

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
FACULDADE DE MEDICINA VETERINÁRIA
CAMPUS DE ARAÇATUBA

TRANSCRIPTOMA DE METACESTÓIDE DE
Taenia saginata

Silvana de Cássia Paulan
Bióloga

ARAÇATUBA – SP
2015

UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
FACULDADE DE MEDICINA VETERINÁRIA
CAMPUS DE ARAÇATUBA

TRANSCRIPTOMA DE METACESTÓIDE DE
Taenia saginata

Silvana de Cássia Paulan

Orientadora: Profa. Adjunto Cárís Maroni Nunes

Co-orientador: Prof. Adjunto José Fernando Garcia

Tese apresentada a Faculdade de Medicina Veterinária – Unesp, Campus Araçatuba, como parte das exigências para obtenção do título de Doutor em Ciência Animal (Medicina Veterinária Preventiva e Produção Animal).

ARAÇATUBA – SP

2015

Dados Internacionais de Catalogação na Publicação (CIP)
Ficha catalográfica elaborada pela Biblioteca da FMVA / UNESP

P324t Paulan, Silvana de Cássia
Transcriptoma de metacestóide de *Taenia saginata* / Silvana de Cássia Paulan. – Araçatuba, 2015.
63 f. : il. + 1 CD-ROM

Tese (Doutorado) – Universidade Estadual Paulista,
Faculdade de Medicina Veterinária de Araçatuba.
Orientadora: Profa. Cárís Maroni Nunes
Coorientador: Prof. José Fernando Garcia

1. Cisticercose. 2. Sequenciamento de nucleotídeos em larga escala. 3. RNA mensageiro. 4. RNAseq. I. T.

CDD 636.20896



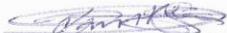
CERTIFICADO DE APROVAÇÃO

TÍTULO: Transcriptoma de metacestóide de Taenia saginata.

AUTORA: SILVANA DE CÁSSIA PAULAN

ORIENTADORA: Dra. CÁRIS MARONI NUNES

Aprovada como parte das exigências para obtenção do Título de DOUTORA em CIÊNCIA ANIMAL (MEDICINA VETERINÁRIA PREVENTIVA E PRODUÇÃO ANIMAL) pela Comissão Examinadora.


Dr. ADAM TAITI HARTH UTSUNOMIYA


Dra. FLÁVIA LOMBARDI LOPES


Dr. DANIEL GUARIZ PINHEIRO


Dr. MARCELO VASCONCELOS MEIRELES


Dra. CÁRIS MARONI NUNES

DATA DA REALIZAÇÃO: 4 de novembro de 2015.


Presidente da Comissão Examinadora
Dra. CÁRIS MARONI NUNES
- Orientadora -

DADOS CURRICULARES DO AUTOR

SILVANA DE CÁSSIA PAULAN – nascida na cidade de São Bernardo do Campo – SP em 23 de fevereiro de 1979. Possui graduação em Ciências Biológicas pela Faculdade de Engenharia de Ilha Solteira (UNESP), campus de Ilha Solteira, em janeiro de 2009, onde iniciou suas atividades na área da ciência em Imunoparasitologia. Em março deste mesmo ano iniciou projeto sobre diagnóstico sorológico para a cisticercose bovina, pelo qual obteve o título de Mestre em Ciência Animal, pela Faculdade de Medicina Veterinária de Araçatuba (UNESP), campus de Araçatuba. Possui 07 artigos publicados na área de imunoparasitologia e 27 resumos apresentados em congressos nacionais e internacionais.

EPIGRAFE

“Amanhã
Eu tenho tanto
Que trabalhar
Eu tenho muito
O que conquistar
Agora eu penso
Amanhã (...)
Eu tenho ideias
Eu tenho metas
Amanhã (...)
Quero amigos
Para encontrar
Amanhã
Quero estranhos
Para descobrir
Amanhã quero bússolas
Para me perder
E quero um norte
Para ser feliz
Amanhã (...)
Eu tenho tanto
Amanhã
Amanhã
Amanhã
acaba de chegar”

Theidy Corrêa

DEDICATÓRIA

Aos meus pais e à minha irmã.

Às amigas Anna C. Fioravante e Irani Ferreira.

AGRADECIMENTOS

À Universidade Estadual Paulista “Júlio de Mesquita Filho”, em especial ao Programa de Pós-Graduação em Ciência Animal.

À CAPES e CNPq, pelas bolsas de estudo durante o Doutorado.

À Dra. Cárís Maroni Nunes, pela oportunidade de desenvolver este projeto de Doutorado, sendo tão competente orientadora e amiga.

Ao Dr. José Fernando Garcia, pelo suporte e orientação.

Ao doutorando Yuri Tani Utsunomiya e ao pós-doutorando Marco Milanesi por toda dedicação, troca de conhecimentos e amizade.

À doutoranda Rafaela Beatriz Pintor Torrecilha, pela inapreciável amizade e apoio.

À equipe do laboratório de Bioquímica e Biologia Molecular, ao graduando Pier Kenji R. K. Ito, às mestrandas Flávia A. Florencio, Tamiris Sayuri Aguiar e Kennya Rozy, à Mestre Dandara Camila Morais Pereira, aos doutorandos Fernanda M. de Oliveira, Anirene G. T. Pereira, Ludmilla Zavarez e aos pós-doutorandos Adam T. H. Utsunomiya e Sarita G. Ferrari, por auxiliarem a execução deste trabalho, por compartilhar pensamentos e por todo apoio.

Ao Dr. Najib M. El-Sayed, da University of Maryland, pelo auxílio para o desenvolvimento deste projeto, disponibilizando o laboratório e acesso ao servidor da Universidade. À equipe do Department of Cell Biology and Molecular Genetics, Ashton Trey Belew, Maria Cecilia F. Dupecher, Keith Hughitt, Yuan Lee, Laura Dillon, Rachel Evans e Ginger Houston-Ludlam, por todo apoio e suporte.

Ao Dr. Rondon P. M. Neto, pelo suporte, incentivo e amizade.

A todos que direta ou indiretamente participaram deste projeto.

SUMÁRIO

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS.....	12
<i>Taenia saginata</i>	13
Ciclo biológico.....	13
Epidemiologia do complexo Teníase-Cisticercose por <i>Taenia saginata</i>	15
Diagnóstico e controle.....	17
Sequenciamento de RNA (RNAseq).....	20
Montagem do transcriptoma.....	21
Objetivo.....	25
Referências.....	25
CAPÍTULO 2 – Montagem do transcriptoma de metacestóide de <i>Taenia saginata</i> *	34
<i>Taenia saginata</i> transcriptome assembly.....	34
Abstract.....	34
Introduction.....	35
Material and methods.....	36
Total RNA extraction and cDNA library.....	36
Sequencing and data processing.....	36
Transcriptome assembly.....	37
Functional annotation.....	37
Results.....	38
Discussion.....	49
Conclusions.....	51
References.....	52
Supporting Information.....	57

ABREVIATURAS E TERMOS

EST – *Expressed sequence tag*. Fragmento de DNA, referente à uma região gênica, utilizada para a identificação de genes completos, bem como referência para mapeamento.

cDNA - DNA complementar. É o DNA sintetizado a partir de uma molécula de RNA mensageiro, cujos íntrons já foram removidos.

mRNA – Molécula de ácido ribonucleico transcrito a partir de um gene, e do qual uma proteína pode ser traduzida pela ação de ribossomos.

k-mer – refere-se à motivo de tamanho k , observado mais de uma vez numa sequência de genômica.

KEGG - *Kyoto Encyclopedia of Genes and Genomes*. É uma coleção de bancos de dados on-line que lidam com genoma, vias metabólicas e substâncias químicas biológicas

COG – *Clusters of Orthologous Groups*. Banco de dados criado para comparar proteínas preditas e conhecidas.

FPKM – Unidade de expressão *fragments per kilobase of exon per million fragments mapped*.

BLAST - *Basic local alignment search tool* é um algoritmo para comparação de informação entre sequências biológicas.

RNAseq – Sequenciamento de RNA.

Contig - Sequência contígua de DNA, criado por sobreposição de montagem de fragmentos de um cromossomo sequenciados.

Grafo – Conjunto de elementos unidos por arcos.

Splicing - Processo que remove os íntrons e une os éxons depois da transcrição do RNA.

Reads – Pequenas sequências obtidas a partir do sequenciamento do DNA fragmentado.

TRANSCRIPTOMA DE METACESTÓIDE DE *Taenia saginata*

RESUMO – *Taenia saginata* representa uma ameaça à segurança alimentar e à saúde pública devido à infecção pela ingestão de carne mal cozida e contaminada. Esta zoonose está amplamente distribuída, afetando humanos, os hospedeiros definitivos, e bovinos, hospedeiros intermediários. Além disto, a cisticercose bovina é responsável por perdas econômicas significativas devido à condenação de carcaças infectadas. A estratégia fundamental para o controle do complexo teníase-cisticercose consiste em interromper o ciclo evolutivo do parasita, evitando assim a infecção nos animais e no homem. Assim, a principal medida praticada no Brasil tem sido a inspeção de carcaças durante o abate, a qual é também a forma mais comum de diagnóstico. A insuficiente informação molecular de *T. saginata* tem dificultado o avanço das pesquisas para o aprimoramento de testes diagnósticos, o desenvolvimento de vacinas e a identificação de novas drogas para tratamento. Com o intuito de adquirir informação sobre o estágio de desenvolvimento metacestóide do parasita, foi utilizada a tecnologia de RNAseq para a montagem do transcriptoma de metacestóide de *T. saginata*. Estes dados serão úteis para estudos futuros envolvendo triagem para diagnóstico e marcadores imunoprolíficos para a cisticercose bovina.

Palavras-chave: cisticercose bovina, mRNA, RNAseq, sequenciamento NGS

Taenia saginata METACESTODE TRANSCRIPTOME

SUMMARY – *Taenia saginata* represents a threat to food security and public health in consequence of human infection by ingestion of contaminated undercooked meat. This zoonosis is worldwide distributed, affecting human, definitive host, and bovine, intermediate host. Besides, bovine cysticercosis is responsible for significant economic losses due to the condemnation of infected carcasses. The main strategy to control taeniasis-cysticercosis complex consists of interrupting the parasite's biological cycle, thus preventing human and animal infection. Therefore, in Brasil the main control measure has been the carcass inspection during the slaughter, which is also the most common diagnostic practice. Insufficient *T. saginata* molecular information makes difficult consistent advances to the improvement of diagnostic tests, vaccine development and the identification of new drugs for treatment. In order to gain insights about the parasite's metacestode developmental stage, RNAseq technology was used to assemble the whole transcriptome of a *T. saginata* metacestode. These data will also be useful for the next generation screening studies of diagnostic and immunoprophylactic markers for bovine cysticercosis.

Keywords: bovine cysticercosis, NGS sequencing, mRNA, RNAseq

CAPÍTULO 1 – CONSIDERAÇÕES GERAIS

A teníase é uma parasitose causada pelo desenvolvimento da forma adulta dos cestódeos *Taenia saginata* ou *Taenia solium* no intestino humano. Esta infecção ocorre pela ingestão de carne bovina ou suína crua ou mal cozida, respectivamente, contendo formas larvares (metacestóides) viáveis. Já a cisticercose é o desenvolvimento da forma metacestóide de *T. saginata* em bovinos e, de *T. solium*, em suínos. Os animais se infectam ao ingerirem alimentos ou água contaminados por ovos, os quais são liberados juntamente com as fezes humanas. No caso de ingestão de ovos de *T. solium* pelo homem, este também pode desenvolver a cisticercose, sendo a infecção do sistema nervoso central (neurocisticercose) a forma mais grave da doença. Portanto, o complexo teníase-cisticercose está diretamente ligado à segurança alimentar e à saúde pública.

Com a globalização, o trânsito internacional de produtos de origem animal eleva o risco de introdução/manutenção do complexo teníase-cisticercose. Sendo assim, programas de controle que envolvam práticas relacionadas à higiene, ao preparo dos alimentos e ao tratamento por anti-helmínticos são necessários, especialmente pela participação do ser humano na manutenção desta zoonose (GAJADHAR et al., 2006). Conseqüentemente, vários países têm regulamentado a produção interna e a importação e exportação de carnes visando o controle deste complexo. A limitação de exportação devido à cisticercose bovina diminui o prestígio dos países produtores bem como o valor de seus produtos, sendo necessárias ações que assegurem a qualidade da carne (ROBERTSON et al., 2014; ROSSI et al., 2014).

O Brasil é um dos maiores produtores de rebanho bovino apresentando, em 2014, um total de 208,3 milhões de cabeças e produção de 10 milhões de toneladas equivalente carcaça (TEC). Deste total, 2 milhões TEC foram exportados, sendo 80% na forma *in natura* (ABIEC, 2015). Esses dados evidenciam a importância econômica, além da segurança alimentar, devido à

desvalorização ou condenação de carcaças infectadas, e das barreiras sanitárias para a exportação desse produto *in natura*.

Taenia saginata

Espécies de helmintos que parasitam humanos datam da antiguidade, cujo início pode ter coincidido com a domesticação de animais ungulados pelo homem, embora análises filogenéticas indiquem que espécies de *Taenia* podem estar associadas ao gênero *Homo* antes mesmo do surgimento do homem moderno e do desenvolvimento de atividades agrícolas e de domesticação (HOBBERG, 2001).

O gênero *Taenia* pertence à Classe Cestoda, família Taenidae, e compreende helmintos hermafroditas, com corpo dorsoventralmente achatado e segmentado, em sua forma adulta. É possível a distinção do corpo em três regiões: na extremidade anterior localiza-se o escólex, possuindo quatro ventosas formadas por tecido muscular, seguido da região do colo, porção mais delgada e com intensa atividade celular, originando as proglótides, e o restante do corpo formado pelo estróbilo, contendo proglótides em diferentes estágios de desenvolvimento (jovens, maduras e grávidas) (SILVA, 2003).

A espécie *T. saginata* pode viver muitos anos no intestino delgado humano e atingir comprimento de seis a sete metros. A liberação de proglótides grávidas pelas fezes do hospedeiro ocorre diariamente sendo o parasita, portanto, transmitido por contaminação fecal do solo e da água, podendo ser disseminado pelo vento, por anelídeos e artrópodes. Mesmo sendo sensíveis à dessecação e à temperatura elevada, ovos de *T. saginata* podem permanecer viáveis no ambiente por aproximadamente um ano (PFUETZENREITER; ÁVILA-PIRES, 2000).

Ciclo biológico

A *T. saginata* requer, para seu ciclo de desenvolvimento, dois tipos de hospedeiros: o bovino como hospedeiro intermediário, causando a cisticercose, e

o homem, como hospedeiro definitivo, causando teníase. Dessa forma, esta zoonose se constitui um complexo denominado teníase-cisticercose (Figura 1).

Pastos e água contaminados com ovos de *T. saginata* compõem a principal via de transmissão para o bovino. Ao ingerir os ovos, estes sofrem a ação de sucos gástricos e da bile, propiciando a ativação da oncosfera (embrião hexacanto), que se evagina e se fixa na parede intestinal do hospedeiro. Num período entre 24 e 72 horas as oncosferas se difundem pelo organismo, através da corrente sanguínea e do sistema linfático, instalando-se nos tecidos musculares esqueléticos e cardíacos, formando os metacestóides (formas larvares). Num período entre 10 a 12 semanas estes metacestóides se tornam infectantes para os humanos (FAO, 2013; PFUETZENREITER; ÁVILA-PIRES, 2000).

A presença de metacestóides viáveis na carne bovina é a principal forma de infecção do homem, se ingerida crua ou mal cozida. Após ingestão, o metacestóide de *T. saginata* é ativado pela ação de sucos gástricos, evaginando e se fixando na parede do intestino delgado, pelo escólex. Após um período de aproximadamente três meses o parasita se desenvolve numa tênia adulta, iniciando a eliminação de proglótides contendo, em média, 40 mil ovos embrionados. Mesmo sendo possível a ocorrência de infecção por múltiplas tênias adultas em humanos, na maioria dos casos de teníase envolve um único parasita (FAO, 2013; RIBEIRO et al., 2012).

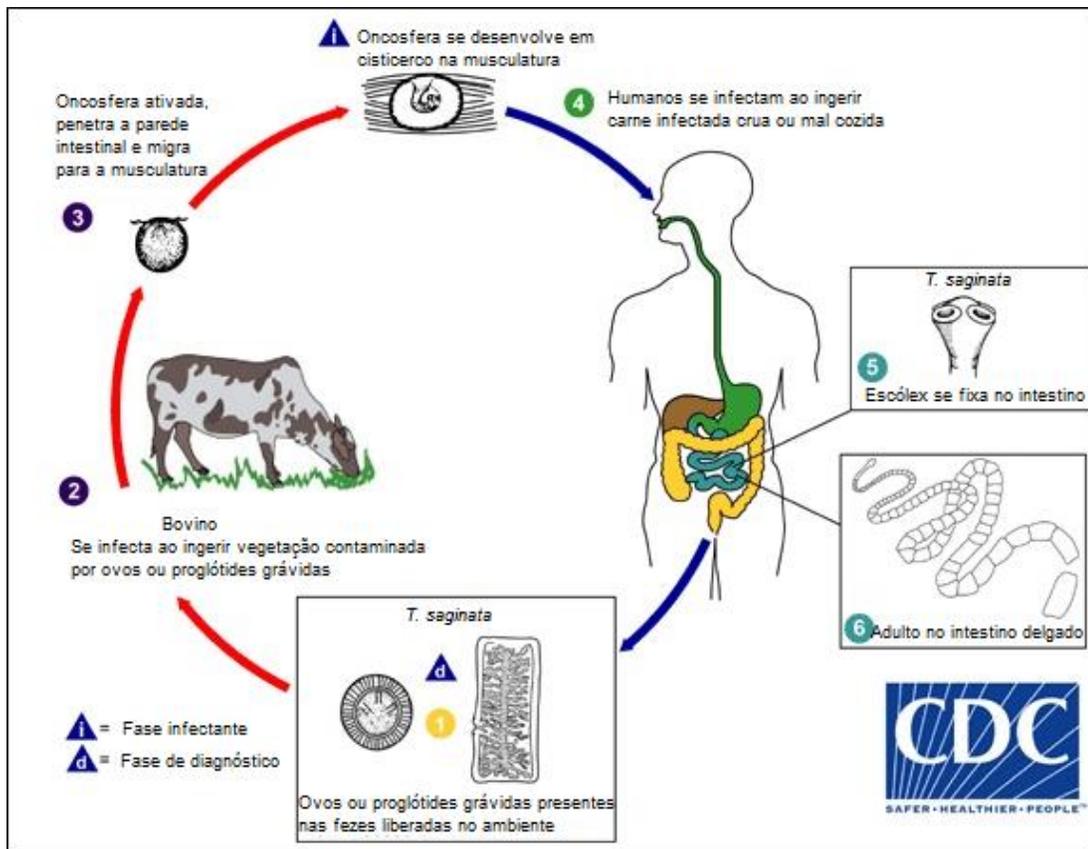


FIGURA 1 - Ciclo biológico de *Taenia saginata*, agente etiológico da cisticercose bovina e da teníase humana. *Oncosfera refere-se à forma metacestóide do parasita.

Epidemiologia do complexo Teníase-Cisticercose por *Taenia saginata*

O complexo teníase-cisticercose ocorre mundialmente, principalmente em países onde há atividade pecuária bovina mais intensa e com inapropriadas práticas de criação animal, além de precárias condições sanitárias para a população humana. Responsável por grandes perdas econômicas devido à cisticercose bovina, a infecção por *T. saginata* reflete também em questões de

saúde pública, pela conseqüente teníase humana. Estima-se que dezenas de milhares de pessoas no mundo sejam acometidas pela doença, apresentando diversos sintomas, dependendo da idade e das condições de saúde do portador, como náuseas, dores abdominais e perda de peso. Em contrapartida, bovinos infectados pela forma metacestóide (cisticerco), em geral não apresentam sinais clínicos (FAO, 2013; RIBEIRO et al., 2012; WHO, 2015).

Atualmente não há dados conclusivos sobre a prevalência da cisticercose bovina no Brasil. Contudo, um levantamento realizado por Rossi et al. (2014), considerando o abate bovino realizado em todas as regiões do país, mostrou que o sudeste apresenta os maiores índices para esta zoonose, sendo o Estado de São Paulo o que apresenta a maior prevalência (3,34%). A região sul apresenta a segunda maior prevalência, com aproximadamente 3%. A inspeção sistemática de carne em abatedouros é a principal medida de controle praticada no Brasil visando a interrupção do ciclo biológico. Durante a inspeção são realizados cortes em locais estratégicos, como na cabeça, onde se incisam os músculos masseteres e os pterigoides internos e externos. Caso haja metacestóides nos locais inspecionados, órgãos como a língua, inspecionado por palpação, e o coração, também recebem cortes durante a inspeção. Se identificadas lesões por metacestóides na língua e no coração, um exame sistemático (cortes em série) é realizado nos músculos mastigadores, diafragma, músculos do pescoço e intercostais. Dentre as normas estabelecidas para o controle do complexo teníase cisticercose destaca-se a condenação total de carcaças intensamente infectadas (contendo um ou mais metacestóides em incisões praticadas em várias partes da musculatura e numa área correspondente a aproximadamente à palma da mão). Em casos de infecção discreta ou moderada, é feito a remoção do metacestóide e do tecido circunvizinho, havendo rejeição parcial, ou seja, direcionamento para câmara frigorífica ou para desossa e tratamento por salmoura, por um período de 21 dias (BRASIL, 2015). Contudo, metacestóides desenvolvidos fora dos locais de predileção (locais definidos para inspeção) podem não ser detectados e as

carcaças são liberadas para consumo (OGUNREMI; BENJAMIN, 2010; MONTEIRO et al., 2006; WANZALA et al., 2007).

Em 2012, conforme dados fornecidos pelo Serviço de Inspeção Federal do Estado de São Paulo, foram abatidos 190.903 bovinos, dos quais 2,26% foram identificados com cisticercose. Considerando as perdas repassadas aos produtores devido à condenação ou tratamento de carcaças infectadas, o prejuízo foi de aproximadamente US\$312.194,52 (ROSSI et al., 2015).

Diagnóstico e controle

Além da inspeção sanitária das carcaças, ao abate, uma outra possibilidade de interrupção do ciclo biológico da *T. saginata* tem sido o tratamento de bovinos, o qual não é massivamente realizado, exceto quando os animais são confinados. Um dos motivos para que esta prática não seja amplamente utilizada pode ser decorrente do fato dos animais não apresentam sinais ou sintomas que justifiquem medidas terapêuticas (PEREIRA et al., 2006).

O Benzimidazol-carbamato (albendazol) apresenta um amplo espectro de atividade contra nematódeos e cestódeos em bovinos, incluindo a forma metacestóide de *T. saginata*. O tratamento de bovinos com o albendazol, além de reduzir a quantidade de metacestóides viáveis, auxilia também na redução das lesões causadas na carne (LLOYD et al., 1978), evitando maiores perdas no abate.

Entretanto, estudo utilizando o sulfóxido de albendazol, um composto sintético derivado do albendazol, para o tratamento de bovinos experimentalmente infectados por *T. saginata* relatou significativa redução da eficácia deste tratamento durante um período de dez anos, redução esta observada mesmo com o aumento da dose e com variações no intervalo de administração do anti-helmíntico (Lopes et al., 2014). Dessa forma, o método mais utilizado como controle e que também é a forma diagnóstica mais frequente do complexo teníase-cisticercose, tem sido a inspeção sanitária realizada durante o abate de bovinos.

Com o intuito de detectar bovinos infectados por *T. saginata* antes do envio para o abate, teste sorológicos têm sido explorados utilizando-se antígenos homólogos ou heterólogos extraídos de metacestóides de espécies relacionadas como *Taenia solium* e *Taenia crassiceps*, devido às reações cruzadas entre os componentes antigênicos destas espécies (OGUNREMI; BENJAMIN, 2010; ZARLENGA et al., 1994).

Dentre os antígenos já avaliados, os produtos de excreção e secreção de *T. saginata* têm apresentado melhor resultado em animais experimentalmente infectados (OGUNREMI; BENJAMIN, 2010). Por outro lado, animais com infecção natural foram mais bem diagnosticados quando se utilizou o antígeno de fluido vesicular de *T. solium* (PAULAN et al., 2013). Todavia, a sensibilidade dos testes ainda se mostra insuficiente quando aplicados a animais naturalmente infectados, geralmente por apresentarem baixa carga parasitária (MONTEIRO et al., 2006).

Devido ao fato dos ensaios para a detecção de anticorpos indicarem exposição à infecção e não necessariamente uma infecção ativa, ensaios sorológicos para a detecção de antígenos circulantes também passaram a ser considerados. Anticorpos reativos com glicoproteínas de superfície/excreção (HARRISON et al., 1989) e específicos para produtos de excreção/secreção (DORNY et al., 2000; BRANDT et al., 1992) de formas metacestóides de *T. saginata* têm sido utilizados em diagnóstico sorológico para bovinos infectados. Contudo, a sensibilidade dos testes ainda deixa a desejar, não tendo sido possível estabelecer um método de diagnóstico sorológico *ante mortem* aplicável para a identificação e tratamento de animais naturalmente infectados, que geralmente apresentam baixa carga parasitária (PAULAN et al., 2013).

Apesar da resposta imune dos bovinos infectados não estar completamente compreendida, o fato de que proteínas de oncosferas de *T. saginata* podem conferir imunidade ao hospedeiro tem incentivado o desenvolvimento de antígenos vacinais já há alguns anos. Os antígenos recombinantes TSA-18 e TSA-9, extraídos de oncosfera de *T. saginata*, foram testados em bovinos infectados

apresentando elevados níveis de proteção, embora não tenha se estabelecido o período pelo qual a imunidade permaneceu ativa (LIGHTOWLERS et al., 1996). Outra abordagem tem sido o desenvolvimento de peptídeos sintéticos como o GK-1, de *T. crassiceps*, formado por 18 aminoácidos, o qual foi testado em camundongos, resultando numa taxa de proteção entre 64,7 e 96% (FRAGOSO et al., 2011).

Estudos em biologia molecular têm auxiliado na identificação de marcadores moleculares, possibilitando o delineamento de ferramentas para o diagnóstico, mesmo quando perdidas as informações morfológicas do parasita, além de permitirem a identificação e caracterização de diferentes isolados dentro de uma mesma espécie. Exemplos disto foram reportados por Sato et al. (2011) que, utilizando os marcadores de DNA nuclear *Ag1V1* e *Ag2*, identificaram dois genótipos de *T. solium* (Afro-Americano e Asiático) e por Yamane et al. (2012), que observaram a ocorrência de hibridização entre as espécies *T. saginata* e *T. asiática* ao estudarem o gene mitocondrial *cox1* e os nucleares *ef1* e *elp* nestes parasitas.

A adaptação ao parasitismo de quatro espécies de cestódeos (*Echinococcus multiloculares*, *E. granulosus*, *T. solium* e um modelo laboratorial, *Hymenolepis micróstoma*), foi descrita em consequência da possibilidade de sequenciamento do genoma e do transcriptoma destas espécies resultando, por exemplo, na identificação de vias de desintoxicação e de metabolismo de nutrientes do hospedeiro, além de novos alvos para drogas (Tsai et al., 2013).

Recentemente, a análise do conjunto de proteínas que compõem o secretoma de *T. solium* identificou proteínas de excreção/secreção de interesse terapêutico e diagnóstico (Gomez et al., 2015). Nesta questão, o desenvolvimento de técnicas para o sequenciamento de RNA (RNAseq) e de ferramentas em bioinformática para a montagem do conjunto completo de transcritos (transcriptoma), produzidos por uma dada célula/organismo, compõem importante estratégia para a identificação de novos antígenos.

Sequenciamento de RNA (RNAseq)

A tentativa de se conhecer o transcriptoma de células/organismos se iniciou com técnicas de baixa abrangência, como o *Northern Blot*, que exigia grande quantidade de RNA para a obtenção de poucos transcritos para cada amostra. Com o desenvolvimento de reação em cadeia da polimerase por transcrição reversa (RT-PCR) a capacidade de análise aumentou para algumas centenas de transcritos, sendo superado pela tecnologia de *microarray*, com a qual é possível a detecção de milhares de transcritos. Além disto, estas técnicas possibilitam a caracterização de níveis de expressão (MOROZOVA et al., 2009).

De forma intermediária, o desenvolvimento da técnica de análise serial da expressão gênica (SAGE) permitiu a análise do perfil de transcrição, chegando próximo da identificação de todo um transcriptoma (VELCULESCU et al., 1995).

Entretanto, uma revolução em estudos sobre transcriptomas ocorreu com o desenvolvimento da técnica de sequenciamento de RNA (RNAseq), a qual utiliza as mesmas plataformas desenvolvidas para o sequenciamento de DNA genômico (Illumina, Roche 454, etc.) para o sequenciamento de fragmentos de DNA complementar (cDNA) (OZSOLAK; MILOS, 2011).

Num modelo típico, o sequenciamento de RNA se inicia com a extração de RNA total de uma dada amostra, seguida de fragmentação e conversão em sequências de cDNA. Adaptadores são então ligados a ambas as extremidades de cada sequência, podendo estas passar por amplificação prévia ao sequenciamento. Cada fragmento pode ser sequenciado em uma ou ambas as extremidades, resultando em sequências *single-end* ou *paired-end*, respectivamente. O tamanho das sequências pode variar entre 30-400 pares de base (pb), dependendo do tipo de tecnologia utilizada (WANG et al., 2009) (Figura 2).

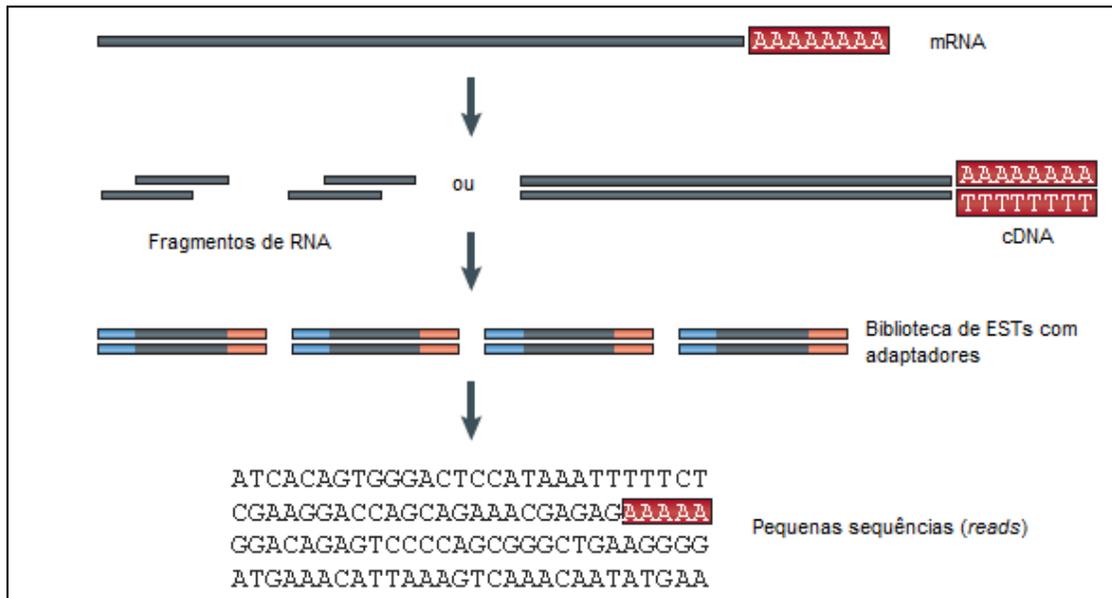


FIGURA 2 - Etapas para o sequenciamento de RNA (RNAseq). Fonte: WANG et al. (2009).

Montagem do transcriptoma

O sequenciamento de RNA representou grande avanço na obtenção da informação funcional de organismos. Porém, também tem resultado em desafios constantes para o desenvolvimento de ferramentas capazes de armazenar e processar grandes quantidades de dados, além de montar transcritos completos e suas variantes (isoformas) (MARTIN; WANG, 2011; WANG et al, 2009).

Havendo o genoma da espécie de interesse disponível, a montagem do transcriptoma pode ser feita de forma guiada. Nesta estratégia os transcritos sequenciados são alinhados contra o genoma de referência, diminuindo o risco de erros, uma vez que seqüências artefatos não se alinharão. Além disto, a existência de uma referência possibilita a detecção de novos transcritos, mesmo com baixa abundância (MARTIN; WANG, 2011) (Figura 3).

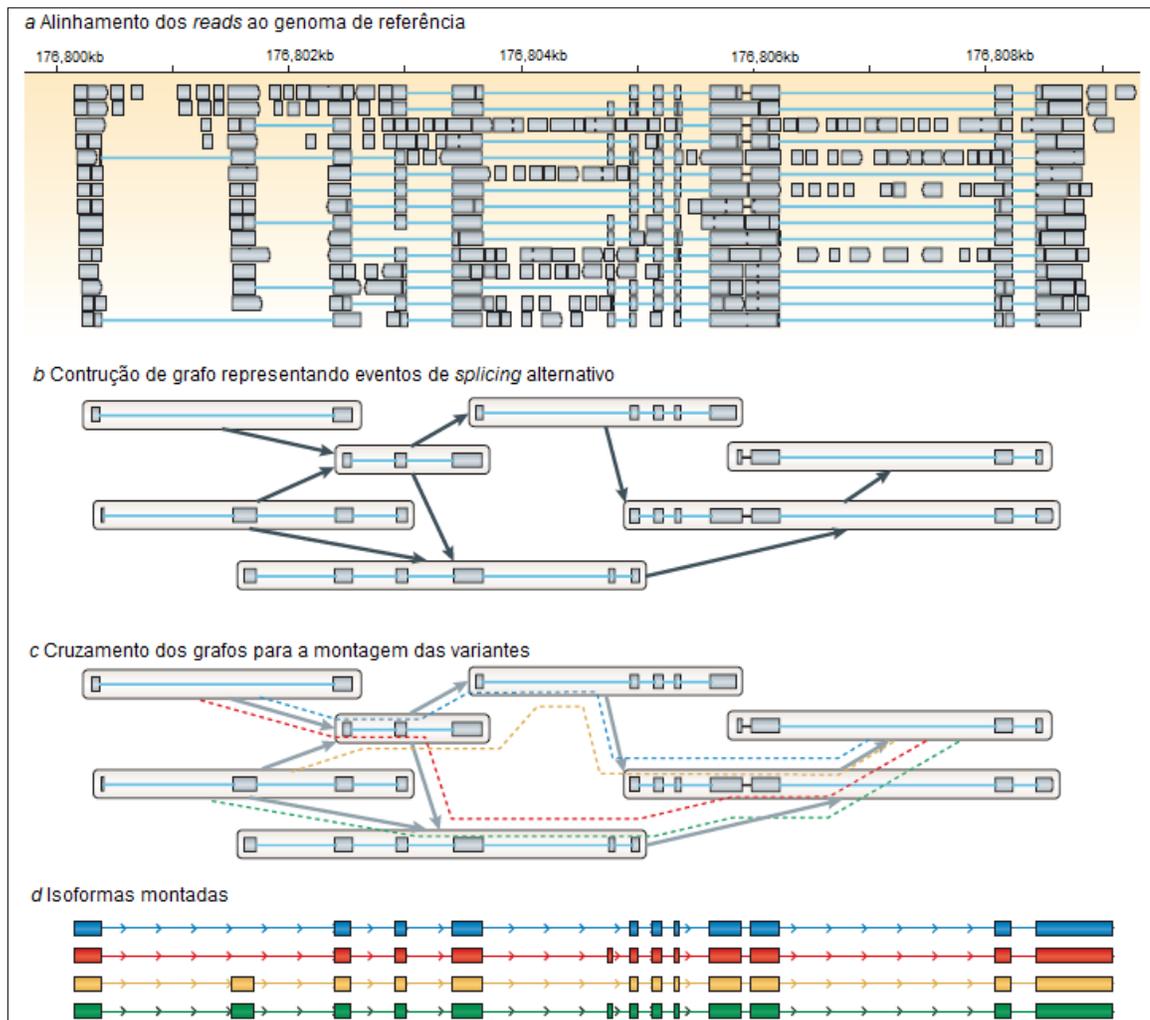


FIGURA 3 - Estratégia de montagem de transcriptoma guiado por referência.

Fonte: MARTIN; WANG (2011).

Entretanto, alguns organismos não possuem um genoma de referência. Para enfrentar este desafio, utiliza-se o método *de novo*, já empregado para genomas. Esta técnica utiliza a redundância de sequências pequenas (*reads*), encontrando sobreposições, e a construção de grafos *de Bruijn* para a montagem completa dos transcritos e suas variantes. Nesta metodologia, cada nó é definido por uma sequência de comprimento determinado de k nucleotídeos (k -*mer*) e

estes nós são conectados por arestas, caso ocorra sobreposição perfeita de $k-1$ nucleotídeos. Cada caminho do grafo representa um possível transcrito (GRABHERR et al., 2013; MARTIN; WANG, 2011).

Há alguns programas que utilizam esta estratégia para montagem *de novo* de transcriptomas. Por exemplo, o Trinity, um pacote de programas que combina três componentes: *Inchworm*, *Chrysalis* e *Butterfly*. No primeiro passo, *Inchworm* reconstrói *contigs* seguindo os passos: 1) construção de dicionário de k -mer para todos os *reads* sequenciados; 2) remoção de k -mers que possam conter erros; 3) seleção do k -mer mais frequente para servir de início da montagem do *contig*; 4) extensão, em ambas as extremidades, utilizando o método de sobreposição $k-1$; 5) finalizada a extensão, o *contig* é reportado; 6) repetição dos passos 3-5 até que todo o dicionário tenha sido analisado. Em seguida, *Chrysalis* agrupa os *contigs* minimamente sobrepostos, formando conjuntos de componentes (*contigs* que derivam de *splicing* alternativos ou que são parálogos) e grafos *de Bruijn* são construídos para cada um destes componentes. Finalmente, *Butterfly* reconstrói transcritos completos através de interligação de cada grafo gerado no passo anterior, utilizando os *reads* originais. Além disto, esta etapa remove ambiguidades decorrentes de erros ou de sequências, compartilhadas entre diferentes transcritos, com tamanho acima de k bases (Figura 4) (GRABHERR et al., 2013).

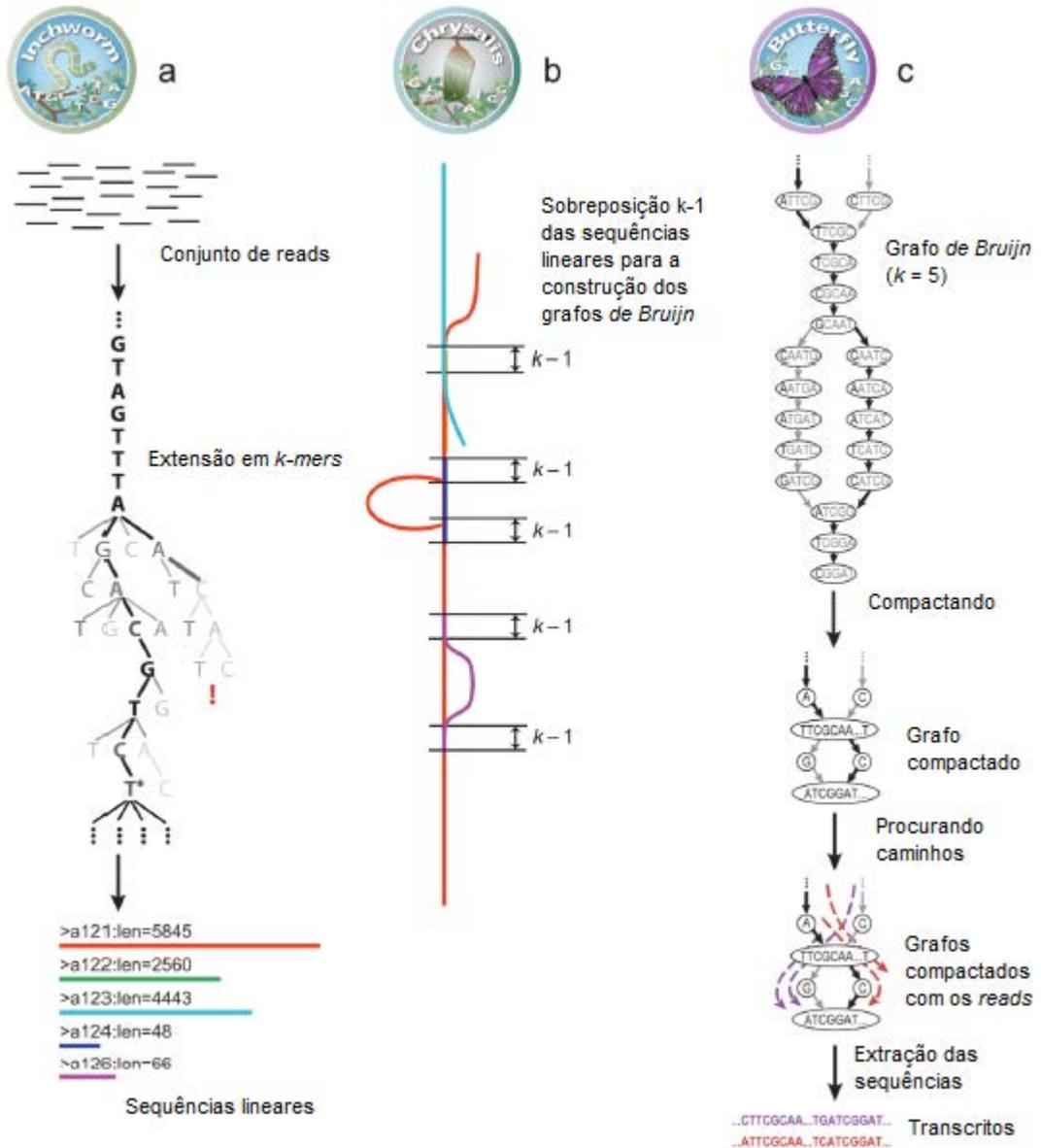


FIGURA 4 - Metodologia Trinity para montagem *de novo* de transcriptoma completo. Fonte: GRABHERR et al., 2013.

Importantes avanços já foram conquistados no campo da parasitologia com o sequenciamento do transcriptoma de espécies relevantes para a saúde pública, como o de *Schistosoma mansoni* (VERJOVSKI-ALMEIDA et al., 2003), de

Trypanosoma brucei (KOLEV et al., 2010) e de *T. solium* (ALMEIDA et al., 2009). Outras espécies de *Taenia* também tiveram seus transcriptomas sequenciados e montados, como o de *T. multiceps* (WU et al., 2012) e de *T. pisiformis* (YANG et al., 2012).

Neste contexto, o presente trabalho propôs a realização do sequenciamento e da montagem do transcriptoma da forma metacestóide de *T. saginata*, um dos agentes etiológicos do complexo teníase-cisticercose, responsável por riscos à segurança alimentar e à saúde pública, bem como por elevadas perdas econômicas na cadeia de carne.

Objetivo

Realizar o sequenciamento e a montagem do transcriptoma da forma metacestóide de *T. saginata*.

Referências

ABIEC – ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS EXPORTADORAS DE CARNE. **Estatística:** balanço da pecuária. Disponível em: <<http://www.abiec.com.br/texto.asp?id=8>>. Acesso em: 5 out. 2015.

ALMEIDA, C. R.; STOCO, P. H.; WAGNER, G.; SINCERO, T. C. M.; ROTAVA, G.; BAYER-SANTOS, E.; RODRIGUES, J. B.; SPERANDIO, M. M.; MAIA, A. A. M.; OJOPI, E. P. B.; ZAHA, A.; FERREIRA, H. B.; TYLER, K. M.; DÁVILA, A. M. R.; GRISARD, E. C.; DIAS-NETO, E. Transcriptome analysis of *Taenia solium* cysticerci using Open Reading Frame ESTs (ORESTES). **Parasites & Vectors**, v. 2, doi: 10.1186/1756-3305-2-35, 2009.

BRANDT, J. R. A.; GEERTS, S.; DE DEKEN, R.; KUMAR, V.; CEULEMANS, F.; BRIJS, L.; FALLA, M. A monoclonal antibody-based ELISA for the detection of circulating excretory-secretory antigens in *Taenia saginata* cysticercosis. **Internacional Journal for Parasitology**, v. 22, p. 471-477, 1992.

BRASIL. Decreto 30.691 de 29/03/52, alterado pelo Decreto 8.444 de 07/05/2015. Dispõe sobre o regulamento de inspeção industrial e sanitária de produtos de origem animal. **Diário Oficial da União**, Atos do Poder Executivo, Brasília, DF, 06 maio 2015, Seção 1, p. 1.

DORNY, P.; VERCAMMEN, F.; BRANDT, J.; VANSTEENKIST, W.; BERKVENS, D.; GEERTS, S. Sero-epidemiological study of *Taenia saginata* cysticercosis in Belgian cattle. **Veterinary Parasitology**. v. 88, p. 43-49, 2000.

FAO/OMS. **Summary risk profile on *C. bovis* in meat from domestic cattle** [Internet]. 2013. Available: http://www.fao.org/fileadmin/user_upload/agns/pdf/Foodborne_parasites/RiskProfT_aeniasaginataOct2013.pdf

FERRER, E.; BONAY, P.; FOSTER-CUEVAS, M.; GONZÁLEZ, L. M.; DÁVILA, I.; CORTÉZ, M. M.; HARRISON, L. J. S.; PARKHOUSE, R. M. E.; GÁRATE, T. Molecular cloning and characterization of Ts8B1, Ts8B2 and Ts8B3, three new members of the *Taenia solium* metacestode 8kDa diagnostic antigen family. **Molecular and Biochemical Parasitology**, v. 152, p. 90-100, 2003.

FERRER, E.; GONZÁLEZ, L. M.; MARTÍNEZ-ESCRIBANO, J. A.; GONZÁLEZ-BARDERAS, M. E.; CORTÉZ, M. M.; DÁVILA, I.; HARRISON, L. J. S.; PARKHOUSE, R. M. E.; GÁRATE, T. Evaluation of recombinant HP6-Tsag, an 18

kDa *Taenia saginata* oncospherical adhesion protein for the diagnosis of cysticercosis. **Parasitology Research**, v. 101, p.517-525, 2007.

FRAGOSO, G.; ESQUIVEL-GUADARRAMA, F.; SANTANA, M. A.; BOBES, R. J.; HERNÁNDEZ, B.; CERVANTES, J.; SEGURA, R.; GOLDBAUM, F. A.; SCIUTTO, E.; ROSAS, G. Heterologous prime-boost oral immunization with GK-1 peptide from *Taenia crassiceps* cysticerci induces protective immunity. **Clinical and Vaccine Immunology**, v. 18, p. 1067-1076, 2011.

GAJAGHAR, A. A.; SCANDRETT, W. B.; FORBES, L. B. Overview of food- and water-borne zoonotic parasites at the farm level. **Revue scientifique et technique (International Office of Epizootics)**, v. 25, p. 595-606, 2006.

GOMEZ, S.; PERALTA, L. A.; FONSECA, H. P.; ROBLES, V. A. C.; SOBERÓN, X.; SCIUTTO, E.; FRAGOSO, G.; BOBES, R. J.; LACLETTE, J. P.; YAUNER, L. P.; LEYVA, A. O. Genome analysis of excretory/secretory proteins in *Taenia solium* reveals their abundance of antigenic regions (AAR). **Scientific Reports**, doi: 10.1038/srep09683, 2015.

GRABHERR, M. G.; HAAS, B. J.; YASSOURT, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACHOEN, N.; GNIRKE, A.; RHIND, N.; PALMA, F.; BIRREN, B. W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. **Nature Biotechnology**, v. 29, p. 644-652, 2013.

HAAS, B. J.; PAPANICOLAOU, A.; YASSOUR, M.; GRABHERR, M.; PHILIP, D.; BOWDEN, J.; COUGER, M. B.; ECCLES, D.; LI, B.; LIEBER, M.; MacMANES, M. D.; OTT, M.; ORVIS, J.; POCHE, N.; STROZZI, F.; WEEKS, N.; WESTERMAN,

R.; WILLIAM, T.; DEWEY, C. N.; HENSCHER, R.; LEDUC, R. D.; FRIEDMAN, N.; REGEV, A. *De novo* transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. **Nature Protocols**, v. 8, p. 1-43, 2013. doi:10.1038/nprot.2013.084.

HARRISON, L. J. S.; JOSHUA, G. W. P.; WRIGHT, S. H.; PARKHOUSE, R. M. E. Specific detection of circulating surface/secreted glycoproteins of viable cysticerci in *Taenia saginata* cysticercosis. **Parasite Immunology**, v.11, p.351-370, 1989.

HOBERG, E. P.; ALKIRE, N. L.; QUEIROZ, A.; JONES, A. Out of Africa: origins of the *Taenia* tapeworms in humans. **The Royal Society**, 2001. doi:10.1098/rspb2000.1579.

KOLEV, N. G.; FRANKLIN, J. B.; CARMI, S.; SHI, H.; MICHAELI, S.; TSCHUDU, C. The transcriptome of the human pathogen *Trypanosoma brucei* at single-nucleotide resolution. **Plos Pathogens**, v. 6, e1001019, 2010. doi:10.1371/journal.ppat.1001090.

LIGHTOWLERS, M. W.; ROLFE, R.; GAUCI, C. G. *Taenia saginata*: vaccination against cysticercosis in cattle with recombinant oncosphere antigens. **Experimental Parasitology**, v. 84, p. 330-338, 1996.

LLOYD, S.; SOULSBY, E. J. L.; THEODORIDES, V.J. Effect of albendazole on the metacestodes of *Taenia saginata* in calves. **Experientia**, v. 34, p. 734-724, 1978.

LOPES, W. D. Z.; CRUZ, B. C.; SOARES, V. E.; NUNES, J. L. N.; TEIXEIRA, W. F. P.; MACIEL, W. G.; BUZZULINI, C.; PEREIRA, J. C. M.; FELIPELLI, G.; SOCCOL, V. T.; OLIVEIRA, G. P.; COSTA, A. J. Historic of therapeutic efficacy of albendazol sulphoxide administered in different routes, dosages and treatment

schemes, against *Taenia saginata* cysticercus in cattle experimentally infected. **Experimental Parasitology**, v. 137, p. 24-20, 2014.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. **Nature Reviews Genetics**, v. 12, p. 671-682, 2011.

MONTEIRO, L.L.; PINTO, P. S. A.; DIAS, F. S. Evaluation of the ELISA test for the antibody detection in cattle naturally and experimentally infected with *Cysticercus bovis*. **Veterinary Parasitology**, v. 141, p. 260-263, 2006.

MOROZOVA, O.; HIRST, M.; MARRA, M. A. Applications of new sequencing technologies for transcriptome analysis. **Annual Review of Genomics and Human Genetics**, v. 10, p. 135-151, 2009.

MONTEIRO, L.L.; PINTO, P. S. A.; DIAS, F. S. Evaluation of the ELISA test for the antibody detection in cattle naturally and experimentally infected with *Cysticercus bovis*. **Veterinary Parasitology**, v. 141, p. 260-263, 2006.

OGUNREMI, O.; BENJAMIN, J. Development and field evaluation of a new serological test for *Taenia saginata* cysticercosis. **Veterinary Parasitology**, v. 169, p. 93-101, 2010.

OZSOLAK, F.; MILOS, P. M. RNA sequencing: advances, challenges and opportunities. **Nature Reviews Genetics**, v. 12, p. 87-98, 2011.

PAULAN, S. C.; GONZÁLEZ, R. M. H.; PERALTA, L. A.; VICNETINI-OLIVEIRA, J. C.; BIONDI, G. F.; CONDE, E. S.; PARKHOUSE, R. M. E.; NUNES, C. M. Usefulness of serological ELISA assay for *Taenia saginata* to detect naturally

infected bovines. **Revista Brasileira de Parasitologia Veterinária**, v. 22, p. 270-275, 2013.

PEREIRA, M. A. V. C.; SCHWANZ, V. S.; BARBOSA, C. G. Prevalência da cisticercose em carcaças de bovinos abatidos em matadouros-frigoríficos do Estado do Rio de Janeiro, submetidos ao controle do serviço de inspeção federal (SI-RJ), o período de 1997 a 2003. **Arquivos do Instituto Biológico**, v.73, p.83-87, 2006.

PFUETZENREITER, M. R.; ÁVILA-PIRES, F. D. Epidemiologia da Teníase/Cisticercose por *T. solium* e *T. saginata*. **Ciência Rural**, v. 30, p. 541-548, 2000.

RIBEIRO, N. S. A.; TELLES, E. O.; BALIAN, S. C. O complexo teníase humana-cisticercose: ainda um sério problema de saúde pública. **Revista de Educação Continuada em Medicina Veterinária e Zootecnia**, v. 10, p. 20-25, 2012.

ROBERTSON, L. J.; SPRONG, H.; ORTEGA, Y. R.; VAN DER GIESSEN, J. W. B.; FAYER, R. Impacts of globalization of foodborne parasites. **Trends in Parasitology**, v. 30, p. 37-52, 2014.

ROSSI, G. A. M.; GRISÓLIO, A. P. R.; PRATA, L. F.; BÜRGER, K. P.; HOPPE, E. G. L. Situação da cisticercose bovina no Brasil. **Semina: Ciências Agrárias**, v. 35, p. 927-938, 2014.

ROSSI, G. A. M.; HOPPE, E. G. L.; MATHIAS, L. A.; MARTINS, A. M. C.V.; MUSSIC, L. A.; PRATA, L. F. Bovine cysticercosis in slaughtered cattle as an indicator of Good Agricultural Practices (GAP) and epidemiological risk factors. **Preventive Veterinary Medicine**, v. 118, p. 504-508, 2015.

SATO, M. O.; SAKO, Y.; NAKAO, M.; WANDRA, T.; NAKAYA, K.; YANAGIDA, T.; ITO, A. A possible nuclear DNA marker to differentiate the two geographic genotypes of *Taenia solium* tapeworms. **Parasitology International**, v. 60, p. 108-110, 2011.

SILVA, A. M. Teníase e cisticercose. In: NEVES, D. P. **Parasitologia humana**. 10^a ed. Ribeirão Preto: Atheneu, 2003. Cap. 25, p. 227-238.

TSAI, I.; ZAROWIECKI, M.; HOLROYD, N.; GARCIARRUBIOS, A.; SANCHES-FLORES, A.; BROOKS, K.; 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-STANCHEVA, N.; RIDDIFORD, N.; ROSENZVIT, M.; SALINAS, G.; WASMUTH, J. D.; ZAMANIAN, M.; ZHENG, Y.; THE *Taenia solium* GENOME CONSORTIUM; CAI, X.; SOBERÓN, X.; OLSON, P. D.; LACLETTE, J. P.; BREHM, K.; BERRIMAN, M. The genomes of four tapeworm species reveal adaptations to parasitism. **Nature**, v. 496, p. 57-63, 2013.

VELCULESCU, V. E.; ZHANG, L. VOGELSTEIN, B.; KINZLER, K. W. Serial Analysis of Gene Expression. **Science**, v. 270, p. 484-487, 1995.

VERJOVSKI-ALMEIDA, S.; DEMARCO, R.; MARTINS, E. A. L.; GUIMARÃES, P. E. M.; OJOPI, E. P. B.; PAQUOLA, A. C. M.; PIAZZA, J. P.; NISHIYAMA, M. Y.; KITAJIMA, J. P.; ADAMSON, R. E.; ASHTON, P. D.; BONALDO, M. F.; COULSON, P. S.; DILLON, G. P.; FARIAS, L. P.; GREGÓRIO, S. P.; HO, P. L.;

LEITE, R. A.; MALAQUIAS, L. C. C.; MARQUES, R. C. P.; MIYASATO, P. A.; NASCIMENTO, A. L. T.O.; OHLWEILWE, F. P.; REIS, E. M.; RIBEIRO, M. A.; SÁ, R. G.; STUKART, G. C.; SOARES, M. B. S.; GARGIONI, C.; KAWANO, T.; RODRIGUES, V.; MADEIRA, A. M. B. N.; WILSON, R. A.; MENCK, C. F. M.; SETUBAL, J. C.; LEITE, L. C. C.; DIAS-NETO, E. Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. **Nature Genetics**, v. 35, p. 148-157, 2003.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, p. 57-63, 2009.

WANZALA, W.; KYULE, N. M.; ZESSIN, K. H.; ONYANGO-ABUJE, A. J.; KANG'ETHE, K. E.; OCHANDA, H.; HARRISON, L. J. S. Evaluation of an antigen-ELISA in the diagnosis of bovine cysticercosis in Kenyan cattle. **Parasitology Research**, v. 100, p.539-548, 2007.

WHO – WORLD HEALTH ORGANIZATION. **Teniasis/Cysticercosis**. Disponível em: <<http://www.who.int/zoonoses/diseases/taeniasis/en>>. Acesso em: 13 out. 2015.

WU, X.; FU, Y.; YANG, D.; ZHANG, R.; ZHENG, W.; NIE, H.; XIE, Y.; YAN, N.; HAO, G.; GU, X.; WANG, S.; PENG, X.; YANG, G. Detailed transcriptome description of the neglected cestode *Taenia multiceps*. **Plos One**, v. 7, e45830, 2012. doi:10.1371/journal.pone.0045830.

YAMANE, K.; SUZUKI, Y.; TACHI, E.; LI, T.; CHEN, X.; NAKAO, M.; NKOUAWA, A.; YANAGIDA, T.; SAKO, Y.; ITO, A.; SATO, H.; OKAMOTO, M. Recent hybridization between *Taenia asiatica* and *Taenia saginata*. **Parasitology International**, 2012, doi:10.1016/j.parint.2012.01.005.

YANG, D.; FU, Y.; WU, X.; XIE, Y.; NIE, H.; CHEN, L.; NONG, X.; GU, X.; WANG, S.; PENG, X.; YAN, N.; ZHANG, R.; ZHENG, W.; YANG, G. Annotation of the transcriptome from *Taenia pisiformis* and its comparative analysis with three Taeniidae species. **Plos One**, v. 7, e32283, 2012. doi:10.1371/journal.pone.0032283.

ZARLENGA, D. S.; RHOADS, M. L.; AL-YAMAN, F. M. A *Taenia crassiceps* cDNA sequence encoding a putative immunodiagnostic antigen for bovine cysticercosis. **Molecular and Biochemical Parasitology**, v. 67, p. 215-223, 1994.

CAPÍTULO 2 – Montagem do transcriptoma de metacestóide de *Taenia saginata**

***Taenia saginata* transcriptome assembly**

Abstract - *Taenia saginata* is a tapeworm worldwide distributed that affects human and cattle, representing a threat to food security and public health due to human infection, following ingestion of contaminated undercooked meat. Bovine cysticercosis, caused by the development of the metacestode stage, is responsible for significant economic losses due to the condemnation of infected carcasses. The lack of *T. saginata* genomic knowledge poses a challenge to the improvement of diagnosis and vaccine research. In order to gain insights about the genomic make up of parasite's metacestode developmental stage, the whole transcriptome of a *T. saginata* metacestode was assembled using RNAseq technology. This work aims to contribute as an initial step towards the identification of novel targets for the diagnosis and control of bovine cysticercosis. A total of 34,021 gene isoforms, related to 16,641 genes were assembled using Trinity package software. Assembled transcripts were aligned against *T. solium* and *T. asiatica* dataset, using BLASTx and BLASTp, resulting in 44% and 36% of the alignments, respectively. Functional annotation was performed using Trinotate pipeline, KEGG and COG databases. Abundance analysis and functional annotation of the predicted transcripts suggest the parasite express a large number of genes which participate in a wide range of cellular maintenance processes, as well as host interactions via. The present results are suitable for further screening studies of diagnostic and immunoprophylactic markers for bovine cysticercosis.

Keywords: RNAseq, next generation sequencing, mRNA, bovine cysticercosis

* De acordo com as normas de formatação da revista PlosOne.
(disponível em: <http://journals.plos.org/plosone/s/submission-guidelines>)

Introduction

Taenia saginata is a cestode tapeworm with worldwide occurrence representing not only a threat for food security and public health but also an important cause of economic losses in the meat industry. Typically, the adult form of *T. saginata* grows in the human small bowel producing eggs which are released to the environment through human waste. Eggs from contaminated pasture or water sources can be ingested by the intermediate host (cattle) where the oncospheres are activated by gastro enteric enzymes. The oncospheres are carried through the blood stream, migrate to skeletal and cardiac muscles, where they encyst and achieve the metacestode stage. Finally, human can be infected by ingesting undercooked meat containing metacestode [1], maintaining the biological cycle.

Depending upon the extent of the infection, bovine carcasses can be completely condemned or treated by physical procedures before human consumption [2], [3]. In both cases, producers have significant economic losses. For instance, even lightly contaminated carcass can be depreciated up to 30% of its market value in Brazil [4].

Efforts have been made to improve diagnostic tests to identify and treat animals before slaughter [5], [6], [7], and to develop vaccines in order to interrupt the parasite biological cycle [8], [9], [10], but molecular information about the *T. saginata* metacestode is still scarce, struggling the consistent advances in this area. The first obvious target for treatment, diagnosis and immunoprophylaxis is the metacestode developmental stage, since it is directly responsible for human infection and bovine carcass condemnation.

In order to gain insight about the parasite's metacestode developmental stage, we employed RNAseq technology and successfully assembled the whole transcriptome of a *T. saginata* metacestode. This information generated herein

contributes as an important step towards the identification of novel targets for diagnosis and treatment of bovine cysticercosis.

Material and methods

Total RNA extraction and cDNA library

A single *T. saginata* metacestode was collected from a naturally infected bovine carcass and conserved in liquid nitrogen. Total RNA was extracted using the standard Trizol[®] (Life Technologies) protocol, following the manufacturer's instructions. Messenger RNA (mRNA) was isolated using magnetic oligo(dT) beads, followed by fragmentation and construction of the complementary DNA (cDNA) library, using the TruSeq RNA Library Preparation Kit v2 (Illumina) with random hexamer primers, followed by fragment ends repair and phosphorylation. In order to facilitate ligation of the adapters, an adenine base was added to the blunted ends. Finally, fragments were denatured, amplified by PCR using random primers, and checked for quality and integrity using Bioanalyzer[®] (Agilent). This library preparation yielded inserts of an average size of 270 bases (**Figure S1**).

Sequencing and data processing

DNA sequencing by synthesis was performed using the MiSeq Reagent kit v3, and the Illumina MiSeq[®] sequencing platform. After cDNA library preparation, fragments ligated to adapters were hybridized to the flow cell and clusters were then generated. A total of 2 x 23,549,932 paired-end reads (47,099,864 reads) with 200 bases in length were generated. Trimmomatic v0.33 software [11] was used to remove adapters from both ends of the reads and also bases with quality score below 10 (base call error probability 0.1). Additionally, poly-A tails and low complexity sequences were removed with Prinseq v0.20.4 [12]. As the read length was approximately three quarters of the average insert size, most of the paired-end reads presented overlapping bases. Therefore, the paired-end reads were merged

to produce longer single sequences. This step was performed using PEAR v0.9.6 software [13], with a minimum overlapping criteria of 100 bases.

Transcriptome assembly

De novo transcriptome assembly of *T. saginata* metacestode was performed using Trinity v2.0.6 software package [14], [15]. Briefly, Trinity analysis consisted of three steps: contigs were assembled by the Inchworm pipeline, Chrysalis was used to cluster the contigs by similarity and *de Bruijn* graphs were built for each cluster. Then, individual graphs were processed by the Butterfly algorithm to construct unique and alternatively spliced transcripts. Finally, prediction of coding regions within transcript sequences was performed using TransDecoder (available at: <https://github.com/TransDecoder/TransDecoder/releases>). Transcript abundance was estimated using the RSEM (RNA-Seq by Expectation-Maximization) [16] algorithm in Trinity. For each transcript, relative abundance was expressed in terms of fragment per kilobase of exon per million mapped fragments (FPKM). We classified the transcripts according to the distribution of FPKM values: poorly represented for $FPKM < 1$, moderately represented for $1 \leq FPKM < x$, and highly represented for $FPKM > x$, where x represents the 90th percentile of the FPKM distribution.

Functional annotation

Assembled transcripts had their nucleotide and predicted amino acid sequences aligned against reference protein sequences using BLASTx and BLASTp functions (available at: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) [17], respectively, considering an e-value threshold of 10^{-20} . The reference set included protein sequences from the closely related cestode species *Taenia solium* [18] (available at: <http://www.genedb.org/Homepage/Tsolium>), *Taenia asiatica* (available at: http://parasite.wormbase.org/Taenia_asiatica_prjeb532/Info/Index/),

Echinococcus multilocularis [18] (available at: <http://www.genedb.org/Homepage/Emultilocularis>), *Echinococcus granulosus* [18] (available at: <http://www.genedb.org/Homepage/Egranulosus>) and *Hymenolepis microstoma* [18] (available at: <http://www.genedb.org/Homepage/Hmicrostoma>), as well as from the model organisms *Caenorhabditis elegans* [19] (available at: http://www.ensembl.org/Caenorhabditis_elegans/Info/Index) and *Schistosoma mansoni* [20] (available at: <http://www.genedb.org/Homepage/Smansoni>). The resulting alignments were used for functional annotation in Trinotate v2.0.2 (available at: <http://www.vcru.wisc.edu/simonlab/bioinformatics/programs/trinity/docs/annotation/Trinotate.html>), which included protein domain identification (Pfam) [21] by hidden Markov models (HMMER) [22], homology to ribosomal RNA sequences (RNAMMER) [23], prediction of signal peptide cleavage sites (signalP) [24] and transmembrane domain (tmHMM) [25], and comparison to curated annotation databases (Uniprot and GO pathway).

Genes were also grouped into functional categories using the Cluster of Orthologous Groups of proteins (COG) database [26]. Complementary gene orthology prediction was performed by aligning the amino acid sequences against the protein library of the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [27] (available at: www.genome.jp/tools/kaas/) (**Table S1**). Additionally, for the set of highly represented genes, we conducted functional analyses also using the KEGG database.

Results

Total RNA from a single *T. saginata* metacestode was isolated to prepare a cDNA library for paired-end Illumina sequencing, generating a total of 2 x 23,549,932 reads, 200 bp in length. For quality control of the reads, adapter contamination was assessed and low complexity reads and poly-A tails were

removed using Trimmomatic and Prinseq software. After this process, a total of 2 x 20,050,839 paired-end reads (85.1%) were retained for further analysis (**Figure S2**).

As most of the paired reads presented overlapping bases, the PEAR software was used to perform the merging of the overlapping reads. Thus, a total of 18,806,272 longer single reads were generated, with an average size of 168 ± 44 bases. The assembly of the merged reads resulted in 34,028 isoforms with N50 of 2,814 bases, which were assigned to 16,641 unique components (genes). In addition to the quality control process of the reads, that improves the quality of the assembly, RSEM algorithm was also used to filter Trinity results by FPKM. Transcripts resulting in FPKM < 1 were excluded, since they may refer to artifacts. After exclusion, a total of 26,001 transcripts remained, representing 15,538 distinct components (**Table 1, Figure 1**).

Table 1 - Summary of the *T. saginata* metacestode transcriptome assembly

Total reads	2 x 23,549,932
Total post-processed reads	18,806,272
Total Trinity 'genes'	16,641
Total Trinity transcripts	34,028
Total FPKM filtered 'genes'	15,538
Total FPKM filtered transcripts	26,001
Percent GC	47.17
Contig N10	6,41
Contig N20	4,904
Contig N30	3,991
Contig N40	3,329
Contig N50	2,814
Median contig length	1,706
Average contig length	2192.38
Total assembled bases	74,602,211

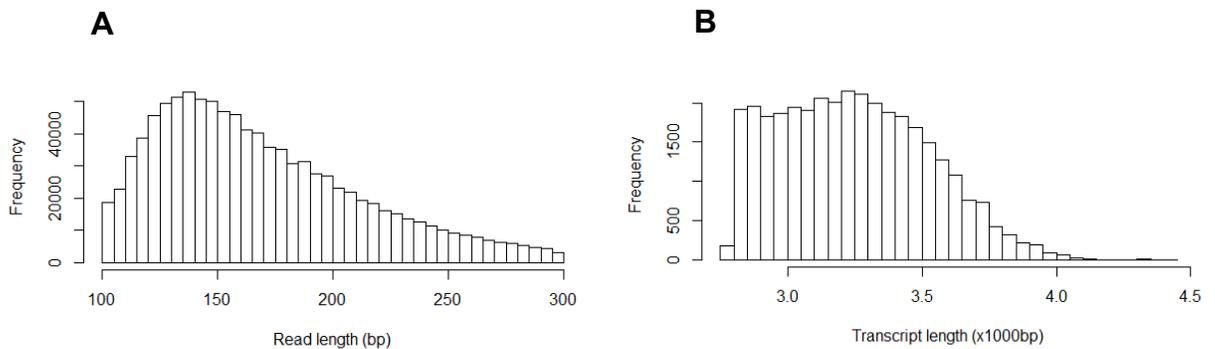


FIGURE 1 - Length distribution of the processed reads (A) and assembled Trinity transcripts (B).

All predicted transcripts and open reading frames (ORFs) were aligned against previously reported protein sequences from four helminth species belonging to the Taeniidae family (*T. solium*, *T. asiatica*, *E. granulosus*, *E. multiloculares*), one additional cestode (*H. microstoma*), one nematode (*C. elegans*) and one trematode (*S. mansoni*). A total of 43,058 hits were found in the BLASTx analysis, from which 44% and 36% were alignments against *T. solium* and *T. asiatica*, respectively (**Figure 2**). Performing BLASTx using only *Taenia* species dataset, a total of 13,310 hits were common to the three species, 2,631 hits were common between *T. saginata* and *T. asiatica*, and 3,940 between *T. saginata* and *T. solium* (**Figure 3**).

The BLASTp analysis reported a total of 25,119 hits, from which 43% of the alignments were against *T. solium* and 41% were against *T. asiatica*. Besides, there were 6,080 proteins matching to a Trinity transcript by > 90% and ≤ 100% of their protein length, and 6,844 proteins represented by nearly full-length transcripts, having > 80% alignment coverage (**Table S2**).

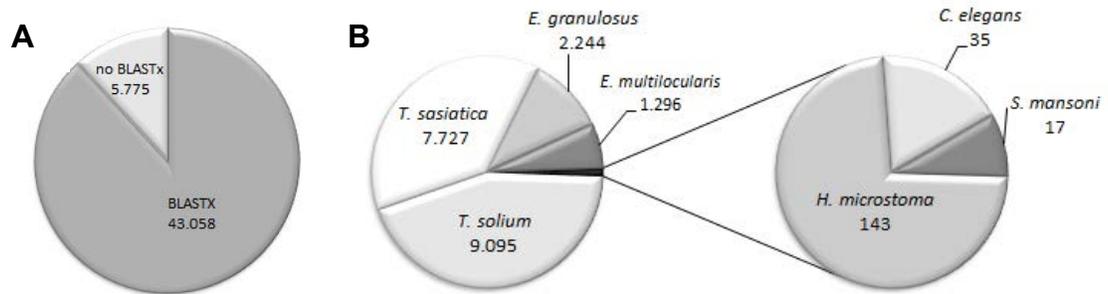


FIGURE 2 - Annotation of *T. saginata* metacestode transcripts using BLASTx. Pie chart illustrating the total annotations (A) and the percentage of best hits per species against the reference set (B).

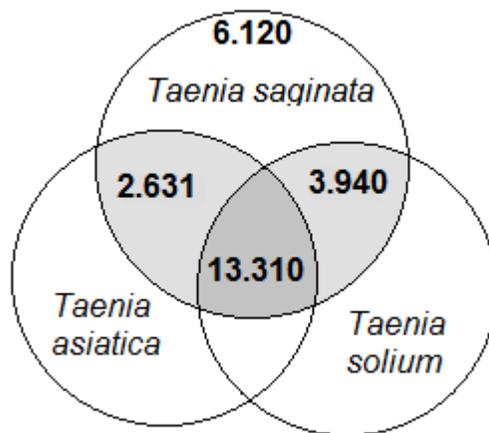


FIGURE 3 - Overlapped sequences among *T. saginata*, *T. solium* and *T. asiatica*.

Alignments using the COG database resulted in 4,687 annotated genes, which could be assigned to 24, out of a total of 26, functional COG categories (**Figure 4**). We also found 1,171 COG terms associated with these genes. The most represented functional category was R (general function prediction only), with 1,370 genes (25.11%). The classes K (transcription), L (replication, recombination

and repair) and T (signal transduction mechanisms) accounted for 593 (10.87%), 534 (9.78%) and 524 (9.60%) of the identified genes, respectively.

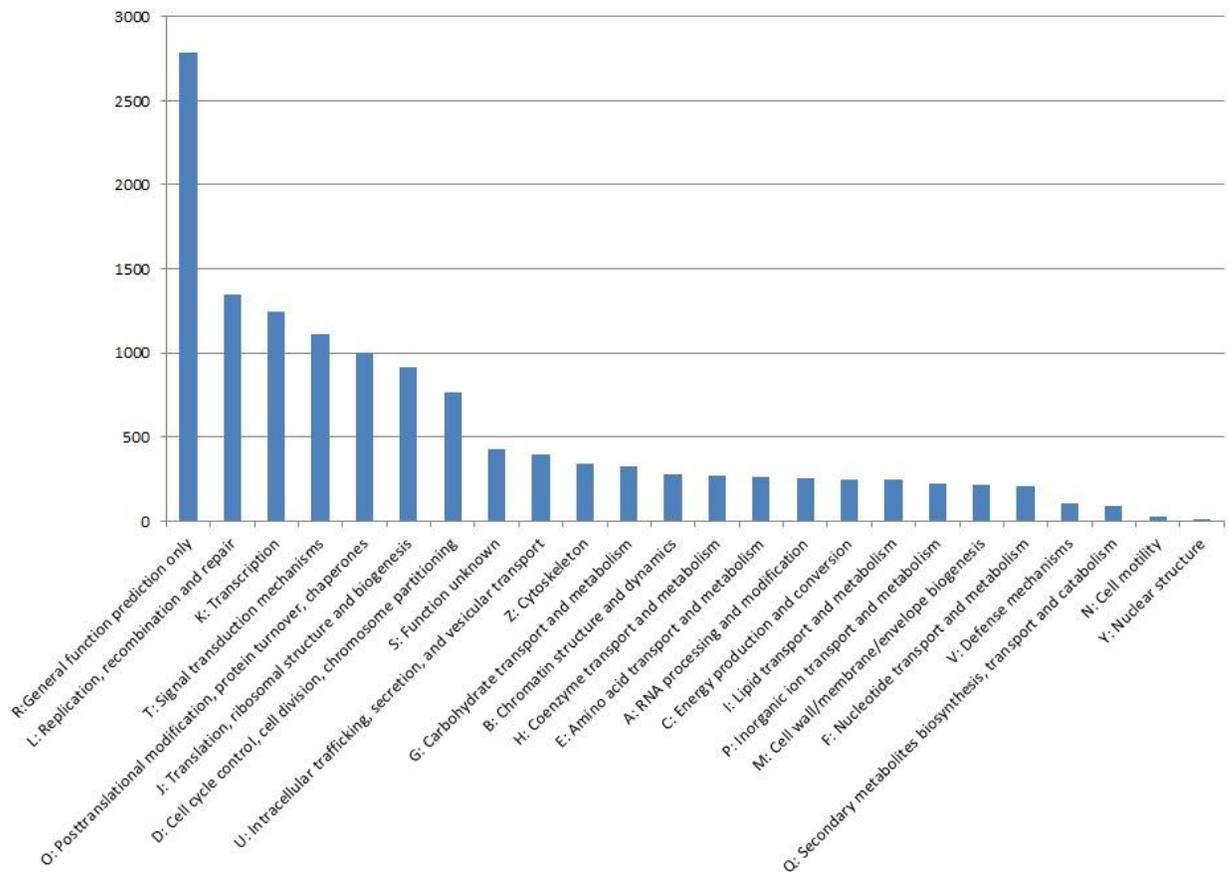


FIGURE 4 - COG functional annotation of putative proteins of *T. saginata* metacestode. A total of 4,687 genes classified into 25 molecular families.

Using Trinotate we were able to retrieve at least one Pfam annotation for 34% of the queried sequences (16,537). These included 3,317 distinct protein domains/families. Among the 50 most represented Pfam domains/families (**Table 2**), “Protein Kinase” (265 annotations), “Protein Tyrosine Kinase” (252

annotations), “WD40 domain, G-bet repeat” (179 annotations), “RNA recognition motif” (455 annotations) and “C2H2-type zinc finger” (115 annotations) were the most abundant.

Table 2 - The 50 most represented Pfam domains/families in *T. saginata* metacestode genes

Pfam entry	Domains/Families	No. of genes
PF00069.20	Protein kinase domain	265
PF07714.12	Protein tyrosine kinase	252
PF00400.27	WD40 domain, G-beta repeat	179
PF14259.1	RRM_6 RNA recognition motif	171
PF00076.17	RRM_1 RNA recognition motif	169
PF13893.1	RRM_5 RNA recognition motif	115
PF13894.1	zf-C2H2_4 C2H2-type zinc finger	113
PF00096.21	zf-C2H2 Zinc finger, C2H2 type	99
PF13895.1	Ig_2 Immunoglobulin domain	99
PF00271.26	Helicase_C Helicase conserved C-terminal domain	98
PF00023.25	Ank Ankyrin repeat	82
PF00071.17	RasRas family	80
PF08477.8	Miro Miro-like protein	79
PF13405.1	EF-hand_6 EF-hand domain	78
PF12796.2	Ank_2 Ankyrin repeats	77
PF13833.1	EF-hand_8 EF-hand domain	77
PF00270.24	DEAD DEAD/DEAH box helicase	75
PF13504.1	LRR_7 Leucine rich repeat	75
PF13202.1	EF-hand_5 EF hand	72
PF13637.1	Ank_4 Ankyrin repeats	71
PF13923.1	zf-C3HC4_2 Zinc finger, C3HC4 type	71
PF13912.1	zf-C2H2_6 C2H2-type zinc finger	70
PF00025.16	Arf ADP-ribosylation factor family	69
PF07719.12	TPR_2 Tetratricopeptide repeat	68
PF13465.1	zf-H2C2_2 Zinc-finger double domain	65
PF00595.19	PDZ PDZ domain	63

PF00515.23	TPR_1 Tetratricopeptide repeat	60
PF13499.1	EF-hand_7 EF-hand domain	60
PF07690.11	MFS_1 Major Facilitator Superfamily	59
PF13606.1	Ank_3 Ankyrin repeat	59
PF13927.1	Ig_3 Immunoglobulin domain	59
PF13857.1	Ank_5 Ankyrin repeats	58
PF07679.11	I-set Immunoglobulin I-set domain	57
PF00560.28	LRR_1 Leucine Rich Repeat	56
PF01926.18	MMR_HSR1 50S ribosome-binding GTPase	56
PF13181.1	TPR_8 Tetratricopeptide repeat	56
PF00001.16	7tm_1 7 transmembrane receptor	55
PF00097.20	zf-C3HC4 Zinc finger, C3HC4 type	54
PF13516.1	LRR_6 Leucine Rich repeat	53
PF00046.24	HomeoboxHomeobox domain	51
PF12874.2	zf-met Zinc-finger of C2H2 type	50
PF13414.1	TPR_11 TPR repeat	49
PF13855.1	LRR_8 Leucine rich repeat	49
PF00018.23	SH3_1 SH3 domain	48
PF13639.1	zf-RING_2 Ring finger domain	48
PF00168.25	C2 domain	45
PF07653.12	SH3_2 Variant SH3 domain	45
PF14531.1	Kinase-like	45
PF00004.24	ATPase family associated with various cellular activities	44
PF00028.12	Cadherin domain	44

The annotation of GO terms using Trinotate included 11,310 transcripts (26%) associated to 5,455 genes, which had 18, 15 and 10 terms under biological process, cellular component and molecular function, respectively (**Figure 5**). Among the child categories, the most predominant ones for biological process were “cellular process” (43%) and “metabolic process” (33%). In the case of cellular component, we found “cell” and “cell part” with 32% each. Finally, the

molecular function component had GO terms “binding” (59%) and “catalytic activity” (33%).

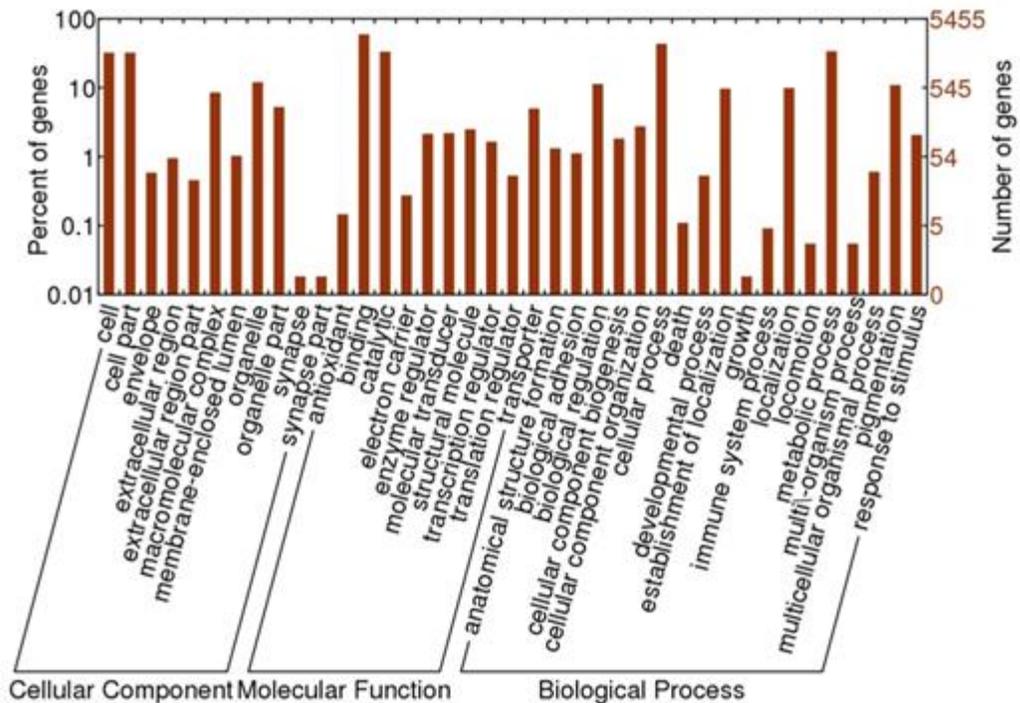


FIGURE 5 - Gene Ontology (GO) terms for *T. saginata* metacestode genes. Classification according to the categories ‘Cellular Component’, ‘Molecular Function’ and ‘Biological Process’ assigned to 5,455 genes from *T. saginata*.

We identified 5,455 genes with complete annotation resulting from the BLASTx, Pfam and GO term analysis. This subgroup of genes was analyzed to find active biological pathways using KEGG database. We have found 2,486 KEGG Orthologs (KO) terms related to *T. saginata* metacestode genes. These entries were grouped into 6 categories: “Metabolism” (485 genes), “Genetic Information Processing” (561 genes), “Environmental Information Processing” (329

genes), “Cellular Processes” (338 genes), “Organismal Systems” (357 genes) and “Human Diseases” (478 genes).

In order to identify highly represented genes, we pruned in genes according to their abundance and selected the top 10% of the FPKM distribution (FPKM > 81,22). The top expressed 1,481 genes mapped to them 5,014 transcripts, which were used for KEGG pathway reconstruction. A total of 1.469 entries were assigned to 871genes. The pathways that accounted for the higher number of transcripts were “Glycolysis/Gluconeogenesis” and “Oxydative phosphorylation” (Metabolism); “Ribosome” and “Protein processing in endoplasmic reticulum” (Genetic Information Processing); “MAPK signaling” and “PI3K-AKT signaling” (Environmental Information Processing); “Endocytosis” and “Phagosome” (Cellular Processes); “Huntington’s disease” and “Parkinson’s disease” (Human Diseases); and “Insulin signaling” and “Glucagon signaling” (Organismal Systems).

Host insulin seems to be an important factor to the development of parasites, as shown by an *in vitro Echinococcus multilocularis* cultivating [28]. The addition of exogenous insulin increased the stem cell differentiation into new metacestodes, stimulated the uptake of glucose and affected the phosphorylation profile of the components of the PI3K-AKT signaling pathway. In the present work, among the highly represented *T. saginata* genes set, 25 are assigned to 16 KEGG insulin signaling pathway entries (**Table 3**).

In addition, from the total 7.313 common genes (13,310 transcripts) among *T. saginata*, *T. asiatica* and *T. solium*, according to BLASTx alignment, a total of 3,466 KEGG entries were assigned (**Figure S3**). From these, 77 genes are related to 35 insulin pathway entries (**Table 3**).

Table 3 - The Insulin signaling pathway common genes among *T. saginata*, *T. asiatica* and *T. solium*. *Genes highly represented for *T. saginata* (90th percentile).

Gene name	KEGG entry	Definition
*PYG	K00688	starch phosphorylase [EC:2.4.1.1]
GYS	K00693	glycogen(starch) synthase [EC:2.4.1.11]
*HK	K00844	hexokinase [EC:2.7.1.1]
PHKG	K00871	phosphorylase kinase gamma subunit [EC:2.7.11.19]
*PEPCK	K01596	phosphoenolpyruvate carboxykinase (GTP) [EC:4.1.1.32]
*CALM	K02183	calmodulin
HRAS	K02833	GTPase HRas
*RPS6	K02991	small subunit ribosomal protein S6e
GSK3B	K03083	glycogen synthase kinase 3 beta [EC:2.7.11.26]
*EIF4E	K03259	translation initiation factor 4E
*FBP	K03841	fructose-1,6-bisphosphatase I [EC:3.1.3.11]
*PKA	K04345	protein kinase A [EC:2.7.11.11]
*GRB2	K04364	growth factor receptor-binding protein 2
BRAF	K04365	B-Raf proto-oncogene serine/threonine-protein kinase [EC:2.7.11.1]
MAP2K1	K04368	mitogen-activated protein kinase kinase 1 [EC:2.7.12.2]
*MAPK1_3	K04371	mitogen-activated protein kinase 1/3 [EC:2.7.11.24]
JNK	K04440	c-Jun N-terminal kinase [EC:2.7.11.24]
INSR	K04527	insulin receptor [EC:2.7.10.1]
RPS6KB	K04688	p70 ribosomal S6 kinase [EC:2.7.11.1]
*CBL	K04707	E3 ubiquitin-protein ligase CBL [EC:2.3.2.27]
*PRKAR	K04739	cAMP-dependent protein kinase regulator
LAR	K05695	receptor-type tyrosine-protein phosphatase F [EC:3.1.3.48]
PTPN1	K05696	tyrosine-protein phosphatase non-receptor type 1 [EC:3.1.3.48]
PRKCI	K06069	atypical protein kinase C iota type [EC:2.7.11.13]
*PPP1C	K06269	serine/threonine-protein phosphatase PP1 catalytic subunit [EC:3.1.3.16]
PDPK1	K06276	3-phosphoinositide dependent protein kinase-1 [EC:2.7.11.1]
SHC1	K06279	SHC- transforming protein 1
HSL	K07188	hormone-sensitive lipase [EC:3.1.1.79]
*PHKA_B	K07190	phosphorylase kinase alpha/beta subunit
FLOT	K07192	flotillin
*AMPK	K07198	5'-AMP-activated protein kinase, catalytic alpha subunit [EC:2.7.11.11]
*PRKAB	K07199	5'-AMP-activated protein kinase, regulatory beta subunit
PRKAG	K07200	5'-AMP-activated protein kinase, regulatory gamma subunit
MTOR	K07203	serine/threonine-protein kinase mTOR [EC:2.7.11.1]
ACACA	K11262	acetyl-CoA carboxylase / biotin carboxylase 1 [EC:6.4.1.2 6.3.4.14]

Another interesting pathway, represented here by 35 common genes among *Taenia* species, is the Transforming growth factor-beta (TGF- β). This pathway belongs to a large family of proteins, including activins/inhibins and bone morphogenic proteins (BMP), besides controlling a wide range of events related to growth, development and immune response [29] (**Table 4**).

Table 4 - The TGF- β signaling pathway common genes among *T. saginata*, *T. asiatica* and *T. solium*. *Genes highly represented for *T. saginata* (90th percentile).

Gene name	KEGG entry	Definition
*SKP1	K03094	S-phase kinase-associated protein 1
PPP2R1	K03456	serine/threonine-protein phosphatase 2A regulatory subunit A
*RBX1	K03868	RING-box protein 1
*MAPK1_3	K04371	mitogen-activated protein kinase 1/3 [EC:2.7.11.24] serine/threonine-protein phosphatase 2A catalytic subunit [EC:3.1.3.16]
*PPP2C	K04382	E1A/CREB-binding protein [EC:2.3.1.48]
EP300	K04498	E1A/CREB-binding protein [EC:2.3.1.48]
*SMAD2_3	K04500	mothers against decapentaplegic homolog 2/3
SMAD4	K04501	mothers against decapentaplegic homolog 4
*RHOA	K04513	Ras homolog gene family, member A
ROCK1	K04514	Rho-associated protein kinase 1 [EC:2.7.11.1]
NOG	K04658	noggin
FST	K04661	follicle-stimulating hormone receptor 1
ACVR2A	K04670	activin receptor type-2A [EC:2.7.11.30]
TGFBR1	K04674	TGF-beta receptor type-1 [EC:2.7.11.30]
ACVR1	K04675	activin receptor type-1 [EC:2.7.11.30]
MADHIP	K04679	MAD, mothers against decapentaplegic interacting protein
RPS6KB	K04688	p70 ribosomal S6 kinase [EC:2.7.11.1]

Discussion

We've assembled a *de novo* whole transcriptome of *T. saginata* metacestode generated from 40,101,678 paired-end reads.

The assembly resulted in 16,641 unigenes branching into 34,028 isoforms. Interestingly, the number of putative genes was higher than expected, considering that deep sequencing of the transcriptomes of multiple developmental stages of four other tapeworm parasite species revealed expression of only 10,000 to 12,000 genes [18]. Even after excluding lowly expressed transcripts (FPKM < 1), which could have arisen from assembly errors, the number of putative genes and associated isoforms was still high, comprising 26,001 isoforms from 15,538 unigenes. We've observed that 75% of the transcripts were low or moderately abundant and less than 1% were highly abundant transcripts.

Although we did not formally perform polymorphism or alternative splicing analysis, the number of isoforms was found to be almost twice the number of unigenes, which predicts substantial exonic nucleotide variation and post-transcriptional rearrangements for this particular sample. This higher than the expected number of isoforms could, in part, be explained by the introduction of systematic errors during the *de novo* assembly process, as the lack of guidance from a high-quality reference genome may lead to assembly artifacts [30]. However, our BLASTx and BLASTp analyses could successfully identify protein similarities for 79% of the assembled transcripts. From the total BLAST hits, 7,480 proteins matched to a transcript by 70% to 100% of their length. Regarding to the full-length analysis, 6,844 proteins had at least 90% alignment coverage (**Table S2**). Remarkably, over 40% of the BLAST hits showed high identity with protein sequences from other Taeniid parasites. These findings support the reliability of our RNA sequencing experiment and *de novo* assembly approach.

A total of 5,775 transcripts presented no BLAST hits. Although no functional information could be retrieved for these transcripts, they are still an important

fraction of the metacestode transcriptome, as they may represent poorly curated gene sequences, novel genes, untranslated regions, non-coding RNAs, or species/stage-specific transcripts. Further investigation is necessary to characterize and validate these unidentified transcripts, and distinguish them from assembly artifacts. Nevertheless, these transcripts are potential species and/or stage-specific markers.

The comprehensive functional annotation carried out in the present study showed that the assembled transcripts participate in a wide range of biological processes. The diversity of annotated gene ontology terms and protein domains makes it difficult to highlight a single fundamental pathway that could be distinctive of the metacestode state, and highlights the complexity of this overlooked developmental stage.

Although the parasite seems to be focused on survival at this stage, the separate analysis performed using KEGG database for the most abundant transcripts (FPKM > 81.22) and for the common genes among three *Taenia* species allowed us to observe genes composing pathways that suggest important interaction between the parasite and its host, such as insulin signaling and TGF- β .

Insulin receptors and receptor-related proteins are conserved through animal kingdom, regulating development and metabolism following activation by insulin-related peptides [29]. This interaction was demonstrated, for instance, by adding human insulin in *Echinococcus multilocularis* metacestode culture, thus activating the insulin signaling pathway. On the other hand, the addition of insulin receptor inhibitor led to impaired development and parasite killing [28].

The TGF- β family includes activin, bone morphogenetic proteins (BMPs), and is involved in processes like cell proliferation, migration, differentiation, adhesion and apoptosis, and playing a role in metazoan development. The downstream signaling protein of receptors that respond to TGF- β and activin is Smad2, whereas the receptors that respond to BMPs signal via Smad1 [31], [32].

In vitro experiments have demonstrated the activation of TGF- β pathway in some parasites, such as *E. multilocularis* ALK1/ALK2-like members of the TGF- β receptor family, through binding to the human bone morphogenetic proteins (BMP) [33]. Another example is the *Schistosoma mansoni* receptor kinase 1 (SmRK1) that is capable of signaling via Smad2 in response to human TGF- β [34].

We also observed enrichment for the GO terms “interspecies interaction between organisms” and “symbiosis, encompassing mutualism through parasitism”. Despite the fact that the majority is composed by ribosome, actin and tubulin proteins, these interactions must be analyzed in details, since it is known that the parasite may interact to the host (i.e., excretion-secretion molecules; transmembrane proteins), thus being candidates for better understanding host-parasite interaction and ultimately providing insights for the development of diagnosis and vaccines.

Conclusions

We reported the successfully whole transcriptome assembly of *T. saginata* metacestode. Although the transcription machinery of the parasite was found to be devoted to the maintenance of processes essential to its survival, some are important via for host-parasite interactions. This information will be useful as a new window for *T. saginata* biology and pathogenicity, as well as a resource to improve gene annotation of reference genomes of related parasites and comparative transcriptome analyses.

Acknowledgments

To CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) for the scholarships. To Dr. Luiz Lehmann Coutinho, Laboratory of Animal Biochemistry – São Paulo University, for libraries construction. To Deoxi Biotecnologia Ltda for data sequencing.

References

1. MA 1. Gemmell, Z M, Z P, E J L S, C L, G S N. Guidelines for surveillance prevention and control of taeniasis/cysticercosis. 1983.
2. FAO. Summary Risk Profile on *C. bovis* in Meat From Domestic Cattle [Internet]. 2013. Available: http://www.fao.org/fileadmin/user_upload/agns/pdf/Foodborne_parasites/RiskProfTaeniasaginataOct2013.pdf
3. OIE. Cysticercosis [Internet]. 2014. Available: http://www.oie.int/fileadmin/Home/eng/Health_standards/tahm/2.09.05_CYSTICERCOSIS.pdf
4. M.A.V. da C. Pereira , V.S. Schwanz CGB. Prevalência Da Cisticercose Em Carcaças De Bovinos Abatidos Em Matadouros-Frigoríficos Do Estado Do Rio De Janeiro , Submetidos Ao Controle Do Serviço De Inspeção Federal. Arq Inst Biol. 2006; 83–87.

5. Ogunremi O, Benjamin J. Development and field evaluation of a new serological test for *Taenia saginata* cysticercosis. *Vet Parasitol. Elsevier B.V.*; 2010;169: 93–101.
6. Ferrer E, Bonay P, Foster-Cuevas M, González LM, Dávila I, Cortéz MM, et al. Molecular cloning and characterisation of Ts8B1, Ts8B2 and Ts8B3, three new members of the *Taenia solium* metacestode 8 kDa diagnostic antigen family. *Mol Biochem Parasitol.* 2007;152: 90–100.
7. Harrison LJS, Garate T, Bryce DM, Gonzalez LM, Foster-Cuevas M, Wamae LW, et al. Ag-ELISA and PCR for monitoring the vaccination of cattle against *Taenia saginata* cysticercosis using an oncospherical adhesion protein (HP6) with surface and secreted localization. *Trop Anim Health Prod.* 2005;37: 103–20. Available: <http://www.ncbi.nlm.nih.gov/pubmed/15742866>
8. Lightowlers MW, Rolfe R, Gauci CG. *Taenia saginata*: vaccination against cysticercosis in cattle with recombinant oncosphere antigens. *Exp Parasitol.* 1996;84: 330–8.
9. Flisser A, Gauci CG, Zoli A, Garza-rodriquez A, Luis J, Maravilla P, et al. Induction of Protection against Porcine Cysticercosis by Vaccination with Recombinant Oncosphere Antigens Induction of Protection against Porcine Cysticercosis by Vaccination with Recombinant Oncosphere Antigens. 2004;72: 5292–5297.
10. Gauci C, Vural G, Öncel T, Varcasia A, Damian V. Vaccination with recombinant oncosphere antigens reduces the susceptibility of sheep to infection with *Taenia multiceps*. *Int J Parasitol.* 2008;38: 1041–1050.

11. Bolger a. M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30: 2114–2120.
12. Schmieder R, Edwards R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics*. 2011;27: 863–864.
13. Zhang J, Kobert K, Flouri T, Stamatakis a. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*. 2014;30: 614–620.
14. Manfred G. Grabherr, Brian J. Haas, Moran Yassour, Joshua Z. Levin, Dawn A. Thompson, Ido Amit, Xian Adiconis, Lin Fan, Raktima Raychowdhury, Qiandong Zeng, Zehua Chen, Evan Mauceli, Nir Hacohen, Andreas Gnirke, Nicholas Rhind, Federica di Palma, Bruce W. N, Friedman and AR. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat Biotechnol*. 2013;29: 644–652.
15. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Philip D, Bowden J, et al. reference generation and analysis with Trinity. *Nat Protocols*. 2014.
16. Bo L, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. 2011; 12:2-16.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool 2Department of Computer Science. *J Mol Biol*. 1990;215: 403–410.

18. Tsai IJ, Zarowiecki M, Holroyd N, Garciarrubio A, Sanchez-Flores A, Brooks KL, et al. The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*. 2013;496: 57–63.
19. The *C. elegans* sequencing consortium. Genome sequencing of the nematode *C. elegans*: A platform for investigating biology. *Science*. 1998; 282:2012-2018.
20. Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature*. 2009; 460:352-361.
21. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, et al. The PFAM protein families database. *Nucleic Acids Research*. 2002; 30:276-280.
22. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*. 2011; 1–9.
23. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Research*. 2007; 35:3100-3108.
24. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*. 2011; 8:785-786.

25. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol.* 2001; 305:567-580.
26. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV. The COG database: an updated version includes eukaryotes. 2003; 4:1-14.
27. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. 2000; 28:27-30.
28. Hemer S, Konrad C, Spiliotis M, Koziol U, Schaack D, Förster S, et al. Host insulin stimulates *Echinococcus multilocularis* insulin signalling pathways and larval development. *BMC Biology.* 2014; 12:1-22.
29. Dissous C, Khayath N, Vicogne J, Capron M. growth factor receptors in heminth parasites: Signalling and host-parasite relationships. *FEBS Letter.* 2006; 580:2968-2975.
30. Martin J a, Wang Z. Next-generation transcriptome assembly. *Nat Rev Genet.* Nature Publishing Group; 2011;12: 671–682.
31. Zavala-Góngora R, Kroner A, Bernthaler P, Knaus P, Brehm K. A member of the transforming growth factor- β receptor family from *Echinococcus multilocularis* is activated by human bone morphogenetic protein 2. *Molecular & Biochemical Parasitology.* 2006; 146:265-271.
32. Beall MJ, Pearce EJ. Human transforming growth factor- β activates a receptor serine/threonine kinase from the intravascular parasite *Schistosoma mansoni*. 2001; 276:31613-31619.

Supporting Information

FIGURE S1 - Electropherogram of *Taenia saginata* metacestode insert size under Bioanalyzer.

FIGURE S2 - Paired-end sequencing quality across bases and over all raw reads (A – forward reads; B – reverse reads) and post processed reads (C).

FIGURE S3 - KEGG annotation of the common genes among *T. saginata*, *T. asiatica* and *T. solium*.

Table S1 - KEGG orthology annotation of the *Taenia saginata* genes

Table S2 - Analysis of full-length *Taenia saginata* metacestode transcripts using BLASTp

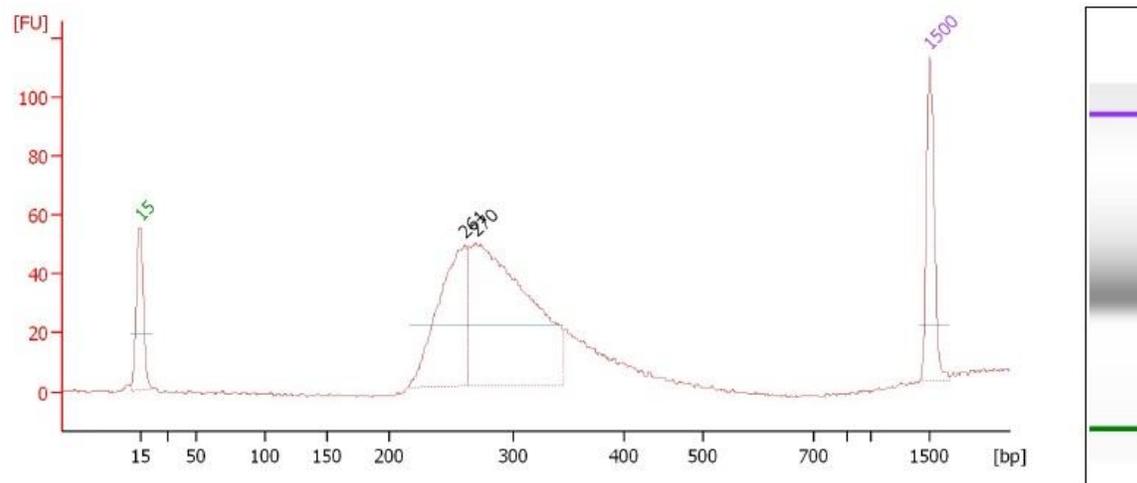


Figure S1. Electropherogram of *Taenia saginata* metacestode insert size under Bioanalyzer.

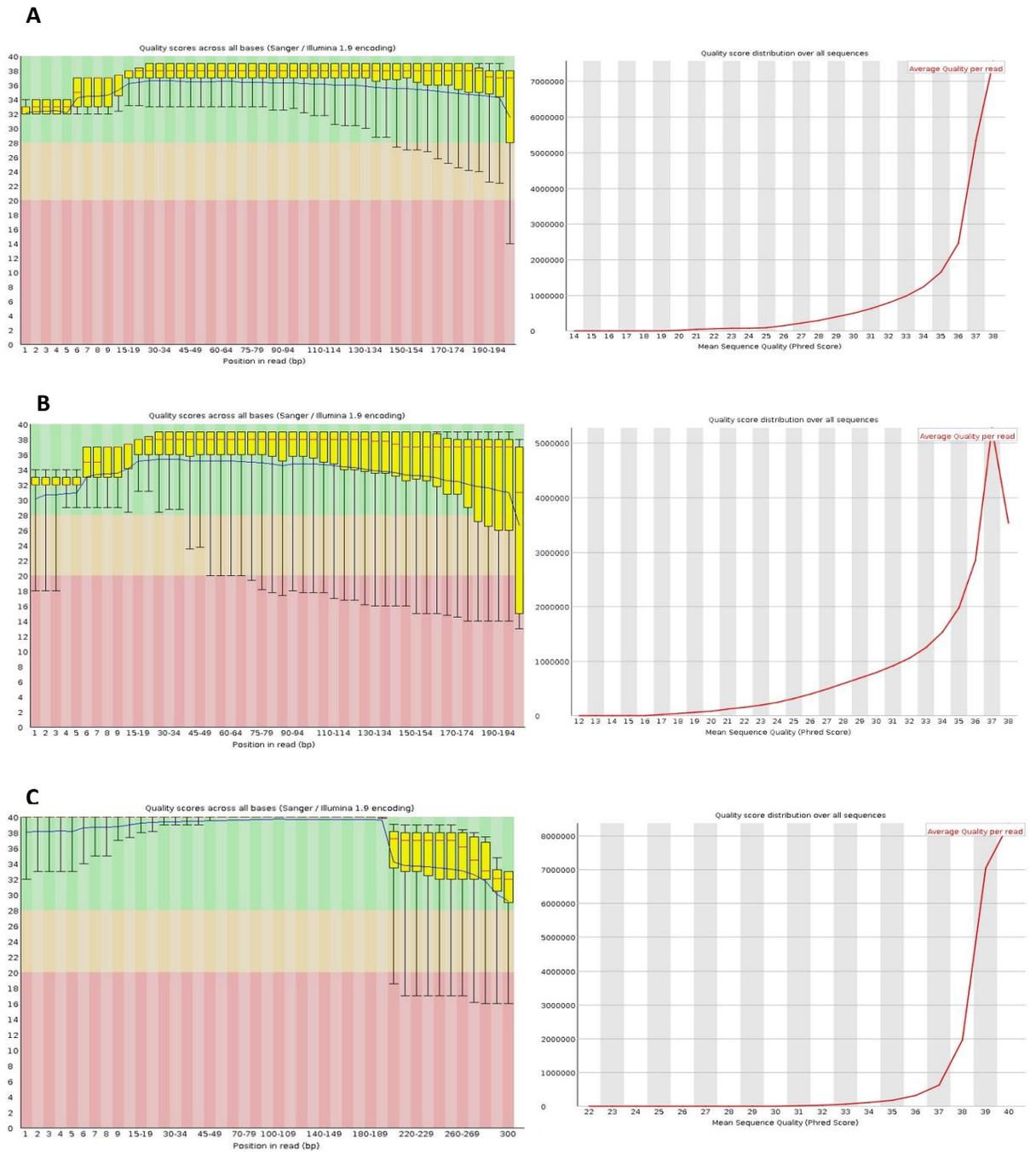


Figure S2. Paired-end sequencing quality across bases and over all raw reads (A – forward reads; B – reverse reads) and post processed reads (C).

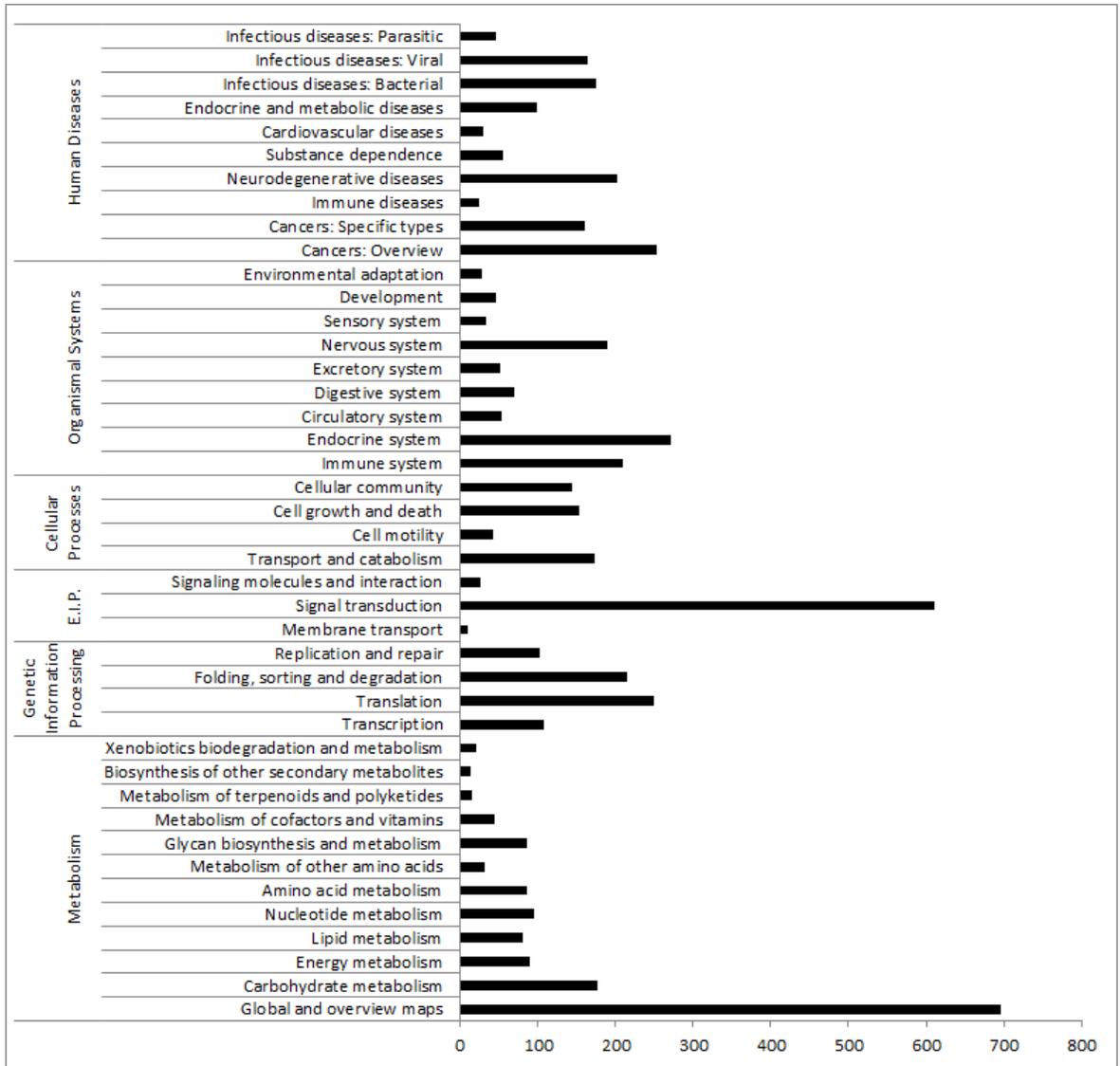


Figure S3. KEGG annotation of the common genes among *T. saginata*, *T. asiatica* and *T. solium*.

Table S1. KEGG orthology annotation of the *Taenia saginata* genes

KEGG pathway categories	Genes	Most abundant KEGG entries
Metabolism		
Carbohydrate Metabolism	22	ko00020 Citrate cycle (TCA cycle) and ko00520 Amino sugar and nucleotide sugar metabolism
Energy Metabolism	44	ko00190 Oxidative phosphorylation
Lipid Metabolism	26	ko00564 Glycerophospholipid metabolism
Nucleotide Metabolism	54	ko00250 Alanine, aspartate and glutamate metabolism
Amino Acid Metabolism	15	ko00310 Lysine degradation
Metabolism of Other Amino Acids	15	ko00480 Glutathione metabolism
Glycan Biosynthesis and Metabolism	17	ko00510 N-Glycan biosynthesis
Metabolism of Cofactors and Vitamins	9	ko00860 Porphyrin and chlorophyll metabolism
Metabolism of Terpenoids and Polyketides	12	ko00900 Terpenoid backbone biosynthesis
Biosynthesis of Other Secondary Metabolites	3	ko00950 Isoquinoline alkaloid biosynthesis
Xenobiotics Biodegradation and Metabolism	10	ko00983 Drug metabolism - other enzymes
Genetic information processing (GIP)		
Transcription	66	ko03040 Spliceosome
Translation	60	ko03013 RNA transport
Folding, Sorting and Degradation	55	ko04141 Protein processing in endoplasmic reticulum
Replication and Repair	22	ko03030 DNA replication
Environmental information processing (EIP)		
Membrane Transport	8	ko02010 ABC transporters
Signal Transduction	44	ko04151 PI3K-Akt signaling pathway
Signaling Molecules and Interaction	18	ko04080 Neuroactive ligand-receptor interaction
Cellular processes		
Transport and Catabolism	75	ko04144 Endocytosis
Cell Motility	43	ko04810 Regulation of actin cytoskeleton
Cell Growth and Death	40	ko04110 Cell cycle

Cell Communication	46	ko04510 Focal adhesion
Organismal systems		
Immune System	27	ko04062 Chemokine signaling pathway
Endocrine System	39	ko04910 Insulin signaling pathway
Circulatory System	23	ko04261 Adrenergic signaling in cardiomyocytes
Digestive System	15	ko04972 Pancreatic secretion
Excretory System	14	ko04962 Vasopressin-regulated water reabsorption and ko04961 Endocrine and other factor-regulated calcium reabsorption
Nervous System	30	ko04722 Neurotrophin signaling pathway
Sensory System	13	ko04722 Neurotrophin signaling pathwa
Development	23	ko04360 Axon guidance
Environmental Adaptation	13	ko04713 Circadian entrainment
Human diseases		
Cancers	69	ko05200 Pathways in cancer
Immune System Diseases	10	ko05323 Rheumatoid arthritis
Neurodegenerative Diseases	57	ko05016 Huntington's disease
Cardiovascular Diseases	22	ko05034 Alcoholism
Metabolic Diseases	9	ko05416 Viral myocarditis
Infectious Diseases	54	ko05166 HTLV-I infection
Endocrine and metabolic diseases	37	ko04932 Non-alcoholic fatty liver disease (NAFLD)

Table S2. Analysis of full-length *Taenia saginata* metacestode transcripts using BLASTp

#hit_pct_cov_bin	count_in_bin	>bin_below
100	6080	6080
90	764	6844
80	606	7450
70	601	8051
60	630	8681
50	569	9250
40	533	9783
30	528	10311
20	774	11085
10	1216	12301