
Universidade Federal de Minas Gerais

**Análise de Transcritos Produzidos por Trans-splicing em
*Schistosoma mansoni***

Dissertação de mestrado apresentada ao
departamento de Bioquímica e
Imunologia da Universidade Federal
de Minas Gerais

**Por
Marina de Moraes Mourão**

Orientadora: Glória Regina Franco

Agosto de 2005

Lista de Figuras	I
Lista de Tabelas	III
Lista de Abreviaturas	IV
Resumo	VI

Índice

1. Introdução	1
1.1 Organismo de Estudo – <i>Schistosoma mansoni</i>	1
1.2 O ciclo de vida de vida do <i>Schistosoma</i>	6
1.3 O Genoma do <i>Schistosoma</i>	8
1.4 <i>Splicings</i>	11
1.4.1. <i>Cis splicing</i>	11
1.4.2. Spliceossomo	12
1.4.3. <i>Trans splicing</i>	14
1.5. Bioinformática	21
2. Objetivos	26
2. 1. Objetivo Geral	26
2.2. Objetivos específicos	26
3. Materiais e Métodos	27
3.1. Materiais Biológicos	27
3.2. Extração de RNA	27
3.3. Transcrição reversa	29
3.4. Amplificação dos transcritos que sofrem <i>trans splicing</i>	30
3.5. Seleção dos cDNAs por tamanho	33
3.5.1. Seleção por Polietilenoglicol (PEG) 8000 15%	33
3.5.2. Seleção por eletroforese em gel de agarose 1%	33
3.6. Ligação dos cDNAs aos vetores	36
3.7. Transformação de bactérias <i>Escherichia coli</i> DH5 α eletrocompetentes e seleção dos clones	36
3.8. Seleção dos clones	38
3.9. Obtenção dos plasmídeos recombinantes	38
3.10. Seqüenciamento das extremidades dos insertos	39
3.11. Análise Bioinformática	40
3.11.1. Classificação dos <i>uniques</i> produzidos por <i>trans splicing</i> em categorias funcionais ...	42
3.13.2. Anotação Manual dos Uniques	44
4. Resultados e Discussão	45
4.1 Extrações de RNA	45
4.2. Amplificação dos transcritos que sofrem <i>trans-splicing</i>	45
4.3. Seleção dos cDNAs por tamanho	46

4.4. Seleção dos clones	46
4.5. Quantificação do DNA plasmidial.....	53
4.6. Análises Bioinformáticas.....	55
4.6.1. Processamento das seqüências geradas	55
4.6.2. Análise das seqüências agrupadas pelo programa CAP3	56
4.6.3. Buscas de homologia dos unives de <i>S.mansoni</i> produzidos por <i>trans splicing</i>	62
4.6.6. Classificação dos <i>uniques</i> produzidos por <i>trans splicing</i> em categorias funcionais	82
4.6.7. Identificação de novas ESTs.....	86
5. Conclusões	93
6. Perspectivas	94
7. Referências Bibliográficas.....	95

1. Introdução

1.1 Organismo de Estudo – *Schistosoma mansoni*

A esquistossomose é uma das parasitoses humanas mais comuns, sendo endêmica em 76 países (Figura 1), principalmente naqueles em desenvolvimento (Engels *et al.*, 2002). A doença é um dos mais sérios problemas de saúde pública, sendo considerada a segunda doença tropical mais prevalente e com alta morbidade. Esta persiste, devido às precárias condições de vida, ao fornecimento inadequado de água e pela inadequação das instalações sanitárias, nas quais a população está inserida. É estimado que 200 milhões de pessoas estão infectadas e outros 600 milhões correm risco de serem infectadas e ainda, estima-se que anualmente morram 11.000 pessoas em todo o mundo (WHO: www.who.int/tdr/diseases/schisto/direction.htm).

Apesar de ser uma doença muito antiga, o estudo científico da esquistossomose começou apenas há 150 anos. Quem primeiro identificou o parasita *Schistosoma mansoni* foi Theodor Bilharz, no Egito, em 1851. A doença, popularmente conhecida como xistose, barriga d'água ou mal do caramujo, chegou às Américas durante o tráfico negreiro através de imigrantes orientais e asiáticos. (WHO: <http://www.who.int/tdr/diseases/schisto/>)

A esquistossomose é causada por organismos do filo Platyhelminthes, classe trematodea, ordem Digenea, da família Schistosomatidae e do gênero *Schistosoma*, que compreende espécies como *Schistosoma mansoni* (África e América Latina), *S. haematobium* (África e Oriente Médio), *S. intercalatum* (10 países do cinturão de florestas equatoriais da África), *S. japonicum* (restrito à região do Pacífico, incluindo a China, e nas Filipinas) e *S. mekongi* (limitado a áreas do Laos e Camboja) (WHO: <http://www.who.int/tdr/diseases/schisto/diseaseinfo.htm>). No Brasil, a esquistossomose é causada pela espécie *S. mansoni* (Figura 2) e seu hospedeiro intermediário é a *Biomphalaria glabrata* (Figura 3).

A doença é determinada, principalmente, pela distribuição do hospedeiro intermediário. O primeiro alvo para erradicação da esquistossomose deve ser impedir a infecção e segundo reduzir a morbidade. Novos focos surgem principalmente em locais onde foram realizados projetos de

direcionamento da água para indústria e agricultura. Este desenvolvimento estendeu o número e a distribuição de áreas alagadas e, portanto habitats dos hospedeiros intermediários. O posterior aumento da concentração de pessoas nestas áreas favoreceu que a esquistossomose se espalhasse (Sturrock, 2001).

Em alguns países, programas nacionais de controle da esquistossomose foram bem sucedidos. Porém, nos países africanos abaixo do Saara (compreendendo 80% das pessoas infectadas do mundo), o controle é praticamente inexistente.

O surgimento de drogas contra a esquistossomose no início dos anos 80, dentre elas o praziquantel, que é efetivo e seguro contra todos os *Schistosomas*, restaurou a intenção de erradicação da doença (<http://www.who.int/tdr/diseases/schisto/>). No entanto, não há prevenção contra reinfecções, o que invariavelmente ocorre, já que a doença dificilmente é controlada. Os programas de controle de maior sucesso foram aqueles que incluíram algum método que interrompesse a transmissão através do uso de moluscidas (Webbe & El Hak, 1990).

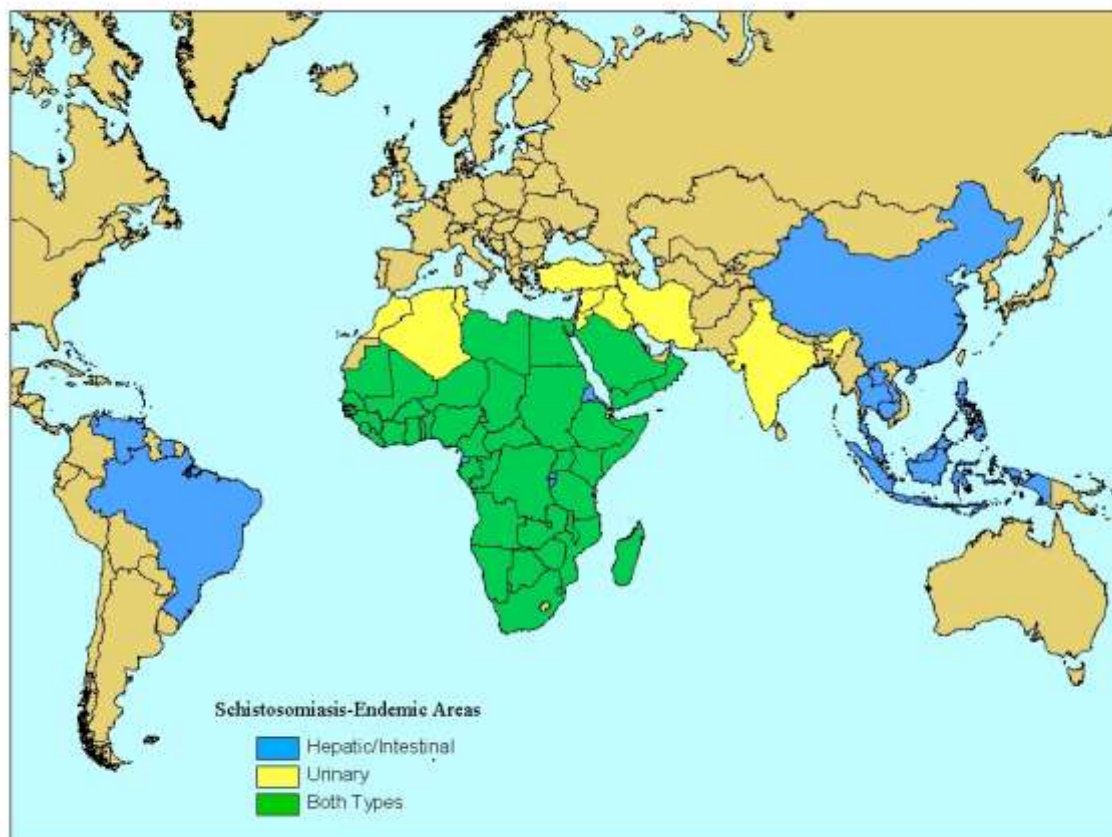


Figura 1. Áreas endêmicas da esquistossomose no mundo. O mapa mostra áreas endêmicas de esquistossomose no mundo, categorizadas por áreas de ocorrência da esquistossomose hepática/intestinal, urinária e onde ambas formas da doença ocorrem. Fonte:http://www.cdc.gov/travel/diseases/maps/schisto_map.htm.



Figura 2: Vermes adultos de *Schistosoma mansoni*. Fonte: www.usuhs.mil/mic/Davies/Research.html



Figura 3: Exemplar de *Biomphalaria glabrata*. Fonte: www.sea-river-news.com/50_2.htm

1.2 O ciclo de vida do *Schistosoma*

O ciclo biológico (Figura 4) foi descrito, na mesma época, por Lutz no Brasil e Lerper no Egito. O *Schistosoma mansoni* apresenta ciclo de vida complexo, envolvendo hospedeiros intermediários (molusco), encontrando-se em sua fase assexuada, e definitivos (mamíferos), onde encontra-se na fase adulta sexuada e vivem como parasitas no homem e em outros vertebrados (Pessoa & Martins, 1982).

Os ovos são eliminados na água através das fezes. Em condições ótimas, estes eclodem liberando o miracídio que nada ao encontro do caramujo específico, penetrando-o que, assim, passa a ser seu hospedeiro intermediário. O estágio no caramujo inclui a transformação em duas gerações de esporocistos e posteriormente em cercária. Um miracídio pode gerar até 300.000 cercárias. Após se liberarem, as cercárias vão para a água nadando e penetram ativamente em seu hospedeiro definitivo, comumente o homem. Durante sua penetração, a cercária perde sua cauda bifurcada e glicocálice, tornando-se o esquistossômulo. Estes migram através do sistema sanguíneo e linfático para os pulmões, até se instalarem no sistema porta hepático, onde atingem a maturidade sexual. Os vermes adultos - dióicos, com ventosa oral e ovíparos - vivem acasalados no plexo venoso mesentérico e portal e, para ovoposição, migram para as veias mesentéricas inferiores (Lichtenberg, 1987). As fêmeas colocam em média 400 ovos por dia. Estes ovos atravessam para o lúmen do intestino sendo, então, eliminados nas fezes. Alguns ovos ficam alojados na mucosa intestinal ou nos capilares do sistema porta, tornando-se circundados por células características de resposta inflamatória granulomatosa do hospedeiro (revisado por Pessoa & Martins, 1982).

Foram estudadas neste trabalho, a fase semelhante à forma de esquistossômulo (cultura de sete dias após transformação “*in vitro*” das cercárias em esquistossômulos), a fase de ovo e a fase de vermes adultos, incluindo separadamente, fêmeas e machos, que são provenientes de infecção bissexual e sendo os casais separados após a perfusão dos camundongos infectados.

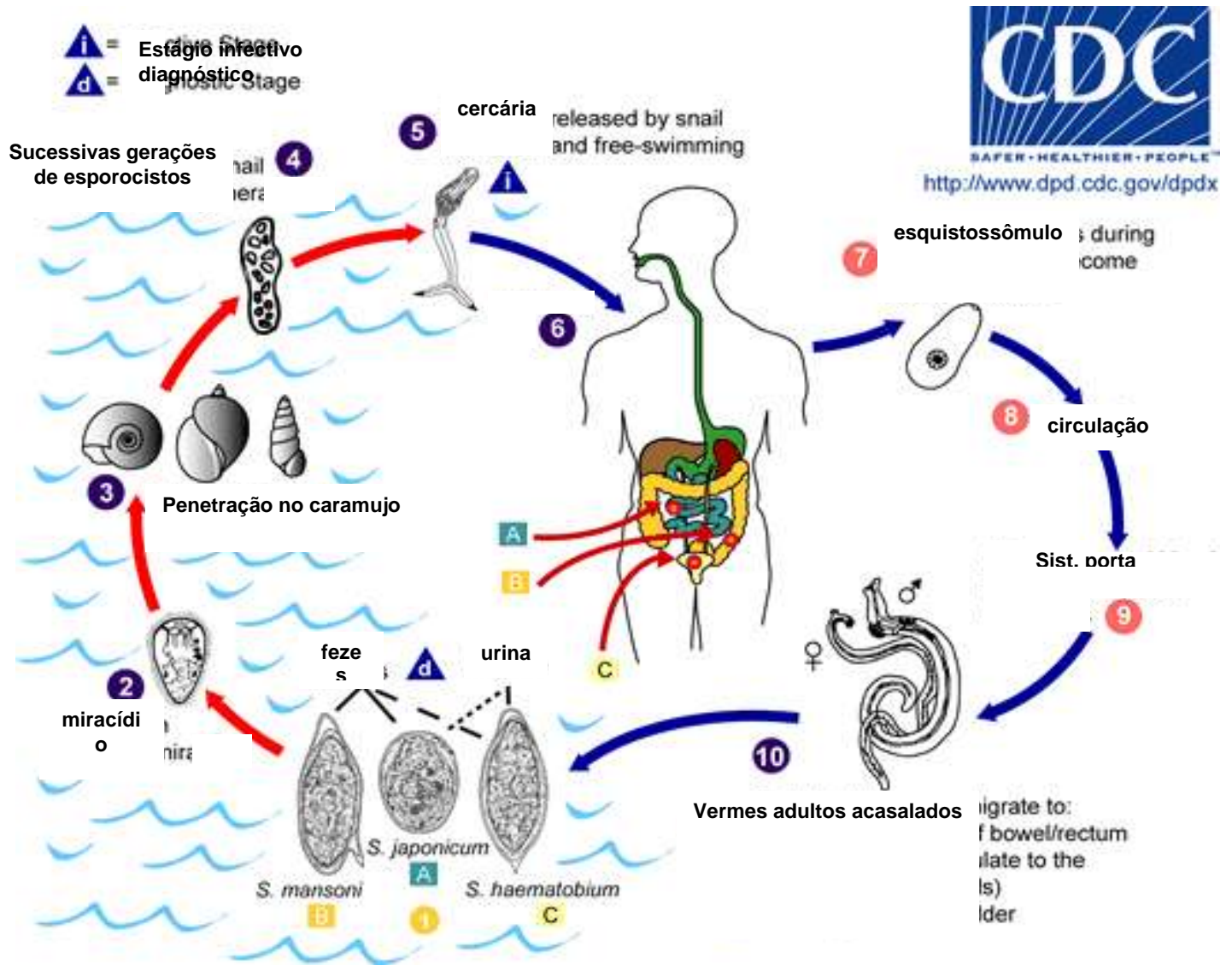


Figura 4: Ciclo de vida do verme *Schistosoma*. Fonte: Adaptado de <http://www.dpd.cdc.gov/dpdx/HTML/Schistosomiasis.htm>.

1.3 O Genoma do *Schistosoma*

O estudo do genoma do parasita é uma forma de se obter informações que podem ser traduzidas em novas ferramentas de controle, além disso, genomas de parasitas oferecem melhores perspectivas para identificação de novos alvos para drogas, vacinas, diagnósticos e bases biológicas para o entendimento da interação parasita-hospedeiro. O estudo do genoma do *Schistosoma*, pode também ajudar a elucidar questões mais fundamentais sobre este organismo e sobre a evolução dos animais, já que este é representante do nível mais simples dos animais que apresentam simetria bilateral com cromossomos heterogaméticos podendo conduzir, assim, a respostas às questões sobre alguns dos genes mais primitivos para tais funções evolutivas.

Estudos iniciais do genoma de *S. mansoni* revelaram um genoma diplóide com a presença de sete pares de cromossomos autossomos e um par de cromossomos sexuais. Neste organismo o sexo homogamético é o macho (ZZ) e a fêmea o heterogamético (ZW) (Short & Grossman, 1981; Short, 1983).

O genoma haplóide estimado deste organismo, compreende 270 Mb, sendo composto de aproximadamente 40% de sequências altamente repetitivas, incluindo aqui os genes ribossomais que estão arranjados seguidamente, podendo ter de 500 a 1000 cópias no genoma e 60% do genoma representa sequências de cópias únicas ou pequenas famílias gênicas (Simpson *et al.*, 1982). Estima-se que esse organismo possua entre 15.000 a 25.000 genes (Franco *et al.*, 2000, Verjovski-Almeida *et al.*, 2003), e 66% de seu genoma é composto de adenina e timina.

O projeto de descoberta gênica do *Schistosoma* foi iniciado em 1992 no Brasil, e a partir de 1995 passou a contar com o suporte da Organização Mundial de Saúde (United Nations Development Programme e World Bank Special Program for Reserch and Training in Tropical Diseases – UNDP/World Bank/TDR). O projeto visava a descoberta e caracterização de novos genes de *S. mansoni* e *S. japonicum* e seqüenciamento e análise do genoma mitocondrial. Após esta data, houve um enorme crescimento do banco de dados de sequências deste organismo (Franco *et al.*, 1995, 2000), visto que em 1993 haviam 600 ESTs depositadas e hoje temos 152.942 depositadas em banco de dados públicos (dbEST release 071505, data 15/07/05).

Franco e colaboradores (1995), inauguraram a era de descoberta gênica em *S. mansoni* utilizando a estratégia de ESTs. ESTs ou etiquetas de seqüências transcritas, são pequenas seqüências, em média de 400 nucleotídeos, geradas das pontas de cDNAs, selecionadas randomicamente de bibliotecas. As ESTs podem ser utilizadas para busca de homologia em banco de dados de DNA e proteína (Adams *et al.*, 1991). Nos estudos das bibliotecas foram observados 60% de clones de cDNA não identificados nos bancos de dados, o que pode indicar genes específicos da espécie, ou que ainda não haviam sido catalogados para outras espécies (Franco *et al.*, 1995).

Recentemente, houve um grande aumento no número de ESTs. Verjovski-Almeida e colaboradores (2003) depositaram 139.064 ORESTES no banco de dados de ESTs. ORESTES são pequenas seqüências de cDNAs, em geral representando a parte central dos transcritos, produzidos por transcrição reversa e PCR usando iniciadores, arbitrariamente selecionados, sob condições de baixo baixo rigor de amplificação. As seqüências foram geradas pelo consórcio ONSA, Organization for Nucleotide Sequencing and Analysis, em São Paulo, financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP. Este projeto foi iniciado em 2001, onde esperava-se gerar 40 a 50 milhões de bases de cDNA de *S.mansoni*, o que representaria algo em torno de 1 vez a 1 vez e meia o transcriptoma do parasita. O projeto, finalizado em 2003, teoricamente, cobriu 92% do transcriptoma do parasita.

Diversas características interessantes de *S.mansoni* puderam ser observadas com este trabalho:

- ❖ Composição gênica geral – obtiveram 37% de genes relacionados a metabolismo, bem como de um número especialmente alto de genes relacionados à comunicação intracelular e controle da transcrição gênica, como se pode esperar de um organismo parasita submetido a diversos habitats durante o seu ciclo de vida.
- ❖ Através da comparação gênica da interseção dos genes comuns às espécies de organismos multicelulares invertebrados (*Drosophila melanogaster*, *Caenorhabditis elegans*, *Ciona intestinalis* e *S. mansoni*) versus a interseção dos

genes comuns aos unicelulares eucariotos (*S. cerevisiae* e *Plasmodium falciparum*), foi possível determinar diversos genes específicos de metazoários e de eucariotos indicando, assim, quais genes responderiam pelas características mais primitivas destes grupos.

- ❖ Diversos genes envolvidos em aspectos biológicos importantes para a compreensão da relação parasita-hospedeiro em *S. mansoni*, tais como sinalização celular, longevidade do parasita, resposta a diversos tipos de stress e evasão ao sistema imune do hospedeiro puderam ser visualizados em um nível de detalhe nunca obtido anteriormente.

- ❖ Diversos novos possíveis candidatos a vacinas puderam ser descobertos e caracterizados.

Esse projeto inclui atualmente a obtenção de mais informações da biologia desse parasita também através o uso de microarranjo de cDNA para a análise temporal do seu desenvolvimento. Mais informações podem ser obtidas no site <http://verjo18.iq.usp.br/schisto/>.

Em breve, a Rede Genoma de Minas Gerais, financiada pela Fundação de Amparo a Pesquisa de Minas Gerais (FAPEMIG) e pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pretende depositar em torno de 70.000 transcritos de cinco bibliotecas, dos estágios de desenvolvimento: vermes adultos, fase similar à fase de esquistossômulo, ovo e cercária.

O programa de sequenciamento do genoma desenvolvido pelo The Institute for Genomic Research (TIGR), nos Estados Unidos, em colaboração com o Sanger Institute, na Inglaterra, se iniciou em Outubro de 2002, adotando uma estratégia de sequenciamento de “Whole genome shotgun”, e pretende cobrir 7,5 vezes o genoma, com 2,8 milhões de leituras (El-Sayed *et al.*, 2004). Nesta estratégia o DNA genômico total é fragmentado, aleatoriamente, em pedaços menores. Uma faixa de tamanho dos insertos é selecionadas e os fragmentos são clonados em vetor plasmídico, constituindo-se bibliotecas. Os insertos dos clones são sequenciados em ambas as

extremidades e as seqüências são agrupadas para gerar os “scaffolds”, que são grandes seqüências genômicas. Em breve o genoma do *S. mansoni* será publicado por estas instituições.

O genoma mitocondrial de duas espécies de *Schistosoma* foi também seqüenciado e possui 36 genes. Viu-se neste estudo, que *S. mansoni* e *S. japonicum* possuem diferenças significantes em seus arranjos gênicos mitocondriais (Le *et al.*, 2001).

Devido ao enorme progresso na área genômica, e sua crescente quantidade de informação gerada, seqüenciamento em larga escala e anotação gênica não são mais problemas para laboratórios em todo o mundo. O déficit existente, no entanto, se encontra nas caracterizações das funções gênicas, que é um processo demorado, devido à necessidade de desenvolvimento de novas ferramentas para pesquisa *in vivo*, *in vitro* e *in silico*.

1.4 Splicings

1.4.1. *Cis splicing*

Em eucariotos em geral, a RNA polimerase II transcreve o DNA, gerando um transcrito primário. No transcrito primário, a seqüência de nucleotídeos codificante do polipeptídeo pode não ser contígua. As seqüências que interrompem a região codificante são os introns e os segmentos codificantes são chamados éxons. O processo que remove os introns do transcrito primário é chamado *cis splicing*, e é através deste processo que os éxons serão postos de forma a produzir uma seqüência contínua de um RNA maduro e funcional (Nelson & Cox, 2000). O *cis splicing* é um antigo processo eucariótico, que está preservado em todas as linhagens, apesar de muito reduzida em alguns grupos, tais como: Kinetoplastidae e em leveduras.

1.4.2. Spliceossomo

Alguns dos eventos moleculares mais interessantes no metabolismo de RNA ocorrem durante o processo pós-sintético, sendo denominado de processamentos pós-transcricionais. O *splicing* é um destes mecanismos. As enzimas que catalisam o *splicing* consistem de RNAs ao invés de proteínas. São as ribozimas, ou RNAs catalíticos (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, e são RNAs pequenos, de 100 a 200 nucleotídeos, muito conservados, possuindo um sítio poliuridínico (Nelson & Cox, 2000). Neste processo, é preciso que haja sítios específicos no transcrito primário, onde os Us se ligarão. Primeiro ocorre a ligação do U1 no sítio específico do transcrito primário, na parte 3' do exon e 5' do intron. Logo após, ocorre a ligação do 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 5):

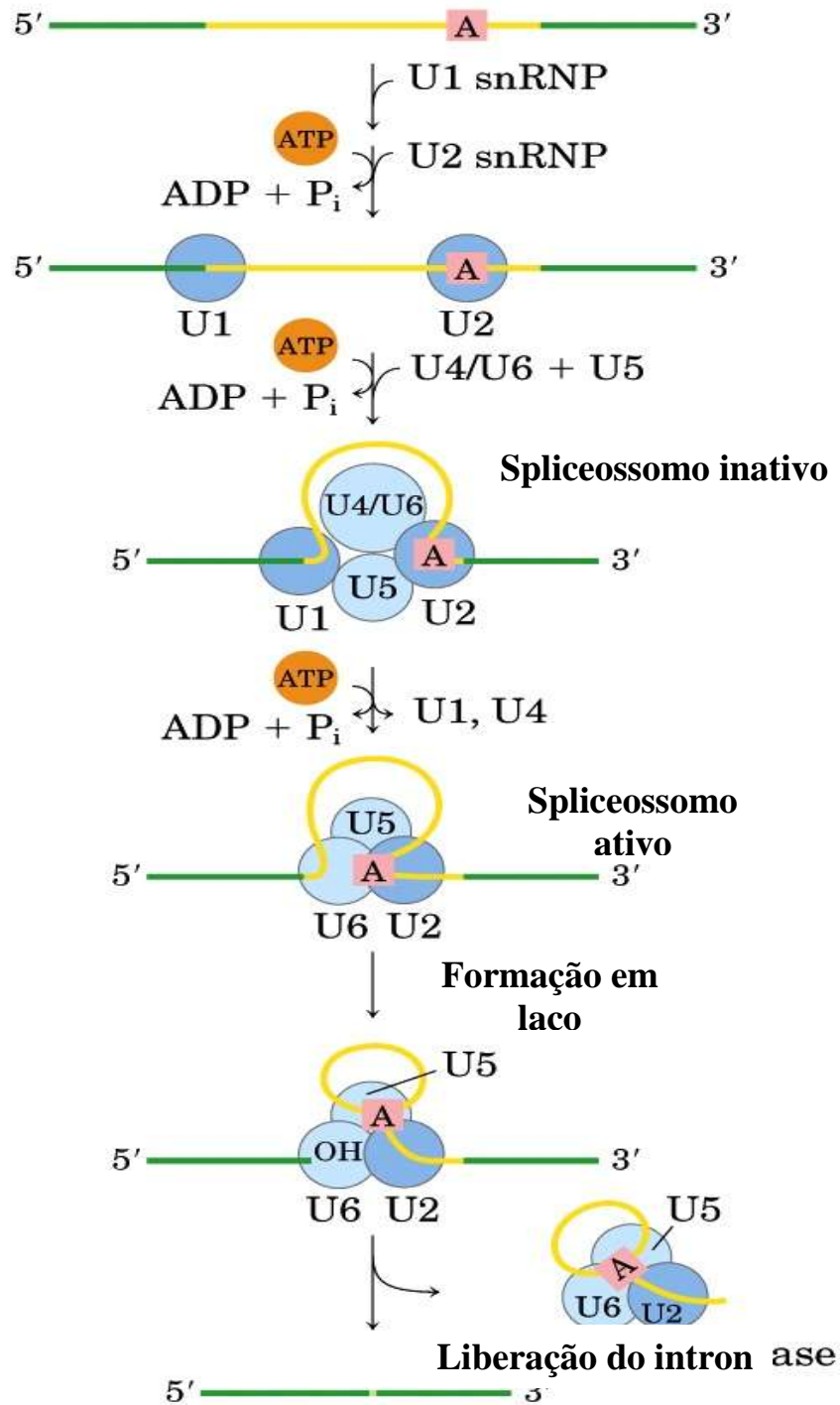


Figura 5: Esquema do processamento do transcrito primário através de *cis* splicing. Fonte: Adaptado de Nelson & Cox, 2000.

1.4.3. *Trans splicing*

Numerosas e complexas regulações gênicas e pós-transcricionais devem ocorrer neste parasita de ciclo tão complexo, que envolve acomodações alternadas a ambientes distintos como a água e o meio interno de seus hospedeiros. Estas adaptações são apenas parcialmente compreendidas e por isso, suas elucidações são um amplo campo para pesquisa, pois em fases críticas do desenvolvimento o parasita pode ser susceptível a medidas de controle e interrupção das infecções (Neves, 2000). O *trans splicing* pode ser uma destas formas de regulação.

Em adição ao *cis splicing*, o spliceossomo é capaz de fazer o *trans splicing* (junção de dois RNAs separados para formar um RNA quimera). Ambos existem juntos em alguns organismos. 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/íntron, este se juntará ao próximo sítio receptor no final do íntron, ocorrendo, assim, *cis splicing*. Caso haja um sítio receptor despareado, sem um sítio doador correspondente, ocorrerá *trans splicing*. Similarmente, a presença de sítios receptores despareados internos, serão alvo de resolução de transcritos policistrônicos. O spliceossomo, no entanto, tem clara preferência pelo *cis splicing*, mas a base deste mecanismo não é ainda conhecida (Harstings, 2005).

Trans splicing é um evento de processamento de RNA onde há precisamente a transferência spliceossômica de seqüências derivadas de RNAs independentemente transcritos. Uma de suas formas é adição de uma seqüência identificada como *spliced leader* (SL), que é doada da extremidade 5' de um RNA pequeno, não poliadenilado, para alguns pré-mRNAs receptores, para formar o éxon 5' terminal dos mRNAs maduros. Assim, desde uma minoria até a totalidade de uma população de pré-mRNAs, em organismos diferentes, adquirem a seqüência comum na extremidade 5' (Davis *et al.*, 1995). Frações variantes da população de mRNA sofrem *trans splicing* em filos diferentes. Nos protistas, aparentemente, todos os transcritos sofrem *trans splicing*. Já em metazoários, a fração de genes vai desde a grande maioria (70 a 90%) em *Caenorhabditis elegans* (Blaxter *et al.*, 1996) e *Ascaris lumbricoides* (Maroney *et al.*, 1995), até uma imprecisa minoria em *Schistosoma* (Davis *et al.*, 1995).

Há duas formas de *trans splicing* que ocorrem naturalmente:

- 1- *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. Estes pré-mRNAs, em geral são transcritos de um mesmo gene, ou de uma mesma região genômica. Este processo está presente em plantas e algas (Bonen,1993);
- 2- *Trans splicing* do SL: neste processo, o doador é especializado, ele é um não pré-mRNA e possui de 15 a 50 nucleotídeos. Este se torna o 5' do mRNA receptor maduro (Harstings, 2005).

Todo organismo que sofre *trans splicing* tem um ou alguns SL RNA. O SL RNA é transcrito pela DNA polimerase III, de genes pequenos, que não possuem introns e que são repetidos inúmeras vezes seguidamente no genoma. Em alguns organismos, entre estes segmentos há o gene que codifica o rRNA 5S. Em *S.mansoni* a localização do gene do rRNA 5S não condiz com a localização do gene do SL e, este, está representado em 200 cópias por genoma haplóide (Rajkovic, 1990). SL RNAs são pequenos RNAs de 40 a 140 nucleotídeos, que têm um sítio doador de *splicing* e um cap 5' supermodificado (2,2,7 trimetilguanosina-TMG em metazoários, ou 7 monometilguanosina hipermetilado em kinetoplastídeos) (Nilsen, 1993). O sítio doador de *splicing* divide o RNA SL em dois segmentos: - o 5', que é a seqüência leader; - o 3' que se comporta como se fosse a parte inicial de um intron. Apesar de não terem seqüências semelhantes, os SL RNAs possuem uma estrutura secundária impressionantemente similar aos snRNAs, que estão presentes nas snRNPs spliceossomais que participam no mecanismo de *splicing* (Harstings, 2005).

SL *trans splicing* tem funções diferentes no processo de maturação do mRNA. O SL *trans splicing* serve para (1) prover o 5' cap para RNAs transcritos pela RNA polimerase I, que são codificadores de proteínas (unicamente em kinetoplastídeos), (2) resolver transcritos policistrônicos da RNA polimerase II em transcritos individuais com 5' cap, (3) acentuar a eficiência traducional por adicionar cap hipermodificado ou a seqüência leader no mRNA. Uma função adicional, e menos entendida, é a de cortar as regiões 5' não traduzidas de pré-mRNAs. A

parte 5' de um pré-mRNA monocistrônico talvez seja descartada quando possua elementos que possam comprometer o mRNA, como por exemplo, um AUG fora da fase de leitura.(Harstings, 2005). (Figura 6)

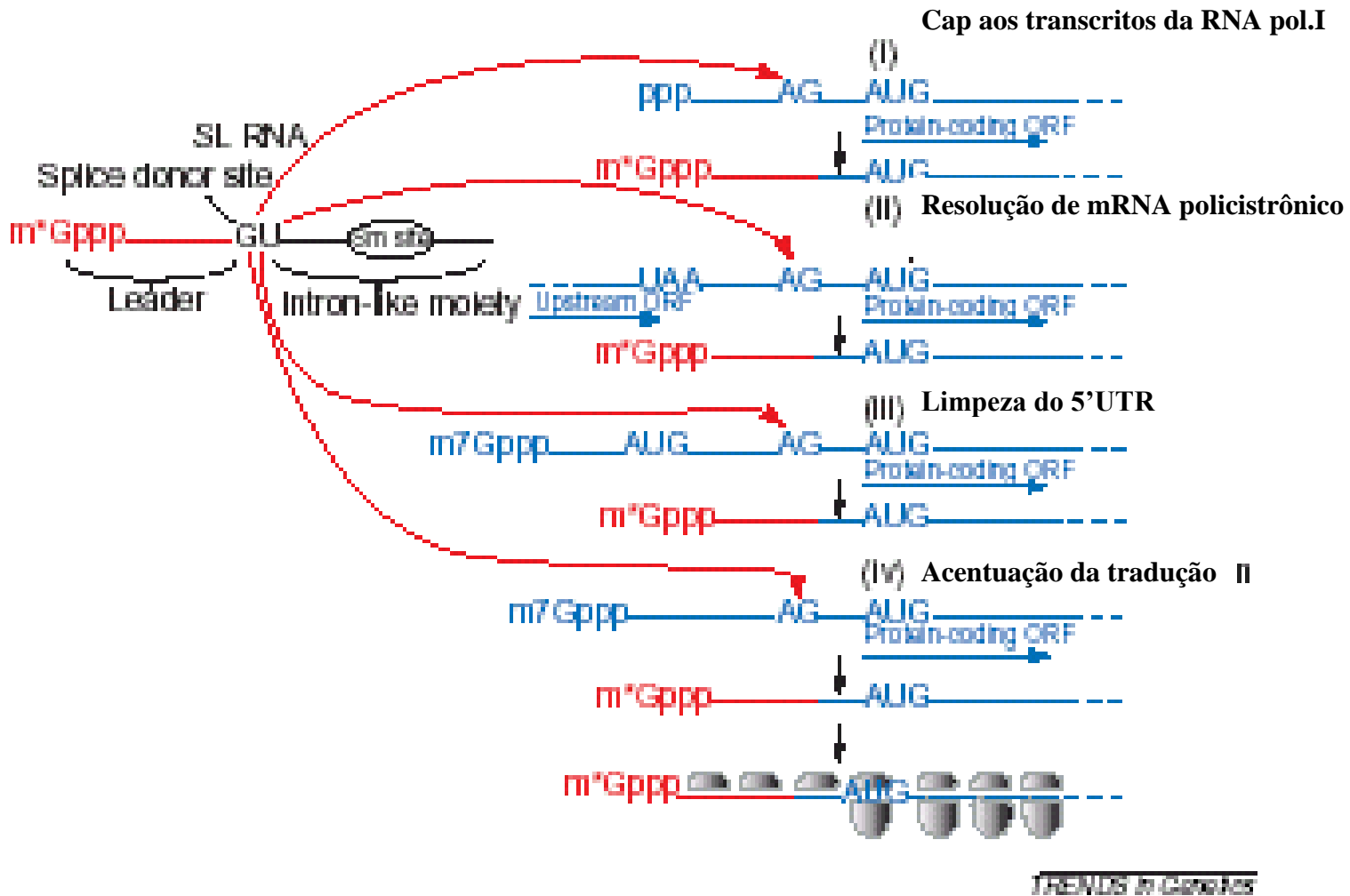


Figura 6: Exemplos de funções do trans splicing. Fonte: Adaptado de Harstings, 2005.

SL *trans splicing* não parece estar especificamente associado a tecidos particulares, estágio de desenvolvimento, sexo (Davis, 1996) ou com classes de genes funcionalmente distintos (Davis, 1995), embora tenha sido visto que uma fração pequena de mRNAs que sofrem *trans splicing* em *C. elegans* codificam proteínas *housekeeping*, e são derivados de operons policistrônicos. Estes genes *housekeeping* são especialmente compatíveis com a organização em operons, porque eles não requerem controles transcricionais independentes, e podem ser expressos constitutivamente. A resolução de operons não é a única razão de 90% dos genes de nematódeos sofrerem *trans splicing*, haja visto que, apenas 15% do genoma destes organismos está organizado em operons (Blumenthal *et al.*, 2003).

O mecanismo de *trans splicing* foi visto pela primeira vez há vinte anos, em tripanossomatídeos, onde 39 nucleotídeos eram adicionados a todos os transcritos policistrônicos (Revisado por Agabian, 1990). Somente após esta adição, estes mRNAs eram considerados maduros e eram traduzidos. Posteriormente, um fenômeno parecido foi visto em nematódeos, mas apenas um subconjunto de pré-mRNAs sofriam a adição de um SL de 22 nucleotídeos (Bektesh, *et al.* 1988). A descoberta de *trans-splicing* em nematódeos aumentou a idéia da prevalência deste fenômeno em outros metazoários, que ao se confirmar, sugere um modelo particular de processamento do RNA, que pode ser uma forma importante de controle da expressão gênica.

SL *trans splicing* nunca tinha sido visto em muitos dos grupos intensamente estudados, tais como: plantas, vertebrados, artrópodes, fungos. Por isso, pensava-se que este mecanismo apenas estaria ligado aos primeiros grupos identificados. Sabemos hoje, que *trans splicing* está presente nos mais variados filós e até mesmo em cordados e cnidários.

Ultimamente, *trans splicing* foi visto na *Ciona intestinalis*, uma ascidiacea da classe Urocordata (Vandenberghe *et al.*, 2001). Logo após, foi identificado em *Oikopleura dioica*, uma apendicularia, em que 12-24% dos pré-mRNAs sofrem *trans splicing* com a função de resolução de seus operons (Ganot *et al.*, 2004).

Estas descobertas tiveram o maior impacto no estudo dos SL *trans splicing*, pois estes últimos organismos foram os primeiros na divisão deuterostômios a apresentarem este processo. A descoberta mais significativa foi que pela primeira vez observou-se SL *trans splicing* dentro do filo dos cordados, que tem como grupo principal os vertebrados, que, no entanto, não fazem *trans splicing*. Isso tem implicações profundas para os estudos evolutivos.

Em *S. mansoni*, calcula-se que algo em torno de 30% dos mRNAs sofrem *trans splicing*, mas nenhum padrão foi ainda observado, provavelmente por ter sido muito pouco estudado. Rajkovic e colaboradores (1990), mostraram que o SL RNA de *S. mansoni* contém 36 nucleotídeos que são advindos de um RNA não poliadenilado de 90 nucleotídeos e que não mostra similaridade em sua seqüência às de outros organismos, sugerindo assim que o SL RNA de *S. mansoni* seja espécie - específico. No entanto, a estrutura secundária do SL RNA de *S. mansoni* se assemelha bastante às estruturas secundárias dos SL RNAs de nematódeos e de *Trypanosoma cruzi* (Figura7).

Explorando o significado biológico do *trans splicing* em trematódeos, Davis e colaboradores (1995), usaram *S. mansoni* como modelo. Neste trabalho, foram utilizados apenas 30 mRNAs e 4 genes para avaliar (1) se haveria algum padrão nas proteínas codificadas por estes; (2) se haveria alguma característica geral evidente no mecanismo; (3) se o AUG que termina a seqüência do SL poderia servir de metionina inicial na tradução e se, finalmente, (4) haveria alguma especificidade entre fêmeas e machos, ou entre tecidos. Estes autores não observaram nenhuma especificidade no processo. Mas estas conclusões podem estar relacionadas à baixa amostragem utilizada neste trabalho.

Em relação a presença do trinucleotídeo AUG no SL e sua possível utilização como códon para iniciação da tradução, observou-se que este codon está presente não só no SL de *S. mansoni*, mas também na seqüência de outros platelmintos (*Echinococcus* e *Faciola*). Calcula-se que 43% dos transcritos que sofrem *trans splicing* em *E. granulosus*, iniciam sua tradução utilizando este códon (Fernández *et al.*, 2002).

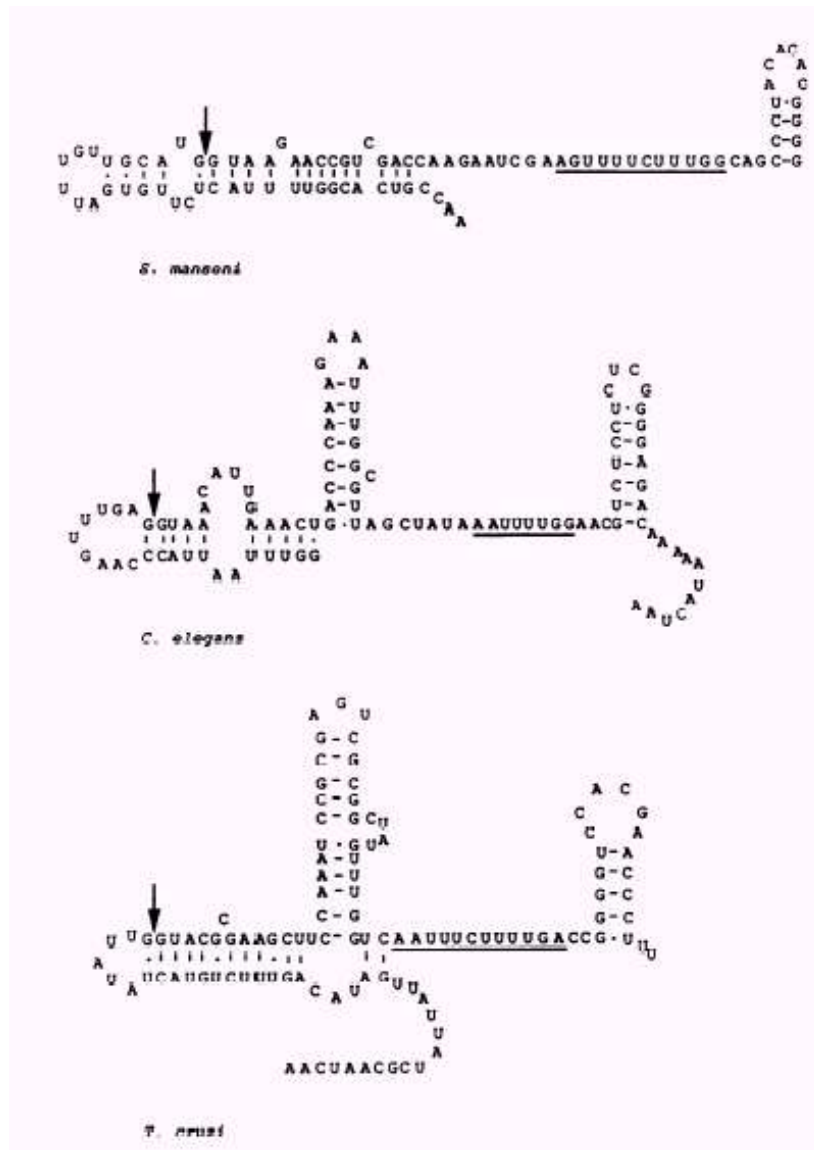


Figura 7: Estrutura secundária dos SLRNAs de diferentes organismos: *S. mansoni*, *C. elegans*, *T. cruzi*. As setas indicam o sítio doador de splicing e a seqüência sublinhada o sítio de ligação a proteínas Sm. Fonte: Davis *et al.*, 1995.

Para o estudo de *trans splicing*, há uma conveniente estratégia utilizando o 5' *spliced leader* para isolar cDNAs completos (Gems *et al.*, 1995 e Gregory *et al.*, 1997). Os mRNAs de parasitas podem ser prontamente amplificados usando iniciadores para o SL e para cauda de poli A 3'. Esta estratégia evita contaminantes de hospedeiros e contaminação mitocondrial (Fernández *et al.*, 2002).

Propomos aqui a análise de uma quantidade significativa de cDNAs advindos de bibliotecas construídas a partir de transcritos completos contendo o SL, avançando nos trabalhos que já foram realizados. Acrescentaremos também comparações entre diferentes bibliotecas *S. mansoni*. Estas novas seqüências completas geradas serão caracterizadas, tornando-se de grande valia para os projetos genomas e transcriptoma deste organismo. Esses projetos encontram-se em fase de anotação e têm se deparado com uma grande carência de genes completos do parasita depositados em banco de dados públicos, o que dificulta enormemente a anotação do genoma. Além disso, a presença da seqüência SL e do *trans splicing* em parasitas protozoários e metazoários oferece um alvo potencial para o desenvolvimento de específicos agentes quimioterápicos capazes de inibir uma grande variedade de patógenos (Davis *et al.*, 1995)

1.5. Bioinformática

A bioinformática surgiu a partir do momento que se iniciou a utilização de ferramentas computacionais para a análise de dados genéticos, bioquímicos e de biologia molecular. A bioinformática envolve várias áreas do conhecimento e tem como uma das finalidades desvendar a grande quantidade de dados que vem sendo obtida através de seqüências de DNA e proteínas. Para a análise dos genomas e transcriptomas, a bioinformática é imprescindível e a biologia molecular moderna não estaria tão avançada hoje, se não fossem os recursos computacionais existentes.

Ferramentas existentes e de grande valia para quaisquer projetos de transcriptoma são os programas PHRED e Cross Match. O primeiro mencionado analisa os cromatogramas de seqüências geradas por sequenciadores automáticos de DNA, e agrega valores de qualidade a cada uma das bases seqüenciadas, além de remover partes espúrias das extremidades das seqüências. O

segundo programa compara as seqüências obtidas com as seqüências dos vetores de clonagem utilizados, mascarando as regiões de vetores, caudas de poliA e poliT, nas seqüências.

Análise de agrupamento é outra ferramenta muito utilizada em bioinformática. O programa é chamado CAP3, e deve comparar as seqüências de ESTs tratadas do organismo entre si, de forma a encontrar quais delas são idênticas, ou contêm regiões parecidas o suficiente para que sejam reunidas em uma só. Assim, o programa apresenta uma saída contendo as seqüências que foram agrupadas - chamadas de consensos ou *contigs* - e as seqüências que não foram reunidas (por não apresentarem similaridade suficiente com nenhuma outra) - chamadas de *singlets*. Cada uma das seqüências resultantes do agrupamento (seja ela uma *singlet* ou *contig*) é chamada de *unique*. (Figura 8). Cada uma das seqüências *unique* deve representar um gene distinto, apesar, de ser possível que mais de um *unique* corresponda a um único gene. No entanto, o agrupamento das seqüências é importante devido, principalmente, aos seguintes fatores: -(1) elimina a redundância das seqüências; (2) aumenta o tamanho das seqüências facilitando a anotação por homologia (Oliveira & Johnston, 2001); (3) aumenta o nível de confiabilidade de cada seqüência (Miller *et al.*, 1999).

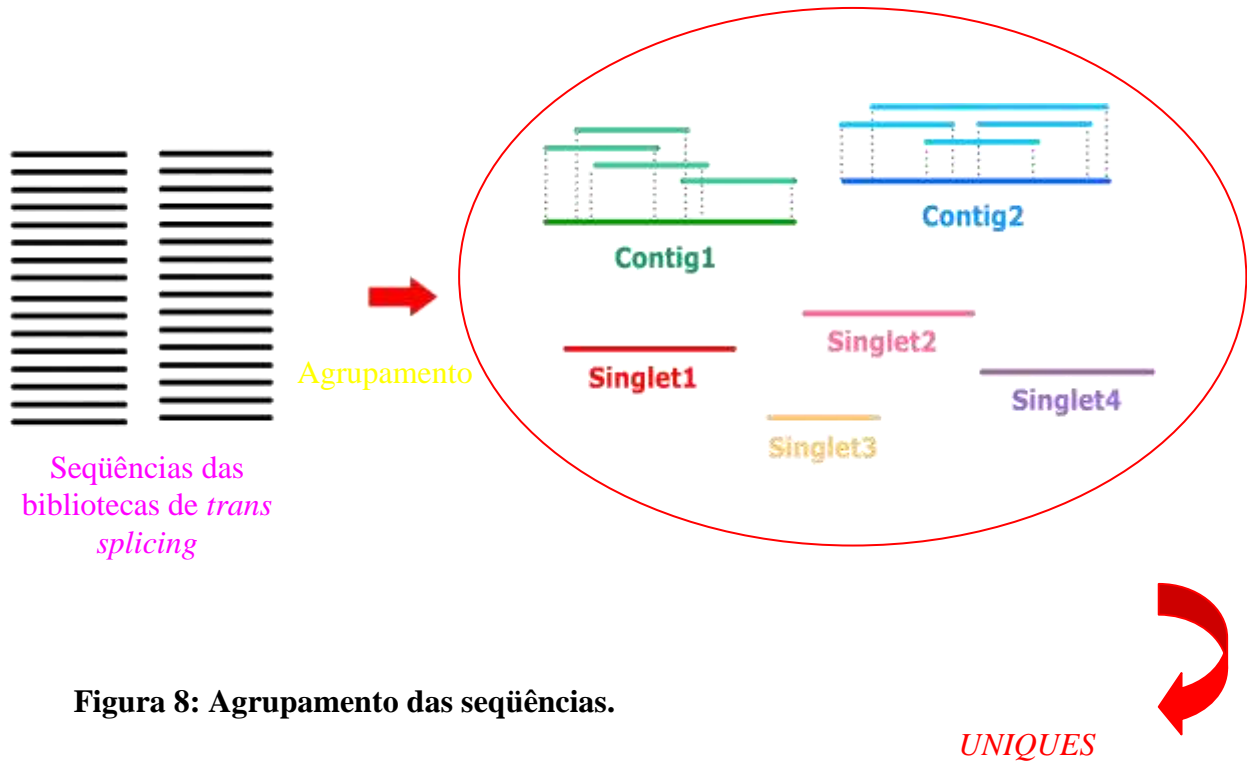


Figura 8: Agrupamento das seqüências.

Um outro aspecto essencial para a análise de seqüências é a construção dos bancos de dados. Os bancos de dados em biologia molecular são importantes principalmente para proporcionar à comunidade científica uma forma de tornar os dados produzidos em todo o mundo mais acessíveis (<http://ncbi.nlm.nih.gov/About?primer/bioinformatics.html>). Com o advento do sequenciamento do DNA e, principalmente a partir da década de 90, do sequenciamento em larga escala, foi necessária a construção de bancos de dados robustos para abrigar tamanho número de seqüências que passaram a ser geradas.

Tendo em mãos os bancos de dados de seqüências, ferramentas bioinformáticas estão disponíveis para o alinhamento de seqüências. Este consiste no processo de comparar duas seqüências (de nucleotídeos ou proteínas) de forma a se observar seu grau de identidade para descobrir o nível de similaridade entre as seqüências. Dessa forma, pode-se inferir, ou não, a uma delas, alguma propriedade já conhecida da outra (Prosdocimi *et al.*, 2002). Foram utilizadas para a busca de seqüências geradas neste projeto, o programa de alinhamento local de seqüências (utiliza pequenas regiões entre duas seqüências para fazer sua comparação) BLAST (Basic Local Alignment Search Tool) (Altshul, *et al.*, 1990), encontrado em <http://www.ncbi.nlm.nih.gov/BLAST/>. Esse programa compreende um conjunto de algoritmos de comparação de seqüências montado de forma a explorar toda a informação contida em bases de dados de DNA e proteínas (http://www.ncbi.nlm.nih.gov/BLAST/blast_overview.html). Este alinhamento funciona da seguinte forma: há a procura de uma região de similaridade entre duas seqüências. Quando essa região é encontrada, são dadas pontuações para similaridades e retiradas pontuações das diferenças, abertura de buracos e extensão destes buracos. A somatória dos pontos desse alinhamento é chamado de score do alinhamento (*score bits*). Existe também o valor E (e-value), que representa o número de alinhamentos com escores iguais que seria de se esperar acaso, numa base de dados do tamanho da utilizada. Assim, quanto melhor for o alinhamento, maior deverá ser o escore e menor o valor E. (<http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/glossary2.html>). O BLAST apresenta diferentes subprogramas que devem ser utilizados de acordo com o tipo de seqüência de entrada e os bancos de dados que se deseja pesquisar.

Neste trabalho, em que utilizamos seqüências de nucleotídeos como entrada, executamos os programas BLASTn (para busca no banco de dados de nucleotídeos), BLASTx (para a busca no banco de dados de proteínas). Nesse último, a seqüência de entrada é traduzida nas seis fases de leitura, produzindo seis seqüências de proteínas, que são, então, utilizadas como entrada para a execução de um BLAST no banco de dados de proteínas (<http://www.ncbi.nlm.nih.gov/blast/html/BLASThomehelp.html>). Executamos também o BLAST contra as seqüências do banco de dados SWISSPROT (banco de dados curado de proteínas, que está contido no banco de dados de proteínas em geral) e também, o BLAST contra os genes completos de *S. mansoni* já depositados. Estes BLASTs serviram para que pudéssemos anotar o gene, ou seja, a partir de suas similaridades com outras seqüências, inferir suas possíveis identidades e funções.

2. Objetivos

2.1. Objetivo Geral

Identificar e anotar genes que sofrem *trans splicing*, através da construção de bibliotecas de cDNAs contendo o *spliced leader* de algumas fases do ciclo de vida do parasita *Schistosoma mansoni* e do sequenciamento em larga escala.

2.2. Objetivos específicos

- ❖ Construção de bibliotecas de cDNA de várias fases do ciclo de vida, contendo transcritos que se iniciam pelo *spliced leader*.
- ❖ Sequenciamento em larga escala, identificação e anotação das ESTs geradas através da utilização de ferramentas de bioinformática (Cross Match, PHRED, CAP3, BLASTn, BLASTx).
- ❖ Classificação das seqüências de clones completos baseado nas categorias dos COG/KOG.
- ❖ Comparação preliminar do perfil de genes regulados por *trans splicing* durante o desenvolvimento do parasita.

3. Materiais e Métodos

3.1. Materiais Biológicos

Os materiais biológicos foram obtidos no GIDE-UFMG cedidos pela Prof. Débora Côrrea do Departamento de Parasitologia do ICB, exceto as amostras da forma que se assemelha a fase de esquistossômulo (obtido por cultura de 7 dias) do verme, que foi gentilmente cedido pelo Centro de Pesquisa René Rachou (CPqRR-FIOCRUZ).

Os vermes adultos da cepa LE, foram obtidos usando o protocolo de rotina do GIDE-UFMG para a perfusão, foram lavados em salina 0,9% estéril e congelados a -80 graus. Através desta extração foram obtidos cerca de 1g de material.

Os ovos dos vermes foram retirados dos intestinos dos camundongos SWISS com 48 dias de infecção. Os intestinos foram lavados, colocados em salina 1,7% por 12 horas e foram então triturados. A separação dos tecidos e ovos foram feitas em um tecido e o filtrado obtido foi deixado para decantar, O precipitado foi lavado várias vezes com salina 1,7% e congelado a -80°C . Foram extraídos a partir desta metodologia 0,3 g de ovos.

3.2. Extração de RNA

Para a extração dos mRNAs dos vermes adultos, da forma que se assemelha à fase de esquistossômulo e dos ovos, foram utilizadas esferas magnéticas Dynabeads Oligo (dT) 25 magnéticas (Dyna), com o protocolo a seguir:

Para o tratamento das *beads*: 250 μl de *beads* em tampão (250 mM Tris-HCl, pH 7.5 20 mM EDTA, 0.1% Tween-20, 0.02% Azida de Sódio) foram transferidas para um novo tubo e colocados num suporte com ímã. O sobrenadante foi removido. O tubo foi retirado do suporte e as *beads* lavadas com 250 μl SSC (0,15 M de Cloreto de Sódio e 0,015 M Citrato de Sódio). Este

processo foi repetido por três vezes, sendo que na última o sobrenadante foi retirado e a amostra lisada adicionada às *beads*.

Para a lise das amostras: Foram colocados 50 mg de vermes em 1ml de TRIZOL (Invitrogen), para evitar a degradação da amostra por RNAses, iniciar a lise celular e promover a dissociação dos complexos de nucleoproteínas. O material foi rompido em sonicador Sonic Dismembrator modelo 500 (Fisher Scientific) três vezes por 30 segundos, com intervalos de 30 segundos, mantido sempre a 4° C. As amostras foram incubadas por 5 minutos em temperatura ambiente. Para a separação em duas fases, uma contendo o RNA e outra contendo DNA e proteínas, foram adicionados 200µl de clorofórmio às amostras, que foram agitadas e incubadas à temperatura ambiente por 2 minutos. Após este período, as amostras foram centrifugadas em microcentrifuga 5402 (Eppendorf) a 12000g, por 15 minutos a 4°C. A fase superior, contendo o RNA, foi transferida para um novo tubo com as *beads* previamente tratadas.

Para extração do mRNA: O tubo contendo as *beads* e a fase superior do lisado foi vigorosamente agitado por 4 minutos em temperatura ambiente e posteriormente colocado no suporte magnético por 3 minutos, para que as *beads* com o mRNA capturado se separassem do resto do lisado. As *beads* foram lavadas três vezes com 0,7ml de SSC e uma vez com o tampão da enzima transcriptase reversa (Invitrogen). As *beads* foram então, ressuspensas em 20µl de tampão 1x da transcriptase reversa (Invitrogen) e usadas diretamente, sem eluição prévia, na reação de RT-PCR.

Foi feita também a extração dos RNAs das formas semelhantes à fase de esquistossômulos, ovos utilizando o kit RNAagents (Promega), seguindo o protocolo do fabricante. Sucintamente, o sistema inclui cristais tiocianato de guanidina e tampão CSB (Citrato/Sarcosina/Mercaptoetanol) na solução desnaturante e também uma solução de Fenol:Cloroformio: Álcool Isoamílico que remove DNA cromossomal contaminante das amostras.

Um terceiro método foi utilizado a extração por fenol ácido (TRIZOL), na tentativa de extração de RNA de ovos, sempre seguindo as instruções do fabricante. TRIZOL é uma solução mono-fásica de fenol e isotiocianato de guanidina, onde foram homogenizadas e lisadas as

amostras, que age em um método de um único passo para extração do RNA. Após o rompimento das células e dissolução dos componentes celulares, o RNA é liberado intacto. É adicionado clorofórmio seguido de centrifugação, o que separa a solução em fase aquosa e orgânica. Após a transferência da fase aquosa, o RNA é recuperado por precipitação com álcool isopropílico.

Os mRNAs de fêmeas e machos haviam sido previamente extraídos no nosso laboratório, e foram provenientes de vermes adultos que foram separados após a perfusão dos camundongos, ou seja, a infecção dos camundongos foi bissexual.

Os RNAs totais e mensageiros de fêmea e de macho foram dosados em espectrofotômetro (Shimadzu), nos comprimentos de ondas de 260nm (é o comprimento de onda absorvido pelas purinas e pirimidinas) e 280nm (absorvido por tirosina e triptofano presentes nas proteínas). A razão 260/280 foi feita para se avaliar o grau de contaminação de proteína nas amostras, sendo que a razão ideal é de 1,8-2,0. Os mRNAs extraídos com *beads* foram diretamente submetidos a transcrição reversa e posteriormente a PCR específico para amplificação dos transcritos que sofrem *trans splicing*, sem prévia dosagem, pois as *beads* interferem na dosagem.

3.3. Transcrição reversa

Para a transcrição reversa foi utilizada a SuperScript II Reverse Transcriptase (Invitrogen). Nas reações de síntese da primeira fita de cDNA, feitas a partir de RNA total, foram utilizados de 2 a 3,5 µg de RNA total para as amostras da forma que se assemelha a esquistossômulo, ovos e também, para RNA mensageiro de fêmeas e machos. Para *beads* contendo mRNA de vermes adultos foi adicionado 1µl das *beads* à reação. Aos RNAs foram adicionados 1µl de dNTP (10mM cada), 1µl de iniciador Oligo dT (500 µg/ml), neste havia uma cauda de um outro iniciador (a seqüência será descrita posteriormente), que foi usado também na reação de PCR de amplificação dos transcritos, em um volume final de 12µl.

A mistura foi aquecida a 65°C por 5 minutos, resfriada em gelo e em seguida foram adicionados 4µl de tampão da primeira fita 5x (250mM Tris-HCl, pH 8,3; 375 mM KCl; 15 mM

MgCl₂), 2 µl de DTT 0,1 M e 40 U de RNaseOUT (Invitrogen). A mistura foi incubada a 42°C por 2 minutos. Foram adicionadas 200U de SuperScript II RT e a mistura incubada a 42°C por 50 minutos e inativada por aquecimento a 70°C por 15 minutos.

3.4. Amplificação dos transcritos que sofrem *trans splicing*

Foi utilizada a estratégia de amplificação completa dos cDNAs produzidos por *trans splicing*, através da utilização da seqüência do SL como iniciador e de uma cauda ligada ao oligo dT (Brehm *et al.*, 2000).

Para a produção dos cDNAs foram desenhados dois iniciadores. Um deles foi construído utilizando a seqüência do *spliced leader* de *S. mansoni*, descrito por Davis e colaboradores, 1995:

5'AACCGTCACGGTTTTACTCTTGTGATTTGTTGCATG3'

Para o outro iniciador, construímos um oligo dT com uma cauda que serviria, tanto na síntese da primeira fita de cDNA, como na reação de PCR. Esta cauda funcionaria estabilizando o iniciador e se anelaria aos produtos no segundo ciclo de PCR. (Figura 9)

5'CGGTATTTTCAGTCGGTGTCAAACCTTTTTTTTTTTTTTTTTTTTTTV3' V=A,G,C

Para a reação de PCR foram utilizados 4µl de cDNA

Para o volume final de 40 µl

O programa utilizado primeiramente, foi:

- desnaturação a 95°C por 5 minutos
- 30 ciclos de: desnaturação a 95°C por 1 minuto
anelamento dos iniciadores a 55°C por 1 minuto
extensão a 72°C por 1 minuto e 30 segundos

Posteriormente, para evitar alinhamentos espúrios, o programa utilizado passou a ser:

- 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
- 19 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

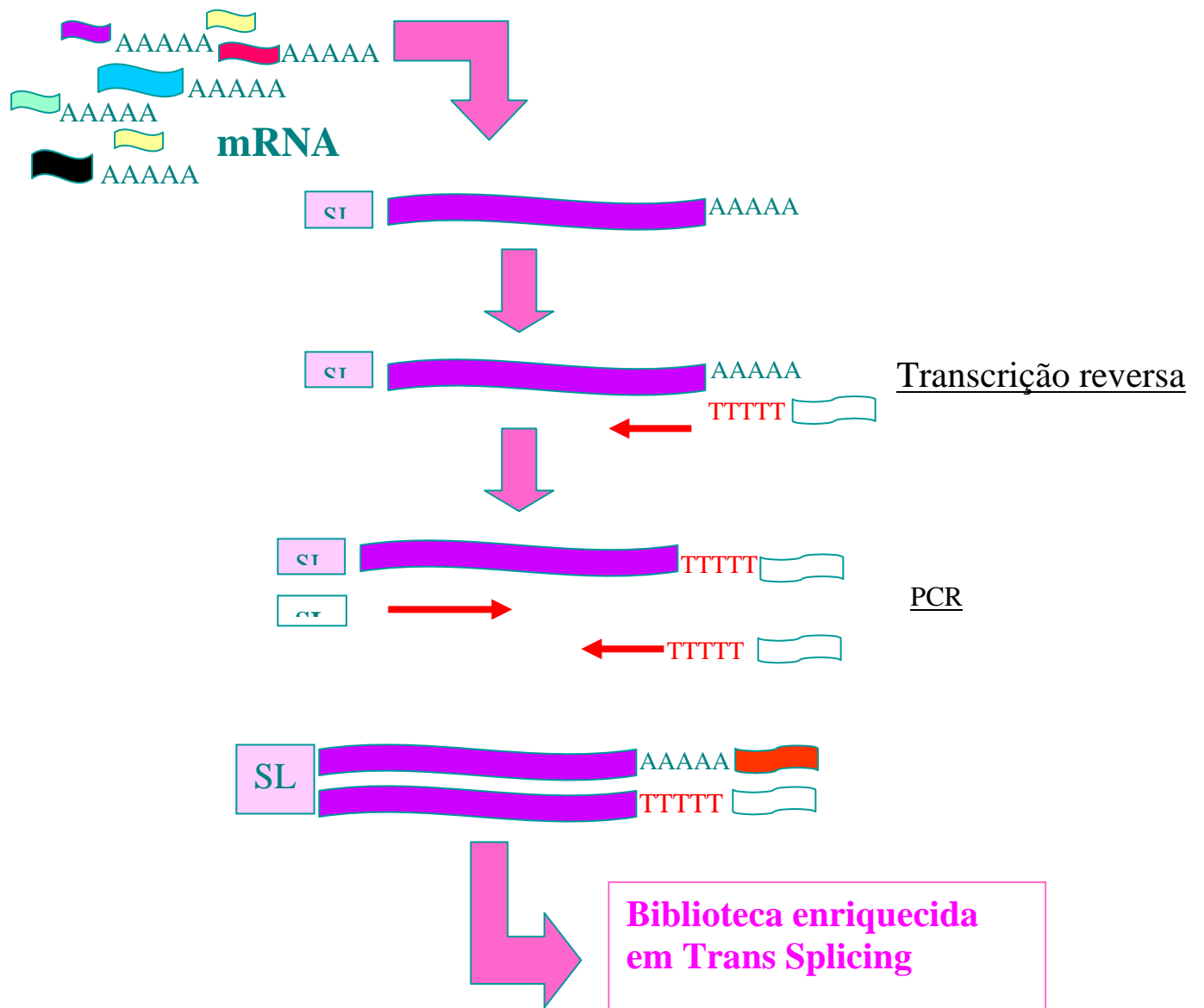


FIGURA 9: Esquema da amplificação dos transcritos produzidos por *trans splicing*, e utilizados para gerar as bibliotecas enriquecidas.

3.5. Seleção dos cDNAs por tamanho

Após observar o produto de uma pequena alíquota da reação de PCR por eletroforese em gel de agarose 1% corado com brometo de etídio, 40µl de todos os produtos de PCR foram submetidos à seleção de tamanho. A seleção foi feita de duas maneiras: (1) utilizando Polietilenoglicol 8000 15% (esta concentração se mostrou eficiente para separar fragmentos maiores que 400 pares de base, segundo padronização feita em nosso laboratório pela aluna de doutorado Simone Santos) para as amostras de fêmeas, machos, vermes adultos, ou (2) utilizando separação em gel de agarose, protocolo este, que foi utilizado para os produtos da fase similar à fase de esquistossômulo e ovo.

3.5.1. Seleção por Polietilenoglicol (PEG) 8000 15%

Foram adicionados 15µl de PEG15% e NaCl 2,5 M em tubos contendo 15µl dos cDNAs de fêmeas, machos e adultos mistos. As amostras foram homogeneizadas e incubadas em banho-maria a 37°C por 15 minutos. Posteriormente as amostras foram centrifugadas a 13000 rpm por 15 minutos a 4°C. O sobrenadante, que continha os insertos menores que 300 pares de bases, foi retirado. Para precipitar os produtos maiores, foram adicionados 125µl de etanol 80% gelado. As amostras foram centrifugadas a 13000 rpm, por 15 minutos a 4°C e o sobrenadante foi novamente descartado. Os passos de precipitação foram repetidos por mais uma vez e, após o descarte do sobrenadante, os tubos foram deixados em temperatura ambiente para secar. O DNA foi ressuspenso em 25µl de água milliQ. Após o protocolo acima, foi feita a eletroforese em gel de agarose 1% corado com brometo de etídio, para que os resultados das seleções dos cDNAs fossem visualizados.

3.5.2. Seleção por eletroforese em gel de agarose 1%

Em gel de agarose 1% foram aplicados 40µl dos produtos de PCR de ovo e fase semelhante à fase de esquistossômulo, juntamente ao padrão de peso molecular 1Kb plus (Invitrogen). Após a separação eletroforética, o gel foi observado em transluminador de luz UV de ondas longas. As partes do gel que continham cDNAs maiores ou iguais a 500 pares de bases foram cortados. As

partes cortadas foram transferidas para um gel de agarose de baixo ponto de fusão 1%, de forma que os cDNAs menores ficassem posicionados na parte superior do gel e os maiores na região inferior. Quando os fragmentos passavam para o gel de agarose de baixo ponto de fusão estes formavam uma banda compacta, tornando possível a sua purificação posterior, após remoção desta região do gel e digestão da agarose utilizando 1 unidade de β -agarase para 100 μ l de amostra, como se segue: (Figura 10)

As bandas retiradas dos géis foram colocadas em um banho a 65°C por 5 minutos, o volume de agarose foi medido, e a ele foi adicionada a mesma quantidade de TE. As amostras foram incubadas por 15 minutos a 65°C, e a 42°C por 5 minutos após a adição de um décimo do volume do tampão da enzima (10 mM Bis-Tris-HCl, 1 mM EDTA pH 6,5 25°C). Foi adicionada 1 unidade de β -agarase (New England Biolabs) para cada 100 μ l de amostra. A mistura foi deixada a 42°C por duas horas para digestão da agarose. Após a digestão, para ver resquícios de agarose sólida, as amostras foram deixadas no freezer -20°C por 15 minutos e centrifugadas a 14000rpm por 10 minutos em centrífuga 5402 (Eppendorf). Os cDNAs foram precipitados utilizando um décimo do volume da amostra de acetato de sódio 5M e dois volumes de etanol absoluto e foram deixadas por doze horas no freezer -20°C. As amostras foram centrifugadas por 20 minutos na velocidade de 14000 rpm a 4°C. O sobrenadante foi descartado por inversão. Foram adicionados 500 μ l de etanol 70% gelado aos tubos contendo o DNA precipitado e estes foram centrifugados por 5 minutos a 14000rpm. Novamente o sobrenadante foi descartado por inversão e os tubos deixados em temperatura ambiente para secar. O DNA foi ressuspendido em 20 μ l de água milliQ.

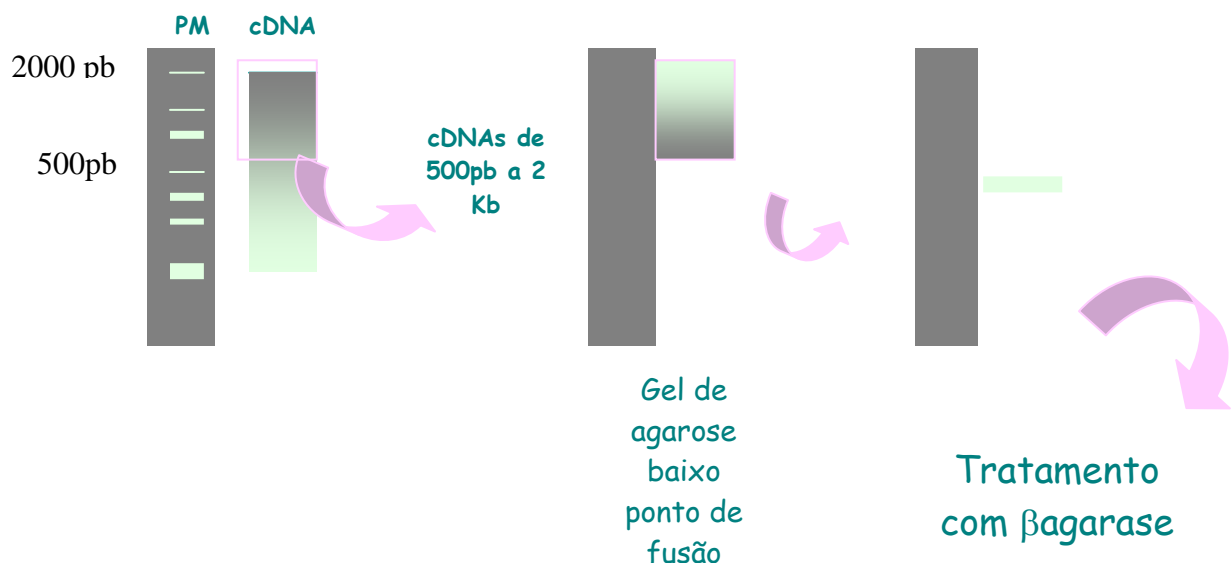


FIGURA 10: Processo de seleção de tamanho dos cDNAs por gel de agarose.

3.6. Ligação dos cDNAs aos vetores

Após a seleção, os cDNAs provenientes de vermes adultos, fêmeas e machos, foram ligados ao vetor pGEM utilizando o kit T-easy Vector System (Promega), conforme instruções do fabricante. Os cDNAs selecionados de ovos e da fase que se assemelha à fase de esquistossômulo foram ligados no vetor pCR2.1 utilizando o kit TA cloning (Invitrogen), de acordo com as especificações do fabricante (Figura 11). Estes dois vetores são T-vetores, ou seja, são linearizados com uma enzima que corta o DNA em extremidades cegas e posteriormente é adicionado um nucleotídeo de timina nas extremidades 3' destes vetores linearizados. Desta forma, é possível ligar produtos de PCR, pois estes possuem em suas extremidades 3' um nucleotídeo adenina, naturalmente adicionado pela Taq DNA polimerase. As ligações foram feitas utilizando 2µl dos cDNAs purificados.

3.7. Transformação de bactérias *Escherichia coli* DH5α eletrocompetentes e seleção dos clones

Os plasmídeos recombinantes foram utilizados para transformar bactérias hospedeiras *E. coli* DH5α eletrocompetentes. Estas bactérias (50µl) foram incubadas com 2µl de cada produto de ligação por 5 minutos, em gelo. Como controle positivo da transformação foi feita uma incubação com 10ng do vetor pUC18 intacto. Como controle negativo da transformação, as *E. coli* DH5α foram incubadas sem plasmídeos. Após a incubação, as bactérias foram submetidas a uma descarga elétrica de 3.400 volts por 2,5 milisegundos em eletroporador Micro Pulser™ (BIORAD). Após o choque elétrico, foram adicionados 250µl de meio de cultura 2XYT (extrato de levedura 1%, peptona 1,6%, NaCl 0,5%, pH 7,0) às bactérias que foram incubadas por 1 hora a 37°C sob agitação. As bactérias transformadas foram plaqueadas em meio 2XYT ágar contendo 100mg/ml ampicilina e incubadas a 37°C por 16 horas.

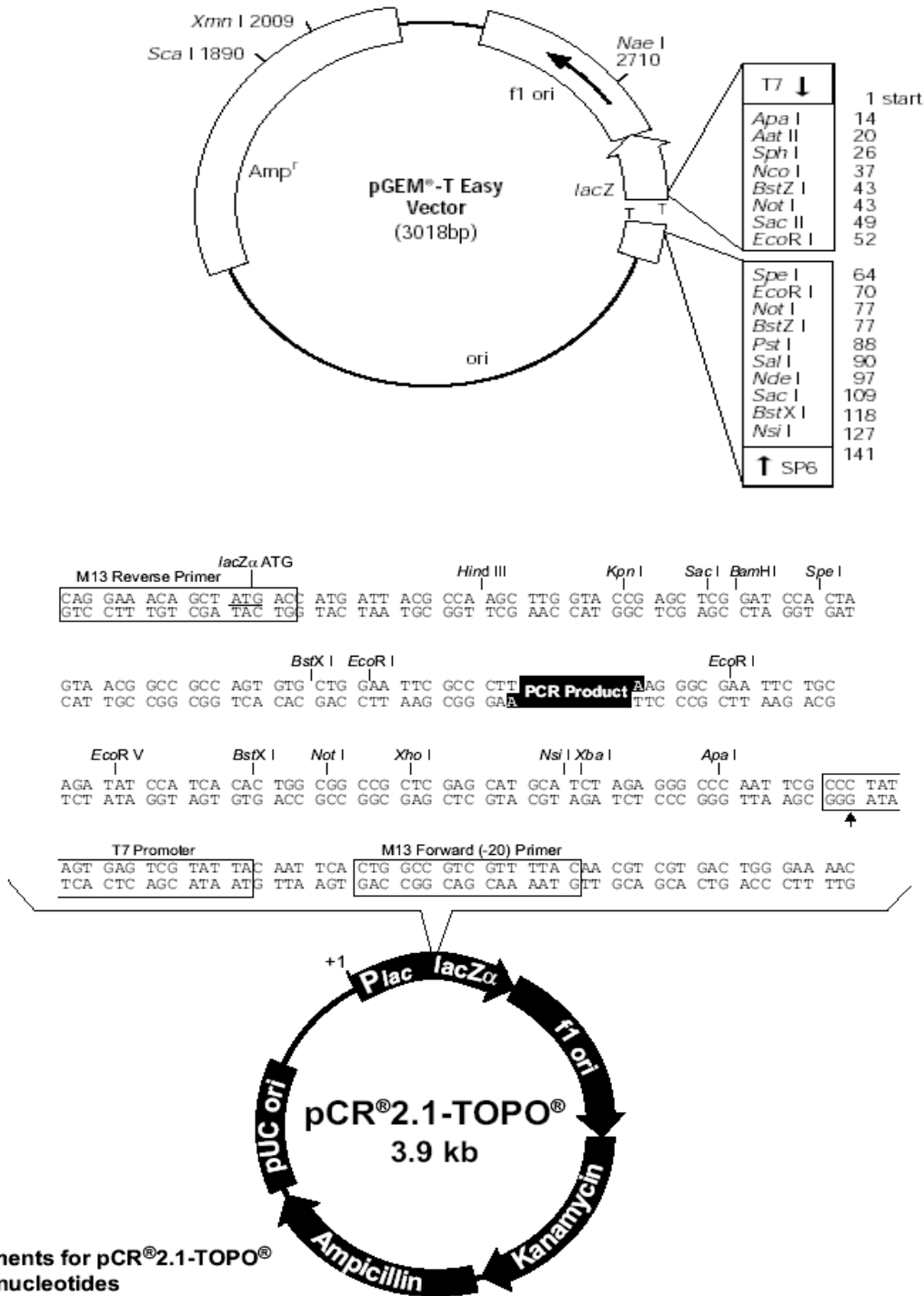


Figura 11: Vetores utilizados para clonagem dos insertos. pGEM (Promega) e pCR2.1 (Invitrogen)

3.8. Seleção dos clones

Para selecionar clones que continham insertos grandes e examinar a qualidade das bibliotecas, os clones de bactérias recombinantes crescidas em placas foram submetidos a reações de PCR de colônia. Nesta reação foram utilizados 0,2 μM do par de iniciadores M13 For-40 e M13 Rev, 200 μM de cada dNTP, 0,5 U de Taq DNA polimerase (Phoneutria), tampão de reação 1B (Phoneutria) (Tris-HCL 10mM pH8,4, KCL 50mM, 0,1% Triton X-100, MgCl_2 1,5mM), em um volume de 10 μl . Os clones escolhidos foram isolados com palito esterelizado e inoculados em placas de 96 poços contendo a mistura citada acima. Na reação de PCR foi utilizado o programa abaixo:

- Desnaturaçãso a 95°C por 5 minutos.
- 25 ciclos de: Desnaturaçãso a 95°C por 5 minutos
anelamento dos iniciadores a 55°C por 1 minuto
extensão a 72° C por 1 minuto
- extensão a 72° C por 8 minutos

A amplificação foi confirmada em gel de agarose 1% corado com brometo de etídio.

3.9. Obtenção dos plasmídeos recombinantes

Os clones de bactéris que continham plasmídeos com insertos de tamanho desejado foram coletados e colocados em placas *deep-well* com 96 poços, contendo 1,2ml de meio 2XYT com ampicilina. As placas foram agitadas por 22 horas, 180rpm a 37° C. Após o tempo necessário para que as bactérias crescessem, 100 μl destas foram colocados em uma placa de 96 *wells* fundo U contendo 100 μl de glicerol 50% e congeladas em um freezer -80° C. O restante do meio contendo as bactérias foi usado para a extração de plasmídeos (mini-prep), seguindo o protocolo de extração utilizado no Projeto Genoma Brasileiro, descrito abaixo.

As placas *deep well* foram centrifugadas, na centrífuga Eppendorf modelo 5804R, por 6 min. a 4000 rpm, para sedimentar as células. Os sobrenadantes foram descartados e as placas

invertidas sobre papel absorvente. Foram, então, adicionados 240µl de GET (9g de glicose, 20ml de EDTA 0,5M pH 8,0, 26ml de Tris HCl 1M pH 7,4, q.s.p. 1000ml de H₂O) em cada poço, e estes foram agitados em um vórtex por 2 min para ressuspender as células e depois centrifugados por 9 min. a 4000 rpm. Posteriormente, os sobrenadantes foram descartados e as placas invertidas em papel absorvente por 5 min. Foram adicionados 80µl de GET e as placas novamente agitadas. Foram adicionados 2,5µl de RNase (10mg/ml) em cada poço de microplacas de 250µl de propileno de fundo redondo (tipo ELISA) e toda a suspensão de células transferida para a mesma. Então, foram adicionados 80µl de NaOH 0,2N/ SDS 1% a cada poço. Misturou-se 30 vezes por inversão e incubou-se por 10 min em temperatura ambiente. Em seguida, foram adicionados, a cada poço, 80µl de acetato de potássio 3M que foram misturados por inversão 30 vezes. Incubou-se por 10min em temperatura ambiente e por 30 min em estufa a 90° C. As placas foram esfriadas em gelo picado por 10 min e centrifugadas por 9 min a 4000 rpm. Transferiu-se todo o volume dos poços das placas de fundo U, para placas filtro que foram fixadas em outras de fundo V. Estas foram centrifugadas por 6 min a 4000rpm, com o objetivo de filtrar detritos celulares e DNA genômico. As placas filtro foram descartadas e foram adicionados 100µl de isopropanol ao filtrado. Misturou-se 30 vezes por inversão e centrifugou-se por 45 min a 4000rpm. Os sobrenadantes foram descartados, e ao precipitado adicionados 190µl de etanol 70% gelado. Centrifugou-se por 5 min a 4000rpm e descartou-se os sobrenadantes. As placas foram invertidas sobre papel absorvente e centrifugadas rapidamente a 900rpm. Após 60 min secando a temperatura ambiente, foram adicionados 60µl de água milliQ e as placas deixadas à temperatura ambiente por 18h.

A quantificação do DNA plasmidial purificado foi feita através de eletroforese em gel de agarose 1% corado com brometo de etídio em TAE 0,5X (20mM Tris Acetato, pH 8,3 and 0,5mM EDTA) por 30 min a 80 volts.

3.10. Seqüenciamento das extremidades dos insertos

Na reação de seqüenciamento das extremidades dos insertos de cDNA, nos sentidos direto e reverso, foram utilizados 10pmoles de iniciador, 4µl de Dyanamic ET – dye terminator Kit –

Mega Bace (GE Healthcare), 300ng de DNA, em um volume final de 10µl. A placa foi colocada no termociclador (Eppendorf Mastercycler) por 1 ciclo a 95° C para desnaturação, 1 ciclo a 50° C para anelamento, 1 ciclo a 60° C para extensão. Estes ciclos foram repetidos por 29 vezes.

Posteriormente foi realizada a precipitação do DNA (“clean-up”), utilizando o protocolo a seguir, também utilizado no Projeto Genoma Brasileiro: Foram adicionados 1µl de acetato de amônio 7M em cada poço e também 30µl de etanol absoluto. Agitou-se por 10 segundos e as placas foram deixadas a temperatura ambiente por 20 min em ambiente escuro (a fluorescência é dos nucleotídeos é fotossensível). As placas foram centrifugadas por 45 min a 4000rpm em centrífuga 5804 R (Eppendorf), e os sobrenadantes desprezados. Foram adicionados 100µl de etanol 70% por poço e as amostras foram centrifugadas novamente por 10 min a 4000rpm. Novamente os sobrenadantes foram descartados, as placas foram invertidas sobre papel absorvente e centrifugadas rapidamente a 800rpm. Deixou-se as placas em temperatura ambiente, até que o etanol secasse completamente. Foram adicionados 10µl de Loading Solution (GE Healthcare) por poço, as amostras foram agitadas por 2 min e rapidamente centrifugadas; as placas estavam prontas para serem aplicadas no Sequenciador Automático Mega BACE 1000 (GE Healthcare). O sequenciamento durava cerca de 4 horas, utilizando 2kv para a injeção das amostras por 60 segundos e 6kv para a corrida.

3.11. Análise Bioinformática

As seqüências obtidas foram submetidas à análise bioinformática que contou com a colaboração do especialista em bioinformática e estudante de doutorado em Bioinformática Francisco Prosdocimi.

Primeiramente, vale ressaltar que apenas as seqüências maiores ou iguais a 100 bases com qualidade PHRED >8 foram aceitas, ou seja, o PHRED 8 com o algoritmo TRIM-alt é calibrado para aceitar 16% de erro (Prosdocimi *et al.*, 2004). As seqüências de saída do programa PHRED foram submetidas ao programa Cross Match. Neste programa foram inseridas as seqüências dos vetores pGEM, pCR2.1, as seqüências de poliA e T, além, da seqüência do *spliced leader* de *S.*

mansoni, pois a presença desta seqüência mascararia os resultados de BLAST que seriam feitos posteriormente.

As seqüências devidamente mascaradas foram submetidas ao programa CAP3, que compara as seqüências formando *contigs* com aquelas que apresentam um certo nível de similaridade pré-definido, e os parâmetros utilizados foram: $-O=40$ e $-S=800$; que indica que para se formar um contig, tem que haver entre as seqüências no mínimo um alinhamento de 40 nucleotídeos com um escore mínimo de 800 (o escore é formado de valores agregados aos *matches*, *mismatches* e *gaps*).

Os *uniques* gerados foram submetidos à pesquisa de similaridade por BLAST e para estes programas usamos um valor de corte de $E < 10^{-6}$. As análises utilizadas foram as seguintes:

- ❖ BLASTx e BLASTn, que comparam as seqüências obtidas contra os bancos de dados não-redundantes de proteínas e nucleotídeos do National Center for Biotechnology Information (NCBI), contendo seqüências de todos os organismos presentes no GenBank.
- ❖ BLAST contra o banco de dados curado de proteínas Swissprot, que por ser curado é mais confiável para anotação.
- ❖ BLAST contra as seqüências gênicas completas de *S. mansoni* depositadas no GenBank.
- ❖ BLAST contra as seqüências produzidas pela Rede Genoma de Minas Gerais e contra as seqüências já depositadas no dbEST, para que pudéssemos comparar qual a contribuição dada por nosso trabalho na descoberta de novos genes do parasita.
- ❖ BLAST contra seqüências depositadas nos bancos de dados COG e KOG (Cluster of Orthologous Groups), que são bancos de dados constituídos de famílias gênicas conservadas presentes em pelo menos três genomas de organismos diferentes (Tatusov *et al.*, 1997), do grupo dos procariotos (COG) e eucariotos (KOG). Este procedimento foi feito para que fosse possível inferir categorias funcionais aos nossos *uniques*. (<http://www.ncbi.nlm.nih.gov/COG>)

3.11.1. Classificação dos *uniques* produzidos por *trans splicing* em categorias funcionais

Utilizamos as categorias funcionais presentes nos COGs (4 categorias e 18 sub-categorias) para agregar possíveis funções aos nossos *uniques* que apresentaram similaridade com seqüências presentes neste banco de dados, após o BLAST.

As categorias presentes nos COGs estão apresentadas na Tabela 1:

Processamento e Armazenamento de Informações	
J	Tradução, estrutura ribossomal e biogênese
K	Transcrição
L	Replicação, recombinação e reparo de DNA
Processos celulares	
D	Divisão celular e particionamento cromossômico
O	Modificação pós-traducional, <i>turnover</i> de proteínas e chaperones
M	Biogênese do envelope celular, membrana externa
N	Mobilidade celular e secreção
P	Metabolismo e transporte de íons inorgânicos
T	Mecanismo de transdução de sinais
Metabolismo	
C	Produção e converção de energia
G	Metabolismo e transporte de carboidratos
E	Metabolismo e transporte de aminoácidos
F	Metabolismo e transporte de nucleotídeos
H	Metabolismo de coenzimas
I	Metabolismo de lipídeos
Q	Biossíntese, catabolismo e transporte de metabólitos secundários
Pobremente caracterizados	
R	Somente predição de função geral
S	Função desconhecida

Quadro 1: Categorias funcionais dos COGs. O projeto COG classifica os genes em 4 categorias funcionais principais, sendo cada uma delas dividida em subgrupos.

3.13.2. Anotação Manual dos *Uniques*

A anotação manual dos *uniques* gerados foi realizada através da observação dos resultados obtidos nas pesquisas de homologia. Para isso estabelecemos alguns critérios:

- ❖ **Prioridade para os resultados do Swiss-Prot**, pois este é um banco de dados mais bem organizado e curado do que os outros e portanto, utilizamos, sempre que possível, suas anotações.
- ❖ **Alta similaridade com genes ou proteínas de *S. mansoni***, nos casos em que houvessem alta similaridade com genes/proteínas de *S. mansoni*, os *uniques* eram classificados como do próprio organismo, ou seja, *Schistosoma mansoni* [nome do gene ou proteína].
- ❖ **Baixa similaridade com genes ou proteínas de *S. mansoni***, quando estes casos eram observados, estes *uniques* eram classificados: similares a gene/proteínas do verme. Estas ocasiões podem estar indicando a presença de genes duplicados ou membros de famílias gênicas.
- ❖ **Alta similaridade com genes ou proteínas de outros organismos**, havendo um consenso de similaridade entre a mesma proteína ou gene em diversos organismos diferentes, as seqüências eram identificadas como: Putative [nome da proteína em questão]
- ❖ **Similaridade com genes ou proteínas de outros organismos**, na falta de um consenso no resultado do Swiss-Prot, identificamos as seqüências através de seus homólogos em outros organismos. Identificávamos o melhor alinhamento entre os bancos de dados, incluindo as similaridades encontradas no banco de dados KOG.

Estas eram classificadas como:

Similar to [organismo] e [nome do gene ou proteína], ou;

Similar to [nome do gene ou proteína], nos casos de anotação utilizando o KOG.

4. Resultados e Discussão

4.1 Extrações de RNA

Foram feitas várias tentativas de extração com as várias técnicas descritas anteriormente (Materiais e Métodos, item 3.1), para as diversas amostras. As melhores extrações que produziram o material, que foi posteriormente utilizado para a síntese de cDNA, foram as seguintes:

Para vermes adultos foram utilizadas Dynabeads Oligo (dT) 25 magnéticas, estas, portanto, não puderam ser dosadas em espectrofotômetro, pois a presença das *beads* magnéticas interferem na dosagem. Os mRNAs extraídos com *beads* foram diretamente submetidos ao PCR, sem a prévia separação das *beads* magnéticas dos mRNAs.

Para o RNA de vermes machos e fêmeas a quantidade total obtida em uma extração prévia feita por Glória Franco, foi de 134,4 µg/ul e 68,8 µg/ul, respectivamente.

Para o RNA da forma semelhante à fase de esquistossômulos, a técnica que rendeu os melhores resultados foi através da utilização do kit RNAGents, produzindo 207 µg/ul de RNA total.

Para a extração de RNA de ovos, a técnica em que se obteve melhores resultados foi através da utilização de TRIZOL reagente, produzindo 99,1 µg/µl de RNA total.

4.2. Amplificação dos transcritos que sofrem *trans-splicing*

Após a síntese da primeira fita dos cDNAs através da transcrição reversa, as mesmas foram submetidas a PCR para amplificação dos transcritos produzidos por *trans splicing*, utilizando os dois iniciadores descritos no item 3.4 da seção Materiais e Métodos. Em seguida, os produtos de amplificação foram analisados em gel de agarose 1% corado pelo brometo de etídio (Figuras 12 a 15).

Para a construção das bibliotecas de vermes adultos, machos e fêmeas, utilizamos o primeiro protocolo de amplificação descrito no item 3.4 da seção Materiais e Métodos. No entanto,

para as fases de ovo e da forma semelhante à fase de esquistossômulo verificamos que este programa de amplificação estava gerando uma grande quantidade de produtos de tamanho pequeno. Assim, decidimos alterar o protocolo para a realização de um PCR *step down*, seguindo o segundo protocolo descrito no item 3.4 da seção Materiais e Métodos.

4.3. Seleção dos cDNAs por tamanho

Obtivemos uma média de 200.000 clones independentes para todas as bibliotecas.

Após observar o resultado das amplificações em gel de agarose, 40µl dos produtos das reações de PCR foram submetidos à seleção de tamanho, por Polietilenoglicol 8000 15% e o precipitado foi analisado em gel de agarose 1% (Figura 16).

Este método de utilização de PEG 8000 na concentração de 15% se mostrou eficiente na separação dos fragmentos maiores que 400 pares de bases, nas bibliotecas de cDNA de vermes adultos, fêmeas e machos. Para as bibliotecas de cDNA de ovos e da forma que se assemelha à fase de esquistossômulo a seleção de tamanho foi feita em gel de agarose, conforme descrito no item 3.8 na seção Materiais e Métodos.

4.4. Seleção dos clones

Para selecionar os clones que continham insertos maiores para o sequenciamento e analisar a qualidade das bibliotecas, clones de bactérias recombinantes foram escolhidos aleatoriamente e submetidos a reações de PCR de colônia.

Podemos observar que, por exemplo, a biblioteca de cDNA de machos continha insertos de tamanhos diferenciados (Figura 17) Após esta análise as bibliotecas foram liberadas para o seqüenciamento em larga escala.

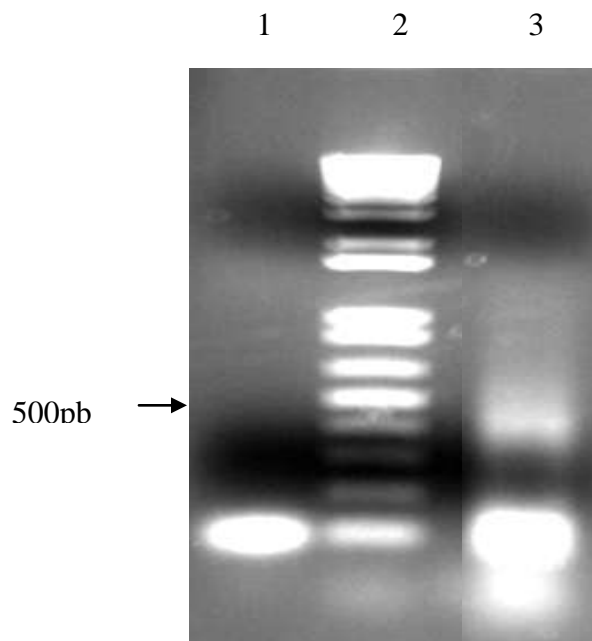


Figura 12: Perfil de amplificação dos transcritos que sofrem *trans splicing* obtido a partir da forma que se assemelha à fase de esquistossômulos. Gel de agarose 1% corado com brometo de etídio. Canaletas: 1- branco da reação de PCR; 2- Padrão de peso molecular 1Kb plus (Invitrogen); 3- amostra de esquistossômulo

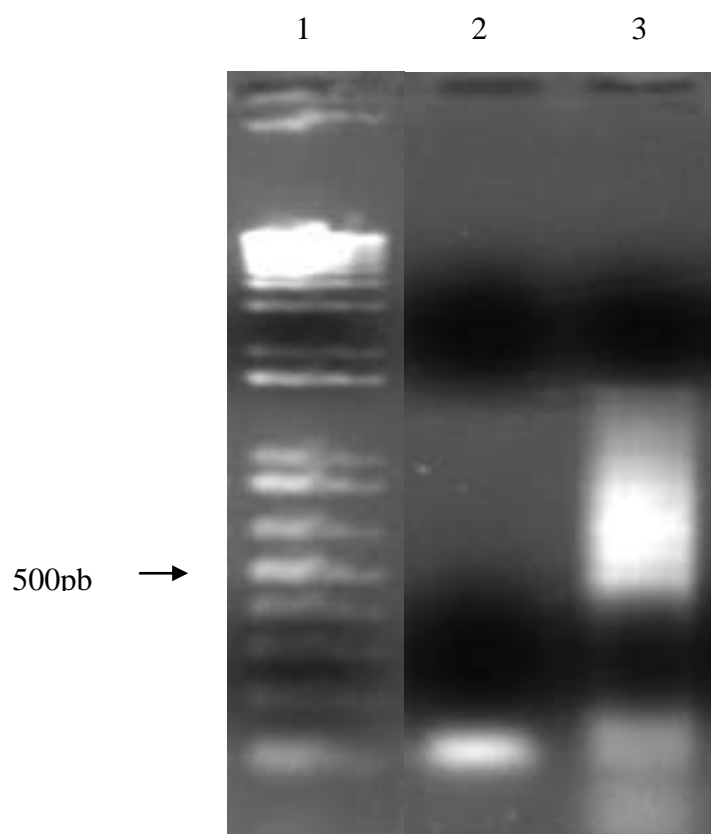


Figura 13: Perfil de amplificação dos transcritos que sofrem *trans splicing* obtidos a partir de RNA de ovos. Gel de agarose 1% corado com brometo de etídio. Canaletas: 1- Padrão de peso molecular 1Kb plus (Invitrogen); 2- branco da reação de PCR; 3- amostra de ovo

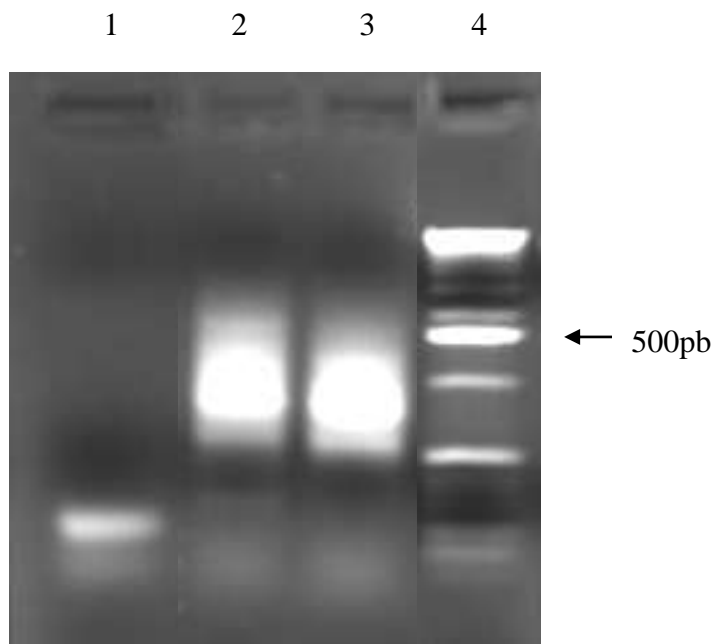


Figura 14: Perfil de amplificação dos transcritos que sofrem *trans splicing* obtidos a partir de RNA de fêmeas e machos. Gel de agarose 1% corado com brometo de etídio. Canaletas 1- branco da reação; 2- amostra de cDNA de fêmea; 3- amostra de cDNA de macho; 4- Padrão de peso molecular 1Kb plus (Invitrogen)

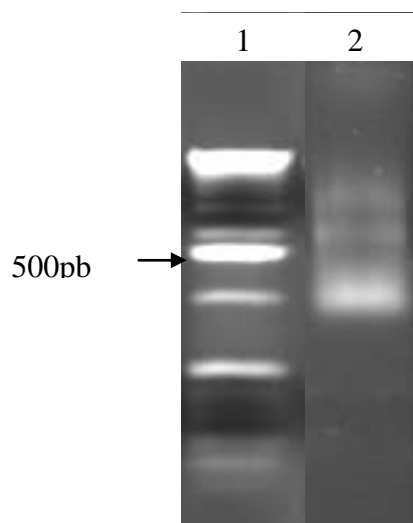


Figura 15 : Perfil de amplificação dos transcritos que sofrem *trans splicing* obtidos a partir de RNA de vermes adultos. Gel de agarose 1% corado com brometo de etídio. Canaletas: 1- Padrão de peso molecular 1Kb plus (Invitrogen); 2- Amostra de vermes adultos

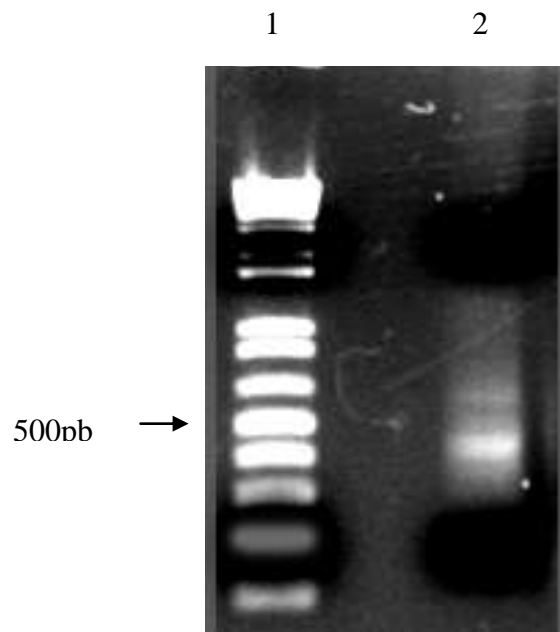


Figura 16 : Perfil de tamanho dos cDNAs após a seleção por PEG 15%. Gel de agarose 1% corado com brometo de etídio. Canaletas: 1- Padrão de peso molecular 1Kb plus (Invitrogen); 2- cDNA de vermes adultos após seleção por tamanho.



Figura 17 : Perfil de PCR de colônia de clones da biblioteca de machos. Gel de agarose 1% corado com brometo de etídio, 108 canaletas. O padrão de tamanho utilizado foi o 1Kb plus (Invitrogen)

4.5. Quantificação do DNA plasmidial

A quantificação do DNA plasmidial purificado foi feita através de eletroforese em gel de agarose 1% corado com brometo de etídio.

A quantificação foi feita para verificar a qualidade e a quantidade dos plasmídios obtidos e se há presença de RNA, o que pode interferir no sequenciamento. Em geral, o protocolo de mini-prep utilizado levou a um rendimento de 150 ng/ μ l de plasmídio (Figura 19). Após a dosagem e análise da qualidade do DNA plasmidial este foi submetido à reação de sequenciamento.

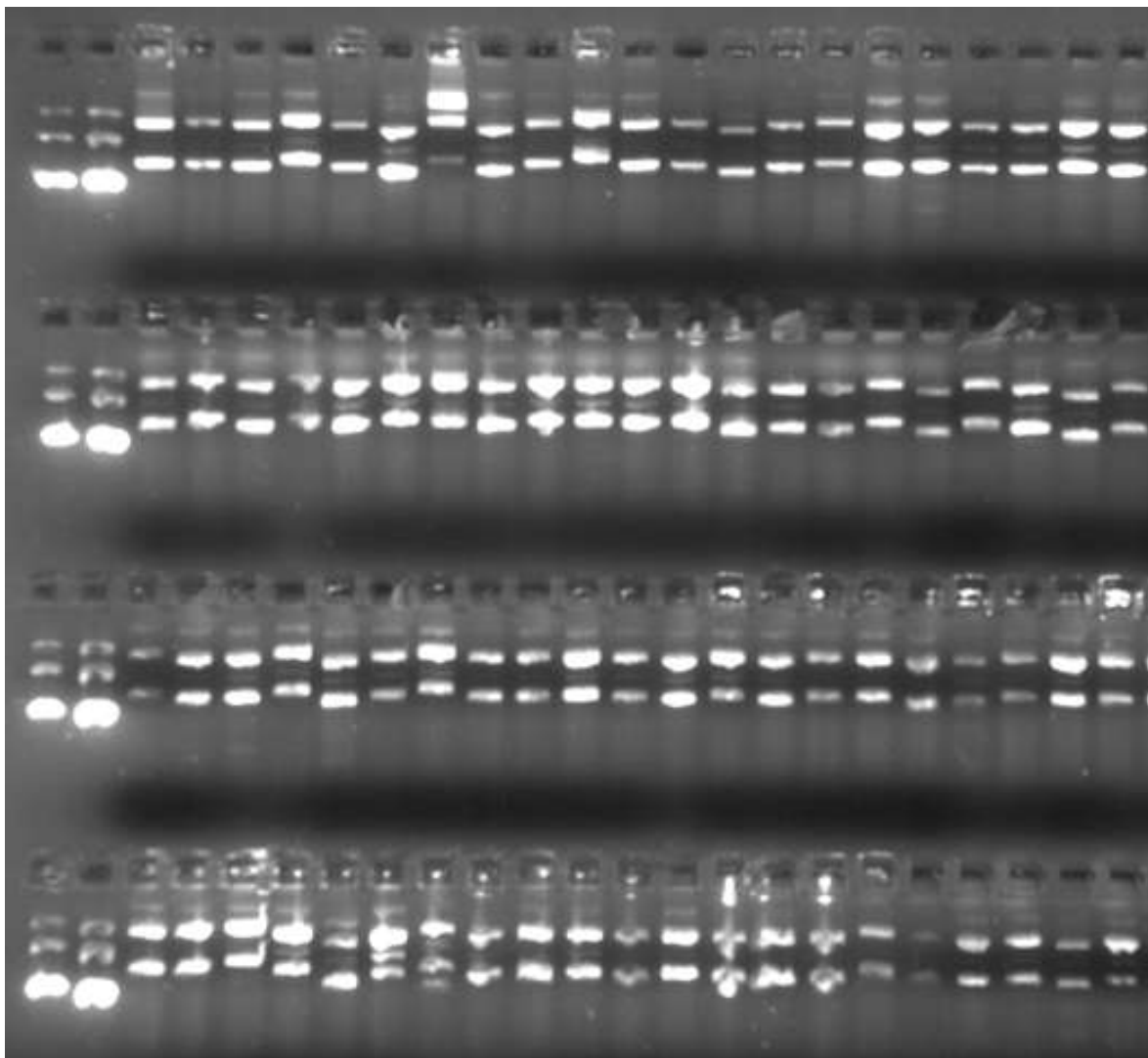


Figura 18: Quantificação de plasmídios derivados da biblioteca de cDNA de vermes adultos. Gel de agarose 1%, corado com brometo de etídio. Nas duas primeiras canaletas de cada andar do gel estão mostrados plasmídios pUC18 nas concentrações de 200 e 400 ng, usados como padrões de quantidade.

4.6. Análises Bioinformáticas

4.6.1. Processamento das seqüências geradas

Foram geradas 4054 seqüências, que foram submetidas à análise bioinformática. Primeiramente, estas foram submetidas ao programa PHRED, onde apenas as seqüências com qualidade PHRED >8 são aceitas. Posteriormente, estas seqüências foram submetidas ao programa Cross Match, onde os vetores, cauda de poli A e *spliced leader* foram mascarados dos arquivos gerados. Removemos todos os arquivos que apresentavam uma seqüência menor que 100 nucleotídeos válidos (após a retirada de vetor, SL e cauda de poli A), ficamos então no final, com 2372 seqüências válidas (374 de adultos, 637 de fêmeas, 645 de machos e 716 de ovos).

As seqüências geradas para a forma que se assemelha à fase de esquistossômulo não entraram na análise de bioinformática, devido às restrições de tempo, mas já estão sendo analisadas no nosso laboratório. As médias das seqüências processadas podem ser visualizadas na Tabela 2:

Tabela 2: Média das seqüências geradas para cada biblioteca.

	Adulto	Fêmea	Macho	Ovo
Média *	285	488	428	292

As seqüências foram previamente processadas pelos programas PHRED e Cross Match e removidas as seqüências menores que 100 bases.

* para o cálculo da média não foi contabilizado o SL e nem a cauda de poli A

É possível que o de tamanho dos insertos das bibliotecas pudesse ser aumentado a partir de modificações nas condições de amplificação (Piao *et al.*, 2001). Nesse sentido, utilizamos dois programas de amplificação. Um deles, o mais simples, foi usado na amplificação dos cDNAs de machos, fêmeas e vermes adultos. O segundo PCR *step down*, foi usado para a amplificação de cDNAs de ovos e da forma semelhante à fase de esquistossômulo. Os resultados foram

semelhantes com os dois programas, não verificando um incremento de tamanho, quando se utilizou PCR "*step down*".

4.6.2. Análise das seqüências agrupadas pelo programa CAP3

As seqüências válidas foram submetidas ao agrupamento pelo programa CAP3, gerando os *uniques* apresentados na Figura 19. Foram gerados um total de 643 *uniques* (*uniques* por biblioteca: adulto 115, fêmeas 195, macho 156, ovo 177), a partir das 2372 seqüências.

Para biblioteca de vermes adultos foi realizado um menor número de sequenciamentos, pois assumimos que as bibliotecas de vermes fêmeas e machos representariam o contingente de transcritos apresentado pelos vermes adultos. A biblioteca de ovos foi a biblioteca mais explorada e apresentou um número maior de *singlets* em relação ao número de *contigs*. No entanto, os *contigs* desta bibliotecas apresentaram um número maior de seqüências do que as outras bibliotecas (*contigs* com mais de 12 ESTs, como pode ser visto na Figura 20).

Para analisarmos a redundância das bibliotecas, observamos o número de seqüências inseridas dentro de *contigs* (Figura 20). Podemos ver com o perfil deste gráfico, que as bibliotecas em geral, apresentaram baixa redundância, podendo ser melhor exploradas. Foi dada uma especial atenção aos transcritos que supostamente estavam mais representados, ou foram mais seqüenciados, nas diversas bibliotecas. Pode-se observar no gráfico que obtivemos, em média, cinco *contigs* compostos por mais de 20 ESTs nas bibliotecas analisadas. Estes transcritos estão enumerados na Tabela 3.

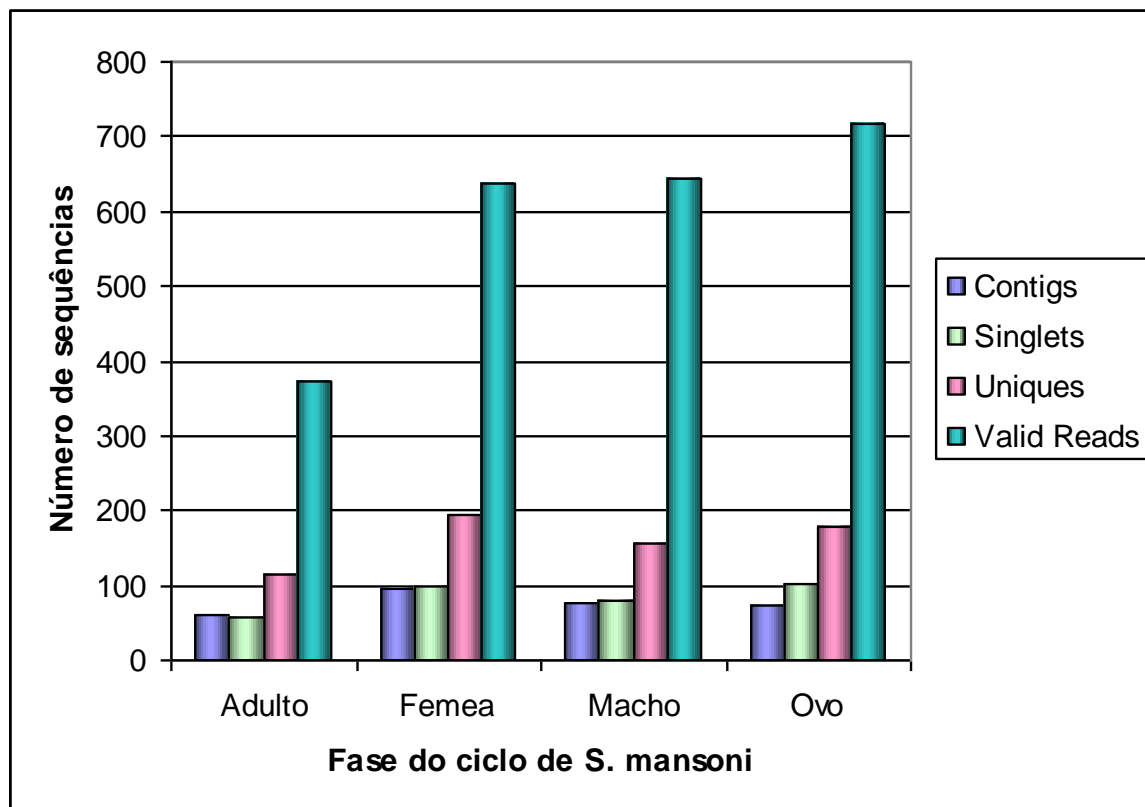


Figura 19: Número de *uniques* gerados para cada biblioteca a partir das sequências válidas, utilizando o programa CAP3.

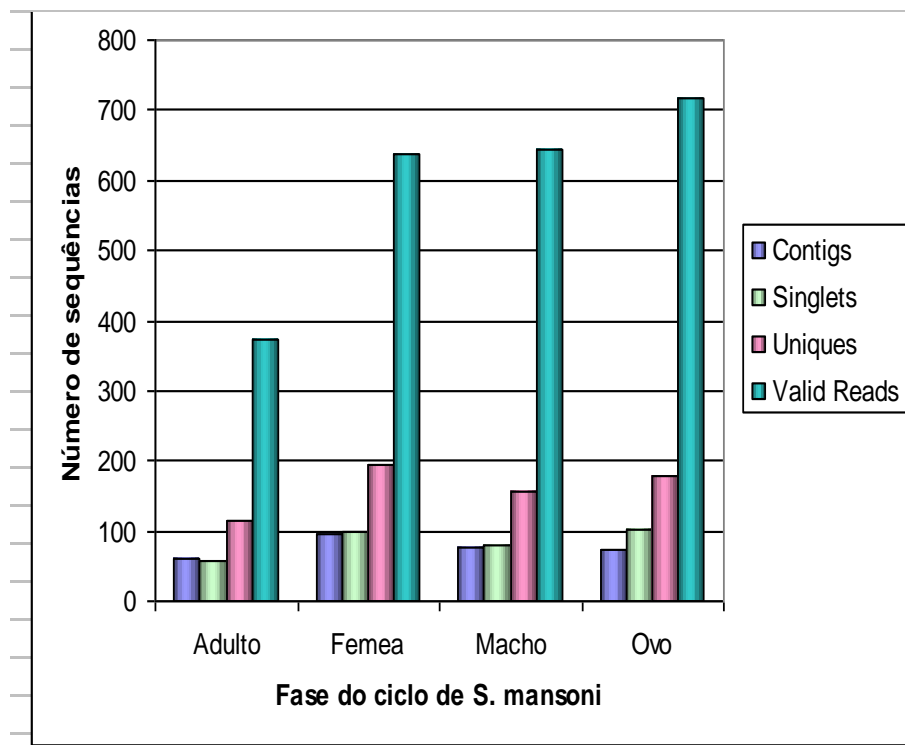


Figura 20: Análise da redundância das bibliotecas de cDNA após agrupamento das seqüências pelo programa CAP3.

Tabela 3: Contigs com número de ESTs superior a vinte.

Contig	Número de EST por contig	Número de contigs			
		Adulto	Fêmea	Macho	Ovo
69 FE; 05 MA; 23 MA; 75 OV	24	0	1	2	1
15 FE; 48 OV	25	0	1	0	1
26 AD; OV 7	26	1	0	0	1
24 OV;	27	0	0	0	1
32 AD; 42 MA	32	1	0	1	0
14 AD	37	1	0	0	0
48 AD	39	1	0	0	0
54 MA	43	0	0	1	0
21 OV; 26 OV	44	0	0	0	2
70 MA	61	0	0	1	0
06 FE	82	0	1	0	0

De acordo com a tabela apresentada, os *contigs* com um número discrepante de seqüências agrupadas foram: um *contig*, da biblioteca de fêmea, que possui 82 ESTs e um *contig* da biblioteca de macho contendo 61 ESTs. Este resultado pode indicar que estes transcritos estavam sendo fortemente expressos no organismo e, portanto, tínhamos muitas cópias presentes para servir de molde na PCR; ou foram simplesmente, tendenciosamente mais amplificados durante os ciclos de PCR para a construção da biblioteca.

Após a análise de homologia por BLAST, pudemos identificar estes *contigs* e anotá-los (Tabela 4). O *contig* com 61 ESTs apresenta homologia à seqüência de *Rattus norvegicus* que possui similaridade a uma proteína que contém um domínio Bola, que possui importante função na resposta à diversos tipos stress celular. Este *contig* também apresenta alta similaridade com uma EST de *S. Japonicum* não anotada. O *contig* de fêmea contendo 82 ESTs não apresentou quaisquer homologia com seqüências depositadas em bancos de dados devendo ser um gene específico de *S. mansoni*.

Entre os outros transcritos mais seqüenciados, encontramos o *erh* (*enhancer of rudimentary homolog*) que é um co-regulador transcricional, altamente conservado entre os eucariotos, desde humanos até protozoários. Funções atribuídas a este regulador transcricional incluem: acentuação da biossíntese de pirimidinas, regulação do ciclo celular e repressão do fator de transcrição tecido-específico HNF-1 (*hepatocyte nuclear factor-1*), através da ligação ao coativador DCoH (*dimerization cofactor of HNF1*) (Wan *et al.*, 2005). Estes *uniques* estão presentes nas bibliotecas de adultos e ovos e têm homologia com a seqüência de *T. solium*, que, interessante, também sofre *trans splicing* neste organismo. Esta seqüência também foi encontrada entre transcritos gerados por *trans splicing* em *E. multilocularis* (Brehm *et al.*, 2000)

As bibliotecas de machos, adultos e fêmeas apresentaram seqüências similares a possível seqüência codificadora proteína associada a setor da ATPase de membrana vacuolar, que é responsável pela acidificação de uma variedade de compartimentos intracelulares em eucariotos. Novamente, esta seqüência já foi identificada em *T. solium*, e é tida como transcrito produzido por *trans splicing* neste organismo (Brehm *et al.*, 2002).

Tabela 4: Anotação dos contigs com mais que vinte seqüências.

Tamanho nt	adultos	
416	Contig14	Putative Enhancer of rudimentary homolog (<i>Taenia solium</i>)
407	Contig26	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced
575	Contig32	Similar to <i>Taenia solium</i> putative vacuolar ATPase membrane sector associated protein
348	Contig48	No match
	machos	
624	Contig23	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04497 unknown mRNA
425	Contig42	Similar to <i>Rattus norvegicus</i> similar to Bola domain-containing protein like
391	Contig54	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced
307	Contig 5	No match
576	Contig70	Similar to <i>Taenia solium</i> putative vacuolar ATPase membrane sector associated
	femeas	
510	Contig15	Similar to <i>Pongo pygmaeus</i> hypothetical protein
563	Contig69	Similar to <i>Taenia solium</i> putative vacuolar ATPase membrane sector associated protein
312	Contig 6	No match
	ovo	
439	Contig7	Similar to <i>Pongo pygmaeus</i> hypothetical protein
418	Contig26	Putative U6 snRNA-associated Sm-like protein LSm8 [mus musculus]
126	Contig21	No match
425	Contig24	No match
416	Contig48	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced mRNA, complete cds
471	Contig75	Putative Enhancer of rudimentary homolog [taenia solium]

Uma seqüência que está entre as mais encontradas na biblioteca de ovo, mas que também foi vista em outras bibliotecas (em *contigs* compostos de menos de 20 seqüências), é o transcrito codificador da proteína LSm8 da família das Sm proteínas que se associa a U6 snRNA (além dela, encontramos durante anotação dos *uniques* em geral a proteína LSm 7, que também se liga ao U6 snRNA) (Pannone *et al.*, 2001). Portanto Lsm8 é uma das proteínas envolvidas na formação do spliceossomo através da ligação a U6. Proteínas associadas à maquinaria de *splicing* também foram encontradas dentre os transcritos produzidos por *trans splicing* em *E. multilocularis* (Brehm *et al.*, 2000).

Um gene muito redundante nas bibliotecas é o gene da proteína semelhante ao inibidor de ATPase. Este tem identidade elevada com a seqüência de *S. mansoni* depositada e foi descrito por Davis e colaboradores (1995) como um transcrito que sofre *trans splicing*. Também em *C. elegans* este transcrito foi identificado como produzido por *trans splicing* (Blumenthal, 1995). Este *unique* foi encontrado em três das quatro bibliotecas em questão. A proteína semelhante ao inibidor de ATPase regula a atividade catalítica da ATP sintase mitocondrial, sem que haja a interrupção na produção de ATP (Zanotti *et al.*, 2004). O fato deste mRNA ter sido encontrado e descrito em *S. mansoni*, dentre poucos transcritos estudados que sofrem por *trans splicing* por Davis e colaboradores (1995), pode indicar que este é muito expresso no organismo.

Seria interessante separarmos, dos bancos de dados disponíveis, as seqüências dos transcritos que sabidamente sofrem *trans splicing* de todos organismos que possuem este processo, pois diferentemente do que tem sido descrito na literatura, sugerimos que possa haver alguma conservação entre os transcritos gerados por *trans splicing* nos diferentes organismos.

4.6.3. Buscas de homologia dos *uniques* de *S.mansoni* produzidos por *trans splicing*

Após a obtenção das 643 *uniques*, foram feitas buscas de homologia com seqüências depositadas de *S. mansoni*, ou de outros organismos. Na análise dos dados foram estabelecidas prioridades. Estas foram as seguintes: -Primeira prioridade foi dada a genes de *S. mansoni* anotados; - Segunda prioridade às seqüências de *S. mansoni* depositadas tanto no dbEST quanto

aquelas obtidas pela RGMG; -Terceira prioridade foi dada a cada banco de dados de ESTs separadamente (dbEST ou RGMG); -Quarta prioridade às seqüências com homologia a outros organismos e às seqüências de *S. japonicum* e, finalmente, como Quinta prioridade as seqüências sem quaisquer homologia significativa (Figura 21).

Em trabalhos anteriores de análises do transcriptoma de *S. mansoni* pela geração de ESTs, foi demonstrado que em torno de 71% das seqüências eram completamente desconhecidas para bibliotecas geradas de diversas fases da vida deste organismo (Franco *et al.*, 1997, Prosdocimi *et al.*, 2002). Em recente publicação do transcriptoma do parasita, onde foram analisados 30.988 *uniques*, observou-se que 55% das seqüências não possuíam homologia em bancos de dados, e que outros 23% eram homologos a seqüências conhecidas de *S. mansoni* (Verjovski-Almeida *et al.*, 2003). Hoje, após a análise deste grande número de seqüências das redes de sequenciamento de São Paulo e de Minas Gerais, chegamos a um total de 73% de *hits* com seqüências já obtidas de *S. mansoni*. No entanto, este número aqui encontrado pode, ainda, incluir novos parálogos, que não foram separados na nossa análise. Este grande número de seqüências com similaridade a seqüências de *S. mansoni* já conhecidas mostra o grande impulso que foi dado nos últimos anos na descoberta gênica.

Apesar da enorme contribuição destes projetos nas descobertas gênicas de *S. mansoni*, obtivemos ainda 27% de possíveis novos fragmentos gênicos, compostos de 24% dos fragmentos com função desconhecida, ou seja, sem qualquer homologia significativa com seqüências presentes nos bancos de dados (estas seqüências possuíam uma média de 420 nt. para a biblioteca de fêmeas), e 3% de possíveis novos ortólogos (Figura 21).

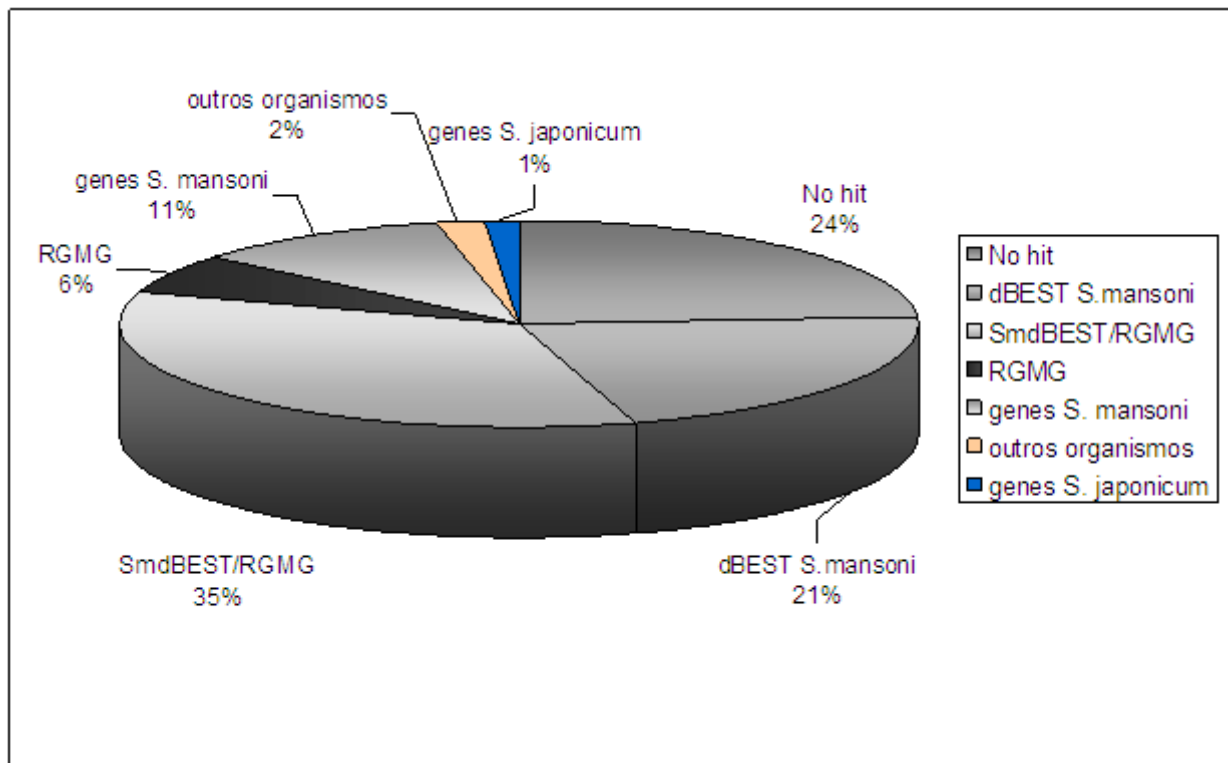


Figura 21: Homologia dos *uniques* gerados com seqüências depositadas em bancos de dados.

Devemos ainda ressaltar que dentre os *uniques* incluídos como *no match* estão ESTs que por terem um tamanho pequeno, podem não apresentar quaisquer *hits* com seqüências em bancos de dados. O que também dificulta a busca de seqüências similares a *S. mansoni* é a escassez de seqüências depositadas de organismos evolutivamente próximos, visto que houve uma divergência antiga e independente deste organismo em relação aos outros metazoários (Hausdorf, 2000). Uma pequena melhora na identificação foi obtida devido ao crescimento do número de seqüências de *S. japonicum* em bancos de dados, que chegou a 43.707 ESTs de várias fases de seu ciclo de vida (Hu *et al.*, 2003). Neste projeto de transcritos produzidos por *trans splicing* foi possível identificar 1% de novos transcritos de *S. mansoni* por homologia a *S. japonicum*.

Para anotação dos *uniques* gerados foram feitas pesquisas utilizando o programa BLASTx contra o nr, SwissProt e contra o KOG e o COG. Foi feito também BLASTn contra as seqüências gênicas de *S. mansoni* depositadas no GenBank e contra o nt. O número de *hits* obtidos contra cada um dos bancos de dados utilizados para anotação, pode ser visto na Figura 22. Na nossa análise é importante ressaltar que os *uniques* podem apresentar similaridade com seqüências de um ou mais bancos de dados, como por exemplo os bancos de proteínas e o SwissProt, já que este último está contido dentro do banco de proteínas.

Como podemos notar nesta análise ocorreram poucos *hits* para biblioteca de vermes adultos, sendo a maioria das seqüências não anotadas. Isto foi de certa forma inesperado, visto que a maioria das seqüências de *S. mansoni* depositadas até hoje são provenientes desta fase. Este achado pode se dever ao fato da média de tamanho das seqüências geradas ser muito pequena (285 nt.), o que dificulta a sua anotação.

O baixo número de *uniques* com similaridade com seqüências completas do banco de dados de *S. mansoni* é atribuído ao pequeno tamanho deste banco de dados. Há, também, alto número de *uniques* da biblioteca de ovo sem similaridade com seqüências depositadas em banco de dados, talvez por esta fase expressar genes mais específicos de *S. mansoni*, não sendo compartilhado com outros organismos, ou também por esta fase ser menos estudada do que adultos.

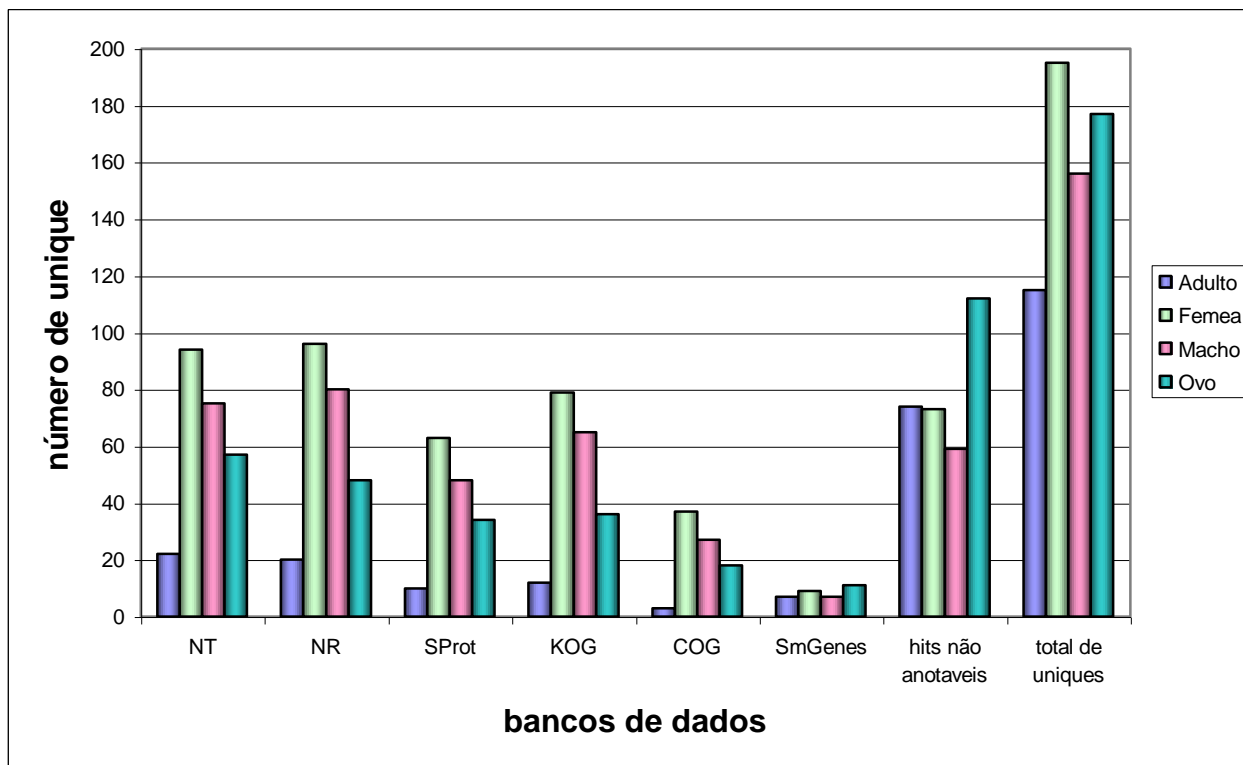


Figura 22: Anotação dos *uniques* gerados de acordo com a similaridade às seqüências depositadas em bancos de dados: **NT**- banco de dados de nucleotídeos, **NR**- banco de dados de proteínas, **Sprot**- SwissProt, **KOG**- *Clusters of orthologous groups* de eucariotos, **COG**- *Clusters of orthologous groups* de procariotos, **SmGenes**- genes completos de *S. mansoni*, e hits não anotáveis- sem homologia a quaisquer seqüências dos bancos de dados anteriormente mencionados.

Os BLASTs contra os bancos de dados KOG e COG foram feitos para que pudéssemos agregar categorias funcionais aos *uniques*. As diferenças de similaridade encontradas entre os bancos KOG e COG (para o COG obtivemos menos da metade dos *hits* encontrados para o KOG), são atribuídas a distância evolutiva que há entre *S. mansoni* e os procariotos, em contraste com os eucariotos (*D. melanogaster* e *C. elegans* são os organismos que fazem parte do KOG mais próximos evolutivamente ao *S. mansoni*).

Crítérios para anotação foram pré-estabelecidos conforme descritos em Materiais e Métodos (item 3.13.2). No entanto, ao iniciarmos a anotação, nos deparamos com um elevado número de seqüências de *S. japonicum* não anotadas, mas que apresentaram ótimos alinhamentos com os *uniques* aqui gerados. Tivemos então que estabelecer um novo critério: quando houvesse um bom alinhamento com seqüências conhecidas de outros organismos, estas seriam as seqüências usadas para anotação; mas quando os alinhamentos entre as seqüências de *S. mansoni* e outros organismos exibissem um escore significativamente menor do que o alinhamento com seqüências *S. japonicum*, utilizaríamos as seqüências deste último para a anotação.

A partir dos resultados de alinhamento retornados pelos BLASTs foi possível anotar uma grande quantidade dos *uniques*, que podem ser vistos nas Tabelas 5, 6, 7 e 8.

Tabela 5: Uniques anotados da biblioteca de adulto

ID	Anotação
AD1f_C11	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05470 unknown mRNA
AD1f_G02	Similar to <i>Rattus norvegicus</i> sulfite oxidase, mitochondrial precursor
AD1r_A02	Similar to <i>Mus musculus</i> Hypothetical UPF0327 protein
AD1r_A05	Putative <i>Schistosoma mansoni</i> fatty acid binding protein
AD2f_A03	Similar to TPA: <i>Schistosoma mansoni</i> Perere non-LTR retrotransposon
AD2f_F04	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced
AD3r_C04	Putative <i>Schistosoma mansoni</i> trans-spliced mRNA L11
AD3r_H03	Similar to <i>Schistosoma japonicum</i> clone SJCHGC06126 unknown mRNA
Contig1	Similar to <i>S. mansoni</i> mRNA for tandem repeat
Contig12	Similar to <i>Schistosoma mansoni</i> LTR-retrotransposon Boudicca
Contig13	Putative <i>Schistosoma mansoni</i> hexokinase gene
Contig14	Putative Enhancer of rudimentary homolog
Contig16	Similar to <i>C. elegans</i> Hypothetical protein ZK353.9 in chromosome III
Contig20	Similar to putative cytochrome c oxidase, subunit COX19
Contig21	Similar to <i>Schistosoma mansoni</i> DNA for repeat sequences
Contig26	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced
Contig29	Similar to <i>Schistosoma mansoni</i> clone Sm16 satellite
Contig32	Similar to <i>Taenia solium</i> putative vacuolar ATPase membrane sector associated protein
Contig35	Similar to TPA: <i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon
Contig38	Similar to <i>Taenia solium</i> putative Mitochondrial 60S ribosomal protein L14
Contig39	Similar to <i>Schistosoma japonicum</i> Sj-Ts1
Contig42	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05364
Contig44	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05920 unknown mRNA
Contig46	Similar to <i>Homo sapiens</i> hypothetical protein HSPC016
Contig49	Similar to <i>Mus musculus</i> putative U6 snRNA-associated Sm-like protein LSm8
Contig5	Putative RNA-binding protein [<i>Schistosoma japonicum</i>]
Contig50	Similar to <i>Schistosoma japonicum</i> clone SJCHGC02526
Contig55	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09784 unknown mRNA
Contig60	Similar to <i>S. japonicum</i> similar to NM_132090 CG3446 gene product in <i>D. melanogaster</i>
Contig62	Similar to <i>Rattus norvegicus</i> BOLA domain-containing protein like
Contig63	Similar to <i>Pongo pygmaeus</i> hypothetical protein

Tabela 6: Uniques anotados da biblioteca de fêmeas

Contig	Anotação
Contig1	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05740 unknown mRNA
Contig4	Similar to <i>Schistosoma mansoni</i> DNA for repeat sequences
Contig7	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04305 unknown mRNA
Contig10	Similar to <i>Rattus norvegicus</i> similar to BolA domain-containing protein like
Contig11	Similar to Chibby protein (PIGEA-14) (Cytosolic leucine-rich protein)
Contig13	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05203 unknown mRNA
Contig14	Similar to <i>Rattus norvegicus</i> Thioredoxin, mitochondrial precursor (Mt-TRX)
Contig15	Similar to <i>Pongo pygmaeus</i> hypothetical protein
Contig16	Similar to <i>Mus musculus</i> U6 snRNA-associated Sm-like protein LSm7
Contig17	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09304 unknown mRNA
Contig18	Similar to <i>Schistosoma japonicum</i> hypothetical protein, putative Saposin type B
Contig19	SMEGG2 <i>S.mansoni</i> DNA for eggshell precursor protein
Contig20	Similar to <i>Xenopus laevis</i> LOC494711 protein
Contig21	Similar to <i>Schistosoma japonicum</i> clone SJCHGC06407 unknown mRNA
Contig22	<i>S. mansoni</i> ATPase inhibitor-like protein trans-spliced mRNA
Contig23	Similar to <i>Schistosoma japonicum</i> hypothetical protein
Contig25	Putative Dynactin 6 (WS-3 protein) [<i>Homo sapiens</i>]
Contig27	Putative Ribulose-phosphate 3-epimerase [<i>Homo sapiens</i>]
Contig30	Similar to <i>Homo sapiens</i> MRG-binding protein
Contig32	Similar to <i>Pan troglodites</i> Glycerophosphoryl diester phosphodiesterase
Contig33	Putative Enhancer of rudimentary homolog [<i>Taenia solium</i>]
Contig34	Putative Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 [<i>Homo sapiens</i>]
Contig36	Similar to <i>Crassostrea gigas</i> cAMP responsive element binding protein-like
Contig37	Similar to <i>Reclinomonas americana</i> Mitochondrial ribosomal protein L14
Contig38	Similar to <i>Caenorhabditis elegans</i> DnaJ homolog dnj-20 precursor
Contig39	Similar to <i>Schistosoma japonicum</i> clone SJCHGC02526 unknown mRNA
Contig40	Similar to <i>Homo sapiens</i> NTF2-related export protein 2 (p15-2 protein)
Contig41	Similar to <i>Caenorhabditis elegans</i> tropomodulin protein 1, isoform a
Contig42	Similar to <i>Schizosaccharomyces pombe</i> putative cytochrome c oxidase, subunit COX19
Contig43	Similar to <i>Schistosoma japonicum</i> clone ZZD1160 mRNA sequence
Contig44	Uncharacterized conserved protein, contains SPRY
Contig45	Similar to <i>Mus musculus</i> trafficking protein particle complex subunit 6B
Contig46	TPA: <i>Schistosoma mansoni</i> Perere non-LTR retrotransposon mRNA,
Contig47	Putative WD-repeat protein SL1-17 [<i>Schistosoma mansoni</i>]
Contig50	Similar to <i>Homo sapiens</i> nucleotide binding protein 2 (NBP 2)
Contig52	TPA: <i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon mRNA
Contig53	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09252 unknown mRNA
Contig56	Putative Peptidyl-prolyl cis-trans isomerase like 2 (PPIase) [<i>Caenorhabditis elegans</i>]
Contig57	Similar to <i>Schistosoma japonicum</i> clone SJCHGC08900 unknown mRNA

Contig58	Similar to <i>Homo sapiens</i> predicted MYND Zn-finger protein/hormone receptor
Contig59	Similar to <i>Emericella nidulans</i> nuclear movement protein nudC
Contig60	Similar to <i>Mesorhizobium loti</i> 50S ribosomal protein L24
Contig61	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05920 unknown mRNA
Contig62	Similar to dual specificity phosphatase
Contig63	Similar to <i>Homo sapiens</i> peripheral-type benzodiazepine receptor (PBR)
Contig64	Similar to <i>Yeast</i> BolA (bacterial stress-induced morphogen)-related [yeast]
Contig65	Similar to <i>Mus musculus</i> thioredoxin-like protein 2
Contig66	Similar to <i>S mansoni</i> uncharacterized conserved protein
Contig67	Putative reductases with broad range of substrate [<i>Hyoscyamus niger</i>]
Contig69	Putative vacuolar ATPase membrane sector associated protein [<i>Schistosoma mansoni</i>]
Contig70	Similar to <i>Schistosoma japonicum</i> clone SJCHGC01021 unknown mRNA
Contig71	Similar to <i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon mRNA, complete sequence
Contig72	Similar to <i>Homo sapiens</i> 28S ribosomal protein S18a, mitochondrial precursor
Contig73	Similar to <i>Mus musculus</i> replication protein A 14 kDa subunit (RP-A)
Contig74	Similar to <i>Rattus norvegicus</i> proteasome non-ATPase regulatory subunit 9
Contig75	Similar to uncharacterized conserved protein
Contig76	Similar to Small Nuclear ribonucleoprotein splicing factor [<i>Mus musculus</i>]
Contig77	Similar to uncharacterized conserved protein [<i>Anopheles gambiae</i>]
Contig79	Similar to <i>Mus musculus</i> postsynaptic protein Cript
Contig80	Similar to <i>Rattus norvegicus</i> BET1 homolog (Golgi vesicular membrane trafficking protein p18)
Contig81	Similar to <i>Tetraodon nigroviridis</i> unnamed protein product
Contig83	Similar to <i>Mus musculus</i> prefoldin subunit 6 (Protein Ke2)
Contig84	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced mRNA, complete cds
Contig85	Similar to <i>Neurospora crassa</i> predicted protein
Contig86	<i>Schistosoma mansoni</i> trans-spliced mRNA L11 gene, complete cds
Contig87	Similar to <i>Mus musculus</i> Protein SYS1 homolog
Contig88	Enolase [<i>S. mansoni</i>]
Contig89	Similar to <i>Mus musculus</i> UXT protein (ubiquitously expressed transcript protein)
Contig90	Similar to <i>Mus musculus</i> probable U3 small nucleolar RNA-associated protein 11
Contig91	Similar to Unnamed protein
Contig92	Putative U2 small nuclear ribonucleoprotein B" [<i>Mus musculus</i>]
Contig95	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03918 unknown mRNA
FE1f_E03	Similar to <i>Schistosoma japonicum</i> hypothetical protein
FE1f_F08	Similar to <i>Homo sapiens</i> B-cell receptor-associated protein 31
FE1f_H02	Similar to <i>Mycosphaerella graminicola</i> 4-hydroxyphenylpyruvate dioxygenase
FE1f_H06	Similar to <i>Schistosoma japonicum</i> clone SJCHGC01436 unknown mRNA
FE1r_G10	<i>Schistosoma mansoni</i> trans-spliced mRNA L11 gene, complete cds
FE2r_F03	Similar to <i>D. melanogaster</i> 39S ribosomal protein L35, mitochondrial
FE2r_G03	Similar to <i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon mRNA, complete sequence

FE2r_G06	Similar to <i>Mus musculus</i> transcription initiation factor TFIID subunit 13
FE2r_H08	Probable O-sialoglycoprotein endopeptidase [<i>Mus musculus</i>]
FE3f_B04	<i>Schistosoma mansoni</i> tropomyosin mRNA, complete cds
FE3f_B07	Similar to <i>Homo sapiens</i> NADH-ubiquinone oxidoreductase B16.6 subunit
FE3f_B10	Similar to <i>Schistosoma japonicum</i> clone ZZD309 mRNA sequence
FE3f_C09	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03581 unknown mRNA
FE3f_C10	Similar to <i>Homo sapiens</i> peroxisomal 3,2-trans-enoyl-CoA isomerase
FE3f_E10	Similar to ubiquitin-like protein [yeast]
FE3f_F02	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05169 unknown mRNA
FE3f_F11	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05942 unknown mRNA
FE3f_G05	Similar to <i>Schistosoma japonicum</i> hypothetical protein
FE3f_G07	Similar to <i>Homo sapiens</i> LYR family protein A-211C6.1
FE3f_H04	Similar to <i>Bos taurus</i> Alpha-(1,6)-fucosyltransferase
FE5f_B06	Similar to SEG_AF448819S <i>Schistosoma mansoni</i> adenylosuccinate lyase (ASL) gene
FE5f_D05	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05364 unknown mRNA
FE5f_H06	Similar to <i>Caenorhabditis elegans</i> hypothetical protein R05H5.5
FE6f_B03	Similar to <i>Danio rerio</i> mitochondrial import inner membrane translocase subunit Tim23
FE6f_B06	Similar to <i>Schistosoma japonicum</i> clone SJCHGC02727 unknown mRNA
FE6f_E04	Similar to <i>Schistosoma japonicum</i> clone SJCHGC00905 unknown mRNA
FE6r_D05	Similar to <i>Danio rerio</i> ectonucleotide pyrophosphatase/phosphodiesterase 3
FE6r_E04	Similar to <i>Schizosaccharomyces pombe</i> mating-type switching protein swi10
FE6r_H10	Similar to <i>Mus musculus</i> trafficking protein particle complex subunit 6B
FE7f_B05	Similar to <i>Schistosoma japonicum</i> clone SJCHGC06284 unknown mRNA
FE7f_B12	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04376 unknown mRNA
FE7r_D08	Similar to yeast thymidylate kinase (dTMP kinase)
FE7r_F12	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09423 unknown mRNA
FE8r_A11	Similar to <i>Homo sapiens</i> anaphase promoting complex subunit 10 (APC10)
FE8r_B02	Similar to predicted hydrolases of HD superfamily [yeast]
FE8r_B05	Similar to <i>Rattus norvegicus</i> dihydropteridine reductase (HDHPR)
FE8r_B08	Similar to <i>Rattus norvegicus</i> predicted alpha/beta hydrolase
FE8r_D06	Putative Peptidyl-prolyl cis-trans isomerase H [<i>Homo sapiens</i>]
FE8r_F05	Putative COP9 signalosome complex subunit 5 [<i>Danio rerio</i>]
FE8r_F06	Putative 6-pyruvoyl tetrahydrobiopterin synthase [<i>Poecilia reticulata</i>]
FE8r_F08	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03998 unknown mRNA
FE8r_G01	Similar to <i>Homo sapiens</i> mitochondrial 28S ribosomal protein S33 (S33mt)
FE8r_G05	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03877 unknown mRNA
FE8r_H02	SMTANREP <i>S.mansoni</i> (Liberia) mRNA for tandem repeat
FE8r_H12	Similar to <i>Schistosoma japonicum</i> clone SJCHGC02392 unknown mRNA

Tabela 7: Uniques anotados da biblioteca de machos

ID	Anotação
Contig1	Putative COP9 signalosome complex subunit 5 [<i>Danio rerio</i>]
Contig12	Similar to <i>Gallus gallus</i> protein similar to ZNF258 protein
Contig14	Similar to <i>Homo sapiens</i> MRG-binding protein
Contig16	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05920 unknown mRNA
Contig18	Putative O-sialoglycoprotein endopeptidase [<i>Xenopus laevis</i>]
Contig19	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04305 unknown mRNA
Contig2	Similar to <i>Schistosoma japonicum</i> clone SJCHGC02526 unknown mRNA
Contig20	Putative Dynactin 6 (WS-3 protein) [<i>Homo sapiens</i>]
Contig22	Similar to <i>Taenia solium</i> putative mitochondrial 60S ribosomal protein L14
Contig23	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04497 unknown mRNA
Contig24	Similar to <i>Schistosoma japonicum</i> clone SJCHGC02734 unknown mRNA
Contig25	Similar to <i>D. melanogaster</i> enoyl-CoA hydratase
Contig27	Similar to <i>Schistosoma mansoni</i> DNA for repeat sequences
Contig28	<i>Schistosoma mansoni</i> myosin light chain mRNA
Contig29	similar to <i>Pongo pygmaeus</i> hypothetical protein
Contig31	Similar to putative cytochrome c oxidase, subunit COX19
Contig32	Putative U6 snRNA-associated Sm-like protein LSm7
Contig33	<i>Schistosoma mansoni</i> methionine sulfoxide reductase
Contig34	Similar to a mitochondrial/chloroplast ribosomal protein S18
Contig35	Similar to <i>Mus musculus</i> mitochondrial/chloroplast ribosomal protein S18
Contig36	Similar to <i>Mus musculus</i> U2 small nuclear ribonucleoprotein B
Contig37	Similar to <i>Schistosoma japonicum</i> clone SJCHGC01021 unknown mRNA
Contig38	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09252 unknown mRNA
Contig4	Similar to <i>S. mansoni</i> TPA: <i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon
Contig40	Putative U6 snRNA-associated Sm-like protein LSm8
Contig41	Putative peptidyl-prolyl cis-trans isomerase 10 (<i>C. elegans</i>)
Contig42	Similar to <i>Rattus norvegicus</i> similar to BOLA domain-containing protein like
Contig43	Putative enhancer of rudimentary homolog [<i>Taenia solium</i>]
Contig44	Similar to <i>D. melanogaster</i> DNA-repair protein complementing XP-A cells homolog
Contig45	Similar to <i>Xenopus laevis</i> U1 snRNP component
Contig46	Similar to <i>Rattus norvegicus</i> 26S proteasome non-ATPase regulatory subunit 9
Contig47	<i>Schistosoma mansoni</i> trans-spliced mRNA L11 gene
Contig48	Similar to <i>Mus musculus</i> transcription initiation factor TFIID subunit 13
Contig49	Putative <i>Schistosoma mansoni</i> homeodomain protein (smox-5) mRNA, complete cds
Contig50	Similar to <i>Rattus norvegicus</i> similar to BOLA domain-containing protein like
Contig51	Similar to G protein gamma subunit
Contig53	Similar to <i>S. japonicum</i> similar to NM132090 CG3446 gene product in <i>D. melanogaster</i>

Contig54	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced
Contig55	Similar to <i>D. melanogaster</i> COP9 signalosome complex subunit 5
Contig56	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03337 unknown mRNA
Contig57	<i>S. mansoni</i> ubiquinol-cytochrome C reductase complex ubiquinol binding protein (UbCRBP) gene, and enolase gene
Contig58	Similar to <i>Schistosoma japonicum</i> clone ZZD1160 mRNA sequence
Contig59	Similar to <i>Homo sapiens</i> NTF2-related export protein 2 (p15-2 protein
Contig61	Similar to <i>Schistosoma japonicum</i> clone SJCHGC06126 unknown mRNA
Contig63	Similar to <i>Schistosoma japonicum</i> hypothetical protein
Contig65	Similar to <i>Schistosoma japonicum</i> hypothetical protein
Contig68	Similar to <i>Crassostea gigas</i> cAMP responsive element binding protein-like
Contig7	Similar to <i>Bos bovis</i> Chibby protein (PIGEA-14) (Cytosolic leucine-
Contig70	Similar to <i>Taenia solium</i> putative vacuolar ATPase membrane sector associated
Contig71	Similar to <i>Schistosoma japonicum</i> clone SJCHGC08089 unknown mRNA
Contig74	Putative thioredoxin, mitochondrial precursor (Mt-TRX) [<i>Rattus norvegicus</i>]
Contig8	Putative <i>Schistosoma mansoni</i> heat shock protein 70
MA2f1_F06	Similar to <i>S. japonicum</i> similar to NM_023140 thioredoxin-like 2; PKC interacting
MA2f2_B05	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09496 unknown mRNA
MA2f2_C09	Similar to <i>Schistosoma japonicum</i> hypothetical protein
MA2f2_D05	Similar to <i>Schistosoma japonicum</i> clone SJCHGC08900 unknown mRNA
MA2f2_D09	Similar to <i>Homo sapiens</i> dehydrogenase/reductase SDR family member 7
MA2f2_E05	Similar to <i>S. japonicum</i> similar to NM_017271 nuclear distribution gene C homolog in <i>R. norvegicus</i>
MA2f2_E09	Similar to <i>Homo sapiens</i> Rho GDP-dissociation inhibitor 2 (Rho GDI 2)
MA2f2_H04	Similar to <i>D. melanogaster</i> uncharacterized conserved protein
MA2r_G03	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03665 unknown mRNA
MA4r_A02	Similar to <i>Mus musculus</i> anaphase promoting complex subunit 11
MA4r_B02	Similar to <i>Schistosoma japonicum</i> clone SJCHGC01828 unknown mRNA
MA4r_C06	Similar to <i>S. mansoni</i> mRNA for tandem repeat
MA4r_C08	Similar to <i>Schistosoma japonicum</i> clone SJCHGC06974 unknown mRNA
MA4r_C12	<i>Schistosoma mansoni</i> fatty acid binding protein
MA4r_D12	Similar to <i>S. japonicum</i> hypothetical protein, putative Saposin type B
MA4r_F02	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04497 unknown mRNA
MA4r_F12	Similar to <i>Mus musculus</i> glutaredoxin-related protein C14orf87 homolog
MA4r_G05	Similar to <i>Schistosoma japonicum</i> clone SJCHGC08017 unknown mRNA
MA4r_H04	Putative 26S protease regulatory subunit S10B (Proteasome subunit p42)
MA4r_H10	Putative Mago nashi protein homolog
MA5r1_A09	Uncharacterized conserved protein
MA5r1_C04	Putative peptidyl-prolyl cis-trans isomerase 10
MA5r1_H03	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced
MA7r_A12	Similar to <i>Schistosoma japonicum</i> IB1 protein mRNA, complete cds

MA7r_B08	Similar to <i>C. elegans</i> DNA-directed RNA polymerase I 13.7 kDa
MA7r_D07	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03322 unknown mRNA
MA7r_E03	Similar to <i>Homo sapiens</i> GDP-fucose transporter
MA7r_F04	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09304 unknown mRNA
MA7r_G12	<i>Schistosoma mansoni</i> clone NS8W-c-10 mRNA sequence
MA7r_H05	Similar to <i>Schistosoma japonicum</i> hypothetical protein
MA8r_A11	Similar to TPA: <i>Schistosoma mansoni</i> Perere non-LTR retrotransposon
MA8r_C03	Putative tRNA (guanine-N(7)-)-methyltransferase
MA8r_D05	Putative predicted hydrolases of HD superfamily [yeast]
MA8r_G02	Putative peptidyl-prolyl cis-trans isomerase H [<i>Homo sapiens</i>]
MA8r_G03	Putative postsynaptic protein Cript [<i>Mus musculus</i>]
MA8r_G09	<i>Schistosoma mansoni</i> trans-spliced mRNA L11
MA8r_H03	Similar to <i>Schistosoma japonicum</i> clone ZZD854 mRNA sequence
MA8r_H07	Similar to reductases with broad range of substrate
MA9rf_G12	Putative Rad51 gene, complete cds [<i>Trypanosoma cruzi</i>]
MAr_A10	Putative <i>Schistosoma mansoni</i> ATPase inhibitor-like protein
MAr_G12	Similar to <i>Mus musculus</i> U6 snRNA-associated Sm-like protein LSm8

Tabela 8: Uniques anotados da biblioteca de ovos

ID	Anotação
Contig2	Similar to <i>Schistosoma japonicum</i> clone SJCHGC02526 unknown mRNA
Contig5	Similar to <i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon mRNA, complete sequence
Contig7	Similar to <i>Pongo pygmaeus</i> hypothetical protein
Contig14	Similar to <i>D. melanogaster</i> RING-box protein 1A (Regulator of cullins 1a)
Contig18	Similar to <i>C. elegans</i> peptidyl-prolyl cis-trans isomerase 10
Contig22	Similar to <i>Mus musculus</i> replication protein A 14 kDa subunit (RP-A)
Contig26	Putative U6 snRNA-associated Sm-like protein LSm8 [<i>Mus musculus</i>]
Contig29	U2 small nuclear ribonucleoprotein B
Contig33	Similar to <i>D. melanogaster</i> molybdenum cofactor synthesis protein cinnamon
Contig34	Similar to <i>Mus musculus</i> 28S ribosomal protein S18a, mitochondrial
Contig36	Similar to <i>Rattus norvegicus</i> thioredoxin, mitochondrial precursor (Mt-TRX)
Contig37	Similar to TPA: <i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon mRNA, complete sequence
Contig38	cytochrome c oxidase subunit 1 [<i>Schistosoma mansoni</i>]
Contig40	U6 snRNA-associated Sm-like protein LSm7 [<i>Mus musculus</i>]
Contig42	Similar to nuclear distribution protein NUDC
Contig46	Similar to <i>Schizosaccharomyces pombe</i> putative cytochrome c oxidase, subunit COX19
Contig48	<i>Schistosoma mansoni</i> ATPase inhibitor-like protein trans-spliced mRNA, complete cds
Contig49	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05920 unknown mRNA
Contig50	Similar to <i>Homo sapiens</i> guanine nucleotide-binding protein
Contig56	Putative acetylcholinesterase [<i>S. hematobium</i>]
Contig57	Putative lactate dehydrogenase
Contig58	Similar to <i>Mus musculus</i> mitochondrial deoxynucleotide carrier
Contig59	<i>Schistosoma mansoni</i> Saci-2 LTR retrotransposon mRNA, complete sequence
Contig61	SCMSLRNA <i>S. mansoni</i> spliced leader RNA gene
Contig63	Putative WD-repeat protein SL1-17
Contig65	Similar to <i>Homo sapiens</i> putative vacuolar ATP synthase membrane sector associated protein M8-9
Contig66	<i>Schistosoma mansoni</i> trans-spliced mRNA L11 gene, complete cds
Contig67	Similar to <i>Homo sapiens</i> eukaryotic translation initiation factor 4E binding protein 2 (eIF4E-binding protein 2)
Contig68	Similar to <i>Agrobacterium tumefaciens</i> chaperone protein dnaJ
Contig69	Similar to <i>Apis mellifera</i> similar to CG8171-PA
Contig70	Similar to <i>Homo sapiens</i> peripheral-type benzodiazepine receptor (PBR)
Contig71	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09496 unknown mRNA
Contig73	Similar to <i>Schistosoma japonicum</i> clone SJCHGC08320 unknown mRNA
Contig74	Putative RNA-binding protein [<i>S. japonicum</i>]

<u>Contig75</u>	Putative enhancer of rudimentary homolog [<i>Taenia solium</i>]
<u>Contig84</u>	Similar to metal-dependent phosphohydrolase HD domain-containing protein
<u>Contig86</u>	Similar to DNA-directed RNA polymerase I 13.7 kDa polypeptide
<u>OV10_C09</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC06725 unknown mRNA
<u>OV10_F08</u>	Similar to 5'-AMP-activated protein kinase, beta-2
<u>OV10r_F04</u>	Enolase [<i>S. mansoni</i>]
<u>OV2f_E11</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05169 unknown mRNA
<u>OV2f_G07</u>	Similar to ATP synthase delta chain, mitochondrial precursor
<u>OV3f_A08</u>	Similar to <i>Schistosoma japonicum</i> clone ZZZ67 mRNA sequence
<u>OV3f_B05</u>	SMTANREP <i>S.mansoni</i> (Liberia) mRNA for tandem repeat
<u>OV3f_D05</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04305 unknown mRNA
<u>OV3f_G02</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05388 unknown mRNA
<u>OV4r_B02</u>	Similar to <i>Schistosoma mansoni</i> endoplasmin
<u>OV4r_D01</u>	Hypothetical 36.2 kDa protein in HAM1-PEM2 intergenic region
<u>OV4r_E05</u>	Similar to 26S proteasome non-ATPase regulatory subunit 9
<u>OV4r_G07</u>	Similar to <i>Mus musculus</i> RIKEN cDNA 1810056O20
<u>OV5f_A04</u>	Similar to mitochondrial 28S ribosomal protein S33 (S33mt)
<u>OV5f_A06</u>	Similar to uroporphyrinogen-III synthase
<u>OV5f_B04</u>	Similar to <i>Schistosoma mansoni</i> hypothetical protein - fluke
<u>OV5f_B06</u>	Similar to <i>Caenorhabditis elegans</i> hypothetical protein T09A5.7
<u>OV5f_D08</u>	Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) [<i>S. mansoni</i>]
<u>OV5f_E02</u>	Uncharacterized conserved protein
<u>OV5f_E07</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05364 unknown mRNA
<u>OV5f_F02</u>	Similar to mitochondrial 28S ribosomal protein S6 (S6mt)
<u>OV7r_C03</u>	Similar to <i>Cricetulus griseus</i> stress-induced-phosphoprotein 1 (Hsc70/Hsp90-organizing protein)
<u>OV7r_D07</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03581 unknown mRNA
<u>OV7r_G06</u>	Similar to <i>Homo sapiens</i> peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 (Rotamase Pin4)(PPIase Pin4) (Par14) (Peptidyl-prolyl cis/trans isomerase EPVH)
<u>OV7r_G10</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC05469 unknown mRNA
<u>OV8_A05</u>	Similar to <i>Homo sapiens</i> DKFZP564K2062 protein
<u>OV8f_E05</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04497 unknown mRNA
<u>OV8f_G12</u>	Similar to <i>Taenia solium</i> enhancer of rudimentary homolog
<u>OV9r_B02</u>	Similar to <i>Schistosoma japonicum</i> clone SJCHGC03741 unknown mRNA

Um dado inusitado foi a identificação de *uniques* que apresentaram homologia com partes de retrotransposons de *S. mansoni*. É algo intrigante de se imaginar que regiões de retrotransposons poderiam sofrer *trans splicing*. Os retrotransposons identificados em meio aos transcritos produzidos por *trans splicing* foram:

- ❖ Saci-2 (bibliotecas: adulto, macho, fêmea, ovo)
- ❖ Pererê (bibliotecas: adulto, macho, fêmea)
- ❖ Boudicca (biblioteca de adulto)

Os retrotransposons Saci-2 e Pererê, possuem alta atividade transcricional onde exibem baixo número de cópias no genoma e uma taxa transcricional por cópia muito maior do que os quatro retrotransposons anteriormente descritos em *S. mansoni* (DeMarco *et al.*, 2004). Pererê não possui região LTR e foram vistos três *uniques* com alta similaridade a este retrotransposon. Destes *uniques*, dois se alinham na mesma região, entre as posições 4570 a 4837 nt., que representa região 5' UTR, e outro entre as posições 820 a 1085 nt., que representa o início da região codificadora da proteína poliproteína. Este retrotransposon possui 4875 nucleotídeos em sua seqüência (Figura 23).

O retrotransposon Saci-2 possui regiões LTR, regiões nas quais *uniques* diferentes se alinham. *Uniques* que cobrem tanto a região LTR 5' da posição 2 até 260 nt (esta região possui 259 nucleotídeos) quanto a LTR 3' da posição 4830 até 4945 nt., (essa LTR que possui 470 nucleotídeos). Todos os quatro *uniques* encontrados nas bibliotecas de *trans splicing*, apenas se alinham com as seqüências das regiões LTR neste retrotransposon (Figura 24).

Somente um *unique* da biblioteca de vermes adultos possui homologia com o retrotransposon Boudicca. Este retroposon também possui região LTR e tem 6038 nucleotídeos (Copeland *et al.*, 2003). Nosso *unique* se alinha com este retrotransposon entre as posições 4408 até 5154 nt, região 3' UTR (Figura 25).


```

gcggtacggt tggagtttgc gttacgatca cacggtatth attacaaagt ccttatatat 61
cgtaaacca cagtttaaaa atcaataata acattcagtt aacccttgaa gaaaagtgg 121
tgactgaca caaaaacgag ctgaaaaag tcctatttct tcagatatta acaactctgg 181
tcttacctat accccatgct ctgcaactcc tgttgacca gataatagtg ggagtggact 241
tattccacac atgttgctgg atgactgcat gtctaactta ataatgagtt ctaactcacg 301
aacagatggt atcgatctta ttaagagcga attagcctct gtctttaagc aattaagtgg 361
agtcacttgc gggctcttta ttaaagcatg atgatcttaa atttgaacta aagacactgt 421
cacccttgc actactgttg tccaaggata tagttacttc tgaattcgaa gacaggatca 481
aagtagaatc cgttattgac ttgatagcta gtgaagtcac caagcgaata atctctcgaa 541
ataatgtggt tatatataat atacctgaca aagttgccat taaaactgta agaaactcga 601
tactaaaggc agttaatctc caggacaatc catgccaatg tatccgactt acaaaaaagc 661
atcaaaaata ctctgcccct atctgttcta gatttgattc ccatttatta gcggaacggt 721
tgaagaatc tgaacagcta gtctgtgccc atacgaagtt taaaaacgct cgtatagtct 781
cagacaaaac caccaatcaa agactaacac aaaagcgtac catgaacgaa aataacattg 841
484 ac catgaacgaa aataacattg
ttgaggtcac aacaaatgtg atcgcgccag aagctgagcc cactggtaca gctaactctat 901
ttgaggtcac aacaaatgtg atcgcgccag aagctgagcc cactggtaca gctaactctat
ctcatgtagc tcagtgtact gatataccaa tgcagtgtgc tagtttgccct ctcatactca 961
ctcatgtagc tcagtgtact gatataccaa tgcagtgtgc tagtttgccct ctcatactca
tagtgccct agataaccag gacatatctg atgtagactc taatgtcatt ctttccagtg 1021
tagtgccct agataaccag gacatatctg atgtagactc taatgtcatt ctttccagtg
taatatcgga agccaaggat cgaacacctg attctattgt gaaggctcat aaaaacactg 1081
taatatcgga agccaaggat cgaacacctg attctattgt gaaggctcat aaaaacactg
ctaaacctaa aaagaatata ccaattaaga agaaagtaa taaaaaacct tctggttcta 1141
ctaaa 218
gaccgaatac caaaaatatt tcaacggcct taccaaaca aagtttaatt tctgaaacca 1201
tgcttaacc taccaccgt aaggagggtt ataaggacac atttctgtca aacgttacac 1261
gagacggcca ctacaatcat catgtaagca ggctactat tagacagacg ggaatgaacc 1321

acaaactctc cttcacatcc agtcaggaga ttcaatatgc taaagccca cattacgaca 4261
ttcgttaact cttgtcttta gcgaacaaa catattctag atcttctctc tacatgaatt 4321
actttacctg tgagttttct atgctctgga ataatttacc ccaaactatc cgtatgtaa 4381
actctctccc attgtttgtt cgtcctaata atgctttttg ctctctgtaa aatgcattaa 4441
atgctctagc gcatgcaagt gtatcttact ccacaagtga gattatagga actttaaatg 4501
ttaaacctct tacttgctat cattgttttg ttgttccgtg ctctttgctt taatcttag 4561
ttctctagtt gtagataaatt aataaataact tgctctggca ctggaaataa ccttacagaa 4621
137 tt gtagataaatt aataaataact tgctctggca ctggaaataa ccttacagaa
caatggtgat caatcatcag gctatccact taataaaaaa tgacaagttc cctggtgttg 4681
caatggtgat caatcatcag gctatccact taataaaaaa tgacaagttc cctggtgttg
cattggttg cgtgacata tgccatattt aagctaatta tcaattacta aaccagcaa 4741
catgggcttg cgtgacata tgccatattt aagctaatta tcaattacta aaccagtaa
cataaaactt tcttatattt catgtttgtt ctctacatta aatggtgatt tttataaca 4801
cataaaactt tcttatattt catgtttgat ctctacatta gatggtgatt attataaca
tctgatcatg cctgtaaact ggtgtgaatt ctgtaaatat tattttcaga atatattatt 4861
tctgatcatg c 379
attattatta tttag

```

Figura 23: Alinhamentos dos *uniques* com o retrotransposon non-LTR Perere. Em cinza início e fim da seqüência codificadora da proteína poliproteína. Em vermelho o alinhamento do *unique* encontrado na biblioteca de fêmeas. Em verde alinhamento dos *uniques* encontrados nas bibliotecas de adultos e machos. Senqüência com ID BK004067.1 depositada no NCBI



Figura 24: Alinhamentos dos *uniques* com o retrotransposon Saci-2 contendo LTR . Em cinza início e fim da seqüência codificadora da poliproteína e em azul região LTR. Em vermelho o alinhamento dos *uniques* encontrados nas bibliotecas de fêmeas, adultos e ovos. Em verde alinhamento dos *uniques* encontrados nas bibliotecas de adultos e fêmeas e em rosa . alinhamento do *unique* encontrado na biblioteca de ovos. Seqüência com ID BK004069 depositada no NCBI.

```

gcccttttgg agagaggcag gaaatactat gtttccgtac actaaatggg gacgaataaa 61
actgttgaag attgtagctg taaatagttc cccaaaatca atagctttct aagtgttcga 121
tattggatgc tgaccacggt gttaaagtgg acttgtttaa ctgaaggctt tcggtcatgc 181
atccgtacaa gatcacggtt caactcggcg tgggacactg ctctacggtt cactagtcag 241
cttagaataa acgttcctca atatattgat aacgtatcaa atatatcttt cctacctctt 301
ctcgtctgac ttctgatttc tatttggcca agaaggcttc gattatggaa ggtgctactg 361
gtagctaatt gtaaagttag aatactagtc gtatcctcag tggtgagcct gacgtgatct 421
ggtacaccaa gccataaac tgcgagtctg ttcctttttg gaatattatt attcattggt 481
attctttatt tgatttttat atattgtgat atactttttt tgcgagttt atcccggata 541
tattttaatt agacatcttt cgttctttat ataaaatga cggaacactc acctaagggt 601
ttaaagttaa aatctattcc ccctatttcg attcagttaa cgccattttg gcttgacaat 661
.....
cgacagcata ccaccacaa gcaaacgggt tggtagaacg ttttcatcga caattgaatg 3901
cttcattatc agctgcaaac gtttcacagt ggacagacgc tcttacactc gtcttgctag 3961
gtatccgcaa tgcagtgaag gctgacattg gataaacggc atctcaactc gtttacggaa 4021
cgacactccg acttccagga gaattcgtgg atccttcggc ttctcattg aacatggatc 4081
.....
.....aagtacttca gtaacagaca acaaaactaa atggttatac ctttaagtatt 4741
cctccatgaa tattgtgtgc tctgcttttc gtccttcctt tttcttacac ctttttcagc 4801
cgtctgaaga tactccaccg aaaacagata tggcacgggt gacttcgtcc tacaccgta 4861
129 aga tacgccaccg aaaacagata tggaacgggt gacttcgtcc tacaccacta
ttaaagcttc cgatattctg gagggatata agttcgaaac ttctaaaagc ccccaacttc 4921
ttaaagcctt agatattcta aagttatata aggttgaagc ttctaaaatc ccccaacttc
gcatgtatct attctcgacc atcagttgta tgttgcaacg ctacctccat aaagtgcgcc 4981
gcatggatct attctcgacc atcagttgta tgtcgcaatg ccaccttcat aaagccctcc
aagtgtg--- ttattgaaat gtttcttacg aaatccattt acattcatac cttatagaat 5041
aactgtgata ttggcaacac gttcctcact aaaacctctt acattcaaat cttataaaat
atcaaccag ctgagtaatt ttcaactgtg aaaattgtaa cttctttaac aatttgatct 5101
atcaaccag ctgagtaatt tactaccgtg agaattgtaa cttttgtagc aattcgatct
tgctgtaaa ccagcatgcc tcatgtaaca agctgttatt tgagaataaa ttattatcat 5161
tgctgtaaa ctgcatgcc tcacgtaaca aactgtaatt tgagaataaa tcattat 358
tattattataaaaaataatttttttaatgtgctcattttttta ttcatttaa aaattattat 5221
tttttgagc gtatttgtgt tttttatata ataaaaaac catttttttc cgatcgcttt 5281
tacgctgttt gtaagtgtat tagttgtttt tttcccgtc catcttgtgt ccttacgcaa 5341
gtacgagtgc attttcgaca tgcactcaca cgttctattt tttctctttt ccaggacggc 5401
tggaaaagaa cgacgaagtt tgcgaaggaa ccgaaatttt cattgtacct tatttctaa 5461
ttgctatggt gtacattttg gtcccttagt gtaggaaag acgaccagac gctgctaaac 5521
gaagcaagct atcagaagag tgtttaccac cgacaaactc gttgggttta aaattctggc 5581
cggtcacgct ttaggggcat cactacctca ctagggggg agtgatectg agctgtaaat 5641
aattcccaa aatcaatagc tttctaagt ttcgatattg gatgctgacc acgttgttta 5701
agtggacttg ttaactgaa ggctttcggc catgcatccg tacaagatca cgtttcaact 5761
cggcgtggga cactgctcct acgttcacta gtcagcktag aataaacggt cctcgatata 5821
ttgataacgt atcaaatata tatattatat atttggccaa gaaggtttcg attatagaag 5881
gtgctactgg tagctaattg taaagttaga aatactagtc tatcctcaag attatgtgga 5941
aagttctacc gtcagattgg tcaaaaatga gcttcaatgt taccagtgca aggtttgatt 6001
gaaaggcgtt tctgacacag ttagcgtagg acaagggc

```

Figura 25: Alinhamento do *unique* com o retrotransposon Boudicca contendo LTR . Em cinza início e fim da seqüência codificadora da poliproteína e da proteína gag não funcional devido a mutação e em azul região LTR. Em vermelho o alinhamento do *unique* encontrado na biblioteca de adultos. Senqüência com ID AY662653 depositada no NCBI

Parece-nos muito inusitado a ocorrência de *trans splicing* em retrotransposons e mesmo em regiões LTR de retrotransposons. Seria difícil imaginar a causa deste processo e as conseqüências advindas da produção de transcritos contendo parte de retroposons ou a LTR inteira ou parcial. Pelo nosso conhecimento nada parecido foi descrito na literatura.

Verificamos a presença de grande quantidade de proteínas ribossomais entre os transcritos produzidos por *trans splicing*. Dado semelhante foi descrito em *C. elegans*, onde estes genes são, inclusive, encontrados em operons (Blumenthal, 1995). Similarmente, em *Oikopleura dióica*, um organismo evolutivamente distante de *S. mansoni*, observou-se que entre 145 ESTs geradas por *trans splicing*, 43 eram proteínas ribossomais (Ganot *et al.*, 2004).

Identificamos, também, a presença do transcrito de *S. mansoni ubiquinol-cytochrome C reductase complex ubiquinol binding protein* (UbCRBP) / enolase que já havia sido descrito como sendo um transcrito policistrônico, e apenas a enolase sendo processada por *trans splicing* (Davis & Hodgson, 1997). No entanto, UbCRBP foi identificada neste trabalho entre os transcritos processados por *trans splicing*. Identificamos que a localização da inserção do *spliced leader* é exatamente na posição 1474 nt., no início do segundo exon do gene, onde encontramos o dinucleotídeo AG que indica um sítio receptor de *splicing*. Tudo isso sugere um *splicing* alternativo de *trans splicing*, assim como ocorre para os transcritos da 3-hidroxi-3-metil-glutaril Coa redutase em *S. mansoni* (Rajkovic *et al.*, 1995), e talvez por essa razão não tenha sido identificado anteriormente. De fato, seria esperado que este mRNA sofresse *trans splicing*, devido a sua localização em um operon, como acontece para a maioria dos mRNAs provenientes de transcritos policistrônicos em outros organismos. Além disso, já foi descrito *trans splicing* para este transcrito em *E. multilocularis*, reforçando a idéia de que existe uma conservação do mecanismo de *trans splicing* em genes ortólogos. Há ainda, a probabilidade deste gene ser processado da mesma forma que o gene de cópia única *elp*, em *E. multilocularis*, ora convencionalmente, ora por *trans splicing*.

Além dos genes discutidos, podemos ressaltar da lista de transcritos que sofrem *trans splicing* aqueles codificadores das diferentes subunidades do proteossomo 26S, diferentes U snRNPs, diferentes peptidil-prolil-isomerases e proteínas que com ela interagem e também

proteínas envolvidas com tráfego intracelular. Dessa forma, começa-se a delinear grupos de proteínas envolvidos em processos similares entre os transcritos estudados.

Pôde ser observado que, mesmo após o estabelecimento de um novo critério para otimizar a anotação das seqüências, ainda restaram um grande número de *uniques* com homologia unicamente com seqüências depositadas de *S. japonicum*. Isto poderia indicar que estes genes são específicos do gênero *Schistosoma*.

Todos os *uniques* anotados que são similares às seqüências de *T. solium*, sofrem *trans splicing* neste cestódeo. Além da similaridade com os transcritos de *Taenia*, *Oikopleura* e *C. elegans*, observamos alta conservação do processo entre os poucos transcritos estudados de *Echinococcus*. Entre estes, além da UbCRBP, anteriormente descrita, foram encontrados em *Echinococcus*: *Enhancer of rudimentary homolog*, U snRNAs e *Sm like proteins*. Estes achados corroboram a hipótese de que há de fato alguma conservação entre os transcritos, pois seria difícil ver ao acaso tamanha similaridade entre mRNAs produzidos por *trans splicing*, em meio a enorme infinidade de genes expressos nestes organismos.

4.6.6. Classificação dos uniques produzidos por *trans splicing* em categorias funcionais

Os *uniques* que tiveram uma similaridade significativa após o BLAST contra os bancos de dados COG e KOG (167 *uniques*), foram submetidos a um processo de classificação funcional, onde foram divididos nas categorias propostas pelo COG. Assim, a divisão destes *uniques* nas principais categorias dos COGs gerou os dados mostrados na Figura 26.

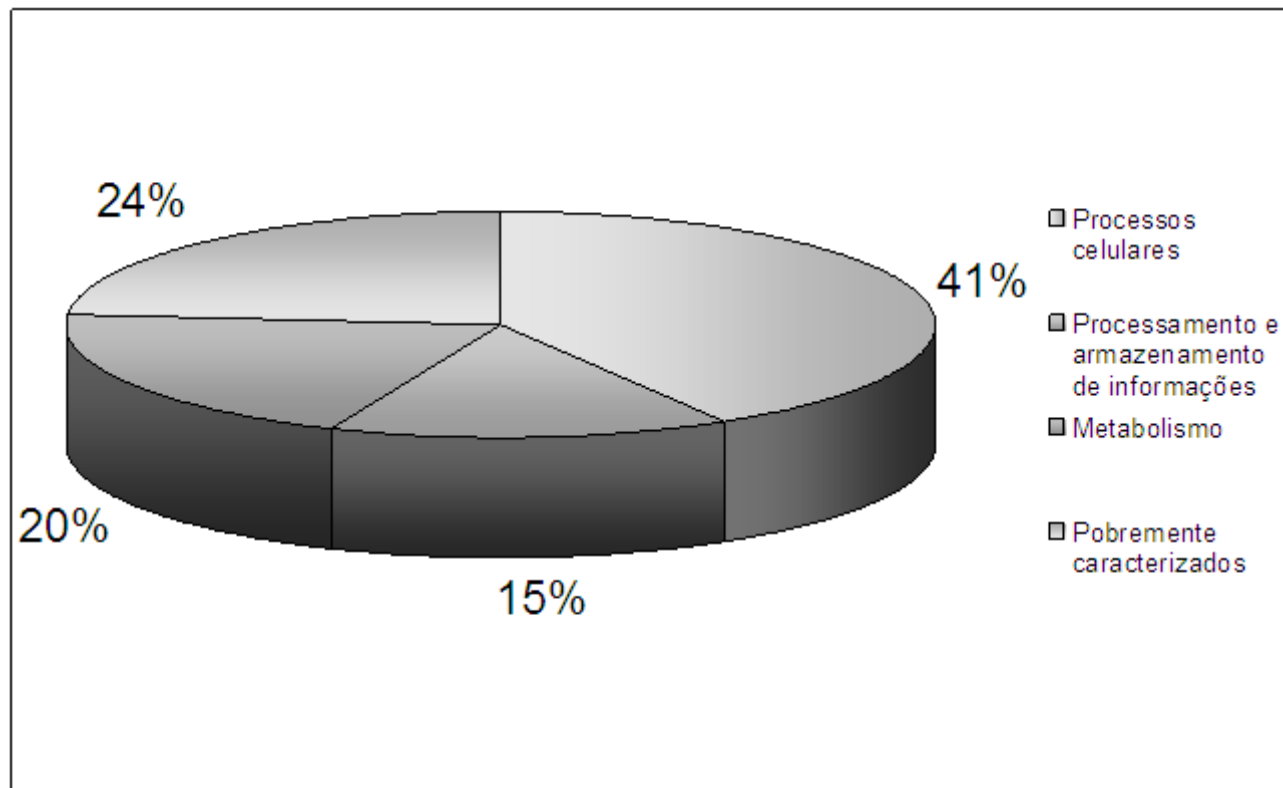


Figura 26: Divisão dos *uniques* anotados nas princ'ipais categorias dos COG/KOGs.

Analisando a Figura 26, vemos que 24% dos *uniques* não têm sua função bem caracterizada, apesar de muitos deles apresentarem similaridade com genes ou proteínas depositadas em banco de dados, mostrando a importância da genômica funcional para tais esclarecimentos.

Verjovski-Almeida e colaboradores (2003), classificaram 30.988 *uniques* nas categorias do Gene Ontology (GO), e observaram que as categorias mais frequentes foram de metabolismo de proteína e de metabolismo de ácidos nucleicos. Ao compararmos essa informação com as informações geradas neste trabalho, vimos que metabolismo não é a categoria em que a maioria dos transcritos que sofrem *trans splicing* está inserida. Para obtermos uma visão mais detalhada da contribuição de cada biblioteca para esta classificação, dividimos os *uniques* em cada uma das subcategorias funcionais dos COGs (Figura 27).

È importante notar que na categorização dos *uniques*, alguns podem estar representados em mais de uma categoria, devido à artificialidade dos grupos de classificação, uma vez que a natureza não exhibe padrões fixos que gostaríamos de criar para melhor entendermos o funcionamento das proteínas.

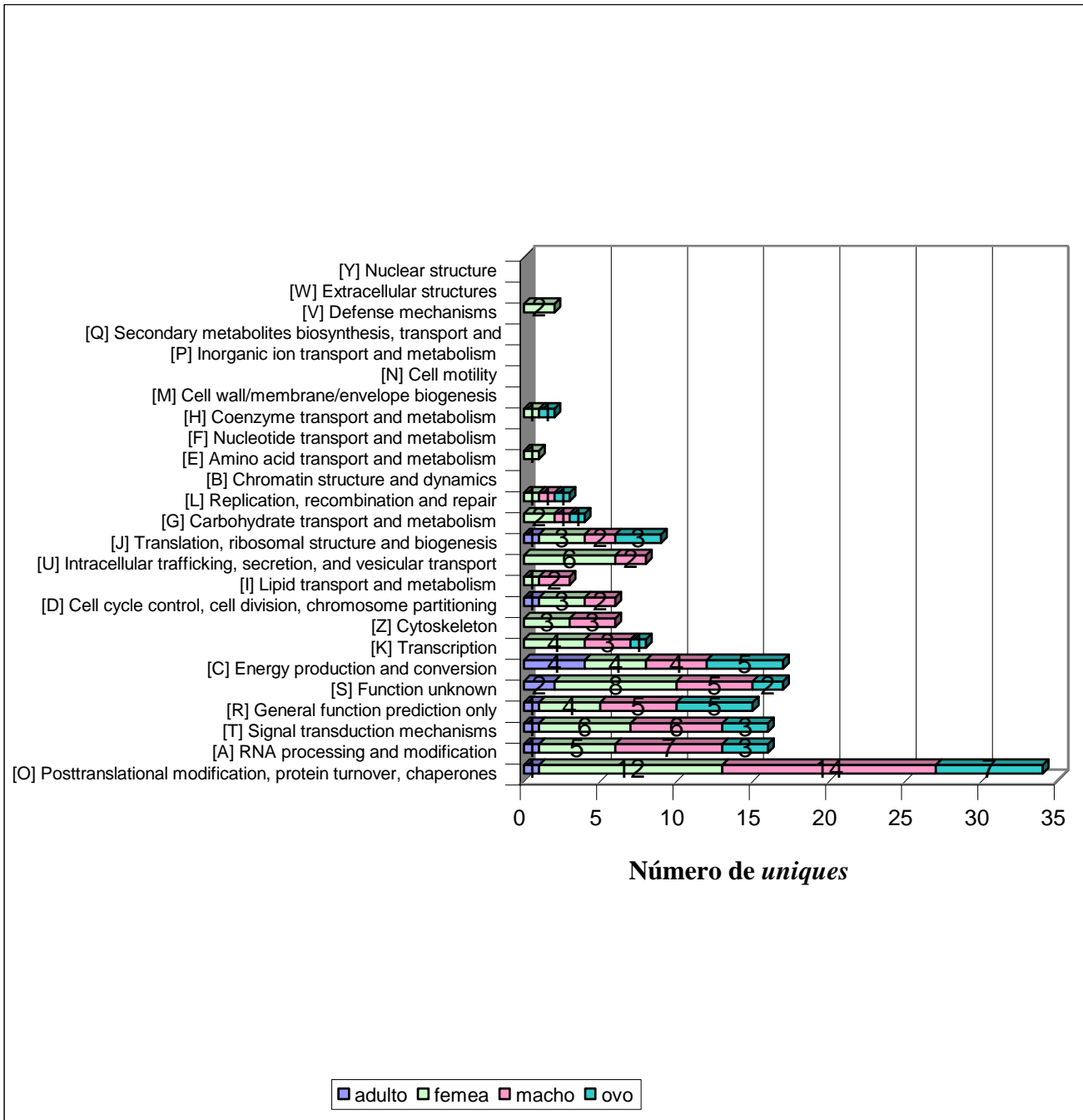


Figura 27: Classificação dos *uniques* de cada biblioteca nas sub-categorias dos COG/KOGs.

Analisando os *uniques* classificados nas categorias dos COGs, vimos que a categoria que inclui proteínas de modificação pós traducional, *turnover* de proteínas e chaperones foi representada por 20% dos transcritos produzidos por *trans splicing*. Este valor é o dobro da representação encontrada para as categorias de produção e conversão de energia e para a categoria de proteínas com funções desconhecidas, que são as segundas mais representadas entre elas. Ao classificar em categorias dos COGs os transcritos depositados de *S. mansoni*, Prosdocimi e colaboradores (2002) observaram que 10,4% (199) dos genes se encaixaram na categoria de modificação pós traducional, *turnover* de proteínas e chaperones. Neste trabalho as categorias mais representadas foram as categorias de funções desconhecidas (18,95% - 360 genes), a de predição da função em geral (12,68% - 241 genes) e a de mobilidade celular, secreção, transporte intracelular e citoesqueleto (10,58% - 201 genes) (Prosdocimi *et al.*, 2002). Estas categorias estão menos representadas entre os transcritos que sofrem *trans splicing*, como visto na nossa análise. Uma grande diferença que pôde ser observada nos dois estudos, foi a apresentação da categoria de produção e conversão de energia, onde verificamos 10,2% dos *uniques* que sofrem *trans splicing*, contra apenas 3,53% no trabalho de 2002. Estas diferenças encontradas devem corresponder à realidade quando consideramos que existe, de fato, neste estudo uma seleção de transcritos que sofrem *trans splicing* e que pertencem a algumas categorias funcionais específicas. Estamos cientes, no entanto, da existência da possibilidade de amplificação preferencial de transcritos, o que poderia produzir um viés na nossa análise.

Proteínas *heat shock* foram encontradas também, entre os transcritos de *O. dioica* (Ganot *et al.*, 2004) podendo ser mais um ponto a ser considerado, enfatizando a conservação deste mecanismo de processamento de RNA nesta categoria.

4.6.7. Identificação de novas ESTs

Para identificarmos novas ESTs que foram geradas por este projeto de sequenciamento de bibliotecas enriquecidas em transcritos produzidos por *trans splicing*, foi feito BLAST contra as seqüências produzidas pela Rede Genoma de Minas Gerais e contra as seqüências já depositadas

no dbEST (incluindo as seqüências depositadas pela rede ONSA). Na Figura 28, vemos o resultado desta análise.

Após o grande avanço na análise do transcriptoma do parasita *S. mansoni* tanto pela Rede ONSA, onde foi descrito que produziu-se uma cobertura de 92% do transcriptoma deste organismo (Verjovsk-Almeida *et al.*, 2003), quanto pela comunidade científica mundial, esperávamos que a grande maioria das seqüências geradas tivessem similaridade com seqüências depositadas no dbEST, e que praticamente não houvessem novos transcritos a serem descobertos. Realmente, 67,2% dos *uniques* gerados por este projeto estão representados nas seqüências depositadas no dbEST, mas, ainda haviam outras ESTs com similaridade às seqüências geradas pela Rede Genoma de Minas Gerais (6,7%), e mais 26,1% de transcritos que foram produzidos exclusivamente através da estratégia aqui utilizada.

Agrupamos os 168 *uniques* das várias bibliotecas que ainda não haviam sido descobertos, para retirarmos a redundância que pudesse haver entre eles e assim verificar quantos transcritos novos havíamos obtido (Figura 29).

Foram gerados 156 *uniques* a partir dos 168 novos transcritos obtidos neste projeto. Esta grande quantidade de singlets geradas após o agrupamento, pode ser devido tanto à pouca representatividade dos transcritos, quanto ao seu pequeno tamanho, o que não contribui para a formação de *clusters*.

Alguns dos novos transcritos puderam ser anotados, e estão descritos na Tabela 9.

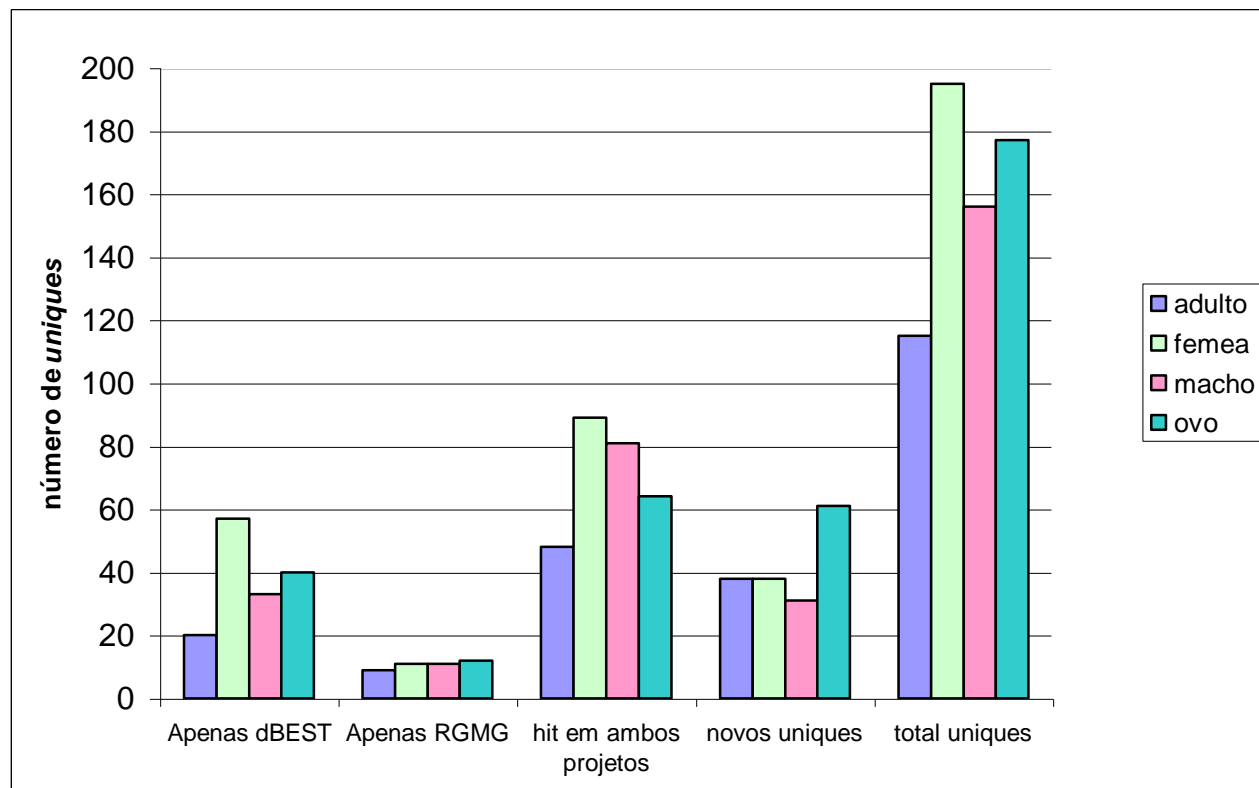


Figura 28: *Uniques* com similaridade às ESTs depositadas no SmdbEST ou produzidas pela RGMG.

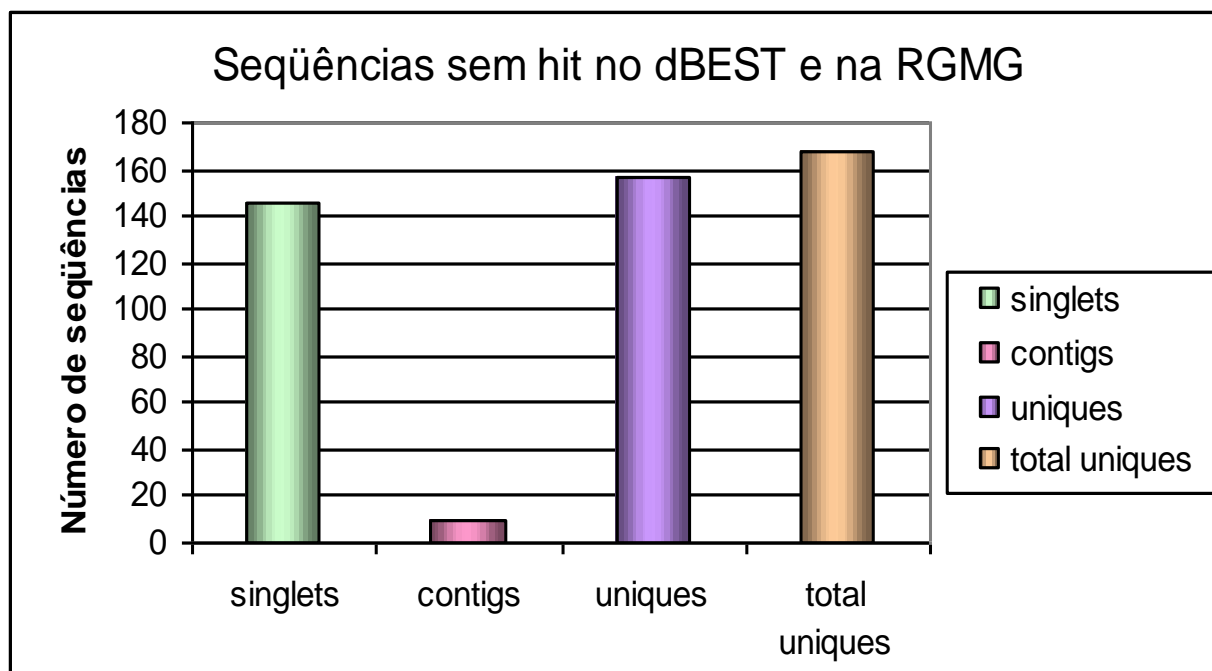


Figura 29: Seqüências geradas após o agrupamento por CAP3 dos novos *uniques*.

Tabela 9: Anotação dos novos genes gerados neste estudo.

nome unique	Anotação
Contig1Ma	Similar to <i>Danio rerio</i> COP9 signalosome complex subunit 5
Contig19Ma	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04305 unknown mRNA
Contig45Ma	Similar to <i>Xenopus laevis</i> U1 snRNP component
Contig55Ma	Similar to <i>D. melanogaster</i> COP9 signalosome complex subunit 5
Contig71Ma	Similar to <i>mus musculus</i> protein SYS1 homolog
MA4r_A02	Similar to <i>Mus musculus</i> anaphase promoting complex subunit 11
MA4r_G05	Similar to <i>Schistosoma japonicum</i> clone SJCHGC08017 unknown mRNA
MA5r1_C04	Similar to <i>C. elegans</i> peptidyl-prolyl cis-trans isomerase 10 (rotamase) (cyclophilin-10)
MA7r_A12	Similar to <i>Schistosoma japonicum</i> IB1 protein mRNA, complete cds
MA8r_C03	Similar to <i>Homo sapiens</i> tRNA (guanine-N(7)-methyltransferase (methyltransferase-like protein 1)
MA9rf_G12	Similar to <i>Trypanosoma cruzi</i> Rad51 gene, complete cds
Contig46AD	Similar to <i>Homo sapiens</i> hypothetical protein HSPC016
Contig7fe	Similar to <i>Schistosoma japonicum</i> clone SJCHGC04305 unknown mRNA
Contig87fe	Similar to <i>Mus musculus</i> protein SYS1 homolog
FE1f_H06	Similar to <i>Schistosoma japonicum</i> clone SJCHGC01436 unknown mRNA
FE3f_B10	Similar to <i>Schistosoma japonicum</i> clone ZZD309 mRNA sequence
FE3f_H04	Similar to <i>Bos taurus</i> alpha-(1,6)-fucosyltransferase
FE8r_F06	Similar to <i>Poecilia reticulata</i> 6-pyruvoyl tetrahydrobiopterin synthase
Contig67o	Similar to <i>Homo sapiens</i> peptidyl-prolyl cis-trans isomerase 10
Contig71o	Similar to <i>Schistosoma japonicum</i> clone SJCHGC09496 unknown mRNA
Contig46Ad	Similar to <i>Homo sapiens</i> hypothetical protein HSPC016

Observamos, após anotação, que entre os novos transcritos analisados, estão a proteína Rad 51, representada por um dos *uniques*. A Rad 51 é uma importante proteína envolvida no reparo de DNA por recombinação homóloga, através da formação de uma nucleoproteína em regiões de fita única de DNA, que catalisa a procura de seqüências homólogas e pareamento das fitas.

Dois *uniques* diferentes, da biblioteca de machos apresentaram homologia com a subunidade 5 da proteína COP9, de *D. melanogaster* e com a de *D. rerio*. Esta proteína, COP9 signalosome (CSN) um complexo multiproteico da via ubiquitina-proteassomo, composta de 8 é subunidades, as quais estão relacionadas às subunidades da partícula regulatória do proteassomo 26S (Schwechheimer, 2004).

Um dos *uniques* foi anotado como componente do U1 snRNP. O U1 snRNP é essencial para o reconhecimento do sítio de *splicing* 5' do pre-mRNA e para a ligação subsequente do spliceossomo. Em levedura, U1 snRNP possui 16 componentes (Gottschalk *et al.*, 1998), e em *S. mansoni*, pelo menos um destes é produzido por *trans splicing*. O interessante é que U1, em Kinetoplastideos, não parece estar presente na formação do spliceossomo para a realização do processo de *trans splicing*, pois não apresenta potencial para interação com o sítio de *splicing* do SL RNA (Liang *et al.*, 2003), e, entretanto, é produzido por este processo.

Outra nova proteína identificada em *S. mansoni*, é a subunidade 11 do complexo promotor de anaphase (APC). Esta é uma ligase proteína-ubiquitina, composta de 11 subunidades, que cataliza a poliubiquitinação para a destruição de reguladores mitóticos para segregação cromossômica durante a mitose (Carroll *et al.*, 2005). Identificamos também a proteína IB1 que participa nas vias de transdução de sinal, principalmente por interação proteína-proteína. A seqüência identificada possui homologia com esta proteína em *S. japonicum*.

Estão entre os transcritos ainda não identificadas de *S. mansoni*, a Alfa-(1,6)-fucosiltransferase que cataliza a transferência de fucose do GDP-Fuc para alguns complexos de glicoproteínas (Miyoshi *et al.*, 1999) e a enzima 6-piruvil tetrahydrobiopterina sintase, que é uma enzima envolvida na complexa química orgânica de transformação do GTP em tetrahydrobiopterina (Auerbach & Nar, 1997). Também o transcrito codificador da proteína

peptidyl-prolyl cis-trans isomerase 10, sofre *trans splicing* e está entre os novos transcritos identificados para *S. mansoni*. Peptidyl-prolyl cis-trans isomerases catalisa a interconversão das ligações peptídicas cis e trans e são consideradas importantes para o enovelamento de proteínas. Essa proteína parece participar nos processos de sinalização, reconhecimento da superfície celular, como chaperone e resposta de choque térmico (Henriksson *et al.*, 2004). Esta, juntamente com a subunidade 5 da proteína COP9 e a subunidade 11 do complexo promotor de anaphase, seriam proteínas pertencentes à categoria de modificação pós traducional, *turnover* de proteínas e chaperones, que foi vista ser a mais representada entre as categorias dos transcritos que sofrem *trans splicing*. Esta proteína foi identificada também entre as proteínas de *E. granulosus* (Fernández *et al.*, 2002).

Um grande número de *uniques* (28,58%) que apresentaram homologia com seqüências depositadas em bancos de dados têm similaridade com seqüências não identificadas de *S. japonicum*. Esta alta porcentagem de seqüências com similaridade à *S. japonicum* nos mostra a importância para identificação gênica do sequenciamento de genomas/transcriptomas de organismos evolutivamente próximos ao *S. mansoni*.

A maioria dos possíveis novos transcritos não têm homologia com seqüências depositadas nos bancos de dados utilizados para anotação. Estes, possivelmente, são genes específicos de *S. mansoni*.

5. Conclusões

A partir das bibliotecas geradas, foram produzidas 2372 seqüências de boa qualidade, ou seja, maiores que 100 pares bases após a retirada das seqüências de vetores, das caudas de poli A e da seqüência do *spliced leader*.

As bibliotecas apresentaram pouca redundância, com uma média de menos de 5% de clusters com mais de 20 seqüências.

Parece haver conservação entre os transcritos que sofrem *trans splicing* entre os organismos em que este mecanismo já foi identificado, e entre as proteínas envolvidas em um mesmo processo dentro de um organismo.

A categoria funcional que inclui proteínas envolvidas em modificação pós-traducional, turnover de proteínas e chaperones foi a categoria do COG mais representada por genes produzidos por *trans splicing*.

Foram gerados 156 novos genes através da técnica de sequenciamento de clones de bibliotecas enriquecidas em transcritos que sofrem *trans splicing*.

6. Perspectivas

- ❖ Análise bioinformática das seqüências da biblioteca da forma que se assemelha a fase de esquistossômulos.

- ❖ Depósito das seqüências no dbEST.

- ❖ Comparação das seqüências produzidas nos diversos estágios do ciclo de vida do *S. mansoni*

- ❖ Comparar através de BLAST, os transcritos que sofrem *trans splicing* entre todos os organismos em que este mecanismo já foi identificado, para confirmarmos a conservação e se há algum padrão entre mRNAs que sofrem *trans splicing* em organismos diferentes.

7. Referências Bibliográficas

ADAMS, M.D., *et al.* (1993) Rapid cDNA sequencing (Expressed Sequence Tags) from a directionally cloned human infant brain cDNA library. *Nature Genet* 4: 373-389.

AGABIAN, N. (1990) Trans splicing of nuclear pre-mRNAs. *Cell* 61:1157-1160.

ALTSCHUL, S.F., *et al.* (1990) Basic local alignment search tool. *J. Mol Biol* 215: 403-410.

AUERBACH, G.; NAR, H. (1997) The pathway from GTP to tetrahydrobiopterin: three-dimensional structures of GTP cyclohydrolase I and 6-pyruvoyl tetrahydropterin synthase. *J. Biol Chem* 378: 185-192.

BEKTESH, S.L.; HIRSH, D.I. (1988) *C. elegans* mRNAs acquire a spliced leader through a trans-splicing mechanism. *Nucleic Acids Res* 16:5692.

BLAXTER, M.; LIU, L. (1996) Nematode spliced leaders-ubiquity, evolution and utility. *Int J Parasitol* 26: 1025-1033.

BLUMENTHAL, T. (1995) Trans splicing and polycistronic transcription in *Caenorhabditis elegans*. *Trends in Genet* 11: 132-136.

BLUMENTHAL, T.; GLEASON, K.S. (2003) *Caenorhabditis elegans* operons: form and function. *Nat Rev Genet* 4: 112-120.

BONEN, L. (1993) Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J* 7:40-46.

BREHM, K., *et al.* (2000) mRNA trans splicing in the human parasitic cestode *Echinococcus multilocularis*. *J Biol Chem* 275: 38311-38318.

- BREHM, K., *et al.* (2002) Characterization of a spliced leader gene and of trans-spliced mRNAs from *Taenia solium*. *Mol Biochem Parasitol* 122: 105-110.
- CARROLL, C.W., *et al.* (2005) The APC subunit Doc1 promotes recognition of the substrate destruction box. *Curr Biol* 15:11-18.
- COPELAND, C.S., *et al.* (2004). Structural and evolutionary analysis of the transcribed sequence of Boudicca, a *Schistosoma mansoni* retrotransposon. *Gene* 329: 103-114.
- DAVIS, R.E. (1996) Spliced leader RNA trans splicing in metazoa. *Parasitol Today* 12: 33-40.
- DAVIS, R.E., *et al.* (1995) RNA trans splicing in flatworms. Analysis of trans-spliced mRNAs and genes in human parasite, *Schistosoma mansoni*. *J Biol Chem* 270: 21813-21819.
- DAVIS, R.E.; HODGSON, S., (1997) Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in *Schistosoma mansoni*. *Mol Biochem Parasitol* 89: 25-39.
- DEMARCO, R., *et al.* (2004) Saci-1, -2, and -3 and Perere, four novel retrotransposons with high transcriptional activities from the human parasite *Schistosoma mansoni*. *J Virol* 78: 2967-2978.
- EL-SAYED, N.M.A, *et al.* (2004) Advances in schistosome genomics. *Trends in Parasitol* 20: 154-157.
- ENGELS, D., *et al.* (2002) The global epidemiological situation of schistosomiasis and new approaches to control and research. *Acta Tropica* 82: 139-146.
- FERNÁNDEZ, C., *et al.* (2002) Full-length-enriched cDNA libraries from *Echinococcus granulosus* contain separate populations of oligo-capped and trans-spliced transcripts and a high level of predicted signal peptide sequences. *Mol Biochem Parsitol* 122: 171-180.
-

- FRANCO, G.R., *et al.* (2000) The *Schistosoma* gene discovery program: state of the art. *Int J Parasitol* 30: 453-463.
- FRANCO, G.R., *et al.* (1997) Evaluation of cDNA libraries from different developmental stages of *Schistosoma mansoni* for production of expressed sequence tags (ESTs). *DNA Res* 4:231-240.
- FRANCO, G.R., *et al.* (1995) Identification of new *Schistosoma mansoni* genes by the EST strategy using a directional cDNA library. *Gene* 152: 141-147.
- GANOT, P., *et al.* (2004) Spliced-leader RNA trans splicing in a chordate, *Oikopleura dioica*, with a compact genome. *Mol Cell Biol* 24: 7795-7805.
- GEMS, D.H., *et al.* (1995) An abundant, trans spliced mRNA from *Toxocara canis* infective larvae encodes a 26 kDa protein with homology to phosphatidylethanolamine binding proteins. *J Biol Chem* 270:18517-18522.
- GOTTSCHALK, A., *et al.* (1998) A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *RNA* 4: 374-393.
- GREGORY, W.F., *et al.* (1997) Differentially expressed, abundant trans spliced cDNAs from larval *Brugia malayi*. *Mol Biochem Parasitol* 87: 85-95.
- HARSTINGS, K.E.M. (2005) SL trans splicing: easy come or easy go? *Trends in Genet* 21: 240-247.
- HAUSDORF, B. (2000) Early evolution of the bilateria. *Syst Biol* 49: 130-142.
- HENRIKSSON, L.M., *et al.* (2004) X-ray structure of peptidyl-prolyl cis-trans isomerase A from *Mycobacterium tuberculosis*. *Eur J Biochem* 271: 4107-4113.

- LE PASLIER, M.C., *et al* (2000) Construction and characterization of a *Schistosoma mansoni* bacterial artificial chromosome library. *Genomics* 65: 87-94.
- LE, T.H., *et al.* (2001) Mitochondrial gene content, arrangement and composition compared in African and Asian schistosomes. *Mol Biochem Parasitol* 117: 61-71.
- LIANG, X. H., *et al.* (2003) Trans and cis splicing in Trypanosomatids: mechanism, factors, and regulation. *Eukariotic Cell* 5: 830-840.
- LICHTENBERG, F. (1987) The biology of schistosomes: from genes to latrines. *Academic Press* London.
- LOVERDE, P.T., *et al.* (2004) *Schistosoma mansoni* genome project: an update. *Parasitol International* 53: 183-192.
- MARONEY, P.A., *et al.* (1995) Most mRNAs in the nematode *Ascaris lumbricoides* are trans-spliced: a role for spliced leader addition in translational efficiency. *RNA* 1: 714-723.
- MILLER, R.T., *et al.* (1999) A comprehensive approach to clustering of expressed gene sequence: The sequence tag alignment and consensus knowledge base. *Genome Res* 9: 1143-1155.
- MIYOSHI, E., *et al.* (1999) The alpha1-6-fucosyltransferase gene and its biological significance *Biochem Biophys Acta* 1473: 9-20.
- NELSON, D.L.; COX, M.M. (2000) Lehninger Principles of Biochemistry. 3ª edição. New York. Worth Publishers
- NEVES, D.P. (2000) Parasitologia Humana. 8ª edição. Rio de Janeiro e São Paulo. Livraria Ateneu.
- NILSEN, T.W. (1995) Trans splicing: an up date. *Mol Biochem Parasitol* 73: 1-6.

OLIVEIRA, G.; JOHNSTON, D.A. (2001) Mining the schistosome DNA sequence database. *Trends Parasitol* 17: 501-503.

PANNONE, B.K, *et al.* (2001) Multiple functional interactions between components of the Lsm2-Lsm8 complex, U6 snRNA, and the yeast La protein. *Genetics* 158: 187-196.

PESSOA, S.B.; MARTINS, A.V. (1982) *Parasitologia Médica*. 11^a edição. Rio de Janeiro, Guanabara Koogan.

PIAO, Y., *et al.* (2001) Construction of long-transcript enriched cDNA libraries from submicrograms amounts of total RNAs by a universal PCR amplification method. *Genome Res* 11: 1553-1558.

PROSDOCIMI, F; PEIXOTO, F.C.; ORTEGA, J.M. (2004) Evaluation of window cohabitation of DNA sequencing errors and lowest PHRED quality values. *Genet Mol Res* 3: 483-492.

PROSDOCIMI, F., *et al.* (2002) Clustering of *Schistosoma mansoni* mRNA sequences and Analysis of the most transcribed genes: implications in metabolism and biology of different developmental stages. *Mem Inst Oswaldo Cruz* 97: 61-69.

RAJKOVIC, A., *et al.* (1990) A spliced leader is present on a subset of mRNAs from the human parasite *Schistosoma mansoni*. *Proc Natl Acad Sci U. S. A.* 87: 8879-8883.

SCHWECHHEIMER, C. (2004) The COP9 signalosome (CSN): an evolutionary conserved proteolysis regulator in eukaryotic development. *Biochim Biophys Acta* 1695: 45-54.

SHORT, R.B.; GROSSMAN, A.I. (1981) Conventional giensa and C-banded Karyotypes of *Schistosoma mansoni* and *S. rodhaini*. *J parasitol* 67: 661-667.

- SIMPSON, A.J.G.; SHER, A.; MeCUTCHAN, T.F. (1982) The genome of *Schistosoma mansoni*: isolation of DNA, its size, bases and repetitive sequences. *Mol Biochem Parasitol* 6: 125-137.
- STURROCK, R.F. (2001) Schistosomiasis epidemiology and control: how did we get here and where should we go? *Mem Inst Oswaldo Cruz* 96: 17-27.
- TATUSOV, R.L.; KOONIN, E.V.; LIPMAN, D.J. (1997) A genomic perspective on protein families. *Science* 278: 631-637.
- VANDENBERGHE, A.E., *et al.* (2001) mRNA 5'-leader trans splicing in a chordate. *Genes Dev* 15: 294-303.
- VERJOVSKI-ALMEIDA, S., *et al.* (2003) Transcriptome analysis to the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics* 35: 148-157.
- WAN, C., *et al.* (2005) Structure of the conserved transcriptional repressor enhancer of rudimentary homolog. *Biochemistry* 44: 5017-5023.
- WEBBE, G.; EL HAK, S. (1990) Progress in the control of schistosomiasis in Egypt 1985-1988. *Trans R Soc Trop Med Hyg* 84: 394-400.
- ZANOTTI, F., *et al.* (2004) Inhibitory and anchoring domains in the ATPase inhibitor protein IF1 of bovine heart mitochondrial ATP synthase. *J Bioenerg Biomembr* 36: 447-457.