGTGGAGGGTGGGAACAGTCTGCTGACC AAGAGGCGAGGACCACGAACCACCATCAAAGCCAAACAGCTGGACACACTGAAAGAGGCCTTCGCGACCACT CCAAAGCCCACGAGACACATCCGGGAACGACTAGCTCAGGAAACTGGCCTGTCAATGCGTGTTATTCAGGTA AACCTGGTTGAGATGATGTTCTGTTTTGATGAC---------CTCAGATATGTC---------------TGG	----------- MDRYYPCVDDRTWHQDCLRCVVCRAQLF GRCYARNGRYFCRNDFIRQSSPRCPTCTE AILPTDMVRLLGS- IAYHADCFLCILCSRRLATGDECRLIGDGV RFICLEHNQVEGGNSLLTKRRGPRTTIKAK QLDTLKEAFATTPKPTRHIRERLAQETGLS MRVIQVNLVEMMFCFDD---LRYV-----W
*Onchocerca volvulus**	OVOC1338 OVP03222 WBGene00238147	-- -- --- GCTAATAAGATAGATGATATGTCAGCATCATCAAAAAGGAGAGGACCGCGAACTACCATTAAAGCTAAACAGTT GGATACTCTAAAAGCAGCATTTGCTTCAACACCAAAACCAACAAGGCATATCAGGGAACAATTAGCACAGGAA ACCGGTCTTAATATGCGAGTAATACAGGTTTCTATCCTGTTT-----------------------------TTGCCTTTTTTTTTGTTT-- ----------	-- -- - ANKIDDMSASSKRRGPRTTIKAKQLDTLKA AFASTPKPTRHIREQLAQETGLNMRVIQVS ILF----------LPFFLF----

<table>
<tr><td></td><td><i>Strongyloides ratti</i></td><td>gi|685831957| emb|CEF66878.1| Lim1 [Strongyloides ratti]</td><td>AGTTCTTCATGTTACTCATGTAAAGCTCCAATAAAAGATCGTTATATGTTGATGAGTGATAATTATTATTGGCATGAAAAATGTTTAAAATGTTATGATTGTAAGATGGAGTTAACGGAAAAATATTTTAAAATTGATGGTGTTCAAGTTTGTAAGAAAGACTATTCAAAAAGAGTAGGTAATAAATGTGGTAATTGCAAAAAACAAATTGAAAAAACAGAGATGGTTAGGCAGATAAGAGGA---AAAATATTTCATATCAGTTGTTTTAAATGTTCAAAATGTTTTAAACTTTTCGATACAGGTGATAAGATTTGTAGTTTA---GAAGATGGGACATTTGCTTGTGAAAAAGATGATAATTCAAAATATGATAATGATTTAATATCAAAACGAAGAGGTCCCAGAACAACAATAAAATCAAGACAATTAGAAATTTTAAAGGCAGCATTTAATGCAACACCAAAACCAACAAGACATATCAGAGAACAATTAGCTCAAGAAACAGGATTAAGTATGCGTGTTATACAGGTTTGGTTTCAAAATAGAAGATCCAAAGAAAGGAGATTAAAACAGATGAGATTAACTGGGGGAAGAAGAGATTCA</td><td>SSSCYSCKAPIKDRYMLMSDNYYWHEKCLKCYDCKMELTEKYFKIDGVQVCKKDYSKRVGNKCGNCKKQIEKTEMVRQIRG-KIFHISCFKCSKCFKLFDTGDKICSL-EDGTFACEKDDNSKYDNDLISKRRGPRTTIKSRQLEILKAAFNATPKPTRHIREQLAQETGLSMRVIQVWFQNRRSKERRLKQMRLTGGRRDS</td></tr>
<tr><td></td><td><i>Taenia solium</i>*</td><td>TsM_000103200</td><td>GCCAACTGCTGCGTTGGCTGCGAGCGACCAATTATGGACAAGTACTATCCATGCATCGACGATCAGATTTGGCATCAGGACTGTCTTCGCTGTGTGGTCTGCCGCGTTCAATTGGTGAGCCGGTGCTTCGTCCGAAACGGGCAGTACTTTTGCAGAAACGACTTTACACGTCTCTTCAGTCCGCGCTGCCCTACCTGCACAGAGACGATCCTGTCGACAGACATGGTCCGACTGCTGGGATCA---GTTGCCTATCATGCTGACTGCTTCCGCTGTGTCCTCTGCGCACGCTGTCTCGCTACGGGCGACGAGTGTCGATCCCTCGGCGATGGCGTGCGATTCGTCTGCATGGAGCACTCGTTGGAAAGCAGTGGAAACACTGCCGTCACGAAGCGACGTGGCCCTCGCACTACTATTAAGGCAAAGCAGTTGGACACCTTAAAACAGGCCTTCGCCACCACGCCGAAACCCACCAGACATATTCGCGAACAATTGGCTCAAGAGACTGGCCTCTCTATGCGGGTCATTCAGGTATGGTTCCAAAATCGTCGAAGTAAGGAGCGTCGGATGAAGCAGCTGTCGGCACTGGGGGTGCGTAGGTCCTTT</td><td>ANCCVGCERPIMDKYYPCIDDQIWHQDCLRCVVCRVQLVSRCFVRNGQYFCRNDFTRLFSPRCPTCTETILSTDMVRLLGS-VAYHADCFRCVLCARCLATGDECRSLGDGVRFVCMEHSLESSGNTAVTKRRGPRTTIKAKQLDTLKQAFATTPKPTRHIREQLAQETGLSMRVIQVWFQNRRSKERRMKQLSALGVRRSF</td></tr>
<tr><td></td><td><i>Trichinella britovi</i></td><td>gi|954386316| gb|KRY49335.1| LIM/homeobox protein Lhx1 [Trichinella britovi]</td><td>ATGACACTGTGCGCTGGTTGCAAGAAACCTATATACGATCGTTATTTATATCATGTGATGGATAAATCTTGGCATGGTTCTTGTATTGTTTGTGAAGTATGCCAAACTCCACTGGATGATCGTTGTTTTACGAGAGATGGTCTAATTTTTTGCAAAACAGACTTTTTGAAAAGGTATGGAGCAAAATGTTCGAGGTGTTCCCAAAATTTTTCTCGTGGTGATTTGGTTCGCTATGCCCGAAAT---AAAGCATTTCACATTGACTGCTTTTGCTGCACAATTTGTCAGAAGCGCCTAAATACTGGAGATCAGCTATACATTATC---AATGACAGTACTTTTGTTTGCAAAACTGACGGAACCAGCGTTGGAACTTCCTCCAACGGGGCTAAACGACGCGGACCTCGGACAACAATCAAGGCCAAGCAGCTGGAGACACTCAAAGCGGCCTTCGCAGCCACACCAAAGCCCACCCGGCATATCCGAGAACAATTGGCCCAGGAGACCGGCCTGAATATGCGCGTTATACAGGTTTGGTTCCAAAATCGACGCTCTAAAGAACGGCGCATTAAGCAGCTGCGATTTGGAGCCTTTCGACCGGGAAGT</td><td>MTLCAGCKKPIYDRYLYHVMDKSWHGSCIVCEVCQTPLDDRCFTRDGLIFCKTDFLKRYGAKCSRCSQNFSRGDLVRYARN-KAFHIDCFCCTICQKRLNTGDQLYII-NDSTFVCKTDGTSVGTSSNGAKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRIKQLRFGAFRPGS</td></tr>
</table>

<table>
<tr><td><i>Trichinella pseudospiralis</i></td><td>gi|954436550|gb|KRY85197.1| LIM/homeobox protein Lhx1 [Trichinella pseudospiralis]</td><td>ATGACACTGTGCGCTGGTTGCAAGAAACCTATATACGATCGTTATTTATATCATCTGATGGACAAATCTTGGCATGGTTCTTGTATTGTTTGTGAAGTATGCCAAACTCCATTGGATGATCGTTGCTTTACGAGAGATGGTCAAATTTTTTGCAAAACAGACTTTTTAAAAAGGTATGGAGCAAAATGTTCGAGGTGTTCACAAAATTTTTCTCGTGGAGATTTGGTTCGCTATGCCCGAAAT---AAAGCATTTCACATTGACTGCTTTTGCTGCACAGTTTGTCAGAAGCGCCTAAACACTGGAGATCAGTTATACATTATC---AATGACAGTACTTTTGTTTGCAAAACTGACGTAGCCAGCGGAGGAACTTCCTCCAACGGGGCTAAACGACGCGGACCTCGGACAACAATCAAGGCCAAGCAGCTGGAGACACTCAAAGCGGCCTTCGCAGCCACACCAAAGCCCACCCGGCATATCCGAGAACAATTGGCCCAGGAGACCGGCCTGAATATGCGCGTTATACAGGTTTGGTTCCAAAATCGACGCTCTAAAGAACGACGCATTAAGCAACTGCGATTTGGAGCCTTTCGACCGGGAAGT</td><td>MTLCAGCKKPIYDRYLYHLMDKSWHGSCIVCEVCQTPLDDRCFTRDGQIFCKTDFLKRYGAKCSRCSQNFSRGDLVRYARN-KAFHIDCFCCTVCQKRLNTGDQLYII-NDSTFVCKTDVASGGTSSNGAKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRIKQLRFGAFRPGS</td></tr>
<tr><td><i>Trichuris muris</i>*</td><td>TMUE_s0059000100|lim:homeobox_protein_lhx</td><td>GTGACCACCTGCGCCGGTTGCGACCGACCCATTTACGACCGCTACCTGTATCGAGTACTGGACAAGCCGTGGCATGGCAACTGCATAGTATGTGAGGTATGTCAGGCTCGACTGGACGACAGATGCTTTACCAGGGACGGGCGGATTTACTGCAAGTCAGACTTTCTGAAGCGCTACGGTGCGAGATGCGCCAGCTGCTCGCAGGGTTTCTCCAGGGGTGATTTGATCCGCCACGCTCGGGAC---AAGACCTTCCATGTTGACTGCTTTTGCTGCACGGTGTGTCGGAAGCGGCTAAACACCGGAGATCAGCTCTACGTCATT---AACGATAGCACCTTCGTCTGCAAGGGCGACTCCTCCTCCGGGGTAGGCGGAGGTGGCGGGGCCAAAAGGCGCGGCCCAAGAACCACCATAAAGGCGAAGCAACTGGAAACCTTGAAAGCCGCCTTTGCAGCCACCCCCAAACCGACCAGGCACATTCGGGAGCAGCTGGCCCAGGAGACCGGCCTGAACATGCGCGTCATTCAGGTGTGGTTCCAGAATCGTCGATCCAAGGAAAGGCGCATAAAGCAGCTACGTTTTGGCGCCTTCCGTCCAGGCAGC</td><td>VTTCAGCDRPIYDRYLYRVLDKPWHGNCIVCEVCQARLDDRCFTRDGRIYCKSDFLKRYGARCASCSQGFSRGDLIRHARD-KTFHVDCFCCTVCRKRLNTGDQLYVI-NDSTFVCKGDSSSGVGGGGGAKRRGPRTTIKAKQLETLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRIKQLRFGAFRPGS</td></tr>
<tr><td><i>Neanthes arenaceodentata</i></td><td>gi|345132131|gb|AEN75258.1| Lim1 [Neanthes arenaceodentata]</td><td>ATGGTGATGTGCGCCGGTTGTGAACGCCCGATCCTTGACCGGTTCTTGCTGAACGTATTGGACCGCGCCTGGCACGCGAAATGTGTGCAGTGCGTCGAGTGCCGGTCCAATCTCACGGATAAGTGTTTTAGCCGAGACGGGAAACTCTATTGCCGGGAGGACTTTTTCAGACGGTTCGGCACGAAATGTGGTGGTTGTTCGCAGGGTATATCCCCTAATGACCTTGTGCGAAGGGCTCGTAAC---AAAGTATTTCACCTCAAATGCTTCACGTGCATGGTTTGCCGGAAGCAGTTGTCGACGGGTGAGGAACTGTACGTTTTG---GACGAGAATAAATTCATCTGTAAAGAAGACGGCAAGGACGGTGAGGCGGCCCCAACCGGAACAAAACGCCGTGGCCCCCGAACGACAATCAAAGCTAAGCAACTGGAAGTTCTTAAGGCGGCGTTCGCAGCGACACCCAAGCCCACGCGGCATATACGTGAACAACTTGCTCAGGAGACCGGTCTCAACATGAGGGTCATTCAGGTTTGGTTCCAAAACCGACGGTCGAAAGAAAGGCGGATGAAACAATTATCGGCACTCGGGGCTCGCCGTCACTTC</td><td>MVMCAGCERPILDRFLLNVLDRAWHAKCVQCVECRSNLTDKCFSRDGKLYCREDFFRRFGTKCGGCSQGISPNDLVRRARN-KVFHLKCFTCMVCRKQLSTGEELYVL-DENKFICKEDGKDGEAAPTGTKRRGPRTTIKAKQLEVLKAAFAATPKPTRHIREQLAQETGLNMRVIQVWFQNRRSKERRMKQLSALGARRHF</td></tr>
</table>

Membrane-associated guanylate kinase protein 2	*Echinococcus granulosus*	gi|674564136|emb|CDS21820.1| membrane associated guanylate kinase ww and pdz [Echinococcus granulosus]	------------------ATGAGATCCCCT--TCTGTCCAGAAG------------GGTACTAACAGCGGTAAAGAAGTCGATATCCTGAGGAGTATCCAGATGCGTGGCATCGGATTAAGCGGACCCTGTAAGACACCCGATTTTGTGCCCGCTTCCATCTACCTCCGAGGGCAAAAATCAGACCAGGATATCGGGTTA---GAAGACACAACGATGACAGACTCTCTGCTTGATTTTTCAACAGCACCGGCG------TCAGCACCGTCGTCTCCAATCCGCATACCCAACGGCGGGACACATCTGGTTGATTTTGAAGTCAAATTACGAAGAGGGAGGAAGGGTCTTGGACTCCGATTAGTTGGAGGAGCAGAGGAAGGCACCCAGGTGCATGTAGGAGCTATCACCCCGGGTGGACCAGCAGAACTCAGCGGTCAAATTTTTCCCGGAGATCAGCTAGTCGCCATCAATGGAGTTTCCGTTCTTGGAGCCACTCATACTGGCGTTGTCCACCTTCTCAACGTCACCTCTCGCAGTCCTTCGGGAAATTTTCCCTCCAACACCACAATCACTCTCTCTTTTCGGGGTTATCGATCTAACCCGACCTCATTTGGCAATGGCTCCTCCATGGAACACATTGACTCAATTCAGGCTTACTCACCCCTCCGTCAATACTCTTTTAGG---CGGTCT---------TGTAGTGAAGATTCCTCTATCAACTTATCATCTTCAATTCTATCAAGTGGACAA---AATTCACCCTTTCACCGATCCCTTGGTACCACCCCCAAACCGGGCACAGCGTCTACG------ATTTACAAAGTTCAGCTTCAAAGAAGAGCCAACGAGAGCTTTGGGTTTGCTATTAATTGCTCCCTGAGCCCCAAGAGAGGATGGCATATTGGTGCAATTACCCCGGGTGGACCAGCCGACAGGAGTAAGCAGTTGCACGTTGGTGACAAAATCACCGAAATGAACGGATATCCGTTGGCGCTGCAGTCACACGCCGATGTGGTGGAACAGCTCTGCACAAAGCACCACCGTCTGGAGCTTATAGTGGAGCGACAACAAGCTAGAAAAGTATTCCAAATTCCCGTTCTAGATGCCAATGATAATACTATGCTAGCTAGATGGTCAGAAAGCCGATCAAACGAGCGTCGAAAGCAACAACAACCAAGCTTTCTTGAACCCGTACCAGAAGCTCTTGATTCTTCCTGGAATTCTCGAACTCCTTCCCCACTAAGCTGCAATTTAAAAGAATCAGGATGCTACGCCGTGGATTTGACAGCGGATGAGCGAGGATTTGGTTTTTCTTTGCGCATGAGTCAGGGCCTTCACCGAGGTACTATGACCATACTGCATATTGAGAAAGGAGGTCCCGCATTTCGTGACGGGAAAATGCAGGTTGGTGATGAAGTTCTGGAGATAAACGGGACTCCAACAGAATCCCTGACATACACACAGGCGGCTAAAGTAGTCAAATATGGTGGAGAACATATTCATCTAAAATTGCAGCGTGTGAACACACGCATCTACGACCTGTCTGTTNNN	------MRSP--SVQK----GTNSGKEVDILRSIQMRGIGLSGPCKTPDFVPASIYLRGQKSDQDIGL-EDTTMTDSLLDFSTAPA--SAPSSPIRIPNGGTHLVDFEVKLRRGRKGLGLRLVGGAEEGTQVHVGAITPGGPAELSGQIFPGDQLVAINGVSVLGATHTGVVHLLNVTSRSPSGNFPSNTTITLSFRGYRSNPTSFGNGSSMEHIDSIQAYSPLRQYSFR-RS---CSEDSSINLSSSILSSGQ-NSPFHRSLGTTPKPGTAST--IYKVQLQRRANESFGFAINCSLSPKRGWHIGAITPGGPADRSKQLHVGDKITEMNGYPLALQSHADVVEQLCTKHHRLELIVERQQARKVFQIPVLDANDNTMLARWSESRSNERRKQQQPSFLEPVPEALDSSWNSRTPSPLSCNLKESGCYAVDLTADERGFGFSLRMSQGLHRGTMTILHIEKGGPAFRDGKMQVGDEVLEINGTPTESLTYTQAAKVVKYGGEHIHLKLQRVNTRIYDLSV?

Echinococcus multilocularis	gi|674266266| emb|CDI98123.1| membrane associated guanylate kinase ww and pdz [Echinococcus multilocularis]	------------------ATGAGATCCCCT--- --TCTGTCCAGAAG------------ GGTACTAACAGCAGTAAAGAAGTCGATATCCTGAGGAGTATCCAGATGCGTGGCATCGGATTAAGCGGACCC TGTAAGACACCCGATTTTGTGCCCGCTTCCATCTACCTCCGAGGGCAAAAATCAGACCAGGATATCGGGTTA-- -GAAGACACAACGATGACAGACTCTCTGCTTGATTTTTCAACAGCACCGGCG------ TCGGCACCGTCGTCTCCAATCCGCATACCCAACGGCGGGACACATCTGGTTGATTTTGTAGTTAAAATACAAA GAGGGAGGAAGGGTCTTGGACTTCGATTGGTTGGAGGAGCAGAGGAAGGCACCCAGGTGCATGTAGGAGCT ATCACCCCGGGTGGACCAGCAGAACGGAGCGGTCAAATTTTTCCCGGAGATCAGTTAGTCGCCATCAATGGA GTTTCCGTTCTTGGAGCCACTCATACTGGCGTTGTCCACCTTCTCAACGTCATCTCTCGCAGTCCTTCGGGAA ATTTTCCCTCCAACACCACAATCACTCTCTCTTTTCGGGGTTATCGATCTAACCCGACCTCATTTGGCAATGGC TCCTCCATGGAACACATTGACTCAATTCAGGCTTACTCACCCCTCCGTCAATACTCTTTTAGG---CGGTCT-------- -TGTAGTGAAGAATCCTCTATCAACTTATCATCTTCAATTCCATCAAGGGGACAA--- AATTCACCCTTTCACCGATCCCTTAGTACCACCCCCAAACCGGGCACAGCGTCTACG------ ATTTACAAAGTTCAGCTTCAAAGAAGAGGCAATGAGAGCTTTGGGTTTGCTATTAATTGCTCCCTGAGCCCCGA GAGAGGATGGCATATTGGTGCAATTACCCCGGGTGGACCAGCCGACAGGAGTAAGCAGTTGCACGTTGGTGA CAAAATCACCGAAATGAACGGATATCCGTTGGCGCTGCAGTCACACGCCGATGTGGTGGAACAGCTCTGCAC AAAGCACCACCGTCTGGAGCTTACAGTGGAGCGACAACAAGCTAGAAAAGTATTCCAAATTCCCGTTCTAGAT GCCAATGATAAGACTATGCTAGCTAGATGGCCAGAAAGCCGATCAAACGAGCGTCAAAAGCAACCACAACCAA GCTTTCTTGAATCCGTACCAGAAGCTCTTGATTCTTCCTGGAATTCTCAAACTCCTTCCCCACTAAGCTGCAAT TTAAAAGAATCAGGATGCTACGCCGTGGATTTGACAGCGGATGAACGAGGATTTGGTTTTTCTATGCGCATGA GTCAGGGCCTTCACCGAGGTACTATGACCATACTGCATATTGAGAAAGGAGGTCCCGCATTTCGTGACGGGA AAATGCAGGTTGGTGATGAAGTTCTGGAGATAAACGGGACTCTAACAGAATCCCTGACATACACACAGGCGG CTAAAGTAGTCAAATATGGTGGAGAACATATTCATCTAAAATTGCAGCGTGTGAACACACGCATCTACGACCTG TCTGTTNNN	------MRSP-- -----SVQK---- GTNSSKEVDILRSIQMRGIGLSGPCKTPDF VPASIYLRGQKSDQDIGL- EDTTMTDSLLDFSTAPA-- SAPSSPIRIPNGGTHLVDFVVKIQRGRKGL GLRLVGGAEEGTQVHVGAITPGGPAERSG QIFPGDQLVAINGVSVLGATHTGVVHLLNVI SRSPSGNFPSNTTITLSFRGYRSNPTSFGN GSSMEHIDSIQAYSPLRQYSFR-RS--- CSEESSINLSSSIPSRGQ- NSPFHRSLSTTPKPGTAST-- IYKVQLQRRGNESFGFAINCSLSPERGWHI GAITPGGPADRSKQLHVGDKITEMNGYPL ALQSHADVVEQLCTKHHRLELTVERQQAR KVFQIPVLDANDKTMLARWPESRSNERQK QPQPSFLESVPEALDSSWNSQTPSPLSCN LKESGCYAVDLTADERGFGFSMRMSQGL HRGTMTILHIEKGGPAFRDGKMQVGDEVL EINGTLTESLTYTQAAKVVKYGGEHIHLKL QRVNTRIYDLSV?

Hymenolepis microstoma	gi|674590367|emb|CDS30687.1| membrane associated guanylate kinase ww and pdz [Hymenolepis microstoma]	ATGTTCCGTGCAAAGTTAATGCGATCCTTT---CCTATTCATAAAACTAACCCAAATGGGACTGAAAATCACAAGGATCTTGATATTATAAAGGCCATTAGAATGCGGGGAATCGGTCTTAGTGGACCCTGTAAGACCCCCGATTTCGTACCAGCCTCCGTCTATGTTAAATCCTATAAGTCAAGTCATGACGTCACTTCTGGAGATGATACTACAGTCGCAGAATCGTTTGTCGACTTTTCGCCATCTCCGGTACCAGTTTCTGCGCCCTCATCTCCTCTTCGAAAGCCAAATGGTGGACCACAGCACATCGATCTCGACGTAAGATTAGTTCGAGGACCTAAGGGATTTGGACTTCGACTACTCGGTGGAGCGGAGGAAGGCACTCAGGTTCGTGTGGGAGCTCTTACGCCAGGAGGACAGGCAGAACTGAGTGGAAACGTCTTTCCTGGGGATTTACTGATCGCTATTAATGGAGTTTCGGTTATTGGGATGACCCATAGTTCCGTTGTACAACTTCTAATGGCAGCAGCGCCAACGATA------------------AATAATCCTGTTACTTTAACGCTAAGAGGTCAACGACCGGCTCTCTATAGATCATTAAATGACTCATCTACGGATCAAATGGAGTCAAGTCGAGCATCACCAACCAACCACCAGCAATCATTCAGA---AGATCT---------TTCAGTGAAAATTCGTCCATTCATTCAAGATTTTCGACATCATCCTCGAGCGAACCGAATTCGCCAATCCGACGACCCCTAAATTTAAGTGGTGACCAAAGTGCAAGGTCCTCC------ATCTTCAAGGTCAAACTCCATCGGCGACCAAACGAAACATTTGGCTTTTCGATGAACCGATCTCTCAGCCCCGATAGAGGTTGTCACATTGGAGCTATTATCCCAGGCGGTCCTGCAGCCAGAAGTAGACGCATTAAAGTTGGAGACAAGATCACTGAAATAAATGGACACTCGCTAGCCAAAATATCTCACTCTGATATTGTGGAACAACTCTGCACTCAGCATCGTCGCTTAGAACTCACTATGGAAAGACAGGAACCGTCCAACACCTTCAGAATTCCAGTTTTAGAAAACAACTGTCGAACTCCAACTTTCCTACGGTCCAGTAGTAGATCGAGTGAGCGAAAAAAGCTTCAGCAAACATCCTTCCTAGAAACAGTTCCTGAATCGATTGATTCATCTTGGAATTCCAGATCTTCGACACCGGTTGGATGCAACCTGAAGGAAGCGGAGAGTTACAACGTTGAACTATTATCAGATAGTCGGGGGTTTGGATTTTCTGTTCGAATGAGTCCAGAACTGCAGAAGGGAGCCATGGTTATTCTTCGAATTGAAAAGGGAAGTCCAGCTTACAACGACTGCAAAATGCAGGTTGGAGACAGGGTTCTGGAAATAAATGGTGTTCCAACAGAAACCCTGACCTACACTCAAGCTGTCAAAATAGTGAAATATGGCGGAGACCATCTTCACCTGAAACTTCGGAGAGTCAACACTCGGGCTTATGATCTATCCATTNNN	MFRAKLMRSF---PIHKTNPNGTENHKDLDIIKAIRMRGIGLSGPCKTPDFVPASVYVKSYKSSHDVTSGDDTTVAESFVDFSPSPVPVSAPSSPLRKPNGGPQHIDLDVRLVRGPKGFGLRLLGGAEEGTQVRVGALTPGGQAELSGNVFPGDLLIAINGVSVIGMTHSSVVQLLMAAAPTI------NNPVTLTLRGQRPALYRSLNDSSTDQMESSRASPTNHQQSFR-RS---FSENSSIHSRFSTSSSSEPNSPIRRPLNLSGDQSARSS--IFKVKLHRRPNETFGFSMNRSLSPDRGCHIGAIIPGGPAARSRRIKVGDKITEINGHSLAKISHSDIVEQLCTQHRRLELTMERQEPSNTFRIPVLENNCRTPTFLRSSSRSSERKKLQQTSFLETVPESIDSSWNSRSSTPVGCNLKEAESYNVELLSDSRGFGFSVRMSPELQKGAMVILRIEKGSPAYNDCKMQVGDRVLEINGVPTETLTYTQAVKIVKYGGDHLHLKLRRVNTRAYDLSI?

*Mesocestoides corti**	MCOS_0000436901-mRNA-1	------------------ATGACCTTTCCG--GGACTAAGGAAA------------AGCTCTAGCTTTAGTGTTAAGTTGGATATTCTGAAAAGCATTCAGATGCGAGGAATTGGGTTAAGTGGACCCTGTCAAACACCTGACTTTGTGCCCGCATCCCTATACCTCAAGGCTCAGAAGTCCGATCTGAACAATGCAGAAGGAGATGACACGACAGTATCAGAATCGATGCTGGACCTGCTGGAGCCCCCAATA------TCAGCACCAACATCACCAGTGAACAAACCAACCAGAAGGTCTCAACTGATGGATTTTGAGGTTAAATTACGACGAGGAAAATTTGGTTTTGGACTTCGATTGATAGGAGGGGCGGAGGAAGGCACCCAGGTGCAGGTGGGTGCTCTCACTCCTGGTGGCTCCGCTGAAAACAGCGGCCAAATCTATCCTGGCGATCTTATCATTGCTATCAATGGAATCTCCGTTCTTGGAGCTACTCACAAAACAGTAATCCGTTTGCTCGACACCCTATCTCGCGACCTCTCGGGCAATATTTCCTTCGACTTGGTGATCACCCTCTCCCTCCGTGGCTACCGAGCATCCCAGAAAAACGGTGAATATGAGGCTGCCATCGGGCACTTCGGCTCAAGCCCAACTTGTGCGTCAATTCACCAACTAAGTAGCCAACCGAGAAGTGTCGGAGATTCTTCTAGTGAAGAATCTCTTCATTCAAAACTCTCGTCGCCACAAAGCAACTCC---AGTTCCCCAATTCGTCGCCTC------ACTTCCCCACGTCGTCCTACAAGGCCGATTCCGCTCATTTACAAGGTGGAACTTCACCGACGGTTCGATGAGAGCTTTGGATTTTCCGTTGATCGACTATTCAGTCCTCACCGGGGATGTCACATTGAATCAATTACTCCAGGTGGTCCGGCCGACAGAAGCAAGCGGCTGCAAATCGGTGACAAAATCATCGAAATAAACGGACACTCCCTAGCTTTGCAGCCTCATGAAACTACAATCGAACAACTCTGCACGAAACGTCATCGCATTCAACTCACAATAGAAAAAAAAGAGTCACGCAATTCCTTTCAAATACCGGAACTCAATGAAAATTCGGAA--------------AGGTCGGAACGCAGATCAAGAGAACATCCGCCTGTTCCACGATCCTGTCTTCTTGAAGCAGTTCCTGAGGCCTATGATTCTTACTGGGACTCCCGCTGTGTGAGTAAATGCAATTCAGTTTTG---TTTGTCTTTTATGTTGTGTTTCAAAAAGACTAT---GTCGAATCATTAATA--ATTCAAATTNNN	------MTFP---GLRK----SSSFSVKLDILKSIQMRGIGLSGPCQTPDFVPASLYLKAQKSDLNNAEGDDTTVSESMLDLLEPPI--SAPTSPVNKPTRRSQLMDFEVKLRRGKFGFGLRLIGGAEEGTQVQVGALTPGGSAENSGQIYPGDLIIAINGISVLGATHKTVIRLLDTLSRDLSGNISFDLVITLSLRGYRASQKNGEYEAAIGHFGSSPTCASIHQLSSQPRSVGDSSSEESLHSKLSSPQSNS-SSPIRRL--TSPRRPTRPIPLIYKVELHRRFDESFGFSVDRLFSPHRGCHIESITPGGPADRSKRLQIGDKIIEINGHSLALQPHETTIEQLCTKRHRIQLTIEKKESRNSFQIPELNENSE-----RSERRSREHPPVPRSCLLEAVPEAYDSYWDSRCVSKCNSVL----------------FVFYVVFQKDY-------------------------------------VESLI----------------------------IQI?

*Taenia solium**	TsM_000272100	---------------ATGTCGCGTTCACCTGCGTGCCGCCACGAGATTGGCTATGCCACGCTTCGAGCAGCACTGGCCGGACTGCTAACCAAGCGGGCGTGCGATTGGCTGGATGCAATCCACCGCCTCCATCACGCAAGCCCGACACACGACTGCTTCATTGCTCACGATGCCTTTGCTTCGTACAAAAAG------------GATGCCGACAGCGGAAAAGATTTTGATATCTTGAAGAGTATCCAAGTGCGTGGCATCGGGTTAAGCGGGCCCTGTAAGACACCCGATTTCGTACCCGCTTCCATCTACCTCCGAGGGCAAAAATCAAACCAGGACGTTGAAATG---GAGGACACAACGATGGCCGACTCTCTACTTGAGTTTTCAACAGCACCAGTG------TCAGCACCGTCGTCTCCAACCCGCAAAGCCAACGAAGGAGGACATCTAGTCGATGTTGAAGTCCAACTACAACGGGGAAGGAAGGGTCTTGGACTTCGACTGGTCGGTGGAGCAGAAGAAGGCACCCAGGTACGAGTTGGAGCCATCACTTCAGGAGGGCCAGCAGAACTGAGTGGTCAGATTTTTCCCGGAGATCAGCTAATTGCCGTCAATGGAGTTTCCGTTTTGGGAGCCACTCACACCGATGTTGTCCACCTTCTCAACGTCGCCTCCCGCAGTGCCTTGGGAACTTTTCCCCCCAACACCACAATCACCCTCTCGCTTCGAGGTTATCGATCAAACCTGACCTCACTGGGCAATGACTCCTCGTTGGAACACTTCAACTCAATCTATGCCTACTCTCCCATTCGTCAGTACTCTTTCAGG---CGGTCT---------TGTAGTGAAGAGTCCTCTATCAACTCCAGT--GCTATCACCCCGGGTGGACCGGCTGAGAGGAGCAAGCAGTTGCGCGTTGGTGACAAAATCACCGAGATGAACGGGTATTCACTGGCGATGCAATCACACGCTGATGTCGTGGAACAGCTCTGTACAAAGCACCACCGCCTGGAGCTCACAGTGGAGCGGCAACAATTTGGGAAAGTATTCCAAATTCCTGTTCTAGATGCTACTGATAAGGCTGTGCCAGTTAGACGATCGGGAAGCCGGTCGGGTGAACGCCGCAAGCAGCAACAGCCATGCCTACTTGAACCAGTACCAGAAGCTCTCGATTCTTCTTGGAATTCGCGATCT---GTTGGTGATGAAGTTTTGGAGATAAACGGAACCCCAGCAGAATCCCTTACCTACGCACAGGCGGTTAGAGTTGTCAAGTGCGGTGGAGATCATCTTCACCTTAAATTGCAGCGTGTGAACACACGCATCTACGACTTGTCT---CTT	-----MSRSPACRHEIGYATLRAALAGLLTKRACDWLDAIHRLHHASPTHDCFIAHDAFASYKK----DADSGKDFDILKSIQVRGIGLSGPCKTPDFVPASIYLRGQKSNQDVEM-EDTTMADSLLEFSTAPV--SAPSSPTRKANEGGHLVDVEVQLQRGRKGLGLRLVGGAEEGTQVRVGAITSGGPAELSGQIFPGDQLIAVNGVSVLGATHTDVVHLLNVASRSALGTFPPNTTITLSLRGYRSNLTSLGNDSSLEHFNSIYAYSPIRQYSFR-RS---CSEESSINSS--AITPGGPAERSKQLRVGDKITEMNGYSLAMQSHADVVEQLCTKHHRLELTVERQQFGKVFQIPVLDATDKAVPVRRSGSRSGERRKQQQPCLLEPVPEALDSSWNSRS--VGDEVLEINGTPAESLTYAQAVRVVKCGGDHLHLKLQRVNTRIYDLS-L

rotein kinase Mark2	*us granulosus*	gi|674565903| emb|CDS18998.1| serine:threonine protein kinase MARK2 [Echinococcus granulosus]	ATGCTTTTACTGGAGAATCCAAAATCAGGTTCTTTTATGTATCCCTACTCGTCTGCTGCGCCCTCAAATCCCCC	MLLLENPKSGSFMYPYSSAAPSNPPVVVT
			TGTGGTGGTCACACCCGATTCCGATTCCACG---AGTGCCATGCTCTTGCCAGGACAG---	PDSDST-SAMLLPGQ-PQSPTAMLY---
			CCGCAGTCGCCTACTGCGATGCTCTAT---------	SPPAYASGPSQQKQAYLNNSGGNGATGP
			TCTCCACCAGCTTATGCCTCTGGGCCTTCCCAACAAAAACAAGCCTATCTCAACAATAGCGGTGGCAATGGCG	AANSGVNPSTTNNGAMLDIMSHSLMQSPS
			CCACTGGCCCTGCGGCCAACAGCGGTGTCAATCCCAGCACTACTAATAATGGTGCTATGTTGGATATAATGTC	TQQA-NSNT---
			TCACTCCTTGATGCAGTCCCCGAGCACTCAACAGGCC---AATAGCAACACC---------	GSFYCWPPPPGTGAPASNVAT---
			GGAAGCTTCTATTGTTGGCCCCCTCCTCCCGGTACGGGTGCACCTGCCTCCAATGTGGCCACT---------	ARLPHLTQRRGGGGGAA-YRISSQNP--
			GCACGACTGCCCCACCTAACCCAGAGAAGAGGAGGTGGTGGTGGTGCAGCC---	PSDFSHA-
			TACAGGATCTCCTCCCAGAACCCC------CCTTCCGACTTCTCTCACGCC---	QYYFRQQQQQLAPHKAASETGGYGCVPS
			CAGTACTATTTCCGACAGCAGCAGCAGCAACTGGCTCCGCATAAGGCGGCTAGTGAGACCGGTGGCTACGG	KLMGSGTSATTTSGSSSSNNWKERPHVG
			CTGTGTTCCCAGTAAGCTCATGGGAAGTGGAACTTCCGCTACCACCACCAGTGGCAGTAGCAGCAGCAATAA	KYSLIRTIGKGNFAKVKLAQHVTTGMEVAV
			TTGGAAGGAACGACCACATGTGGGGAAGTATAGCCTGATCCGCACCATAGGCAAAGGCAACTTTGCCAAGGT	KVIDKTQLNPTSLRKLFREVRIMKTLDHPNII
			CAAGCTAGCCCAGCATGTGACAACGGGCATGGAGGTCGCCGTCAAAGTTATAGACAAGACACAGCTGAATCC	KLLEVIESEKHLYLVMEYASNGEVFDYLVT
			GACGAGCCTAAGAAAGCTCTTCCGGGAAGTACGGATAATGAAGACCTTGGACCATCCCAATATTATCAAACTT	HGKMKEKDARIKFRQIVSAVQYCHAKNIVH
			CTAGAGGTCATCGAGTCGGAGAAGCACCTGTACCTTGTCATGGAGTATGCTAGCAACGGCGAGGTGTTCGAC	RDLKAENLLLDESMNLKIADFGFSNNYSAG
			TATCTCGTTACCCACGGAAAAATGAAGGAGAAGGATGCACGGATAAAGTTTCGTCAGATTGTCTCTGCTGTAC	RKLDTFCGSPPYAAPELFLGRKYDGPEVD
			AATACTGCCATGCTAAAAACATTGTACATCGTGATCTCAAGGCGGAGAATCTGTTGTTGGACGAGTCGATGAA	VWSLGVILYTLVSGSLPFDGKNLKELRECV
			CCTGAAGATAGCGGACTTCGGGTTTTCCAACAACTACTCAGCTGGTCGAAAATTGGACACCTTCTGCGGGTCA	LRGSYRVPFYMSHECEMLLRKMLVLNPTK
			CCGCCATACGCGGCGCCGGAGCTCTTCCTTGGCCGCAAGTACGACGGCCCAGAAGTGGACGTTTGGTCTCT	RASLLDIMKDKWLNTTFEDNILQPFKEDLP
			CGGGGTCATCCTCTACACCCTCGTCTCTGGCTCACTCCCTTTTGACGGCAAAAATCTCAAGGAACTGCGGGAA	NYNDPERIQWMVQMGFSRSDIHDSLMKQ
			TGTGTGCTCCGAGGCTCCTATCGTGTGCCCTTCTACATGTCTCACGAGTGCGAGATGCTACTCAGGAAGATGC	RFNNITATYILLGQRKRKSLPWPPTLSGV--
			TCGTTCTTAATCCGACCAAACGCGCTTCTCTGCTAGATATTATGAAGGACAAGTGGTTGAACACAACTTTCGAG	MQPRSLPSDMPMSDGSASSSTAAGSGRQ
			GACAACATCTTGCAGCCCTTCAAGGAGGACCTCCCCAATTACAACGATCCGGAGAGAATTCAATGGATGGTGC	-PSSQARQ-LTSTTTSTAAA-
			AGATGGGTTTCTCGCGGAGTGACATTCACGACTCTCTGATGAAGCAGCGGTTCAACAACATCACTGCTACGTA	SFRRPFHNPSSLA----------
			CATTTTGTTGGGTCAACGAAAAAGAAAGTCGCTACCTTGGCCTCCCACCCTGTCTGGCGTC------	DATNRKHSNADIDSN---
			ATGCAGCCTCGAAGCCTCCCATCGGATATGCCCATGAGCGATGGGAGTGCTTCGTCATCAACAGCGGCAGGA	YASDGYRKTTISGASTTGGGGQSKVTPT--
			AGCGGTCGACAA---CCATCTTCACAGGCTCGTCAA---CTAACCTCCACCACAACTTCCACCGCCGCCGCC---	LTA----
			TCCTTCCGTCGTCCTTTCCACAATCCCAGCTCACTGGCC-----------------------------	ANAEGDASIVNAPLAITDNDDVNNVSIPTNT
			GACGCCACCAATCGCAAGCACTCCAACGCTGATATTGACAGCAAC---------	IVTTTAGTIVSTT--
			TACGCCTCTGATGGGTATAGAAAAACTACTATTTCCGGTGCATCCACCACAGGAGGAGGAGGTCAGAGCAAG	GTLKRQQTRPDSQQPTTNLTTSPAPSPPP-
			GTGACACCAACC------TTGACGGCA------------	PESSSTSKPLKNQFVTLSTS--------
			GCCAATGCGGAAGGTGACGCTAGCATTGTAAATGCGCCTCTGGCGATAACGGATAATGACGATGTGAATAAT	ANSPSTVNPVVGTFKDSRNGHTNSRPSAT
			GTCTCCATCCCGACTAACACGATAGTGACTACCACCGCGGGCACCATCGTCTCTACTACA------	SAQTSSRGVPAVSKHPSFTHRNILTGTE----
			GGAACCTTGAAGCGGCAACAAACGCGGCCAGACTCCCAACAACCAACAACTAACCTCACCACCTCTCCTGCT	------DDFNVPILAPRASDAAVA-
			CCCTCACCACCCCCT---	GAEASLGGRVSKRDDTTSNPSTIVATTRA
			CCGGAGTCGTCCTCCACTTCGAAACCGCTGAAAAACCAATTCGTGACCTTATCCACCTCT-----------------------	QPYELINPAALESSLQ-------NASDMAG-
			GCCAATTCTCCCAGCACCGTTAATCCTGTTGTAGGCACCTTCAAAGATTCGAGAAATGGTCATACAAACTCCC	ISSNTNQSLFLRNHPSLGYRSLRLPADAAA
			GTCCCAGTGCTACTTCAGCTCAGACATCAAGTCGAGGTGTCCCGGCCGTTTCCAAGCATCCATCCTTCACTCA	ELRAAAAAATLSGTATMGEAHPSTYYNT---

s multilocularis	gi|674576610| emb|CDS3704 7.1| serine:threonin e protein kinase MARK2 [Echinococcus multilocularis]	ATGCTTTTACTGGAGAATCCAAAATCAGGTTCTCCTATGTATCCCTACTCGTCTGCTGCGCCCTCAAATCCCCC TGTGGTGGTCACACCCGATTCCGACGCTACG---AGTGCCATGCTCTTGCCAGGACAG--- CCGCAATCGCCTACTGCGATGCTCTAT--------- TCTCCACCAGCTTATGCCTCCGGACCTTCCCAACAAAAACAAGCCTATCTCAACAATAGTGGCGGCAATGGCG CCACTGGCCCTGCGGCCAACAGCGGTGTCAATCCCAGCACTACTGATAATGGTGCTATGTCGGATATAATGTC TCACTCCTTGATGCAGTCCCCGAGCACCCAACAGGCC---AATAGCAACATC--------- GGAAGCTTCTATTGTTGGCCCCCTCCTCCCGGTACGGGTGCACCTGCCTCCACTGTGGCCACT--------- GCACGACTGCCCCACCTAACCCAGAGGCGAGGAGGTGGTGGTGGTGCAGCC--- TACAGGATCTCCTCCCAGAACCCC------CCTTCCGACTTCTCTCATGCC--- CAGTACTATTTCCGACATCAGCAGCAGCAATTGGCTCCGCATAAGGCGGCTAGTGAGACCGGTGGCTACGGC TGTGTTCCTAGTAAGCTCATGGGAAGTGGAACTTCCGCTACCACCACCAGTGGCAGTAGCAGCAGCAACAATT GGAAGGAACGGCCACATGTGGGGAAGTATAGCCTAATCCGCACCATAGGCAAAGGCAACTTTGCCAAGGTCA AGCTAGCCCAGCATGTGACAACGGGCATGGAAGTCGCCGTCAAAGTTATAGACAAGACGCAGCTGAATCCGA CGAGCCTAAGAAAGCTCTTCCGGGAAGTACGGATAATGAAGACCTTGGACCATCCCAATATTATCAAACTTCT GGAGGTCATCGAGTCGGAGAAGCACCTGTACCTTGTCATGGAGTATGCTAGCAACGGCGAGGTGTTCGACTA TCTCGTTACCCACGGAAAAATGAAGGAGAAGGATGCACGGATAAAGTTTCGTCAGATTGTCTCTGCTGTACAA TACTGCCATGCTAAAAACATTGTACATCGTGATCTCAAGGCAGAGAATCTGTTGTTGGACGAGTCGATGAACC TGAAGATAGCGGACTTCGGATTTTCCAATAACTACTCAGCTGGTCGAAAATTGGACACCTTCTGCGGGTCACC GCCGTACGCGGCGCCGGAGCTCTTCCTTGGCCGCAAGTACGACGGCCCAGAAGTGGACGTTTGGTCTCTCG GGGTCATCCTCTACACCCTCGTCTCTGGCTCACTCCCTTTTGACGGCAAAAATCTCAAGGAACTGCGGGAATG TGTGCTCCGAGGCTCCTATCGTGTGCCCTTCTACATGTCTCACGAGTGCGAGATGCTACTCAGGAAGATGCTC GTTCTTAATCCGACCAAGCGCGCTTCTCTGCTAGATATTATGAAGGACAAGTGGTTGAACACAACTTTCGAGG ACAACATCTTGCAGCCCTTCAAGGAGGACCTCCCCAATTACAACGATCCGGAGAGAATTCAATGGATGGTGCA GATGGGTTTCTCGCGGAGTGACATTCACGACTCTCTGATGAAGCAGCGGTTCAACAACATCACTGCTACGTAC ATTTTGTTGGGTCAGCGAAAAAGAAAGTCGCTACCTTGGCCTCCCACCCTGTCTGGCGTC------ ATTCAGCCTCGAAGCCTCCCGTCAGATATGCCCACGAACGATGGGAGTGCTTCGTCATCAACAGCGGCAGGA AGCGGTCGACAA---CCATCCTCACAGGCTCGTCAA---CTAACCTCCACCACAACTTCTACTGCCGCCGCC--- TCCTTCCGTCGTCCTTTCCACAATCCCAGCTCACTGGCC----------------------------- GACGCCACCAATCGCAAGCACTCCAACGCTGATATTGACAGCAAC--------- TACACCTCTGATGGGTATAGAAAAACTAGTATTTCCGGTGCATCCACCACAGGAGGAGGAGGTCAGAGCAAG GTGACACCAACC------TTGACGGCA------------ GCCAATGCGGAAGGAGACGCTAGCATTGCAAATGCGCCTCTGGCGATAACGGATAATGACGATGTGAATAAT GTCTCCATCCCGACTAACGCGACAGTGACTACCACCGCGGGCACCATCGTCTCTACTACA------ GGAACCTTGAAGCGG--------------------- CAACAACCAACAACTAACCTTACCATTTCTCCTGCTCCCTCACCACCCCCT--- CCGGAGTCGTCCTCCACTTCGAAACCGCTGAAAAACCAATTCGTGACCTTATCCACCTCT----------------------- GCCAATTCTCCCAGCACCGTTAATTCTGTTGTAGGCACCTTCAAAGATTCGAGAAATGGTCATACAAACTTCCG TCCCAGTGCTACTTCAGCTCAGACATCAAGTCGAGGTGTCCCGGCCGTTTCCAAGCGTCCATCCTTCACTCAC	MLLLENPKSGSPMYPYSSAAPSNPPVVVT PDSDAT-SAMLLPGQ-PQSPTAMLY--- SPPAYASGPSQQKQAYLNNSGGNGATGP AANSGVNPSTTDNGAMSDIMSHSLMQSPS TQQA-NSNI--- GSFYCWPPPPGTGAPASTVAT--- ARLPHLTQRRGGGGGAA-YRISSQNP-- PSDFSHA- QYYFRHQQQQLAPHKAASETGGYGCVPS KLMGSGTSATTTSGSSSSNNWKERPHVG KYSLIRTIGKGNFAKVKLAQHVTTGMEVAV KVIDKTQLNPTSLRKLFREVRIMKTLDHPNII KLLEVIESEKHLYLVMEYASNGEVFDYLVT HGKMKEKDARIKFRQIVSAVQYCHAKNIVH RDLKAENLLLDESMNLKIADFGFSNNYSAG RKLDTFCGSPPYAAPELFLGRKYDGPEVD VWSLGVILYTLVSGSLPFDGKNLKELRECV LRGSYRVPFYMSHECEMLLRKMLVLNPTK RASLLDIMKDKWLNTTFEDNILQPFKEDLP NYNDPERIQWMVQMGFSRSDIHDSLMKQ RFNNITATYILLGQRKRKSLPWPPTLSGV-- IQPRSLPSDMPTNDGSASSSTAAGSGRQ- PSSQARQ-LTSTTTSTAAA- SFRRPFHNPSSLA---------- DATNRKHSNADIDSN--- YTSDGYRKTSISGASTTGGGGQSKVTPT-- LTA---- ANAEGDASIANAPLAITDNDDVNNVSIPTN ATVTTTAGTIVSTT--GTLKR------- QQPTTNLTISPAPSPPP- PESSSTSKPLKNQFVTLSTS-------- ANSPSTVNSVVGTFKDSRNGHTNFRPSAT SAQTSSRGVPAVSKRPSFTHRNILTGTE---- ------DDFNVPILAPRASDAAVA- GAEASLGGRVSKRDDTTSNPSTIVATNRA QPYELINPAALESSLQ-------NASNMAG- ISSNTNQSLFLRNHPSLGYRSLRLPTDAAA ELRAAAAAAATLSGTATMGEAHSSTYYNT---

pis microstoma	gi|674595624|emb|CDS25677.1| serine:threonine protein kinase MARK2 [Hymenolepis microstoma]	ATGCTTTTATTGGAGAATCCAAAACCAGGTTCTCCCATGTATTCATACACATCACCACCAGTTAATAAC------------ATCTCAAGTGAGTTTGATTCCACTCCCAGCGAAATGCTTCTTCCGGGGCAACTGCCTCAAAATCCGCCCTCAATGCTCTATCCCTCTGCTTCTCCCTCGAACTACACC------CCCAACCAACCAAAATCTGGATACCTCAATCCTGCTGCCAATAATACCAGCACCATGAATAGCAATAATAGCGGTGCCAAC------------AACAATGGAGCAATGTCAGATATAATGTCTCACTCTATGATGCAAACGCCTCTGCAACCAGGGGCGAGTAATTCTAATATTATTACTGCAGGAAGCTTTTATCGTTGGCCGCCACCTTCTAATGGGCCAAATACTCCTTCCAATCCACCTACAAATAATTTCACTCGAATGCCCCACATGGGCCAGAGACGCACTAGTGGTGCGTCAGCGACGGGGTATAGAATCTCTCCTCATAATCCCCCACCTCCATCGGACTATTCTCATGCTCAGCAGTACTATTTTCGTCAGCAACAGCAGCAG------CAGCACAAAACGATTAGTGAAACTGGAGCTTAT------ACTGGCAGCAAATTGTCTGCTCCTGGAAATCCTTCTGCAAATCCCTCTGGTGTTAGTGGTAGCAGTAATTGGAAAGAGAGACCACATGTGGGAAAGTATAGCCTCATTCGGACAATTGGAAAGGGAAATTTTGCCAAAGTGAAGCTGGCTCAACATGTAACTACGGGGATGGAGGTCGCTGTGAAGGTTATCGATAAAACCCAACTCAACCCAACAAGCCTTCGAAAGCTCTTCCGAGAAGTGCGAATAATGAAGACACTGGATCATCCAAATATAATCAAATTATTGGAGGTGATTGAATCTGAGAGGCACCTCTACCTGGTTATGGAATATGCTAGCAACGGTGAAGTTTTTGACTACCTCGTAACTCATGGAAAGATGAAGGAGAAGGATGCGCGAATAAAGTTCCGTCAAATCGTATCCGCTGTGCAATATTGCCACGCCAAGAATATTGTACACCGTGACCTCAAGGCGGAAAACCTCCTGCTTGACGAATCAATGAACTTGAAGATCGCGGACTTTGGGTTTTCCAATAACTACTCTGCCGGACGCAAGTTGGATACTTTCTGTGGTTCACCACCCTATGCTGCTCCTGAACTTTTCCTTGGACGGAAGTACGATGGACCGGAAGTGGATGTTTGGTCACTTGGAGTCATCCTCTACACCCTGGTCTCGGGATCTCTTCCTTTTGATGGAAAGAATCTCAAAGAACTGCGTGAATGCGTTCTCCGAGGTTCCTATCGTGTGCCCTTCTACATGTCTCACGAATGTGAAATGCTTCTCAGAAAGATGCTCGTCCTCAATCCCACTAAAAGAGCCTCCCTTCTTGAGATTATGAAGGACAAGTGGTTGAACACTACCTTTGAAGACAATACTCTACAGCCTTACAAAGAAGACCTTCCCAATTACAATGATCCTGAGAGAATTCAATGGATGGTCAACATGGGTTTCTCCCGCAGTGACATCCACGATTCCCTGACCAAACAACGCTTCAACAACATCACGGCCACGTACATTCTACTCGGTCAAAGACGTCAGAAATCTCTTCCCTGGCCACAAACCCTCTCTGGAGCT------ATTCAACCACGAAGCCTCTCTTCGGAAGCCTCCACATCTAATGGAAATTCCACAGTG---------------AATGGTCGACAGTCTAACCCTCGTCAACTACAGCAGCCTCTCACTTCGGCTTCAACGTCAACC---GCCGCG---TCCTTCCGACGACCTTTTCATGCTCCCAATATCTCTGCTACGACTGGTCAGTCGAATGATAATGCTACTCCCGTGACTTCTAGAAAATATTCGAATGTGGAGAGTGGTAGTAGTGTTAACAGTTATGGAAGTGAGGGATATAGAAAAAGCTGTGTTCCT------TCAACTTCTGCAACAGGTGGTCAGAACAAAATTTCCTCGTCA------ACAACGCCGGGAAGAAGCGGCGGTGGAGCAGAAGTTGATGCCACTCTGCCAAATGCTCCCTTAGCAACATCAGACAACGATGATGTGAATAACGTCTCCAAT------------------TCTCCAGCTCCG------------ACCACTTCT------GGAACCCTGAAACGTCAGCAGACC--------------------------------------TCCCCTGTCTCTTCTCCCGTCCCTCAAGACCCATCATCG--------------------------AAAGCCTTGACAACAACGACAACAAAGAACCTCACTTCTGCTATCGCCACATCTAGCAGTACAACTGATCCTTCTATAGGCAATGTTTTAGAATCGAGA------AATGGTAATTCTCGTACTGGTACAACACCCACAAGTCGCCCAAAT---------------GTCCCCAAACGCTCCTCCTTCAATCATCGAAATATCCTAGCCGGTACAGAGAGTGCCAATAACAACAGCAACA	MLLLENPKPGSPMYSYTSPPVNN----ISSEFDSTPSEMLLPGQLPQNPPSMLYPSASPSNYT--PNQPKSGYLNPAANNTSTMNSNNSGAN----NNGAMSDIMSHSMMQTPLQPGASNSNIITAGSFYRWPPPSNGPNTPSNPPTNNFTRMPHMGQRRTSGASATGYRISPHNPPPPSDYSHAQQYYFRQQQQQ--QHKTISETGAY--TGSKLSAPGNPSANPSGVSGSSNWKERPHVGKYSLIRTIGKGNFAKVKLAQHVTTGMEVAVKVIDKTQLNPTSLRKLFREVRIMKTLDHPNIIKLLEVIESERHLYLVMEYASNGEVFDYLVTHGKMKEKDARIKFRQIVSAVQYCHAKNIVHRDLKAENLLLDESMNLKIADFGFSNNYSAGRKLDTFCGSPPYAAPELFLGRKYDGPEVDVWSLGVILYTLVSGSLPFDGKNLKELRECVLRGSYRVPFYMSHECEMLLRKMLVLNPTKRASLLEIMKDKWLNTTFEDNTLQPYKEDLPNYNDPERIQWMVNMGFSRSDIHDSLTKQRFNNITATYILLGQRRQKSLPWPQTLSGA--IQPRSLSSEASTSNGNSTV-----NGRQSNPRQLQQPLTSASTST-AA-SFRRPFHAPNISATTGQSNDNATPVTSRKYSNVESGSSVNSYGSEGYRKSCVP--STSATGGQNKISSS--TTPGRSGGGAEVDATLPNAPLATSDNDDVNNVSN------SPAP----TTS--GTLKRQQT-------------SPVSSPVPQDPSS---------KALTTTTTKNLTSAIATSSSTTDPSIGNVLESR--NGNSRTGTTPTSRPN-----VPKRSSFNHRNILAGTESANNNSNNAGGDFNVPILAPRASDATVAPGE------RISRRDEAAVNSTTGGGQSRPQAYELINTAALENTLQ-------NAG-ASQKNSNTGQSMFSRNHSSLGYRSLRLPTDTAAELRSVAMAAATNGASPMGDTHPP-NYYTANTNNSTPSGA-

*Mesocestoides corti**	MCOS_0000801201-mRNA-1	---ATGAAAGAAAAGGATGCACGAGTAAAGTTTCGTCAAATCGTGTCTGCGGTGCAATATTGCCATGCCAAAAACATTGTTCATCGTGATCTCAAGGCGGAGAATCTGTTGCTGGATGAATCAATGAATCTGAAGATAGCAGATTTCGGATTTTCCAATAACTATTCATCGGGTCGAAAGCTGGATACATTTTGTGGTTCCCCACCTTACGCAGCGCCTGAGCTCTTCCTCGGCAGGAAATACGACGGTCCTGAAGTGGATGTCTGGTCTCTCGGTGTCATTCTTTACACTCTCGTTTCCGGTTCTCTCCCGTTTGACGGGAAAAATCTCAAAGAACTGCGTGAATGCGTCCTTCGAGGCTCCTACCGCGTGCCCTTCTACATGTCCCATGAATGTGAGATGCTGCTCAGGAAGATGCTTGTTCTCAATCCAACCAAGCGCGCTTCTCTTCTGGATATTATGAAAGACAAGTGGCTGAATACGGCCTTTGAGGATAACACTCTGCAACCCTTTAAAGAGGACATCCCTGACTACAATGATCCTGAGAGAATACAATGGATGGTACAGATGGGCTTCTCTAGAAGCGATATCCACGACTCGCTGACGAAGCAGCGGTTCAACAATATCACCGCGACGTACATCCTGCTAGGTCAACGAAGGCAGAAGTCCCTTCCCTGGCCTCCAACCCTGACGGGCAGTAGTGTCGTGCCACAGCGGAGCCTCTCGTCGGAGGTGCCGGCACCA---------------GTG---------------AACGGCCGCCAA---AACCCT--GCGACCACAGGATCCTCCCAA------------CCGCCACAATCTCGCCAACTTACTCCCACAACCACCACCACTAATGCTGCTTCCTTCCGG------------CGGCCCTACCACCCTCCTCCCTCATCCTCCTCAAACACTAGTGCTCAGGTCAAAGTGATGTCAACTGCCTTGACCGCGCCT------------GGTAACATGGAAGGTGACGCAAGCTTGGCGAATGCACCGCTTGCGACGACAGACATTGACGACGTAAAT---------AAC------------------TCTCCCGCAAAC------------ACTACCACAATTATAGGCACCTTAAAGAGGCAGCAGGCGCGTACGGAACCTCACCAACCTATCACC---------ATTTCCCCCGTGCCCTCACCCTCTCCC---GACTCCTCTACTGGCACGGTGAAGCCTTCAAGGAATCAGTCTGCACTTTCGTCAACCTCTACT---------------------GCCGTAGAATCTAGCACTACGACCGGAAGT------ACGGTTCAAGATTCGAGGAGCGAGCATGCAGATTCTCGTGGGGGTGCGACCTCAACCCAGACACCGAATCGTGGTGCTCCTACTGTGCCCAAACGCCCATCCTTCACACACAGGAACATCACTGCCGGTATTGAG------------------------------GGTGAGTTTCACGTGCCGATTTTGGCGCCACGTGCATCAGACGCTGCT---GCTCCCGGCTCCGAAACACCGCTCACTGGTCGACTGCCTCGGAGGGACGAGCCAAAT---TCGTCTTCGTCTGGAGCCACAGCTAACAGACCCCAGCCGTACGAACTCATAAATTCAAACATCGAGTCATCTCTCCTCCAGCAACAACAACCAAACACATCTAATGCTTCTAACGTCGCCAATCGAAAAAGTGGCAATTCCAACCAGACTGCTTTCTCCAAGACCCATTCTTCCCTTGGTTATAGGTCACTTCGCCTCCCAGCAGATACCACTGCGGAGTTGAGAGCTCTAGCTAGCGGT--------------------------GGTGACACGCAGTCATCG---TACTACAATACA------------------------TCAGGTGCCAGTTTTGCTGCGCACCACCAGCCGACTATAGCCGGTGTCTCCCCG------------GCACCAGTGAGCCACTCGCCCTCTGTGTCCTCGTCCACTCAGGAG------	---MKEKDARVKFRQIVSAVQYCHAKNIVHRDLKAENLLLDESMNLKIADFGFSNNYSSGRKLDTFCGSPPYAAPELFLGRKYDGPEVDVWSLGVILYTLVSGSLPFDGKNLKELRECVLRGSYRVPFYMSHECEMLLRKMLVLNPTKRASLLDIMKDKWLNTAFEDNTLQPFKEDIPDYNDPERIQWMVQMGFSRSDIHDSLTKQRFNNITATYILLGQRRQKSLPWPPTLTGSSVVPQRSLSSEVPAP-----V-----NGRQ-NP-----------------------------ATTGSSQ----PPQSRQLTPTTTTTNAASFR----RPYHPPPSSSSNTSAQVKVMSTALTAP----GNMEGDASLANAPLATTDIDDVN---N------SPAN----TTTIIGTLKRQQARTEPHQPIT---ISPVPSPSP-DSSTGTVKPSRNQSALSSTST-------AVESSTTTGS--TVQDSRSEHADSRGGATSTQTPNRGAPTVPKRPSFTHRNITAGIE----------GEFHVPILAPRASDAA-APGSETPLTGRLPRRDEPN-SSSSGATANRPQPYELINSNIESSLLQQQQPNTSNASNVANRKSGNSNQTAFSKTHSSLGYRSLRLPADTTAELRALASG---------GDTQSS-YYNT--------SGASFAAHHQPTIAGVSP----APVSHSPSVSSSTQE--GESRHSTAASAV---NSGFYVPFGRNVPGRSTIQHVPASETRDRIALERPDQWSRPLVRHPGTAAHMDQKTTLADLYSQTVNQVRSRKISWPITMILAFSLLV--SQSQGRRSPSDLSVISVTSSSATAKAESN

*Taenia solium**	TsM_000072000	--	---
		--	--MRS---
		-----------------------ATGCGGTCT---	---
		--	---
		--	---
		--	-----------------------------------VF----------------------
		--GTATTT---	------------------
		---	QAENLLLDESMNLKIADFGFSNNYSAGRKL
		CAGGCAGAGAATCTGTTGTTGGACGAGTCGATGAACCTGAAGATAGCGGACTTCGGCTTTTCCAACAACTACT	DTFCGSPPYAAPELFLGRKYDGPEVDVWS
		CAGCTGGTCGGAAATTGGACACCTTCTGCGGGTCGCCACCATACGCGGCGCCGGAGCTCTTCCTTGGCCGC	LGVILYTLVSGSLPFDGKNLKELRECVLRG
		AAGTACGACGGCCCAGAGGTGGATGTCTGGTCTCTCGGGGTCATCCTCTACACCCTCGTCTCTGGCTCACTC	SYRVPFYMSHECEMLLRKMLVLNPTKRAS
		CCTTTTGACGGCAAAAATCTCAAGGAACTGCGGGAGTGTGTACTTCGAGGCTCCTATCGTGTGCCCTTCTACA	LIEIMKDKWLNTTFEDNILQPFREDLPNYND
		TGTCTCACGAATGCGAGATGCTACTCAGGAAGATGCTCGTTCTCAATCCCACCAAACGCGCCTCGCTGATAGA	PERIQWMVQMGFSRSDIHDSLTKQRFNNI
		GATTATGAAGGATAAGTGGTTAAATACAACTTTCGAGGACAACATTCTGCAGCCCTTCAGGGAGGATCTCCCC	TATYILLGQRKRKSLPWPPTLAGV--
		AATTACAACGATCCGGAGAGAATTCAATGGATGGTACAGATGGGCTTCTCACGGAGTGATATTCACGACTCTC	MQPRNLPSDMPAADGSVSSSAASGSGRQ
		TGACGAAGCAGCGGTTCAATAACATCACTGCTACGTACATTTTGTTGGGTCAACGAAAACGAAAGTCGCTACC	-PTSQARQ-
		TTGGCCTCCCACCCTGGCTGGTGTC------	LTSTTTSTAAASSFRRPFHNPSSTA---------
		ATGCAACCTCGAAATCTACCATCGGATATGCCTGCGGCTGATGGGAGTGTCTCATCATCGGCAGCCTCAGGA	DATSRKHSNAEINSN---
		AGCGGTCGACAA---CCAACCTCACAGGCTCGTCAA---	YTSDGHRKTSISGASTTG-GGQSKLTSV--
		CTCACCTCCACCACAACTTCTACCGCCGCCGCCTCCTCCTTCCGCCGTCCCTTCCACAATCCCAGTTCAACGG	SMA----
		CC------------------------------GACGCCACCAGTCGCAAGCACTCCAACGCTGAGATCAACAGCAAC---------	DNAEGDASIVNAPLAITDNDDVNNVSVPS-
		TACACTTCTGATGGGCATAGAAAAACTAGTATTTCCGGTGCATCCACCACAGGA---	TTVSTTA----TTA--
		GGGGGTCAGAGCAAGCTGACATCAGTC------TCGATGGCG------------	GTLKRQQTRPESQQPITN--TSPVPSPPP-
		GACAATGCAGAAGGAGATGCGAGCATTGTAAATGCGCCTCTGGCGATAACAGATAATGACGATGTGAATAATG	PESSSTSKTLKNQFVTLSAS--------
		TCTCTGTCCCGTCT---ACCACTGTGAGCACCACTGCC------------ACCACTGCA------	GNPPGAVNPIVGTFKDSRNGHTNSRSSAT
		GGAACTTTGAAGCGGCAACAAACGCGACCTGAGTCCCAACAACCAATAACCAAC------	SAQTSNRGVTAASKLPSFTHRSILTGTE----
		ACCTCTCCCGTCCCTTCTCCGCCTCCT---	------DDFNVPILAPRASDATVA-
		CCCGAGTCGTCCTCCACTTCAAAAACGCTGAAAAACCAGTTCGTGACTTTATCCGCTTCT------------------------	GAETSLGGRVSKWDDVTSNPSTIVTTTRA
		GGCAACCCTCCCGGTGCTGTCAACCCTATTGTAGGCACCTTCAAAGATTCGAGAAATGGTCATACAAACTCCC	QPYELINPAALESSLQ-------NPSNVTV-
		GTTCCAGTGCCACTTCAGCTCAGACATCAAACCGAGGCGTCACGGCCGCTTCCAAGCTCCCATCCTTCACTCA	ISSNTKQSLFSRKNPSLGYRSLRLPADAAS
		CAGAAGCATCCTTACTGGCACCGAA------------------------------	ELRAAAVAATLSGTATMVEAHPTTYYNA---
		GATGACTTTAACGTGCCGATTTTGGCACCGCGAGCCTCCGACGCTACGGTGGCT---	-----GGV-
		GGAGCAGAGACATCGTTGGGCGGGCGTGTGTCGAAGTGGGACGATGTGACGTCCAATCCTTCGACTATCGT	YVSHLHQPGGGTGSPGYTVTKSASPTQEL
		GACGACGACTAGAGCGCAACCCTACGAGCTAATCAACCCTGCGGCACTCGAGTCCTCTTTGCAG-----------------	-TMAAGGETKRS------
		----AACCCCTCCAACGTGACAGTA---	AGGGGGFSVPFSRNVPGRSTIQHVPASET
		ATCAGCAGCAACACGAAGCAGTCACTGTTTTCGAGGAAAAACCCCTCACTTGGATATCGGTCTCTCCGGTTGC	RDRLALERADQWSRPLVRHPGTAAHMDQ
		CAGCAGATGCGGCGTCGGAACTAAGAGCTGCAGCAGTGGCAGCAACGCTGTCAGGAACAGCGACGATGGTT	KTTLADLYNQN----------------------
		GAAGCGCACCCCACTACATACTACAACGCG------------------------GGTGGTGTC---	LIQTQTTGRQSPSDLSEISVTSNSTTAKAP

Atrial natriuretic peptide receptor 1	*Biomphalaria glabrata*	gi|908451201| ref|XP_013082359.1| PREDICTED: atrial natriuretic peptide receptor 1-like, partial [Biomphalaria glabrata]	---AACAGACAAAATCAGAACGGT--- ------------------------------------ CACGATGTTTTGGAGAATGAAGAAATCAAATTAGACTGGATGTTCAGATATTCCATAATGCAAGACATTGTCAG AGGAATGGCCTATTTGCATAGCACTGAGATCAAAAGCCACGGTCACTTAAAGTCCAGCAACTGCGTGGTGGAC AGTAGGTTCGTGGTCAAGATCACGGACTTTGGGCTGCATTACTTCAGA---------GAAAAGGAAGAAGAG--- CTGGAAGAAAACCTCTACGCC---AAACAC--- AGAAGTCAATTGTGGACAGCGCCCGAGCTTCTTAGGATGAGCAATCCTCCAGCAGGCGGTACTCAGAAA------- --------GCAGACGTGTACAGTTTCGCCATCATCTGTCAGGAGATAGTCTAC--- AGATCTGGCGTGTTTCACTTACGAAACATTGACCTTTCA--------------- CCCAACGAAATCCTTGATAAACTTAAGACTGGGGTAAAGCCCTACTTCCGT------CCAACGCTGGAAGAA--------- TTCGATTGCCCCAAC---------------GAT---------------------------------- GAGCTGGCTGTCACCATCAGAGTGTGTTGGTCTGAAGACCCAGCAGAGAGGCCGGACTTCCAACAGTTAAGG ACAAGCATTAAAAAGCTGAATAAGGATGGGGACAAAGGAGATATTCTGGATAATCTGCTTTCTCGAATGGAGC AGTACGCTAACAACTTGGAGGCACTAGTAGAGGAAAGAACATCT--- GACTACCTCCAAGAAAAGAAGAAAGCCGAGGAACTCTTATACAATATGTTGCCCAAGTTTGTGGCTTCCCAAC TAATACGTGGTGAGACAGTCAGTGCTGAATGGTACGATGGTGTCACCATATACTTCAGTGACATCTGTGGGTT CACTGCCATGTCTGCAGAAAGCTCGCCTATGCAGGTCGTTGACCTTCTAAATGATCTGTACACATGTTTTGATT CCATCCTTGAAAGCTTTGATGTCTACAAAGTGGAAACAATAGGAGATGCCTACATGGTTGTCTCTGGCTTACCT GTCAGGAATGGGAACCTTCATGCCAGGGAGATAGCCAGGATGTCATTGGCCTTGCTTCATGCTGTGTTCAAGT TTAAGATACGACACAGGCCCGGAGACCAGCTGAAACTCAGGATAGGCATGCACAGTGGTCCAGTTTGTGCTG GTGTGGTTGGTTTGAAGATGCCTCGATACTGCCTCTTTGGAGACACTGTGAATACGTCCTCAAGAATGGAGTC TAATGGCTTGCCTCTACGCATTCATGTGAGTCCAGCCACTAAAGAAATTTTGGAAACTTTTGGAACATTCCAAC TAGAGATGAGAGGAGCTGTTGAAATGAAGGGTAAAGGTGCCATCACGACTTACTGGCTCCTTGGC--------- GAGAAAGATTCCCCG	-NRQNQNG--- ---- HDVLENEEIKLDWMFRYSIMQDIVRGMAYL HSTEIKSHGHLKSSNCVVDSRFVVKITDFG LHYFR---EKEEE-LEENLYA-KH- RSQLWTAPELLRMSNPPAGGTQK----- ADVYSFAIICQEIVY-RSGVFHLRNIDLS----- PNEILDKLKTGVKPYFR--PTLEE---FDCPN-- ---D----------- ELAVTIRVCWSEDPAERPDFQQLRTSIKKL NKDGDKGDILDNLLSRMEQYANNLEALVE ERTS- DYLQEKKKAEELLYNMLPKFVASQLIRGET VSAEWYDGVTIYFSDICGFTAMSAESSPM QVVDLLNDLYTCFDSILESFDVYKVETIGDA YMVVSGLPVRNGNLHAREIARMSLALLHA VFKFKIRHRPGDQLKLRIGMHSGPVCAGV VGLKMPRYCLFGDTVNTSSRMESNGLPLR IHVSPATKEILETFGTFQLEMRGAVEMKGK GAITTYWLLG---EKDSP

Capitella teleta	gi|443723624|gb|ELU11951.1| hypothetical protein CAPTEDRAFT_165186 [Capitella teleta]	GTGCGTTCCGTTTGCAAGACACAGTTTACGCTGACAAAACAAGTACGCCAAGAAGTGAAAACTTTGAGGTCCATTGACCATCATAATGTCTGCAAATTTGTCGCTGCGTGTTTGGATCCAGAGAAATTCTGCATCATGATGGAGTATTGCCCTAAGGGCAGTCTGGCTGATGTTTTGCAGAATCCAGACGTTCCTCTGAATTGGGGATTTAGATTTTCGATGGCGAGTGATGTTGCCCGTGGTATGATCCAACTCCACACCCATCACATC---ATCCACGGTAGACTAAGCTCGAACAATTGCGTCATTGACGACAGGTGGACCGTCAAAATCACAGATCTGGATGGCAAAAGC---AGA---------------CAGGATGAGAGA---GACGACGCGTTTCACAAGGAGCGCCTCATGCAG------GTGTATAAACCACCCGAGTGCTACGAGAAAGGCTAC------ACGATCGGC---CCAGAG---------------GCAGACTCGTACGCCTTTGGCATTATACTAGTGGAACTTGCAACT---CGAAATGATGCATATGGGGTACAC------GATGAAGAT---------------ACGTACGACCTGTCAGAGACCTGGAAGCCTGACCTACCCGAACTGGAGGAT------GAAGTTGACAAAGAC---------ATGAAATGTCCCAGTCCT------------GTA---------------------------------CAGTATAACCAGTTGATTGCACAGTGCATTAGCGACAATGCCCACACCAGACCTACGTTCGAAGTCATCAAAAGGATGATCACAAAAATGAACCCAAGC------ATACAAAGTCCTGTCGATCTCATGATGAACATGATGGAGAAGTACTCGAAACATCTGGAAGCAGTTGTTGGAGAAAGAACAGCC---GACTTGGTGGTGGAGAAACAGAAAACAGATCGACTTTTATACAGCATGCTTCCGAAGCCCGTTGCCGACGATTTAAGAGTTGGAAAAACCATCGCTTGCGAGCAGTTCGACGTCTGCACAATTTACTTCAGTGACATCGTTGGATTCACGGTCATCTCCAGCAAGAGCACACCTTTTGAGATCGTAGGGTTGTTGAACAAGTTGTACACTACTTTCGATTCTATCATCGAGAAATACGATGTGTACAAAGTGGAAACCATTGGAGATGCTTACATGGTCGTATCTGGCGTTCCTCAACGCAATGGGGATAGGCATGCATCAGAAACTGCCGGCATGGCTGTCGATCTCGTCGCAGCATCGGAAGTCTTCGTCATCCCCCACATGCCCAAGGAACCGCTCAAGATCCGCGTCGGCATGCACAGCGGACCGGTTTGCGCTGGTGTGGTGGGACTGAAGATGCCGCGGTACTGCCTCTTTGGAGACACGGTGAACACTGCGTCGCGCATGGAATCCAATGGAGAAGCATACAGAATTCACATGAGTAATCCGACGTATGAGGTGTTGAAAAAGTGTGGCGGTTTCAAGATGGAGGAAAGAGGAGTCATTCCAGTCAAGGGAAAGGGGGACATGAGGACTTGGTGGTTGACGGGA---------AGAATGGAGTCACAG	VRSVCKTQFTLTKQVRQEVKTLRSIDHHNVCKFVAACLDPEKFCIMMEYCPKGSLADVLQNPDVPLNWGFRFSMASDVARGMIQLHTHHI-IHGRLSSNNCVIDDRWTVKITDLDGKS-R-----QDER-DDAFHKERLMQ--VYKPPECYEKGY--TIG-PE-----ADSYAFGIILVELAT-RNDAYGVH--DED-----TYDLSETWKPDLPELED--EVDKD---MKCPSP----V-----------QYNQLIAQCISDNAHTRPTFEVIKRMITKMNPS--IQSPVDLMMNMMEKYSKHLEAVVGERTA-DLVVEKQKTDRLLYSMLPKPVADDLRVGKTIACEQFDVCTIYFSDIVGFTVISSKSTPFEIVGLLNKLYTTFDSIIEKYDVYKVETIGDAYMVVSGVPQRNGDRHASETAGMAVDLVAASEVFVIPHMPKEPLKIRVGMHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESNGEAYRIHMSNPTYEVLKKCGGFKMEERGVIPVKGKGDMRTWWLTG---RMESQ

Crassostrea gigas	gi|762098840| ref|XP_011431822.1| PREDICTED: atrial natriuretic peptide receptor 1-like [Crassostrea gigas]	GTTAAACGCATCAATAAATACAACTTTGGACTGTCTAAATCTCTCCGAATAGAAGTTAAGGAAATTAGAGAACTGAGGCACCCAAACCTGTGTCAGTTTGTGGGAGCGTGTACAGAGACCCCCAATGTGTGTATCCTGATGGAGTATTGTCCAAAGGGGGCGCTGGCAGACGTCCTCCTGAATGACGACATCCCACTCACGTGGTCCTTCCGGTTCTCATTTGCCGCTGACATCGCAAACGGAATGGACTACCTCCACAGCCATGGACTC---GTCCACGCTCGACTGAACTCCAGTAATTGCGTGGTGGACGATCGATGGTCGGTGAAAATTACAGATTACGGACTTCCAATCCTGCGCAAAAACGATTTTAAATCAGAGGAAATGACATCAGAT---TTTCAAAGCAGACGA------CGAGTT---GTGTACAATGCTCCAGAGGTTTGCGGT---TCGTTT------CCAGTGTTCACAAAGTCC---------------TCTGATGTCTACTCTTATGGGATTATTTTGGTCGAAATAGCTAAC---AGAAGCGATCCATACGGG------------GATGAAGAC---------------CCAGCTTTTTTACCGCCTCAGTGGAAGCCGCCACTACCGAATCTAAAGAGG------GACAACGAAGACGAA---------AAC---TGCCCCTCCCCC------------ACG--------------------------------GCTCTCTGTGCTCTGATTGACGAATGTCTGGACTTCAGATCACAAGAAAGACCAACGTTCGTCAACATCAGAAAGATTCTGTATAAAATCAACCCCAAT------AAACAGAACCCTGTCGACCTGATGATGGCGATGATGGAGAAGTACTCGAAACACTTGGAACAGATTGTGACCGAGCGAACTAAT---GATTTAACGATAGAAAAACAGAGGACGGATCGATTGCTATACAGTATGCTTCCCAAAGAGGTAGCAGACGTTCTGAGGCGGGGACGCCCGGTAGAGGCGCGGTACCTCGATGACGTCACGATCTATTTCAGTGACATCGTCGGATTCACCACTCTCTGCTCCAACAGTAGTGCTATGGAGGTTGTGAACCTCCTCAACAAACTATACATCACTTTTGACGAAGTCATTGAACTGTATCATGTCTACAAAGTGGAAACCATTGGGGATGCATACATGGTAGCCTCGGGTGTTCCCGAAGCTTACCCCACC---CACGCAATAGAAGTCGCCCGTATGGCCATCAGTCTAGTCAACAAGTGTAAATCGTTTGTGATTCCCCATTTTCCGGATCAAAAACTGAAGATAAGAGTCGGCATTCACTCGGGCCCTGTGTGTGCTGGAGTGGTCGGGTCCAAAATGCCCCGATACTGTTTGTTTGGAGATACCGTCAACACTGCCTCAAGGATGGAGTCCAATGGGGAAGCTTATAAAATTCACATAAGTGCTAATACTTACGACTTGCTACAAACG---GGGACATTTCAGTTTGAGGCGCGGGATAAGATTTCAGTCAAAGGTAAAGGAGAGATGCAGACCTACTGGCTCCTTAAA---------GAAAGAGAAAGCCCG	VKRINKYNFGLSKSLRIEVKEIRELRHPNLCQFVGACTETPNVCILMEYCPKGALADVLLNDDIPLTWSFRFSFAADIANGMDYLHSHGL-VHARLNSSNCVVDDRWSVKITDYGLPILRKNDFKSEEMTSD-FQSRR--RV-VYNAPEVCG-SF--PVFTKS-----SDVYSYGIILVEIAN-RSDPYG----DED-----PAFLPPQWKPPLPNLKR--DNEDE---N-CPSP----T-----------ALCALIDECLDFRSQERPTFVNIRKILYKINPN--KQNPVDLMMAMMEKYSKHLEQIVTERTN-DLTIEKQRTDRLLYSMLPKEVADVLRRGRPVEARYLDDVTIYFSDIVGFTTLCSNSSAMEVVNLLNKLYITFDEVIELYHVYKVETIGDAYMVASGVPEAYPT-HAIEVARMAISLVNKCKSFVIPHFPDQKLKIRVGIHSGPVCAGVVGSKMPRYCLFGDTVNTASRMESNGEAYKIHISANTYDLLQT-GTFQFEARDKISVKGKGEMQTYWLLK---ERESP

Echinococcus granulosus	gi|674561972|emb|CDS23693.1| atrial natriuretic peptide receptor 1 [Echinococcus granulosus]	GTAAAATACGTTGAGCGTGAACAATTTCTTCTGACAAAGAACATTAGAAAGGAGATCAAGGCAATGCGGCAACTGAGTCATAGGAATTTGTGTCAACTAGTTGGAATTTGTCTGGAACCTCCTGAACTTGCAATCTACATGGAATATTGCCCAAAACGAAGTCTTCGGGATGTGTGTCACAACGAGGTGATGCCTTCCAGTTGGGCCTTCAAACTCTCTTTGATTCAAGATATCATCTGCGGCGTTGAATTTCTTCATGCTCATGGCTTC---ATTCATGGACGTCTGAATTCACAAAACTGCGTTGTCGATGACCGTTTGACTTGCAAGATCACTGATTTTGGGCTGGAATCCATCCGC------TATAACAAACCGGAGGAAAAGCTG---GAAACCTTTCTAGAGGATCCC------AGAAATTGGGCATTTATTGCACCGGAGTATCGGGGTAAT---------------------GCACCTGCTCCGCCGAATATTCACATGGACTCGTTCAGCTATGGAACGATTATGTGTGAGGTAGCCCAGTCCCGTGAAGACCCA------------------GAAGAT------------------TTCAACGCG------------------CCTGACCTGACAGAAGGCGAAAGA---CAGGAACTGATGGTCCAATGGTGGGACTACCCTCTACCAAAT---GTGGATGCATTTGAAGGTTGCGCTGATGACAGCACCCCAAATATGACTGAATACATTAACCTTATCAGATTGTGCTGGAGTCCAGTT---GATGTTAGACCCGCTTTTGACGTCATCAGAGCAAAGATGGACCTTATCAACCCAAGA------AGGAAAAACCCAATTGACATTATTCTCAGTCTTATGGAAAAATATTCGGCACATCTTGAAAGCATTGTCAGTGAAAGGACACAA---GACCTCATCGCAGAGAAACAACGCACAGACGAATTGCTGCACAGTATGCTTCCAAAGACAATCGCAAATCAACTGCGCAGTGGGCAGGCGGTTCCAGCCGAGGCCTACTCCTCCTGTACCATATACTTTAGCGACATTGTTGGCTTTACCAACATCTCCTCGGATTCAACGCCCTTTCAGGTTGTTGCACTGCTTAACAAACTCTACAGCGAGTTTGACCAAATCATTGACCGATATGACGTCTACAAAGTGGAGACCATCGGAGACGCTTACATGGTAGCCTCCGGAGTTCCGAGAAGGAATGGTCAACGGCATGCAGTGTCTATAACGGATATGGCACTGGATTTGGTCGAGGTCTCGCACTCCTTTATCATCCCCCATATGCCAAATGAACCACTTAAAATTCGAGTCGGCTTACATTCAGGACCTGTTTGTGCGGGTGTTGTTGGCTTGAAAATGCCGCGGTACTGCTTGTTTGGGGACACTGTCAACACAGCAAGTCGGATGGAGAGCAACGGAGAGGCGTACAAAATTCATTGCAGTGATGCCACACATGACATACTAAGTACACTTGGTGGCTTTCACTTCGAGGAAAGGGGAACAATTGAGGTCAAGGGTAAGGGAACAATGCGCACATGGTGGGTGACAGGT---------CGAACTCGGCCACCG	VKYVEREQFLLTKNIRKEIKAMRQLSHRNLCQLVGICLEPPELAIYMEYCPKRSLRDVCHNEVMPSSWAFKLSLIQDIICGVEFLHAHGF-IHGRLNSQNCVVDDRLTCKITDFGLESIR--YNKPEEKL-ETFLEDP--RNWAFIAPEYRGN-------APAPPNIHMDSFSYGTIMCEVAQSREDP------ED------FNA------PDLTEGER-QELMVQWWDYPLPN-VDAFEGCADDSTPNMTEYINLIRLCWSPV-DVRPAFDVIRAKMDLINPR--RKNPIDIILSLMEKYSAHLESIVSERTQ-DLIAEKQRTDELLHSMLPKTIANQLRSGQAVPAEAYSSCTIYFSDIVGFTNISSDSTPFQVVALLNKLYSEFDQIIDRYDVYKVETIGDAYMVASGVPRRNGQRHAVSITDMALDLVEVSHSFIIPHMPNEPLKIRVGLHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESNGEAYKIHCSDATHDILSTLGGFHFEERGTIEVKGKGTMRTWWVTG---RTRPP

Echinococcus multilocularis	gi|674572779| emb|CDS41614.1| atrial natriuretic peptide receptor 1 [Echinococcus multilocularis]	GTAAAATACGTTGAGCGTGAACAATTTCTTTTGACAAAGAACATTAGAAAGGAGATCAAGGCAATGCGGCAACTGAGTCATAGGAATTTGTGTCAACTAGTTGGAATTTGTCTGGAACCTCCTGAACTTGCAATCTTCATGGAATATTGCCCAAAACGAAGTCTTCGGGATGTGTGTCACAACGAGGTGATGCCTTCCAGTTGGGCCTTCAAACTCTCTTTGATTCAAGATATCATCTGCGGCGTTGAATTTCTTCATGCTCATGGCTTC---ATTCATGGACGTCTGAATTCACAAAACTGCGTTGTCGATGACCGCTTGACTTGCAAGATCACTGATTTTGGGCTGGAATCCATCCGC------TATAACAAACCGGAGGAAAAGCTG---GAAACATTTCTAGAGGATCCC------AGAAATTGGGCATTTATTGCACCGGAGTATCGGGGTAAT---------------------GCACCTGCTCCGCCGAATATTCACATGGACTCGTTCAGCTATGGAACGATTATGTGTGAGGTAGCCCAGTCCCGTGAGGACCCA------------------GAAGAT------------------TTCAACGCA------------------CCTGACCTGACAGAAGGCGAAAGG---CAGGAACTGATGGTCCAATGGTGGGACTACCCTCTACCAAAT---GTGGATGCATTTGAAGGTTGCGCTGATGACAGCACCCCAAATATGACTGAATACATCAACCTTATCAGATTGTGCTGGAGTCCAGTT---GATGTTAGACCCGCTTTTGACGTCATCAGAGCAAAGATGGACCTTATCAACCCAAGA------AGGAAAAACCCAATTGACATTATTCTCAGTCTCATGGAAAAATATTCGGCACATCTTGAAAGCATTGTCAGTGAAAGGACACAA---GACCTCATCGCAGAGAAACAACGCACAGACGAATTGCTGCACAGTATGCTTCCAAAGACAATCGCAAATCAACTGCGCAGTGGACAGGCGGTTCCAGCCGAGGCCTACTCCTCCTGTACCATCTACTTTAGTGACATTGTTGGCTTTACCAACATCTCCTCGGACTCAACACCCTTTCAGGTTGTTGCACTGCTTAACAAACTCTACAGCGAGTTTGACCAAATCATTGACCGATATGACGTCTACAAAGTGGAGACCATCGGAGACGCTTACATGGTAGCCTCTGGAGTTCCGAGAAGAAATGGCCAACGGCATGCAGTGTCTATAACGGATATGGCACTGGATTTGGTCGAGGTCTCGCACTCCTTTATCATCCCCCATATGCCAAATGAACCACTTAAAATTCGAGTCGGCTTGCATTCAGGACCTGTTTGTGCGGGTGTTGTTGGCTTGAAAATGCCGCGGTACTGCTTGTTTGGGGACACTGTCAACACAGCAAGTCGGATGGAGAGCAACGGAGAGGCGTACAAAATTCATTGCAGTGATGCCACACATGACATACTAAGTACACTTGGTGGCTTTCACTTCGAGGAAAGGGGAACAATTGAGGTCAAGGGTAAGGGAACAATGCGCACATGGTGGGTGACAGGT---------CGAACTCGGCCACCG	VKYVEREQFLLTKNIRKEIKAMRQLSHRNLCQLVGICLEPPELAIFMEYCPKRSLRDVCHNEVMPSSWAFKLSLIQDIICGVEFLHAHGF-IHGRLNSQNCVVDDRLTCKITDFGLESIR--YNKPEEKL-ETFLEDP--RNWAFIAPEYRGN-------APAPPNIHMDSFSYGTIMCEVAQSREDP------ED------FNA------PDLTEGER-QELMVQWWDYPLPN-VDAFEGCADDSTPNMTEYINLIRLCWSPV-DVRPAFDVIRAKMDLINPR--RKNPIDIILSLMEKYSAHLESIVSERTQ-DLIAEKQRTDELLHSMLPKTIANQLRSGQAVPAEAYSSCTIYFSDIVGFTNISSDSTPFQVVALLNKLYSEFDQIIDRYDVYKVETIGDAYMVASGVPRRNGQRHAVSITDMALDLVEVSHSFIIPHMPNEPLKIRVGLHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESNGEAYKIHCSDATHDILSTLGGFHFEERGTIEVKGKGTMRTWWVTG---RTRPP

Hymenolepis microstoma	gi|674593514| emb|CDS27743.1| atrial natriuretic peptide receptor 1 [Hymenolepis microstoma]	GTAAAGTACGTTGAAAGAGAGCAGTTCCTGCTCACTAAGAATATTCGTAAGGAAATTAAAGCCATTCGTCAACTAAATCACCGAAATCTTTGCCAACTAGTTGGAATTTGTCTCGATCCTCCTGAAATGGCCATCTACATGGAATATTGTCCAAAAAGAAGCCTCAAAGATGTATTCCGCAATGAGGTTATGCCTTTGAGCTGGGCATTCAAATTATCTCTAATTCAAGACATAGTCTCTGGCATGGAATACATGCACTCCCATGGGTTC---ATTCATGGACGCTTAAACACCCAAAACTGTGTGGTTGACGACCGTCTAACTTGCAAAATTGCTGACTTTGGATTAGAATCGATTCGA------TATGATAGACCGCAAGAAAAACTT---GAGACATTTTTGGAGAACCCA------CAGAATTGGGCATTCGTTGCTCCGGAATATCGAGGCGAT--------------------AATCCTGCTGCACCGAATCCCCACATGGACTCGTACAGCTACGGAACAATTATGTGTGAGGTTGCACAGCTTCGCGAAGAGCCA------------------GAAGAC------------------TTCACTGAC------------------CCCGAGTTGACGGAGATGGAGCGG---CAAGAGTTGAAAATGCTGTGGTGGGAGTATCCCTTACCCACT---CTAGAAGCTTTCGAAAGTAACGCAGATGACACTTCTCCTAACATGACTGAATACCTTAATTTAATCAAACTGTGTTGGAGCCCCGTG---GATACGAGACCTGGTTTTGACGGCATCAGAGGAAAGATGGATCTCATCAACCCCAGG------CGAAAGAACCCAATAGATATTATTCTCAGTCTAATGGAGAAGTATTCCGCTCACCTTGAGAGTATTGTGAGCGAAAGAACTCAA---GATCTCATTGTAGAAAAGCAACGGACTGATGAACTTCTTCACAGCATGCTCCCGAAAACTATCGCAAATCAGCTGAGAAGTGGTCAGTCAGTCCCCGCCGAAGCCTATTCCTCCTGTACCATCTACTTCAGTGATATTGTCGGATTTACAAATATTTCTTCAGATTCAACACCCTTTCAGGTTGTGGCACTCCTCAACAAACTCTACAGTGAATTTGATCAAATTATCGATCGATATGATGTCTACAAAGTCGAAACTATTGGTGATGCCTACATGGTGGCGTCAGGTGTTCCTAGGCGAAATGGTCAACGACATGCCGTCTCTGTCACAGATATGGCTCTGGATTTGGTCGAAGTCTCTCATTCGTTTGTGATTCCTCATATGCCCAATGAACCTCTCAAAATTCGCGTTGGTATACACTCAGGTCCTGTATGTGCTGGAGTGGTGGGTTTGAAGATGCCTAGATACTGTCTCTTTGGCGATACAGTCAACACTGCAAGCAGAATGGAGAGCAATGGAGAAGCCTATAAGATCCACTGCAGCGACGCCACTCATGGCATCCTAAAAAATCTCGGTGGTTTTCTGTTTGAGGAAAGAGGGACAATCGAGGTTAAGGGCAAGGGAACGATGCGTACCTGGTGGGTGACTGGT---------CGAACGCGTCCACCC	VKYVEREQFLLTKNIRKEIKAIRQLNHRNLCQLVGICLDPPEMAIYMEYCPKRSLKDVFRNEVMPLSWAFKLSLIQDIVSGMEYMHSHGF-IHGRLNTQNCVVDDRLTCKIADFGLESIR--YDRPQEKL-ETFLENP--QNWAFVAPEYRGD-------NPAAPNPHMDSYSYGTIMCEVAQLREEP------ED------FTD------PELTEMER-QELKMLWWEYPLPT-LEAFESNADDTSPNMTEYLNLIKLCWSPV-DTRPGFDGIRGKMDLINPR--RKNPIDIILSLMEKYSAHLESIVSERTQ-DLIVEKQRTDELLHSMLPKTIANQLRSGQSVPAEAYSSCTIYFSDIVGFTNISSDSTPFQVVALLNKLYSEFDQIIDRYDVYKVETIGDAYMVASGVPRRNGQRHAVSVTDMALDLVEVSHSFVIPHMPNEPLKIRVGIHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESNGEAYKIHCSDATHGILKNLGGFLFEERGTIEVKGKGTMRTWWVTG---RTRPP

<table>
<tr><td><i>Lingula anatina</i></td><td>gi|919029633| ref|XP_013398539.1| PREDICTED: atrial natriuretic peptide receptor 1-like [Lingula anatina]</td><td>ATAAAAAGAATAGAGAAGAGGTATTTCAGCTTAACCAAAGTTATCAGGTTGGAAGTCTCACAAGTCAGACAGCTGGATCATGTTAACCTGGTAAAATTTATTGGAGGTTGTGTTGAGATCCCTACTGTGGCCATTATCACAGAGTACTGCCCTAAAGGAGGCCTTAATGATGTGCTGCAAAATGATGAGATCCCTCTCACCTGGGCTTTCAGATTTTCTTTTATACACGACATAGCTCGTGGTTTGCATTTTCTCCACTATAATAAAATA---ACCCATGGGCGTCTCAAGTCTCCAAACTGTGTGATAGATGACAGATGGACTGTGAAAATATCTGACTTTGGACTTGCTACATATCGT------------------GAAGAT---GTGGAAGAAGACAAGTACAGGAGTAAGGCCTGCAGA------GTGTATAGAGCCCCAGAGCTGACACACCTTCCTGCT------GCTCAACCAACACCAGAG--------------GCAGATGTTTATGCATTTGCCATCATTCTTGTAGAAATTGCCACA---AGGAATGGACCATATGGG------------GAGGAAGAT--------------ATTGATGACCTACCTGACCACTGGAAGCCCAGCCTACCAGATTTACAAAGTTCTGGTAAAACAAGTAAAGAA---------TACTCTTGTCCTTGTGGT------------GAC--------------------------------CAGTACATTCAGTTAATAAAGCGCTGCTGGAGTGACAATCCATTTGACAGACCAAACTTTGAACAAATTAAGCGACAAATCCACAAGATCAACCCCAAT------AAACAGAGCCCAGTGGATATGATGATGACCATGATGGAGAAATATTCAAAGCATCTGGAAGTGATGGTTGCAGAGAGAACTCAG---GATCTGATGGCAGAGAAGCAGAAAACAGACAGACTATTATACAGCATGTTACCTAAAAGTGTAGCAGATGCCCTGCGTCTTGGAAAACCAGTTCAAGCAGAGTCCTTTGAGTCCTGCACCATTTTCTTTAGTGACATTGTGGGTTTCACTGAGCTGGCAGGACACAGCACACCTCTGGAGGTGGTCACACTGCTCAATAAACTGTACACATGCTTCGATGAAATCATTGACAGATACAGTGTATATAAAGTGGAAACAATAGGGGATGCATATATGGTTGTATCAGGAGTTCCAATAAGAAATGCAAACCACCACTGTAAAGAAATAGCTAATCTGGCTATTGACTTGGTAAAGGAAAGTGAAATGTATGTTATACCTCACAAACCTTATGAGTCACTCAGGATAAGAGTGGGACTGCACTCAGGTCCAGTGTGTGCAGGAGTAGTTGGTTTGAAGATGCCCAGGTACTGTCTGTTTGGTGATACTGTGAACACTGCCTCCAGGATGGAGTCTACAGGGGAAGCTGGGAAAATCCACATAAGTGATACCACATACCAGCTGCTACAGGAGTATCAGGGGTTCTGCTGTCAATCCAGGGGGACCATCCCTATCAAGGGAAAAGGGGACATGAAAACCTGGTGGTTAGCAACA---------TTCCCGGCAAACAGA</td><td>IKRIEKRYFSLTKVIRLEVSQVRQLDHVNLVKFIGGCVEIPTVAIITEYCPKGGLNDVLQNDEIPLTWAFRFSFIHDIARGLHFLHYNKI-THGRLKSPNCVIDDRWTVKISDFGLATYR------ED-VEEDKYRSKACR--VYRAPELTHLPA--AQPTPE-----ADVYAFAIILVEIAT-RNGPYG----EED-----IDDLPDHWKPSLPDLQSSGKTSKE---YSCPCG----D----------QYIQLIKRCWSDNPFDRPNFEQIKRQIHKINPN--KQSPVDMMMTMMEKYSKHLEVMVAERTQ-DLMAEKQKTDRLLYSMLPKSVADALRLGKPVQAESFESCTIFFSDIVGFTELAGHSTPLEVVTLLNKLYTCFDEIIDRYSVYKVETIGDAYMVVSGVPIRNANHHCKEIANLAIDLVKESEMYVIPHKPYESLRIRVGLHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESTGEAGKIHISDTTYQLLQEYQGFCCQSRGTIPIKGKGDMKTWWLAT---FPANR</td></tr>
</table>

<table>
<tr><td>Lollita gigantea</td><td>gi|676439332| ref|XP_009048819.1| hypothetical protein LOTGIDRAFT_112412, partial [Lottia gigantea]</td><td>GTAAAGAAAATTCAGAAGAACGACTTCAAACTTTCTCTTGAGATTAGATCAGAAGTGAAAGCCGTAAGAGAAATGGATCACCCTAATTTATGTAAATTTGTTGGTGGATGTATAGATATTCCTGATGTTGCTATAGTAACAGAATATTGTCCTAAAGGTAGTTTAAACGACGTATTGCTGAATGATGAAATACCCCTCAACTGGGCTTTCAGGTTTTCACTGGCAAGCGACATAAGTCGTGGGATGAGTTACCTACACAGTCGGGGTATG---GTTCATGGTCGACTTACATCCAGTAATTGTGTAGTAGATGATAGGTGGACGGTAAAAGTTACTGATTTCGGTCTACCAACATATCGAGCAGTTGATGTAACATGTGATGAGGACAAACAAGAACAGGTTTATAAAGAAACAGCT---AGAGAT---GTGTATATTGCACCAGAAATCAGAAAG---GGTGTG------ACACGGTCGAGCTGTCCT---------------GGAGATGTATACTCATTTTCCGTCTTACTGGTGGAAATAGCAAAT---AGAAATGATCCATATGGG------------GATGAAGAT---------------AGAAGTGTCCTACCAGGGGATTGGAAACCACCTCTACCAGAAGAAGATGCA------ACTGTTGATAAAGAG---------AGCCGATGTCCTTGTCCG------------TTT---------------------------------GAATATTGCTCGTTAATAAAAGACTGTTGGAATAACGAACCAGAGGAAAGACCAACTTTTGACACCATCAAGAAAACTATATATAAAATAAACCCTAAT------AAACTCAGTGCTGTCGATCTCATAATGCATATGATGGAAAAATACTCGAAGCATTTAGAATCCATAGTTGTTGACAGAACACGT---GATCTGGTGGCTGAGAAACAAAAAACTGACAAATTATTATACAGTATGTTACCAAAACCAGTAGCGGATCAGCTGCGACAAGGTACCGTGGTAAACGCTGAATCGTTTGATGAATGTACCATCTACTTCAGTGATGTGGTAGGATTTACAACTCTCTCGGGTAAAAGTACTGCTATGCAAGTTATTGCCCTGCTAAATAAACTCTATACAACTTTCGATGAAATTATAGACCAGTTTGATGTGTATAAAGTTGAGACTATAGGTGACGCTTACATGGTTGTATCTGGTGTTCCGATAAAAACA---GAATTCCACGCCCGAGAGATCAGCAATATGTCTTTAGAAATAGTGGCAGCGTGTAAAAAATTTGTTATACCACATCTTCCTGACGAACTATTACAGATTCGTGTTGGCTTACATTCAGGTCCAGTATGTACTGGTGTAGTTGGGTTAAAGATGCCTAGATATTGTTTATTCGGTGATACTGTTAATACAGCTTCTAGAATGGAATCAAACGGTGCAGCTTACAGAATCCACATCAGTTCGTCAACACATGACCATCTAGAGATGATAGGAGGTTACATTTTCGAGTGTAGAGGAGCCATACCCATAAAGGGAAAGGGTGAAATGGTCACTTGGTGGCTTCTATCCAAAACACCAGAAATATCAAGTTCA</td><td>VKKIQKNDFKLSLEIRSEVKAVREMDHPNLCKFVGGCIDIPDVAIVTEYCPKGSLNDVLLNDEIPLNWAFRFSLASDISRGMSYLHSRGM-VHGRLTSSNCVVDDRWTVKVTDFGLPTYRAVDVTCDEDKQEQVYKETA-RD-VYIAPEIRK-GV--TRSSCP-----GDVYSFSVLLVEIAN-RNDPYG----DED-----RSVLPGDWKPPLPEEDA--TVDKE---SRCPCP----F-----------EYCSLIKDCWNNEPEERPTFDTIKKTIYKINPN--KLSAVDLIMHMMEKYSKHLESIVVDRTR-DLVAEKQKTDKLLYSMLPKPVADQLRQGTVVNAESFDECTIYFSDVVGFTTLSGKSTAMQVIALLNKLYTTFDEIIDQFDVYKVETIGDAYMVVSGVPIKT-EFHAREISNMSLEIVAACKKFVIPHLPDELLQIRVGLHSGPVCTGVVGLKMPRYCLFGDTVNTASRMESNGAAYRIHISSSTHDHLEMIGGYIFECRGAIPIKGKGEMVTWWLLSKTPEISSS</td></tr>
</table>

*Mesocestoides corti**	MCOS_0000821001-mRNA-1	ATCAAAAGCGTCGAACGTGACCAAGTTCTCCTGACCAAAAAACTTCGCAAGGAGATAAAGACGATGAGACAACTGAACCATCGCAATGTGTGTCAACTCGTGGGCATTTGCCTGGAACCGCCCGAAATAGTAATCTACATGGAATACTGCCCCAAGAGAAGCTTGCGAGACGTCTTGCGCAATGAGGTGATGCCAACGAGTTGGGCATTCAAGCTCTCGTTAATTCATGATATCGTATCTGGGGTCGAACACCTACACACACACGGCTTC---ATTCATGGCCGTCTCAACACACAAAATTGTGTTGTTGATGACCGTTTAACTTGCAAAATTACTGACTTTGGACTGGAGTGTCTTCGC------TACAATAAACCCGAAGACAAGTTG---GAGACGTTTTTGGACGACCCA------CGAAACTGGGCATTTATTGCGCCGGAGTATCGCGGTGAA--------------------ACCGTCACACCACCGCACATTCGGATGGATTCTTACAGTTATGGAACGATCATGTGCGAAGTCGCTCAACTTCGTGAA---CTTATAAAGTTCTGCTGGAGCCCGCCG---GAGACACGGCCTTCATTTGATGTGATTCGAATGAAAATGGAACTCATCAATCCCAGG------CGAAAAAATCCAATTGACATCATACTCAGCCTTGTAAGTCTATCACCACGCTTTAACTGTTTCCGACTGCAATTTGGATCGGCGGCTGATTTT---------AAA------ACT------------CATNNNATGCTTCCGAAGACTATCGCCAATCAGCTGCGAAGCGGTCAGGCGGTGCCAGCTGAAGCGTACTCCTCGTGCACCATTTACTTCAGCGATATCGTCGGCTTTACCAACATCTCATCCGATTCAACGCCCTTCCAGGTTGTTGCTCTGCTCAACAAACTCTACAGTGAATTTGACCAAATCATCGATCGGTATGCTGTCTATAAAGTGGAGACCATTGGAGATGCATACATGGTTGCCTCCGGTGTACCACGAAGGAATGGACAGCGTCATGCGGTTTCCATAACCGACATGGCCTTGGACTTGGTGGAGGTTTCGCATTCCTTCGTCATCCCCCACATGCCCAACGAGCCACACAAGATCCGAGTGGGTCGGCATTCAGGACCCGTTTGTGCAGGAGTTGTTGGCTTAAAAATGCCGCGGTACTGCTTGTTCGGCGACACCGTCAACACCGCGAGCCGAATGGAAAGCAACGGCGAAGCTTACAAAATTCACTGCAGTGATGCTACCCACAGCATTTTGGACCCTCTTGGCAGTTTCCTTTTTGAAGAAAGAGGCACAATTGAAGTCAAGGGAAAGGGAACGATGCGCACGTGGTGGGTAATA---------GGGCGCACTCGACCC	IKSVERDQVLLTKKLRKEIKTMRQLNHRNVCQLVGICLEPPEIVIYMEYCPKRSLRDVLRNEVMPTSWAFKLSLIHDIVSGVEHLHTHGF-IHGRLNTQNCVVDDRLTCKITDFGLECLR--YNKPEDKL-ETFLDDP--RNWAFIAPEYRGE-------TVTPPHIRMDSYSYGTIMCEVAQLRE---LIKFCWSPP-ETRPSFDVIRMKMELINPR--RKNPIDIILSLVSLSPRFNCFRLQFGSAADF---K--T----H?MLPKTIANQLRSGQAVPAEAYSSCTIYFSDIVGFTNISSDSTPFQVVALLNKLYSEFDQIIDRYAVYKVETIGDAYMVASGVPRRNGQRHAVSITDMALDLVEVSHSFVIPHMPNEPHKIRVGRHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESNGEAYKIHCSDATHSILDPLGSFLFEERGTIEVKGKGTMRTWWVIG---RTRPP

<table>
<tr><th></th><th></th><th></th><th></th></tr>
<tr><td>Schistosoma haematobium</td><td>gi|844842302| ref|XP_012793513.1| Atrial natriuretic peptide receptor 2 [Schistosoma haematobium]</td><td>GTTGAATTTACTGAG---------------TTAACAAGCGATATGCAAAGACAACTATGGGAAGTTAAGAAAATGAAACATAATAATTTAGTGAAGCTCATTGGGGTCACATTTATATCACCTGTCTTATCATTGTATACAGAGTTTTGTGATAGAGGAAGTTTGTGTTACGTACTTCGACGTGATTCTATCCCGCTAAGTTGGAGTTTAAGAATAGGTTTTTTAACTGGTCTTGCAAATGGTTTAGCTTATTTACATCATTGTCAGATT---GTACATGGTCGATTGAATTCATCAAATTGTGTG---CCTAGATTGATCATGACATTTAGCTTAATGTCTGATGTTAAT---GAT---TTTAATTCCATTGTT------CAGAATGCATCCAAT--------------------------TATAAAGAAACACTATTCGTGCCAGCA---------------GTAGACATATACAGCTTTGGTACAATCATGTGGGAAACGGCAAGT---CGGAGTGATCCG------------TCTCAAGATGACGAATTTGTGGCAGATCCAATGTATAATCAC---------------CCGGAATTGCCGACTAAAAGACGT---TTTGAAACTGATGTTAAA---------TACGAGGTTGCAACA---GTACCGCCTTTTGAA--------------------------------GAATACAACAATTTGATGGAATCTTGTTGGAGTGAAAAT---ACACTTAGACCAAACTTAAATGCAATTATAGAATGGTTGACAAAGATTAATCCAAAA---------AATATTGGAGTTGATGAAGATACAATACTTACGAAAGAATATGCTAAATGTCTTGAATCTATTATTGAAGATAGAACACAA---GCGCTTCGTAGTGAACAAAAAATGGCAGACACTTTACTTAACAGTATGTTACCAAAACAAGTTGTAGAAATGCTAAGACATGGTGAGAATGTACCGCCTGAAGCATTTGAACAATGTACAATATACTTTAGTGATATTGTCGGTTTTACCACAATTTCATCAAGTTCTACACCATTTGAAGTTGTAGAATTTCTAAACAAATTGTACACTCAATTCGATGATATCATTGATCGATACGACGTTTATAAAGTGGAAACAATTGGTGATGCATATATGGTTGCATCAGGTGTTCCAAGAAGAAATGGTGAACGTCATGCGATAGCAATAGCTGACATGTCACTTGATCTAGTCAGTGTTTCACACAGTTTCGTAATTCCTCACAAACCTGATGAACCGTTAAAAATTCGAGTTGGTTTACATTCAGGTCCAGTATGTGCTGGTGTTGTTGGTTTGAAAATGCCAAGATACTGTCTTTTTGGTGATACAGTCAACACAGCAAGTCGAATGGAAAGTACTGGCGAAGCTTACAAAATACACTGCTCAGAAACAACACACGCTATATTGGATCGACTTGGTGGTTTCACTTTTGAAAAACGTGGTACAATAACTGTGAAAGGAAAAGGGGATATGCAGACATGGTGGATCACAGGA---------CGTACTAGAGCTGAT</td><td>VEFTE-----LTSDMQRQLWEVKKMKHNNLVKLIGVTFISPVLSLYTEFCDRGSLCYVLRRDSIPLSWSLRIGFLTGLANGLAYLHHCQI-VHGRLNSSNCV--------------PRLIMTFSLMSDVN-D-FNSIV--QNASN---------YKETLFVPA-----VDIYSFGTIMWETAS-RSDP----SQDDEFVADPMYNH-----PELPTKRR-FETDVK---YEVAT-VPPFE-----------EYNNLMESCWSEN-TLRPNLNAIIEWLTKINPK---NIGVDEDTILTKEYAKCLESIIEDRTQ-ALRSEQKMADTLLNSMLPKQVVEMLRHGENVPPEAFEQCTIYFSDIVGFTTISSSSTPFEVVEFLNKLYTQFDDIIDRYDVYKVETIGDAYMVASGVPRRNGERHAIAIADMSLDLVSVSHSFVIPHKPDEPLKIRVGLHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESTGEAYKIHCSETTHAILDRLGGFTFEKRGTITVKGKGDMQTWWITG---RTRAD</td></tr>
</table>

Schistosoma mansoni	gi|353233119|emb|CCD80474.1| serine/threonine RGC [Schistosoma mansoni]	GTAGAGTTTACTGAG---------------TTAACAAGCGATATGCAAAAACAACTATGGGAAGTTAAAAAAATGAAACATAATAATTTAGTGAAGCTCATCGGAGTCACATTTATATCACCTGTCTTATCATTGTATACAGAATTTTGTGATAGAGGAAGTTTATGTTATGTACTTCGACGTGATTCGATCCCGCTGAGTTGGAGTTTAAGAATAGGTTTTCTTACTGGTCTTGCAAATGGTTTAGCTTATTTACATAATTATCACATT---GTACATGGTCGATTGAATTCATCAAATTGTGTAGTTAGTGATACATGGACATGTAAAATAACGGATTATGGATTAGATAGTTTAATTTGG------TCAAATAATTTTGAAAAACAT---AAAACATTTTTAGATAAACCA------GAAAATCTGCCATATATCCCACCTGAATATAGAGGTAAGTATTATAAAGAAACATTATTCGTGCCAGCA---------------GTAGACATATACAGTTTTGGTACAATCATGTGGGAAACAGCAAGT---CGTAGTGATCCG------------TCTCAAGATGACGAATTTGTGGCAGAACCAATGTATAATCAC---------------CCGGAATTGCCGACTAAAAGACGT---TTTGAAACTGATGTTAAA---------TACGAGGTTACAACA---GTACCGCCTTTTGAA--------------------------------GAATACAACAACTTGATGGAATCTTGTTGGAGTGAAGTC---ACACTTAGACCAAACTTAAATACAATTATAGAATGGTTGACTAAGATTAATCCAAGG---------AATATTGGAGTTGATACAGGTACAATACTTGTAAGTAAATATGCTAGATGTCTTGAATCTATTATTGAAGATAGAACACAA---GCGCTTCGTAGTGAACAGAAAATGGCAGATACTTTACTTAACAGTATGTTACCAAAACAAGTTGTAGAAATGCTTAGACATGGTGAAAATGTACCACCTGAAGCATTTGAACAATGTACAATATACTTTAGTGATATTGTCGGTTTTACTACAATTTCATCAAGTTCTACACCATTTGAAGTTGTAGAATTCCTAAACAAATTGTATACTCAATTCGATGATATCATTGATCGATATGACGTTTACAAAGTGGAAACAATCGGTGATGCATATATGGTTGCTTCAGGTGTTCCAAGAAGAAATGGTGAACGTCATGCGATAGCAATAGCTGATATGTCACTTGATCTGGTCAGTGTTTCACACAGTTTCGTAATTCCTCACAAACCTGATGAACCGTTAAAAATTAGAGTCGGTTTACATTCAGGTCCAGTGTGTGCTGGTGTCGTTGGTTTGAAAATGCCAAGATATTGTCTTTTTGGTGATACAGTCAACACAGCAAGTCGAATGGAAAGTACTGGCGAAGCTTACAAAATACATTGTTCAGAAACAACACACGCTATATTGGATCGACTTGGTGGTTTCACTTTTGAAAAACGTGGTACAATAACTGTGAAAGGAAAAGGGGATATGCAGACATGGTGGATCACAGGA---------CGTACTAGAGCTGAT	VEFTE-----LTSDMQKQLWEVKKMKHNNLVKLIGVTFISPVLSLYTEFCDRGSLCYVLRRDSIPLSWSLRIGFLTGLANGLAYLHNYHI-VHGRLNSSNCVVSDTWTCKITDYGLDSLIW--SNNFEKH-KTFLDKP--ENLPYIPPEYRGKYYKETLFVPA-----VDIYSFGTIMWETAS-RSDP----SQDDEFVAEPMYNH-----PELPTKRR-FETDVK---YEVTT-VPPFE-----------EYNNLMESCWSEV-TLRPNLNTIIEWLTKINPR---NIGVDTGTILVSKYARCLESIIEDRTQ-ALRSEQKMADTLLNSMLPKQVVEMLRHGENVPPEAFEQCTIYFSDIVGFTTISSSSTPFEVVEFLNKLYTQFDDIIDRYDVYKVETIGDAYMVASGVPRRNGERHAIAIADMSLDLVSVSHSFVIPHKPDEPLKIRVGLHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESTGEAYKIHCSETTHAILDRLGGFTFEKRGTITVKGKGDMQTWWITG---RTRAD

	*Taenia solium**	TsM_000575800	GTGAAATACGTCGAGCGCGAACAATTTCTTCTGACAAAGAAAATTAGGAAGGAGATCAAGGCAATGCGGCAACTGAGCCATAGGAATTTGTGCCAGCTAGTTGGAGTATGTCTGAAACCGCCTGAAATTGCGATCTTCATGGAATACTGCCCGAAACGAAGCCTTCGAGATGTGTTTCGCAATGAAGTGATGCCATCCAGTTGGGCTTTTAAACTCTCCTTGATTCAAGATATCGTCTGTGGAGTTGAATTTCTCCACACCCACGGCTTC---ATTCATGGACGTCTCAATTCGCAAAACTGCGTTGTTGATGATCGCTTGACTTGTAAAATTACAGATTTTGGACTGGAATCTATCCGC------TACAACAAACCGGAGGAGAAGTTG---GAGACCTTTTTGGAGGATCCC------AGAAATTGGGCATTTATTGCACCCGAGTATCGAGGTAAC--------------------ACACCAGCTCCACCAAATATTCACATGGACTCGTTCAGCTATGGAACCATTATGTGTGAAGTGGCTCAGCCTCGTGAAGATCCC------------------GAAGAT------------------TTCAATGCG------------------CCTGACCTGACAGAGGGCGAGAGG---CAGGAATTGATGGTCCAGTGGTGGGACTACCCTCTACCAAAT---GTGGATGCATTTGAAGGCTCGGCTGATGACAGCACTCCAAATATGAACGACTACCTTAACCTTATTAAGTTGTGCTGGAGTCCGGTT---GATGTGAGGCCTTCATTTGACGTGATTAGATCAAAGATGGACCTCATTAACCCAAGA------AGGAAGAATCCGATCGACATCATACTTAGTCTCATGGAAAAATATTCGGCACATCTTGAAAGCATAGTCAGTGAAAGGACACAA---GACCTCATTGCAGAGAAGCAGCGCACAGACGAATTACTGCACAGTATGCTCCCGAAGACAATCGCAAATCAACTGCGCAGCGGACAGGCAGTTCCAGCCGAGGCCTACTCCTCCTGCACTATCTACTTTAGCGACATTGTTGGCTTCACTAACATTTCTTCGGATTCAACGCCCTTTCAGGTCGTTGCACTGCTTAATAAACTCTACAGTGAGTTCGATCAAATCATTGACCGGTATGATGTCTACAAAGTGGAGACCATCGGAGACGCTTATATGGTAGCTTCGGGAGTTCCAAGAAGAAATGGCCAACGGCACGCAGTTTCTATAACGGACATGGCACTGGATTTGGTCGAGGTGTCACACTCCTTTGTCATTCCCCACATGCCTAACGAACCACTAAAAATTCGAGTTGGCTTGCATTCAGGACCTGTTTGTGCGGGTGTCGTCGGTTTAAAAATGCCAAGGTACTGCCTGTTTGGAGACACAGTCAATACAGCAAGTCGGATGGAGAGCAACGGGGAAGCATACAAAATTCATTGCAGTGATGCCACACATGAAATACTAAGTACACTTGGTGGCTTCCACTTCGAGGAAAGGGGAACAATTGAGGTCAAGGGAAAGGGTACAATGCGAACATGGTGGGTGACAGGT---------CGGACCCGACCACCA	VKYVEREQFLLTKKIRKEIKAMRQLSHRNLCQLVGVCLKPPEIAIFMEYCPKRSLRDVFRNEVMPSSWAFKLSLIQDIVCGVEFLHTHGF-IHGRLNSQNCVVDDRLTCKITDFGLESIR--YNKPEEKL-ETFLEDP--RNWAFIAPEYRGN-------TPAPPNIHMDSFSYGTIMCEVAQPREDP------ED------FNA------PDLTEGER-QELMVQWWDYPLPN-VDAFEGSADDSTPNMNDYLNLIKLCWSPV-DVRPSFDVIRSKMDLINPR--RKNPIDIILSLMEKYSAHLESIVSERTQ-DLIAEKQRTDELLHSMLPKTIANQLRSGQAVPAEAYSSCTIYFSDIVGFTNISSDSTPFQVVALLNKLYSEFDQIIDRYDVYKVETIGDAYMVASGVPRRNGQRHAVSITDMALDLVEVSHSFVIPHMPNEPLKIRVGLHSGPVCAGVVGLKMPRYCLFGDTVNTASRMESNGEAYKIHCSDATHEILSTLGGFHFEERGTIEVKGKGTMRTWWVTG---RTRPP

<table>
<tr><td>RNA binding motif single stranded interacting</td><td><i>Echinococcus granulosus</i></td><td>gi|674569503|emb|CDS15567.1| RNA binding motif single stranded interacting [Echinococcus granulosus]</td><td>ATG---------GAGAAAAGAGAGGATACTGCGAGT---ACAGAAATGCAGTCTCTTGCAGACAATGTGGTCAAGAGAAGCCATTCATTGGAATTGCAGGGTGAAGTCCATCAATCTACTAATGAAAATAGCAAGGAACGATTGCAAGCTGCCAAAATATATTGCATGGATTCCCAATCACTTGCGATTACTTTCCCTGCTGTGACGACCGCCAATACCCCAAAACGCCATCATCCCTCCTCACCT---------------------GTCATTAGCTGTACAACAGACAAGACCAGTGCCAACAATGCT---TCTCAATCTAACAGCGAAGACTTGGCCTCCAGTGCAGCTTCCCTTTCCACTAACTCTGCTGCATCTACACCACTTGAAGATTTGGAGCAGGAAAAACCCGTC---ACAAACTGTGATATGGCCACAACCACCCCTTTGTTGAATGAAGATTCTTAT---GTGAATGGACAAACACCACCATCTTCCCTACGACAAAGTGGGGGTAAC---ACCACCGTCGACGCTGCTGCCTGTGGCACTAAGACGGCAGCATGTGTTTCTACT---------------------------GGGAACAATAGAAACGGATCTCTTCTACCCAGACCTAGTGCAATGGGATTGAACCCCTCAACTTCTTCGAGAACCAAAACTGTGTCGAGGTCTAAGAGGCATTCTTCCCGAACAAATCTTTACATTCGAGGGCTTCCCAAATCAATGTCAGAAAACGACCTTGTTAGTCTTGTTCCCGATGCCTCTGCGATCCGCTCAGTGAAGCTTGTGGTAAATAATGACGGCGAAGGCTACGGGTTCATCGACTTCGTCACCAACGAGGCCGCTCTAATCGCGATGCAGCACATAAAACTTCAGAACTCTGGCCAGTATGTTAACTTTGCCTACGAGTCGGAAAAGGATCCTTTGAATGTCTACGTCACAAATATTCCGGAATCATGGAATTCAGATAATGTGGAGAATTTGAAGCAGATTTTTGCTCCCTACGGCAAAATAACCTCTGCGCTTGTCATGACACGGAGGTCAACAAACACATGCACAGGAACGGGGTTTGTTCGCTATCTCACCTCTGAGGAAGCGCAAAGAGCCATTGATGGCATAAGGAAGGCGAAGATCACGTTGCCTGGAGCAAAGCGGCCTCTGGAACTTAAGCTGGCTGACAGACAACGGGCTCGAGAACACAAAAGTGAATCAATTGGGTCCGAAAACACTACACTGCCACTATTTCAGCAGTTTATGCACGACGAAACTCAGCGACACCTATGCCAGGCATTGAAGACGCAA---CATCACAACCAAATGGCTGAGGACATTTTTAATGTCGCACCTCTGCCAGAGCGCACCATGTTTGATCAGGCCAAGATGGCTGGCTATGTCACATCGGGCGCTTCGCTTATGAATTCGAGTTCGGTGGACATCACACCAACTTCAAGTGATGAAATTCCCGCTCCTGCACCACCTCAGCTCACCACTCCCACCGTCATGCCTGTCGCTCCAGCTGTTGCCGCTGCAACAGCTGCAGCACTGGCAACGCCGCTCTACCAATTTCCGTGTATTCCTTGGATACAGCAGGACTCCACAGCGGCG---CATGGTAATCCT------ACTGCAGCTTACATTCATCCCCAAATAGCGGCTACAGTGCCCCAGAAGGCTGTGACTGCAACAGCA---CCACTGGATCTAAACTCTCTCGCTGCAATCTACCAAATGGCCAGTCTTAATGCTCTCAGCTATGGAGCAACAGCTCCTGCTCAATCTTTTGACATGGTCCTTCCCCTTGTTAATGGGCAGCTTAATGAACATCCCATCGAT--GCTACAAACCTTAATCCATTTGGACTGAATCAGTCGCTTTGGCCAAAAGCATGTTTCTTCGGCNNN</td><td>M---EKREDTAS--TEMQSLADNVVKRSHSLELQGEVHQSTNENSKERLQAAKIYCMDSQSLAITFPAVTTANTPKRHHPSSP-------VISCTTDKTSANNA-SQSNSEDLASSAASLSTNSAASTPLEDLEQEKPV-TNCDMATTTPLLNEDSY-VNGQTPPSSLRQSGGN-TTVDAAACGTKTAACVST---------GNNRNGSLLPRPSAMGLNPSTSSRTKTVSRSKRHSSRTNLYIRGLPKSMSENDLVSLVPDASAIRSVKLVVNNDGEGYGFIDFVTNEAALIAMQHIKLQNSGQYVNFAYESEKDPLNVYVTNIPESWNSDNVENLKQIFAPYGKITSALVMTRRSTNTCTGTGFVRYLTSEEAQRAIDGIRKAKITLPGAKRPLELKLADRQRAREHKSESIGSENTTLPLFQQFMHDETQRHLCQALKTQ-HHNQMAEDIFNVAPLPERTMFDQAKMAGYVTSGASLMNSSSVDITPTSSDEIPAPAPPQLTTPTVMPVAPAVAAATAAALATPLYQFPCIPWIQQDSTAA-HGNP--TAAYIHPQIAATVPQKAVTATA-PLDLNSLAAIYQMASLNALSYGATAPAQSFDMVLPLVNGQLNEHPID---ATNLNPFGLNQSLWPKACFFG?</td></tr>
</table>

Echinococcus multilocularis	gi|961440407| emb|CDS40388.2| RNA binding motif single stranded interacting [Echinococcus multilocularis]	ATG---------GAGAAAAGAGAGGATACTGCGAGT-- ---------------------------------- ACAGAAATGCAGTCTCTTGCAGACAATGTGGTCAAGAGAAGCCATTCATTGGAATTGCAGGGTGAAGTCCATC AATCTACTAATGAAAATAGCAAGGAACGATTGCAAGCTGCTAAAATAAATTGCATGGATTCCCAATCACTTGCG ATTACTTTCCCTGCTGTGACGACCGCCAATACCCCAAAACGTCATCATCCCTCCTCACCT--------------------- GTCATTAGCTGTACAACAGACAAGACCAGTGCCGACAATGCT--- TCTCAATCTAACAGCGAAGACTTGGCCTCCAGTGCAGCTTCCCTTTCCACTAACTCTGCTGCATCCACACCAC TTGAAGATTCAGAGCAGGAAAAACCCGTC--- ACAAACTGTGATATGGCCACAACCACCCCTTTGTTGAATGAAAATTATTAT--- GTGAATGGACAAACACCACCATCTTCCCTACGACAAAGTGGAGGTAAC--- ACCACCGTCGACGCTGCTGCCTGTAGCACTAAGGCGGCAGCATGTGGTTCTACT--------------------------- GGGAACAATAGAAACGGATCTCTTCTACCCAGACCTAGTGCAATGGGATTGAACCCCTCAACTTCTTCGAGAA CCAAAACTATGTCGAGGTCTAAGAGGCATTCTTCCCGAACAAATCTTTACATTCGAGGGCTTCCCAAATCAATG TCAGAAAACGACCTTGTTAGTCTTGTTCCCGATGCCTCTGCGATCCGCTCAGTGAAGCTTGTGGTAAATAATG ACGGCGAAGGCTACGGGTTCATCGACTTCGTCACCAACGAGGCCGCTCTAATAGCGATGCAGCACATAAAAC TTCAGAACTCTGGCCAGTATGTTAACTTCGCCTACGAGTCAGAAAAGGATCCTTTGAATGTATACGTCACAAAT ATTCCGGAATCATGGAATTCAGATAATGTGGAGAATTTGAAGCAGATTTTTGCTCCCTACGGCAAAATAACCTC TGCGCTTGTCATGACACGGAGGTCAACAAACACATGCACAGGAACTGGGTTTGTTCGCTACCTCACCTCTGAG GAAGCGCAAAGAGCCATTGATGGCATAAGGAAGGCGAAGATCACGTTGCCTGGGGCAAAGCGGCCTCTGGA ACTTAAGCTGGCTGACAGACAACGGGCTCGAGAACACAAAAGTGAATCAATTGGGTCCGAAAACACTACACTG CCACTATTTCAGCAGTTTATGCACGACGAAGCTCAGCGACATCTATGCCAGGCATTGAAGACGCAA--- CATCACAACCAAATGGCTGAGGACATTTTTAATGTTGCACCTCTGCCGGAGCGCACCATGTTTGATCAGGCCA AGATGGCTGGCTTTGTCACATCGGGCGCTTCGCTTATGAATTCGAGTTCGGTGGACATCACACCAACTTCAAG TGATGAAATTCCCGCTCCTGCACCACCTCAGCTCACCGCTCCCACCGTTGTGCCTGTCGCTCCAGCTGTTGC CGCTGCAACAGCTGCAGCACTGGCAACGCCGCTCTACCAATTTCCGTGTATCCCTTGGATGCAGCAGGACCC CACAGCGGCG---CATGGTAATCCT------ ACTGCAGCTTACATTCATCCCCAAATAGCGGCTACAGTGCCCCAGAAGGCTGTGACTGCAACAGCA--- CCACTGGATCTACACTCTCTCGCTGCAATCTACCGAATGGCCAGTCTTAATGCTCTTAACTATGGAGCAACAG CTCCTGCTCAATCTTTTGACATGGTCCTTCCCCTTGTTAATGGGCAGCTTAATGAACATCCCATCGAT---------- --------------------------------GTCGCT--------------------------------- TTGGCCAAAAGCATGTTTCTTCGGCTAGTCAAATCTTCGGATTCG-- ---NNN	M---EKREDTAS-------------------------------------- -- TEMQSLADNVVKRSHSLELQGEVHQSTNE NSKERLQAAKINCMDSQSLAITFPAVTTAN TPKRHHPSSP-------VISCTTDKTSADNA- SQSNSEDLASSAASLSTNSAASTPLEDSE QEKPV-TNCDMATTTPLLNENYY- VNGQTPPSSLRQSGGN- TTVDAAACSTKAAACGST--------- GNNRNGSLLPRPSAMGLNPSTSSRTKTMS RSKRHSSRTNLYIRGLPKSMSENDLVSLVP DASAIRSVKLVVNNDGEGYGFIDFVTNEAA LIAMQHIKLQNSGQYVNFAYESEKDPLNVY VTNIPESWNSDNVENLKQIFAPYGKITSALV MTRRSTNTCTGTGFVRYLTSEEAQRAIDGI RKAKITLPGAKRPLELKLADRQRAREHKSE SIGSENTTLPLFQQFMHDEAQRHLCQALK TQ- HHNQMAEDIFNVAPLPERTMFDQAKMAGF VTSGASLMNSSSVDITPTSSDEIPAPAPPQ LTAPTVVPVAPAVAAATAAALATPLYQFPCI PWMQQDPTAA-HGNP-- TAAYIHPQIAATVPQKAVTATA- PLDLHSLAAIYRMASLNALNYGATAPAQSF DMVLPLVNGQLNEHPID--------------VA------ -----LAKSMFLRLVKSSDS---------------------?

<table>
<tr><td></td><td><i>Hymenolepis microstoma</i></td><td>gi|674588650| emb|CDS3235 0.1| RNA binding motif single stranded interacting [Hymenolepis microstoma]</td><td>ATGAAATATTCAAAATCTGAAGAAAATTCTGTTGGAATAAAAATTCGGAGAAGACAAAATGATCAACCGAGTGA
GGTTCTTACCAATGGTTGCCACGAATCGGAATCTCCAATGGATCATGACGGTTCTCTGAAATCTGACCACCGC
AAGCCAACTACAGATGCGCTATCAGTTTCGGATAATGTTGTCAAGAGAAGCCATTCCCTTGAATTGCAGAATGA
ATCCCACTCCCGTTCTAGCGAT------
AAAGTTTTATTGCAGGTGGAAAAAAAGAACAATCTTGAATCTATATCTGCAACAGTAATATCTATACCTACGACC
GCAGCCAGTAGCCCCAAACGTCATCATCCATCTTCACCG--------------------
GTGAGCAAATACGGAATGGATGAGATAGCCACAGAGTCTACAAATAACCAACGAAGGACAAAAGACTCAACTG
TCTCT------------
AATACGTATTCTCCGCATTTCACATCTCTTCAAAATGGCGATCAAGAGAAACCGATGGATAATGACTGTTCTAT
GAGTACAGGTACTCCGCTTGGTGATGACAGTTCGTCAAGTGGAAATGGAAAAACATCATCGCCTTCCCTGAGG
CAAGCTGATCCCAATGTATCAACTACTAATAGTGTTGCGAGTAGCACTAAAGTCTCCTTACCAGGAGTACCA----
----
ATTAACCGGAATGGCTCCCTTCTCTCTAGATCTGGAGTTTCAAGCTCGAATCAACCAAACTCCTCCCGTAGCAA
GACAGCATCTAGATCTAAGAAGCATTCATCACGAACCAACTTATACATTCGAGGTCTTCCAAAATCCATGTCAG
AGAATGACCTAGTGAGTCTTGTTCCAGACGCCTCAGCAATCCGCTCCGTAAAATTGGTTGTGAATAACGATGG
AGAAGGCTATGGATTTATTGATTTCGTTTCAAATGAGGCTGCTCTTATAGCTATGCAGCACATTAAACTCCAAA
ACTCCAACCAATATGTCAACTTTGCATATGAGTCAGAAAAGGATCCGTTGAATGTTTATGTGACAAATATTCCA
GAGTCCTGGAGTTCAGATAACGTTGAGACCCTCAAGCAAATTTTCGCTCCCTACGGCAAAATCACCTCTGCCC
TTGTCATGACTAGGAGATCGACAAATACCTGTACTGGAACTGGATTTGTGCGCTATCTCACATCTGAAGAAGCT
CAACGGGCAATTGATGGTATAAGGAAAGCCAAAATAACTCTTCCTGGGGCTAAACGGCCACTCGAACTCAAAC
TGGCGGATAGACAACGAGCCCGTGAGCACAAAAGTGAATCAATTGGATCAGAAAATAACGCTCTTCCCTTATT
CCAACAATTTATGCGTGAGGATGCCCAACAGCAATTAAGTCAAGCAATAAAGAATCAGCAACACCACAAACAA
ATGGTGGACGACATTTTCAATGCTGCTTCCATATCCGATCGAGGTATGTTCGATCAGACAAAGATGGCAAACT
ATGTCAATTCTGGTGCCTCTCTCATGAGTCCAAGCCCTGTGGATATAACACCAAAATCTAGCGATGAGATTTCA
GCTACTGCGCCTTCACAACTGACAGCTCCAACCGTTATGCCTGTTTCTCCAGCTGTTGCAGCAGCCACG------
GCACTTGCTTCTCCAGTGTATCAATTTCATTGTTTTCCATGGATACAACAAGACCACACGACCGCGCCACATGC
GAACTCCCAAGCAGCAGCTGCCTACCTTCACTCTCAGTTGGCGGCAGCGGTCCAAAAGGCTAACCCAGCTGC
CACCGCGACCCCTCTGGATCTGCACTCACTAGCTGCTGTCTATCAGATGGCTGGTCTAGGAGTCCTCAACTAC
GGTGCCACGGCTCCTTCTCAATCGTTCGATATGATTCTTCCTTTGCTTAACGGCCAACTTAATGGCCATCCGAT
TGAT--
GCCGCCAATCTATATCCCTTCGGGCTGAATCAATCTGTGTGGCCTAAAGCGTGTTTCTAC--------------------
GGA---NNN</td><td>MKYSKSEENSVGIKIRRRQNDQPSEVLTN
GCHESESPMDHDGSLKSDHRKPTTDALSV
SDNVVKRSHSLELQNESHSRSSD--
KVLLQVEKKNNLESISATVISIPTTAASSPK
RHHPSSP-------
VSKYGMDEIATESTNNQRRTKDSTVS----
NTYSPHFTSLQNGDQEKPMDNDCSMSTG
TPLGDDSSSSGNGKTSSPSLRQADPNVST
TNSVASSTKVSLPGVP----------
INRNGSLLSRSGVSSSNQPNSSRSKTASR
SKKHSSRTNLYIRGLPKSMSENDLVSLVPD
ASAIRSVKLVVNNDGEGYGFIDFVSNEAALI
AMQHIKLQNSNQYVNFAYESEKDPLNVYV
TNIPESWSSDNVETLKQIFAPYGKITSALVM
TRRSTNTCTGTGFVRYLTSEEAQRAIDGIR
KAKITLPGAKRPLELKLADRQRAREHKSES
IGSENNALPLFQQFMREDAQQQLSQAIKN
QQHHKQMVDDIFNAASISDRGMFDQTKMA
NYVNSGASLMSPSPVDITPKSSDEISATAP
SQLTAPTVMPVSPAVAAAT--
ALASPVYQFHCFPWIQQDHTTAPHANSQA
AAAYLHSQLAAAVQKANPAATATPLDLHSL
AAVYQMAGLGVLNYGATAPSQSFDMILPL
LNGQLNGHPID---------------
AANLYPFGLNQSVWPKACFY-------G---------
------------?</td></tr>
</table>

	*Mesocestoides corti**	MCOS_0000139701-mRNA-1	ATG---------CACGAA--ACCTTAACTTCTCCGAAACGTCACCATCCCGCCTCACCGTCTTCACTGGGACCAAGATCAGTCATCCTTTATGCAAAAGAGGAAAGTGCCACTACAAGCCCT---ATTGAGGACAAAATGCAAGGCACAAACCCTAGCACTACTTCTGTTTATTGTGAATCTTCAGCATCAATGTGCAGAGAAGAGGTAAATCGAGAGAAACCGGTC---ATTGGCTGTGACATGGCTGTGACGAAAGTACTAACGGATGAGCCCCTACCC---GGTTACGAGAGGGTATTTCCATCACCTTACCAGCAGAGTGAAGATGAT--GCCCCTGGTGTTAGTGAAAAGCAGAAAACAAATCAGTTGAATGGGCTCAACAGATCAGGATCACTGCTGCCTAGACCTGGTCATGGTGGAACGGCTTCCTTGATGTCCTCCCGGAATAAAACAGTTGCGAGGTCTAAGCGGCACTCCTCAAGAACAAATCTCTACATCCGAGGGCTTCCGAAATCGATGTCAGAAAATGACCTTGTTAGTCTTGTGCCGGATGCCTCAGCGATTCGGTCAGTGAAGCTCGTTGTTAATAACGACGGTGAAGGT------------------------AACGCTGCCTCATTTATT------AAAACCATAAAA-----------------------------------GAGTCAGAAAAAGATCCGCTAAATGTCTACGTAACAAACATTCCAGAATCCTGGAACTCGGACAATGTGGAGAATCTTAAGCAAATTTTTGCTCCTTATGGCAAGATAACGTCAGCACTTGTCATGACACGGAGATCGACAAATACATGTACAGGAACTGGATTTGTCCGCTATCTGACCTCTGAAGAAGCCCAAAGAGCCATTGATGGGATTCGGAAAGCAAAGATTACGCTGCCTGGTGCAAAGCGGCCTCTTGAACTCAAGTTGGCTGACAGACAACGGACACGTGAACACAAAAGGNNN---	M---HE---TLTSPKRHHPASPSSLGPRSVILYAKEESATTSP-IEDKMQGTNPSTTSVYCESSASMCREEVNREKPV-IGCDMAVTKVLTDEPLP-GYERVFPSPYQQSEDD-------------APGVSEKQKTNQLNGLNRSGSLLPRPGHGGTASLMSSRNKTVARSKRHSSRTNLYIRGLPKSMSENDLVSLVPDASAIRSVKLVVNNDGEG--------NAASFI--KTIK------------ESEKDPLNVYVTNIPESWNSDNVENLKQIFAPYGKITSALVMTRRSTNTCTGTGFVRYLTSEEAQRAIDGIRKAKITLPGAKRPLELKLADRQRTREHKR?--

	*Taenia solium**	TsM_000485200	ATG---------GCGAAAAGAGAGGATACTGTGGGT--ACAGAAATGCAGTCCCTTTCTAAAAATGTGGTCAAGGGAAGCCATTTGTTGGAATCGCAAAGTGAAGATCATCGTTCTGCTAACGAATACAGCAAGGACCAAGTGCAAGTAGATAAAATAAGTTGTATGGACTCACAATCACTTGCGATTGCTTCGCCTTCCGTAATGACCGTCAATACCCCAAAACGCCATCACCCCGCCTCACCT----------------------GTCATTAGCTATGCAACAGACGAGACCAGTACTGACAATGCT---GCTCATTCTAAGAGCGAAGACTCTGTCTCCAGTGCAGCTTCCCTTTCCACCGACTCAGCTGCATCCACACCACATAAAGATTTGGAGCAGGAGAAGCACGTC---ACAGACTGTGATATGGCCACAGCCATTCGCCTATTGGATGAAAATTCTCAT---ATAAATGTAAAAACATCAGTATCTGCTCTAAGACAGTGTGCAGGAAAT---TCTGCCGTCCTCGCTGCTGCCTCCGGCAGCAAAGCAACAGCATGTGGTTCAAGT----------------------------GGAAACAATCGAAACGGATTTCTTTTACCTAGACCCGGAGCATCAGGACCGAATCCCTTGCCTTCTTCCCGAACAAAAGCTGTGTCGAGGTCTAAGAGGCATTCATCCCGGACAAATCTCTACATTCGAGGGCTTCCCAAATCGATGTCAGAAAACGACCTTGTTAGTCTTGTTCCTGACGCCTCTGCAATTCGTTCAGTGAAGCTTGTGGTAAATAATGAGGGCGAAGGCTACGGGTTTATCGACTTTGTCTCCAATGAGGCCGCTCTGATCGCAATGCAGCACATTAAACTTCAGAATTCTAGCCAATATGTTAACTTTGCCTATGAATCGGAAAAGGATCCTTTGAATGTCTACGTTACAAACATTCCTGAATCATGGAATTCAGATAATGTGGAGAATTTAAAGGAGATCTTTGCTCCATACGGCAAAATAACCTCTGCACTTGTCATGACACGGAGATCAACAAACACCTGCACAGGAACGGGATTTGTTCGGTATCTCACCTCTGAGGAAGCGCAAAGAGCCATTGATGGTATAAGAAAGGCGAAGATTACGTTGCCCGGAGCAAAGCGGCCCCTGGAACTTAAATTGGCCGACAGACAACGAGCTAGGGAACACAAAAGTGAATCAATTGGCTCTGAAAACAATACGCTGCCACTGTTTCAGCAGTTTATGCGCGATGGTACTCAGCGACACCTTTGCCAGGCATTGAGGACGCAG---CATCACAACCAAATGGCTGAAGACATTTTTAATGTACCATCAGTGCCAGAACGCACCATGTTTGATCAAACCAAGGTGGCTGGCTATGTGGCATCGGGCTCTTCACTTATGAATACAAGTCCAGTGGACATCACACCAACTTCAAGTGATGAGATTTCTGCTCCTGCACCACCTCAGCTCACTACTCCAGCCGTTTTGCCTATCGCCCCAGCTGTTGCCGCTGCAACGGCTGCAGCATTGGCAACGCCACTCTACCAATTCCCATGCATCCCTTGGATACAACAGGACCCCACAGCGGCC---CATGGCAATCCT------ACTGCGGCCTACATTCATCCCCAAATGGCGGCTACAGTACCCCAGAAGGCCGTGACTGCGACAGCA---CCACTGGATCTGAACTCTCTCGCTGCAGTCTACCAAATGGCCAGTCTTAATACTCTCAACTATGGAGTAACAGCTCCTGCTCAATCTTTTGACATGGTCCTTCCCCTTATTAATGGGCAGCTTAATGAACATCCCATCGATGAGGTCGGATCTGGCTACAAACCTTTATCCATTTGGACTGAATCAGTCGCT----------------------------------TTGGCCAAAAGCCTGTTTCTTCGGCTAGCCAAATCTCCGGATTCG--NNN	M---AKREDTVG---TEMQSLSKNVVKGSHLLESQSEDHRSANEYSKDQVQVDKISCMDSQSLAIASPSVMTVNTPKRHHPASP-------VISYATDETSTDNA-AHSKSEDSVSSAASLSTDSAASTPHKDLEQEKHV-TDCDMATAIRLLDENSH-INVKTSVSALRQCAGN-SAVLAAASGSKATACGSS---------GNNRNGFLLPRPGASGPNPLPSSRTKAVSRSKRHSSRTNLYIRGLPKSMSENDLVSLVPDASAIRSVKLVVNNEGEGYGFIDFVSNEAALIAMQHIKLQNSSQYVNFAYESEKDPLNVYVTNIPESWNSDNVENLKEIFAPYGKITSALVMTRRSTNTCTGTGFVRYLTSEEAQRAIDGIRKAKITLPGAKRPLELKLADRQRAREHKSESIGSENNTLPLFQQFMRDGTQRHLCQALRTQ-HHNQMAEDIFNVPSVPERTMFDQTKVAGYVASGSSLMNTSPVDITPTSSDEISAPAPPQLTTPAVLPIAPAVAAATAAALATPLYQFPCIPWIQQDPTAA-HGNP--TAAYIHPQMAATVPQKAVTATA-PLDLNSLAAVYQMASLNTLNYGVTAPAQSFDMVLPLINGQLNEHPIDEVGSGYKPLSIWTESVA-----------LAKSLFLRLAKSPDS---------------------?

Serine:threonine protein kinase	*Echinococcus granulosus*	gi|674561902|	TTGAAGCTCTATGAGGTGATTGAATCTGATCGTCACGTCTATCTGGTAATGGAATTCGCCGCAAATGGTGAGC	LKLYEVIESDRHVYLVMEFAANGELFEYLV
		emb|CDS2386	TCTTTGAATACCTCGTGTCCAATGGTCGGATGCGTGAGAAGGATGCTCGCATCAAGTTCCGTCAGATTGTCTC	SNGRMREKDARIKFRQIVSAVQYCHQKNIV
		9.1|	TGCGGTACAGTACTGCCACCAAAAAAATATCGTCCACCGTGATCTCAAGGCAGAGAATCTACTCCTGGATGCC	HRDLKAENLLLDADYNIKLADFGFSNTFRA
		serine:threonin	GATTACAACATCAAATTGGCCGACTTCGGTTTCTCCAATACATTCCGCGCGGATAAGAAATTGGACACATTCTG	DKKLDTFCGSPPYAAPELFLGKKYIGPEVD
		e protein	TGGGTCGCCACCGTACGCAGCGCCCGAGTTGTTCCTTGGCAAGAAGTATATCGGTCCTGAGGTGGATGTTTG	VWSLGVILYTIVAGYLPFDAQNLRDLRERV
		kinase	GTCACTGGGCGTCATTCTCTACACCATTGTTGCCGGGTACCTTCCCTTTGATGCACAGAACTTGCGGGATCTT	LRGKYRIPFFMSTDCEMLLKRMLVLNPEKR
		[Echinococcus	CGTGAACGTGTTCTCCGTGGCAAATATCGCATCCCCTTTTTCATGTCCACCGATTGCGAGATGCTGCTGAAGC	YSLLSVMEDKWTNINMEDNILRPYQEPPP
		granulosus]	GCATGCTTGTCCTTAACCCAGAGAAAAGGTATTCACTCTTATCTGTCATGGAGGACAAATGGACGAATATCAAC	DFKDPIRLAKMVEMGFTLEEVKDSLENNKF
			ATGGAAGACAATATTCTCCGACCCTACCAAGAACCACCGCCTGACTTCAAAGATCCTATTAGACTTGCCAAGAT	NNVTATYFLLGTDRSTSSNSSLCPQLPSSV
			GGTCGAGATGGGCTTCACGCTGGAGGAGGTTAAGGACTCTCTGGAGAATAATAAATTCAACAACGTGACAGC	SISNRPAAFTTTDDGSASVSTSDPA-------
			TACCTACTTCCTCTTAGGGACGGATCGCTCCACCTCCTCTAACTCATCCCTCTGTCCCCAATTACCCTCTTCCG	KSVSTPKGSAA------
			TCAGCATTTCCAACCGCCCCGCCGCCTTCACCACTACCGATGACGGCTCCGCATCTGTGTCCACCTCCGATC	VGTTPAAVADGGSALTTLEDS-----
			CCGCC--------------------AAGTCCGTGTCTACACCCAAAGGATCTGCCGCC-----------------	DTPQLLASSASIRKSSSST----------
			GTTGGCACCACACCTGCTGCTGTTGCTGATGGCGGTTCTGCCTTGACCACCCTCGAAGATTCA--------------	RANGST---NSTVPAAE-------STSDSTSS-----
			GACACACCTCAATTGCTGGCCTCTTCAGCCTCCATCAGAAAGAGTTCCTCTTCTACG-----------------------------	ASGK----
			CGTGCAAACGGATCTACT---------AACTCCACCGTTCCCGCTGCAGAG--------------------	QKSAWAQKRGETVSVESTTVTSGVRPDK
			TCTACCTCCGATTCTACCTCGTCT---------------GCCAGCGGAAAG------------	PKV--------
			CAGAAATCCGCTTGGGCACAAAAACGCGGGGAGACTGTTTCCGTTGAATCGACGACTGTCACCAGTGGCGTA	AIPTTHFDTPKSSEAPRDSPTNRPP-------
			AGACCTGATAAGCCAAAGGTG-----------------------	SKPSGGIGSGSGVRRTRTFTSTDKRRNTV
			GCTATTCCAACTACTCACTTTGATACTCCCAAGTCATCCGAGGCCCCAAGGGACTCTCCCACTAATCGACCTC	AVGGPGG------
			CG--------------------	AEGDSDVAKALIRVPNMDNIVD---------------
			TCGAAGCCGTCTGGCGGCATTGGCAGTGGTAGTGGTGTTCGGAGAACGCGAACCTTCACGTCCACGGATAAA	--
			CGCCGCAACACAGTTGCTGTCGGAGGCCCTGGTGGT-----------------	QVGAK--SGGERDASDRELDV----
			GCTGAGGGTGACTCTGATGTCGCAAAGGCTCTTATTCGCGTCCCCAATATGGACAACATCGTCGAC-------------	DGLGVHSSASGRCAASGSSDLRLGKTLKQ
			---	ATTMPSPTS----
			---CAAGTGGGTGCTAAA------	PRGCADDSECSTLKRQDSFWKRVRR-------
			TCAGGGGGAGAGCGGGACGCATCGGACCGAGAGCTGGATGTC------------	------------------------SMSRKHQRNK-SR----
			GACGGTTTGGGGGTTCACTCCTCGGCTAGTGGACGCTGTGCTGCTAGTGGTTCGAGTGATCTCCGGCTGGGA	PAIHFEAADEVLSADHPSHAVAKVE----
			AAGACCCTGAAACAGGCCACCACTATGCCTAGTCCTACTTCC------------	PNTVPTFMRGAPDRSTAPAARIVRH---
			CCCCGTGGCTGCGCCGACGATAGCGAATGCAGCACCCTCAAACGCCAGGACTCCTTCTGGAAGCGGGTTAG	ATNTAVEHSQRFRPP-
			AAGG---	VTSSSPLTHSSQQQRQAEV-
			AGCATGTCACGGAAGCATCAACGCAACAAG---TCGCGT------------	IESPVRVTTKPTEYHH---
			CCGGCAATCCACTTTGAGGCAGCGGATGAGGTGCTTTCCGCCGATCATCCTTCACACGCAGTTGCCAAAGTC	SSSFLRSVSSRLSKSFRRKRNRSHSRTTP
			GAG------------	TRG----------ASIQAAVERGEAGKQ-----
			CCCAACACAGTACCCACCTTCATGCGGGGCGCACCAGACCGCAGTACCGCCCCGGCAGCTCGCATCGTCCG	GDSLMATDSPIPAHHRAFSS-ERKH-------
			CCAC---------GCCACCAACACTGCAGTGGAGCACTCGCAGCGATTCAGGCCCCCC---	AVITT---
			GTCACCTCCTCTTCACCTCTGACGCACTCCTCCCAGCAACAGCGGCAGGCGGAAGTG---	DDPWSSMLPEKGHSEGEALVASISNSLEP
			ATAGAGTCGCCGGTGCGGGTGACCACCAAACCCACTGAGTACCACCAC---------	LLPACFDKKPRNISVFSGTWRKA------

<table>
<tr><th></th><th></th><th></th><th></th></tr>
<tr>
<td>*Echinococcus multilocularis*</td>
<td>gi|674577520|emb|CDS36968.1| serine:threonine protein kinase [Echinococcus multilocularis]</td>
<td>TTGAAGCTCTATGAGGTGATTGAATCTGATCGTCACGTCTATCTGGTAATGGAATTCGCCGCAAATGGTGAGC
TCTTTGAATACCTCGTGTCCAACGGTCGGATGCGTGAGAAGGATGCTCGCATCAAGTTCCGTCAGATTGTCTC
TGCGGTACAGTACTGCCACCAAAAAAATATTGTCCACCGTGATCTCAAGGCAGAGAATCTACTCCTGGATTCC
GATTACAACATCAAATTGGCCGACTTCGGTTTCTCCAATACATTCCGCGCGGATAAGAAATTGGACACATTCTG
TGGGTCGCCACCGTACGCAGCGCCCGAGTTGTTCCTTGGCAAGAAGTATATCGGCCCTGAGGTGGATGTTTG
GTCACTGGGCGTCATTCTTTACACAATTGTTGCCGGGTACCTTCCTTTTGATGCACAGAACTTGCGGGATCTTC
GTGAACGTGTTCTCCGTGGCAAATATCGCATCCCCTTTTTCATGTCCACCGATTGCGAGATGCTGCTGAAGCG
CATGCTTGTCCTTAACCCAGAGAAAAGGTATTCACTCTTATCTGTCATGGAGGACAAATGGACGAATATCAATA
TGGAAGGCAATATTCTCCGACCCTACCAAGAACCACCGCCTGACTTCAAAGATCCTATTAGACTTGCCAAGAT
GGTCGAGATGGGCTTCACGCTGGAGGAGGTTAAGGACTCTCTGGAGAATAATAAATTCAACAACGTGACAGC
TACCTACTTCCTCTTAGGGACGGATCACTCCACCCCCTCTAACTCGTCCCTCTGTCCCCAATTACCCTCTTCCG
TCAGCATTTCCAACCGCCCCGCCGCCTTCACCACTACCGATGACGGCTCCGCATCTGTGTCCACTTCCGATC
CCGCT--------------------AAGTCTGCGTCTACACCCAAAGGATCTGCCGCC-----------------
GTTGGCACCACACCTGCTGCTGTTGCTGATGACGGTTCTGCCTTGACCACCCTCGAAGATTCA--------------
GACACACCTCAATTGCTGGCCTCTTCAGCCTCCATCAGAAAGAGTTCCTCTTCTACG----------------------------
CGTGCAAACGGATCTACT---------AACTCCACCGTTCCCGCTGCGGAG--------------------
TCTACCTCCGATTCTACCTCGTCT--------------GCCAGCGGAAAG-----------
CAGAAATCCGCTTGGGCACAAAAACGTGGGGAGACTGTTTCCGTTGAATCGACGACTGTCACCAGTGGCGTA
AGACCTGATAAGCCAAAGGTG----------------------
GCTATTCCAACTACTCACTTTGATACTCCCAAGTCATCCGAGGCCTCAAGGGACTCTCCCACTAATCGACCTC
CG--------------------
TCGAAGCCGTCTGGCGGCATTGGCAGTGGTAGTGGTGTTCGGAGAACACGAACCTTTACGTCCACGGATAAA
CGCCGCAACACAGTTGCTGTCGGAGGCCCTGGTGGT-----------------
GCTGAGGGTGACTCTGATGTCGCAAAGGCTCTTATTCGCGTCCCCAATATGGACAACATCGTCGAC-------------
----
--CAAGTGGGTGCTAAA------
TCAGGAGGAGAGCGAGACGCATCGGACCGAGAGCTGGATGTC-----------
GACGGTTTGGGGGTTCACTCCTCGGATAGTGGGCGCTGTGCTGCTAGTGGTTCGAGTGATCTCCGGCTGGG
AAAGACCCTGAAACAGGCCACCACTATGCCTAGTCCTACTTCC-----------
CCCCGTGGCTGCGCCGACAATAGCGAATGCAGCACCCTCAAACGCCAGGACTCCTTCTGGAAGCGGGTTAG
AAGG--
AGCATGTCACGGAAGCATCAGCGCAACAAG---TCGCGT-----------
CCGGCAATCCACTTTGAGGCAGCGGATGAGGTGCTTTCCGCCGATCATCCTTCACACGCAGTTGCCAAAGTC
GAG-----------
CCCAACACCGTACCCACCTTTATGCGGGGCGCACCAGACCGCAGTACCGCCCCGGCAGCTCGCATCGTCCG
CCAC---------GCCACCAACACTGCAGTGGAGTACTCGCAGCGATTCAGGCCCCCC---
GTCACCTCCTCTTCACCTCTGACGCACTCCTCCCAGCAACAGCGGCAGGCGGAAGTG---
ATAGAGTCGCCGGTGCGGGTGACCACCAAACCCACTGAGTACCACCAC--------</td>
<td>LKLYEVIESDRHVYLVMEFAANGELFEYLV
SNGRMREKDARIKFRQIVSAVQYCHQKNIV
HRDLKAENLLLDSDYNIKLADFGFSNTFRA
DKKLDTFCGSPPYAAPELFLGKKYIGPEVD
VWSLGVILYTIVAGYLPFDAQNLRDLRERV
LRGKYRIPFFMSTDCEMLLKRMLVLNPEKR
YSLLSVMEDKWTNINMEGNILRPYQEPPP
DFKDPIRLAKMVEMGFTLEEVKDSLENNKF
NNVTATYFLLGTDHSTPSNSSLCPQLPSSV
SISNRPAAFTTTDDGSASVSTSDPA-------
KSASTPKGSAA------
VGTTPAAVADDGSALTTLEDS-----
DTPQLLASSASIRKSSSST----------
RANGST---NSTVPAAE-------STSDSTSS-----
ASGK----
QKSAWAQKRGETVSVESTTVTSGVRPDK
PKV--------
AIPTTHFDTPKSSEASRDSPTNRPP-------
SKPSGGIGSGSGVRRTRTFTSTDKRRNTV
AVGGPGG------
AEGDSDVAKALIRVPNMDNIVD----------------

QVGAK--SGGERDASDRELDV----
DGLGVHSSDSGRCAASGSSDLRLGKTLKQ
ATTMPSPTS----
PRGCADNSECSTLKRQDSFWKRVRR-------
------------------------SMSRKHQRNK-SR----
PAIHFEAADEVLSADHPSHAVAKVE----
PNTVPTFMRGAPDRSTAPAARIVRH---
ATNTAVEYSQRFRPP-
VTSSSPLTHSSQQQRQAEV-
IESPVRVTTKPTEYHH---
SSSFLRSVSSRLSKSFRRKRNRSHSRTTP
TRG----------ASIQAAVQRGKAGKQ-----
GDSLMATDSPIPAHHRAFSS-ERKH-------
AVITT---
DDPWSPMLPEKGHSEGEALVASISNSLEP
LLPACFDKKPRNISVFSGTWRKA------</td>
</tr>
</table>

Hymenolepis microstoma	gi|674591923|emb|CDS29252.1| serine:threonine protein kinase [Hymenolepis microstoma]	GTGAAACTCTACGAAGTGATCGAGTCCGACCGACACGTTTACCTCGTAATGGAATTCGCTGCAAACGGTGAGCTCTTTGAATACCTCGCGTCGAACGGTCGAATGCGTGAGAAGGATGCCCGCATCAAGTTCCGCCAAATCGTTTCCGCTGTGCAATATTGCCATCAGAAAAACATCGTTCATCGGGATCTCAAAGCAGAGAATTTACTGTTAGATGCGGACTTCAACATTAAGCTGGCAGACTTCGGTTTCTCTAATGCCTTCCGAGCTGACAAGAAACTGGACACATTCTGCGGCTCCCCGCCTTACGCAGCACCGGAACTTTTTCTGGGCAAGAAGTACATTGGTCCTGAGGTGGACGTATGGTCCTTAGGGGTCATTCTCTACACAATCGTCGCTGGTTACCTCCCATTCGACGCCCAGAACCTTCGGGATCTGCGTGAGAGGGTACTGAGAGGCAAATATCGAATTCCTTTCTTCATGTCAACCGACTGTGAACTCCTGCTCAAGAAAATGCTCGTTCTTAATCCCGAAAAGCGCTATTCCCTCTTGAATGTGATGGCGGACAAGTGGACAAATATTGGTATGGATGATAATCCCCTTCAGCCCTACCAGGAACCCCCTCCCGACTACAATGATCCCGTCAGACTAAAGAAAATGGAGGAGATGGGATTCGCGCTGGAGGAAATAAGAGACTCCCTGGAGAACAACAAATTTAACAACGTGACTGCCACGTACTTCCTTCTTGGCACTGATCGACCCGCTTCGTCCTCC-----------------------TCCTCC--CCCGCCATCACCACCACCACTGCTACCGCCACTGTAAGCAAATCTGGGGAAAAGAAGGACACCACCTCTTTCGGTGTCGCGGCAGCTAGCAAGAGCTCTTCCACG-----------------------------------GCTTGCACCGATTGTGATCCTTCA---GCTGAAGCTCCTGATACCCCCCAAGTGCTGGTGTCCTCCACTTCTATGCGCAGGAAGGATTCGTCGTCAGTTTCGGCTGCTGCGTCGACAACTGCACAACGGGCGAATGGATCGGCTGGTGTGTCTACCTCAGCAGCCGCGGCAGCAGCTCCTCCATCCACTGAAAATAATTCCACTTCAGAATCTTCTCCTTCCACCACTGCCGCCATCGCAAATGCCAAACACCCACAACAGCAAAAGTCTACTTGGTCTCAAAAGATAGAGGATACAGCTACAGGAGATTCAGGAACA---ACTGGAGGAGTGCATTTCGACAAACCCAAGATCTCAGCCGTAAGTTCTGCTGCCCCTGCAATGCCGCTGACACGATTCGACCCTGTCCTCACGACAACCGCAAAAAAGGAT------------CGTTCTCCGGCCACTACCGTCACCAACAGTTCAAATAGT--------------------GCTGGAGTCCGGCGGACTCGGACCTTCACTTCTGTAGACAAGAGGCGTAGCGCAGCTGGTATC-------------------------------GCTGCAAGTGATAACGACTCCTCAAAGGCCCTTGTTCGAGCCCCTAACATGGATAACATTGTCGAT--CAAGTT---GATGAAGCTGCGTCTGGGGTTGAAGGGCACGATTCTGGCAGCGGTTTGGACGTGTTGGTCGATGGTGGAGGAGGAGGAAAACATTCCTTCACAACCACTCGT------CCCACTGGTTCAGCTGAAATCCGTCGTCTCACCTCCACCAATGTCACAGCAACCCTCCCCAGCCAATCTTCT------------CGT--------------GACAGCGCCTCCAACACCCTCAAACGGCATGATTCATTCTGGAAGCGTGTTCGAAGG---AGTATGTCGAGGAAGAAGGACCGGGAGAAC---AGGAAGTTGGCACAGCAACCAACGATTCGTTTTGAAGCGGCGGACACAGTGCTCTCCGGA--------------------------GCCCCA------------CCGAACACCGTTCCAGCTTTCATGCGAGGTGCGCCCGATCGGTCCACAGCGCCACCCTCTCGCGTCACCCGCGATCGCAAAGCCGCTCTTGAACCCGCACTTGAGCACTCTCAGCGATTCCGTCCTCCTGCCACTGCTCTTCAACAGCAGTCTACGGAAGATGATGATGAGCAG---------------	VKLYEVIESDRHVYLVMEFAANGELFEYLASNGRMREKDARIKFRQIVSAVQYCHQKNIVHRDLKAENLLLDADFNIKLADFGFSNAFRADKKLDTFCGSPPYAAPELFLGKKYIGPEVDVWSLGVILYTIVAGYLPFDAQNLRDLRERVLRGKYRIPFFMSTDCELLLKKMLVLNPEKRYSLLNVMADKWTNIGMDDNPLQPYQEPPPDYNDPVRLKKMEEMGFALEEIRDSLENNKFNNVTATYFLLGTDRPASSS--------SS----------------PAITTTTATATVSKSGEKKDTTSFGVAAASKSSST------------ACTDCDPS-AEAPDTPQVLVSSTSMRRKDSSSVSAAASTTAQRANGSAGVSTSAAAAAAAPPSTENNSTSESSPSTTAAIANAKHPQQQKSTWSQKIEDTATGDSGT-TGGVHFDKPKISAVSSAAPAMPLTRFDPVLTTTAKKD----RSPATTVTNSSNS-------AGVRRTRTFTSVDKRRSAAGI-----------AASDNDSSKALVRAPNMDNIVD--QV-DEAASGVEGHDSGSGLDVLVDGGGGGKHSFTTTR--PTGSAEIRRLTSTNVTATLPSQSS----R-----DSASNTLKRHDSFWKRVRR--------------------------------SMSRKKDREN-RKLAQQPTIRFEAADTVLSG---------AP----PNTVPAFMRGAPDRSTAPPSRVTRDRKAALEPALEHSQRFRPPATALQQQSTEDDDEQ-----EAESPVRVTTKPTEYHSAASSFSFLRSVSSRLSKSFHRKRNRSHSRATPTRGPGNGVDEPFATGKQVTVTDG------------------SPLPVHQRVHSNAERKHHNTANAVSV------DDPWSPMLPEHGRLE------DLNNSLEPLLPACFDKKPRNISIFSGTWRKAAASSHHSSSTTSQSENAGTTTTASLFGS

*Mesocestoides corti**	MCOS_0000070101-mRNA-1	GTCAAACTTTACGAGGTGATCGAATCTGAGCGCCATGTTTACCTCGTCATGGAATTCGCCGCCAATGGTGAAC TGTTTGAGTACCTCGTGTCCAACGGTCGAATGCGCGAAAAAGACGCTCGAATCAAGTTTCGGCAAATTGTCTC CGCCGTGCAGTACTGCCATCAAAAAAACATTGTCCATCGAGACCTCAAGGCAGAGAATCTCCTCTTGGACGCG GACTATAACATTAAATTGGCCGACTTCGGTTTCTCCAACACCTTTCGGCCGGACAAGAAGCTGGACACATTTT GCGGCTCACCGCCCTACGCCGCGCCCGAGTTGTTCCTGGGCAAGAAGTACACCGGACCCGAGGTCGATGTC TGGTCCTTGGGTGTCATTCTCTACACCATCGTCGCCGGATACCTACCCTTCGATGCGCAGAATCTTCGCGATC TCCGTGAGCGGGTTCTCCGTGGCAAATACAGAATCCCCTTCTACATGTCCACAGATTGTGAGATGTTGCTTAA ACGCATGCTTGTCCTTAACCCAGAAAAACGACATTCCCTCTTGTCCGTGATGGAAGACAAATGGACGAACATC AACATGGAGGACGACATTCTTCGTCCATACCAGGAGCCTCCACCCGATTACAACGATCCTGTGCGTCTTGCCA AAATGACTGAAATGGGCTTCACGCTGGAGGAAATCAAGGACTCACTGGAGAATAACAAATTCAACAATGTGAC TGCCACCTACCTCTTGTTGGGCACCGACCGAAAGAACTCCAGCTCCACC---------CCCCCACCTCCGTCCTCT--- --- TCCGCCGCTGCGGCTAACAATCAACCAGCTGTTGTTGATGAGTCGCCCAAACAGAAGACTTCCACCCAGAAG GAATCCGGTGCC------------------ACC------------------------------------ GCGCTAGCCGCGGTGGAAGATTCCATGGTCGAAGCTCCTGACACTCCTCAGGTTCTAGTCTCATCTACCTCCA TGCGGAAGGGATCCTCCACGACA------------------------------CGCGCGAATGGGTCCGCC--------- ACGTCCACAGCCCCTGTCACCGAGAACCCC------------CAATCAACTCCCGAGTCCACCTCGTCT--------------- GCCAGCGGCAAA------------ CAAAAGCCCACTTGGTCTGCCAAACGAGGCGAAACCGCTTCAGTTGACTCGACCACCGGATCGGCCAAGGTC ------GACAAGCCAAAAACA------------------------ GCAATTCCAACCACCCATTTCGACACCCCTCGCCCCACAGAGACTGGAAAGGACTCCAGCTCCAATCGATCA CCC---------------------GTCAAATCT--------------------- GGAGGTGTCCGGCGGACCCGCACATTCACCTCGTCAGACAAACGACGCAACACCGTGGTTGTTAGCGGCAC GGGTGGAAGAATTGGCGTCTCCGGCGCCAATGGTGATGCTGACACCTCAAAGGCCCTTGTCCGAGTTCCTAA TATGGACAATATTGTAGAGTGTTTTGTGGACGATCTAGTCCATACGGAAGGTAGGTTCCGCCCTTGCTATGGT GCGTTTTGCTCCGTTTGTTTTTGTGTATTGACGCTCTGGGCTCAGTGGCTTGCGGTGCGCTGGCCTGCCCCTC GCTTCGCCGCACCATTCGGTGCCTCCACCGGGGACGGTTCCATCAGCTCCGTCCTGGTGGACCGCGAGATG ACATTTTCCAGGTGCCAGGTCGGCGACGAG------ TCCTATGCCGAAAGGGACGCCTCAGACAAGGAGTTGGCTGTC------------ GACGGTCTGGGAATTCACTCTTCCTCTAGTGGTCGGTGTGTTGGCGTTGGTTCGAGCGATCCGCGTTTGGCC AAGGCGCTGACGCAGGCGGCCACACTCACCGGTCCCGTTTCCGCCGACGCGCCGACGTCGTCTGAGAAGGA GGGCAGCGAGTCCACCGCCCTCAGACGCCATGATTCC--------------------- TCCACGGCCGAAAGCGTCTCTCAGCTTTCGACCACATCTCCGTCCTCCCAACTGGGAGCCGGCCGTGGTGTT GGTGGCAGCGCTCTCTTCTGTGGAAGCATGTCGCGCAGGCCCCAGCGCAACAAGGTTTCGCGT------------ CCAGCTATTCGTTTTGAGGCTGCCGACGAGATCATCTCCGCCGACCACCCTTCACACTCGACATCTAACGCCG AAGGCTCACCCCCCACCAACGTAGTCCCCGCCTTCATGCGCGGCGCTCCAGACCGCAGTACGGCTCCGGCG GCACGCTTAACGCATCGC------------------------ GTTGAATACTCACAGCGCTACCGGCCTCCCGAGGTCGCTAAACCTCGACCGTCCGAGGAAGAGACGGAGGC	VKLYEVIESERHVYLVMEFAANGELFEYLV SNGRMREKDARIKFRQIVSAVQYCHQKNIV HRDLKAENLLLDADYNIKLADFGFSNTFRP DKKLDTFCGSPPYAAPELFLGKKYTGPEV DVWSLGVILYTIVAGYLPFDAQNLRDLRER VLRGKYRIPFYMSTDCEMLLKRMLVLNPEK RHSLLSVMEDKWTNINMEDDILRPYQEPP PDYNDPVRLAKMTEMGFTLEEIKDSLENN KFNNVTATYLLLGTDRKNSSST---PPPPSS- --------------- SAAAANNQPAVVDESPKQKTSTQKESGA-- ----T------------ ALAAVEDSMVEAPDTPQVLVSSTSMRKGS STT----------RANGSA---TSTAPVTENP---- QSTPESTSS-----ASGK---- QKPTWSAKRGETASVDSTTGSAKV-- DKPKT-------- AIPTTHFDTPRPTETGKDSSSNRSP------- VKS------- GGVRRTRTFTSSDKRRNTVVVSGTGGRIG VSGANGDADTSKALVRVPNMDNIVECFVD DLVHTEGRFRPCYGAFCSVCFCVLTLWAQ WLAVRWPAPRFAAPFGASTGDGSISSVLV DREMTFSRCQVGDE-- SYAERDASDKELAV---- DGLGIHSSSSGRCVGVGSSDPRLAKALTQ AATLTGPVSADAPTSSEKEGSESTALRRH DS------- STAESVSQLSTTSPSSQLGAGRGVGGSAL FCGSMSRRPQRNKVSR---- PAIRFEAADEIISADHPSHSTSNAEGSPPTN VVPAFMRGAPDRSTAPAARLTHR-------- VEYSQRYRPPEVAKPRPSEEETEAEVEPE VVPETPIKETTKPTEYHS- SGSSAFIRSVSSRLSKSNRMPAAQS--------- ---------------- VEGGKGDKESGGGGGGHLVASDSPIPA- HRAFSS-

*Taenia solium**	TsM_000914500	TTTAAACTCTATGAGGTGATTGAATCCGATCGCCACGTCTATCTGGTAATGGAATTCGCCGCAAACGGTGAAC	FKLYEVIESDRHVYLVMEFAANGELFEYLV
		TTTTTGAATACCTTGTATCCAACGGTCGAATGCGTGAGAAGGATGCTCGCGTCAAGTTTCGTCAGATTGTCTCT	SNGRMREKDARVKFRQIVSAVQYCHQKNI
		GCGGTACAGTACTGCCACCAAAAAAACATCGTCCACCGTGATCTCAAGGCAGAGAATCTTCTCCTGGACGCC	VHRDLKAENLLLDADYNIKLADFGFSNTFR
		GATTATAACATCAAGTTGGCCGACTTTGGTTTTTCCAACACATTCCGCGCGGATAAGAAATTGGATACATTCTG	ADKKLDTFCGSPPYAAPELFLGKKYVGPE
		TGGGTCGCCACCATATGCGGCGCCCGAGTTGTTCCTTGGCAAGAAGTATGTCGGTCCTGAGGTGGATGTCTG	VDVWSLGVILYTIVAGYLPFDAQNLRDLRE
		GTCGTTGGGCGTCATCCTCTACACAATTGTTGCCGGTTATCTTCCCTTTGATGCACAGAACTTGCGGGATCTT	RVLRGKYRIPFFMSTDCEMLLKRMLVLNPE
		CGTGAACGCGTTCTCCGTGGCAAGTATCGGATCCCCTTTTTCATGTCTACCGACTGCGAAATGCTGCTAAAGC	KRYSLLSVMEDKWTNINMEDNVLRPYQEP
		GCATGCTCGTCCTTAACCCAGAGAAGAGGTATTCTCTCCTGTCTGTCATGGAGGACAAGTGGACGAATATCAA	APDFKDPARLAKMVEMGFTLEEIKDSLENN
		CATGGAGGACAATGTTCTCCGGCCCTACCAAGAACCAGCACCTGACTTTAAAGATCCAGCTAGGCTTGCCAAG	KFNNVTATYFLLGTDRPASSSSSLSHQLSS
		ATGGTCGAGATGGGCTTCACACTGGAGGAAATCAAGGACTCTCTGGAGAATAATAAGTTCAACAACGTGACGG	STSISNRPTTSITADDDSVSV--PDPS-------
		CGACCTATTTCCTCTTGGGGACGGATCGCCCCGCCTCCTCAAGCTCGTCCCTCTCCCACCAATTATCCTCTTC	KSVSTPRGSAA------
		CACTAGCATTTCCAACCGCCCCACCACATCCATCACTGCTGATGATGACTCCGTGTCCGTA------	TSTTAAIAVGGGSALTTGEDS-----
		CCCGATCCCTCC---------------------AAGTCGGTCTCCACACCCAGGGGATCTGCCGCC------------------	DTPHILVSSASMKKGSSST----------
		ACCAGCACTACAGCCGCTATTGCTGTTGGTGGTGGTTCTGCCCTGACCACCGGCGAAGATTCA---------------	HANGSA---NSTAPTAE-------STSDSTSS-----
		GACACCCCTCACATACTGGTCTCTTCAGCCTCCATGAAGAAGGGCTCCTCTTCCACG------------------------------	ANGK----
		CATGCAAACGGATCCGCT---------AACTCCACTGCTCCCACCGCGGAG---------------------	QKSAWSQKRGETVSVESTAATSGVRFDK
		TCTACCTCTGATTCTACCTCGTCT---------------GCCAACGGGAAG------------	PKV--------
		CAGAAATCTGCTTGGTCACAAAAACGCGGGGAGACTGTTTCTGTCGAATCGACAGCTGCCACCAGTGGCGTA	VVPTTHFDTPRSSEAARDSPTSRSP-------
		AGATTTGACAAGCCAAAAGTG------------------------	SKPSGAIGSGGGVRRTRTFTSADKRRNTV
		GTTGTTCCAACCACCCACTTTGATACTCCCAGGTCATCTGAAGCTGCAAGGGACTCTCCCACCAGTCGATCGC	AVGGPGG-----
		CG---------------------	GAECDSDVSKALVRVPNMDNIVD-------------
		TCGAAGCCGTCTGGTGCAATTGGTAGTGGCGGTGGTGTTCGGAGAACGCGAACCTTTACGTCTGCGGATAAA	---
		CGGCGCAACACAGTCGCTGTTGGCGGCCCTGGTGGT---------------	QVGDE--SGGERDASDRELDV----
		GGTGCTGAGTGTGACTCAGATGTCTCCAAGGCCCTTGTCCGTGTCCCAAATATGGACAACATTGTCGAT---------	DGLGVHSSTSGRCVAGGSSDLRLGKPLN
		--	QATTLPSPTS----
		--CAAGTTGGTGATGAG------	TRVRGDDSDSTTLKHQDS----------------------
		TCAGGAGGGGAGCGGGATGCATCGGACCGAGAACTGGATGTC------------	-----------------SMSRKQQRNK-SR----
		GACGGTTTGGGCGTTCACTCCTCGACTAGTGGGCGCTGTGTTGCTGGTGGCTCGAGCGATCTCCGGTTGGG	PAIRFEAADEVLSADHPSHATAKAE----
		AAAGCCTCTGAACCAAGCTACCACTCTGCCTAGTCCTACCTCC------------	PNTVPAFMRGAPDRSTAPAARVVHH---
		ACTCGCGTACGCGGAGACGATAGCGACTCCACCACCCTCAAACACCAGGACTCG--------------------------------------	ATNTPVEYSQRFRPP-PAPSSPTPHSSQ-
		---AGCATGTCACGGAAGCAGCAGCGTAACAAA---	QRQVEVGTESPVRVTTKPTEYHHH--
		TCACGT------------	SSSFLRSVSSRLSKSFRRKRDRSHSRTTP
		CCAGCGATCCGCTTTGAAGCAGCGGACGAGGTGCTTTCCGCCGATCACCCTTCACACGCGACAGCCAAAGCC	TQG----------ISRQATEERDEAGKQ-----
		GAG------------	GAHLVTTDSPIPAHHRAFSS-ERKH-------
		CCCAACACGGTACCAGCCTTCATGCGGGGCGCTCCAGACCGCAGCACCGCTCCGGCAGCTCGCGTTGTCCA	AVITT---
		CCAT---------GCCACCAACACTCCAGTGGAGTACTCTCAGCGGTTCAGACCCCCC---	DDPWSPMLPEKGHSEGEALVTNISNSLEP
		CCTGCCCCCTCCTCACCTACGCCGCATTCCTCCCAA---	LLPACFDKKPRNISVFSGTWRKA------
		CAGCGACAAGTGGAAGTGGGAACAGAATCGCCGGTGCGGGTGACCACAAAACCTACTGAGTACCACCACCA	SSSTTA---------TT-

Mothers against decapentaplegic homolog 4-like	*Crassostrea gigas*	gi|562890029|gb|AHB37077.1| smad4 protein [Crassostrea gigas]	ACGTCTAACGATGCCTGTCTGAGCATAGTCCATAGTCTGATGTGCCATAGACAGGGAGGAGAGAGTGAGAGC---TTCGCCAAGCGCGCCATCGAGAGTCTAGTGAAAAAACTGAAGGAAAAACGGGACGAGCTTGACAGCCTCATTACGGCAATCACCACCAATGGCGCACACTCCACGAAATGTGTGACCATTCAAAGGACCCTCGACGGACGGTTACAGGTGGCAGGAAGGAAGGGCTTTCCTCATGTCATTTACGCCCGTATTTGGAGATGGCCTGATCTTCACAAAAACGAGTTGAAACACTGCAAATTCTGTCATTATGCTTTTGACCTCAAGCAAGACAGCGTCTGTGTGAATCCATATCACTAT---------------------------------GAGAGGGTGGTTTCCCCTGGC---------ATA---------GATCTATCTGGACTGAAT------ATCCAAAATCCGCCAGAGTTTTGGTGTACCATTACTTACTTTGAGTTGGACCAGCAGGTCGGCGAAACGTTCAAAGTC------CCTTACAGCTACGCCCGGGTCACTGTGGACGGCTATACAGACCCATCCAGTCTGGACCGGTTCTGTCTGGGTCAGCTCTCAAATGTCCACCGGACCGAGACTAGTGATAAAGCCAGGTTACACATTGGTAAAGGTGTACAGTTAGATTACAACGGTGAGGGAGACGTGTGGATCCGCTGTGTCAGTGACCACAGTGTGTTTGTACAGTCCTACTATCTGGACAGAGAGGCAGGCAGACAGCCAGGGGATGCGGTCCACAAAATCTACCCCAGCGCTTATATCAAGGTGTTTGACATCCGTCAGTGTCACAGACAGATGCAGGAACAG---------GCTGCCACCGCACTTAGTGCTGCC---------GCTGGTATAGGAGTGGACGATCTTCGTCGCCTCTGT---ATCTTGCGCCTCAGCTTC---------------GTCAAGGGCTGGGGACCAGACTACCCTCGACACAGCATCAAGGAGACCCCATGTTGGATCGAAGTTCAGCTTCACCGACCACTCCAACTCCTGGACGAGGTT------------------------CTACAGACCATGCCA	TSNDACLSIVHSLMCHRQGGESES-FAKRAIESLVKKLKEKRDELDSLITAITTNGAHSTKCVTIQRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHCKFCHYAFDLKQDSVCVNPYHY-----------ERVVSPG---I---DLSGLN--IQNPPEFWCTITYFELDQQVGETFKV--PYSYARVTVDGYTDPSSLDRFCLGQLSNVHRTETSDKARLHIGKGVQLDYNGEGDVWIRCVSDHSVFVQSYYLDREAGRQPGDAVHKIYPSAYIKVFDIRQCHRQMQEQ---AATALSAA---AGIGVDDLRRLC---------------------ILRLSF-----VKGWGPDYPRHSIKETPCWIEVQLHRPLQLLDEV--------LQTMP

<table>
<tr><td>Echinococcus granulosus</td><td>gi|674565437|emb|CDS19986.1| Smad4 [Echinococcus granulosus]</td><td>ACCTCTTCAGACGCCTGTATGAACATCGTGCACAGCCTCATGTGTCACCGGAAGGGTGGCGAGTCGGAGGAG---TTTTCCAAATTTGCTATCGAAAGTCTCATTAAGAAGCTGAAGGATCGCAGGGACGAGCTGGATGCGCTCATCGCCGCGGTGACTAGCAATGGAGCCACACAGACCAGCTGCGTGACCATCCAGCGAACCCTCGACAGTCGAATGCAGATTGCTGGTCGGAAGTGCTTTCCTCATCTCATTTACGCTCGACTCTGGCGGTGGTCGGATGCCCACAAAACGGAGCTGCGTCACCTGCCTTTTTGCCACTTCGGGTTTGACAAGAAGCTTGACTGGGTGTGCGTCAACCCCTATCACTAC--------------------------------GAGCGCACCGTCTCCTCCGCT---------CTC---------GACATCTCGTCGCTGGCT------CTCCAGAGGCCGCCGGAGTACTGGTGCAACATAGCCTACTTCGAGTTGGATCAGCAGGTGGGCGAGTTGTTCAAGGTG------CCCAGTCACTACACACGAGTAATTGTGGACGGCTATACCGACCCCTCCAGCCGAAATCGCTTCTGTCTGGGCCAGCTATCCAACGTGCACAGGTCGGAGCAGTCGGAGAAGTCGCGTCTCTACATTGGGAAGGGCGTGGAGCTAGACATAGTGGGCGAAGGTGACGTCTGGATCCGCTGTCTCTCCGAGTTCTCCATCTTTGTACAAAGCTACTACCTTGACCGCGAGGCAGGCAGGGCACCGGGTGATGCTGTGCACAAAATTTATCCCGGTGCTTACATTAAGGTGTTCGACATACGTCAGTGCCACGAACAGATGCGTCATCTG---------GCCCACATGACACCAATGGGCACATGCGAAGCCGCCGGGGTAGGCGTGGACGACTTCCGTCGACTCTGT---AACCTTCGTCTCAGTTTC---------------GTCAAGGGCTGGGGTCCGGACTATCCCCGCCACGACATTAAGGAGACTCCCTGTTGGATTGAAATCCAACTTCACAGACCACTGCAACTACTGGACGAGGTT------------------------TTGCAAGCAATGCCA</td><td>TSSDACMNIVHSLMCHRKGGESEE-FSKFAIESLIKKLKDRRDELDALIAAVTSNGATQTSCVTIQRTLDSRMQIAGRKCFPHLIYARLWRWSDAHKTELRHLPFCHFGFDKKLDWVCVNPYHY-----------ERTVSSA---L---DISSLA--LQRPPEYWCNIAYFELDQQVGELFKV--PSHYTRVIVDGYTDPSSRNRFCLGQLSNVHRSEQSEKSRLYIGKGVELDIVGEGDVWIRCLSEFSIFVQSYYLDREAGRAPGDAVHKIYPGAYIKVFDIRQCHEQMRHL---AHMTPMGTCEAAGVGVDDFRRLC---------------------NLRLSF-----VKGWGPDYPRHDIKETPCWIEIQLHRPLQLLDEV--------LQAMP</td></tr>
</table>

	Echinococcus multilocularis	gi|674265824| emb|CDI98328.1| Smad4 [Echinococcus multilocularis]	ACTTCTTCGGACGCCTGCATGAACATCGTGCACAGCCTCATGTGTCACCGGAAGGGTGGCGAGTCGGAGGAG---TTTTCCAAATTTGCTATCGAAAGTCTCATTAAGAAGCTGAAGGATCGCAGGGACGAGCTGGATGCGCTCATCGTCGCGGTGACTAGCAATGGAGCCACACAGACCAGCTGCGTGACCATCCAGCGAACCCTCGACAGTCGAATGCAGATTGCTGGTCGAAAGTGCTTTCCTCATCTTATTTACGCGCGACTCTGGCGGTGGTCGGATGCCCACAAAACGGAGCTGCGTCACCTGCCTTTTTGCCACTTCGGGTTTGACAAGAAGCTTGACTGGGTGTGCGTCAACCCCTATCACTAC--------------------------------GAGCGCACCGTCTCCTCCGCT---------CTC---------GACATCTCGTCGCTGGCT------CTCCAGAGGCCACCGGAGTACTGGTGCAACATAGCCTACTTCGAGTTGGATCAGCAGGTGGGCGAGTTGTTTAAGGTG------CCCAGTCACTACACACGAGTAATTGTGGACGGCTATACCGACCCCTCCAGCCGAAATCGCTTCTGTCTGGGCCAGCTCTCCAACGTGCACAGGTCGGAGCAGTCGGAGAAGTCGCGTCTCTACATTGGGAAGGGCGTGGAGCTAGACATAGTGGGCGAAGGTGACGTCTGGATCCGCTGTCTCTCCGAGTTCTCCATCTTTGTACAAAGCTACTACCTTGACCGCGAGGCAGGCAGGGCACCGGGTGATGCTGTGCACAAAATTTATCCCGGTGCTTACATTAAAGTGTTCGACATTCGTCAGTGCCACGAACAGATGCGTCATCTG---------GCCCACATGACACCAATGGGCACATGCGAAGCCGCAGGGGTAGGCGTGGACGACTTCCGTCGACTCTGT---AACCTTCGTCTCAGTTTC---------------GTCAAGGGCTGGGGTCCGGACTATCCCCGGCACGACATTAAGGAGACTCCATGTTGGATTGAAATCCAACTTCACCGACCACTGCAACTACTGGACGAGGTT------------------------TTGCAAGCAATGCCA	TSSDACMNIVHSLMCHRKGGESEE-FSKFAIESLIKKLKDRRDELDALIVAVTSNGATQTSCVTIQRTLDSRMQIAGRKCFPHLIYARLWRWSDAHKTELRHLPFCHFGFDKKLDWVCVNPYHY-----------ERTVSSA---L---DISSLA--LQRPPEYWCNIAYFELDQQVGELFKV--PSHYTRVIVDGYTDPSSRNRFCLGQLSNVHRSEQSEKSRLYIGKGVELDIVGEGDVWIRCLSEFSIFVQSYYLDREAGRAPGDAVHKIYPGAYIKVFDIRQCHEQMRHL---AHMTPMGTCEAAGVGVDDFRRLC---------------------NLRLSF-----VKGWGPDYPRHDIKETPCWIEIQLHRPLQLLDEV--------LQAMP

	*Globodera pallida**	GPLIN_000772200	TCCAGCGGCGACGCCAACACAACAATCACTCAATTTTTGATGAATTATGTGGTGGGATCGGAT---CGCGAG--- TTCAACAAAAAGGCCATCGAAAGTTTAATCAAAAAACTGAAGGACAAAGGCGACGAATTGGACGATTTTATCGC TTCGGTTAGCGCAATGGGCCAATTGTGCACCAAATGCGTGACGACGCCGCGCACTTTGGACGGCCGTCTACA GGTGGCAGGACGAAAAGGCTTTCCGCACGTCGTTTACTCGAAGATCTTTCGCTGGCCCGACCTGCACAAGAA CGAGCTGAAGCACAAAAACTTTTGCATTTATGCGTTTGACCTCAAAAAGGACCAGGTTTGTGTGAACCCGTAC CATTAC--------------------------------GAGCGAGTCGTT-----------------------------------TCTGCGGAAGGCATTGAGTTG-- ------- CCACTCCATTGGTTAGCGGCCAATTACTACGAATTCGACCGCAGAATCGGCGAAACGTTTCAAGCGGTCGCG- ----- GAATGTCCTCAGATTTTCGTGGACGGCGGTTTGGACAGCACGGGAAACGCCCGCTTCTGTTTGGGTCCGTTG ACCAACACTGAACGCGGGGAGGCGGCCGAAAAGTGCAGGCGAAACATTGGCCTCGGCATTCGATTGGATCT GAAGGGCGAAGGCGACGTTTGGCTGACGGTCCTCTCTAAAGGGCCCGTGTTCGTGCAGAGTCATTATTTGGA CGTGCTCACGGAACGCGAGGAATTGGGCCACGCTCACAAATTTGTCCAATACACCACCGTTAAGATTTTCGAC CTGTTCAAGTGCTACGAGTGTTGGAAAGTGACCCATTTGGAGCGAATTATGGCC----------------------- GATCCGGGCGTGGATGATTTTCGCACTCTGTGC-- ACCATGCGCATTTCGTTC--------------- TTCAAGGGATTTGGTCTCAGCTATCCAAAACGGACAATTCAGGAAACGCCTTGCTGG--------------------- GCTCTTCAATTGTTGGACGAAGTGATGAATACTCCGCTGATCGATCACTTGACT---------	SSGDANTTITQFLMNYVVGSD-RE- FNKKAIESLIKKLKDKGDELDDFIASVSAMG QLCTKCVTTPRTLDGRLQVAGRKGFPHVV YSKIFRWPDLHKNELKHKNFCIYAFDLKKD QVCVNPYHY-----------ERVV------------ SAEGIEL--- PLHWLAANYYEFDRRIGETFQAVA-- ECPQIFVDGGLDSTGNARFCLGPLTNTER GEAAEKCRRNIGLGIRLDLKGEGDVWLTVL SKGPVFVQSHYLDVLTEREELGHAHKFVQ YTTVKIFDLFKCYECWKVTHLERIMA-------- DPGVDDFRTLC---------------------TMRISF----- FKGFGLSYPKRTIQETPCW------- ALQLLDEVMNTPLIDHLT---

	Helobdella robusta	gi|675858096| ref|XP_009014608.1| hypothetical protein HELRODRAFT_76435 [Helobdella robusta]	ACCTCAAGTGATGCATGCATCAACATTGTGCAGAGTTTAATGTTCTACAGACAGGGTGAGGATGCAGAGGCAG GTTTCTCAAAGAAGGCCATAGAAAGTCTGGTGAAGAAGTTGAAGGAGAAGAGGGAGGAGCTGGACAACTTGA TCGCAGCTATCACATCAAATGGAAGTCAGCCAACGAAATGTGTCACCATACCGAGGACGCTGGACGGCAGAC TACAGGTGGCCAGTCGAAAAGGTTTTCCCCACGTCATATACTCTCGGATATGGAGATGGCCCGACGTACATAA GAATGAGCTGAAGCACCTGAAGTTCTGTCAGTTTGCGTTCGACCTGAAGCAAGATGGCATCTGCGTCAATCCT TATCATTAT---------------------------------GAACGCGTCCCACCCACTGGA---------TTA--------- GAATCAAGTTACATGGTT------ GATCACCCAGTACCAGAATTTTGGTGCAGCATAACTTACTTTGAACTGGATCAAAAGGTAGGTGAGATCTTCAA AGTT------ CCATCGTCGAATCACACAATCTCAGTTGACGGATACACGGATCCATCTAGTCTGGATAGGTTTTGTTTGGGCA AGTTAACAAACGTTCACAGAACTGAATCAATTGAAAAAGCAAGGTTATACATTGGCAAAGGGGTGCAGCTGGT GTTAGAAGGAGAGGGTGATGTGTGGGTGAGGTGTCTCAGTGAGCACAGCATTTTCGTTCAGAGTTTCTACCTG GACAGAGAGGCAGGGAGAGCACCAGGAGATGCGGTTCATAAGATTTATCCCGCTGCTTTCATTAAAGTATTCG ATTTAAGTCAGTGTCAGAGTCAAATGCAGCAACTG---------GTGGTTCAGGCGTTATCACCGGCA--------- GCAAGCATTGGAGCCGACGATCTCAGACGATTGTGC--- GTGCTACGTTTAAGTTTT--------------- GTGAAGGGCTGGGGTCTGGACTACCCACGACCGACTATAAAGGACACGCCCTGTTGGATTGAGGTGCAGTTG AATCGGCCCCTGCAATTTCTTGACGAGTTT------------------------CTTCAAGCTATGCCG	TSSDACINIVQSLMFYRQGEDAEAGFSKKA IESLVKKLKEKREELDNLIAAITSNGSQPTK CVTIPRTLDGRLQVASRKGFPHVIYSRIWR WPDVHKNELKHLKFCQFAFDLKQDGICVN PYHY-----------ERVPPTG---L---ESSYMV-- DHPVPEFWCSITYFELDQKVGEIFKV-- PSSNHTISVDGYTDPSSLDRFCLGKLTNVH RTESIEKARLYIGKGVQLVLEGEGDVWVR CLSEHSIFVQSFYLDREAGRAPGDAVHKIY PAAFIKVFDLSQCQSQMQQL--- VVQALSPA---ASIGADDLRRLC----------------- ----VLRLSF----- VKGWGLDYPRPTIKDTPCWIEVQLNRPLQ FLDEF--------LQAMP

	Hymenolepis microstoma	gi|674594093|emb|CDS27166.1| Smad4 [Hymenolepis microstoma]	TCATCTTCAGACGCCTGTATGAACATTGTTCATAGTCTAATGTGCCATCGTAAGAATGGTGAATCAGAAGAG---TTTTCCAAATTCGCTATCGAAAGTCTAATAAAGAAACTCAAGGATCGTAGGGAGGAGTTGGACTCTCTTATCGTTGCAGTTACTAGTAATGGAGCTACCCAAAGTGGATGCGTTACCATTCAAAGGACGCTTGACAGTCGAATGCAGATTGCCGGCCGAAAGTGTTTCCCTCATCTTATCTACGCTCGTCTTTGGCGTTGGTCAGATGCTCACAAAACAGAGTTGCGCCATCTACCCTTCTGCCATTATGGTTTTGACAAAAAACTCGATTGGGTTTGCGTAAACCCCTATCACTAT--------------------------------GAACGAATTGTTTCATCAGCT---------TTG---------GATATCTCATCTTTAGCA------TTACAGAGGCCTCCTGAGTATTGGTGCAACATCGCTTACTTTGAGTTGGACCAGCAAGTTGGCGAGCTCTTTAAGGTT------CCATCTCACTACACCCGTGTCATTGTGGACGGCTATACAGATCCCTCGAGTCGCAATCGCTTCTGCCTTGGTCAACTCTCCAATGTCCACAGATGTCCAATGTTCACTAATGCGCATAGATTT--ACACGGGTGTCCATCTTTGCTCCTCTCACTTGTCCTTCGATTGAGGCC--ACCTGTCAGTCACTGACGGTCTCACACACTGGGCGTGTCAACAATGGAGTAGGAGATGGGAGTGGCAAGAGATTGATTCCCTCTTTTCGCTCACATCTTGATNNN--	SSSDACMNIVHSLMCHRKNGESEE-FSKFAIESLIKKLKDRREELDSLIVAVTSNGATQSGCVTIQRTLDSRMQIAGRKCFPHLIYARLWRWSDAHKTELRHLPFCHYGFDKKLDWVCVNPYHY-----------ERIVSSA---L---DISSLA--LQRPPEYWCNIAYFELDQQVGELFKV--PSHYTRVIVDGYTDPSSRNRFCLGQLSNVHRCPMFTNAHRF--------------------TRVSIFAPLTCPSIEA--TCQSLTVSHTGRVNNGVGDGSGKRLIPSFRSHLD?--

	Lingula anatina	gi|919043938|ref|XP_013405146.1| PREDICTED: mothers against decapentaplegic homolog 4-like isoform X1 [Lingula anatina]	ACCTCTGCTGATGCCTGCCTCAGCATTGTCCATAGCCTCATGTGTCACCGTCAGGGTGGGGAGAGTGAAAGC---TTTGCTAAAAGAGCCATCGAAAGCTTGGTGAAAAAGCTGAAAGAAAAGCGCGATGAACTGGACAGTTTGATTACTGCTATTACGACAAATGGTGCTCATCCAACAAAGTGCGTTACAATTCAGCGGACGCTTGATGGAAGATTGCAGGTGGCTGGCAGGAAAGGTTTCCCTCATGTTATCTACGCCAGGATTTGGCGCTGGCCAGATTTGCACAAAAATGAACTGAAGCATGCCAAGTTTTGCCAGTTTGCATTTGATCTGAAACAGGACAGTGTCTGTGTCAATCCATATCACTAT---------------------------------GAGAGGGTGGTGTCTCCAGGG---------ATA---------GATTTGTCAGGCTTATCA------CTACAGCCAATGCCCGAGTTCTGGTGTACCATAGCCTATTTTGAGCTGGATCAGCAGGTGGGTGAAACCTTCAAAGTG------TCCAGCAGCTGTATGACAGTGACAGTGGATGGTTACACAGACCCTTCCAGTATTGACAGGTTCTGCCTGGGACAGCTGTCTAATGTGCACAGAACAGAGGCCAGTGAGAGGGCAAGGTTACATATTGGCAAGGGCGTTCAGCTTGACTTACGCGGAGAGGGAGATGTTTGGATCAGGTGCCTCAGTGATCACAGCGTATTTGTACAGAGTTACTACTTGGACAGGGAGGCAGGCAGGGCGCCAGGAGACGCAGTTCATAAAATCTACCCAAGTGCCTACATAAAGGTGTTTGACATACGTCAGTGTCACCATCAGATGCAGCAACAA---------GCAGCTACCGCACTGTCGGCAGCA---------GCAGGCATTGGGGTAGATGACCTAAGGAGATTATGC--ATTTTACGGCTTAGCTTT---------------GTGAAAGGCTGGGGACCTGACTACCGCCGCCACAGCATCAAAGAGACGCCATGTTGGATTGAGGTGCAGTTGCATCGCCCTCTACAGTTGTTGGACGAAGTA------------------------TTACAAGCAATGCCA	TSADACLSIVHSLMCHRQGGESES-FAKRAIESLVKKLKEKRDELDSLITAITTNGAHPTKCVTIQRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHAKFCQFAFDLKQDSVCVNPYHY-----------ERVVSPG---I---DLSGLS--LQPMPEFWCTIAYFELDQQVGETFKV--SSSCMTVTVDGYTDPSSIDRFCLGQLSNVHRTEASERARLHIGKGVQLDLRGEGDVWIRCLSDHSVFVQSYYLDREAGRAPGDAVHKIYPSAYIKVFDIRQCHHQMQQQ---AATALSAA---AGIGVDDLRRLC---------------------ILRLSF-----VKGWGPDYRRHSIKETPCWIEVQLHRPLQLLDEV--------LQAMP

<table>
<tr><th></th><th></th><th></th><th></th><th></th></tr>
<tr><td></td><td>*Lollita gigantea*</td><td>gi|676487206|ref|XP_009064219.1| hypothetical protein LOTGIDRAFT_131078 [Lottia gigantea]</td><td>ACATCTGCAGATGCATGTTTAAGTATTGTTCATAGTTTGATGTGTCATAGACAGGGTGGTGAAAGTGAAAGT---TTTGGCAAAAGAGCAATAGAAAGTTTGGTTAAGAAATTAAAAGAGAAACGGGATGAATTAGATAGTTTGATAACAGCTATAACAACAAATGGGGCTTTACCATCAAAATGTGTTACTATACAGAGGACGTTAGATGGTAGATTGCAGGTTGCTGGTAGAAAAGGTTTCCCTCATGTCATTTATGCTAGAATATGGAGATGGCCAGATTTGCATAAAAATGAGTTGAAACATGTCAAGTTTTGTCAATATGCCTTTGATTTAAAACAAGACAGTGTTTGTGTTAATCCATACCATTAT--------------------------------GAAAGAGTTGTTTCCCCTGGA---------ATT---------GATTTATCAGGTTTAACT------CTA---------CCAGAATTTTGGTGTACCATAACATATTTTGAGTTAGATCAGCAAGTTGGTGAAACATTCAAAGTA------CCATATAGTTGTTCTACTGTCACTGTTGATGGTTATACTGATCCTTCTAGTATTGATAGATTCTGTTTAGGACAGTTATCCAATGTTCATCGTACTGAGGCTAGTGAAAGAGCTCGATTACATATAGGTAAAGGAGTTCAGTTAGATTATCGGGGTGAAGGTGATGTATGGATAAGATGTGTTAGTGATCATAGTGTCTTTGTACAGAGTTATTATTTAGATAGAGAAGCTGGTCGAGCTCCTGGTGATGCTGTACATAAAATTTATCCTAGTGCTTATATAAAAGTATTTGATATTCGACAGTGTCATCGTCAAATGCAGCAGCAA---------GCAGCTACAGCACTGTCAGCTGCA---------GCTGGTATTGGTGTAGATGATTTAAGACGTTTGTGT--ATTTTACGATTAAGTTTT---------------GTGAAAGGTTGGGGACCTGATTACCCTCGCAAAAGTATCAAAGAAACTCCATGTTGGATTGAAGTCCAATTACATCGTCCTTTGCAGCTGTTAGATGAAGTT------------------------CTTCAAGCCATGCCA</td><td>TSADACLSIVHSLMCHRQGGESES-FGKRAIESLVKKLKEKRDELDSLITAITTNGALPSKCVTIQRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHVKFCQYAFDLKQDSVCVNPYHY-----------ERVVSPG---I---DLSGLT--L---PEFWCTITYFELDQQVGETFKV--PYSCSTVTVDGYTDPSSIDRFCLGQLSNVHRTEASERARLHIGKGVQLDYRGEGDVWIRCVSDHSVFVQSYYLDREAGRAPGDAVHKIYPSAYIKVFDIRQCHRQMQQQ---AATALSAA---AGIGVDDLRRLC-----------------ILRLSF-----VKGWGPDYPRKSIKETPCWIEVQLHRPLQLLDEV--------LQAMP</td></tr>
<tr><td></td><td>*Mesocestoides corti**</td><td>MCOS_0000775001-mRNA-1</td><td>ACCAGTTCGGACGCCTGCATGAACATTGTTCACAGCCTCATGTGCCACCGAAAAGGTGGCGAATCTGAAGAA---TTTTCGAAATTCGCTATTGAAAGTCTAATAAAAAAACTAAAAGACCGGAGAGATGAGTTAGATGCTCTTATTGTCGCCGTAACTAGCAACGGTGCTACGCAAACCAGCTGTGTAACAATTCAAAGAACCCTTGACAGTCGGATGCAGATTGCTGGTCGAAAGTGTTTTCCCCACCTTATTTACGCTCGGCTTTGGCGATGGTCTGATGTTCACAAGACAGAGCTGCGCCACCTGCCATTTTGTCACTTTGGGTTCGACAAGAAACTTGACTGGGTATGTGTCAATCCCTATCACTAT---------------------------------GAGCGCACGGTTTCCTCTGCT---------CTC---------GACATCTCGTCATTAGCT------CTC------------------------------GCCTTTTTC---TCTGTCGAGGCC---CTAAGC---AAATTTATTGTACCATTT---------------NNN---</td><td>TSSDACMNIVHSLMCHRKGGESEE-FSKFAIESLIKKLKDRRDELDALIVAVTSNGATQTSCVTIQRTLDSRMQIAGRKCFPHLIYARLWRWSDVHKTELRHLPFCHFGFDKKLDWVCVNPYHY-----------ERTVSSA---L---DISSLA--L----------AFF---SVEA--LS---------------------KFIVPF-----?--</td></tr>
</table>

*Onchocerca volvulus**	OVOC4813 OVP10163 WBGene00241622	ACAAGTTCTGATGCATGTGCTACAATCACACAGTATCTGATGATGTATCATACGGGCCGGGAT---GAAGAG---TTTAGTCGCAAAGCTATTGAAAGTCTGATTAAAAAACTGAAAGATAAACGTGATGAATTGGATGCACTAATCACCACTGTAGCATCACATGGGAAGATTTCTCCGAAATGTATCACCATTCAGCGTACTTTGGATGGACGACTGCAGGTGGCTGGAAGAAAAGGCTTTCCGCACGTCGTTTACGCACGTATTTGGAGATGGCCTGATTTGCATAAGAATGAGCTGAAACATTTGCCAATTTGTCAGTGTGCGTTTGACCTGAAATGTGACCTTGTTTGCGTTAATCCATATCATTAC--------------------------------GAGCGTGTTGTTCCGCCAGGTATCGGCACTATT---------GATTTGTCTAATTTAAAAATTGAACAC---------CCGGCAAATTGGTGTGTCATTTCTTATTATGAGTTCAACACAAAAGTTGGTGAAACATTTGCTGTA------------AGTGCACCTGCGGTCTACATTGATGGAGGAGTTGATCCATCTGCTCCTGGTCGTTTTTGTTTAGGATCTCTTTCCAACGTTCAGCGTACTGATGAGAGTGAGCGATGCAGGAAGCATATTGGTCGTGGGATTCGATTGGACGTGAAAGGAGAGGGTGATGTTTGGTTGACATGCTTGTCGGATCGGCCAGTTTTTGTTCAAAGTTCGTATCTGGATCGCGAAGCAGGGCGAGTTCCAGGCGATGCTGTGCACAAGATTTATTCCCAAGCAACACTAAAAGTTTTCGACTTGCGTCAGTGTTACCATCAGCTACGACAGCAGAATATGTACCAGCTGATAGCCTTAAATCAGGCT---------GCAAATGTTGGTGTCGACGAATTGCGGAACTTGTGT--TCGCTAGCAGTATCTTTT---------------GTTAAAGGATGGGGGCCAGATTATGATCGTAAATCAATTAAAGAAACACCGTGTTGGATTGAAGTCCAGATAAACAGGGCTCTTCAACTGCTTGATGAAGTTCTTCATAATCCA---------ACTTTAAACAATTTTCCT	TSSDACATITQYLMMYHTGRD-EE-FSRKAIESLIKKLKDKRDELDALITTVASHGKISPKCITIQRTLDGRLQVAGRKGFPHVVYARIWRWPDLHKNELKHLPICQCAFDLKCDLVCVNPYHY-----------ERVVPPGIGTI---DLSNLKIEH---PANWCVISYYEFNTKVGETFAV----SAPAVYIDGGVDPSAPGRFCLGSLSNVQRTDESERCRKHIGRGIRLDVKGEGDVWLTCLSDRPVFVQSSYLDREAGRVPGDAVHKIYSQATLKVFDLRQCYHQLRQQNMYQLIALNQA---ANVGVDELRNLC--------------------SLAVSF-----VKGWGPDYDRKSIKETPCWIEVQINRALQLLDEVLHNP---TLNNFP

	Pinctada fucata	gi|552954716|gb|AGY49100.1| TGF beta signaling pathway factor [Pinctada fucata]	ACGTCAGCCGACGCTTGTCTTAGTATAGTACACAGTCTTATGTGTCACAGACAAGGAGGAGAAAGTGAAAGC---TTTGCCAAACGAGCCATAGAAAGTCTGGTTAAGAAGCTTAAAGAGAAAAGAGATGAGTTGGACAGTCTTATTACAGCTATAACAACAAATGGAGCCCATCCTACCAAGTGTGTCACCATACAAAGAACTCTGGATGGCAGGTTACAGGTTGCTGGTAGGAAAGGATTTCCCCATGTTATTTATGCCAGGATTTGGAGATGGCCAGACCTACATAAGAATGAGCTGAAACATGCCAAGTTCTGTCAGTATGCGTTTGACCTTAAACAGGACAGTGTCTGTGTTAATCCATATCATTAT---------------------------------GAGAGAGTTGTTTCACCAGGC---------ATT---------GATTTGTCTGGTTTAACC------ATACAAAATCCTCCAGAGTTCTGGTGTACGATAACATATTTTGAGTTAGATCAGCAAGTGGGGGAGACATTTAAGGTA------CCGTACAGTTATACCACAGTGACAGTCGATGGTTACACAGATCCTTCTAGCCTGGATAGGTTCTGTTTGGGTCAGCTGTCCAATGTACATAGAACGGATACTAGTGAAAAAGCCAGACTACACATAGGTAAAGGTGTACAGTTAGATTACCGTGGTGAGGGAGACGTGTGGATACGATGTGTCAGTGATCATAGCGTGTTTGTACAGAGTTATTACCCGGACAGGGAGGCAGGTAGATCACCTGGTGATGCAGTACACAAAATATATCCCAGTGCTTACATTAAGGTATTTGATATTCGTCAGTGTCACCGTCAGATGCAGCAGCAG---------GCGGCTACGGCATTAAGTGCCGCG---------GCAGGTATAGGAGTAGATGATCTCAGGAGGTTGTGT---ATATTAAGACTGAGTTTT---------------GTAAAAGGTTGGGGACCAGACTACCCCCGTCACAGCATTAAGGAGACTCCATGTTGGATAGAAGTTCAACTACACCGACCTCTACAACTTCTAGATGAAGTT------------------------CTACAGACCATGCCA	TSADACLSIVHSLMCHRQGGESES-FAKRAIESLVKKLKEKRDELDSLITAITTNGAHPTKCVTIQRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHAKFCQYAFDLKQDSVCVNPYHY-----------ERVVSPG---I---DLSGLT--IQNPPEFWCTITYFELDQQVGETFKV--PYSYTTVTVDGYTDPSSLDRFCLGQLSNVHRTDTSEKARLHIGKGVQLDYRGEGDVWIRCVSDHSVFVQSYYPDREAGRSPGDAVHKIYPSAYIKVFDIRQCHRQMQQQ---AATALSAA---AGIGVDDLRRLC------------------ILRLSF-----VKGWGPDYPRHSIKETPCWIEVQLHRPLQLLDEV--------LQTMP

Schistosoma haematobium	gi|844874982| ref|XP_012801409.1| Mothers against decapentaplegic 4 [Schistosoma haematobium]	--------------------------------------ATGTGTCATCGGCAGAGTGGAGAATCTGAAGAG---TTTGCAAGACGCGCGATTGAGAGTCTTGTGAAAAAATTGAAAGAAAGGCAGGAAGACTTGGAATCCTTGATT--GCTGGGCGGAAATGTCTTCCTCACATTATTTATTCACGAATATGGCGCTGGCCTGATTTGCATCGAAATGAGTTAAGACACTCCAAACATTGTTTATTTGGATTTGAACTGAAACAAGATTGCGTGTGTATTAATCCTTACCACTAT--------------------------------GAAAGAGTTGTTTCTCCT------------GTT---------GACTTGGCTACTCTGAGC------CTTCAACGTCCTCCTGAATACTGGTGCAACATTGCTTACTTTGAACTGGACCAACAGGTCGGTGAATTGTTTAAAGTT------CCAAGCCAGTATTCACGTGTCACCGTTGACGGTTACACGGATCCTTCTAGTCCAAATCGTTTTTGTCTCGGACAACTATCAAATGTACACCGGTCGGAGCAATCCGAAAAGTCGAGACTTTATATTGGCAAAGGTGTAGAACTAGATAACGTTGGAGAAGGTGATGTTTGGATTCGCTGTCTTTCAGAGTTCTCTGTATTCGTACAAAGCTATTATTTGGATAGAGAGGCTGGCCGTGCACCTGGTGACGCTGTCCATAAAATATATCCTGGTGCTTATATCAAGGTTTTCGATATAAGACAGTGTCATGAAGAAATGAAATCCCTT---------GCTCAGTCTTCTATTATGGCTACT---------GCTGGGGTCGGGGTGGATGATCTTAGACGTCTTTGT---ATGCTTCGGTTGAGTTTC---------------GTTAAAGGATGGGGACCAGATTATCCTAGACGTAGCATCAAAGAAACACCTTGTTGGATTGAAATACAGTTGCACAGACCACTACAATTGCTGGATGAAGTT-----------------------CTCCAGGCTATGCCT	-------------MCHRQSGESEE-FARRAIESLVKKLKERQEDLESLI--------------------------AGRKCLPHIIYSRIWRWPDLHRNELRHSKHCLFGFELKQDCVCINPYHY-----------ERVVSP----V---DLATLS--LQRPPEYWCNIAYFELDQQVGELFKV--PSQYSRVTVDGYTDPSSPNRFCLGQLSNVHRSEQSEKSRLYIGKGVELDNVGEGDVWIRCLSEFSVFVQSYYLDREAGRAPGDAVHKIYPGAYIKVFDIRQCHEEMKSL---AQSSIMAT---AGVGVDDLRRLC--------------------MLRLSF-----VKGWGPDYPRRSIKETPCWIEIQLHRPLQLLDEV--------LQAMP

	Strongyloides ratti	gi|685834770| emb|CEF6969 1.1| Smad4 protein [Strongyloides ratti]	TTATCATCAGACACGTGCTCCACAATTGCTTCATATTTAATGGAATATAATGTGTCTGATGAT---ATTGAA--- TTTTCAAGAAAAGCTATTGAATCATTAATTAAAAAATTAAAAGATAACTCATCACAATTAGATGAACTTATTAATT CTATTTCTTCTAAAGGATCAATTACAACAAAATGTATAACTATACCAAGAACATTAGATGGTAGACTTCAAGTTG CTGGAAGAAAAGGTTTTCCTCATGTTGTTTATGTTAAAACTTTTGTCTATCCAGATGCTTCTAAAAATGATCTTA AACATAAAGATATATGTCAAAATGGTTTTGATGAAAAAACAGAACAAGTTTGTGTTAATCCATATCATTATGATA AAGCATGTAATTTTTTAACACGAAAAGAAGAAAGAGTACCA------------------------------------ TCAAATGAAAGTATAATGTCA--------- CCTGAAATACAATTATCAATAACATATTATGAGTATCATAAAGTTTTATGTGATACAGTTAATGTT------------ GATATTATACCATATTATGTTGATGGTGGTCTTAATATATCATCAAGAAATCGTTTATGTTTAGGTGCTATAACAA ATGTTCTTAGAGAAGTTAGTACAGATAAGGTACTTCAATCAATTGGTAGAGGTGTACGTTTTGATGTCAAAGGT GAAGGTAATATATGGATCTCAAATCTTTCAAATCATCCAGTTTTTGTACAAAGTAATTATTTAGATGGTGATTCT GAAACT--------- GGTATTGTTTATAAAATATCACCATTAGCTACATTTAAAGTTTTTGATCTTGATCATTGTTATAGACAATTAAAAA GAATTAATATGTATAAAAATTTAGCTTTAACTAAAGCT--------- ATTGATACTGGTGTTGATGATATGCGTAATATATGT--- TCAATAAAATTATCATTT--------------- GTTAAAGGATGGGGTGAAGGATATGGAAGAGAAAGAATCTCTGAAGTTCCTTGTTGGATTGATGTTACTGTTAA TAGAGCACTCCAAATATTAGATCATATATTAAATTCTCCT---------AACCTCAAA---------	LSSDTCSTIASYLMEYNVSDD-IE- FSRKAIESLIKKLKDNSSQLDELINSISSKGS ITTKCITIPRTLDGRLQVAGRKGFPHVVYVK TFVYPDASKNDLKHKDICQNGFDEKTEQV CVNPYHYDKACNFLTRKEERVP------------ SNESIMS---PEIQLSITYYEYHKVLCDTVNV- --- DIIPYYVDGGLNISSRNRLCLGAITNVLREV STDKVLQSIGRGVRFDVKGEGNIWISNLSN HPVFVQSNYLDGDSET--- GIVYKISPLATFKVFDLDHCYRQLKRINMYK NLALTKA---IDTGVDDMRNIC-------------------- -SIKLSF----- VKGWGEGYGRERISEVPCWIDVTVNRALQ ILDHILNSP---NLK---

*Taenia solium**	TsM_000635600	ACCTCTTCAGACGCCTGTATGAACATCGTGCACAGCCTTATGTGTCACCGAAAGGGTGGCGAGTCGGAGGAG---TTCTCCAAATTCGCCATTGAAAGTCTTATTAAGAAGCTGAAGGATCGCAGGGACGAGTTAGATGCGCTTATCGTTGCGGTGACTAGCAATGGGGCCACGCAGACCAGCTGCGTGACCATCCAGCGAACTCTCGACAGTCGAATGCAGATTGCTGGTCGAAAGTGCTTTCCCCATCTCATCTACGCCCGCCTCTGGCGGTGGTCGGATGCCCATAAAACAGAGCTGCGCCACCTACCTTTTTGTCACTTCGGATTTGACAAGAAGCTTGACTGGGTTTGCGTCAACCCCTACCCACTAC--------------------------------GAGCGCACCGTCTCCTCTGCT---------CTC---------GACATCTCATCGCTGGCT------CTCCAAAGGCCGCCGGAGTACTGGTGCAACATAGCCTACTTCGAGTTGGATCAGCAGGTGGGCGAATTATTCAAGGTA------CCCAGTCACTATACACGCGTAATTGTGGACGGCTATACCGATCCTTCCAGCCGAAACCGCTTCTGTCTCGGCCAACTCTCTAATGTGCACAGATCGGAGCAGTCGGAGAAGTCGCGCCTCTACATTGGGAAGGGTGTGGAATTAGACATAGTAGGCGAAGGCGATGTGTGGATTCGCTGTCTCTCCGAGTTCTCCATCTTTGTACAAAGCTACTACCTTGATCGCGAGGCTGGCAGGGCACCGGGCGATGCTGTACACAAAATTTACCCCGGTGCCTACATTAAGGTGTTCGACATTCGTCAGTGCCACGAACAGATGCGTCATCTG---------GCCCACATGACA--------------------------------GTGGACGACTTCCGTCGACTCTGT--AACCTTCGTCTCAGTTTC---------------GTCAAGGGCTGGGGTCCGGACTATCCCCGCCACGACATTAAGGAGACTCCCTGCTGGATTGAAATTCAACTACATAGACCACTGCAACTATTAGACGAGGTT-----------------------TTGCAAGCGATGCCA	TSSDACMNIVHSLMCHRKGGESEE-FSKFAIESLIKKLKDRRDELDALIVAVTSNGATQTSCVTIQRTLDSRMQIAGRKCFPHLIYARLWRWSDAHKTELRHLPFCHFGFDKKLDWVCVNPYHY-----------ERTVSSA---L---DISSLA--LQRPPEYWCNIAYFELDQQVGELFKV--PSHYTRVIVDGYTDPSSRNRFCLGQLSNVHRSEQSEKSRLYIGKGVELDIVGEGDVWIRCLSEFSIFVQSYYLDREAGRAPGDAVHKIYPGAYIKVFDIRQCHEQMRHL---AHMT-----------VDDFRRLC---------------------NLRLSF-----VKGWGPDYPRHDIKETPCWIEIQLHRPLQLLDEV--------LQAMP

	*Trichuris muris**	TMUE_s0006012700|mothers_against_decapentaplegic_4	ACCTCATCGGACGCCTGTGCGAGCATTGTTCATAGCTTGATGTGCCACCGACAAGGCGGCGAC---GAACAG--- TTCAGCCGGCGAGCCATTGAAAGCTTGATTAAAAAGTTGAAGGACAAGCGAGAGGAATTGGATGCTCTCATTC AAGCCATAACGACCAGCGGATCGCATCCGACAAAGTGCGTCACAATTCAGCGCACACTGGACGGCCGTCTTC AGGTTGCCGGTCGTAAGGGTTTTCCTCACGTTGTCTATGCCAGAATATGGCGCTGGCCGGATTTGCATAAAAA CGAACTGAAAAGTTCCAAGTATTGCCAGTATGCGTTTGATCTTAAAGTGGACCTCGTTTGCGTCAATCCTTATC ATTAC---------------------------------GAGCGTGTTGTCTCTCCTGGA--------- ATAAGCAACCTCGATTTCTCCGCCCTCCGC------ CTTCAGCCTCTGCCCGACTTTTGGTCCTCCATTGCCTACTACGAATTGGATACGCAAGTAGGCGAGACGTTCA AAACG------ CCGTCCAGTCATCCGTCGGTCACCGTTGATGGTTACGTGGACCCATCTGGGGTCAGCAGATTTTGCCTTGGC GCTTTGTCCAATGTTCACCGAACGGAAGTGAGCGAGAAAGCTAGGATACACATTGGTCGAGGCGTTCGGCTG GACTTAAAAGGAGAAGGCGACGTTTGGCTGTGCTGCCTAAGCGACTACAGCGTTTTCGTGCAGAGTTACTATT TGGATCGAGAAGCGGGTCGCGCGCCCGGCGACGCAGTGCACAAGATATACCCGAAGGCTTACATTAAGGTG TTCGACCTTAGGCAATGCCATCGACAGATGCTTCAGCAG---------GCAGCGACAGCGCTGTCGGCTGCC--------- GCCGGAATCGGCGTTGACGATCTGCGAAGATTGTGC-- ATATTGAGAATGTCCTTT--------------- GTGAAAGGTTGGGGACCTGACTACCCTCGACAAAGCATAAAGGAGACGCCATGCTGGGTTGAGGTGCACCTG CACAGGGCGTTGCAATTGCTTGACGAAGTG------------------------CTGCACACAATGCCC	TSSDACASIVHSLMCHRQGGD-EQ- FSRRAIESLIKKLKDKREELDALIQAITTSGS HPTKCVTIQRTLDGRLQVAGRKGFPHVVY ARIWRWPDLHKNELKSSKYCQYAFDLKVD LVCVNPYHY-----------ERVVSPG--- ISNLDFSALR-- LQPLPDFWSSIAYYELDTQVGETFKT-- PSSHPSVTVDGYVDPSGVSRFCLGALSNV HRTEVSEKARIHIGRGVRLDLKGEGDVWL CCLSDYSVFVQSYYLDREAGRAPGDAVHK IYPKAYIKVFDLRQCHRQMLQQ--- AATALSAA---AGIGVDDLRRLC----------------- ----ILRMSF----- VKGWGPDYPRQSIKETPCWVEVHLHRAL QLLDEV--------LHTMP

Pangolin J	*Echinococcus granulosus*	gi|674565049| emb|CDS20600.1| protein pangolin J, partial [Echinococcus granulosus]	ATGGAGAAGACGGCGAATCGGATGTATGGAAGCCTACTGTCACCGGGACTGATCGGTTCCTTTGGAGGCTTTATGCCCGCACCGCCTCCACCACCG---TCGAACATCTACGATCTGACGGCATTTCAACAACAGCAACAGCAGCATCAGCAGCAACGCCTTCTACAACAACAACAACAACAGCAGCAGCAGCAGCACCAGCAGCGACAACAACGACCCTCTTCCAATTCGGCCACCTCCACCTCTTCTACTTCAGGCTCCTCTGCCACGCCTTCCACAACTTCGGCGTCCGGCAGCGGCAGTGGTGCTGGCGTCGGCAGT---------------------GTCGATGCTTGGTCCCCCTCCTCCACCTCTTCCGCCTCCGCTCTTCATTCCGAACTATTTAATGCTGCCACCGGATTCTACGGGATGGCAGCCTATGGTCAGCCGATGTCCGGCCACTTCGGTGTAGCCCCGCCTTCGCCTTACAATGCACCTTCAGCG---------------------------------GGTCGTCACTTCGCAGCTGCAGCGGCAGCAGCAGCAGCTGTGTTGGGCAGCTGCTTCTATCCAACTCGCTCTCCACCTGAA---------GCT------GCAGTCTTCTCACCACCACCACCACCGCCTCCTCCA------GCAGCAGTGACATCGGGGCCCTCTGCTGTCCCCCCGTCCCTTTCCACCCCTGGCACCGGCACTTCCTCCGGTACTACTCTCTCGGCACCCGCTACGCCAGCTGCTCAACATTCCTCTGCATCTTCCGCCTCCGCTAGCGCTGCTGCTGCGGCTGCCGCCACTGCTGCCTCTGCCACCGCTGCTGCC------------GCCGCTGCTGCGTTGCTCTCCAGCCCTCTGGCTGCTGTTGCGGCTGCCTCTTCACCGTCCTTCACCCCC---------------------AACATCTTCAATCACGTTCCGCCACACAACCGACAAGCAGCAGCAGTGCATTTCCTCTCCAGTCTCACCGCAGCTGCGGCAGCGTCGTCAGCAGCCTCCTCCTCC---GCCAGTTCGCCCCTCCCACCCCCATCCGCGCCGGTGGCA---------CCGACAGCTGCGGCCGCCGCCCCGCCGCCACCTCCACCACAGGCCTCGCCCTTCGCC---------GCAGCTTTGCATCACTCCCTCGCCGTCGCCACCAATGGAGCCAGCGAGAAAATGGAC------GCGAGCAGTCCACATGCCACCGCGCTCTCCCTTCTCGCCCTCTCACCTGCCTCGCCTGCTGCAGCTGCTGTTGCTGCTGCC---TATTGTCAATCAAAGACGGGTATGCAGTTCGGGTCCTCGATGCTAGCTGCAGGAGAGGCTCAGGCCGCTGGCTCTCCGCTAACCTACACACAGCTTCAACCAGCTTCACAACAGCATGTCCCACCTAAGGACAGTAGCAATGTGGGTGGAGACTACGGAGCAGACTCAAGAGCTTCGCGATACCTTGTGGATGCTGGA------------TCGTCGAAGAGATGCACCAACAATCGATCCTCCAGTTCACTCAGTGGCGCTGGA---------------------GACAAGCATTCGCCCTCTCGAAAACTCATGCGCATTGGATCGGCCGGCAGTGGGGGTTCGCCACTTTCT------GCAGCAGCCTCATCGCTAGACCAATGCTGTTCACCCGGTTCTCCACACACC---ACGGGTAATGCTAGTTCACCTTCTAAAGCAACGCAATCCACTTCTGCTTCAATCTCATCTCCAGGGCTGACGATTAACGCCGCTACATTAGCGCAGTATCGCCGTGAAATCACTACAGAGAACAGCGGAACAGGGGTGCCCACAAAACGGGTGCACATCAAGAAACCATTGAATGCTTTCATGCTCTTTATGAAGGAGATGAGGCCCAAGATTCAGGAGGAGTGCACATTAAAGGAGTCTGCAGCGATCAATCAGATTCTGGGCAAAAAGTGGCATGAATTGTCAAAGCCGGAGCAATCAAAGTACTACGAGCTGGCGCGAAAGGAGAAGGAAATTCACCGCCAG---------------CTCTTCCCTGGTTGGTCTGCTCGCGACAACTATGCAATTCATTCGAAACGCAAACGCAAGCGAAAACTCGCTGCTGCTGTGGCTGCCGCCTCGGCCATGAACGCG---------------GGTGCAGGTGGAGAGGGTGGTGCGGGC------------------------TGTCGAGACCTATACGATGGAGACGGCTACTTTGGTGGA---GGCAGTGGCGGCAGCGGTGGCTGTGGTGGTGGCTGTGGCGGTAGCACCTCTCTCAACGTCTCCACTTCCGC	MEKTANRMYGSLLSPGLIGSFGGFMPAPPPPP-SNIYDLTAFQQQQQQHQQQRLLQQQQQQQQQQHQQRQQRPSSNSATSTSSTSGSSATPSTTSASGSGSGAGVGS-------VDAWSPSSTSSASALHSELFNAATGFYGMAAYGQPMSGHFGVAPPSPYNAPSA-----------GRHFAAAAAAAAAVLGSCFYPTRSPPE---A--AVFSPPPPPPPPP--AAVTSGPSAVPPSLSTPGTGTSSGTTLSAPATPAAQHSSASSASASAAAAAAAATAASATAAA----AAAALLSSPLAAVAAASSPSFTP-------NIFNHVPPHNRQAAAVHFLSSLTAAAAASSAASSS-ASSPLPPPSAPVA---PTAAAAAPPPPPPQASPFA---AALHHSLAVATNGASEKMD--ASSPHATALSLLALSPASPAAAAVAAA-YCQSKTGMQFGSSMLAAGEAQAAGSPLTYTQLQPASQQHVPPKDSSNVGGDYGADSRASRYLVDAG----SSKRCTNNRSSSSLSGAG-------DKHSPSRKLMRIGSAGSGGSPLS--AAASSLDQCCSPGSPHT-TGNASSPSKATQSTSASISSPGLTINAATLAQYRREITTENSGTGVPTKRVHIKKPLNAFMLFMKEMRPKIQEECTLKESAAINQILGKKWHELSKPEQSKYYELARKEKEIHRQ-----LFPGWSARDNYAIHSKRKRKRKLAAAVAAASAMNA-----GAGGEGGAG--------CRDLYDGDGYFGG-GSGGSGGCGGGCGGSTSLNVSTSAAALAAAAAAAAGVDLGNPKKCRARFGLEQQTRWCKPCRRKKKCVRFLTDAEYDEALKAGKLQSEPSSP----------QQVASKSTATT-ST------VTGRAKDQW-SE--A------HHL--PTPKSPNP-SFLSP--SASSTGGSGFSHTVPAASLTHPPAGS-

Echinococcus multilocularis	gi|674570891|emb|CDS43516.1| protein pangolin J [Echinococcus multilocularis]	ATGGAGAAGACGGCGAATCGGATGTATGGAAGCCTACTGTCACCGGGACTGATCGGTTCCTTTGGAGGCTTTATGCCCGCACCGCCTCCACCACCG---TCGAACATCTACGATCTGACGGCATTTCAACAACAGCAACAGCAGCATCAGCAGCAACGCCTTCTACAACAACAACAACAACAGCAGCAGCAG------------CGACAACAACGACCCTCTTCCAATTCGGCCACCTCCACCTCTTCTACTTCAGGTTCCTCTGCCACGCCTTCCACAACTTCGGCGTCCGGCAGCGGCAGTGGTGCTGGCGTCGGCAGT---------------------GTCGATGCTTGGTCCCCCTCCTCCACCTCTTCTGCCTCCGCCCTTCATTCCGAACTATTTAATGCTGCCACCGGATTCTACGGGATGGCAGCCTATGGTCAGCCGATGTCCGGCCACTTCGGTGTAGCCCCGCCTTCGCCTTACGCCGCACCTTCAGCG---------------------------------GGTCGTCACTTCGCAGCTGCAGCGGCTGCAGCAGCAGCAGTGTTGGGCGGCTGCTTCTATCCAACTCGCTCTCCACCTGAA---------GCT------GCAGTCTTCTCACCA---CCACCACCGCCTCCTCCA------GCAGCAGTGACATCGGGGCCCTCTGCCGTCCCCCCGTCCCTTTCCACCCCTGGCACCGGCACTTCCTCCGGTACTACTCTCTCGGCACCCGCTACGCCAGCTGCGCAACATTCCTCTGCATCTTCTGCCTCCGCTACCGCCGCTGCTGCGGCTGCCGCCACTGCTGCCTCCGCCACCGCCGCTGCC------------GCCGCTGCTGCGTTGCTCTCCAGCCCTCTGGCTGCTGTTGCGGCTGCCTCTTCACCGTCCTTCACCCCC---------------------AACATTTTCAATCACGTTCCGCCACACAACCGACAAGCAGCAGCAGTGCATTTCCTCTCCAGTCTCACCGCAGCTGCGGCAGCGTCGTCAGCAGCCTCCTCCTCC---GCCAGTTCGCCCCTCCCACCTCCATCCGCGCCGGTAGCA---------CCGACAGCTGCGGCTGCCGCTCCGCCGCCACCTCCACCACAGGCCTCGCCCTTCGCC---------GCAGCTTTGCATCACTCCCTCGCCGTCGCCACCAATGGAGCCAGCGAGAAAATGGAC------GCGAGCAGTCCACATGCCACCGCGCTCTCCCTTCTCGCCCTCTCACCTGCCTCGCCTGCTGCAGCTGCTGTTGCTGCTGCC---TATTGTCAATCAAAGACGGGTATGCAGTTTGGCTCCTCGATGCTAGCTGCAGGAGAGGCTCAGGCCGCTGGCTCTCCGCTAACCTACACACAGCTTCAACCAGCCTCACAACAGCATGTCCCGCCTAAGGACAGTAGCAATGTAAGTGGAGACTACGGAGCAGACTCAAGAGCTTCGCGATACCTTGTGGATGCTGGA------------TCGTCAAAGAGATGCACCAACAACCGATCATCCAGTTCACTCAGTGGCGCTGGA---------------------GACAAGCATTCGCCCTCTCGAAAACTCATGCGCATTGGATCGGCCGGCAGTGGGGGTTCGCCACTTTCT------GCAGCAGCCTCGTCGCTAGACCAATGCTGTTCGCCCGGTTCTCCACACACC---ACGGGTAACGCCAGTTCACCTTCTAAAGCAACGCAATCCACTTCTGCTTCAATCTCATCTCCAGGGCTGACGATTAACGCCGCTACATTAGCGCAGTATCGCCGTGAAATCACTACAGAGAACAGCGGAACAGGGGTGCCCACAAAACGGGTGCACATCAAGAAACCATTGAATGCTTTCATGCTCTTTATGAAGGAGATGAGGCCCAAGATTCAGGAGGAGTGCACATTAAAAGAGTCTGCAGCGATCAATCAGATTCTGGGCAAAAAGTGGCATGAATTGTCAAAGCCGGAGCAATCGAAGTACTACGAGCTGGCGCGAAAGGAGAAGGAAATTCACCGCCAG---------------CTCTTCCCTGGTTGGTCTGCTCGCGACAACTATGCAATTCATTCGAAACGCAAACGCAAGCGAAAACTCGCTGCTGCTGTGGCTGCCGCCTCGGCCATGAACGCG--------------GGTGCAGGTGGAGAGGGTGGTGCGAGC-----------------------CGTCGAGACCTATACGATGGAGACGGCTACTTTGGTGGA---GGCAGTGGCGGCAGCAGTGGTTGTGGTGGTGGCTGCGGCGGTAGCACCTCTCTAAACGTCTCCACTTCCGC	MEKTANRMYGSLLSPGLIGSFGGFMPAPPPPP-SNIYDLTAFQQQQQQHQQQRLLQQQQQQQQQ----RQQRPSSNSATSTSSTSGSSATPSTTSASGSGSGAGVGS-------VDAWSPSSTSSASALHSELFNAATGFYGMAAYGQPMSGHFGVAPPSPYAAPSA-----------GRHFAAAAAAAAAVLGGCFYPTRSPPE---A--AVFSP-PPPPPP--AAVTSGPSAVPPSLSTPGTGTSSGTTLSAPATPAAQHSSASSASATAAAAAAATAASATAAA----AAAALLSSPLAAVAAASSPSFTP-------NIFNHVPPHNRQAAAVHFLSSLTAAAAASSAASSS-ASSPLPPPSAPVA---PTAAAAAPPPPPPQASPFA---AALHHSLAVATNGASEKMD--ASSPHATALSLLALSPASPAAAAVAAA-YCQSKTGMQFGSSMLAAGEAQAAGSPLTYTQLQPASQQHVPPKDSSNVSGDYGADSRASRYLVDAG----SSKRCTNNRSSSSLSGAG-------DKHSPSRKLMRIGSAGSGGSPLS--AAASSLDQCCSPGSPHT-TGNASSPSKATQSTSASISSPGLTINAATLAQYRREITTENSGTGVPTKRVHIKKPLNAFMLFMKEMRPKIQEECTLKESAAINQILGKKWHELSKPEQSKYYELARKEKEIHRQ-----LFPGWSARDNYAIHSKRKRKRKLAAAVAAASAMNA-----GAGGEGGAS--------RRDLYDGDGYFGG-GSGGSSGCGGGCGGSTSLNVSTSAAALAAAAAAAAGVDLGNPKKCRARFGLEQQTRWCKPCRRKKKCVRPETDAEYDEALKAGKLQSEPSSP----------QQVASKSTATT-ST------VTGRAKDQW-SE--A------HHL--PTAKSPNP-SFLSP--

Hymenolepis microstoma	gi|674594630|emb|CDS26655.1| protein pangolin J [Hymenolepis microstoma]	ATGGAAAAGACA---AGCCGAATGTATGGTGGCCTTCTTCCTCCAGGAATGATCGGATCTTTTGGCGGATTTATGGGAGCGCCACCTCCTTCACCGGCGTCCAATTTCTACGACTTTTCTGCACTTCAA------------CACCATCAACAACAGCGACTCCTTCATCAGCAGGCT--------------------------------CAGCGACCTTCTTCCAACTCCGCCACTTCAAATTCCTCCACATCAGGGTCCTCTTTGACACCTTCCACTACTTCAGCAGCTGGA---GGTCCTGGAGGTTCACTTGGAAATACTGTCAGTGGAAATCCTGCTCTCGACGCTCTCTCCCCATGTTCAACATCATCTGCTTCTGCTTTGCATTCCGAAATATTCAATGCTGCAAGTGGATTCTATGGAATGGCAGCCTACGGTCAACCGATGTCAAGCCATTTTGGAATGGGACCGACATCTCCGTACACAACTCCCGGAGCA--------------------------------GGTCGTCATTTCGCAGCAGCTGCTGCAGCCGCAGCGGCGGTTCTTGGGAGTTGCTTCTATCCAACACGCTCCCCTCCTGAT---------GCCTCAAACAGCGTTTTTCCATCGGCTTATCCTCCCTCACAGCAACCAAGTGCATCTCAACAGCAGAGCTCCAACGCC------CCTCCACAACCACAAAACAGCGCTACTACTTCCAGCTCGCAGTCT---------CCAGCGCAACATTCCGCCAACTCCTCCTCT--------------------------------------GCTGTAGCTGCGTCAGCGTCGGCTGCTGCTTCTGCTGCTGCCGCCGCTGCTGCATGGCTATCAAGTCCTTTAGCCGCTGTTGCCGCCGCTAATTCACCATCATTCACTCCC---------------------AATATCTTCAATCACGTTCCTCCGCACAATCGGCAGGCAGCAGCTGTACATTTCCTGACAAGCTTGACAGCAGCTGCAGCGGCTTCTTCTGCAGCCTCTTCTTCCTCTGCAAGCTCGCCTCTACAGCCGCCTTCAGCTTCCGTTAGTGGTCCGAATCATACCGGTATCACTGGGGCTCCTCCA---------------GTGTCGCCTTTCGCCAGTGCT---GCTGCCTTCCAACATTCGCTCGCTGTGGCCACAAATGGCGGCGCAGATAAACTGGATATGGGGGCGGGTAGTCCACATGCTGCGGCCCTCTCGCTTTTTGCTCTGTCTCCAGCATCACCTGCTGCCGCCGCAGTTGCCGCA------TACTGTCAATCCAAGACCAGCATGCAGTTTGGATCCTCAATGTTGAACGCAGGAGACTCCCAGAATGACGGATCTCCTTTAACCTACACTCAATTACAGCCAGCTTCTCAGCAACAG---CCACAACAGGACTCG------ATGAGTGGAGATTATGGAGCAGATCCAAGATCCTCGCGGTATCATGCAGATTCCAAT------------TCGTCAAAGAGGTGTGCAAACAATCGG---TCAAGTTCATTAAGTGGTGCAGGAGGA---------GGAGGGGGAGATAAAAATTCTCCTTCGAGAAAACTCATGCGAATCGGATCAGTAGGTAGTGGTGGGTCTCCGTTATCCTCAGCAGCAGCGTCATCATCCTTGGACCAATGTTGCTCCCCGGGATCTCCTCAAACCCTAACTGGTGCTGGCGGATCACCGTCTAAGACAGGAACATCGACATCAGCATCAGCATCATCTCCAGGTCTGACAATAAATGCGGCAACTTTGGCGCAATATCGTCGGGAAATTACAACTGAGAATAGTGGAACTGGTGTGCCCACGAAGCGCGTGCATATAAAGAAGCCGCTGAATGCGTTTATGCTATTCATGAAAGAAATGCGACCCAAAATTCAGGAGGAATGCACACTGAAAGAGTCTGCAGCAATCAACCAAATTCTTGGCAAAAAGTGGCATGAGCTGTCTAAACCAGAACAGTCAAAATACTACGAGTTAGCGAGGAAAGAGAAGGAAATTCATCGACAG---------------CTGTTCCCCAGCTGGTCTGCTCGGGATAACTATGCAATACATTCGAAGCGTAAACGCAAACGAAAATTAGCCGCAGCAGTGGCCGCAGCCTCAGCAATGAATTCA---------------GGC---GGAGGGGATGGTGGTTCAGGT-----------------------CGTCGTGACCTTTTCGACGGTGACGGATATTTTGCAGGT---GGTGGAAATGGGAGCAATGTGAGTGGAGGG------------------TCCTTGAATTCATCGAATTCGGCTGCTGCTTTGGCGGCAGCAGCCGCAGCT---------GGAGTTGATCTTGGAAACCCCAAAAAATGTCGCGCCCGGTTCGGTCTTGAACAGCAAACCCGGTGGTGTAAG	MEKT-SRMYGGLLPPGMIGSFGGFMGAPPPSPASNFYDFSALQ----HHQQQRLLHQQA-----------QRPSSNSATSNSSTSGSSLTPSTTSAAG-GPGGSLGNTVSGNPALDALSPCSTSSASALHSEIFNAASGFYGMAAYGQPMSSHFGMGPTSPYTTPGA-----------GRHFAAAAAAAAAVLGSCFYPTRSPPD---ASNSVFPSAYPPSQQPSASQQQSSNA--PPQPQNSATTSSSQS---PAQHSANSSS-------------AVAASASAAASAAAAAAAWLSSPLAAVAAANSPSFTP-------NIFNHVPPHNRQAAAVHFLTSLTAAAAASSAASSSSASSPLQPPSASVSGPNHTGITGAPP-----VSPFASA-AAFQHSLAVATNGGADKLDMGAGSPHAAALSLFALSPASPAAAAVAA--YCQSKTSMQFGSSMLNAGDSQNDGSPLTYTQLQPASQQQ-PQQDS--MSGDYGADPRSSRYHADSN----SSKRCANNR-SSSLSGAGG---GGGDKNSPSRKLMRIGSVGSGGSPLSSAAASSSLDQCCSPGSPQTLTGAGGSPSKTGTSTSASASSPGLTINAATLAQYRREITTENSGTGVPTKRVHIKKPLNAFMLFMKEMRPKIQEECTLKESAAINQILGKKWHELSKPEQSKYYELARKEKEIHRQ-----LFPSWSARDNYAIHSKRKRKRKLAAAVAAASAMNS-----G-GGDGGSG--------RRDLFDGDGYFAG-GGNGSNVSGG------SLNSSNSAAALAAAAAA---GVDLGNPKKCRARFGLEQQTRWCKPCRRKKKCVRFLTDAEYDEALKAGRLQSEPSSP----------QQQQQQSQQSQQNT------NQSGTKDQWISD--G------NQLNKSNPKSPNSGAFLSPAPSSSSTG--------------FVPQPS-GDFVGTTNPL---

*Mesocestoides corti**	MCOS_0000647601-mRNA-1	--ATGGTTGACGCCGCAGGC---	----------------MVDAAG----------------------
		CAGCGACAGCGTCACGGGAGTGTGCGCATG--	QRQRHGSVRM--------------------------------------
		--	----------------------------------SAYAFFSPSF---
		TCCGCTTATGCCTTCTTTTCTCCCTCCTTC---------	AGFYGISAYGQSMPSHFGVSSTSPYSSPA
		GCTGGTTTCTACGGGATTTCCGCCTACGGCCAGTCAATGCCGAGTCATTTTGGCGTTTCTTCAACGTCGCCGT	AGAVGGPASSVAGRHFAAAAAAAAAMLG
		ACTCATCGCCGGCAGCGGGTGCTGTGGGGGGTCCAGCATCGTCTGTTGCGGGACGCCATTTCGCGGCTGCG	SCFYPTRSPPETLGAPGSMLSSPPPPPPH
		GCGGCCGCGGCGGCGGCGATGCTAGGGAGTTGCTTTTACCCCACGCGCTCCCCGCCTGAGACTCTTGGAGC	PAPPQQ-------------SASVSSPST---
		GCCCGGGTCGATGTTATCGTCGCCGCCGCCGCCTCCTCCCCATCCAGCGCCGCCTCAGCAG---------------------	PASAASSSSS-----------------
		------------------TCCGCGTCCGTGTCGTCGCCCTCCACG---------CCGGCCTCTGCCGCCTCCTCCTCCTCC-----	SSTSQQHSAATAAAWLSSPLAAVAAATSP
		--	SFTP-------
		TCCTCCACCAGCCAGCAACATTCCGCTGCGACCGCGGCAGCCTGGCTGTCCAGTCCTCTGGCCGCTGTGGC	NLFNHVPPHNRQAAAVHFLSSLTAAAAAS
		AGCTGCCACCTCGCCATCCTTCACACCA---------------------	SAASSSSASSPLPPPPPPQA---PTASPSA--
		AATCTCTTTAACCACGTCCCGCCGCACAATCGACAGGCAGCAGCGGTGCATTTCCTCTCAAGCCTGACTGCG	-----VSPFV--GGAFHHSLSVATNGSGEKMD-
		GCAGCGGCAGCTTCGTCCGCGGCCTCGTCATCTTCCGCCAGTTCGCCCCTTCCACCGCCACCACCACCACA	-
		GGCG---------CCGACGGCTTCGCCCTCGGCC---------------------GTCTCGCCGTTCGTG------	AGSPHATALSLLALSPASPAAAAVAAAAYC
		GGAGGCGCGTTCCACCACTCCCTCTCCGTGGCAACCAATGGAAGTGGCGAGAAGATGGAC------	QSKSGLQFGSSMLTA-
		GCTGGCAGTCCCCATGCCACTGCGCTTTCGCTTCTCGCGTTGTCCCCCGCGTCTCCTGCCGCCGCTGCTGTC	DGQAANSPLTYTQLQPASQT---
		GCCGCTGCCGCTTACTGCCAATCAAAATCCGGGCTACAGTTTGGGTCGTCGATGCTCACGGCG---	SKDSDRTLPGFAADGRHQRYLSESGRSC
		GACGGCCAGGCCGCCAATTCACCCCTCACCTACACCCAACTTCAGCCAGCCTCACAAACC---------	DSTKRC-
		TCGAAGGACAGTGACAGGACCCTTCCAGGTTTCGCAGCAGACGGACGTCATCAGCGATATCTCAGTGAGTCT	NNRSASSTGGGGGGSGGGGGKHSPSWK
		GGGAGATCCTGCGACTCCACGAAAAGATGC---	IMRINSGGSGGSPLS-AAASSSLDQC----
		AACAATCGATCTGCGAGTTCAACGGGTGGTGGTGGCGGCGGCTCCGGAGGTGGTGGAGGAAAGCATTCTCC	SPGC-
		GTCGTGGAAGATCATGCGAATAAACTCGGGCGGAAGTGGGGGCTCTCCGCTGTCA---	TGSPRTGSPRTASTAASSSSPGLTINAATL
		GCAGCAGCATCGTCCTCCCTTGATCAGTGT------------TCGCCAGGCTGC---	AQYRREITTENSGTGVPTKRVHIKKPLNAF
		ACGGGATCTCCACGCACTGGATCACCCCGCACAGCTAGCACTGCGGCCTCTTCGTCGTCGCCTGGCTTGACA	MLFMKEMRPKIQEECTLKESAAINQILGKK
		ATTAACGCCGCCACACTTGCCCAGTACCGGCGGGAGATCACCACGGAGAACAGTGGAACTGGTGTGCCCAC	WHELSKPEQSKYYELARKEKEIHRQASLG
		AAAACGGGTGCATATTAAGAAACCCCTCAACGCTTTCATGCTCTTCATGAAGGAGATGCGACCGAAGATTCAG	KLFPGWSARDNYAIHSKRKRKRKLAAAVA
		GAGGAGTGCACTCTGAAGGAGTCGGCTGCGATTAATCAGATTTTGGGCAAGAAGTGGCACGAGCTTTCCAAG	AASVINAAAAAGG-
		CCCGAACAATCCAAGTACTATGAACTGGCGCGGAAGGAGAAGGAAATCCATCGTCAGGCAAGTCTTGGGAAG	GGGGGAGEGSCGGGNRRDLFDADGYFA
		CTGTTCCCTGGCTGGTCGGCTCGCGACAACTATGCAATCCACTCGAAGCGGAAACGCAAACGCAAACTCGCC	GAGGNGM----G------SLNTSS---ALAAA-------
		GCTGCGGTAGCTGCTGCCTCTGTGATAAACGCCGCCGCCGCCGCTGGCGGA---	VDLGNPKKCRARFGLEQQTRWCKPCRRK
		GGGGGTGGCGGTGGGGCCGGAGAAGGTTCCTGTGGCGGTGGCAACCGACGCGACCTCTTCGACGCTGACG	KKCVRFLTDAEYDEALQSGKLQSEPSSPS
		GCTACTTCGCTGGTGCCGGAGGCAATGGCATG------------GGC--------------------TCCCTCAACACCTCCTCC---------	AAAAGGGNPQAPSSHPQAST-
		GCGTTGGCAGCCGCT----------------------	PAPSGLCGAAGRVKDQW-SDQLP------
		GTCGACCTTGGCAACCCCAAGAAGTGTAGAGCGCGTTTCGGACTGGAGCAACAGACGCGTTGGTGCAAACCA	KSL--SNPKTPNS-TFLSPNPTGAYSH---------
		TGCAGACGAAAGAAGAAGTGCGTCCGCTTCCTCACGGACGCCGAGTACGACGAGGCTTTGCAATCGGGCAAA	-----PPPAPSGSDYAASTS--
		CTGCAGTCTGAGCCCTCTTCTCCGTCGGCGGCGGCGGCAGGCGGTGGAAATCCACAGGCGCCGTCTTCGCA	GSNEDYSAAFSAPS----
		CCCACAAGCTTCCACG---CCAGCACCCAGTGGCTTGTGCGGTGCAGCAGGAAGGGTGAAAGATCAGTGG---	YNQFSPYGDFQRPTLAHSTLSFLEHH-

*Taenia solium**	TsM_001136700	-- -- -- ATGGCAGCCTATGGTCAACCAATGTCCAGCCACTTTAGTGTAGCCCCGCCCTCGCCTTACGCCGCACCATCT GCG-------------------------------- GGTCGTCATTTCGCAGCTGCTGCGGCCGCAGCAGCAGCCGTTTTGGGCAGCTGCCTCTACCCCACTCGCTCT CCACCCGAA---------GCT------ GCAGTCTTTTCACCACCACCACCACCACCGCCTCCACCAGCAGCAGCTGTGCCGCCGGGACCCTCCGCCGT ACCTCCGTCCCTTCCTACCGCTGGTGCAGTCAGCTCCTCAAGTGCTACTCTCTCAGCACCCGCCACGCCGGC CGCGCAACATTCCTCTGCATCCTCCGCCTCCGCTAGCGCCGCTGCTGCGGTCGCTGCCACTGCCGCCTCCG CCACCGCCGCTGCCTCCGCTGCTGCCGCAGCTGCTGCGTTGCTCTCTAGCCCTCTGGCTGCTGTTGCAGCTG CCTCCTCGCCGTCCTTCACTCCCAAGATTTGTATTAACCCCTCCAGCATCTTCAACCACGTTCCGCCACACAA CCGACAAGCTGCAGCAGTGCATTTCCTCTCCAGCCTCACCGCAGCTGCGGCGGCGTCGTCAGCAGCCTCTTC TTCCTCCGCCAGCTCACCCCTGCCACCCCCATCCGCGTCGGTGCCA--------- CCGACGGCTGCGGCCGCCGCCCCGCCACCACCCACACCACAGGCGTCGCCCTTCGCC--------- GCAGCATTCCATCACTCTCTCGCCGTTGCTACTAATGGAGCCAGTGAGAAAATGGAC------ GCAGGCAGTCCGCATGCCACTGCACTCTCTCTTCTTGCTCTGTCACCTGCCTCACCTGCTGCAGCCGCTGTT GCTGCTGCTGCCTATTGCCAATCAAAGACGGGCATGCAGTTTGGCTCCTCGATGCTGGCCGCAGGAGAGGCT CAAGCCGCTGGCTCTCCGCTGACCTACACACAACTGCAACCCGCCTCACAGCAGCACGTTCCCTCCAGGGAG GGTGGCAATGTGGGGGGAGACTATGGAGCTGATGCCAGAGCCTCGCGATACCTTGCGGATTCTGGA------------ TCTTCAAAGAGATGCACCAACAATCGGTCCTCCAGTTCACTCAGTGGCACCGGA-------------------- GACAAGCATTCACCGTCTCGAAAACTCATGCGTATTGGATCGGCCGGCAGTGGAGGTTCACCGCTTTCT------ GCAGCAACTTCATCTATCGACCAGTGCTGTTCACCGGCCTCGCCACACACT--- ACCGGTGCCGCCAGTTCGCCCTCCAAAACAACTCAATCCACATCCGCTTCTGTCTCATCACCAGGGTTAACGA TTAATGCCGCGACACTGGCACAGTATCGTCGTGAAATCACCACGGAGAACAGCGGAACTGGGGTGCCTACAA AGCGAGTACACATTAAGAAACCGTTGAATGCTTTCATGCTCTTTATGAAGGAGATGAGGCCCAAGATTCAAGA AGAGTGCACCCTAAAAGAGTCCGCAGCCATTAACCAGATTCTGGGTAAAAAGTGGCATGAATTGTCAAAACCG GAGCAATCAAAGTACTATGAGCTGGCGCGAAAGGAGAAGGAGATTCACCGCCAG--------------- CTATTTCCTGGTTGGTCTGCCCGTGACAACTATGCAATTCACTCGAAACGCAAACGCAAACGAAAACTCGCCG CCGCTGTGGCCGCCGCCTCAGCCATGAACGCT---------------GGCGCGGGTGGGGAGGGTGGTGCAGGT--------- ---------------CGTCGAGACCTCTACGACGGTGATGGCTACTTTGGTGGT--- GGCAGTGGCGGTGGCAGTGGTTGCGGTGGTGGTTGTGGTGGTAGCACCTCCCTCAACGCCTCCACTTCTGC GGCTGCTCTTGCCGCTGCTGCAGCTGCAGCAGCGGCTGGTGTAGACCTTGGGAATCCCAAAAAGTGCCGAG CTCGTTTTGGTCTCGAGCAACAGACCCGGTGGTGCAAGCCCTGTCGACGAAAGAAGAAGTGCGTCCGATTTC TCACCGACGCTGAGTACGAAGAAGCTCTGAAGGCTGGTAAATTGCAGTCGGAGCCGTCGTCACCG--------------- ---------------CAACAGGTTGCTGGTAAATCCACTGCTACAACC---AGCACA------------------ GTAACGGGAGGGGCGAAAGACCAGTGG---TCGGAG------GCAACTCTAACCAAGTCCACACCCCATCTG------ CCTACCCCAAAGAGTCCAAATTCG---ACCTTCCTCTCTCCT---	-- -- ------------------------ MAAYGQPMSSHFSVAPPSPYAAPSA-------- ---GRHFAAAAAAAAAVLGSCLYPTRSPPE-- -A-- AVFSPPPPPPPPPAAAVPPGPSAVPPSLPT AGAVSSSSATLSAPATPAAQHSSASSASA SAAAAVAATAASATAAASAAAAAAALLSSP LAAVAAASSPSFTPKICINPSSIFNHVPPHN RQAAAVHFLSSLTAAAAASSAASSSSASSP LPPPSASVP---PTAAAAAPPPPTPQASPFA- --AAFHHSLAVATNGASEKMD-- AGSPHATALSLLALSPASPAAAAVAAAAYC QSKTGMQFGSSMLAAGEAQAAGSPLTYT QLQPASQQHVPSREGGNVGGDYGADARA SRYLADSG----SSKRCTNNRSSSSLSGTG-- -----DKHSPSRKLMRIGSAGSGGSPLS-- AATSSIDQCCSPASPHT- TGAASSPSKTTQSTSASVSSPGLTINAATL AQYRREITTENSGTGVPTKRVHIKKPLNAF MLFMKEMRPKIQEECTLKESAAINQILGKK WHELSKPEQSKYYELARKEKEIHRQ----- LFPGWSARDNYAIHSKRKRKRKLAAAVAA ASAMNA-----GAGGEGGAG-------- RRDLYDGDGYFGG- GSGGGSGCGGGCGGSTSLNASTSAAALA AAAAAAAAGVDLGNPKKCRARFGLEQQTR WCKPCRRKKKCVRFLTDAEYEEALKAGKL QSEPSSP----------QQVAGKSTATT-ST------ VTGGAKDQW-SE--ATLTKSTPHL-- PTPKSPNS-TFLSP- PSASSTGGSGFSHTLPAASLTHPPAGS- SDFVSSNS-----EEYSATASTSSG- FGYGHFSPYADFQRPALAHSTLSFLEHH- RPSPSFGS-VQLSQTTTA------GNN-T- EVEE-- EDDMKNETLCVFATEASPAGSCHHDN---

---

* Identification from reference genome

APÊNDICE 21: SUPPLEMENTARY FILE 18

Supplementary File 18. The putative proglottisation-related proteins alignment features and parameters for the evolutionary analysis.

Protein name	Align software[1]	Alignment cover[2]	Align length[3]	NSeqs [4]	NDom [5]	Best NT model[6]	Best Prot model[7]
Bone morphogenetic protein 2	Prank (translated codon)	Partial	165	9	1	K2 + G	JTT+G
Cyclin-g-associated kinase	Prank (translated codon)	Partial	827	14	4	T92+G	LG+G+I
Groucho protein	Prank (translated codon)	Partial	977	13	2	K2+G	JTT+G
Homeobox protein Hox B4a	Prank (translated codon)	Partial	744	8	1	HKY+I	LG+I
Lim homeobox protein lhx1	Clustal Omega	Partial	202	15	2	K2+G	Dayhoff+G
Membrane-associated guanylate kinase protein 2	Prank (translated codon)	Total	570	5	1	K2+I	JTT+G
Serine:threonine protein kinase mark2	Prank (translated codon)	Total	1638	5	3	K2+G	JTT+G+F
Atrial natriuretic peptide receptor 1	Prank (translated codon)	Partial	329	15	3	GTR+G+I	LG+G
RNA binding motif single stranded interacting	Prank (translated codon)	Partial	723	5	1	HKY+I	JTT+I
Serine:threonine protein kinase	Prank (translated codon)	Partial	1187	5	2	HKY+G	JTT+G
Mothers against decapentaplegic homolog 4-like	Prank (translated codon)	Partial	329	15	2	T92+G	JTT+G
Pangolin j	Prank (translated codon)	Partial	1192	5	1	HKY+I	JTT+G+F

[1] Software that generated the best proteins/nucleotides alignment

[2] Partial: Alignments with remotion of low aligned regions; Total: Alignments without remotion of low aligned regions

[3] Final protein alignment size

[4] Number or species/orthologous analysed

[5] Number of different domain types

[6] The Best-fit model of codon evolution

[7] Best-fit model of protein evolution

APÊNDICE 22: PARÂMETROS PAML

Parâmetros aplicados para o cálculo dos comprimentos de ramos pelo modelo M0.

```
      seqfile = <Name>.phy          * sequence data filename
      treefile = <Name>_tree        * tree structure file name
      outfile = <Name>_M0           * main result file name

        noisy = 9   * 0,1,2,3,9: how much rubbish on the screen
      verbose = 1   * 0: concise; 1: detailed, 2: too much
      runmode = 0   * 0: user tree;  1: semi-automatic;  2: automatic
                    * 3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise

      seqtype = 1 * 1:codons; 2:AAs; 3:codons-->AAs
    CodonFreq = 2   * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table

*        ndata = 10
        clock = 0   * 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis
       aaDist = 0   * 0:equal, +:geometric; -:linear, 1-6:G1974,Miyata,c,p,v,a
   aaRatefile = dat/jones.dat  * only used for aa seqs with model=empirical(_F)
                    * dayhoff.dat, jones.dat, wag.dat, mtmam.dat, or your own

        model = 0
                    * models for codons:
                        * 0:one, 1:b, 2:2 or more dN/dS ratios for branches
                    * models for AAs or codon-translated AAs:
                        * 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
                        * 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)

      NSsites = 0   * 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
                    * 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;
                    * 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
                    * 13:3normal>0

        icode = 0   * 0:universal code; 1:mammalian mt; 2-10:see below
        Mgene = 0
                    * codon: 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff
                    * AA: 0:rates, 1:separate

    fix_kappa = 0   * 1: kappa fixed, 0: kappa to be estimated
        kappa = 2   * initial or fixed kappa
    fix_omega = 0   * 1: omega or omega_1 fixed, 0: estimate
        omega = .4 * initial or fixed omega, for codons or codon-based AAs

    fix_alpha = 1   * 0: estimate gamma shape parameter; 1: fix it at alpha
        alpha = 0.  * initial or fixed alpha, 0:infinity (constant rate)
       Malpha = 0   * different alphas for genes
*        ncatG = 8   * # of categories in dG of NSsites models

        getSE = 0   * 0: don't want them, 1: want S.E.s of estimates
 RateAncestor = 1   * (0,1,2): rates (alpha>0) or ancestral states (1 or 2)

   Small_Diff = .5e-6
    cleandata = 0   * remove sites with ambiguity data (1:yes, 0:no)?
  fix_blength = -1   * 0: ignore, -1: random, 1: initial, 2: fixed
       method = 1   * Optimization method 0: simultaneous; 1: one branch a time

* Genetic codes: 0:universal, 1:mammalian mt., 2:yeast mt., 3:mold mt.,
* 4: invertebrate mt., 5: ciliate nuclear, 6: echinoderm mt.,
* 7: euplotid mt., 8: alternative yeast nu. 9: ascidian mt.,
* 10: blepharisma nu.
* These codes correspond to transl_table 1 to 11 of GENEBANK.
```

Parâmetros aplicados para Pr1A para análise pelos modelos M1a, M2a, M3, M7 e M8.

```
      seqfile = <Name>.phy                 * sequence data filename
      treefile = tree_M0                   * tree structure file from M0 analysis
      outfile = <Name>_allmodels           * main result file name

        noisy = 9  * 0,1,2,3,9: how much rubbish on the screen
      verbose = 1  * 0: concise; 1: detailed, 2: too much
      runmode = 0  * 0: user tree;  1: semi-automatic;  2: automatic
                   * 3: StepwiseAddition; (4,5):PerturbationNNI; -2: pairwise

      seqtype = 1  * 1:codons; 2:AAs; 3:codons-->AAs
    CodonFreq = 2  * 0:1/61 each, 1:F1X4, 2:F3X4, 3:codon table

*        ndata = 10
        clock = 0  * 0:no clock, 1:clock; 2:local clock; 3:CombinedAnalysis
       aaDist = 0  * 0:equal, +:geometric; -:linear, 1-6:G1974,Miyata,c,p,v,a
   aaRatefile = dat/jones.dat  * only used for aa seqs with model=empirical(_F)
                   * dayhoff.dat, jones.dat, wag.dat, mtmam.dat, or your own

        model = 0
                   * models for codons:
                       * 0:one, 1:b, 2:2 or more dN/dS ratios for branches
                   * models for AAs or codon-translated AAs:
                       * 0:poisson, 1:proportional, 2:Empirical, 3:Empirical+F
                       * 6:FromCodon, 7:AAClasses, 8:REVaa_0, 9:REVaa(nr=189)

      NSsites = 1 2 3 7 8  * 0:one w;1:neutral;2:selection; 3:discrete;4:freqs;
                   * 5:gamma;6:2gamma;7:beta;8:beta&w;9:beta&gamma;
                   * 10:beta&gamma+1; 11:beta&normal>1; 12:0&2normal>1;
                   * 13:3normal>0

        icode = 0  * 0:universal code; 1:mammalian mt; 2-10:see below
        Mgene = 0
                   * codon: 0:rates, 1:separate; 2:diff pi, 3:diff kapa, 4:all diff
                   * AA: 0:rates, 1:separate

    fix_kappa = 0  * 1: kappa fixed, 0: kappa to be estimated
        kappa = 2  * initial or fixed kappa
    fix_omega = 0  * 1: omega or omega_1 fixed, 0: estimate
        omega = .4 * initial or fixed omega, for codons or codon-based AAs

    fix_alpha = 1  * 0: estimate gamma shape parameter; 1: fix it at alpha
        alpha = 0. * initial or fixed alpha, 0:infinity (constant rate)
       Malpha = 0  * different alphas for genes
        ncatG = 8  * # of categories in dG of NSsites models

        getSE = 0  * 0: don't want them, 1: want S.E.s of estimates
 RateAncestor = 1  * (0,1,2): rates (alpha>0) or ancestral states (1 or 2)

   Small_Diff = .5e-8
    cleandata = 0  * remove sites with ambiguity data (1:yes, 0:no)?
  fix_blength = 1  * 0: ignore, -1: random, 1: initial, 2: fixed
       method = 0  * Optimization method 0: simultaneous; 1: one branch a time

* Genetic codes: 0:universal, 1:mammalian mt., 2:yeast mt., 3:mold mt.,
* 4: invertebrate mt., 5: ciliate nuclear, 6: echinoderm mt.,
* 7: euplotid mt., 8: alternative yeast nu. 9: ascidian mt.,
* 10: blepharisma nu.
* These codes correspond to transl_table 1 to 11 of GENEBANK.
```