

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

**Análise em larga escala de transcritos processados por *Spliced leader trans-splicing*
em esporocistos de *Schistosoma mansoni***

por

Núbia Monteiro Gonçalves Soares Fernandes

Belo Horizonte

Fevereiro/2015

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

**Análise em larga escala de transcritos processados por *Spliced leader trans-splicing*
em esporocistos de *Schistosoma mansoni***

por

Núbia Monteiro Gonçalves Soares Fernandes

**Dissertação apresentada com vistas à obtenção
do Título de Mestre em Ciências na área de
concentração Biologia Celular e Molecular.**

Orientação: Dra. Rosiane A. da Silva Pereira

**Co-orientação: Dra. Marina de Moraes
Mourão**

Belo Horizonte

Fevereiro/2015

Catálogo-na-fonte
Rede de Bibliotecas da FIOCRUZ
Biblioteca do CPqRR
Segemar Oliveira Magalhães CRB/6 1975

F363a Fernandes, Núbia Monteiro Gonçalves Soares.
2015

Análise em larga escala de transcritos processados por *Spliced leader trans-splicing* em esporocistos de *Schistosoma mansoni* // Núbia Monteiro Gonçalves Soares Fernandes. – Belo Horizonte, 2015.

XVI, 65 f: il.: 210 x 297 mm.

Bibliografia: 75-81

Dissertação (mestrado) – Tese para obtenção do título de Mestre em Ciências pelo Programa de Pós-Graduação em Ciências da Saúde do Centro de Pesquisas René Rachou. Área de concentração: Biologia Celular e Molecular.

1. Esquistossomose/genética 2. *Schistosoma mansoni*/genética 3. Análise de Sequência de RNA/métodos
I. Título. II. Pereira, Rosiane Aparecida da Silva (Orientação). III. Mourão, Marina de Moraes (Co-orientação)

CDD – 22. ed. – 616.963

Ministério da Saúde
Fundação Oswaldo Cruz
Centro de Pesquisas René Rachou
Programa de Pós-graduação em Ciências da Saúde

**Análise em larga escala de transcritos processados por *Spliced leader trans-splicing*
em esporocistos de *Schistosoma mansoni***

por

Núbia Monteiro Gonçalves Soares Fernandes

Foi avaliada pela banca examinadora composta pelos seguintes membros:

Presidente: Prof. Dra. Rosiane Aparecida da Silva Pereira

Prof. Dra. Marina de Moraes Mourão

Prof. Dr. Elio Hideo Babá

Prof. Dra. Debora de Oliveira Lopes

Suplente: Prof. Dr. Carlos Eduardo Calzavara Silva

Dissertação defendida e aprovada em: 23/02/2015.

*Ouçã os conselhos e esteja
pronto para aprender, e assim
um dia vocẽ serã sãbio.
(Provãrbios 19:20)*

Dedicatórias

À minha querida família!

Agradecimentos

Primeiramente agradeço à Deus que me permitiu iniciar, caminhar e concluir mais esta etapa em minha vida. Às pessoas que mais amo: meus pais, irmãos e ao Fred, agradeço-lhes todo o apoio e carinho.

À minha orientadora Dra. Rosiane A. da Silva Pereira pela confiança, amizade, paciência e, sobretudo, pelo conhecimento que tanto contribuiu para a conclusão desta dissertação.

À minha coorientadora Dr. Marina de Moraes Mourão agradeço-lhe imensamente por esses quatro anos de orientação. Obrigada pela paciência, dedicação, por me ensinar “mil” coisas, transmitir seus valiosos conhecimentos e por acreditar em mim.

Dra. Ângela Volpini e ao Dr. Guilherme Corrêa Oliveira agradeço a oportunidade de iniciar a minha carreira científica, pelo apoio e incentivo.

À Dra. Mariana Boroni que tanto me ajudou nas análises de bioinformática e trocas de experiências.

Aos meus amigos: Sandra, obrigada por tudo, pelo apoio, ensinamentos e pela enorme paciência. Paola e Fernanda pelo incentivo e momentos agradáveis que passamos no laboratório. Erick e Naiara agradeço-lhes pela amizade e apoio.

À Maíra pela atenção e ajudas.

À Dra. Liana Konovaloff Jannotti Passos, Dr. Omar dos Santos Carvalho e as colegas do moluscário pela concessão do material biológico.

Ao Flávio e Ana pelo apoio técnico, pelo sequenciamento e pela amizade.

À Kênia, que tanto contribuiu na limpeza dos materiais do laboratório e organização dos materiais.

Ao Centro de Pesquisas René Rachou que proporcionou a execução deste projeto.

À Pós-graduação do Centro de Pesquisas René Rachou pelo apoio durante este período.

Aos membros da banca examinadora por aceitarem prontamente o nosso convite.

Aos amigos do Laboratório de Parasitologia Celular e Molecular pelos momentos agradáveis nesta caminhada.

Ao CNPq pela bolsa concedida.

Sumário

| | |
|---|------|
| Lista de figuras | XII |
| Lista de tabelas | XIII |
| Lista de abreviaturas e símbolos | XIII |
| Resumo | XV |
| Abstract | XVI |
| 1 INTRODUÇÃO | 17 |
| 1.1 A Esquistossomose | 17 |
| 1.1.2 Aspectos Gerais | 17 |
| 1.2 Organismo de estudo - <i>Schistosoma mansoni</i> e seu ciclo de vida | 18 |
| 1.3 O genoma do <i>Schistosoma mansoni</i> | 20 |
| 1.4 Mecanismos de processamento de RNAs | 21 |
| 1.5 <i>Trans-splicing</i> | 23 |
| 1.6 Possíveis funções do mecanismo de <i>Spliced leader trans-splicing</i> | 26 |
| 1.7 Sequenciamento de transcritos por RNA-Seq | 27 |
| 2 JUSTIFICATIVA | 29 |
| 3 OBJETIVOS | 31 |
| 3.1 Objetivo Geral | 31 |
| 3.2 Objetivos Específicos | 31 |
| 4 MATERIAL E MÉTODOS | 32 |
| 4.1 Obtenção de material biológico | 32 |
| 4.1.1 Transformação <i>in vitro</i> de miracídios em esporocistos primários | 32 |
| 4.2 Construção da biblioteca de cDNA enriquecida em transcritos processados por SL <i>trans-splicing</i> | 33 |
| 4.2.1 Extração de RNA total e tratamento com DNase | 33 |
| 4.2.2 Purificação dos mRNAs e síntese de cDNAs | 33 |
| 4.2.3 Amplificação de transcritos processados por SL <i>trans-splicing</i> | 35 |
| 4.3 Sequenciamento - Plataforma Ion Torrent PGM™ System | 36 |
| 4.4 Análises de bioinformática - Processamento das sequências geradas | 37 |
| 4.4.1 Identificação da sequência do <i>Spliced leader</i> | 37 |
| 4.4.2 Tratamento de qualidade das sequências | 37 |
| 4.4.3 Mapeamento e contagem das sequências | 38 |
| 4.4.4 Anotação e análise funcional das sequências | 39 |
| 4.4.5 Identificação das vias metabólicas em que atuam as proteínas codificadas pelos transcritos processados por <i>Spliced leader trans-splicing</i> | 42 |
| 4.4.6 Análise de abundância dos transcritos anotados | 42 |

| | | |
|-----|---|----|
| 5 | RESULTADOS | 43 |
| 5.1 | Análise do RNA total extraído | 43 |
| 5.2 | Amplificação dos transcritos processados por <i>Spliced leader trans-splicing</i> | 44 |
| 5.3 | Obtenção das sequências na Plataforma Ion Torrent PGM™ System | 44 |
| 5.4 | Identificação e remoção da sequência <i>leader</i> e tratamento de qualidade das <i>reads</i> geradas | 45 |
| 5.5 | Mapeamento das <i>reads</i> no genoma de referência | 52 |
| 5.6 | Anotação e análise funcional dos transcritos processados por <i>Spliced leader trans-splicing</i> | 53 |
| 5.7 | Identificação das vias metabólicas onde atuam enzimas codificadas pelos transcritos processados por <i>SL trans-splicing</i> na fase de esporocisto | 56 |
| 5.8 | Análise de abundância dos transcritos anotados | 59 |
| 6 | DISCUSSÃO | 61 |
| 7 | CONCLUSÕES | 73 |
| 8 | PERSPECTIVAS | 74 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 75 |

Lista de figuras

| | |
|--|----|
| Figura 1: O ciclo de vida do parasito <i>Schistosoma mansoni</i> | 20 |
| Figura 2: Montagem do spliceossomo | 22 |
| Figura 3: Diferentes tipos de <i>splicings</i> | 23 |
| Figura 4: Esquema da captura dos mRNAs e amplificação dos transcritos processados por SL <i>trans-splicing</i> para a construção da biblioteca de cDNA. | 35 |
| Figura 5: Perfil eletroforético do RNA total extraído da fase de esporocisto do parasito <i>Schistosoma mansoni</i> | 43 |
| Figura 6: Amplificação dos transcritos processados por SL <i>trans-splicing</i> | 44 |
| Figura 7: Valor de qualidade por base | 46 |
| Figura 8: Valor de qualidade média por sequência. | 47 |
| Figura 9: Conteúdo de bases nitrogenadas ao longo da sequência | 48 |
| Figura 10: Conteúdo de GC por sequência | 49 |
| Figura 11: Conteúdo de GC por base | 50 |
| Figura 12: Sequências duplicadas | 51 |
| Figura 13: Percentual de genes processados por SL <i>trans-splicing</i> na fase de esporocisto | 53 |
| Figura 14: Percentual de transcritos identificados no proteoma predito do <i>Schistosoma mansoni</i> e dos transcritos processados por SL <i>trans-splicing</i> na fase de esporocisto classificados por categorias funcionais | 54 |
| Figura 15: Análise comparativa de classes funcionais de transcritos expressos e anotados no proteoma predito do <i>Schistosoma mansoni</i> e processados por SL <i>trans-splicing</i> na fase de esporocisto | 56 |
| Figura 16: Vias metabólicas nas quais enzimas derivadas de transcritos processados por <i>Spliced leader trans-splicing</i> estão presentes na fase de esporocisto | 58 |
| Figura 17: Análise de abundância dos transcritos processados por SL <i>trans-splicing</i> | 59 |

Lista de tabelas

| | |
|--|----|
| Tabela 1: Sequência do <i>Spliced leader</i> em diferentes filis. | 25 |
| Tabela 2: Tipos de análise por BLAST, bases de dados e parâmetros utilizados para anotação dos transcritos de esporocistos de <i>S. mansoni</i> | 41 |
| Tabela 3: Dados gerados na Plataforma Ion Torrent PGM™ System | 45 |
| Tabela 4: <i>Reads</i> sequenciadas na plataforma Ion Torrent PGM™ System após tratamento de qualidade..... | 52 |
| Tabela 5: Análise do alinhamento das <i>reads</i> geradas a partir das bibliotecas de esporocisto no genoma de referência de <i>Schistosoma mansoni</i> | 53 |
| Tabela 6: Anotação dos 20 transcritos mais abundantes na fase de esporocisto processados por <i>SL trans-splicing</i> | 60 |

Lista de abreviaturas e símbolos

µg - Micrograma

µL - Microlitro

µM - Micromolar

B.O.D - Demanda bioquímica do oxigênio

BLAST - *Basic local alignment search tool*

cDNA - DNA complementar

DEPC - Diethylpyrocarbonate

dNTP - Desoxirribonucleosídeo trifosfato

DTT - Dithiothietol

EDTA - Ácido etilenodiamino tetra - acético

ESTs - Etiquetas de sequências transcritas

g - Força centrífuga

Kb - Kilobase, 10³ bases

KEGG - *Kyoto Encyclopedia of Genes and Genomes* (Enciclopédia de Genes e Genoma de Kioto)

LE - Cepa de *S. mansoni* Luiz Evangelista

M - Molar

Mb - Megabase

mL - Mililitro

mM - Milimolar

mRNA - RNA mensageiro

NCBI - *National Center of Biotechnology Information* (Centro Nacional de Informação Biotecnológica)

ng - Nanogramas

nM - Nanomolar

nt - nucleotídeo

°C - Grau centígrado

OD - Densidade óptica

Oligo-d(T) - Oligonucleotídeo de deoxitimina

OMS - Organização Mundial de Saúde

pb - Pares de bases

pmol - picomoles

PZQ - Praziquantel

q.s.p - Quantidade suficiente para

RNA-seq - *RNA Sequencing*

RPKM - *Reads Per Kilobase of exon model Per Million Mapped Reads*

RPMI - *Roswell Park Memorial Institute medium*

RT-PCR - PCR de um Produto de transcrição reversa

siRNA - *small interfering RNA*

SL - *Spliced leader*

SNPs - Polimorfismo de nucleotídeo único

snRNA - *Small nuclear ribonucleic acid*

TA - Temperatura ambiente

TBE - Tampão Tris -borato EDTA, pH 8,0

Tris – Tris(hidroximetil)aminometano

Tris-HCL - Tris(hidroximetil)aminometano hydrochloride

U - Unidade

WHO - *World Health Organization* (Organização Mundial de Saúde)

Resumo

A esquistossomose é uma das mais importantes doenças parasitárias sendo endêmica em 76 países. No Brasil, esta representa um dos mais sérios problemas de saúde pública, persistindo devido às precárias condições de vida nas quais a população está inserida. O *Schistosoma mansoni* é a única espécie descrita no Brasil responsável por causar a esquistossomose. Este parasito é um metazoário digenético com várias características únicas em sua morfologia, fisiologia e ciclo de vida. Diante disto, é provável uma complexa regulação da expressão gênica em *S. mansoni* favorecendo mudanças morfológicas e bioquímicas atendendo as suas necessidades fisiológicas e de adaptação aos diversos ambientes. Assim, o mecanismo de regulação pós-transcricional, *Spliced leader (SL) trans-splicing*, existente no parasito, pode ser importante para viabilizar tais adaptações. Este mecanismo é apenas parcialmente compreendido sendo um amplo campo para pesquisa, auxiliando no desenvolvimento de possíveis ferramentas para o controle da esquistossomose. O *SL trans-splicing* ocorre através da adição de uma sequência identificada como *Spliced leader*, que é doada da extremidade 5' de um RNA pequeno, para alguns pré-mRNAs receptores, formando o éxon 5' terminal dos mRNAs maduros. Neste trabalho, foi realizado a identificação de transcritos processados por *SL trans-splicing* na fase de esporocisto do ciclo de vida do *S. mansoni* através da construção de biblioteca de cDNA enriquecida neste mecanismo. Assim, por meio de um estudo pioneiro de transcriptômica envolvendo a fase de esporocisto, foram construídas bibliotecas do tipo fragmento e utilizado um sequenciador de segunda geração para identificação de 1.191 transcritos processados por *SL trans-splicing* nesta fase. Neste trabalho foi observado que 10% dos transcritos expressos na fase de esporocisto são processados por *SL trans-splicing* se comparado à 5ª versão do proteoma predito de *S. mansoni* e 15%, se comparado com os 6.677 genes expressos identificados no transcriptoma da fase de esporocisto. Ainda, a partir da classificação dos transcritos em categorias funcionais e identificação de vias metabólicas, foi observado que o mecanismo de *SL trans-splicing* não está particularmente enriquecido, caracterizando-se como um mecanismo ubíquo. Em conjunto, estes dados enriquecem os estudos de transcriptômica do parasito *S. mansoni*. Compreender as reais funções deste mecanismo pode auxiliar no desenvolvimento futuro de uma ferramenta de intervenção terapêutica para o controle da esquistossomose mansônica.

Abstract

Schistosomiasis is a major parasitic disease, which is endemic in 76 countries. In Brazil, this is one of the most serious public health problem persisting due to the precarious living conditions in which the population is inserted. *Schistosoma mansoni* is the only specie described in Brazil responsible for causing schistosomiasis. This parasite is a digenetic metazoan with several unique features in its morphology, physiology and life cycle. Therefore, it is plausible a complex regulation of gene expression in *S. mansoni*, allowing morphological and biochemical changes that attend their physiological needs and to adapt to different environments. Therefore, the mechanism of post-transcriptional regulation, Spliced leader (SL) trans-splicing, existent in the parasite, may be important to enable these adaptations. This mechanism is only partial understood and thus a wide field for research towards the development of tools for schistosomiasis control. The SL trans-splicing occurs by the addition of a sequence identified as Spliced leader which is donated from the 5' end of a small RNA to receptor pre-mRNAs, forming the exon 5' end of mature mRNAs. Here, we carried out the identification of transcripts processed by SL trans-splicing in sporocyst of *S. mansoni* by constructing cDNA libraries enriched. Thus, by means of a pioneer study of transcriptomics involving the sporocyst stage, fragment libraries were constructed and used next-generation sequencing to identify 1.191 transcripts processed by SL trans-splicing in this stage. In this work, we found that 10% of transcripts expressed in the sporocyst stage are processed by SL trans-splicing when compared to the 5th version of the predicted proteome of *S. mansoni* and 15%, when compared to the 6,677 expressed genes identified in the transcriptome of the sporocyst stage. After the classification of transcripts in functional categories and identification of metabolic pathways we observed that the SL trans-splicing mechanism does not seem to be particularly enriched, being characterized as an ubiquitous mechanism. Together, our data improve knowledge acquired on transcriptomics studies of the parasite *S. mansoni*. Understanding the real function of this mechanism can assist in the future development of a therapeutic intervention tool for schistosomiasis control.

1 INTRODUÇÃO

1.1 A Esquistossomose

1.1.2 Aspectos Gerais

A esquistossomose é uma das infecções parasitárias mais importantes e difundidas no mundo. Após apenas da Malária, é considerada a segunda doença parasitária mais significativa em termos de saúde pública e impacto econômico. A infecção é prevalente em áreas tropicais e subtropicais, atingindo comunidades desprovidas de água potável e instalações sanitárias adequadas. Segundo a Organização Mundial de Saúde (OMS), estima-se que 200-209 milhões de pessoas no mundo estejam infectadas e que 600-779 milhões correm risco de serem infectadas em países endêmicos (WHO, 2013).

A humanidade convive com a esquistossomose desde a antiguidade, fato comprovado por estudos que verificaram a presença de ovos de *Schistosoma* em vísceras de múmias egípcias cuja origem remonta a 3.500 a.C. Originando-se provavelmente no Egito, essa endemia espalhou-se por vasta área do território africano seguindo o curso dos grandes rios. No Brasil, os especialistas acreditam que a esquistossomose tenha sido introduzida através de escravos, originários da costa da Guiné, Angola e antigo Congo, da África Ocidental e de Moçambique, em parte oriental do continente africano, estabelecendo-se inicialmente nas áreas de produção canavieira do Nordeste brasileiro, para onde drenava a maior parte da mão-de-obra escrava e onde existiam condições bioecológicas para que se completasse o ciclo evolutivo do parasito (CHIEFFI; WALDMAN, 1988).

A doença é causada por organismos do filo Platyhelminthes, classe *Trematodea*, ordem *Digenea*, da família *Schistosomatidae* e do gênero *Schistosoma*, que compreende espécies denominadas *Schistosoma mansoni*, *Schistosoma haematobium*, *Schistosoma japonicum*, *Schistosoma intercalatum*, *Schistosoma mekongi* (limitado a áreas do Laos e Camboja), *Schistosoma malayense* (Malásia), *Schistosoma sciammarella* (WHO: <http://www.who.int/tdr/diseases/schisto/diseaseinfo.htm>). As três espécies *S. mansoni* (prevalente no Brasil), *S. haematobium* e *S. japonicum* são consideradas mais importantes para a saúde humana devido a sua alta prevalência.

Os fatores sociais, culturais, comportamentais e econômicos, juntamente com os fatores ambientais e ecológicos, resultam na variação epidemiológica da esquistossomose, inclusive com relação à prevalência e à intensidade da infecção e ao potencial de controle (HUANG, Y.; MANDERSON, 1992).

Uma das estratégias de controle da esquistossomose que está sendo utilizada com sucesso nos últimos 25 anos, é o tratamento com a droga praziquantel (CARABIN; GUYATT; ENGELS, 2000). Apesar da utilização deste fármaco em muitos países, várias discussões têm sido levantadas após o aparecimento de casos de resistência ao medicamento (WANG, W.; WANG; LIANG, 2012). Assim, estes relatos reforçam a importância do estudo científico das espécies causadoras da esquistossomose e de novas ferramentas e estratégias adicionais adaptadas para a eliminação desta doença no Brasil e em outros países afetados.

1.2 Organismo de estudo - *Schistosoma mansoni* e seu ciclo de vida

O *Schistosoma mansoni* é um trematódeo acelomado, de simetria bilateral e dimorfismo sexual durante a fase adulta (ROLLINSON & SIMPSON, 1987). O parasito tem um ciclo de vida complexo possuindo dois hospedeiros: o homem, como hospedeiro definitivo, e um molusco pertencente à família *Planorbidae* do gênero *Biomphalaria*, como hospedeiro intermediário (DIAS *et al.*, 1994; SANTOS, 2002).

A transmissão da esquistossomose mansônica, depende necessariamente da existência dos hospedeiros intermediários que, no Brasil, são três espécies de caramujos do gênero *Biomphalaria*: *Biomphalaria straminea*, *Biomphalaria glabrata* e *Biomphalaria tenagophila* (CARVALHO *et al.*, 2005).

O ciclo de vida do *S. mansoni* (Figura 1) se desenvolve no hospedeiro definitivo vertebrado, onde ocorre maturação dos vermes e reprodução sexuada, e no hospedeiro intermediário invertebrado, onde ocorre a reprodução assexuada (DAVIS, 1985). Os ovos do *S. mansoni* são eliminados juntamente às fezes do hospedeiro definitivo infectado e, quando alcançam uma coleção hídrica, eclodem e liberam larvas ciliadas denominadas miracídios, que nadam ativamente e penetram nos moluscos. Após 48 horas, os miracídios perdem as placas ciliares e transformam-se em esporocistos primários, possuindo em seu interior as células germinativas, em número de 50 a 100, que iniciam um processo intenso de multiplicação (poliembrião), fazendo com que, após 72 horas, a larva chamada de esporocisto primário, esporocisto mãe ou, simplesmente, esporocisto I, dobre de tamanho. Na segunda semana de infecção, observa-se no interior do esporocisto, uma série de ramificações tubulares, que preenchem todos os espaços intercelulares do tecido conjuntivo. No interior dessas ramificações, as células germinativas encontram-se em franca multiplicação. Em condições ideais de temperatura (entre 25° C e 28° C) ocorre a formação dos esporocistos secundários, esporocistos filhos ou esporocisto II, que se inicia a partir do 14° dia após a

penetração do miracídio. A saída dos esporocistos do local de penetração do miracídio, onde a maioria se desenvolve, até as glândulas digestivas, ou hepatopâncreas, leva de dois a três dias, sendo que a sua migração processa-se ativamente através dos tecidos do molusco. A localização final dos esporocistos será nos espaços intertubulares da glândula digestiva, local com riqueza nutritiva onde começam, então, sofrer profundas modificações anatômicas no seu conteúdo de células germinativas. O ovotestis, ou glândula reprodutiva, poderá também abrigar os esporocistos migrantes, mas com frequência menor, principalmente nas infecções com poucos miracídios. Após atingirem o destino final, os esporocistos secundários apresentam células embrionárias e cercárias desenvolvidas ou em desenvolvimento que, por sua vez, sairão pelo poro de nascimento em condições ideais de temperatura (cerca de 28 °C), após um período de 27 a 30 dias. Esta última geração de células embrionárias poderão dar origem aos novos esporocistos chamados de esporocistos terciários. Impressionantemente, um único miracídio pode gerar cerca de 300 mil cercárias (NEVES, 2005). Liberadas na água pelos caramujos, as cercárias nadam ao encontro do hospedeiro vertebrado, liberam secreções da glândula acetabular facilitando a sua penetração pela epiderme do homem e, em seguida, iniciará o processo de transformação em esquistossômulos (STIREWALT *et al.*, 1983; WILSON & COULSON, 1986). Os esquistossômulos migram, via circulação sanguínea e linfática, para o coração, pulmão, fígado e veias mesentéricas transformando-se em vermes adultos 28 a 48 dias após a penetração. Nas veias mesentéricas inferiores ocorre a cópula, seguida de ovoposição (LOVERDE & CHEN, 1991). Os ovos são eliminados para o meio ambiente juntamente com o bolo fecal do hospedeiro e quando alcançam coleções hídricas os miracídios são liberados, estimulados pelas temperaturas mais altas, luz intensa e oxigenação da água, reiniciando então, o ciclo de vida do parasito. Em condições favoráveis, o ciclo evolutivo do *S. mansoni* se completa em torno de 80 dias (KATZ, 2003). O agravamento da doença se dá devido à presença dos ovos do parasito, uma vez que grande parte dos ovos não ultrapassam a parede intestinal e, através da corrente sanguínea, vão se alojar nos órgãos (principalmente no fígado) e tecidos do hospedeiro resultando em reações ectópicas e na formação do granuloma. Esse consiste da resposta inflamatória do sistema imune do hospedeiro contra antígenos solúveis do ovo e é o principal responsável pelas variações clínicas e pelas complicações digestivas e circulatórias da esquistossomose (NEVES, 2005).

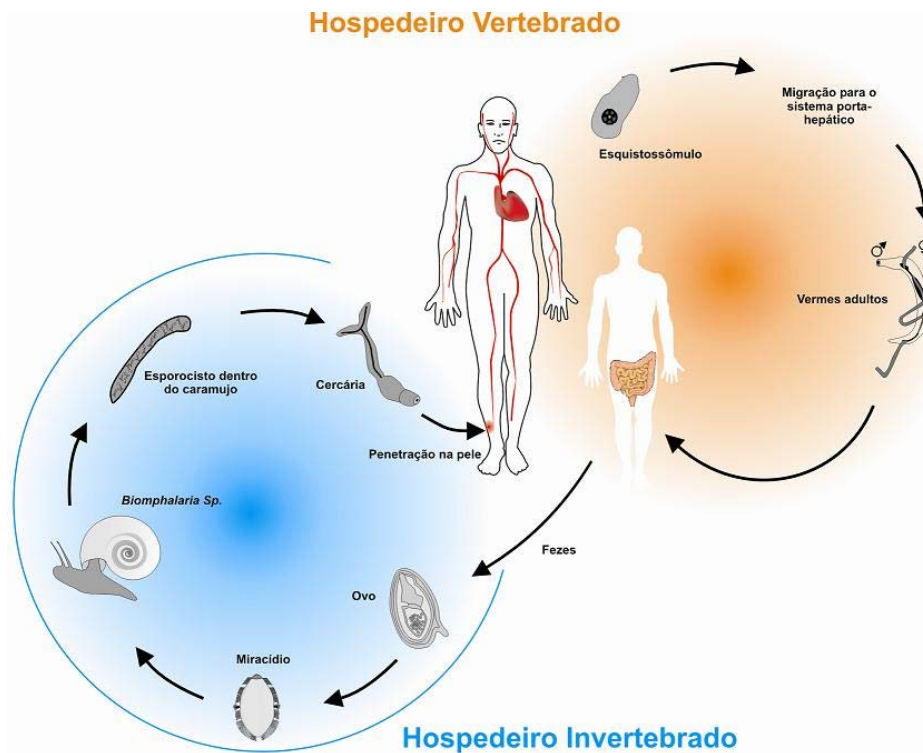


Figura 1: O ciclo de vida do parasito *Schistosoma mansoni*. As cercárias, ao saírem do caramujo, penetram pela pele do hospedeiro e perdem a cauda, iniciando-se o processo de transformação em esquistossômulos. Os esquistossômulos migram até as veias e, através da circulação sanguínea, chegam aos pulmões. Em seguida, são enviados pela circulação geral a todas as partes do corpo do hospedeiro. Quando alcançam o sistema porta intra-hepático completam seu desenvolvimento e se acasalam. As fêmeas depositam seus ovos nas pequenas vênulas dos sistemas porta e perivesical, e são eliminados com as fezes. Os ovos que não são eliminados nas fezes são levados pela corrente sanguínea principalmente para o fígado, dando origem aos granulomas. Já os que são liberados nas fezes, ao caírem em água fresca, eclodem e liberam os miracídeos que nadam e penetram no hospedeiro intermediário, o caramujo do gênero *Biomphalaria*. Os estágios dentro do caramujo incluem as gerações de esporocistos, produção e liberação de cercárias.

Fonte: <http://rgmg.cpqrr.fiocruz.br/content/schistosoma-mansoni-transcriptome-project>.

1.3 O genoma do *Schistosoma mansoni*

Em 1992, foi iniciado o projeto de descoberta gênica em *S.mansoni* coordenado pelo Dr. Sérgio Pena e Dr. Andrew Simpson. A estratégia básica do projeto visava o sequenciamento em um único passo de clones de cDNA selecionados aleatoriamente de

diferentes bibliotecas, para obtenção das Etiquetas de Sequências Transcritas (ESTs) (FRANCO *et al.*, 1997).

Estudos iniciais sobre o genoma de parasitos do gênero *Schistosoma* revelam que estes possuem um genoma diplóide com a presença de sete pares de cromossomos autossomos e um par de cromossomos sexuais, sendo o sexo homogamético o macho (ZZ) e o heterogamético a fêmea (ZW) (GROSSMAN; SHORT; CAIN, 1981) (SHORT, 1983). Curiosamente, o genoma do *S. mansoni* é aproximadamente 10 vezes maior do que o genoma de protozoários.

O primeiro rascunho do genoma de *S. mansoni* foi publicado em 2009 com mais de 360 milhões de bases. O genoma foi montado em 19.022 *scaffolds* e anotado com 11.809 genes correspondendo a 13.197 transcritos (BERRIMAN *et al.*, 2009). Entretanto, mais recentemente, utilizando os dados originais do rascunho do genoma, novos dados obtidos por sequenciamento de Sanger e também sequenciamento de nova geração, a nova montagem do genoma resultou em uma versão menos fragmentada de 885 *scaffolds* e 364.5 milhões de bases contendo 10.852 genes (PROTASIO *et al.*, 2012).

O estudo genômico em larga escala deste parasito é particularmente interessante, pois se conhece muito pouco sobre os mecanismos moleculares dos quais o *S. mansoni* utiliza-se para adaptar aos seus hospedeiros e ao ambiente aquático. Além disso, trata-se de uma abordagem promissora para o entendimento da biologia do parasito, de mecanismos de resistência às drogas, de variação antigênica, de variações na dinâmica populacional e dos mecanismos de controle de expressão gênica, como ocorre durante o processamento pós-transcricional de RNA no parasito: o *Spliced leader trans-splicing*. Assim, existe um grande campo a ser explorado com vistas à identificação de novos alvos de drogas, vacinas e moléculas diagnósticas, auxiliando na compreensão dos mecanismos moleculares envolvidos na infecção, patologia e resistência às drogas contra a esquistossomose.

1.4 Mecanismos de processamento de RNAs

Alguns eventos moleculares importantes no metabolismo de RNA ocorrem durante o processo pós-sintético. Os processamentos pós-transcricionais são responsáveis pelo aumento da variabilidade dos produtos gênicos, assim, o mecanismo de *splicing* é um exemplo disto. As enzimas que catalisam o *splicing* de RNAs consistem de moléculas de RNAs agregadas a proteínas, formando complexos ribonucleoprotéicos. Estes RNAs catalíticos são identificados como ribozimas (NELSON; COX, 2000).

Para que ocorra o *splicing* é necessária a formação de uma maquinaria chamada spliceossomo, que é composta de 5 ribozimas, chamadas U1, U2, U4, U5 e U6, e de mais de 50 proteínas, formando as ribonucleoproteínas. Os RNAs que constituem o spliceossomo são chamados snRNAs (*Small nuclear ribonucleic acid*). São RNAs pequenos compostos de 100 a 200 nucleotídeos, muito conservados, que possuem um sítio poliuridínico (NELSON; COX, 2000). Neste processo, é preciso que haja sítios específicos no transcrito primário, onde as ribozimas se ligarão. Primeiro ocorre a ligação de U1 no sítio específico do transcrito primário, na parte 3' do éxon e 5' do intron. Logo após, ocorre a ligação de U2 em um sítio polipirimídico presente internamente ao intron. Ocorre então, o recrutamento das outras ribonucleoproteínas, clivagem dos éxons e introns, junção dos éxons e liberação dos introns em forma de laços (NELSON; COX, 2000) (Figura 2).

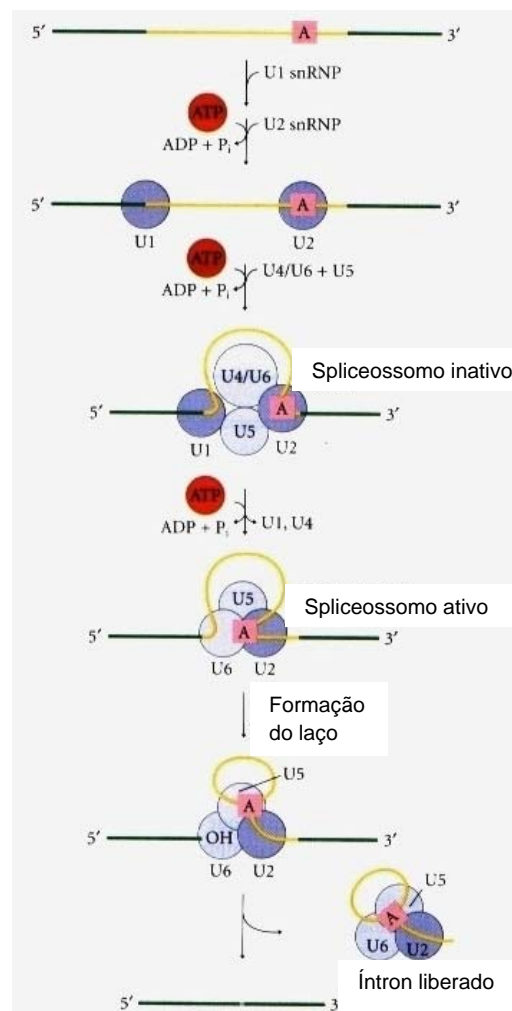


Figura 2: Montagem do spliceossomo. Os snRNP U1 e U2 e os remanescentes snRNOPs (o complexo U4/U6 e U5) ligam-se para formar um spliceossomo inativo. Rearranjos internos convertem essa espécie em um spliceossomo ativo, no qual U1 e U4 foram expelidos e U6 está pareado tanto com o sítio de emenda 5' como com o sítio A. Isso é seguido pelas etapas catalíticas, que correm paralelas àquelas da emenda dos introns do grupo II. Fonte: Nelson e Cox, 2000.

O tipo de *splicing* que irá acontecer em um sítio receptor de *splicing* é determinado pela presença ou ausência de um sítio doador anterior a este. Se houver um sítio doador na junção éxon/intron, este se juntará ao próximo sítio receptor no final do intron ocorrendo assim, o *cis-splicing*. Caso haja um sítio receptor desprovido de um sítio doador a montante, ocorrerá o *trans-splicing*. Ainda, a presença de sítios receptores despareados serão alvo da inserção do SL e, assim, resolução de transcritos policistrônicos em monocistrônicos. O spliceossomo, no entanto, tem clara preferência pelo *cis-splicing*, mas a base deste mecanismo não é ainda conhecida (HASTINGS, 2005) (Figura 3).

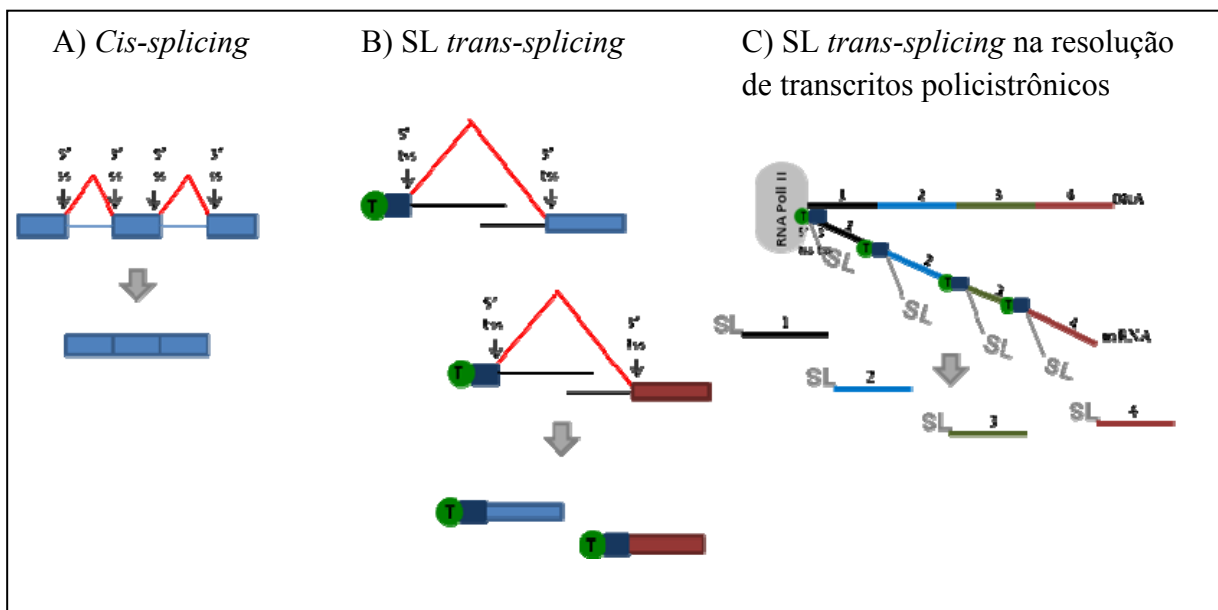


Figura 3: Diferentes tipos de *splicings*. (A) *Cis-splicing* - une os exons de um mesmo transcrito de pré-mRNA através dos sítios de *splicing* (ss) (B) *SL trans-splicing* - une o sítio de *splicing* 5' localizado no SL RNA ao sítio 3' do *trans-splicing* localizado no pré-mRNA. (C) *SL trans-splicing* na resolução de transcritos policistrônicos - Realiza a resolução de transcritos policistrônicos da RNA polimerase II em transcritos monocistrônicos. Os círculos verdes representam o cap trimetilguanosina no exon do SL RNA e os tss representa os sítios de *Spliced leader trans-splicing*. Fonte: Adaptada de Lasda e Blumenthal, 2011.

1.5 *Trans-splicing*

Como dito anteriormente, além do *cis-splicing*, o spliceossomo também é capaz de realizar o *trans-splicing*, que é a transferência spliceossômica de pequenas sequências derivadas de RNAs transcritos independentemente.

O mecanismo de *trans-splicing* ocorre naturalmente de duas maneiras. A primeira, ocorre através da adição de uma sequência identificada como *Spliced leader* (SL), que é

doador da extremidade 5' de um RNA pequeno, não poliadenilado, para alguns pré-mRNAs receptores, para formar o éxon 5' terminal dos RNAs mensageiros (mRNA) maduros.

A segunda maneira de ocorrência é o *trans-splicing* doador de pré mRNA, no qual o sítio de *splicing* doador em um pré-mRNA é transferido para um sítio receptor em um outro. Este processo pode ocorrer mais de uma vez no mesmo gene e está presente em plantas e algas (BONEN, 1993).

Ao contrário do mecanismo de *cis-splicing*, o SL *trans-splicing* não é um mecanismo universal em eucariotos. A porcentagem dos genes que são sujeitos a este processamento varia entre as diferentes espécies. Nos protistas, aparentemente, todos os transcritos sofrem SL *trans-splicing*. Já em metazoários, a fração de genes processados varia de 70-90% em *Caenorhabditis elegans* (BLAXTER; LIU, 1996) e *Ascaris lumbricoides* (MARONEY *et al.*, 1995). Através da metodologia de pirosequenciamento, utilizando como modelo o cordado *Ciona intestinalis*, foram identificados 8.790 genes com a sequência do SL, o que corresponde a aproximadamente 58% do total de genes desse organismo (MATSUMOTO *et al.*, 2010). Ainda, foi demonstrado recentemente que 90% dos transcritos são processados por SL *trans-splicing* no nematóide *Pristionchus pacificus* (SINHA *et al.*, 2014).

Na tentativa de compreender a origem e diversidade da sequência do SL, Bitar e colaboradores (2013) demonstraram o número de sequências variadas nos diferentes grupos, incluindo os nematódeos (182), platelmintos (98), dinoflagelados (95), cnidários (4), rotíferos (2) e cordados (1). Sendo o maior número de sequências presente no grupo dos kinetoplastídeos (757) e quase a metade especificamente de *Trypanosoma spp* (544).

Todo organismo que sofre SL *trans-splicing* tem um ou alguns SL RNA, contando com várias cópias inseridas no genoma (HASTINGS, 2005). Em *C. elegans*, por exemplo, tem sido identificado dois tipos de SL possuindo funções distintas : SL1 e SL2 (HUANG, X.; HIRSH, 1989).

Segundo Pettitt e colaboradores (2008), a presença de múltiplas sequências de SL RNAs em organismos multicelulares é comum. Em um trabalho recentemente realizado pelo nosso grupo foi identificado 5 genes de SL RNA anotados na 5ª versão do genoma de referência do parasito *S. mansoni* e 7 outros genes não anotados, além de que foi detectado 32 *loci* no genoma contendo a sequência SL (MARTINS, 2014). A necessidade de múltiplos genes de SL RNA, presumivelmente, decorre do fato de que cada evento de SL *trans-splicing* consome uma molécula de SL RNA, de modo que, a produção contínua de mRNA maduro requer uma produção de SL RNA de alto nível (YEATS *et al.*, 2010).

Nos platelmintos, o mecanismo de SL *trans-splicing* tem sido observado em todas as espécies estudadas até o momento. Neste filo, além das sequências dos SL RNAs serem longas, variando de 36-37 nucleotídeos, as sequências são bastante variadas entre as diferentes espécies estudadas. E, diferentemente da maioria dos outros filios, o códon de terminação das sequências dos SL RNAs em platelmintos é ATG (BITAR *et al.*, 2013).

De acordo com a tabela a seguir, o parasito *S. mansoni* possui somente uma única sequência do SL contendo 36 nucleotídeos advindos de um RNA não poliadenilado de 93 nucleotídeos. Esta sequência é espécie-específica, não apresentando similaridade às de outros organismos (RAJKOVIC *et al.*, 1990) (Tabela 1).

TABELA 1: SEQUÊNCIA DO *SPLICED LEADER* EM DIFERENTES FILOS

| | Espécies | Tamanho do SL | Tamanho do SL RNA | % Genes <i>Trans-spliced</i> |
|-----------------|------------------------------------|-----------------------|--------------------------|--------------------------------|
| Euglenozoa | <i>Trypanosoma brucei</i> | 39 nt | 141 nt | 100% |
| | <i>Euglena gracilis</i> | 26 nt | 101 nt | |
| Dinoflagellates | <i>Amphidinium carterae</i> | 22 nt | | 67 or 100% |
| | <i>Karlodinium micrum</i> | 22 nt | 56 nt | ~100% |
| Sponges | <i>Heterochone sp.</i> | 39-41 nt | | |
| Cnidarians | <i>Clytia hemisphaerica</i> | 33-37 nt | | 23% |
| | <i>Hydra vulgaris</i> | SL-A 24 nt SL-B 46 nt | SL-A 80 nt SL-B 107 nt | 30% |
| Ctenophores | <i>Mnemiopsis leidyi</i> | 39-43 nt or 55 nt | 128 nt or 138 nt | 3% |
| | <i>Pleurobrachia pileus</i> | 37 nt | | 40% |
| Flatworms | <i>Schistosoma mansoni</i> | 36 nt | 93 nt | |
| | <i>Stylochus zebra</i> | 51 nt | 110 nt | |
| | <i>Echinococcus multilocularis</i> | 36 nt | 104 nt | ~25% |
| Crustaceans | <i>Parhyale hawaiiensis</i> | 33-35 nt | 97-98 nt | 10% |
| Chaetognaths | <i>Spadella cephaloptera</i> | 36 nt | 104 nt | 41% |
| Tunicates | <i>Oikopleura dioica</i> | 40 nt | 93 nt | 12-24% |
| | <i>Ciona intestinalis</i> | 16 nt | 46 nt | 50-58% (48% freq. 19% infreq.) |
| | <i>Halocynthia roretzi</i> | 24 nt | | |
| Rotifers | <i>Adineta ricciae</i> | 23 nt | 105-106 nt | 50-60% |
| Nematodes | <i>Caenorhabditis elegans</i> | 22 nt | SL1 95 nt SL2 107-114 nt | 70% |
| | <i>Ascaris sp.</i> | 22 nt | 106 nt | 80-90% |
| | <i>Trichinella spiralis</i> | 22 nt | 97-99 nt | 1% |

Fonte: Adaptado de John Wiley & Sons, 2011

Em *S. mansoni*, Protasio e colaboradores (2012) identificaram 1.178 genes processados por SL *trans-splicing* nas fases de cercária, esquistossômulo e verme adulto do parasito, correspondendo a 11% do número total de transcritos. Entretanto, dados preliminares do nosso grupo utilizando a metodologia de RNA-seq e envolvendo as fases de cercária, esquistossômulo, verme adulto e miracídio indicam que este percentual é ainda maior, 63% dos transcritos são processados por SL *trans-splicing* (MARTINS, 2014).

Com a execução desse trabalho realizado pelo nosso grupo, várias informações importantes sobre o mecanismo de SL *trans-splicing* puderam ser esclarecidas ou confirmadas com outros trabalhos já realizados. Em contraste com dados anteriormente publicados, de que o primeiro éxon é o substrato exclusivo para a inserção da sequência do SL (CONRAD R.; THOMAS J.; SPIETH J, 1991) foi observado neste trabalho que, apesar de o mecanismo de SL *trans-splicing* ocorrer de forma mais significativamente no primeiro éxon, ele também pode ocorrer nos demais éxons. Adicionalmente, foi possível observar e confirmar os dados de outros estudos com *Trypanosoma brucei* (SIEGEL; TAN; CROSS, 2005) na existência de características dos íntrons que são processados por *cis-splicing* ou SL *trans-splicing*, fato observado para um único transcrito *ubiquinol-cytochrome C reductase complex ubiquinol binding protein* (UbCRBP) de *S. mansoni*, em um trabalho prévio (MOURÃO *et al.*, 2013). Assim, através de um grupo de genes que apresentam pelo menos um éxon interno foi possível confirmar que, os íntrons que servem como substrato para os eventos de SL *trans-splicing* são geralmente mais longos do que os de eventos de *cis-splicing* (MARTINS, 2014).

Em relação a frequência do mecanismo de SL *trans-splicing* e taxas de expressão gênica, não foi possível estabelecer uma correlação direta entre estes dois quesitos através do trabalho realizado. Uma vez que, foram encontrados tanto genes com alto e baixo valores de expressão sendo processados por SL *trans-splicing* (MARTINS, 2014).

Por fim, diante das tentativas de elucidar a importância do mecanismo de SL *trans-splicing* para o desenvolvimento do *S. mansoni*, Mourão e colaboradores (2013) através de RNAi analisaram um conjunto diversificado de transcritos que são processados por SL *trans-splicing* durante diferentes fases do ciclo de vida do *S. mansoni*, com exceção da fase de esporocisto. A tentativa de silenciamento do mecanismo nessa fase gerou uma redução de, em média, 50% no nível de expressão dos genes analisados, sabidamente processados por SL *trans-splicing* e as larvas dos parasitos tratados com siRNA (*small interfering RNA*) se mostraram com tamanho reduzido (MOURÃO *et al.*, 2013).

1.6 Possíveis funções do mecanismo de *Spliced leader trans-splicing*

Acredita-se que o SL *trans-splicing* participe no processo de maturação do mRNA (BITAR *et al.*, 2013), tendo como objetivos:

1. Prover o 5' *cap* para RNAs transcritos pela RNA polimerase I, que são codificadores de proteínas (exclusivo em Kinetoplastídeos);

2. Resolver transcritos policistrônicos da RNA polimerase II em transcritos individuais contendo o 5' *cap*;

3. Acentuar a eficiência de tradução através de adição do *cap* hipermodificado ou da sequência *leader* no mRNA.

Entretanto, uma função adicional, e menos entendida, é a de retirar as regiões 5' não traduzidas de pré-mRNAs. A porção 5' de um pré-mRNA monocistrônico talvez seja retirada quando contenha elementos que possam comprometer a tradução do mRNA, como por exemplo, um códon inicial (AUG) fora da fase de leitura (HASTINGS, 2005).

Apesar da existência de diversos trabalhos que identificam o mecanismo de SL *trans-splicing* em diferentes organismos, ainda são necessários mais estudos sobre este processo, uma vez que suas funções ainda permanecem não esclarecidas.

1.7 Sequenciamento de transcritos por RNA-Seq

Através dos recentes avanços tecnológicos no sequenciamento de DNA, atualmente é possível realizar o sequenciamento massivo de cDNA a partir de RNA celular, e este processo é conhecido como RNA-seq (GARBER *et al.*, 2011). O termo RNA-seq tem sido utilizado para representar o transcriptoma revelado por sequenciamento de cDNA por sequenciadores de segunda geração.

Esta nova tecnologia possibilita a quantificação dos níveis de expressão gênica, mesmo em transcritos que possuam níveis mais baixos de expressão, devido a sua alta sensibilidade, além de propiciar economia de tempo e redução de custos.

Em experimentos de RNA-Seq, os fragmentos de cDNA são sequenciados e mapeados nos genes e, idealmente, em posições únicas. Sendo adequadamente normalizadas, a contagem de fragmentos pode ser usada como uma medida da abundância relativa dos transcritos. Em alguns pacotes estatísticos de análise de RNA-Seq, a abundância dos transcritos é medida em fragmentos por quilo bases de exons por milhões de fragmentos mapeados (FPKM), que é análoga à de uma única leitura "RPKM" (*Reads Per Kilobase of exon model Per Million Mapped Reads*), proposta originalmente por Mortazavi e colaboradores (2008): as contagens de *reads* são divididas pelo tamanho do transcrito (kb) vezes o número total de milhões de sequências mapeadas.

Ainda, o RNA-Seq permite a detecção de SNPs (polimorfismo de nucleotídeo único) nos transcritos, identificação de fusão de transcritos, descoberta de novas classes de RNA e novos eventos de *splicing* alternativo, além da análise de expressão de alelos (FLINTOFT,

2008) (MARIONI *et al.*, 2008) (WANG, Z.; GERSTEIN; SNYDER, 2009) (WILHELM; LANDRY, 2009) (COSTA *et al.*, 2010) (ROBERTS *et al.*, 2011). O grande sucesso das novas tecnologias de sequenciamento de transcriptomas se deve também ao fato de que estas possibilitam a superação de uma das maiores limitações dos projetos de sequenciamento aleatório de cDNAs, derivados de bibliotecas, que geram ESTs - a grande redução no número de sequências novas amostradas com o aumento na quantidade de informação sequenciada (WANG, Z.; GERSTEIN; SNYDER, 2009) (GARBER *et al.*, 2011).

Portanto, os sequenciadores de segunda geração são uma ferramenta útil e ideal para alcançar o objetivo proposto neste trabalho que é a identificação de transcritos que são processados por SL *trans splicing* em *S. mansoni*, possibilitando conhecer este mecanismo e determinar a sua importância na regulação pós-transcricional da expressão gênica na fase de esporocisto do parasito.

2 JUSTIFICATIVA

A esquistossomose é uma das parasitoses humanas mais difundidas no mundo e sua ocorrência está relacionada à ausência ou precariedade de saneamento básico. Considerada uma doença crônica parasitária, a esquistossomose está associada a uma variedade de formas clínicas com alta morbidade (CAFFREY, 2007). De acordo com os dados do Inquérito Nacional de Prevalência da Esquistossomose e Geo-helmintos que começou a ser realizado em 2010, coordenado pelo pesquisador Naftale Katz do Centro de Pesquisas René Rachou/FIOCRUZ, atualmente, existem um milhão e quatrocentas mil pessoas com esquistossomose no Brasil (KATZ, 2014).

Ao longo dos últimos 25 anos de tratamento e controle da doença, a droga que tem sido mais utilizada é o praziquantel (CARABIN; GUYATT; ENGELS, 2000). Entretanto, apesar da droga ser eficaz, apenas a quimioterapia nas regiões endêmicas é pouco efetiva, pois é menos efetiva sobre as fases imaturas do ciclo de vida do parasito, dificultando a erradicação da doença e sendo necessária a repetição do tratamento (SABAH *et al.*, 1986). Além disso, já foi descrita a baixa eficácia do tratamento no Egito e Senegal, onde linhagens resistentes à droga foram isoladas (FALLON *et al.*, 1995) (ISMAIL *et al.*, 1999). Desse modo, fica evidente a necessidade do desenvolvimento de novas drogas efetivas contra o parasito e de novas abordagens para erradicar a doença.

Visto que o *S. mansoni* apresenta um ciclo de vida complexo, envolvendo adaptação em ambientes distintos como a água e o meio interno de seus hospedeiros, numerosos e complexos mecanismos de regulação de expressão gênica devem ocorrer neste parasito, como o mecanismo de *SL trans-splicing*. Entretanto, estes mecanismos são apenas parcialmente compreendidos e por isso, tornam-se um amplo campo para pesquisa. Assim, o mecanismo de *SL trans-splicing* pode ter um papel fundamental no processo de adaptação do parasito, tornando-se potencialmente um mecanismo alvo para ação de medidas de bloqueio, devido a sua especificidade ao parasito, o que poderá contribuir para o controle e interrupção das infecções.

Propomos nesse trabalho a identificação de transcritos processados por *SL trans-splicing* na fase de esporocisto do parasito *S. mansoni* visando enriquecer o conhecimento, ainda insipiente, sobre o papel desse mecanismo no processo de expressão gênica do parasito. Os transcritos identificados neste trabalho serão valiosos para outros estudos de análise da expressão gênica na fase de esporocisto, uma vez que, até o momento não existem estudos de

transcriptômica, sobretudo do mecanismo de *SL trans-splicing*, envolvendo esta fase do ciclo de vida do *S. mansoni*.

A fase de esporocisto apresenta algumas características peculiares que podem servir como importantes alvos para o controle da esquistossomose, como o modo de reprodução assexuada, o intenso metabolismo dentro do molusco e uma enorme capacidade na geração da fase infectante, as cercárias. Segundo Carvalho e colaboradores (2008) um dos fatos relacionados aos riscos da aquisição da esquistossomose durante o contato da população com os ambientes hídricos é a reprodução assexuada do parasito durante a etapa do desenvolvimento intramolusco. A capacidade de multiplicação das formas infectantes, as cercárias, a partir da infecção dos caramujos por um único miracídio, além da perpetuação de *S. mansoni* nos hospedeiros definitivos, com a expectativa do crescimento do número de parasitos, garante a manutenção dos focos da doença, mesmo que o controle implique na redução da sua prevalência.

3 OBJETIVOS

3.1 Objetivo Geral

Identificar e analisar genes que são processados pelo mecanismo de *Spliced leader trans-splicing* na fase de esporocisto do parasito *Schistosoma mansoni* através do sequenciamento massivo de RNA.

3.2 Objetivos Específicos

- Obter esporocistos de *S. mansoni* por meio da transformação *in vitro* de miracídios;
- Construir bibliotecas de cDNAs da fase de esporocisto enriquecidas em transcritos que são processados por SL *trans-splicing*;
- Obter sequências de cDNAs em larga escala para o estudo do transcriptoma;
- Identificar, agrupar e anotar as sequências geradas;
- Identificar e classificar funcionalmente os genes processados por SL *trans-splicing* e buscar por possíveis funções e participação em processos biológicos na fase de esporocisto do parasito.

4 MATERIAL E MÉTODOS

4.1 Obtenção de material biológico

Os animais de experimentação, hamsters da linhagem Golden, foram obtidos no Biotério de Produção do Centro de Pesquisa René Rachou (FIOCRUZ) e transferidos para o Biotério de Experimentação Animal do Centro de Pesquisa René Rachou (BIOTEX). A utilização desses animais é aprovada pela Comissão de Ética para Uso Animal (CEUA) da Fundação Oswaldo Cruz, sob o número LW-30/13.

A obtenção dos miracídios foi realizada empregando-se a técnica descrita por Pelegrino e Katz (1968), conforme a seguir: 10 hamsters foram infectados com 120 cercárias da cepa LE de *S. mansoni* e após 7-8 semanas os animais foram sacrificados para retirada do fígado. Em ambiente estéril, os fígados foram colocados em um béquer contendo salina 0.85% e em seguida, utilizando liquidificador autoclavado, foram triturados. O material foi transferido para um cálice de 1 litro contendo gaze (estéril) em sua abertura para retenção de material não triturado. Após duas lavagens por sedimentação com salina 0.85% e água desclorada, o cálice contendo os ovos foi mantido na ausência de luz por 30 minutos. Após a última lavagem, o sobrenadante foi descartado e os 200 mL restantes foram transferidos para um balão volumétrico. Este balão foi parcialmente coberto com papel alumínio, mantendo a parte restante em contato com a luz. O balão contendo os ovos foi colocado sob luz artificial para estimular a eclosão dos miracídios, que ocorre após, aproximadamente, 15 minutos. Devido ao fototropismo positivo, os miracídios nadam em direção à luz.

4.1.1 Transformação *in vitro* de miracídios em esporocistos primários

Os miracídios obtidos foram coletados com auxílio de pipetador e pipeta graduada em tubos e mantidos no gelo para sedimentação. Após esse período, o sobrenadante foi descartado e o volume restante de cada tubo foi transferido para um único tubo e mantido no gelo por mais 30 minutos. Após o tempo de sedimentação foi retirado o sobrenadante, restando aproximadamente 5 mL. Cada 1 mL da suspensão contendo os parasitos foi colocada em garrafa para cultura de 25 cm² contendo 15 mL de meio RPMI 1640 suplementado (5% de soro fetal bovino e 100 µg/mL de antibiótico penicilina-estreptomicina). A garrafa de cultura foi mantida por 48 horas em estufa B.O.D a 27°C a fim de garantir a transformação dos miracídios em esporocistos primários através da perda das placas ciliares.

Foram feitas duas réplicas biológicas, realizadas em experimentos independentes.

4.2 Construção da biblioteca de cDNA enriquecida em transcritos processados por *SL trans-splicing*

4.2.1 Extração de RNA total e tratamento com DNase

Para a extração de RNA total da fase de esporocisto foi utilizado o método de extração pelo *TRIzol*[®] *Reagent* (Invitrogen) associado a purificação com o kit de extração de RNA *RNeasy*[®] Mini kit (Qiagen). Sucintamente, foram adicionados aos esporocistos 50 µL de TRIzol e homogeneizados com auxílio de pistilo. Em seguida, foram adicionados 200 µL de clorofórmio, as amostras foram homogeneizadas em vortex por 15 segundos e mantidas por 3 minutos à temperatura ambiente. Posteriormente, as amostras foram centrifugadas a 12000 g por 15 minutos a 4°C e a parte aquosa, na qual se encontra o RNA, foi transferida para um novo tubo. Nesta etapa, foram adicionados 280 µL de etanol 70%. Em seguida, as amostras foram transferidas para as colunas do kit RNeasy Mini Kit (QIAGEN), centrifugadas por 15 segundos a 10.000 g e o eluato descartado. Foram adicionados às colunas 700 µL de Buffer RW1 e centrifugadas a 10.000 g por 15 segundos. Em seguida, 500 µL de Buffer RPE foram adicionados e as colunas novamente centrifugadas a 10.000 g por 15 segundos. Foram adicionados 500 µL de Buffer RPE e as colunas centrifugadas por 2 minutos a 10.000 g. Finalmente, o RNA foi eluído em 30 µL de água livre de RNase pré-aquecida a 37 °C, as colunas centrifugadas a 10.000 g por 1,5 minutos.

Com o objetivo de remover o DNA genômico contaminante, o RNA total foi tratado com DNase por 30 minutos a 37°C, utilizando o kit Ambion[®] TURBO DNA-free[™] (Invitrogen), segundo manual do fabricante. Posteriormente, a quantidade e qualidade do RNA total foram analisadas utilizando o *Qubit*[®] 2.0 *Fluorometer* (Invitrogen) e *Bioanalyzer 2100* (Agilent), respectivamente.

4.2.2 Purificação dos mRNAs e síntese de cDNAs

Para a purificação dos mRNAs foram utilizadas esferas magnéticas *Dynabeads Cl Streptavidin* (Invitrogen- Life Technologies). Essas esferas possuem uma monocamada de estreptavidina ligada covalentemente à sua superfície. Assim, com a utilização de um oligo dT biotilado, os mRNAs são capturados através da cauda de poli-A, conforme o protocolo a seguir:

-Lavagem e tratamento das esferas: Para remover a solução conservante foram adicionados 50 μ L de tampão B&W 1x (10 mM TRIS-HCl, pH 7.5; 1 mM EDTA; 0.1% Tween-20 e 2 M NaCl) em 50 μ L de esferas e colocados em suporte magnético. O sobrenadante foi removido e os passos repetidos em um total de 3 lavagens. Em seguida, para a preparação das esferas para a manipulação com RNA, as mesmas foram lavadas com 50 mL de Solução A (0,05 M de NaCl e 0,1 M de Hidróxido de Sódio). Este processo foi repetido por três vezes.

- Ligação do mRNA ao oligo dT: Foram adicionados em tubo de 1,5 mL, 50 μ L de RNA total extraído da fase de esporocisto e 1,5 μ L de oligo dT biotilado (200 pmol/ μ L). Para a ligação da sonda ao mRNA o tubo foi incubado por 15 minutos a 34°C e resfriado lentamente.

-Imobilização do mRNA ligado ao oligo dT biotilado: As esferas foram ressuspendidas em 100 mL de tampão B&W 2x e foi adicionada a solução de ligação do mRNA ao oligo dT, contendo o RNA total e o oligo dT biotilado. O tubo foi incubado por 15 minutos à temperatura ambiente em constante agitação. Após separação das esferas na parede do tubo com auxílio do suporte magnético, o sobrenadante foi removido e as esferas lavadas dez vezes com 50 mL de tampão B&W 1x. Após a última lavagem, foram adicionados 20 μ L de tampão 1X da Transcriptase Reversa (Invitrogen- Life Technologies) e a solução contendo o mRNA imobilizado nas esferas foi utilizada, sem eluição prévia do mRNA, diretamente na síntese de cDNAs. Para a reação de transcriptase reversa foi utilizado o kit *SuperScript III Reverse Transcriptase* (Invitrogen- Life Technologies).

Aos mRNAs foram adicionados 1 μ L de oligo(dT) com cauda (10 pmol), 1 μ L de dNTP (10 mM) e água estéril para um volume final de 12 μ L. As amostras foram incubadas a 65 °C por 5 minutos e então transferidas para o gelo. Foram acrescentados 1 μ L de tampão 10X *First-Strand*, 2 μ L de DTT (0,1 M), 1 μ L de *RNaseOUTTM* (40 U/ μ L), 4 μ L de Cloreto de Magnésio (25 mM), 1 μ L de *SuperScriptTM III RT* (200 U/ μ L) e as amostras incubadas a 50 °C por 50 minutos e 85 °C por 5 minutos. A seguir, foi acrescentado 1 μ L de RNase H e as amostras incubadas a 37 °C por 20 minutos. O cDNA foi armazenado a -20 °C para posterior amplificação por PCR (Figura 4).

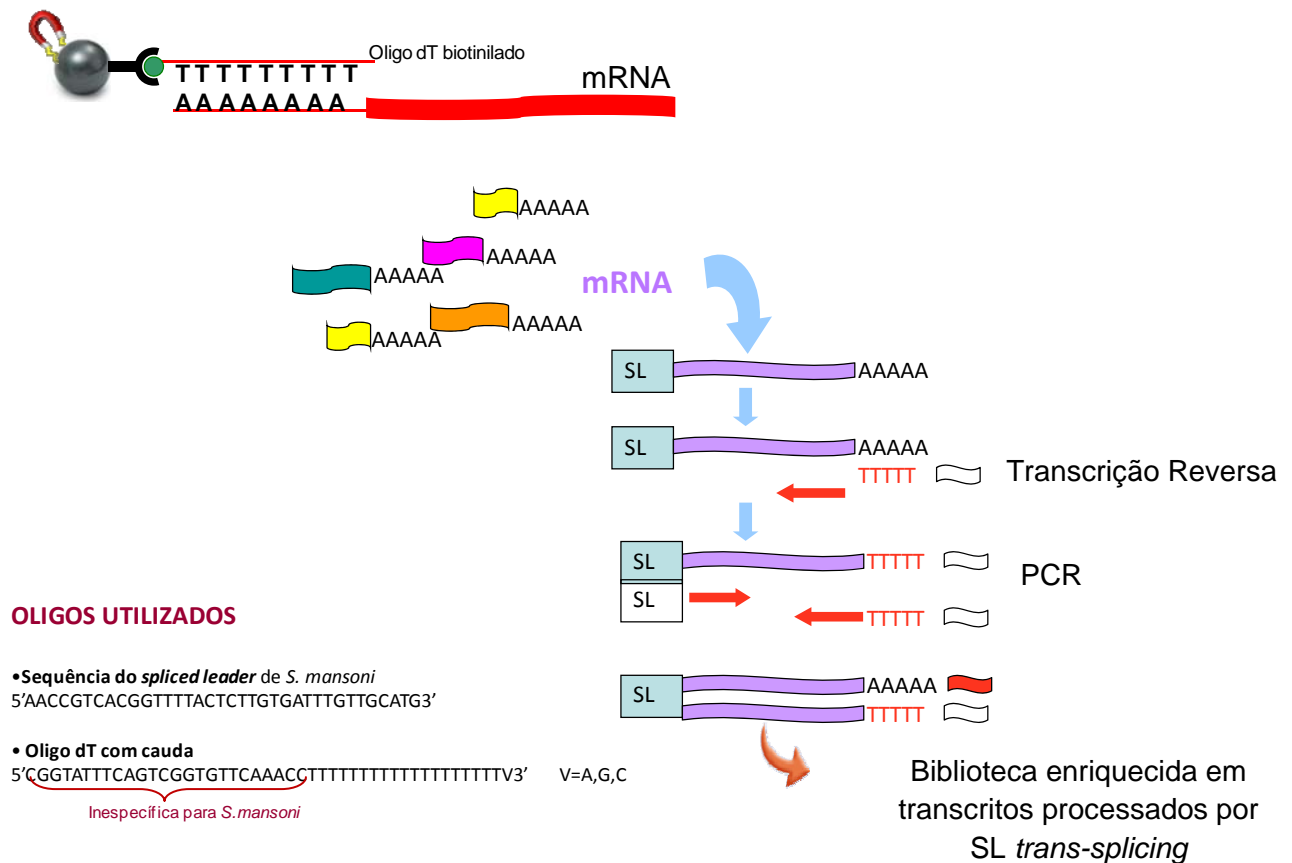


Figura 4 : Esquema da captura dos mRNAs e amplificação dos transcritos processados por SL *trans-splicing* para a construção da biblioteca de cDNA. Ligação dos mRNAs pela cauda poli-A ao oligo dT biotilado acoplado às esferas magnéticas. Em seguida, os mRNAs capturados são submetidos a síntese de cDNA e PCR com a utilização da sequência do *Spliced leader* como iniciador *forward* e de uma cauda ligada ao oligo dT como iniciador *reverse*.

4.2.3 Amplificação de transcritos processados por SL *trans-splicing*

Foi utilizada a estratégia de amplificação dos cDNAs completos processados por SL *trans-splicing*, através da utilização da sequência do SL como iniciador *forward* e de uma cauda ligada ao oligo dT, como iniciador *reverse* (BREHM *et al.*, 2000). A sequência utilizada do *Spliced leader* de *S. mansoni* foi descrita por Davis e colaboradores, 1995:

5'AACCGTCACGGTTTTACTCTTGTGATTTGTTGCATG 3'

Para o iniciador *reverse*, foi desenhado um oligo dT com uma cauda que serviu tanto para a síntese da primeira fita de cDNA, como para a reação de PCR. Esta cauda funcionaria estabilizando o iniciador e se anelaria aos produtos no segundo ciclo da PCR.

5' CGGTATTTTCAGTCGGTGTTCAAACCTTTTTTTTTTTTTTTTTTTTTTV 3' V=A,G,C

Para um volume final de 80 μL , na reação de PCR foram utilizados: 8 μL da reação de síntese de cDNA, 16 μL de tampão de reação 10X (GoTaq Flexi – Promega), 5,2 μL de cada iniciador (10 pmol), 1,6 μL de dNTP (40 mM), 4,8 μL de MgCl_2 (25 mM), 1 unidade de GoTaq DNA Polymerase (Promega) e 38,2 μL de água milli-Q autoclavada, utilizando o programa descrito a seguir:

- desnaturação a 95°C por 5 minutos.
- 5 ciclos de: desnaturação 95°C por 1 minuto, anelamento dos iniciadores a 60°C por 1 minuto, extensão a 72°C por 1 minuto e 30 segundos.
- 5 ciclos de: desnaturação 95°C por 1 minuto, anelamento dos iniciadores a 59°C por 1 minuto, extensão a 72°C por 1 minuto e 30 segundos.
- 5 ciclos de: desnaturação 95°C por 1 minuto, anelamento dos iniciadores a 58°C por 1 minuto, extensão a 72°C por 1 minuto e 30 segundos.
- 23 ciclos de: desnaturação 95°C por 1 minuto,
- anelamento dos iniciadores a 57°C por 1 minuto,
- extensão a 72°C por 1 minuto e 30 segundos.

4.3 Sequenciamento - Plataforma Ion Torrent PGM™ System

Para realizar o sequenciamento da biblioteca de cDNA enriquecida em transcritos processados por SL *trans-splicing* da fase de esporocisto foram construídas bibliotecas do tipo fragmento.

Inicialmente, 100 ng de DNA (amplificado por PCR a partir do cDNA para enriquecimento de transcritos processados por SL *trans-splicing*) foi aleatoriamente fragmentado, por ação enzimática, em fragmentos de 200-300 pb utilizando o sistema Ion

Xpress™ Plus gDNA Fragment Library Preparation. Após a etapa de purificação, para remover o tampão e o restante da enzima, as extremidades dos fragmentos de DNA foram reparadas e ligadas a adaptadores específicos fornecidos nos kits de sequenciamento. Em seguida, objetivando otimizar o sequenciamento, foi realizada uma seleção de tamanho de fragmentos de 330 pb utilizando o sistema E-Gel® SizeSelect™ Gels (Life Technologies). Posteriormente, os fragmentos selecionados foram ligados a esferas magnéticas e amplificados através de PCR em emulsão. Finalmente, as esferas magnéticas positivas foram enriquecidas e preparadas para deposição no Ion 316™ Chip Kit- Ion Torrent™ (100 Mb) (Life Technologies) e sequenciadas em equipamento Ion Torrent PGM™ System (Life Technologies) utilizando-se o Ion PGM™ Sequencing 200 Kit (Life Technologies).

4.4 Análises de bioinformática - Processamento das sequências geradas

As análises de bioinformática e processamento das sequências geradas foram realizadas em colaboração com a doutora em Bioinformática Mariana Lima Boroni Martins, Universidade Federal de Minas Gerais (UFMG).

4.4.1 Identificação da sequência do *Spliced leader*

Primeiramente, para que ocorra o correto alinhamento das sequências de DNA geradas com as sequências de DNA da 5ª versão do proteoma predito do *S. mansoni* disponível no repositório GeneDB (v5.07.08.12.chado.raw.gff), é necessária a remoção da sequência do *Spliced leader*. Assim, as sequências que continham a sequência do SL foram aparadas utilizando-se o algoritmo Fastq-mcf do pacote ea-utils (ARONESTY, 2013) com os seguintes parâmetros: -m 25 (número mínimo de pares de base (pb) a ser identificado na sequência) -p 4 (porcentagem máxima de discrepância entre as sequências: 96%, ou seja, 100-4) -l 20 (descarte das sequências de tamanho menor que 20).

4.4.2 Tratamento de qualidade das sequências

O sequenciador Ion Torrent PGM™ System (Life Technologies) gera sequências com qualidade média em torno de 20 na escala Phred, no entanto é preciso garantir a qualidade dos dados a serem analisados. Portanto, as sequências foram submetidas a um tratamento criterioso de controle de qualidade para que as bases de qualidade abaixo de Phred 20 fossem

removidas, melhorando a qualidade global dos dados. Nesta etapa foi utilizado o algoritmo PRINSEQ (SCHMIEDER; EDWARDS, 2011) com os parâmetros: `-min_len 20` (tamanho mínimo da sequência remanescente); `-min_qual_mean 18` (qualidade média mínima da sequência); `-ns_max_p 60` (porcentagem máxima nas sequências de nucleotídeos desconhecidos -N); `-trim_ns_right 5` (apara a cauda poli-N, com tamanho mínimo de 5 na extremidade 3'); `-trim_tail_right 5` (apara caudas poli-A/T com tamanho mínimo de 5 na extremidade 5'); `-trim_tail_left 5` (apara caudas poli A/T com tamanho mínimo de 5 na extremidade 3'); `-trim_qual_step 1` (tamanho do passo utilizado para mover a janela deslizante); `trim_qual-right 20` (apara a sequência pelo índice de qualidade mínima a partir da extremidade 3'); `-trim_qual_type mean` (tipo de cálculo a ser utilizado no índice de qualidade); `_trim_qual_rule lt` (regra de comparação para o índice de qualidade de valor calculado na janela deslizante); `-trim_qual_window 5` (tamanho da janela deslizante).

4.4.3 Mapeamento e contagem das sequências

Após passar pelo criterioso controle de qualidade, as sequências remanescentes foram utilizadas para alinhamento no proteoma predito de referência (5ª versão) de *S. mansoni*. Os arquivos de extensão `.fastq` foram submetidos ao programa TopHat 2.0 (TRAPNELL; PACHTER; SALZBERG, 2009). Dessa forma, foram utilizados os seguintes parâmetros: `-i 10` (tamanho mínimo do íntron); `-I 30000` (tamanho máximo do íntron); `--coverage-search` (permite a busca por junções exon-exon baseado em cobertura); `-G v5.07.08.12.chado.raw.gff` (fornece ao programa um conjunto de anotações baseado no modelo gênico disponível para *S. mansoni*, para que a primeira etapa do alinhamento ocorra no transcriptoma virtual do parasito) o arquivo utilizado foi obtido em `ftp://ftp.sanger.ac.uk/pub/pathogens/Schistosoma/mansoni/genome/Gene_models/`; `-j combined.juncs` (alimenta o programa com um conjunto de junções exon-exon conhecidas); `--read-mismatches 4` (número máximo de *mismatches* permitido nas sequências alinhadas); `--read-gap-length 4` (número máximo de *gaps* permitido nas sequências alinhadas); `--read-edit-dist 4` (alinhamentos de sequências com mais do que esta distância de edição são descartados); `-a 5` (comprimento “âncora” para encontrar junções exon-exon); `-m 1` (número máximo de *mismatches* que podem aparecer na região de “âncora” de um alinhamento que ocorreu numa junção exon-exon); `--max-insertion-length 5` (tamanho máximo de uma inserção) e `--max-deletion-length 5` (tamanho máximo de uma deleção).

Os arquivos unmmapped.bam contendo as sequências não mapeadas nessa primeira etapa foram convertidos para o formato .fastq utilizando-se a ferramenta SamToFast.jar do pacote Picard (Li *et al.*, 2009). As sequências não mapeadas foram então submetidas a um segundo alinhamento utilizando o alinhador Bowtie2 (LANGMEAD; SALZBERG, 2012), com os parâmetros: --local (neste modo, Bowtie2 executa alinhamento local das sequências, caso isso maximize a pontuação de alinhamento da mesma) e --very-sensitive-local (modo projetado para que o programa seja executado de forma mais veloz, sensível e preciso). O arquivo de saída no formato .sam foi convertido para o formato.bam e fundido com o arquivo de mapeamento obtido na primeira etapa com o programa TopHat 2.0 (TRAPNELL; PACHTER; SALZBERG, 2009), utilizando-se o pacote Samtools (LI *et al.*, 2009).

As métricas dos alinhamentos foram coletadas utilizando-se o programa CollectAlignmentMetrics do pacote de ferramentas Picard (LI *et al.*, 2009). O alinhamento foi verificado também visualmente utilizando o visualizador IGV 2.1 (Broad Institute, obtido em <http://www.broadinstitute.org/software/igv/download/>).

As contagens brutas do número de sequências alinhadas para cada gene presente no modelo gênico de *S. mansoni* foram obtidas com o pacote HTSeq Python versão 0.5.3p3 (disponível em <http://www-huber.embl.de/users/anders/HTSeq/doc/index.html>), utilizando as opções “-stranded=no” (A *read* é considerada alinhada independente do fato se ser mapeada na mesma ou na fita oposta) e “-mode=intersection-strict” (A *read* deve ser completamente alinhada em um determinado gene. Se uma *read* sobrepõe vários genes, mas alinha completamente em apenas um dos genes, a *read* é considerada. As *reads* alinhadas em mais de um gene foram descartadas). Para as análises posteriores, foram aceitos somente os genes encontrados em ambas as replicatas biológicas e as isoformas não foram incluídas nos dados.

4.4.4 Anotação e análise funcional das sequências

Para anotação funcional dos transcritos, foram utilizadas metodologias desenvolvidas pelo grupo do Dr. José Marcos C. Ribeiro (National Institute of Allergy and Infectious Diseases), que possui ampla experiência na análise de dados e desenvolvimento de ferramentas para anotação de sequências geradas em experimentos de transcriptômica pelas novas plataformas de sequenciamento.

Todas as sequências de nucleotídeos referentes aos genes contidos no modelo gênico de *S. mansoni* (5ª versão) foram utilizadas para realizar buscas por similaridade em diversos

bancos de dados (Tabela 2) utilizando os programas BLAST (ALTSCHUL *et al.*, 1997): BLASTX, BLASTN OU RPS-BLAST, conforme realizado por Karim e colaboradores (2011).

Os diferentes tipos de análise por BLAST, as bases de dados e os parâmetros utilizados estão listados na Tabela 2. As sequências de aminoácidos preditas para os genes identificados foram submetidas ao SignalP (NIELSEN *et al.*, 1997) para identificar os produtos de tradução que poderiam ser secretados, ao TMHMM (SONNHAMMER; VON HEIJNE; KROGH, 1998) para detectar as hélices transmembrana, ao NetOglyc para detectar possíveis O-glicosilações do tipo mucina (HANSEN *et al.*, 1998) e ao servidor ProP (DUCKERT; BRUNAK; BLOM, 2004) para identificar possíveis sítios de clivagem (contendo Arg e Lys) para furinas, que são enzimas conhecidas por converter pré-proteínas em seus produtos biologicamente ativos.

Os resultados foram tabulados em uma planilha Excel com *hyperlinks* que permitiram a anotação das sequências de forma automatizada, assim como, sua classificação funcional, utilizando o programa Classifier, escrito em Visual Basic pelo Dr. Ribeiro. Este programa de classificação funcional considera palavras-chave dos *macthes* de todos os resultados de BLAST, assim como, os *e-values*, os resultados para SignalP, domínios transmembranas e glicosilação para classificar os transcritos em 27 categorias funcionais. O programa Class Table Maker (KARIM; SINGH; RIBEIRO, 2011), também escrito pelo Dr. Ribeiro, foi utilizado para calcular a frequência dos transcritos em cada classe funcional, utilizando os valores de contagens obtidos. Para verificar se existe o enriquecimento de determinada classe nos transcritos que são processados por SL *trans-splicing*, foi realizado um teste estatístico χ^2 ($p < 0,05$) para averiguar se a diferença de frequência encontrada em cada classe é estatisticamente significativa entre o grupo de genes expressos no proteoma predito de *S. mansoni* (5ª versão) e os genes processados por SL *trans-splicing* na fase de esporocisto do parasito.

TABELA 2: TIPOS DE ANÁLISE POR BLAST, BASES DE DADOS E PARÂMETROS UTILIZADOS PARA ANOTAÇÃO DOS TRANSCRITOS DE ESPOROCISTOS DE *SCHISTOSOMA MANSONI*

| TIPO | DB | PARÂMETROS |
|--------|--|---------------------------------|
| BLASTN | Mit-pla (GenBank) | -IT -JT -v10 -b3 - e1e -10 -FF |
| | rRNA (GenBank) | -IT -JT -v10 -b3 - e1e -10 -FF |
| | Rfam (Sanger) (GARDNER <i>et al.</i> ,2011) | -IT -JT -v10 -b3 - e1e -10 -FF |
| BLASTX | NR (GenBank) | -IT -JT -v10 -b3 - e100 -FF -CF |
| | NR-Acelomata (Subset do GenBank) | -IT -JT -v10 -b3 - e100 -FF -CF |
| | Swiss-prot (UniProtKB) | -IT -JT -v10 -b3 - e100 -FF -CF |
| | WormBase (HARRIS <i>et al.</i> , 2010) | -IT -JT -v10 -b3 - e100 -FF -CF |
| | Gene Ontology – GO (LEWIS; ASHBURNER; REESE, 2000) | -IT -JT -v10 -b3 - e100 -FF -CF |
| | KEEG Orthology (KANEHISA; GOTO, 2000) | -v1 - b1 - e1e -5 -FF |
| | RPS-BLAST | |
| # | COG (TATUSOV <i>et al.</i> , 2003) | -IT -JT -v10 -b10 - e10 -FF -pF |
| # | Pfam (PUNTA <i>et al.</i> , 2012) | -IT -JT -v10 -b10 - e10 -FF -pF |
| # | CDD (MARCHLER- BAUER, 2002) | -IT -JT -v10 -b10 - e10 -FF -pF |
| # | SMART (SCHULTZ <i>et al.</i> , 2000) | -IT -JT -v10 -b10 - e10 -FF -pF |
| # | TIGRFAMS (J. Craig Venter Institute) | -IT -JT -v10 -b10 - e10 -FF -pF |
| # | PRK | |
| # | TE-CLASS | -IT -JT -v10 -b10 - e15 -FF -pF |

4.4.5 Identificação das vias metabólicas em que atuam as proteínas codificadas pelos transcritos processados por *Spliced leader trans-splicing*

Para identificar as vias metabólicas em que atuam as proteínas codificadas pelos transcritos processados por SL *trans-splicing* foi construído um mapa metabólico através da ferramenta online iPath2.0 (*Interactive Pathways Explorer*) (YAMADA *et al.*, 2011). O iPath2.0 permite que os usuários naveguem e explorem facilmente as complexas vias do mapa metabólico (<http://pathways.embl.de/>). O mapa gerado fornece um conjunto de vias metabólicas específicas para cada espécie disponível e utiliza enzimas para a construção dessas vias. A construção do mapa foi baseada no banco de dados KEGG (Kyoto Encyclopedia of Genes and Genomes) (KANEHISA, M; GOTO, 2000) (KANEHISA *et al.*, 2012), tendo como referência a espécie *S. mansoni* e destaca as vias associadas a proteínas codificadas pelos transcritos processados por SL *trans-splicing*.

4.4.6 Análise de abundância dos transcritos anotados

Para verificar a abundância dos transcritos anotados foi realizado o método de RPKM (*Reads Per Kilobase of exon model Per Million Mapped Reads*). Esse é utilizado para normalizar as *reads* de cada amostra considerando o tamanho do transcrito, o tamanho da biblioteca (amostra) e a quantidade de *reads* do transcrito em questão (MORTAZAVI *et al.*, 2008b).

$$RPKM = \frac{10^9 * C}{N * L}$$

Na equação do RPKM, o valor 10^9 representa o resultado de 1 mil pares de bases multiplicado por 1 milhão de *reads*, enquanto C representa o total de *reads* do transcrito analisado, N o total de *reads* da amostra analisada e L o tamanho do transcrito. Ao final da equação, o valor é representado como abundância do transcrito anotado.

5 RESULTADOS

5.1 Análise do RNA total extraído

Para a construção das bibliotecas de cDNA enriquecidas em transcritos processados por SL *trans-splicing* foram realizadas extrações de RNA total da fase de esporocisto. A qualidade destas amostras de RNA foi verificada no equipamento Agilent 2100 Bioanalyzer, através de eletroforese capilar microfluídica (tecnologia *Lab-on-a-chip* da *Agilent*) (Figura 5)

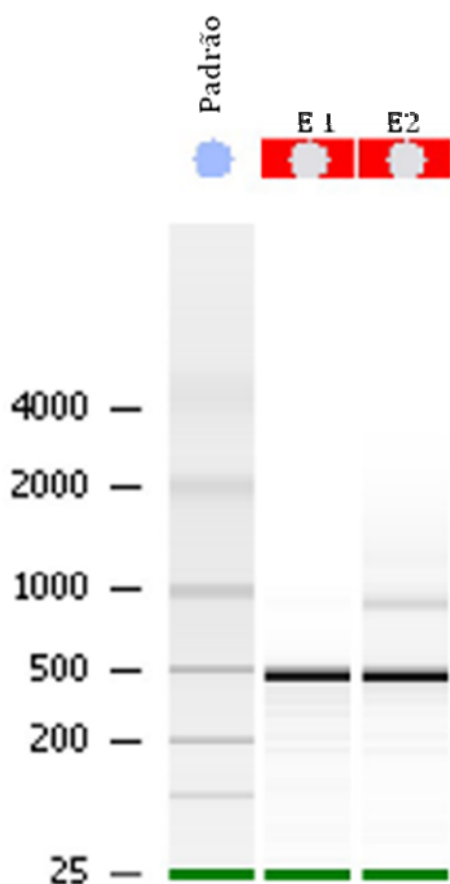


Figura 5: Perfil eletroforético do RNA total extraído da fase de esporocisto do parasito *Schistosoma mansoni*. Gel virtual das duas amostras de RNA total extraídas de esporocistos (amostra E1 e amostra E2) gerado pelo Agilent 2100 Bioanalyzer. Padrão de peso molecular utilizado para amostras de RNA.

Podemos observar na Figura 4, a presença somente de uma banda bem definida, de aproximadamente 500pb, correspondente ao RNA ribossomal 18S nas amostras de RNA total extraídas de esporocistos de *S. mansoni*. Não foi observada a presença de arraste.

5.2 Amplificação dos transcritos processados por *Spliced leader trans-splicing*

Após a síntese de cDNA a partir do mRNA extraído de esporocistos de *S. mansoni*, parte da reação de cDNA foi utilizada como molde em reações de PCR para o enriquecimento dos transcritos processados por SL trans-splicing utilizando parte da sequência do SL como iniciador *forward* e de uma cauda ligada ao oligo dT como iniciador *reverse*. Os produtos de amplificação foram analisados em gel de poliacrilamida 8% (Figura 6).

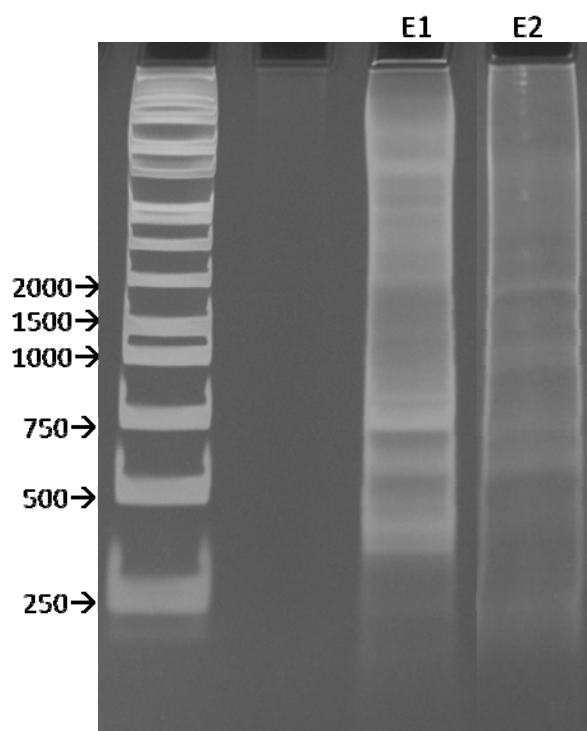


Figura6: Amplificação dos transcritos processados por SL *trans-splicing*. Gel de poliacrilamida 8% corado pela prata. Amostras E1 e E2: fragmentos gerados na PCR, enriquecidos em transcritos que são processados por SL *trans-splicing*. Padrão de peso molecular 1Kb.

O perfil do produto de PCR visualizado no gel de poliacrilamida 8% apresenta a amplificação dos transcritos processados por SL *trans-splicing*, com fragmentos de 100pb a 10.000pb.

5.3 Obtenção das sequências na Plataforma Ion Torrent PGM™ System

Produtos de PCR enriquecidos de sequências de transcritos processados por SL *trans-splicing* de esporocistos foram sequenciados na plataforma Ion Torrent PGM™ System.

Foram geradas duas bibliotecas do tipo *single* referente às réplicas biológicas E1 e E2 da fase de esporocisto. Foram obtidas 3.479.656 *reads* na biblioteca 1 e 3.127.173 *reads* na biblioteca 2, variando em tamanhos de 8-372 pb e 8-376 pb. O conteúdo de GC em ambas as bibliotecas foi de 36%. Os resultados obtidos nesta plataforma estão sendo mostrados na Tabela 3.

TABELA 3: DADOS GERADOS NA PLATAFORMA ION TORRENT PGM™ SYSTEM

| Réplica | Tipo de Biblioteca | Tamanho das <i>reads</i> | %GC | Número de <i>reads</i> |
|---------|--------------------|--------------------------|-----|------------------------|
| E1 | * <i>Single</i> | 8-372 pb | 36 | 3.479.656 |
| E2 | <i>Single</i> | 8-376 pb | 36 | 3.127.173 |

**Single*: o sequenciador faz a leitura dos fragmentos de apenas um lado para o outro.

5.4 Identificação e remoção da sequência *leader* e tratamento de qualidade das *reads* geradas

Através do programa FastQC foi verificada a qualidade das *reads* e os dados gerados podem ser visualizados nos gráficos a seguir:

A figura 7 mostra uma visão geral dos valores de qualidade das bases em cada posição nos arquivos fastq. gerados após o sequenciamento das bibliotecas. Assim, nota-se que a média de qualidade por base gerada para as duas bibliotecas foi em torno de 29 (Phred) e, apesar de apresentar uma queda nos valores de qualidade na porção 3' terminal das sequências, a mesma permanece estável ao longo dos ciclos.

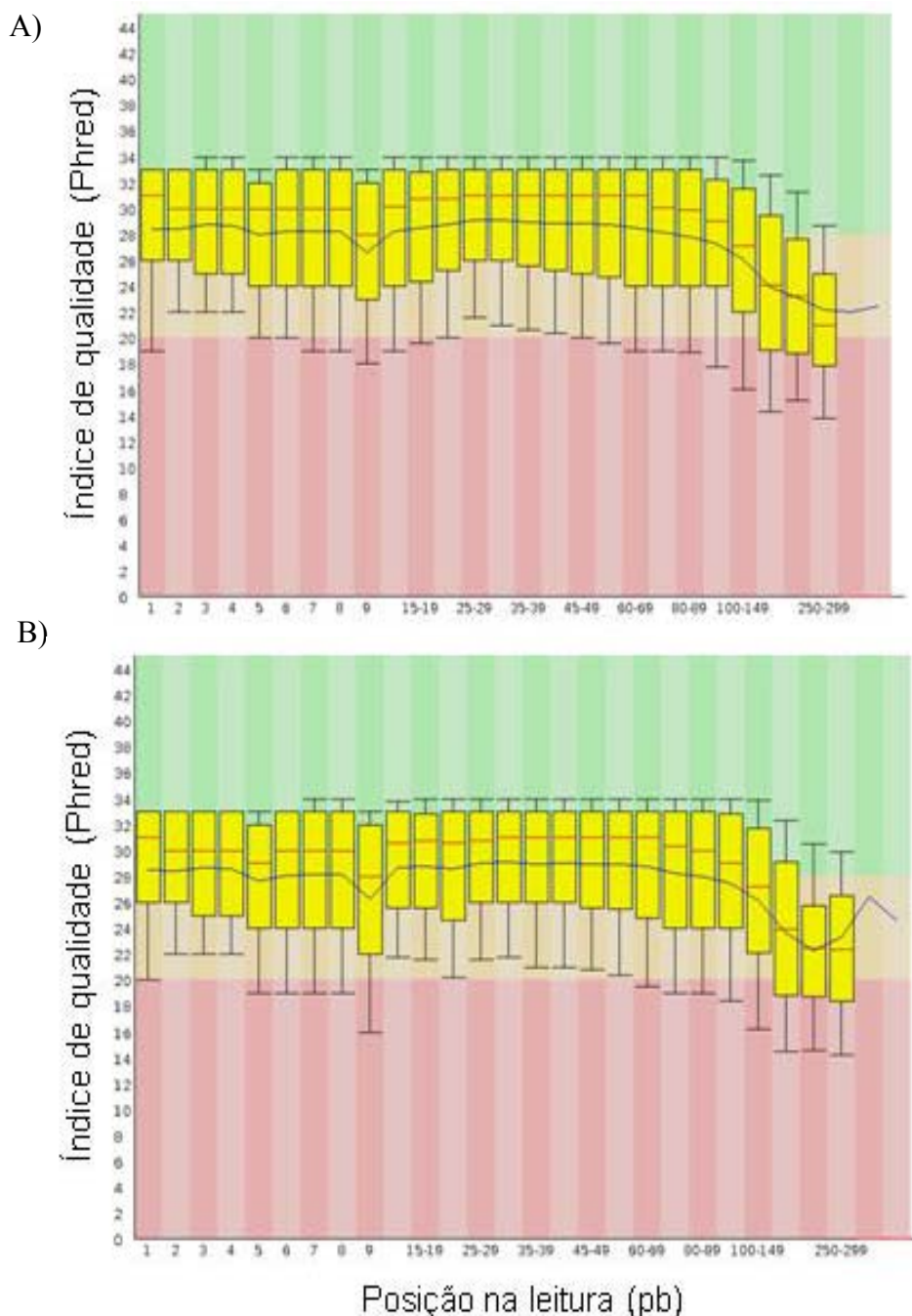


Figura 7: Valor de qualidade por base. Diagrama de extremos e quartis. Para cada posição, a caixa amarela representa a amplitude inter-quartil (25-75%) e as linhas verticais que dela partem unem-na aos extremos inferior e superior. A linha vermelha representa a mediana e a

linha preta representa a qualidade média. Os diagramas representam os dados relativos às réplicas E1 (A) e E2 (B) gerados na plataforma Ion Torrent PGM™ System.

A análise do valor de qualidade média por sequência (Figura 8) permite verificar se um subconjunto das sequências tem valor de qualidade muito baixo, levando-se em consideração a média de qualidade por base de cada sequência. Verifica-se uma distribuição unimodal do valor de qualidade para as duas bibliotecas geradas, sendo o valor médio de Phred em torno de 27.

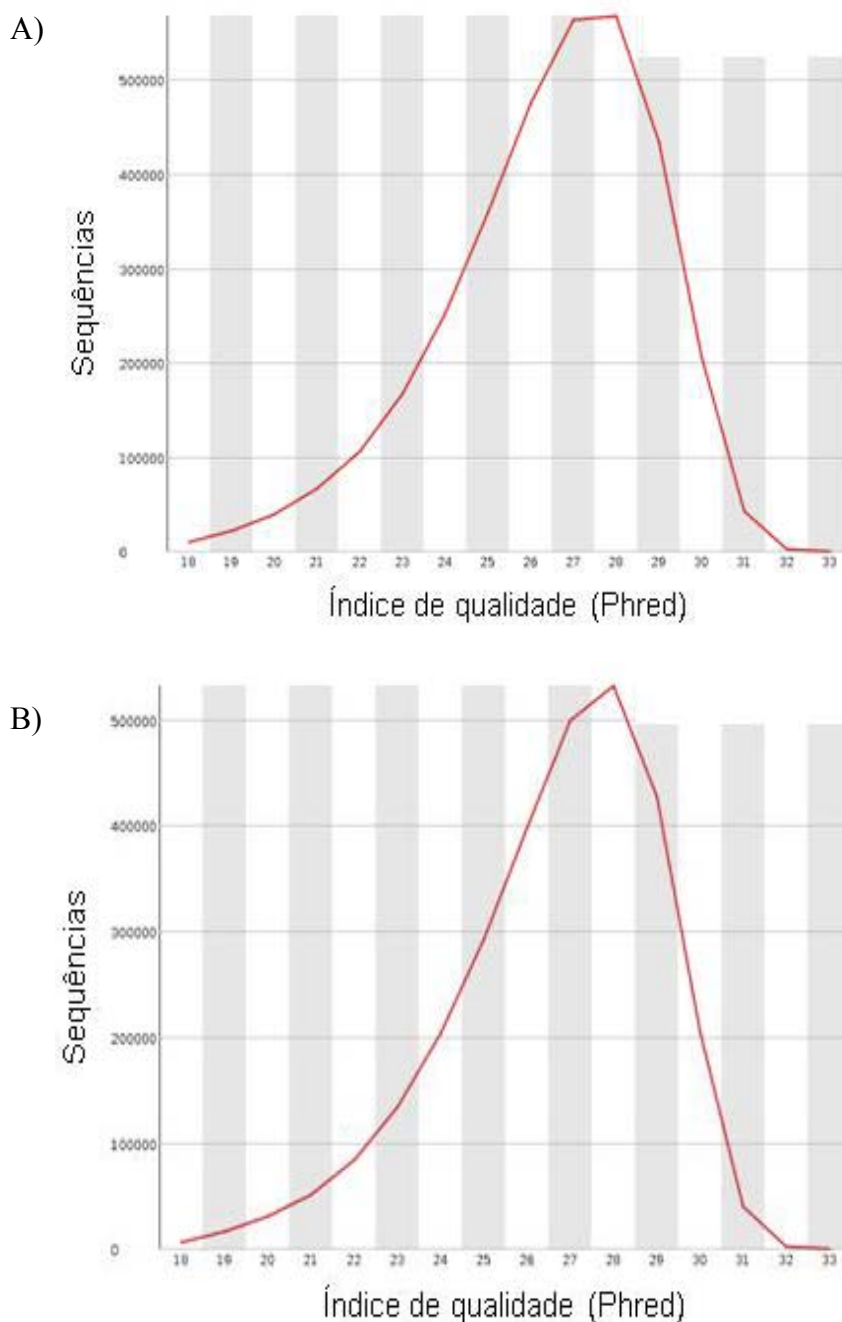
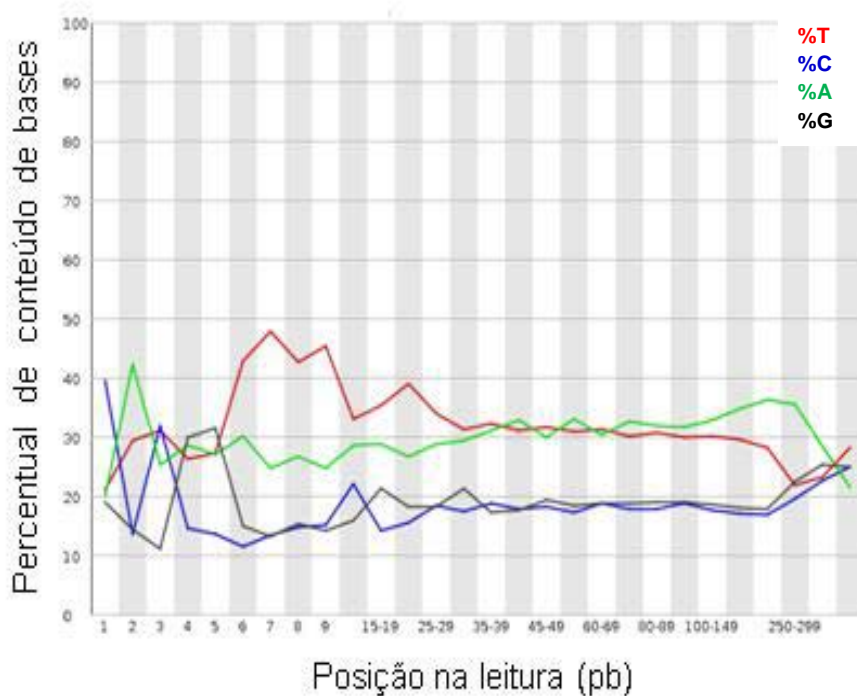


Figura 8: Valor de qualidade média por sequência. O gráfico permite avaliar a qualidade universal de vários subconjuntos de sequências. Os diagramas representam os dados relativos às réplicas E1 (A) e E2 (B) gerados na plataforma Ion Torrent PGM™ System.

A partir do conteúdo de bases nitrogenadas (Figura 9) verifica-se que nas duas bibliotecas geradas existe uma pequena oscilação para cada base nitrogenada ao longo da sequência e um moderado pico da base Timina no início da posição das leituras. Entretanto, observa-se uma tendência em manter o conteúdo das bases G, C, A e T constante ao longo das leituras.

A)



B)

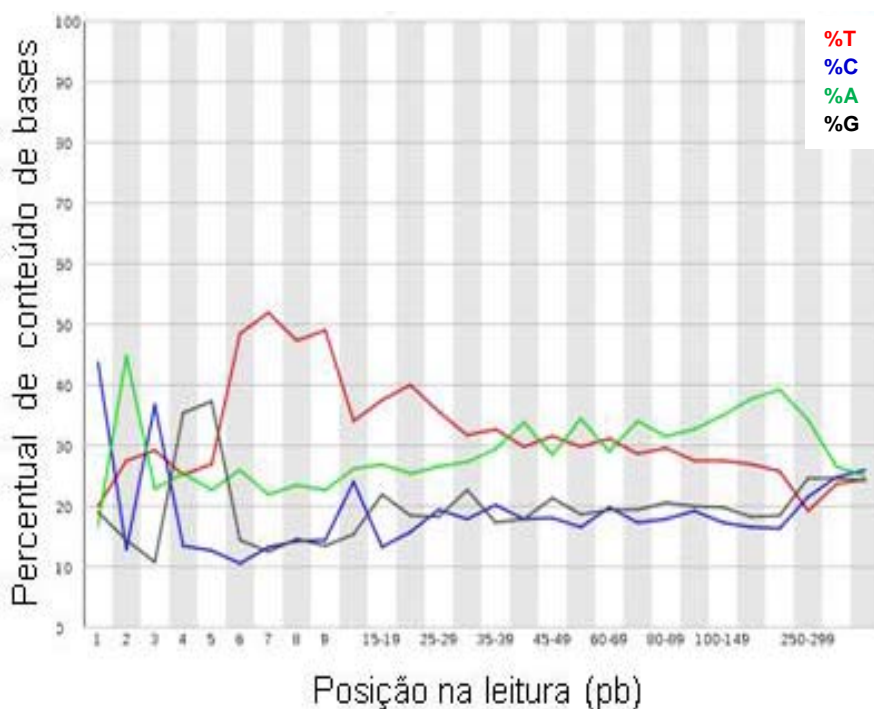


Figura 9: Conteúdo de bases nitrogenadas ao longo da sequência. O gráfico retrata as frequências médias (eixo Y) de A,T,C e G ao longo das leituras (eixo X). Os diagramas representam os dados relativos às réplicas E1 (A) e E2 (B) gerados na plataforma Ion Torrent PGM™ System.

O modelo de distribuição do conteúdo de GC por sequência é uma curva normal onde o valor central corresponde ao teor global médio de GC do genoma. Assim, nota-se que o conteúdo médio de GC corresponde a média central de 36% nas duas bibliotecas geradas (Figura 10).

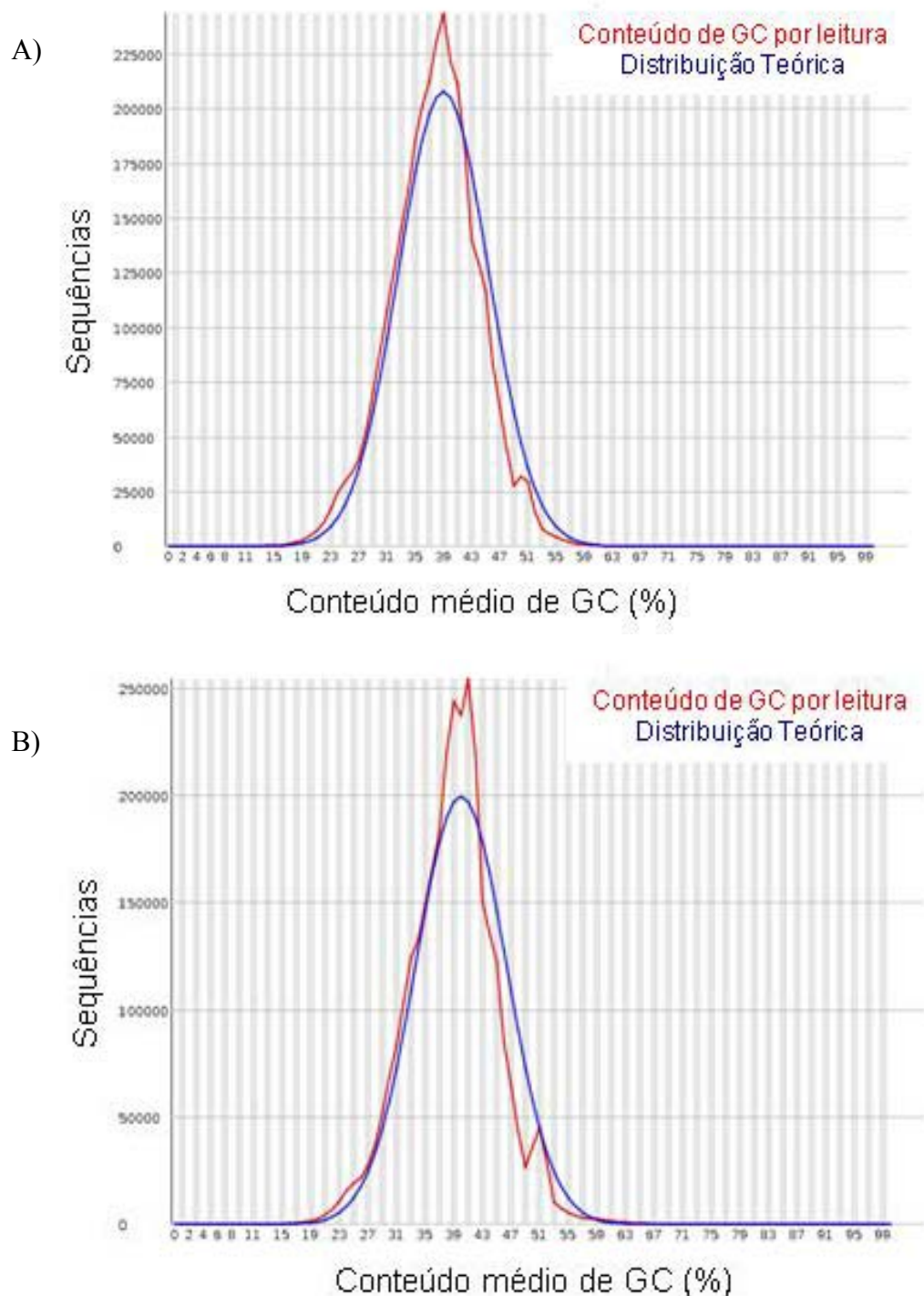


Figura 10: Conteúdo de GC por sequência. O gráfico representa a média do conteúdo de GC por sequência gerada comparado com uma distribuição normal modelada a partir do conteúdo de GC do genoma de *Schistosoma mansoni*. Os diagramas representam os dados relativos às réplicas E1 (A) e E2 (B) gerados na plataforma Ion Torrent PGM™ System.

A análise de qualidade por conteúdo de GC por base permite avaliar o conteúdo das bases G e C ao longo das leituras. Verifica-se que nas duas bibliotecas geradas há uma variação do conteúdo de GC, principalmente para as primeiras e últimas bases, porém permanece estável durante os ciclos (Figura 11).

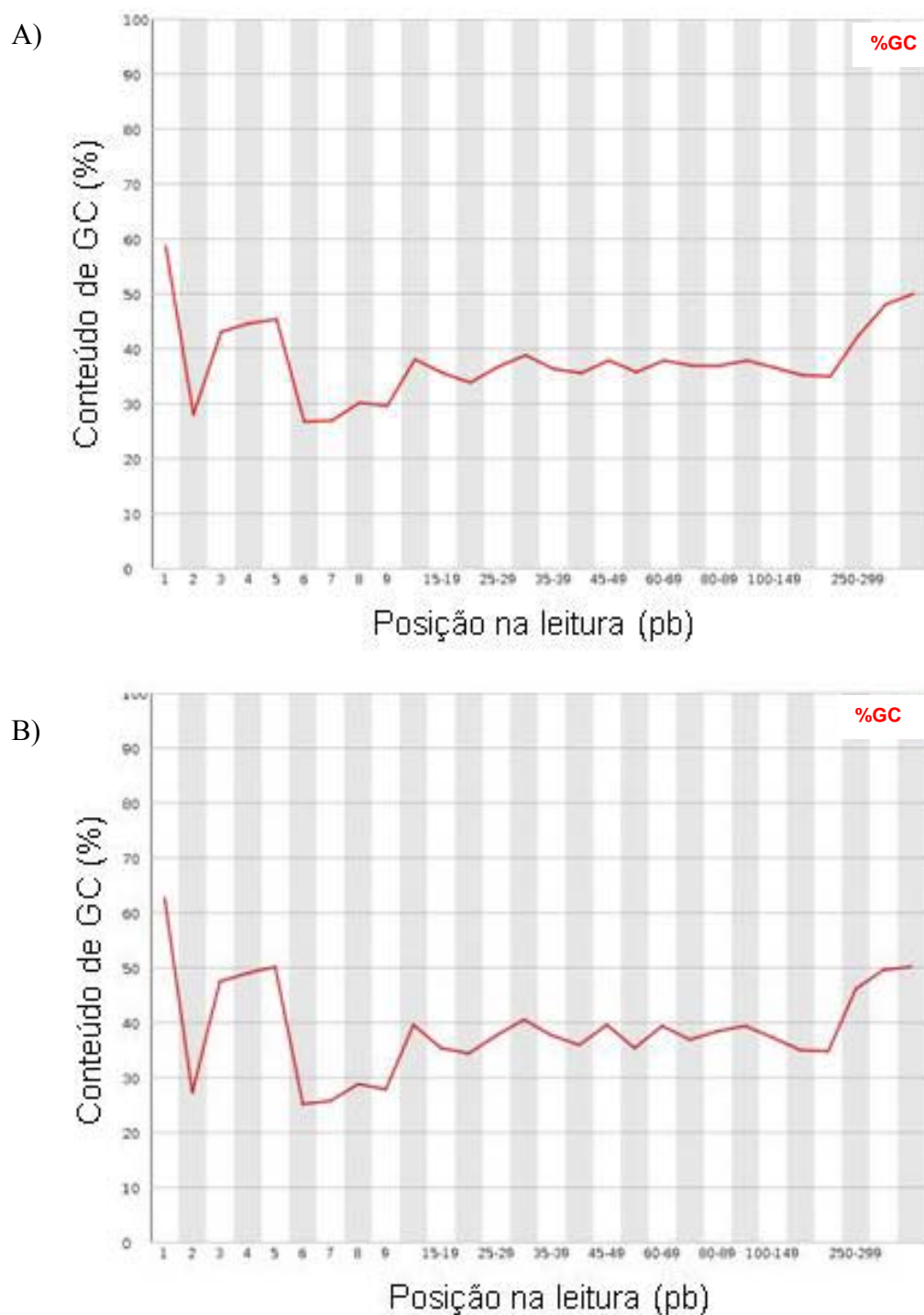


Figura 11: Conteúdo de GC por base. O gráfico indica o conteúdo de GC ao longo da *read* por base avaliada. Os diagramas representam os dados relativos às réplicas E1 (A) e E2 (B) gerados na plataforma Ion Torrent PGM™ System.

A análise de redundância das sequências é realizada em relação a um subconjunto dos dados, ou seja, utilizando 200.000 sequências que representam um conjunto de sequências únicas geradas (*unique*). O valor obtido para o nível de redundância das duas bibliotecas foi em torno de 80%. Cerca de 40% das *reads* referentes a biblioteca 1 possui um nível de duplicação acima de 10. Já para a biblioteca 2, esse percentual é menor, em torno de 20% (Figura 12).

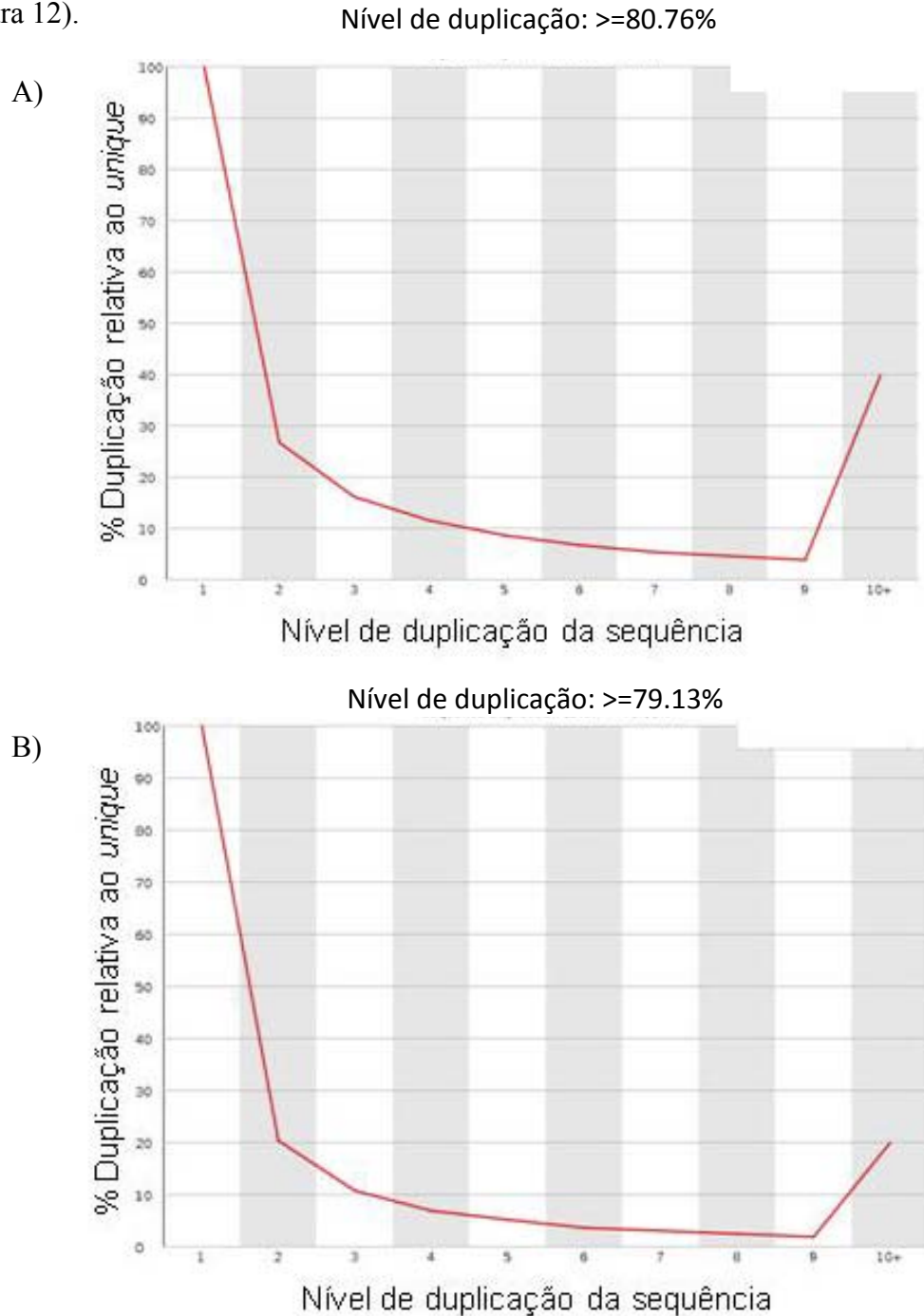


Figura 12: Sequências duplicadas. Os gráficos reportam os graus de duplicações (redundâncias) em um subconjunto de dados, mostrando o número relativo de *reads* com diferentes graus de duplicação. Os diagramas representam os dados relativos às réplicas E1 (A) e E2 (B) gerados na plataforma Ion Torrent PGM™ System.

Após a avaliação de qualidade das *reads* foi realizado um tratamento para remoção das bases de baixa qualidade, de *reads* menores do que 20pb e da sequência SL. Após este tratamento foram excluídas 16.175 e 19.787 *reads* relativas às duas bibliotecas geradas. O número de *reads* contendo a sequência *leader* foi aproximado nas duas bibliotecas geradas (em torno de 43 mil *reads*) significando 1.3% do total das *reads* geradas em cada biblioteca. As *reads* obtidas foram de boa qualidade para as duas bibliotecas (94%) e tiveram tamanho médio de 160 pb (Tabela 4).

TABELA 4: READS SEQUENCIADAS NA PLATAFORMA ION TORRENT PGM™ SYSTEM APÓS TRATAMENTO DE QUALIDADE

| Réplica | Número inicial de <i>reads</i> | Número de <i>reads</i> com SL | % de <i>reads</i> com SL | Número final de <i>reads</i> após tratamento de qualidade | % de <i>reads</i> de boa qualidade | Tamanho médio da <i>read</i> |
|---------|--------------------------------|-------------------------------|--------------------------|---|------------------------------------|------------------------------|
| E1 | 3.479.656 | 42.872 | 1.23% | 3.317.906 | 95,35% | 163,04 pb |
| E2 | 3.127.173 | 44.064 | 1.41% | 2.929.283 | 93,67% | 156,70 pb |

5.5 Mapeamento das *reads* no genoma de referência

Após a execução do primeiro alinhamento, em torno de 60% das *reads* foram mapeadas no proteoma predito (5ª versão) do *S. mansoni*. Entretanto, através do alinhador Bowtie 2 foi realizado um segundo mapeamento das *reads* não mapeadas anteriormente. Essa porcentagem de mapeamento aumentou em torno de 97% para as duas bibliotecas, com aproximadamente 7% (E1: 125.787 *reads* e E2: 89.374 *reads*) de alinhamentos múltiplos (Tabela 5).

TABELA 5: ANÁLISE DO ALINHAMENTO DAS *READS* GERADAS A PARTIR DAS BIBLIOTECAS DE ESPOROCISTO NO GENOMA DE REFERÊNCIA DE *SCHISTOSOMA MANSONI*

| Réplica | Alinhador | Categoria das reads | Número inicial de reads | Reads alinhadas | % Reads alinhadas | * Múltiplos alinhamentos | % Múltiplos alinhamentos |
|---------|-------------------|---------------------|-------------------------|-----------------|-------------------|--------------------------|--------------------------|
| E1 | TopHat2 & Bowtie2 | Single | 3.325.680 | 3.243.957 | 98% | 125.781 | 7% |
| E2 | TopHat2 & Bowtie2 | Single | 2.722.168 | 2.640.772 | 97% | 89.374 | 6% |

* Alinhamento de uma *read* em diferentes regiões no genoma de referência.

5.6 Anotação e análise funcional dos transcritos processados por *Spliced leader trans-splicing*

Após montagem dos fragmentos foram obtidas as sequências e anotados 1.191 transcritos processados por *SL trans-splicing* na fase de esporocisto. Este número corresponde a 15% do total de transcritos expressos na fase de esporocisto se comparado com dados recentes de transcriptoma desta fase (WANG, et al., 2013). Contudo, se comparado à 5ª versão do proteoma predito de *S. mansoni*, esse percentual corresponde a 10% (Figura 13).

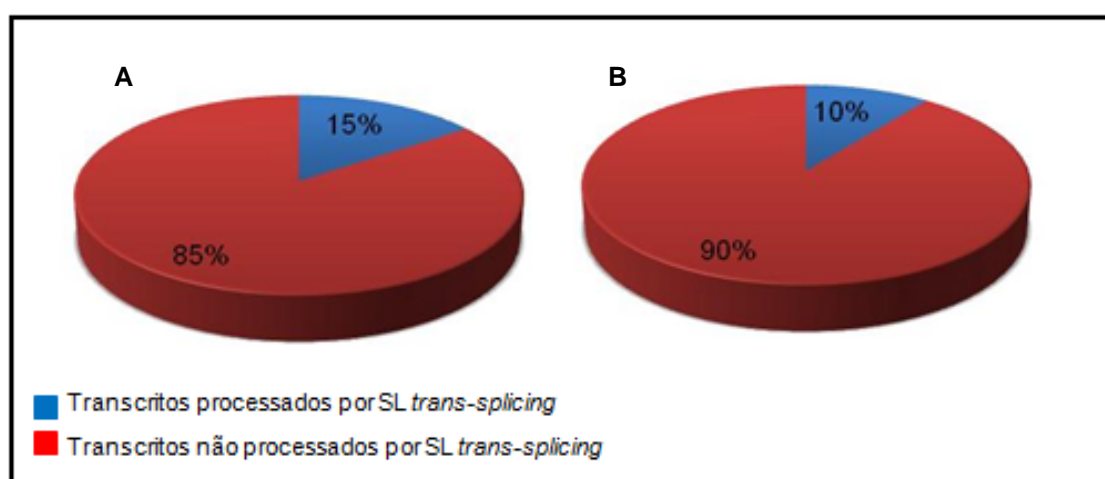


Figura 13: Percentual de genes processados por *SL trans-splicing* na fase de esporocisto. Os gráficos representam os percentuais de transcritos processados por *SL trans-splicing* na fase de esporocisto relativos ao transcriptoma de esporocisto (A) e ao proteoma predito de *S. mansoni* (B).

Após a classificação funcional, os transcritos foram agrupados em 27 categorias funcionais. Em seguida, foi realizado o teste estatístico χ^2 ($p < 0,05$) para averiguar se a diferença de frequência encontrada em cada classe é estatisticamente significativa entre os transcritos expressos no proteoma predito do *S. mansoni* e os transcritos expressos na fase de esporocisto processados por SL *trans-splicing*.

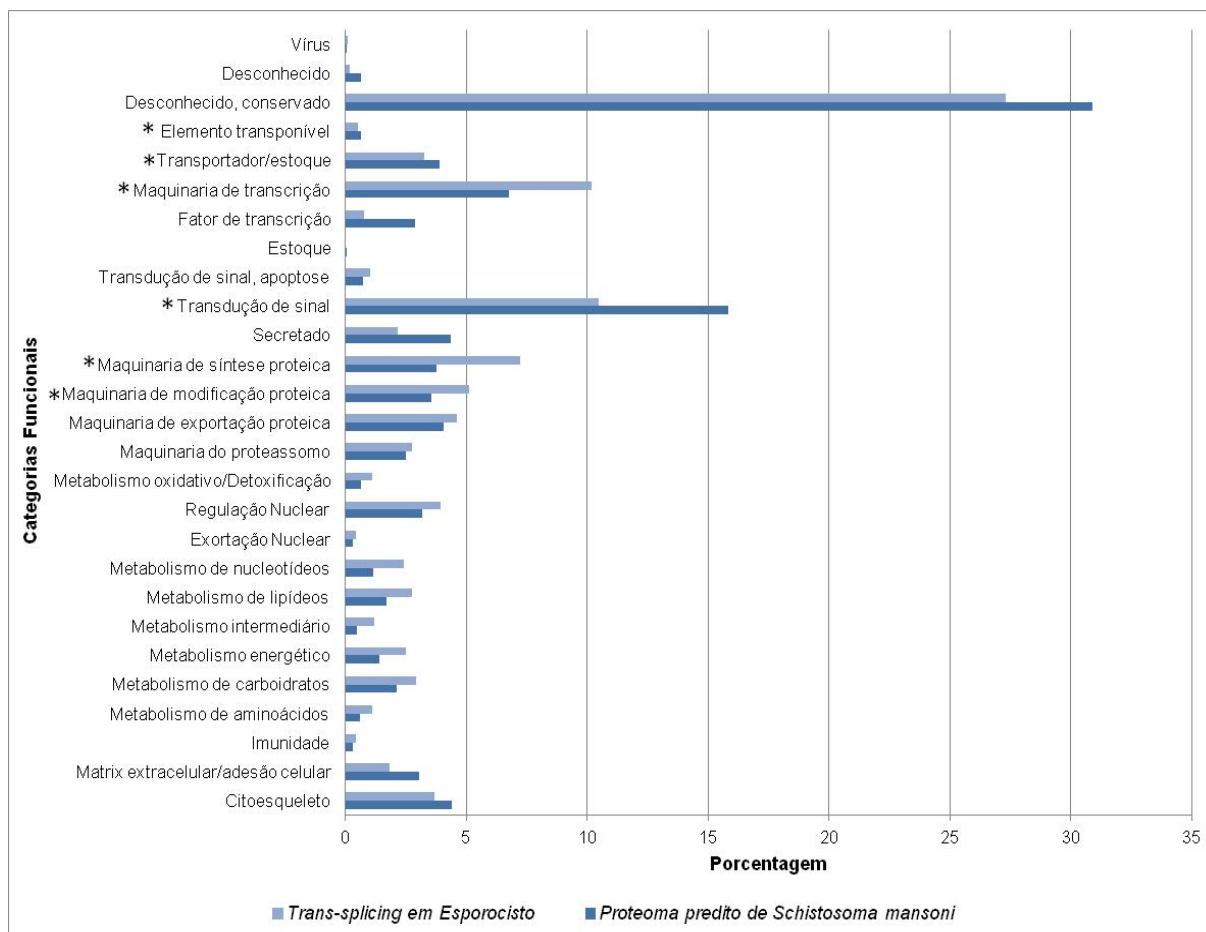


Figura 14: Percentual de transcritos identificados no proteoma predito do *Schistosoma mansoni* e dos transcritos processados por SL *trans-splicing* na fase de esporocisto classificados por categorias funcionais. As barras em azul claro representam a frequência de transcritos que são processados por SL *trans-splicing* na fase de esporocisto e em azul escuro a frequência de todos os transcritos expressos, anotados de acordo com a 5ª versão do proteoma predito de *S. mansoni*. Os asteriscos destacam as categorias funcionais estatisticamente significativas (p -value < 0.05 , teste χ^2).

Conforme apresentado na Figura 14, somente seis categorias funcionais apresentaram frequência diferencial dentre os transcritos processados por SL *trans-splicing*, são elas:

- Elemento transponível
- Transportador/estoque

- Maquinaria de transcrição
- Transdução de sinal
- Maquinaria de síntese proteica
- Maquinaria de modificação proteica

Adicionalmente, verifica-se que dentre essas categorias funcionais três estão super-representadas na fase de esporocisto:

- Maquinaria de transcrição
- Maquinaria de síntese proteica
- Maquinaria de modificação proteica

As das categorias funcionais de Elemento transponível, Transportador/estoque e Transdução de sinal apresentam significância estatística, no entanto, elas não são representativas para a fase de esporocisto, ou seja, apresentam menor frequência de transcritos (0.5%, 3.27% e 10.49%) em relação aos transcritos expressos no proteoma predito de *S. mansoni* (0.6%, 3.9% e 15.82%).

Após a categorização funcional dos transcritos, foi realizada uma análise comparando o perfil das categorias representadas dentre os transcritos processados por SL *trans-splicing* na fase de esporocisto em relação às categorias encontradas no proteoma predito de *S. mansoni*. Verifica-se similaridade de perfil nas porcentagem de transcritos em cada classe funcional entre os transcritos expressos e anotados no proteoma predito do *Schistosoma mansoni* e processados por SL *trans-splicing* na fase de esporocisto (Figura 15).

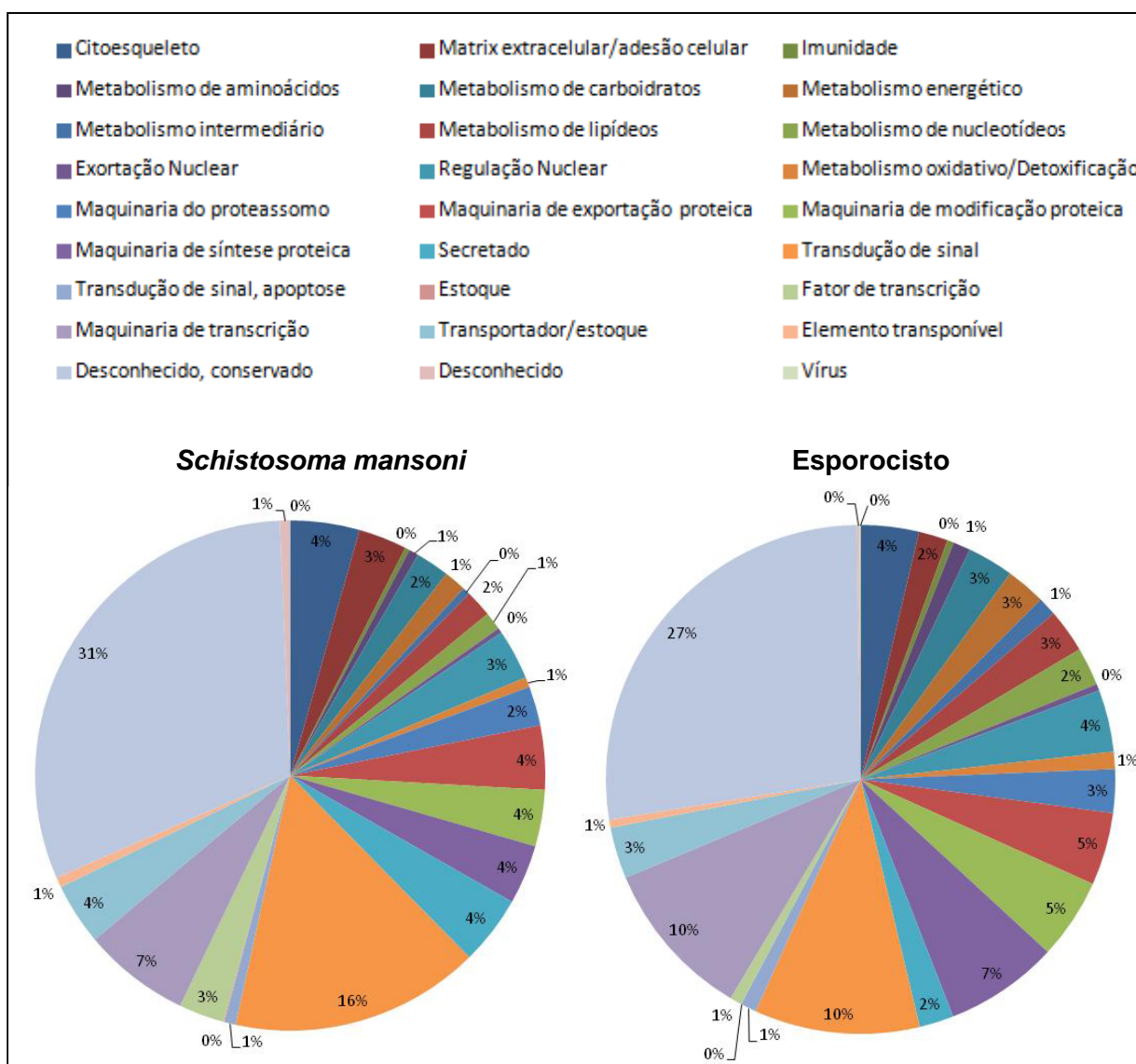


Figura 15: Análise comparativa de classes funcionais de transcritos expressos e anotados no proteoma predito do *Schistosoma mansoni* e processados por SL *trans-splicing* na fase de esporocisto. Os gráficos representam os percentuais de transcritos correspondentes a cada classe funcional, relativos ao proteoma predito de *S. mansoni* e aos transcritos processados por SL *trans-splicing* expressos na fase de esporocisto.

5.7 Identificação das vias metabólicas onde atuam enzimas codificadas pelos transcritos processados por SL *trans-splicing* na fase de esporocisto

Através dos dados disponíveis no banco de dados KEGG (KANEHISA; GOTO, 2000) foi possível associar as enzimas codificadas pelos transcritos processados por SL *trans-splicing* da fase de esporocisto às vias metabólicas do parasito *S. mansoni* e elaborar um mapa

metabólico utilizando o programa iPath2.0 (YAMADA *et al.*, 2011). O mapa metabólico específico de *S. mansoni* é constituído por várias vias pertencentes ao Metabolismo de Carboidrato, Metabolismo de Aminoácido, Metabolismo Energético, Metabolismo de Lipídeos, Metabolismo e Biosíntese de Glicanos, Metabolismo de outros Aminoácidos, Biosíntese de outros Metabólitos Secundários, Metabolismo de Nucleotídeos e Metabolismo de Co-fatores e Vitaminas. Como representado na Figura 16, o mecanismo de SL *trans-splicing* em esporocisto possui transcritos pertencentes às todas as vias metabólicas do parasito *S. mansoni*.

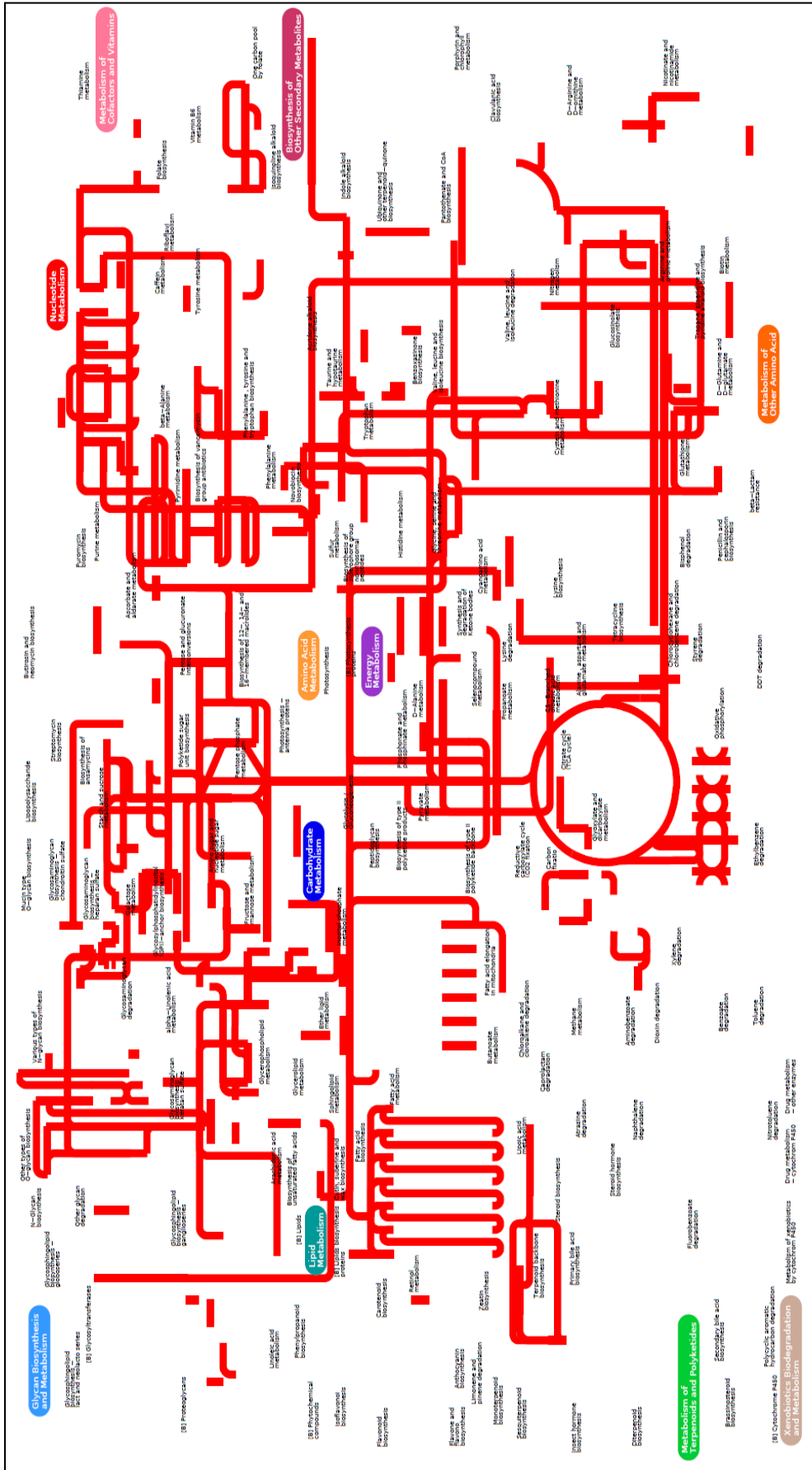


Figura 16: Vias metabólicas nas quais enzimas derivadas de transcritos processados por *Spliced leader trans-splicing* estão presentes na fase de esporocisto. Este mapa metabólico foi construído utilizando o programa iPath. As linhas destacadas em vermelho representam as vias metabólicas cujas proteínas participantes são derivadas de transcritos que são processados por SL *trans-splicing* na fase de esporocisto.

5.8 Análise de abundância dos transcritos anotados

A partir da contagem do número bruto de *reads* alinhadas para cada transcrito presente no modelo gênico de *S. mansoni* e da média do número de *reads* das duas bibliotecas de esporocisto foi possível calcular a abundância dos transcritos processados por SL *trans-splicing* utilizando o método de RPKM (*Reads Per Kilobase of exon model Per Million Mapped Reads*) (MORTAZAVI *et al.*, 2008b). A Figura 17 representa os 20 transcritos mais abundantes na fase de esporocisto processados por SL *trans-splicing* expressos em RPKM.

Dentre os 20 transcritos selecionados é possível observar um elevado nível de abundância dos sete primeiros transcritos em relação ao restante dos transcritos, que apresentaram valores de RPKM semelhantes. O transcrito mais abundante possui valor de abundância de 400359,4 RPKM, já o menos abundante, o valor de 12718,3 RPKM.

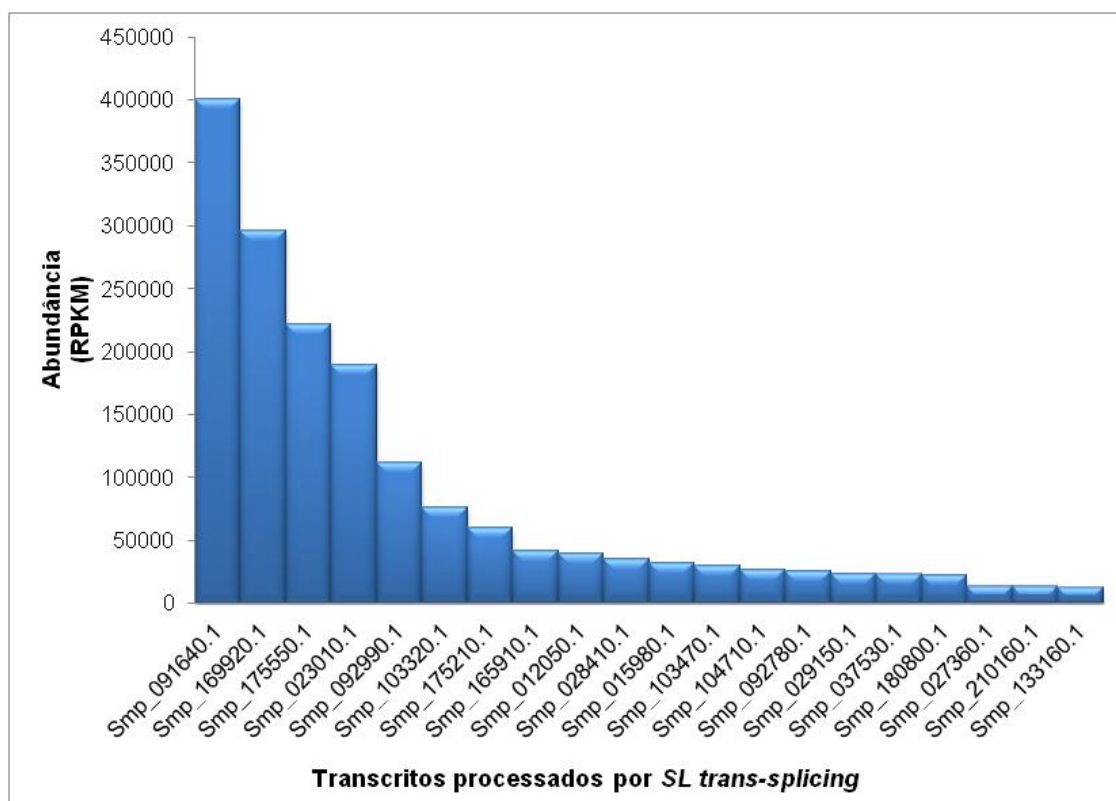


Figura 17: Análise de abundância dos transcritos processados por SL *trans-splicing*. As barras representam a abundância dos 20 transcritos processados por SL *trans-splicing* expressos na fase de esporocisto utilizando o método de RPKM.

TABELA 6: ANOTAÇÃO DOS 20 TRANSCRITOS MAIS ABUNDANTES NA FASE DE ESPOROCISTO PROCESSADOS POR SL *TRANS-SPLICING*

| Transcrito | Identificação |
|---|----------------------|
| 40S ribosomal protein S24 <i>Schistosoma mansoni</i> | Smp_091640 |
| Translation machinery-associated protein <i>Schistosoma mansoni</i> | Smp_169920 |
| Nuclear movement protein nudc, putative | Smp_103320 |
| Small Nuclear ribonucleoprotein splicing factor | Smp_175550 |
| Atpase inhibitor <i>Schistosoma mansoni</i> | Smp_023010 |
| Hypothetical protein <i>Schistosoma mansoni</i> | Smp_092990 |
| Hypothetical protein | Smp_092780 |
| Peptidase <i>Schistosoma mansoni</i> | Smp_165910 |
| Enhancer of rudimentary protein <i>Schistosoma mansoni</i> | Smp_175210 |
| Cytochrome C oxidase copper chaperone <i>Schistosoma mansoni</i> | Smp_029150 |
| Exon-exon junction complex Magoh component | Smp_103470 |
| Hypothetical protein <i>Schistosoma mansoni</i> | Smp_104710 |
| BolA (bacterial stress-induced morphogen)-related protein | Smp_012050 |
| Hypothetical protein | Smp_173410 |
| Hypothetical protein | Smp_015980 |
| Small nuclear ribonucleoprotein U)1a,U)2b, putative | Smp_069880 |
| mRNA splicing factor ATP-dependent RNA helicase | Smp_156060 |
| Transmembrane protein 167A | Smp_028410 |
| Thioredoxin-like protein | Smp_037530 |
| Hypothetical protein <i>Schistosoma mansoni</i> | Smp_180800 |

6 DISCUSSÃO

Após 30 anos de sua descoberta, o mecanismo de SL *trans-splicing* continua sendo um enigma para a biologia (LASDA; BLUMENTHAL, 2011). Em alguns organismos, determinadas funções deste mecanismo já estão bem estabelecidas, porém ainda pouco se sabe sobre a origem filogenética da sequência do SL nos diversos filos, sobre as funções atribuídas ao mecanismo de SL *trans-splicing*, bem como sobre as consequências deste mecanismo na expressão de genes no *S. mansoni*.

Desde o início do projeto de descoberta gênica em *S. mansoni*, em 1992, há um aumento crescente de estudos envolvendo este parasito, que particularmente, possui um ciclo de vida complexo, representando uma notável interação adaptativa entre o parasito e seus hospedeiros com o ambiente natural onde o ciclo ocorre (FRANCO *et al.*, 1995). Deste modo, com o uso crescente de tecnologias de sequenciamento e de novas metodologias para estudos de genômica e transcriptômica, várias funções gênicas e processos transcricionais como o mecanismo de regulação pós-transcricional SL *trans-splicing*, podem ser detalhadamente estudados.

Através do uso da metodologia de RNA-seq foi possível realizar este estudo envolvendo a identificação de transcritos processados por SL *trans-splicing* na fase de esporocisto do parasito *S. mansoni* visando enriquecer o conhecimento ainda insipiente sobre o papel desse mecanismo no processo de expressão gênica do parasito. Os transcritos identificados neste trabalho serão valiosos para outros estudos de análise da expressão gênica na fase de esporocisto, uma vez que, até o momento não existem estudos de transcriptômica, sobretudo do mecanismo de SL *trans-splicing*, envolvendo esta fase do ciclo de vida do *S. mansoni*. Ainda, a fase de esporocisto apresenta algumas características peculiares que podem servir como importantes alvos para o controle da esquistossomose, como o modo de reprodução assexuada, o intenso metabolismo dentro do molusco e uma enorme capacidade na geração da fase infectante, as cercárias.

Durante a construção das bibliotecas, ao analisar as amostras de RNA total de esporocistos de *S. mansoni* por eletroforese capilar foi observado somente a banda correspondente ao RNA ribossomal 18S. Isso ocorre devido a uma clivagem no rRNA 28S gerando dois fragmentos (28S alpha e 28S beta) de, aproximadamente, mesmo tamanho do rRNA 18S e, portanto, não permite a detecção de ambas as bandas (TENNISWOOD; SIMPSON, 1982). Sendo assim, as amostras de RNA total extraído de esporocistos não

apresentaram degradação, visto que pela eletroforese capilar foi observada somente uma banda íntegra e com ausência de arraste.

O tratamento das amostras com DNase após a extração do RNA total remove contaminações com DNA genômico presentes no RNA. A purificação dos mRNAs realizada através da utilização de esferas magnéticas, têm se mostrado um método eficiente para a construção de bibliotecas de cDNA enriquecidas em SL *trans-splicing* como já descrito em trabalhos publicados pelo nosso grupo (MOURÃO *et al.*, 2013). Essas esferas magnéticas possuem uma monocamada de estreptavidina ligada covalentemente à sua superfície e se liga a biotina presente no oligo dT biotilado. Através da cauda de poli-A, os mRNAs se ligam ao oligo dT biotilado e são capturados.

Através do gel de poliacrilamida 8% (Figura 6) foi possível verificar o perfil dos produtos de PCR gerado após a construção da biblioteca enriquecida utilizando o iniciador *Spliced leader*. Esse perfil de arraste confirma a amplificação dos transcritos processados por SL *trans-splicing* uma vez que foram amplificados vários fragmentos de tamanhos diferentes (>100pb).

Após a amplificação dos transcritos, os produtos de PCR foram utilizados na construção da biblioteca do tipo fragmento para posterior sequenciamento utilizando a Plataforma Ion Torrent PGM™ System (Life Technologies). Assim, foi analisado a qualidade das leituras produzidas, já que este passo é importante para estudos qualitativos de transcriptômica, uma vez que quanto maior qualidade dos dados, maior a sua confiabilidade nas análises. O objetivo do controle de qualidade das *reads* é realizar uma avaliação das mesmas, geralmente, com ferramentas de visualização indicadas e, em seguida, proceder-se à correção, remoção e corte das leituras que não apresentem os valores mínimos de qualidade pretendidos (PABINGER *et al.*, 2013). Desse modo, a qualidade foi avaliada através do programa FastQC seguindo os parâmetros: Valores de qualidade por base; Valores de qualidade média por sequência; Conteúdo de bases nas sequências; Distribuição do conteúdo de GC por sequência e Conteúdo de GC em cada posição ao longo da sequência.

Atualmente, o programa que tem sido mais utilizado para visualização de qualidade dos dados, e foi o utilizado neste trabalho, é o FastQC (SCHMIEDER; EDWARDS, 2011). Quanto a análise de qualidade por base, em geral, um *score* de qualidade é considerado aceitável quando é superior ao valor 20 na escala de Phred, o que indica a probabilidade de ocorrência de 1 erro a cada 100 pb. Verifica-se que a média de qualidade por base gerada para as duas réplicas foi considerada ótima (em torno de 29), ao longo dos ciclos permanece estável apesar de que há uma queda nos valores de qualidade na porção 3' terminal das

sequências, entretanto, trata-se de artefatos dos sequenciadores utilizados (DOHM *et al.*, 2008) (Figura 7). Ainda, é possível observar que as leituras obtidas possuem tamanho variado de até 299 pb, tamanho esperado para o sequenciador utilizado.

Em seguida, os valores de qualidade média por sequência foram analisados. Este parâmetro permite verificar se um subconjunto das sequências tem valores de qualidade muito baixos, levando-se em consideração, a média de qualidade por base de cada sequência. Caso seja essa a realidade, ao invés de realizar um corte das sequências, o melhor a fazer seria descartar as sequências de qualidade baixa. Observou-se uma distribuição unimodal do valor de qualidade para as duas bibliotecas geradas, sendo o valor médio obtido de Phred em torno de 27 indicando que as bibliotecas tinham uma boa qualidade (Figura 8).

Em uma biblioteca aleatória é esperado pouca ou nenhuma diferença na proporção (conteúdo) de bases por posição na sequência, de modo que, as linhas nos gráficos devem permanecer paralelas ao eixo das abscissas. Entretanto, foi possível observar uma pequena oscilação para cada uma das bases nas duas bibliotecas geradas, mas, nada que indique um grande viés para uma posição em específico das leituras. É possível verificar um moderado pico da base Timina no início da posição de leitura, o que indica possível contaminação na biblioteca como a presença de adaptadores (Figura 9). Além disso, nas duas bibliotecas avaliadas, o conteúdo de A e T é maior que o de G e C. Já é descrito na literatura que o genoma de *S. mansoni* possui >60% de seu conteúdo em Adenina e Timina, exatamente o que foi observado nas bibliotecas construídas neste trabalho (RINALDI *et al.*, 2012).

O modelo de distribuição que é atribuído ao conteúdo de GC por sequência é uma curva normal em que o valor central corresponde ao teor global médio de GC do genoma. Apesar das duas curvas apresentarem formato e valor do conteúdo médio de GC semelhantes, foi observado alguns picos de maior conteúdo GC na curva de distribuição real do conteúdo médio de GC (curva em vermelho), mas nada que indique um viés nos dados gerados, podendo ser devido ao sequenciador utilizado (Figura 10).

Em relação ao conteúdo de GC em cada posição ao longo da sequência, ao se tratar de uma biblioteca randômica, o esperado é que o conteúdo de GC não varie consideravelmente ao longo das posições de uma sequência. Verifica-se que há uma pequena variação, principalmente, para as primeiras e últimas bases nas duas réplicas, o que é comum tratando-se dos iniciadores utilizados na construção das bibliotecas na plataforma Ion Torrent PGM™ System. No entanto, observa-se uma tendência em manter o conteúdo GC constante ao longo da sequência, o que é esperado devido ao fato de que a medida do conteúdo das bases

costuma ser espécie-específico, e em *S. mansoni* o conteúdo de GC é ser 36% (PROTASIO *et al.*, 2012) (Figura 11).

A análise de redundância das sequências é realizada em relação a um subconjunto dos dados, ou seja, utilizando 200.000 sequências que representam um conjunto de sequências únicas geradas (*unique*). O nível de duplicação das duas bibliotecas geradas foi semelhante (em torno de 80%), este alto valor de redundância das sequências é justificado pelo fato das sequências terem se originado de amplificação por PCR para o enriquecimento das bibliotecas em transcritos processados por SL *trans-splicing* (Figura 12) (MARTINS, 2014). O mesmo foi observado no sequenciamento de outras bibliotecas de SL *trans-splicing* envolvendo as fases de cercária, verme adulto, esquistossômulo e miracídio, podendo indicar a alta profundidade de sequenciamento, levando a exaustão e fidelidade dos dados gerados.

Em um estudo recente de transcriptoma envolvendo a fase de esporocisto nas mesmas condições de cultura aqui realizadas, foram identificados 6.677 genes expressos (WANG, B.; NEWMARK, 2013). Em nosso trabalho, após montagem e anotação das *reads* foram identificados 1.191 transcritos processados por SL *trans-splicing* na fase de esporocisto. De acordo com estes dados publicados, é possível inferir que, provavelmente, 15% do total de transcritos expressos na fase de esporocisto são processados por SL *trans-splicing* (Figura 13-A). No entanto, se o cálculo de porcentagem de transcritos processados por SL *trans-splicing* na fase de esporocisto for feito envolvendo os 10.407 genes expressos na 5ª versão do proteoma predito de *S. mansoni* e não somente o transcriptoma de esporocisto, este percentual é reduzido para 10% (Figura 13-B). Adicionalmente a este estudo, porém, envolvendo as fases de cercária, esquistossômulo e verme adulto do parasito *S. mansoni*, Protasio e colaboradores (2012) identificaram 1.178 genes processados por SL *trans-splicing*, correspondendo a 11% do número total de genes anotados a partir da 5ª versão do genoma de *S. mansoni*, neste trabalho, porém, não houve enriquecimento das bibliotecas em transcritos processados por SL *trans-splicing* e nem cobertura suficiente para detecção de transcritos raros (PROTASIO *et al.*, 2012). Deste modo, é importante ressaltar que o tipo de metodologia a ser utilizado para estudos de transcriptômica, sobretudo, estudos de mecanismo de regulação gênica como o SL *trans-splicing* deve ser avaliado criteriosamente.

Embora tenha sido observado que 10% dos transcritos expressos em esporocistos sejam processados por SL *trans-splicing*, dados preliminares do nosso grupo, utilizando três diferentes tipos de bibliotecas construídas para diferentes fases do parasito (cercária, verme adulto, esquistossômulo e miracídio), mostraram que esse percentual é ainda maior quando analisado globalmente. Foram identificados 8.457 genes processados por SL *trans-splicing*

representando 63% de todos os 13.222 genes anotados na 5ª versão do genoma de referência de *S. mansoni* (MARTINS, 2014). Essa proporção aumenta para 77% se considerando apenas os genes codificadores de proteínas (8.344 genes processados por SL *trans-splicing* em 10.787 genes codificadores de proteínas). Assim, vale ressaltar que o enriquecimento da biblioteca de cDNA através da metodologia utilizada neste trabalho é considerada eficiente e fundamental para estudos de SL *trans-splicing* permitindo profundidade e confiabilidade dos dados gerados.

Tendo em vista que as funções associadas ao mecanismo de SL *trans-splicing* e o seu papel no processamento de transcritos que codificam enzimas pertencentes à várias vias metabólicas nos diferentes organismos, permanecem ainda não esclarecidos, torna-se necessário elucidar os aspectos da estrutura e/ou funções dos genes que podem influenciar o impacto funcional de genes que sofrem SL *trans-splicing* em relação à expressão e tradução “convencional” de genes não processados por este mecanismo (MATSUMOTO *et al.*, 2010). Seria de grande valia no esclarecimento do papel deste mecanismo, um estudo direcionado para a tradução gênica.

A partir da classificação dos transcritos em categorias gênicas proposto por Karim e colaboradores (2011) foi possível atribuir funções biológicas aos genes processados por SL *trans-splicing* na fase de esporocisto. Os resultados de BLAST foram utilizados como dados de entrada para classificação de cada gene conhecido do parasito, incluindo os transcritos que são processados por SL *trans-splicing*. Em seguida, foi realizado o teste χ^2 (p-value <0.05) para averiguar se a diferença de frequência encontrada em cada classe é estatisticamente significativa entre os transcritos expressos no proteoma predito do *S. mansoni* e os transcritos expressos na fase de esporocisto processados por SL *trans-splicing*.

No total, foram encontrados transcritos pertencentes a 27 categorias funcionais (Figura 14). Quando realizado o teste χ^2 foram encontradas poucas categorias diferencialmente expressas estatisticamente significativas. As seis categorias funcionais que apresentaram alteração na frequência com significância estatística são:

- Elemento transponível;
- Transportador/estoque;
- Maquinaria de transcrição;
- Transdução de sinal;

- Maquinaria de síntese proteica e
- Maquinaria de modificação proteica

Interessantemente, a frequência de transcritos expressos processados por SL *trans-splicing* em esporocisto pertencentes as categorias de Maquinaria de transcrição, Maquinaria de síntese proteica e Maquinaria de modificação proteica é maior em relação a frequência de transcritos expressos no proteoma predito de *S. mansoni*, ou seja, são expressos 3% a mais de transcritos pertencentes as categorias funcionais de Maquinaria de transcrição e Maquinaria de síntese proteica e 1 % a mais de transcritos pertencentes a categoria de Maquinaria de modificação proteica (Figura 14).

Segundo Jolly e colaboradores (2007), os transcritos expressos na fase de esporocisto possuem funções biológicas que contribuem na diferenciação e, conseqüentemente, no desenvolvimento de um grande número de cercárias. Assim, estudos de expressão gênica revelam que, em geral, os genes mais expressos na fase de esporocisto estão relacionados à síntese de proteínas e dobramento de proteínas pós-tradução. Desse modo, é possível que o mecanismo de SL *trans-splicing* contribua para acentuar e potencializar a tradução e modificação de proteínas envolvidas no metabolismo de esporocisto, já que, como dito anteriormente, as categorias de Maquinaria de transcrição, Maquinaria de síntese proteica e de Maquinaria de modificação estão super-representadas nesta fase.

Entretanto, nas categorias funcionais Elemento transponível, Transportador/estoque e Transdução de sinal, a frequência de transcritos expressos e processados por SL *trans-splicing* na fase de esporocisto é menor do que a frequência de transcritos expressos no proteoma predito de *S. mansoni*.

Apesar da categoria funcional Elemento Transponível estar sub-representada dentre os transcritos processados por SL *trans-splicing* na fase de esporocisto, foram encontrados 6 elementos transponíveis nas bibliotecas geradas. Inseridos no grupo de sequências repetitivas curtas e longas nos genomas, os elementos transponíveis possuem habilidade de se mover no genoma (JURKA *et al.*, 2005). Em 1989, Finnegan propôs o primeiro sistema de classificação dos elementos transponíveis que discriminava duas classes (Classe I ou retrotransposons e Classe II ou transposons) que diferem basicamente no mecanismo de mobilização. Os elementos da Classe I possuem um mecanismo de mobilização baseado em um intermediário de RNA que utiliza duas enzimas transcritas pelos elementos: transcriptase reversa e integrase. Primeiro o mRNA do elemento transponível é transcrito pela maquinaria de transcrição do hospedeiro, após, o mRNA é revertido para DNA pela enzima transcriptase

reversa e, em seguida, o fragmento de DNA é inserido em outra região do genoma pela enzima integrase. Já nos elementos da Classe II, o mecanismo de mobilização é baseado em um intermediário de DNA, onde somente uma enzima chamada transposase retira o elemento de uma região do genoma e insere em outra (CAPY *et al.*, 1997).

Em eucariotos, os elementos transponíveis constituem uma grande parte do genoma e desempenham um papel importante na estrutura do genoma e evolução (CHARLESWORTH; SNIEGOWSKI; STEPHAN, 1994) (KIDWELL; LISCH, 1997). Através do seu movimento de excisão e inserção, a adição de cópias em outro local do genoma pode causar mutações resultando ou não em mudanças fenotípicas. Essas mutações muitas vezes podem ser favoráveis e acabam contribuindo para aumentar a diversidade existente na natureza, impulsionando a força evolutiva (MASIDE *et al.*, 2001). Em *S. mansoni* estima-se que 40% do genoma seja composto por sequências repetitivas, sendo mais de 20% constituído de retrotransposons (LAHA *et al.*, 2002). Em outros trabalhos realizados utilizando sequenciadores convencionais para identificação de transcritos processados por SL *trans-splicing* foram encontrados *uniques* apresentando homologia com partes de retrotransposons de *S. mansoni*: Saci-2: nas fases de verme adulto e ovo; Pererê: verme adulto e esquistossômulo; Boudicca: verme adulto (MOURÃO, 2005) (SOARES-FERNANDES, 2011). Parece inusitado a ocorrência de SL *trans-splicing* em elementos transponíveis, de tal modo, que são necessários mais estudos para elucidar e justificar as causas/consequências da participação deste mecanismo na produção de transcritos contendo parte de transposons e/ou retrotransposons.

Dentre as categorias funcionais diferencialmente expressas encontra-se também a categoria funcional de Transporte/estoque. Sabe-se que, os schistosomas são conhecidos por expressar uma vasta variedade de transportadores de soluto na superfície de seu tegumento, facilitando a absorção de uma grande variedade de moléculas de seus hospedeiros (BOYLE *et al.*, 2003). Dentre os poucos representantes dessa categoria encontrados nas bibliotecas, foi identificado um transportador de monocarboxilatos (Smp_168000) e um transportador de glicose (Smp_065770), que possui a função de facilitar o transporte de lactato dentro e fora das células. Sabe-se que genes pertencentes a esta categoria possuem um papel fundamental na relação do parasito e seus hospedeiros, haja vista as intensas trocas de moléculas que, ora favorecem a evasão do sistema defesa do hospedeiro, ora o reconhecimento do parasito pelas células de defesa, no entanto o SL *trans-splicing* não parece regular moléculas pertencentes a esta categoria (SALZET; CAPRON; STEFANO, 2000).

Adicionalmente, diferentes tipos de respostas celulares nos organismos como a manutenção da homeostase ou proliferação, são em grande parte, estimuladas por sinais extracelulares específicos. A célula processa a informação proveniente destes sinais a fim de construir uma resposta apropriada ao meio em que está inserida: viver ou morrer, proliferar ou permanecer em repouso, entre outras (ALBERTS *et al.*, 2002). É necessário um sistema elaborado de sinalização para que a célula responda de maneira específica ao sinal recebido e, em *S. mansoni* não seria diferente. Em cada fase do ciclo de vida do parasito são expressos vários sinalizadores que compõe a categoria funcional Transdução de sinal e estes, interceptam sinais diferentes dos ambientes em que o parasito está inserido (WALKER; RESSURREIÇÃO; ROTHERMEL, 2014). Assim, o mecanismo de Transdução de sinal possui funções essenciais no controle celular, principalmente, nas mudanças de estado de proteínas (ANDRADE *et al.*, 2014). Dentre os transcritos processados por SL *trans-splicing* pertencentes a esta categoria foram encontradas importantes proteínas quinases como: a Tirosina quinase (Smp_157300), que são moléculas-chave no controle da diferenciação e proliferação celular e representam alvos importantes de terapia molecular de doenças como a esquistossomose e o câncer (DISSOUS; AHIER; KHAYATH, 2007); uma Timidilato quinase (Smp_001360), reconhecida como um importante alvo potencial para drogas contra o HIV (CUI *et al.*, 2013); uma Serina/Treonina (Smp_142990) que são enzimas responsáveis pela defosforilação de resíduos de fosfoserina e/ou fosfotreonina. Apesar da importância desta categoria, bem como a categoria de transcritos envolvidos em Transporte/estoque, poucos transcritos pertencentes a mesma são processados por SL *trans-splicing* em esporocistos.

Após análise comparativa do perfil das classes funcionais de transcritos expressos e anotados no proteoma predito do *S. mansoni* e processados por SL *trans-splicing* na fase de esporocisto (Figura 15), verifica-se um perfil similar nas porcentagens de transcritos em cada classe funcional entre os transcritos expressos e anotados no proteoma predito do *S. mansoni* e processados por SL *trans-splicing* na fase de esporocisto.

Posteriormente, procedeu-se à identificação das vias metabólicas onde atuam as enzimas cujos transcritos são processados por SL *trans-splicing* através do iPath2.0 que é uma ferramenta disponível online (<http://pathways.embl.de>) para visualização e análise de vias celulares. O mapa gerado fornece um conjunto de vias metabólicas específicas para cada espécie disponível. A partir dessa interface *web* e dos dados disponíveis no banco de dados KEGG (KANEHISA; GOTO, 2000) (YAMADA *et al.*, 2011) foi gerado um mapa metabólico a partir das enzimas codificadas pelos transcritos processados por SL *trans-splicing* da fase de esporocisto às vias metabólicas do parasito *S. mansoni*.

O mapa metabólico específico de *S. mansoni* é constituído por várias vias pertencentes ao Metabolismo de Carboidrato, Metabolismo de Aminoácido, Metabolismo Energético, Metabolismo de Lipídeos, Metabolismo e Biosíntese de Glicanos, Metabolismo de outros Aminoácidos, Biosíntese de outros Metabólitos Secundários, Metabolismo de Nucleotídeos e Metabolismo de Co-fatores e Vitaminas (Figura 16).

Sendo assim, o mecanismo de SL *trans-splicing* está associado a todas as vias metabólicas existentes no parasito *S. mansoni*. Até o momento, dentre os transcritos anotados de *S. mansoni* não foram identificadas enzimas pertencentes ao Metabolismo de Terpenóides e Policetídeos e Metabolismo e Biodegradação de Xenobióticos, tampouco, estas foram identificadas nas bibliotecas construídas nesse trabalho. Entretanto, não podemos afirmar a inexistência de enzimas codificadas por transcritos processados por SL *trans-splicing* nestas vias. Na versão atual do iPath2.0 não são representados todos os genes de um organismo devido a pobre caracterização de muitos deles, e mais importante, nas bibliotecas geradas foram identificados 327 transcritos sem anotação funcional bem caracterizada sendo impossível associar a alguma via metabólica.

Neste trabalho foram encontrados transcritos processados por SL *trans-splicing* de diferentes enzimas envolvidas em todas as vias metabólicas do *S. mansoni*, assim, o mecanismo de SL *trans-splicing* apresenta como um mecanismo ubíquo, podendo ser importante para a regulação da expressão de vários genes associados com o desenvolvimento e adaptação das diferentes fases em seus diversos ambientes (MOURÃO *et al.*, 2013), ou simplesmente, poderia ocorrer aleatoriamente.

Como discutido acima, o mecanismo de SL *trans-splicing* em esporocistos ocorre de forma ubíqua, no entanto foi observado que as categorias Maquinaria de síntese proteica, Maquinaria de modificação proteica e Maquinaria de transcrição estão super-representadas nesta fase. Foi observado, em um trabalho do nosso grupo para as fases de miracídio, verme adulto e esquitossômulo que este mecanismo não é restrito apenas a uma categoria funcional (MARTINS, 2014). Isso confirma as informações já descritas na literatura de que este mecanismo parece não estar associado a nenhum tecido específico, fase de desenvolvimento ou gênero, tão pouco com genes específicos ou famílias gênicas (DAVIS *et al.*, 1995) (DAVIS, 1996) (MOURÃO *et al.*, 2013). No entanto, no trabalho de Martins e colaboradores (2014), foi observado que cercárias apresentam duas categorias sub-representadas dentre os transcritos processados por SL *trans-splicing*, são elas: Fator de Transcrição e Desconhecido.

Em seguida foi feita a análise de abundância dos transcritos anotados através do método de RPKM, determinado. A densidade de *reads* em uma região gênica de interesse, que é normalizado a partir do tamanho original do transcrito.

Após aplicação do método de RPKM, foram identificados dentre os 20 transcritos mais abundantes, vários transcritos de enzimas importantes do metabolismo para o desenvolvimento do parasito *S. mansoni* (Tabela 6).

O gene codificante da proteína ribossomal 40S foi identificado como um dos transcritos mais abundantes. Essa proteína é requerida para o processamento do pré-RNA ribossomal e para maturação da subunidade 40S. Uma vez que a regulação da expressão gênica ocorre, também, em nível traducional, a proteína ribossomal 40S pode ser um fator importante para o desenvolvimento celular (XU *et al.*, 1994).

Os transcritos correspondente ao inibidor de ATPase são abundantes nas bibliotecas de esporocistos geradas neste trabalho, este é um transcrito previamente descrito em estudos utilizando bibliotecas convencionais enriquecidas em SL *trans-splicing* (SOARES-FERNANDES, 2011) (MOURÃO *et al.*, 2013). Esta proteína é responsável por regular a atividade catalítica da ATP sintase mitocondrial, sem que haja a interrupção na produção de ATP (ZANOTTI *et al.*, 2004). Este transcrito também foi identificado em *Caenorhabditis elegans* como sendo processado por SL *trans-splicing* (BLUMENTHAL, 1995).

Relacionados ao ciclo celular, foram encontrados os transcritos que codificam as proteínas *enhancer of rudimentary protein*, bola-like (Bola) e distribuição nuclear (NudC). Estes três transcritos já haviam sido encontrados em outros trabalhos do nosso grupo, porém, utilizando outras fases do ciclo de vida do parasito (esquistossômulo, verme adulto e ovo) e metodologia de sequenciamento convencional (MOURÃO, 2005) (SOARES-FERNANDES, 2011). Em *Drosophila melanogaster*, a proteína *enhancer of rudimentary protein* é uma pequena proteína de 104 resíduos de aminoácidos, sem função conhecida. No entanto, tem sido proposto que ela esteja envolvida na manutenção do ciclo celular. Do ponto de vista evolucionário, esta proteína é altamente conservada e tem sido encontrada em todos os organismos multicelulares estudados (GELSTHORPE *et al.*, 1997). Curiosamente, este transcrito também foi encontrado em *T. solium* e *E. multilocularis* sendo processado por SL *trans-splicing* (BREHM; JENSEN; FROSCH, 2000). As proteínas Bola são amplamente encontradas nos organismos procariotos e eucariotos e estão intimamente envolvidas no processo de proliferação celular, entretanto, a função molecular dessas proteínas, ainda é desconhecida (KASAI, *et al.*, 2003). Já a proteína NudC é uma proteína conservada e capaz de regular a divisão celular (CHUANG *et al.*, 2013).

Ainda, entre os transcritos mais abundantes foi identificado um transcrito codificando uma peptidase. As peptidases, também conhecidas como proteinases, proteases ou enzimas proteolíticas, são enzimas que quebram ligações peptídicas entre os aminoácidos das proteínas. No parasito *S. mansoni* as peptidases são cruciais para o sucesso do parasitismo, incluindo aspectos de invasão, maturação e reprodução podendo exercer várias funções conforme a fase do ciclo de vida do parasito (TRAP; BOIREAU, 2000) (GARBER *et al.*, 2011). Segundo Sajid e colaboradores (SAJID *et al.*, 2003), na fase de verme adulto, são expressas várias peptidases associadas ao intestino que degradam proteínas do sangue do hospedeiro, incluindo a hemoglobina como um meio de nutrição. Na fase de miracídio, as proteinases possuem importante função na penetração no caramujo (YOSHINO *et al.*, 1993) e na transformação dos esporocistos (LODES; YOSHINO, 1989).

Outro transcrito abundante e importante na via oxidativa de *S. mansoni* é o codificante da Tiorredoxina. Além de ser processada por SL *trans-splicing* esta enzima é muito importante para a sobrevivência do parasito e esta foi até mesmo apontada como alvo para desenvolvimento de droga anti-*Schistosoma* (ALGER; WILLIAMS, 2002). Durante o período de transição da transformação dos miracídios e o início do desenvolvimento dos esporocistos primários, as larvas são vulneráveis ao estresse oxidativo. Esse estresse é gerado a partir de produtos da oxidação da hemoglobina no plasma ou espécies reativas de oxigênio/nitrogênio (ROS e RNS, respectivamente), resultante da resposta imune mediada por hemócitos (BENDER *et al.*, 2002).

Interessantemente, o mecanismo de SL *trans-splicing* não está restrito somente aos genes codificados pelo genoma nuclear, mas também está presente nos transcritos do genoma mitocondrial. Dentre os 20 transcritos mais abundantes na fase de esporocisto encontra-se um transcrito de um componente da cadeia respiratória Citocromo C oxidase (COX) – chaperona transportadora de cobre (Smp_029150) (Figura 17). Portanto, o mecanismo de SL *trans-splicing* parece estar envolvido na manutenção de alguns genes mitocondriais identificados até o momento.

Dentre os transcritos mais abundantes processados por SL *trans-splicing* em esporocisto foi encontrada uma proporção significativa (6 dos 20 transcritos selecionados, ou seja, 30%) de transcritos que codificam proteínas hipotéticas, com função desconhecida. Apesar de serem altamente expressos no transcriptoma de esporocisto, não é possível relacionar o mecanismo de SL *trans-splicing* a esses transcritos desconhecidos. Atualmente, sabe-se que mais de 42% do proteoma predito de *S. mansoni* não possui função conhecida consistindo de proteínas hipotéticas. Assim, tal fato reforça a importância de mais estudos de

genômica funcional e caracterização proteômica do *S. mansoni*. Adicionalmente, as funções dessas proteínas possam ser reveladas e talvez associadas aos mecanismos de regulação de expressão gênica como o *SL trans-splicing*.

Neste contexto, estudos funcionais para a caracterização do papel do mecanismo de *SL trans-splicing* em esporocisto de *S. mansoni*, demonstraram que o efeito de *knockdown* por RNA de interferência (RNAi) avaliado em alguns genes processados por *SL trans-splicing*, resulta na diminuição do comprimento de esporocistos (MOURÃO *et al.*, 2013). Compatível com o seu ciclo de vida, envolvendo complexos mecanismos de regulação de expressão gênica, estudos demonstram que durante o seu desenvolvimento o *S. mansoni* apresenta plasticidade, o que o torna capaz de alterar sua expressão gênica em resposta às mudanças nos diferentes ambientes em que vive (BOYLE *et al.*, 2003). Durante a transformação do parasito em esporocisto, este passa por intensas modificações fenotípicas caracterizadas pela perda dos movimentos e das placas epidérmicas, que é coincidente com a expansão das cristas intercelulares e na formação do tegumento do esporocisto (WU *et al.*, 2009).

Portanto, através do uso de tecnologias de sequenciamento e de novas metodologias para estudos de genômica e transcriptômica, várias funções gênicas e processos transcricionais como o mecanismo de regulação pós-transcricional *SL trans-splicing* podem ser detalhadamente estudados como o trabalho aqui realizado. Assim, por meio deste estudo pioneiro de transcriptômica envolvendo o mecanismo de *SL trans-splicing* na fase de esporocisto foi possível anotar e classificar funcionalmente 1.191 transcritos processados por *SL trans-splicing* nesta fase. Observou-se que o *SL trans-splicing* se apresenta como um mecanismo presente em diversas vias do metabolismo de esporocistos do parasito *S. mansoni*. Estes dados enriquecem os estudos de transcriptômica deste parasito e auxiliam na compreensão das reais funções desempenhadas pelo mecanismo de *SL trans-splicing*, auxiliando no desenvolvimento de possíveis ferramentas para o controle da esquistossomose mansônica.

7 CONCLUSÕES

- Com este trabalho foi confirmada a existência de genes processados por SL *trans-splicing* na fase de esporocisto do parasito *S. mansoni*.
- De acordo com os parâmetros de qualidade de sequências utilizados foram obtidas *reads* de boa qualidade e um alto percentual de alinhamento no genoma de referência das duas bibliotecas geradas.
- O percentual de transcritos processados por SL *trans-splicing* identificados no transcriptoma da fase de esporocisto foi maior do que o percentual encontrado na 5ª versão do proteoma predito de *S. mansoni*.
- Uma vez que o percentual de transcritos processados por SL *trans-splicing* identificados em cada classe funcional foi similar em relação ao proteoma predito de *S. mansoni*, podemos inferir que as bibliotecas de esporocisto geradas neste trabalho são representativas do transcriptoma do parasito.
- Dentre os transcritos processados por SL *trans-splicing* mais abundantes na fase de esporocisto foram identificados transcritos de enzimas importantes do metabolismo de desenvolvimento do parasito.
- O mecanismo de SL *trans-splicing* está representado em todas as vias metabólicas existentes no parasito *S. mansoni*, caracterizando-se como um mecanismo ubíquo.

8 PERSPECTIVAS

- Analisar a expressão dos transcritos processados por SL *trans-splicing* nas diferentes fases do ciclo de vida do parasito;
- Realizar ensaios de RNA de interferência para tentar silenciar o mecanismo de SL *trans-splicing* na fase de esporocisto;
- Comparar o perfil de proteínas expressas por parasitos silenciados e não silenciados para o mecanismo de SL *trans-splicing*.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. *Molecular Biology of the Cell, 4th edition*. [S.l: s.n.], 2002.
- ALGER, H. M.; WILLIAMS, D. L. The disulfide redox system of *Schistosoma mansoni* and the importance of a multifunctional enzyme, thioredoxin glutathione reductase. *Molecular and Biochemical Parasitology*, v. 121, n. 1, p. 129–139, 2002.
- ALTSCHUL, S F.; MADDEN, T L.; SCHÄFFER, A A.; ZHANG, J.; ZHANG, Z.; MILLER, W. LIPMAN, D J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, v. 25, n. 17, p. 3389–402, 1 set. 1997.
- ANDRADE, L F; MOURÃO, M M.; GERALDO, J A.; COELHO, F S.; SILVA, L L.; NEVES, R H.; VOLPINI, A.; MACHADO-SILVA, J R.; ARAUJO, N.; NACIF-PIMENTA, R.; CAFFREY, C R.; OLIVEIRA, G; Regulation of *Schistosoma mansoni* development and reproduction by the mitogen-activated protein kinase signaling pathway. *PLoS neglected tropical diseases*, v. 8, n. 6, p. e2949, jun. 2014.
- ARONESTY, E. Comparison of Sequencing Utility Programs. *The Open Bioinformatics Journal*, v. 7, n. 1, p. 1–8, 31 jan. 2013.
- BENDER, R C.; BIXLER, L M.; LERNER, J P.; BAYNE, C. *Schistosoma mansoni* sporocysts in culture: host plasma hemoglobin contributes to in vitro oxidative stress. *The Journal of parasitology*, v. 88, n. 1, p. 14–18, 2002.
- BERRIMAN, M. *et al.* The genome of the blood fluke *Schistosoma mansoni*. *Nature*, v. 460, n. 7253, p. 352–358, 16 jul. 2009.
- BITAR, M.; BORONI, M.; MACEDO, A M.; MACHADO, C R.; FRANCO, G R. The spliced leader trans-splicing mechanism in different organisms: molecular details and possible biological roles. *Frontiers in genetics*, v. 4, n. October, p. 199, jan. 2013.
- BLAXTER, M.; LIU, L. Nematode spliced leaders--ubiquity, evolution and utility. *International journal for parasitology*, v. 26, n. 10, p. 1025–33, out. 1996.
- BLUMENTHAL, T. Trans-splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends in genetics : TIG*, v. 11, n. 4, p. 132–6, abr. 1995.
- BONEN, L. Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, v. 7, n. 1, p. 40–6, jan. 1993.
- BOYLE, J. P.; WU, X J.; SHOEMAKER, C B.; YOSHINO, T P. Using RNA interference to manipulate endogenous gene expression in *Schistosoma mansoni* sporocysts. *Molecular and Biochemical Parasitology*, v. 128, n. 2, p. 205–215, maio 2003.
- BREHM, K.; JENSEN, K.; FROSCH, M. mRNA trans-splicing in the human parasitic cestode *Echinococcus multilocularis*. *The Journal of biological chemistry*, v. 275, n. 49, p. 38311–8, 8 dez. 2000.

- CAFFREY, C. R. Chemotherapy of schistosomiasis: present and future. *Current opinion in chemical biology*, v. 11, n. 4, p. 433–9, ago. 2007.
- CAPY, P.; LANGIN, T.; HIGUET, D.; MAURER, P.; BAZIN, C. Do the integrases of LTR-retrotransposons and class II element transposases have a common ancestor? *Genetica*, v. 100, n. 1-3, p. 63–72, jan. 1997.
- CARABIN, H.; GUYATT, H.; ENGELS, D. A comparative analysis of the cost-effectiveness of treatment based on parasitological and symptomatic screening for *Schistosoma mansoni* in Burundi. *Tropical Medicine and International Health*, v. 5, n. 3, p. 192–202, mar. 2000.
- CARVALHO, O. DOS S.; COELHO, P. M. Z.; LENZI, H. L. *Schistosoma mansoni e Esquistossomose: uma visão multidisciplinar*. [S.l: s.n.], 2008.
- CARVALHO, O. S. .; PASSOS, L. K. J. .; MENDONÇA, C.L.F.G.; CARDOSO, C. M. & R. L. C. *Moluscos de importância médica no Brasil*. [S.l: s.n.], 2005.
- CHARLESWORTH, B.; SNIÉGOWSKI, P.; STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, v. 371, n. 6494, p. 215–20, 15 set. 1994.
- CHIEFFI, P. P.; WALDMAN, E. A. Aspectos particulares do comportamento epidemiológico da Esquistossomose Mansônica no Estado de São Paulo, Brasil. *Cad. Saúde Pública*, v. 4, n. 3, p. 257–275, 1988.
- CHUANG, C.; PAN, J.; HAWKE, D H.; LIN, S.; YU-LEE, L. NudC deacetylation regulates mitotic progression. *PLoS one*, v. 8, n. 9, p. e73841, jan. 2013.
- CONRAD R, THOMAS J, SPIETH J, Blumenthal T. Insertion of part of an intron into the 5' untranslated region of a *Caenorhabditis elegans* gene converts it into a trans-spliced gene. 1991.
- COSTA, V. ANGELINI, C.; DE FEIS, I.; CICCODICOLA, A. Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of biomedicine & biotechnology*, v. 2010, p. 853916, jan. 2010.
- CUI, Q.; SHIN, W S.; LUO, Y.; TIAN, J.; CUI, H.; YIN, D. Thymidylate kinase: an old topic brings new perspectives. *Current medicinal chemistry*, v. 20, n. 10, p. 1286–305, jan. 2013.
- DAVIS, A. H. *Epidemiology and Community Control of Disease in Warm Climate Countries*. [S.l: s.n.], 1985.
- DAVIS, R. E.; HARDWICK, C.; TAVERNIER, P.; HODGSON, S.; SINGH, H. RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, *Schistosoma mansoni*. *The Journal of biological chemistry*, v. 270, n. 37, p. 21813–9, 15 set. 1995.
- DAVIS, R. E. Spliced leader RNA trans-splicing in metazoa. *Parasitology today (Personal ed.)*, v. 12, n. 1, p. 33–40, jan. 1996.
- DISSOUS, C.; AHIER, A.; KHAYATH, N. Protein tyrosine kinases as new potential targets against human schistosomiasis. *BioEssays: news and reviews in molecular, cellular and developmental biology*, v. 29, n. 12, p. 1281–8, dez. 2007.

- DOHM, J. C.; LOTTAZ, C.; BORODINA, T.; HIMMELBAUER, H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Research*, v. 36, n. 16, p. e105, set. 2008.
- DUCKERT, P.; BRUNAK, S.; BLOM, N. Prediction of proprotein convertase cleavage sites. *Protein engineering, design & selection : PEDS*, v. 17, n. 1, p. 107–12, jan. 2004.
- FALLON, P. G.; STURROCK, R. F.; NIANG, A. C.; DOENHOFF, M. J. Short report: diminished susceptibility to praziquantel in a Senegal isolate of *Schistosoma mansoni*. *The American journal of tropical medicine and hygiene*, v. 53, n. 1, p. 61–2, jul. 1995.
- FLINTOFT, L. Transcriptomics: Digging deep with RNA-Seq. *Nature Reviews Genetics*, v. 9, n. 8, p. 568–568, ago. 2008.
- FRANCO, G. R.; RABELO, E. M.; AZEVEDO, V.; PENA, H. B.; ORTEGA, J. M.; SANTOS, T. M.; MEIRA, W. S.; RODRIGUES, N. A.; DIAS, C. M.; HARROP, R.; WILSON, A.; SABER, M.; ABDEL-HAMID, H.; FARIA, M. S.; MARGUTTI, M. E.; PARRA, J. C.; PENA, S. D. Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). *DNA research : an international journal for rapid publication of reports on genes and genomes*, v. 4, n. 3, p. 231–40, 30 jun. 1997.
- FRANCO, G. R.; ADAMS, M. D.; SOARES, M. B.; SIMPSON, A. J. G.; VENTER, J. C.; PENA, S. D. J. Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene*, v. 152, n. 2, p. 141–147, jan. 1995.
- GARBER, M.; GRABHERR, M. G.; GUTTMAN, M.; TRAPNELL, C. Computational methods for transcriptome annotation and quantification using RNA-seq. *Nature methods*, v. 8, n. 6, p. 469–77, jun. 2011.
- GELSTHORPE, M.; PULUMATI, M.; MCCALLUM, C.; DANG-VU, K.; TSUBOTA, S. I. The putative cell cycle gene, enhancer of rudimentary, encodes a highly conserved protein found in plants and animals. *Gene*, v. 186, n. 2, p. 189–195, 1997.
- GROSSMAN, A. I.; SHORT, R. B.; CAIN, G. D. Karyotype evolution and sex chromosome differentiation in Schistosomes (Trematoda, Schistosomatidae). *Chromosoma*, v. 84, n. 3, p. 413–30, jan. 1981.
- HANSEN, J. E.; LUND, O.; TOLSTRUP, N.; GOOLEY, A. A.; WILLIAMS, K. L.; BRUNAK, S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconjugate journal*, v. 15, n. 2, p. 115–30, fev. 1998.
- HASTINGS, K. E. M. SL trans-splicing: easy come or easy go? *Trends in genetics : TIG*, v. 21, n. 4, p. 240–7, abr. 2005.
- HUANG, X.; HIRSH, D. A second trans-spliced RNA leader sequence in the nematode *Caenorhabditis elegans*. v. 86, n. November, p. 8640–8644, 1989.
- HUANG, Y.; MANDERSON, L. Schistosomiasis and the social patterning of infection. *Acta tropica*, v. 51, n. 3-4, p. 175–94, ago. 1992.

- ISMAIL, M.; BOTROS, S.; METWALLY, A.; WILLIAM, S.; FARGHALLY, A.; TAO, L F.; DAY, T A.; BENNETT, J L. Resistance to praziquantel: direct evidence from *Schistosoma mansoni* isolated from Egyptian villagers. *The American journal of tropical medicine and hygiene*, v. 60, n. 6, p. 932–5, jun. 1999.
- JURKA, J.; KAPITONOV, V V.; PAVLICEK, A.; KLONOWSKI, P.; KOHANY, O.; WALICHIEWICZ, J. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, v. 110, n. 1-4, p. 462–467, jan. 2005.
- KANEHISA, M.; GOTO, S.; SATO, Y.; FURUMICHI, M.; TANABE, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic acids research*, v. 40, n. Database issue, p. D109–14, jan. 2012.
- KANEHISA, M.; GOTO, S. Kegg: kyoto encyclopedia of genes and genomes. 2000.
- KARIM, S.; SINGH, P.; RIBEIRO, J. M. C. A deep insight into the sialotranscriptome of the gulf coast tick, *Amblyomma maculatum*. *PLoS ONE*, v. 6, n. 12, p. e28525, jan. 2011.
- KATZ, N. *Inquérito Nacional de Prevalência da Esquistossomose e Geo-helmitos*. . [S.l.: s.n.], 2014.
- KIDWELL, M. G.; LISCH, D. Transposable elements as sources of variation in animals and plants. *Proceedings of the National Academy of Sciences of the United States of America*, v. 94, n. 15, p. 7704–11, 22 jul. 1997.
- LAHA, T.; BRINDLEY, P J.; VERITY, C K.; MCMANUS, D P.; LOUKAS, A. Pido, a non-long terminal repeat retrotransposon of the chicken repeat 1 family from the genome of the Oriental blood fluke, *Schistosoma japonicum* q. v. 284, p. 149–159, 2002.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods*, v. 9, n. 4, p. 357–9, abr. 2012.
- LASDA, E. L.; BLUMENTHAL, T. Trans-splicing. *Wiley interdisciplinary reviews. RNA*, v. 2, n. 3, p. 417–34, 2011.
- LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNEL, T.; RUAN, J.; HOMER, N. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, v. 25, n. 16, p. 2078–9, 15 ago. 2009.
- LODES, M. J.; YOSHINO, T. P. Characterization of excretory-secretory proteins synthesized in vitro by *Schistosoma mansoni* primary sporocysts. *The Journal of parasitology*, v. 75, n. 6, p. 853–62, dez. 1989.
- MARIONI, J. C.; MASON, C E.; MANE, S M.; STEPHENS, M.; GILAD, Y. RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome research*, v. 18, n. 9, p. 1509–17, set. 2008.
- MARONEY, P A.; DENKER, J A.; DARZYNKIEWICZ, E.; LANEVE, R.; NILSEN, T W. Most mRNAs in the nematode *Ascaris lumbricoides* are trans-spliced: a role for spliced leader addition in translational efficiency. *RNA (New York, N.Y.)*, v. 1, n. 7, p. 714–23, set. 1995.

- MARTINS, M. L. B. *O spliced leader trans-splicing no parasito Schistosoma mansoni*. 2014. Universidade Federal de Minas Gerais, 2014.
- MASIDE, X. BARTOLOMÉ, C.; ASSIMACOPOULOS, S.; CHARLESWORTH, B. Rates of movement and distribution of transposable elements in *Drosophila melanogaster*: in situ hybridization vs Southern blotting data. *Genetical research*, v. 78, n. 2, p. 121–36, out. 2001.
- MATSUMOTO, J.; DEWAR, K.; WASSERSCHIED, J.; WILEY, G B.; MACMIL, S L.; ROE, B.; ZELLER, R W.; SATOU, Y.; HASTINGS, K E M. High-throughput sequence analysis of *Ciona intestinalis* SL trans-spliced mRNAs: Alternative expression modes and gene function correlates. *Genome Research*, v. 20, n. 5, p. 636–645, maio 2010.
- MORTAZAVI, A .; WILLIAMS, B A.; MCCUE, K.; SCHAEFFER, L.; WOLD, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods*, v. 5, n. 7, p. 621–628, 2008a.
- MOURÃO, M M., BITAR, M.; LOBO, F P.; PECONICK, A P.; GRYNBERG, P.; PROSDOCIMI, F.; WAISBERG, M.; CERQUEIRA, G C.; MACEDO, A M.; MACHADO, C R.; YOSHINO, T.; FRANCO, G R. A directed approach for the identification of transcripts harbouring the spliced leader sequence and the effect of trans-splicing knockdown in *Schistosoma mansoni*. [S.l: s.n.], 2013. v. 108.
- MOURÃO, M. D. M. Obtenção e Análise de transcritos produzidos por trans splicing em *Schistosoma mansoni*. 2005.
- NELSON, D. L.; COX, M. M. *Lehninger principles of Biochemistry*. [S.l: s.n.], 2000.
- NEVES, D. P. Parasitologia Humana. [S.l: s.n.], 2005.
- NIELSEN, H.; ENGELBRECHT, J.; BRUNAK, S.; VON HEIJNE, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein engineering*, v. 10, n. 1, p. 1–6, jan. 1997.
- PABINGER, S.; DANDER, A.; FISCHER, M.; SNAJDER, R.; SPERK, M.; EFREMOVA, M.; KRABICHLER, B.; SPEICHER, M R.; ZSCHOCKE, J.; ZLATKO, T. A survey of tools for variant analysis of next-generation genome sequencing data. v. 15, n. 2, p. 256–278, 2013.
- PETTITT, J.; MÜLLER, B.; STANSFIELD, I.; CONNOLLY, B. Spliced leader trans-splicing in the nematode *Trichinella spiralis* uses highly polymorphic, noncanonical spliced leaders. *RNA (New York, N.Y.)*, v. 14, n. 4, p. 760–70, abr. 2008.
- PROTASIO, A. V.; TSAI, I J.; BABBAGE, A.; NICHOL, S.; HUNT, M.; ASLETT, M A.; SILVA, N .; VELARDE, G S.; ANDERSON, T J C.; CLARK, R C.; DAVIDSON, C.; GARY, P.; HOLROYD, N E.; LOVERDE, P T.; LLOYD, C.; MCQUILLAN, J.; OTTO, T D.; PARKER-MANUEL, S J.; QUAIL, M A.; WILSON, R A.; DUNNE, D W.; BERRIMAN, M. A Systematically Improved High Quality Genome and Transcriptome of the Human Blood Fluke *Schistosoma mansoni*. v. 6, n. 1, 2012.
- RAJKOVIC, A.; DAVIS, R E.; SIMONSEN, J N.; ROTTMAN, F M. A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proceedings of the National Academy of Sciences of the United States of America*, v. 87, n. 22, p. 8879–83, nov. 1990.

RINALDI, G.; ECKERT, S E.; TSAI, I J.; SUTTIPRAPA, S.; KINES, K J.; TORT, J F.; MANN, V H.; TURNER, D J.; BERRIMAN, M.; BRINDLEY, P J. Germline transgenesis and insertional mutagenesis in *Schistosoma mansoni* mediated by murine leukemia virus. *PLoS pathogens*, v. 8, n. 7, p. e1002820, jan. 2012.

ROBERTS, A.; TRAPNELL, C.; DONAGHEY, J.; RINN, J L.; PACHTER, L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome biology*, v. 12, n. 3, p. R22, jan. 2011.

SABAH, A. A.; FLETCHER, C.; WEBBE, G.; DOENHOFF, M J. *Schistosoma mansoni*: chemotherapy of infections of different ages. *Experimental parasitology*, v. 61, n. 3, p. 294–303, jun. 1986.

SAJID, M.; MCKERROW, J H.; HANSELL, E.; MATHIEU, M A.; LUCAS, K D.; HSIEH, I.; GREENBAUM, D.; BOGYO, M.; SALTER, J P.; LIM, K C.; FRANKLIN, C.; KIM, J.; CAFFREY, C R. Functional expression and characterization of *Schistosoma mansoni* cathepsin B and its trans-activation by an endogenous asparaginyl endopeptidase. *Molecular and biochemical parasitology*, v. 131, n. 1, p. 65–75, set. 2003.

SALZET, M.; CAPRON, A.; STEFANO, G. B. Molecular crosstalk in host-parasite relationships: schistosome- and leech-host interactions. *Parasitology today (Personal ed.)*, v. 16, n. 12, p. 536–40, dez. 2000.

SANTOS, O. &. Aspectos epidemiológicos da esquistossomose mansônica nos Bairros Novo Horizonte e Campo Limpo, Feira de Santana, Bahia. [S.l: s.n.], 2002.

SCHMIEDER, R.; EDWARDS, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics (Oxford, England)*, v. 27, n. 6, p. 863–4, 15 mar. 2011.

SHORT, R. B. Presidential address: Sex and the single schistosome. *The Journal of parasitology*, v. 69, n. 1, p. 3–22, fev. 1983.

SIEGEL, T. N.; TAN, K. S. W.; CROSS, G. A. M. Systematic study of sequence motifs for RNA trans splicing in *Trypanosoma brucei*. *Molecular and cellular biology*, v. 25, n. 21, p. 9586–94, nov. 2005.

SINHA, A. *et al.* Genome-wide analysis of trans-splicing in the nematode *Pristionchus pacificus* unravels conserved gene functions for germline and dauer development in divergent operons. *RNA (New York, N.Y.)*, p. rna.041954.113–, 2014.

SOARES-FERNANDES. Identificação e silenciamento de transcritos processados por Spliced leader trans-splicing em *Schistosoma mansoni*. 2011.

SONNHAMMER, E. L.; VON HEIJNE, G.; KROGH, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology*, v. 6, p. 175–82, jan. 1998.

TENNISWOOD, M. P.; SIMPSON, A. J. The extraction, characterization and in vitro translation of RNA from adult *Schistosoma mansoni*. *Parasitology*, v. 84, n. Pt 2, p. 253–61, abr. 1982.

TRAP, C.; BOIREAU, P. [Proteases in helminthic parasites]. *Veterinary research*, v. 31, n. 5, p. 461–71, 2000.

TRAPNELL, C.; PACHTER, L.; SALZBERG, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics (Oxford, England)*, v. 25, n. 9, p. 1105–11, 1 maio 2009.

WALKER, A. J.; RESSURREIÇÃO, M.; ROTHERMEL, R. Exploring the function of protein kinases in schistosomes: perspectives from the laboratory and from comparative genomics. *Frontiers in genetics*, v. 5, n. July, p. 229, jan. 2014.

WANG, B.; III, J. J. C.; NEWMARK, P. A. Functional genomic characterization of neoblast-like stem cells in larval *Schistosoma mansoni*. p. 1–15, 2013.

WANG, W.; WANG, L.; LIANG, Y.-S. Susceptibility or resistance of praziquantel in human schistosomiasis: a review. *Parasitology research*, v. 111, n. 5, p. 1871–7, nov. 2012.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, v. 10, n. 1, p. 57–63, 2009.

WHO. Weekly epidemiological record. *World Health Organization technical report series*, v. 88, n. 8, p. 81–88, 2013.

WILHELM, B. T.; LANDRY, J.-R. RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods (San Diego, Calif.)*, v. 48, n. 3, p. 249–57, jul. 2009.

XU, L.; HE, G P.; LI, A.; RO, H S. Molecular characterization of the mouse ribosomal protein S24 multigene family: a uniquely expressed intron-containing gene with cell-specific expression of three alternatively spliced mRNAs. *Nucleic acids research*, v. 22, n. 4, p. 646–55, 25 fev. 1994.

YAMADA, T.; LETUNIC, I.; OKUDA, S.; KANEHISA, M.; BORK, P. iPath2.0: interactive pathway explorer. v. 39, n. May, p. 412–415, 2011.

YEATS, B.; MATSUMOTO, J.; MORTIMER, S I.; SHOGUCHI, E.; SATOH, N.; HASTINGS, K E M. SL RNA genes of the ascidian tunicates *Ciona intestinalis* and *Ciona savignyi*. *Zoological science*, v. 27, n. 2, p. 171–80, fev. 2010.

YOSHINO, T P.; LODES, M J.; REGE, A A.; CHAPPELL, C L. Proteinase activity in miracidia, transformation excretory-secretory products, and primary sporocysts of *Schistosoma mansoni*. *The Journal of parasitology*, v. 79, n. 1, p. 23–31, fev. 1993.

ZANOTTI, F.; RAHO, G.; GABALLO, A.; PAPA, S. Inhibitory and anchoring domains in the ATPase inhibitor protein IF1 of bovine heart mitochondrial ATP synthase. *Journal of bioenergetics and biomembranes*, v. 36, n. 5, p. 447–57, out. 2004.