



DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA
PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

TESE DE DOUTORADO

**Identificação e caracterização de microRNAs
de *Schistosoma mansoni***

Por

Mariana Crivellari Machado Simões

Belo Horizonte, MG
2009

Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa de Pós-graduação em Bioinformática
Departamento de Bioquímica e Imunologia

Mariana Crivellari Machado Simões

**“Identificação e caracterização de microRNAs de
Schistosoma mansoni”**

Tese apresentada ao Programa de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais como requisito parcial à obtenção do título de Doutor em Bioinformática.

ÁREA DE CONCENTRAÇÃO: BIOINFORMÁTICA GENÔMICA

Orientador: Dr. Guilherme Oliveira
Co-orientador: Dr. Najib El-Sayed

Agosto, 2009

*Dedico esse trabalho
à minha mãe Silvia e ao meu
pai Henrique pelo exemplo
de vida, apoio, determinação
e amor.*

Esse trabalho foi iniciado no The Institute for Genomic Research (TIGR - EUA), durante estágio de doutorado sanduíche realizado no período de Setembro de 2005 à Novembro de 2006 em colaboração com o grupo do Dr. Najib El-Sayed e foi concluído no Laboratório de Parasitologia Celular e Molecular do Centro de Pesquisas René Rachou, Fiocruz, sob a supervisão do Dr. Guilherme Oliveira. O projeto contou com o suporte financeiro da CAPES, FIOCRUZ e Fogarty-NIH.

AGRADECIMENTOS

Agradeço primeiramente ao Dr. Guilherme Corrêa Oliveira que sempre esteve presente nos meus agradecimentos desde a minha iniciação científica, mostrando assim o seu apoio durante toda esta jornada, com amizade, profissionalismo, comprometimento e carinho. Obrigada pela confiança que demonstrou por mim em todas as suas atitudes, me proporcionando grandes oportunidades.

Ao Dr. Najib El-Sayed, meu orientador no TIGR, pela oportunidade e por toda amizade, incentivo e cobrança, que foram imprescindíveis durante a minha caminhada.

Ao Dr. Appolinaire Djikeng, a quem tenho extrema admiração e amizade, por sua participação fundamental nos meus experimentos de bancada.

Ao Dr. Phillip Lo-Verde, pelo suporte financeiro durante meu doutorado sanduíche no exterior.

Ao Dr. Gustavo Cerqueira e Adhemar Zerlotini que foram meu braço direito nos experimentos de bioinformática.

A Dra. Rosiane A. Silva-Pereira, pela colaboração, boa vontade e disponibilidade durante os experimentos de *northern blot*.

Ao Centro de Pesquisas René Rachou e ao Dr. Álvaro Romanha, pelas facilidades concedidas para o desenvolvimento deste trabalho.

Ao pessoal do setor administrativo e da Pós-graduação da Bioinformática (UFMG) que sempre procurou viabilizar com eficiência e boa vontade as questões burocráticas relacionadas a este trabalho, em especial ao Alberto e ao Carlos.

A Dra. Glória Franco, uma das grandes responsáveis pelo sucesso do curso de doutorado em Bioinformática (UFMG), pela ajuda, amizade e apoio durante o curso e participação em congressos.

Ao meu pai, Henrique, pelo exemplo e incentivo e a minha mãe, Silvia, pela presença, cumplicidade e dedicação. Ao meu irmão Felipe, meu grande amigo. E a minha Lulu, por todo carinho. A Aurinha e a Dalva pela ajuda do dia-a-dia. E a Mega, pela fidelidade.

Ao Rodrigo, meu marido, por toda admiração, compreensão e respeito. Obrigada por todo amor! A nossa Brigittinha, nosso xodozinho.

Aos queridos colegas do Laboratório de Parasitologia Celular e Molecular, Marcela, Paty, Karlinha, Elisângela, Fernanda Barbosa, Kelly, Silvane, Maíra, Paola, Raul, Lu, Jerônimo, Rosana, Roney, Laila, Betânia, e todos os presentes, pelos agradáveis

momentos que passamos no laboratório. A Silvia e a Kênia que tanto contribuíram na limpeza do laboratório e organização dos materiais, facilitando muito o nosso trabalho.

Ao pessoal da salinha e grandes amigos, Fernanda Ludolf, Luiza, Livia, Regina, Flávio, Anderson, Rômulo e Adhemar.

A Dra. Angela Volpini e Maureen Rodarte, pela ajuda nos processos burocráticos, boa vontade e amizade.

Ao Dr. Rodrigo Corrêa de Oliveira e ao pessoal do Laboratório de Imunologia, pela amizade e pela constante disponibilidade. Em especial a Clari e a Ana Pacheco.

Ao pessoal do Laboratório de Malacologia, em especial a Liana pela disponibilidade de material biológico.

Aos meus colegas do TIGR, Dr. Michael Waisberg, Dra. Daniella Bartholomeu, Dr. David Spiro, Dr. Brian Haas, e todos em geral, em especial a Tinu.

Aos meus colegas de pós-graduação, em especial ao Eduardo, Gabriel e Priscila que estiveram presentes durante a realização das disciplinas.

As minhas queridas amigas que estiveram comigo durante todo esse tempo proporcionando muitas alegrias, Júlia, Lorena, Ines, em especial a minha “sister”, Fernanda Almeida, pelos ótimos momentos durante nossa vida no “egg” em Dupont Circle, DC.

Ao meu amigo Chuong, pelo desprendimento e carinho com que me acolheu quando cheguei aos USA.

Às agências que financiaram o projeto, CNPq e Forgy-NIH e a CAPES, pelo fornecimento da bolsa de estudo.

ÍNDICE

	PÁGINA
LISTA DE TABELAS.....	iii
LISTA DE FIGURAS.....	iv
LISTA DE ABREVIATURAS.....	v
RESUMO.....	vi
ABSTRACT.....	vii
I – INTRODUÇÃO.....	01
1.1 – Aspectos gerais da esquistossomose.....	02
1.2 – RNAs não-codificadores (ncRNAs).....	09
1.3 – Interferência mediada por RNA.....	11
1.4 – microRNAs.....	15
II – JUSTIFICATIVA & OBJETIVOS.....	21
2.1 – Justificativa.....	22
2.2 – Objetivo geral.....	23
2.3 – Objetivos específicos.....	23
III – MATERIAIS & MÉTODOS.....	24
3.1 – Parasitos.....	25
3.2 – Amostra biológica e extração de RNA.....	26
3.3 – Clonagem de pequenos RNAs de <i>S. mansoni</i>	27
3.4 – Análise e identificação dos candidatos a miRNA.....	30
3.5 – Validação da expressão de miRNAs.....	31
3.6 – Identificação da estrutura secundária.....	33
IV – RESULTADOS & DISCUSSÕES.....	34
4.1 – Clonagem de pequenos RNAs de <i>S. mansoni</i>	35
4.2 – Análise da expressão dos miRNAs de <i>S. mansoni</i>	43

4.3 – Estrutura secundária dos miRNAs.....	48
V – CONCLUSÕES.....	55
VI – CONSIDERAÇÕES FINAIS.....	57
VII – ENDEREÇOS ELETRÔNICOS.....	60
VIII – REFERÊNCIAS BIBLIOGRÁFICAS.....	62
IX – ARTIGOS.....	75
1. Cloning and characterization of <i>Schistosoma mansoni</i> microRNAs (Submetido BMC Genomics).....	76
2. Identification of novel miRNA genes in the <i>Schistosoma mansoni</i> genome (Submetido BMC Bioinformatics)	96
3. Glucose transporters are critical for schistosome survival in the vertebrate host (Em preparação).....	121
X – APÊNDICE.....	123
Candidatos a miRNA de <i>S. mansoni</i>	124

LISTA DE TABELAS

		PÁGINA
Tabela I	Candidatos a miRNA de <i>S. mansoni</i> que apresentaram similaridade com miRNAs de diferentes espécies presentes no banco de dados, miRBase.....	42
Tabela II	miRNAs de <i>S. mansoni</i> capazes de adquirir estrutura secundária, e sua respectiva posição genômica.....	50

LISTA DE FIGURAS

		PÁGINA
Figura 1	Ciclo de vida do <i>Schistosoma mansoni</i>	05
Figura 2	Via de biogênese de miRNAs.....	17
Figura 3	Representação esquemática da metodologia utilizada para a realização de clonagem de pequenos RNAs de <i>S. mansoni</i>	29
Figura 4	Representação esquemática da metodologia utilizada para a predição secundária em forma de grampo.....	33
Figura 5	Resultado da clonagem dos pequenos RNAs.....	37
Figura 6	Resultado da concatenação dos pequenos RNAs no experimento de clonagem.....	38
Figura 7	Resultado das análises após a clonagem dos pequenos RNAs de <i>S. mansoni</i>	38
Figura 8	Análises da expressão de miRNAs pela técnica de <i>northern blot</i>	47
Figura 9	Estrutura secundária dos miRNAs que tiveram sua expressão confirmada pela técnica de <i>northern blot</i>	51
Figura 10	Estrutura secundária dos miRNAs cuja expressão não foi detectada pela técnica de <i>northern blot</i>	52

LISTA DE ABREVIATURAS

BLAST	Ferramenta de Alinhamento Local – <i>Basic Local Alignment Search Tool</i>
cDNA	DNA Codificante
cm	Centímetro
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
DALYS	Disability Adjusted Life Years
DNA	Ácido Desoxirribonucléico
dsRNA	RNA dupla fita – <i>Double strand RNA</i>
EST	Etiqueta de Seqüência Expressa – <i>Expressed Sequence Tag</i>
Fiocruz	Fundação Oswaldo Cruz
Kb	Quilobases
M	Molar
Mb	Megabases
mg	miligrama
min	minutos
mRNA	RNA mensageiro
miRNA	microRNA
ncRNA	RNA não-codificador
NIH	Instituto Nacional de Saúde – <i>National Institute of Health</i>
pb	Pares de base
PCR	Reação em cadeia da polimerase – <i>Polymerase Chain Reaction</i>
pH	Potencial Hidrogeniônico
pmol	Picomoles, 10^{-12} moles
rRNA	RNA ribossomal
RNA	Ácido Ribonucléico
RNAi	Interferência mediada por RNA
rpm	Rotações por minuto
tRNA	RNA transportador
TDR	Programa Especial para Pesquisa e Treinamento em Doenças Infecciosas
WHO	Organização Mundial de Saúde
WTSI	<i>Wellcome Trust Sanger Institute</i>

RESUMO

O *Schistosoma mansoni* é um importante parasito humano, que infecta milhões de pessoas em todo mundo. Assim como outros parasitos patogênicos como *Plasmodium* e trypanosomas, o seqüenciamento do genoma de *S. mansoni* deu-se início com a antecipação de que a era genômica possibilitasse uma maior compreensão da biologia do parasito e o desenvolvimento de novas vacinas, drogas e métodos diagnósticos. Os avanços na biologia molecular possibilitaram não apenas o aumento da quantidade de dados biológicos, como também o aumento dos conhecimentos através das análises desses dados. Uma descoberta foi o mecanismo de controle da expressão gênica mediado pelos microRNAs (miRNAs), uma nova classe de RNAs reguladores, recentemente descritos. Essa regulação ocorre devido a ligação por complementaridade do miRNA maduro com o mRNA do gene alvo, interrompendo a tradução. É possível, portanto, que os miRNAs apresentem papel fundamental durante o desenvolvimento e crescimento do parasito. Nesse contexto, nosso principal objetivo foi analisar a presença de miRNAs em *S. mansoni*. A identificação foi realizada através da clonagem e seqüenciamento de biblioteca de cDNA construída a partir de pequenos RNAs de verme adulto. 211 novos candidatos a microRNA foram identificados de 582 seqüências de cDNAs clonadas. Os níveis de expressão de 30 candidatos a miRNA foram analisados e 14 novos miRNAs foram validados pela técnica de *northern blot*. Nossos resultados disponibilizam informações para futuros estudos que auxiliarão na elucidação de mecanismos de regulação gênica do parasito e um melhor entendimento do seu funcionamento celular.

ABSTRACT

Schistosoma mansoni is an important human parasite, infecting millions of people worldwide. As with other pathogenic parasites such as *Plasmodium* and trypanosomes, the genome sequencing of *S. mansoni* was initiated with the anticipation that such research may provide new resources to better understand its biology and to control the diseases it causes. One of the recently described mechanisms that regulate the expression of the genome is a family of non-coding RNAs, microRNAs, with approximately 21-25 nucleotides that down-regulate post-transcriptional gene expression. MicroRNAs are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to downregulate gene expression in a variety of manners, including translational repression and mRNA cleavage. MicroRNAs may have essential roles in the development and growth of the parasite. Based on that, our main objective was to identify miRNAs in *S. mansoni* by sequencing of small-RNA cDNA libraries from adult worms. 211 potential miRNAs candidates were identified from 582 cloned sequences. The expression of 14 of the 30 miRNAs was detected by northern blot. Our results provide information for future studies to elucidate the mechanism of gene regulation and the cooperative nature of the miRNA expression with their possible function in the regulation of various cellular processes.

I - INTRODUÇÃO

1.1 - Aspectos gerais da esquistossomose

Os parasitos platemintos do gênero *Schistosoma* são causadores da esquistossomose, uma doença crônica e debilitante e, em alguns casos, fatal. A esquistossomose afeta principalmente indivíduos em áreas rurais, sendo endêmica em países tropicais e subtropicais (WHO, 2008). A esquistossomose foi introduzida no Brasil, no período colonial, com a vinda de escravos africanos (Files, 1951). Apesar dos esforços para controlar essa endemia, ela permanece ainda como uma grande causa de morbidade, afetando aproximadamente 200 milhões de pessoas em todo o mundo, sendo 85% dos casos na África (King *et al.*, 2005; King & Dangerfield-Cha, 2008). Atualmente, a esquistossomose é endêmica em 76 países (Engels *et al.*, 2002), atingindo principalmente crianças e adolescentes com faixa etária entre 10 e 19 anos (TDR, 2009; <http://apps.who.int/tdr/svc/diseases/schistosomiasis>). De acordo com a Organização Mundial de Saúde, o crescimento da doença, calculado pelo DALYS (*Disability Adjusted Life Years*), é de 1.760.000, maior que doenças como tripanossomiase africana (1.598.000), dengue (653.000), chagas (649.000) e lepra (177.000). No Brasil, estima-se que existam até 8,3 milhões de infectados e 70 milhões de indivíduos expostos ao risco de contrair a infecção (Katz & Peixoto, 2000), atingindo quase todos os estados brasileiros, principalmente nas regiões Nordeste, Sudeste e Centro-Oeste (Oliveira *et al.*, 2004).

Várias espécies do gênero *Schistosoma* são importantes parasitos humanos: *S. mansoni*, *S. japonicum*, *S. haematobium*, *S. intercalatum*, *S. mekongi*, *S. mattheei* e *S. malayensis*, sendo as três primeiras as de maior relevância médica. Em geral, as espécies que infectam o homem são facilmente identificadas através do tamanho e morfologia do ovo, a origem geográfica do isolado e a especificidade pelo hospedeiro intermediário (Rollinson *et al.*, 1986; McManus & Hope, 1993). Os ovos de *S. mansoni* possuem uma

espinha lateral, os das espécies *S. haematobium* e *S. intercalatum* possuem espinha terminal, enquanto a espécie *S. japonicum* possuem ovos sem espinha.

Das várias espécies conhecidas de *Schistosoma*, sabe-se que *S. mansoni* é a espécie com maior distribuição global e a única espécie causadora da esquistossomose no Brasil (Bergquist, 2002). A infecção causada por esta espécie é denominada esquistossomose mansônica ou intestinal, pela localização dos parasitos nas vênulas da parede do intestino grosso, sigmóide e reto, com sintomas predominantemente intestinais. Nos casos mais graves, há envolvimento hepatoesplênico e hipertensão no sistema porta (Rey, 1991). Os principais sintomas da doença são febre, dor de cabeça, apatia por hepatoesplenomegalia, e em alguns casos morte (Van der Werf *et al.*, 2003).

Os parasitos do gênero *Schistosoma* pertencem ao Filo Platyelmintho e Classe Trematódea. São digenéticos, apresentam dimorfismo sexual na fase adulta e possuem o corpo achatado dorso-ventralmente. A fêmea mede cerca de 1,5 cm e possui o tegumento liso. O macho mede cerca de 1cm, tem o tegumento coberto por tubérculos e espinhas e um canal ginecóforo para albergar a fêmea e fecundá-la. A fêmea não é capaz de completar a sua maturação sem o acasalamento com o parasito macho. O macho possui um mecanismo ainda desconhecido que é capaz de regular a expressão de genes na fêmea (LoVerde *et al.*, 2009). Foi visto também que a fêmea necessita do macho não só no processo de fertilização, mas também na estimulação de fatores imprescindíveis para o seu crescimento e desenvolvimento (Grevelding *et al.*, 1997; Kunz, 2001).

A transmissão da doença ocorre pelo contato do homem com águas onde existam moluscos infectados. O ciclo biológico de *S. mansoni* apresenta uma alternância de gerações entre o hospedeiro intermediário, moluscos do gênero *Biomphalaria spp.*, e os hospedeiros definitivos vertebrados, dentre eles o homem (Figura 1). O ciclo de vida de *S. mansoni* inicia-se quando as fezes de indivíduos contaminados, contendo ovos do parasito,

entram em contato com a água doce. Os ovos, em contato com a água, eclodem, liberando miracídios, que são a forma infectante do hospedeiro invertebrado. Os miracídios infectam novos caramujos e cada miracídio se transforma em esporocisto I. Cada esporocisto I, por poliembrionia, origina 150 a 200 esporocistos II, que migram para as glândulas digestivas e ovoteste do caramujo, originando as cercárias que serão liberadas na água. A cercária, forma infectante para o hospedeiro vertebrado, infecta o homem por penetração ativa na pele. Ao penetrar na pele, as cercárias perdem a cauda, transformando-se em esquistossômulos. Os esquistossômulos migram para os pulmões cerca de 7 dias após a penetração e, posteriormente, completam o seu desenvolvimento no sistema porta intra-hepático (Rollinson & Southgate, 1987). Após a maturação, aproximadamente 45 dias após a infecção, os vermes adultos se alojam no plexo mesentérico e vivem por vários anos, podendo viver até 20 anos, no hospedeiro definitivo (Coelho, 1970). O ciclo se completa com a postura de ovos pela fêmea, aproximadamente 300 ovos por dia (Pellegrino & Coelho, 1978; Valadares *et al.*, 1981). Grande parte dos ovos é eliminada junto às fezes. Porém, alguns ficam retidos na mucosa intestinal e nos capilares do sistema porta do hospedeiro, onde desencadeiam uma reação inflamatória granulomatosa. A reação granulomatosa que se forma ao redor dos ovos é a principal causa da patogenia da esquistossomose. A reação granulomatosa resulta em fibrose do tecido (Gryseels *et al.*, 2006). A fibrose dos órgãos e a obstrução do plexo venoso podem levar à hipertensão portal, hepatomegalia, esplenomegalia, ascite (aumento do volume abdominal) e formação de varizes esofágicas (Gryseels *et al.*, 2006).

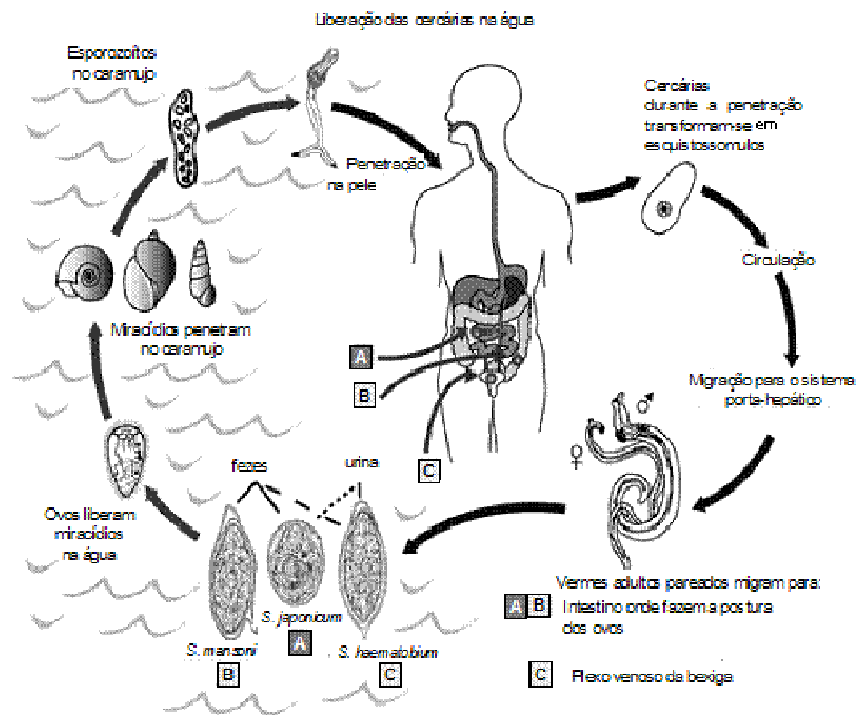


Figura 1 – Ciclo de vida de *Schistosoma*. O ciclo biológico de *Schistosoma mansoni* apresenta uma alternância de gerações entre o hospedeiro intermediário, moluscos do gênero *Biomphalaria spp.*, e os hospedeiros definitivos vertebrados, dentre eles o homem (WHO, 2008).

O controle da esquistossomose vem sendo realizado através de processos químicos e/ou biológicos de eliminação do hospedeiro intermediário, através do saneamento e educação para evitar o contato das pessoas com águas contaminadas e, principalmente, através do tratamento com drogas (<http://apps.who.int/tdr/svc/diseases/schistosomiasis>). Oxaminiquine e Praziquantel são as drogas mais utilizadas no tratamento da esquistossomose mansoni. O Oxaminiquine aumenta a mobilidade do parasito (Hillman & Senft, 1975) e inibe a síntese de ácido nucléico (Pica-Mattoccia *et al.*, 1989). A droga é mais efetiva contra parasitos machos do que em parasitos fêmeas. Contudo, apesar de ter apresentado 96% de cura parasitológica, quando realizada biópsia retal o tratamento apresentou apenas 38,3% de cura (Cunha, 1982). Atualmente, o medicamento de escolha para tratar a esquistossomose é o Praziquantel. É a única droga usada em larga escala para o tratamento da doença, o que tem contribuído para a diminuição da morbidade em áreas endêmicas (Chitsulo *et al.*, 2000; Kheir *et al.*, 2000; Ferrari *et al.*, 2003). O Praziquantel afeta principalmente parasitos fêmeas, causando uma alteração no seu tegumento (Redman *et al.*, 1996) e uma redução no nível de concentração de glutathione (Ribeiro *et al.*, 1998). Um dos possíveis alvos do Praziquantel é um canal de cálcio existente nas células da superfície do parasito, gerando um descontrole no fluxo de íons para dentro e para fora delas (Cioli & Pica-Mattoccia, 2003), levando à morte do parasito.

Apesar da disponibilidade da droga de ação rápida, administrada em dose única por via oral, a situação da esquistossomose ainda se apresenta bastante grave (Katz & Peixoto, 2000; Van der Werf *et al.*, 2003). A prevalência da doença permanece inalterada em muitas regiões endêmicas, devido, na maioria dos casos, aos altos níveis de reinfecção de indivíduos já tratados, seguido do possível aparecimento de populações de parasitos naturalmente resistentes ao tratamento (Doenhoff & Pica-Matoccia, 2006).

Na tentativa de se encontrar possíveis candidatos à vacina, genes foram estudados como a glutathione-S-transferase (GST) (Rao *et al.*, 2003), paramiosina (Al-Sherbiny *et al.*, 2003), IrV-5 (miosina) (Ribeiro de Jesus *et al.*, 2000), a triose fosfato isomerase (TPI) (Reynolds *et al.*, 1994), o antígeno de membrana 23 kDa (Sm23/MAP3) (Da'dara *et al.*, 2002), o antígeno de membrana 14 kDa (Sm14) (Fonseca *et al.*, 2005) e a enzima catepsina B (Noya *et al.*, 2001), entre outros. Não obstante, apesar de grandes avanços, nenhum resultado efetivo foi ainda alcançado. Assim, desde 1993, a Organização Mundial de Saúde (WHO) induziu o estudo da genômica como uma nova abordagem para o desenvolvimento de novas ferramentas de controle do parasito. O estudo da genômica do parasito gera expectativa para um melhor entendimento da biologia, do metabolismo destes organismos e também na busca de novos candidatos para diagnóstico, produção de vacinas e principalmente, novos alvos de drogas (Degrave *et al.*, 2001).

Em relação ao genoma, o *S. mansoni* é um organismo diplóide que possui oito pares de cromossomos (Short & Menzel 1960, Short *et al.*, 1979), sendo 7 pares autossômicos e 1 par sexual. Telômeros típicos foram identificados pela técnica de FISH nestes cromossomos (Hirai & LoVerde, 1996). O sexo heterogamético é a fêmea (ZW), enquanto o macho é homogamético (ZZ) (Short & Grossman 1981). O tamanho do genoma haplóide é estimado em aproximadamente 300 Mb, constituído de um conteúdo A+T bastante elevado (~66%) (Simpson *et al.*, 1982; Marx *et al.*, 2000). A presença da metilação não foi observada no DNA do parasito (Fantappie *et al.*, 2001). Com base no tamanho do seu genoma e posição evolutiva, estima-se que *S. mansoni* possua cerca de 15 a 20 mil genes expressos (Simpson *et al.*, 1982). O genoma é constituído 4% a 8% de seqüências de DNA altamente repetitivas (>1000 cópias), 35% a 40% de seqüências de DNA de média repetitividade (~100 cópias) e aproximadamente 60% representam famílias de genes ou regiões de cópia única (Simpson *et al.*, 1982). Grandes avanços foram publicados

recentemente, como a construção do primeiro mapa de ligação de *S. mansoni* apresentando cobertura de 5.6 cM contendo 243 marcadores do tipo microssatélites (Criscione *et al.*, 2009). Além disso, Berriman e colaboradores (2009) publicaram o genoma de *S. mansoni*. Foram analisados 363 Mb do genoma nuclear, identificando 11.809 genes. Os genes apresentam uma média de 4.7 kb, com introns muito maiores (~1692 pb) que os exons (~217 pb). Logo após a comemoração do centenário de descoberta do *S. mansoni* no Brasil, em 1908, estas informações representam um grande avanço que permitirá um conhecimento mais detalhado do parasito e a busca por novos alvos para o desenvolvimento de drogas mais eficazes e vacinas contra a doença.

Atualmente a abordagem mais promissora para a identificação de novos alvos de drogas, vacinas e reagentes para diagnóstico, assim como para a compreensão da resistência a drogas, diversidade antigênica, infectibilidade e patologia, consiste em compreender e decifrar as informações nos genomas dos organismos. Atualmente está disponível uma grande quantidade de informações em relação ao *S. mansoni*, tais como: seqüências de cDNA, grandes fragmentos genômicos completos, genoma mitocondrial, predição proteômica, etc. Esses dados podem ser visualizados no GeneDB (<http://www.genedb.org/genedb/smanson>), produzido pelo WTSI, e no SchistoDB (<http://www.schistodb.org>), banco de dados genômico desenvolvido pelo nosso grupo para incorporar todos os dados disponíveis sobre o *S. mansoni*. O SchistoDB possui, além de uma variedade de ferramentas, uma série de resultados pré-computados que podem ser consultados através do sistema de buscas como: o algoritmo BLAST (Altschul *et al.*, 1990), análises de peptídeo-sinal e regiões transmembrana, *gene ontology* (<http://www.geneontology.org>), famílias de proteínas, informação estrutural de proteínas, metaboloma, etc.

1.2 – RNAs não-codificadores (ncRNAs)

Até pouco tempo atrás, os RNAs eram considerados somente intermediários entre o genoma e as proteínas, como rege o dogma central da biologia. Recentes descobertas envolvendo uma variedade de novos genes de RNAs não-codificadores (ncRNAs), seus papéis biológicos e seus mecanismos de ação têm mostrado que a diversidade e a importância dos ncRNAs foram subestimadas (Storz, 2002; Mattick, 2009). Atualmente sabe-se que os RNAs que não codificam proteínas desempenham importantes funções relacionadas à regulação da expressão de genes, alterando o fluxo da informação genética tradicional. Até os meados dos anos 80, as famílias de ncRNAs conhecidas eram apenas as de rRNA (ribossomal) e tRNA (transportador). Contudo, atualmente o número de famílias descritas no banco de dados Rfam (<http://www.sanger.ac.uk/Software/Rfam>; versão 9.1) já ultrapassa 1.300. Recentemente foi demonstrado que a maioria dos RNAs transcritos em células humanas são ncRNAs originados de introns ou regiões intergênicas (Saini *et al.*, 2007). ncRNAs estão associados à regulação de diversos processos celulares, tais como silenciamento de genes, edição alternativa, controle da arquitetura da cromatina (formação de heterocromatina), metilação de DNA e edição de RNA (conversão de adenosinas em inosinas), expandindo bastante o vasto repertório de papéis dos RNAs nas células (Mattick & Makunin, 2006; Navarro *et al.*, 2009; Ponting *et al.*, 2009).

Apesar de toda evidência, ainda existe uma dificuldade sobre o conceito formal do que são os ncRNAs. Sabe-se que existem características que são comuns a todos como o fato de que além de não serem codificadores de proteínas, ncRNAs possuem uma atividade biológica relevante, e que não atuam como intermediários entre DNA e proteínas (Ponting *et al.*, 2009). A dificuldade em classificar os ncRNAs deve-se à variedade de estruturas, funções e interações que exercem *in vivo*. Por exemplo, ncRNAs podem ou não passar por processos de poliadenilação, se localizar no núcleo, citoplasma ou ambos, ligar-se a outros

RNAs, a proteínas, ou serem ativos sozinhos (Costa *et al.*, 2007). Além disso, muitas vezes são expressos de forma específica de acordo com o tecido, estágio, e/ou condição fisiológica (Ravasi *et al.*, 2006; Ginger *et al.*, 2006). A grande variedade de funções e estruturas dificulta um consenso para a classificação de ncRNAs na literatura. Além disso, a quantidade de ncRNAs identificados cresce rapidamente e estima-se que apenas uma fração tenha sido descrita até o momento. Acredita-se que estas moléculas possam também explicar melhor as diferenças na complexidade dos diferentes organismos (Amaral & Mattick, 2008). Surge assim um novo cenário revelando que o dogma central pode ser muito mais complexo, no qual ncRNAs desempenham relevantes papéis.

Em particular, algumas categorias de RNAs têm sido bastante estudadas, estando presentes em diversos organismos. Dentre a variedade de ncRNAs existentes, os mais estudados, que causaram uma revolução com a sua descoberta na década de 90, pertencem às classes dos pequenos RNAs do tipo RNA estruturais e reguladores, cujos representantes são os siRNAs (pequenos RNAs de interferência) e os microRNAs (miRNAs). Estes RNAs têm papel fundamental na regulação das vias de desenvolvimento de plantas e animais. Os miRNAs atuam de forma análoga aos siRNAs regulando negativamente RNAs mensageiros (mRNAs) alvo, porém diferenciam-se destes pela origem e pela natureza do gene alvo. Os siRNAs são derivados do próprio mRNA alvo transcrito a partir de um transgene, vírus, transposon ou gene endógeno. Os miRNAs são processados a partir de transcritos endógenos que não codificam proteínas e possuem como alvo RNAs mensageiros endógenos (Zeng *et al.*, 2003). Contudo, ambos, siRNAs e miRNAs, estão envolvidos no processo de silenciamento pós-transcricional e, ou interferência mediada por RNA (Kawasaki *et al.*, 2004).

1.3 – Interferência mediada por RNA

O mecanismo de interferência mediada por RNA tem sido extensivamente usado para a descrição da função de genes, ou para estabelecer as relações em diferentes vias regulatórias intracelulares. O mecanismo de silenciamento pós-transcricional mediado por RNA de fita dupla (dsRNA) é conhecido como RNA de interferência (RNAi). O RNAi é um fenômeno que ocorre naturalmente nas células e em diversos organismos eucarióticos (Hannon, 2002). Tal processo, descrito primeiramente em plantas, foi denominado silenciamento gênico pós-transcricional, ou PTGS (Jorgensen *et al.*, 1996). No entanto, a primeira descrição de silenciamento gênico em animais, assim como sua melhor compreensão, foi obtida no organismo modelo *Caenorhabditis elegans*, nematóide de vida livre (Fire *et al.*, 1998). Foi esta descoberta que premiou os pesquisadores norte-americanos Andrew Z. Fire e Craig C. Mello com o Nobel de Fisiologia ou Medicina de 2006.

O silenciamento de RNA é considerado uma modificação epigenética. Modificações epigenéticas na expressão gênica são características herdáveis que não podem ser explicadas por alterações na sequência de DNA, e que podem resultar na repressão (silenciamento gênico) ou ativação (ativação gênica) da expressão do gene alvo (Vaucheret *et al.*, 2006). Até o final da década de 1980, somente modificações na estrutura da cromatina ou de proteínas eram classificadas como epigenéticas (Lewin, 1998). Entretanto, durante a década de 1990, um grande número de fenômenos de silenciamento gênico que ocorriam a nível transcricional e pós-transcricional foram descritos em plantas, fungos, animais e protozoários, introduzindo o conceito de silenciamento de RNA (Baulcombe, 2000; Matzke *et al.*, 2001).

Logo após os primeiros estudos com plantas transgênicas resistentes a vírus (Lindbo & Dougherty, 1992), acreditava-se que o mecanismo de RNAi era apenas utilizado como defesa contra efeitos da movimentação de transposons, ou ainda, contra infecção por

vírus (Lindbo *et al.*, 1993; Waterhouse *et al.*, 2001). Entretanto, atualmente sabe-se que este mecanismo participa de forma integral na regulação gênica, principalmente no controle de genes envolvidos no desenvolvimento do organismo e na manutenção da integridade do genoma (Begemann *et al.*, 2008; Jinek & Doudna, 2009).

O componente unificador dos diferentes processos de silenciamento de RNA já estudado em diversos organismos é o RNA de fita dupla (dsRNA) (Fire *et al.* 1998). A presença desse tipo de molécula pode induzir a degradação de mRNA alvo. Participam deste processo a enzima Dicer, que pertence à família das ribonucleases do tipo III (RNase III), e as proteínas argonautas, que são as principais componentes do complexo de silenciamento RISC (*RNA Induced Silencing Complex*). Apesar de existirem algumas diferenças entre espécies nos mecanismos de silenciamento gênico por siRNA e miRNAs, tais processos são muito semelhantes (Zeng *et al.*, 2003). As semelhanças foram demonstradas com a clonagem e caracterização de diversos genes que codificam componentes do sistema em *Arabidopsis thaliana*, *C. elegans*, *Drosophila melanogaster*, *Neurospora crassa*, *Schizosaccharomyces pombe*, *Tetrahymena thermophila*, camundongos e humanos (Denli & Hannon, 2003; Cerutti & Casas-Mollano, 2006; Shabalina & Koonin, 2008). Moléculas de dsRNA são processadas pela enzima Dicer, e em seguida os pequenos RNAs gerados são englobados pelo complexo RISC, no qual devido à presença de uma helicase, sua dupla-fita é aberta e a fita antisenso do duplex guia o complexo até o mRNA alvo. Ocorre então a inibição traducional ou a degradação do mRNA, de acordo com o tipo de molécula e especificidade do pareamento (Shabalina & Koonin, 2008).

Em parasitos, como *Trypanosoma brucei*, RNAi tem sido bastante estudado desde que a maquinaria foi descoberta nesse organismo (Ngo *et al.*, 1998). *T. brucei* foi usado para elucidar o papel de proteínas envolvidas na maquinaria de RNAi como Argonauta 1

(Ago 1) e Dicer (Shi *et al.*, 2004, Shi *et al.*, 2006, Shi *et al.*, 2007). Além disso, foi construída uma molécula de dsRNA com formato de grampo, no qual o alvo era o gene da α -tubulina sob o controle do promotor induzível pela tetraciclina. Este sistema foi capaz de inibir a expressão do gene da α -tubulina em *T. brucei* (Shi *et al.*, 2000). Curiosamente, *T. cruzi*, *Leishmania donovani* e *Leishmania major* perderam a maquinaria de RNAi, apesar de pertencerem à mesma família de *T. brucei* (Zhang & Matlashewski, 2000; Robinson & Beverley, 2003; DaRocha *et al.*, 2004). Ao se pesquisar o banco de dados genômico destas espécies foi verificado uma deficiência dos genes ortólogos da enzima Ago1, assim como do domínio PAZ (normalmente presente em Ago 1 e Dicer), confirmando a possível ausência do mecanismo de RNAi uma vez que esta enzima possui papel fundamental na supressão da expressão de genes alvo (Ullu *et al.*, 2004). Em Apicomplexas, como *Plasmodium*, a função da via de RNAi ainda é um ponto controverso. Métodos de eletroporação foram usados para investigar o papel de dsRNA na inibição de diidroorotato desidrogenase (DHODH), enzima relacionada à biosíntese de pirimidina. O tratamento com dsRNA levou uma redução dos níveis de transcritos de DHODH, o que provocou a inibição do crescimento do parasito (McRobert & McConkey, 2002). Outros estudos mostraram também a presença de RNAi na inibição do gene da cisteína protease em *P. falciparum*. Além disso, foi demonstrado que quando introduzido na célula, o dsRNA é processado em pequenos fragmentos de 25 pb, gerando fortes evidências da presença da maquinaria de RNAi em *Plasmodium* (Malhotra *et al.*, 2002). Contudo, Ullu e colaboradores (2004) não acreditam que estas evidências sejam suficientes para demonstrar a presença de RNAi em *Plasmodium*. Eles questionam o fato de que McRobert e McConkey (2002) não realizaram experimentos de *northern blot* para comprovar os resultados. Eles também questionam a hipótese de que o siRNA encontrado em *Plasmodium* (Malhotra *et al.*, 2002) talvez seja fragmento de dsRNA do gene da cisteína

protease, resultante da contaminação de células do hospedeiro (Ullu *et al.*, 2004). Esta hipótese também foi apoiada por outros autores (Rathjen *et al.*, 2006), que não conseguiram confirmar a existência de siRNA em *Plasmodium*. Buscas por genes de RNAi em bancos de dados de *Toxoplasma gondii* sugerem que a via de RNAi possa também existir nesse parasito. Foi demonstrada uma redução nos níveis do transcrito da enzima uracil fosforibosiltransferase na presença de dsRNA, resultado também comprovado por *northern blot* (Al-Anouti & Ananvoranich, 2002). Contudo, não existem evidências de siRNA induzindo RNAi em *T. gondii* tratados, o que gera algumas incertezas sobre a existência de RNAi no parasito (Ullu *et al.*, 2004). Alguns pesquisadores sugerem que oligonucleotídeos anti-senso podem reduzir a expressão de mRNAs alvo mesmo na ausência da maquinaria de RNAi, como demonstrado em *Plasmodium* (Vidal *et al.*, 2005; Crooke *et al.*, 2006), sugerindo que o mesmo possa ocorrer em *Toxoplasma*.

Em parasitos helmintos, o efeito do RNAi, até o presente momento, foi pouco explorado, apresentando a maioria dos trabalhos em nematódeos e poucos em trematódeos (Geldhof *et al.*, 2007). Este fato é surpreendente uma vez que foi em *C. elegans* a primeira evidência de que o processo de RNAi ocorre através da injeção de dsRNA. Em *Schistosoma*, uma das primeiras evidências de RNAi teve como gene alvo o SGTP1 (Glicose-1 Fosfato desidrogenase). O resultado foi analisado pela técnica de PCR Quantitativa em Tempo Real (qPCR) e mostrou uma redução de 70-80% a nível transcricional do gene responsável pelo transporte de glicose no parasito (Boyle *et al.*, 2003). Outros estudos utilizaram a mesma técnica tendo como o gene alvo a Catepsina B, enzima responsável por degradar a hemoglobina (Hb) obtida do sangue do hospedeiro vertebrado que constitui a principal fonte de aminoácidos essenciais para o crescimento e maturação sexual de esquistossomos (Brindley *et al.*, 1997). Os parasitos foram tratados com dsRNA em cultura e foi sugerido um método alternativo para a inclusão do dsRNA, a

eletroporação. Foi confirmada uma redução nos níveis de transcritos do gene da Catepsina B na presença de dsRNA em esquistossômulos, e também foi demonstrado que a Catepsina B é essencial para o desenvolvimento do parasito (Correnti *et al.*, 2005). O método também foi eficiente na redução dos transcritos de receptores “scavenger” (um importante regulador na absorção do colesterol do fígado), resultando em 60-70% de silenciamento e alteração fenotípica no desenvolvimento do parasito (Dinguirard & Yoshino, 2006). Recentemente, a estrutura e expressão do gene Dicer foi caracterizada em *S. mansoni* (Krautz-Peterson & Skelly, 2008). Além disso, vários estudos mostraram que *Schistosoma* possui a maquinaria de silenciamento de RNA e que a adição de RNA exógeno de fita dupla (dsRNA) pode suprimir a expressão de genes alvo (Cheng *et al.*, 2005; Pereira *et al.*, 2008; Zhao *et al.*, 2008; Rinaldi *et al.*, 2009). Apesar dos diferentes estudos utilizando a técnica de RNAi no genoma de diferentes organismos, em *S. mansoni* ainda sabe-se muito pouco sobre o seu papel fisiológico na regulação de genes alvos. Em especial a nova classe de RNAs reguladores, miRNAs, tem despertado bastante interesse nos últimos anos e cuja maquinaria se faz presente no genoma do parasito (Gomes *et al.*, 2009).

1.4 – microRNAs

MicroRNAs (miRNAs) constituem uma classe de pequenos RNAs endógenos de aproximadamente 22 nucleotídeos, que atuam como silenciadores pós-transcricionais, inibindo a tradução de RNAs mensageiros alvo (Ambros, 2001; Lee & Ambros, 2001). Descobertos há pouco mais de uma década em *Caenorhabditis elegans* e associados à regulação do desenvolvimento larval (Lee *et al.*, 1993), os miRNAs são hoje reconhecidos como reguladores fundamentais da expressão gênica em plantas e animais. Eles encontram-se distribuídos no genoma em regiões intergênicas, intrônicas ou exônicas (Lee & Ambros, 2001; Pasquinelli & Ruvkun, 2002).

A biogênese do miRNA, esquematizada na Figura 2, inicia-se com a transcrição do gene pela RNA polimerase II, gerando um longo transcrito de miRNA primário (pri-miRNA) contendo cap 5' e cauda poli (A) a 3' (Lee *et al.*, 2004). O pri-miRNA apresenta uma estrutura característica dos miRNAs em forma de grampo em seu arcabouço que é clivada ainda no núcleo pela RNase III, Drosha, e seu cofator DGCR8, gerando uma molécula precursora denominada pré-miRNA, com cerca de 70-200 nucleotídeos (Lee *et al.*, 2003). Em seguida, o pré-miRNA é transportado rapidamente ao citoplasma pela exportina-5 (Exp5), proteína de exportação nuclear que utiliza Ran-GTP como co-fator (Lund *et al.*, 2004). No citoplasma, o pré-miRNA é processado pela RNase III, Dicer, gerando um miRNA fita dupla de aproximadamente 22 nucleotídeos (Bernstein *et al.*, 2001). Apenas uma das fitas do duplex de miRNA, correspondente à seqüência do miRNA maduro, é incorporada ao complexo multimérico denominado RISC (complexo de silenciamento induzido por RNA), que inclui as proteínas Argonautas como principais componentes, e que será direcionado para controlar a expressão pós-transcricional de genes alvo (Schwarz *et al.*, 2003; Gregory *et al.*, 2006). Os miRNAs exercem seus efeitos regulatórios ligando-se à região 3' não traduzida do mRNA alvo. Este mecanismo de atuação permite a redução dos níveis protéicos de seus genes alvo, raramente afetando o nível de expressão transcricional (Kim, 2005). Embora a presença de miRNAs em plantas e animais sugira que esta classe de RNAs não codificadores esteja envolvida na regulação da expressão gênica desde pelo menos o último ancestral comum destas linhagens por meio de um mecanismo conservado, algumas particularidades importantes existem na biogênese e no modo de ação entre os miRNAs de plantas e animais. Em plantas, a estrutura do pri-miRNA é muito variável e tipicamente mais longa do que em animais (Reinhart *et al.*, 2002), além disso a complementaridade entre o miRNA e o mRNA alvo em plantas ocorre de forma perfeita, diferente em animais (Hutvagner & Zamore, 2002).

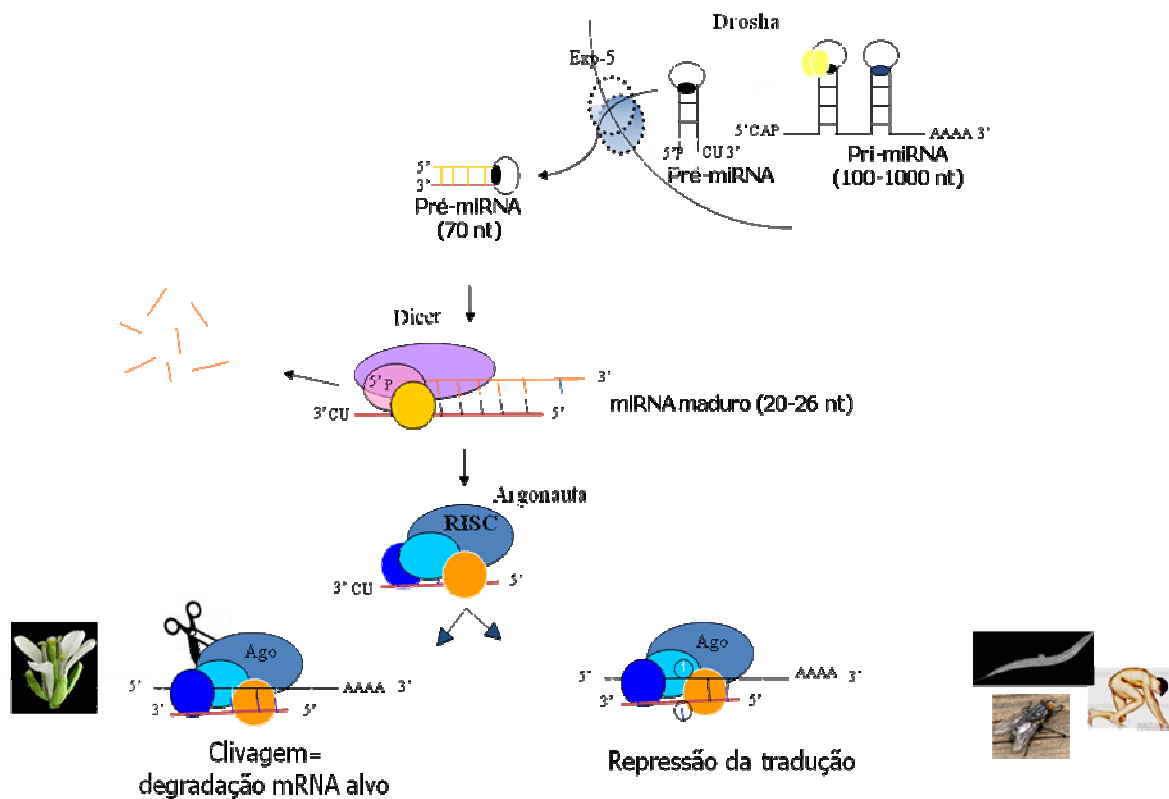


Figura 2 – Via de biogênese dos miRNAs. O gene de miRNA é transcrito pela RNA polimerase II. O transcrito primário (pri-miRNA) apresenta uma estrutura hairpin que é processada pela enzima RNase III, Drosha, formando o miRNA precursor (pre-miRNA) de 70 nucleotídeos. A proteína exportina-5 leva esse produto ao citoplasma para ser processado pela RNase III, Dicer, gerando um miRNA fita dupla de ~22 nucleotídeos. Uma das fitas do *duplex* de miRNA é degradada enquanto a outra permanece no complexo RISC para controlar a expressão pós-transcricional de genes-alvo.

A regulação pós-transcricional exercida pelos miRNAs na região 3' não traduzida depende do grau de complementaridade com o mRNA alvo, podendo ocorrer por inibição traducional ou degradação do mRNA. O pareamento de modo imperfeito com o mRNA acarreta a inibição traducional do alvo, sendo o mecanismo principal de atuação dos miRNAs em mamíferos (Brennecke *et al.*, 2005). Em plantas, a regulação dos miRNAs ocorre principalmente através de sua interação perfeita com o mRNA, levando-o à degradação (similar ao mecanismo de RNAi) (Brennecke *et al.*, 2005). No entanto, já foram descritos exemplos da ocorrência deste silenciamento gênico também em mamíferos (Valencia-Sanchez *et al.*, 2006).

Atualmente, miRNAs já foram descritos nos mais diversos organismos, como insetos, plantas, camundongos e humanos (Lagos-Quintana *et al.*, 2001; Lau *et al.*, 2001; Llave *et al.*, 2002; Reinhart *et al.*, 2002; Lagos-Quintana *et al.*, 2003; He & Hannon, 2004). A maioria dos miRNAs são conservados entre espécies relacionadas e aproximadamente 30% são altamente conservados, com ortólogos em vertebrados ou invertebrados, sugerindo uma conservação evolutiva com base em sua função biológica. Embora miRNAs estejam presentes em plantas e animais, ainda não foram descritos miRNAs ortólogos nos dois reinos (Bartel, 2004).

A identificação de miRNAs e seus genes alvo em organismos modelo são de fundamental importância para se compreender o mecanismo de regulação gênica e seu silenciamento (Stark *et al.*, 2003; Sunkar & Zhu, 2004; Sunkar *et al.*, 2005; Subramanian *et al.*, 2008). Estudos recentes utilizando análises computacionais revelaram um número significativo de novos miRNAs comparando o genoma de diferentes espécies (Berezikov *et al.*, 2005; Zhang *et al.*, 2006; Heikkinen *et al.*, 2008; Baev *et al.*, 2009). Até o momento, mais de 400 miRNAs diferentes foram identificados em humanos, representando de 1% a 4% dos genes expressos, e estima-se que este número supere 1000 miRNAs, constituindo

uma das maiores classes de reguladores gênicos (Berezikov *et al.*, 2005). Os mecanismos de interação utilizados pelo miRNAs na regulação da transcrição do RNA alvo ainda não estão completamente esclarecidos. Contudo, a função de alguns miRNAs foi determinada experimentalmente *in vivo*, incluindo os miRNAs *lin-4* e *let-7* em *C. elegans*, *bantam* e *mir-14* em *Drosophila* e *mir-23* em humanos, apresentando papel crucial na proliferação celular, diferenciação e no processo de apoptose (Reinhart *et al.*, 2000; Brennecke *et al.*, 2003; Kawasaki & Taira, 2003).

A maioria dos miRNAs apresenta seqüências conservadas (Lagos-Quintana *et al.*, 2003), e muitos possuem sua expressão modulada espacial ou temporalmente (Kasashima *et al.*, 2004; Liu *et al.*, 2004; Wienholds *et al.*, 2005) participando da regulação de uma enorme gama de processos fisiológicos (Ambros, 2004; Lau & Lai, 2005). Devido ao fato dos miRNAs possuírem seqüências pequenas e agirem sem a necessidade de pareamento completo, análises computacionais indicam que um único miRNA atue em diversos RNAs mensageiros, ou um mRNA pode ser alvo de vários miRNAs (Brennecke *et al.*, 2005), influenciando múltiplas vias de sinalização concomitantemente e apresentando enorme potencial regulatório (Farh *et al.*, 2005; Lim *et al.*, 2005).

Apesar da biologia dos miRNAs não estar ainda completamente esclarecida, essas moléculas já foram relacionadas a diversos processos biológicos. Além disso, a expressão anômala destes pequenos RNAs tem sido associada a diferentes patologias humanas (Esquela-Kerscher *et al.*, 2006). A expressão alterada dos componentes da maquinaria de biogênese dos miRNAs como Drosha, Dicer, e Argonautas foi associada a diferentes tumores humanos, destacando a importância desta via no funcionamento celular adequado (Fabbri, 2008; Zhang *et al.*, 2009). Nesse contexto, estudos envolvendo a identificação destas moléculas e posterior entendimento funcional é muito promissor, e pode auxiliar na busca de novos alvos terapêuticos, como para o câncer (Petrocca & Lieberman, 2009), ou

para doenças associadas ao desenvolvimento cerebral, como autismo, esquizofrenia, epilepsia, entre outras, uma vez que a expressão dos miRNAs é regulada dinamicamente durante a neurogênese e no desenvolvimento cerebral embrionário (Schratt *et al.*, 2006).

Até o presente momento, poucos trabalhos foram publicados envolvendo a identificação de miRNAs em parasitos e o seu possível papel em processos celulares, relacionados ao desenvolvimento, defesa contra patógenos e resposta ao estresse através da regulação da expressão gênica. Em *S. mansoni*, análises computacionais identificaram as principais proteínas envolvidas na maquinaria de miRNA através de buscas por homologia (Gomes *et al.*, 2009). Nas análises foram utilizadas seqüências de aminoácidos das proteínas conhecidas depositadas em banco de dados públicos contra o banco de dados de *S. mansoni*, identificando o total de 13 proteínas do parasito (Gomes *et al.*, 2009). Além disso, recentemente, Xue e colaboradores (2008) identificaram, pela primeira vez, miRNAs de *S. japonicum* através de clonagem e seqüenciamento de bibliotecas de pequenos RNAs. 106 clones foram selecionados aleatoriamente, gerando 227 seqüências candidatas a miRNA. Das análises dos dados foram identificados 5 miRNAs em *S. japonicum*. Hertel e colaboradores (2006) também fizeram a predição de 4 possíveis miRNAs utilizando métodos computacionais. Não obstante, até o presente momento nenhum trabalho havia sido realizado envolvendo a identificação de miRNAs em *S. mansoni* utilizando critérios de expressão e biogênese, apesar do seu potencial uso ser reconhecido. Portanto, o presente estudo visa a identificação de pequenos RNAs em *S. mansoni*.

II - JUSTIFICATIVA & OBJETIVOS

2.1 – Justificativa

A revelação de que moléculas de RNA possuem atividade catalítica, em meados dos anos 80, alertou para o fato de que estas moléculas possuem funções além de mensageiros de proteínas. As iniciativas de seqüenciamento genômico e o conseqüente aumento do número de seqüências disponíveis em bancos de dados públicos geraram a expectativa de que todo este avanço fosse útil para cura de doenças de origem genética. Uma das razões para a demora na obtenção de resultados concretos em termos de produtos que poderiam resultar em melhoria de saúde foi devido à falta de ferramentas que permitissem intervir no genoma. Desta forma, a descoberta do mecanismo de RNAi abriu uma nova perspectiva para novas possibilidades de estudos. Dentre as possibilidades está a presença especial da nova classe de RNA reguladores, os microRNAs (miRNAs). miRNAs ganharam relevância nos últimos anos devido as recentes e contínuas descobertas revelando a sua diversidade e importância. Apesar de não terem suas funções totalmente esclarecidas, a descoberta dos miRNAs atraiu a comunidade científica pelas evidências sugestivas de que estas moléculas apresentam papel fundamental em diversos processos biológicos. Nesse contexto, nos propomos a identificar potenciais candidatos a miRNA de *S. mansoni* através da clonagem e seqüenciamento de bibliotecas de pequenos RNAs, e verificar a expressão de alguns através da técnica de *northern blot*. Buscamos caracterizar miRNAs de *S. mansoni* seguindo os critérios de expressão e biogênese, o que será útil na compreensão da fisiologia molecular desta espécie. Acreditamos que uma estratégia completa que combine a avaliação simultânea da genética dos organismos e a influência de fatores ambientais será um importante meio para o entendimento dos aspectos biológicos da doença, do parasito e, conseqüentemente, para o desenvolvimento de métodos de controle.

2.2 – Objetivo geral

O objetivo deste trabalho é identificar genes de microRNA presentes no genoma do parasito *Schistosoma mansoni* através da clonagem e seqüenciamento de pequenos RNAs.

2.3 – Objetivos específicos

- 1 Construir bibliotecas contendo pequenos RNAs;
- 2 Identificar computacionalmente miRNAs nas bibliotecas seqüenciadas;
- 3 Caracterizar os miRNAs estruturalmente por meio de modelos preditos;
- 4 Mapear os miRNAs selecionados no genoma de *S. mansoni*;
- 5 Validar a expressão dos candidatos a miRNAs mais freqüentes utilizando a técnica de *northern blot* nas fases de verme adulto e esquistossômulo;
- 6 Verificar a expressão dos candidatos a miRNAs identificados em *Schistosoma japonicum* utilizando a técnica de *northern blot* nas fases de verme adulto e esquistossômulo;
- 7 Identificar miRNAs ortólogos aos de outros organismos.

III - MATERIAIS & MÉTODOS

3.1 – Parasitos

Os vermes adultos de *S. mansoni* foram obtidos de perfusão de camundongos infectados, realizada segundo Pellegrino & Siqueira (1956). Os vermes foram isolados do sistema porta e veias mesentéricas de camundongos infectados após 6-7 semanas da realização da infecção. Em seguida, os vermes foram manualmente separados e lavados em solução salina tamponada com fosfatos (PBS; 10 mM Na₂HPO₄, 1.7 mM KH₂PO₄, 137 mM NaCl, 2.7 mM KCl, pH 7.4). As amostras foram imediatamente usadas para extração de RNA total ou estocadas a -70°C para posterior utilização.

Parasitos na fase de esquistossômulo foram também obtidos através da transformação de cercárias (Ramalho-Pinto *et al.*, 1974). Cercárias foram ressuspensas em 6 ml de meio RPMI 1640 (GIBCO) e agitadas a velocidade máxima, com o auxílio do agitador Vortex (Reax-Top), por aproximadamente 1 minuto. Esse passo é necessário para que a cauda das cercárias se separe do corpo. Em capela de fluxo laminar, aproximadamente 1ml de cercárias em meio RPMI foi transferido para novo tubo. As cercárias foram incubadas em 5 ml de meio RPMI durante 5 minutos por 5 vezes, sendo dois tubos agrupados a cada lavagem. O sobrenadante foi retirado com auxílio de pipeta Pasteur e os parasitos foram ressuspensos em 20 ml de meio RPMI suplementado com 5% de soro fetal bovino (Gibco), 1% de gentamicina (Sigma) e 1% de glutamicina (Sigma). Para o cultivo, os parasitos foram distribuídos em placa de cultura de 6 poços e mantidos em estufa (Forma Scientific) a 37°C, com ambiente de 5% de CO₂ e 95% de umidade. Após uma semana de cultura, os poços foram agrupados, o sobrenadante foi descartado e os esquistossômulos foram armazenados a -70°C para posterior utilização.

3.2 – Amostra biológica e extração de RNA

As extrações de RNA provenientes dos estágios de vida do parasito verme adulto e esquistossômulo foram realizadas utilizando o reagente Trizol[®] (Invitrogen), conforme instruções do fabricante, apresentando pequenas modificações. Primeiramente, os tubos foram mergulhados em nitrogênio líquido por 10 minutos e homogeneizados em Trizol com auxílio de pistilos. Para cada 50-100 mg de tecido, 1 ml de reagente foi utilizado. 0,2 ml de clorofórmio foram adicionados às amostras, que foram agitadas vigorosamente por 15 segundos. Os tubos foram centrifugados (5415 R Eppendorf) a 13.000 rpm na temperatura de 4°C por 15 minutos, e a fase aquosa, contendo o RNA, foi transferida para um novo tubo. Foram adicionados 0,5 ml de isopropanol, seguido de incubação por 2 horas a -20°C. Novamente, os tubos foram centrifugados a 13.000 rpm na temperatura de 4°C por 15 minutos e o sobrenadante descartado. Ao sedimento de RNA foram adicionados 500µl de etanol 75%, seguido de centrifugação 13.000 rpm por 10 minutos a 4°C. O sobrenadante foi descartado e novamente a mesma etapa foi repetida. O excesso de etanol foi retirado e as amostras foram mantidas à temperatura ambiente por 10 minutos. O RNA foi eluído em 50µl de água livre de RNase previamente tratada com 1% DEPC (Invitrogen). A concentração do material foi determinada através do espectrômetro NanoDrop[®] ND-1000 (NanoDrop Technologies, Inc.) a 260nm, sendo o grau de pureza das amostras determinado pela relação da absorbância: $A_{260/280nm}$. A qualidade da extração do RNA foi analisada por eletroforese em gel de agarose desnaturante a 1,5% corado em brometo de etídeo (0,5µg/ml) (Sigma) e a imagem digitalizada pelo aparelho *Eagle Eye II* (Stratagene).

3.3 – Clonagem de pequenos RNAs de *S. mansoni*

A clonagem dos pequenos RNAs e análises foram realizadas com base em protocolos publicados anteriormente (Lau *et al.*, 2001; Lagos-Quintana *et al.*, 2001; Chappell *et al.*, 2005) (Figura 3). Aproximadamente 1 miligrama de RNA total de vermes adultos machos e fêmeas foram separados em gel de poliacrilamida desnaturante 15% adicionado de urea 8M (Sequel™ NE, American Bioanalytical). Os pequenos fragmentos de RNA contendo aproximadamente 17-26 nucleotídeos foram cortados do gel e eluídos em NaCl 0,3M *overnight* a 4°C, sob constante agitação. A solução foi precipitada adicionando-se etanol absoluto (2,5 vezes o volume total), e incubada a -20°C por 2 horas. Em seguida, o RNA foi desfosforilado utilizando enzima fosfatase alcalina (1U, APEX, Epicentre). Foi realizada extração utilizando Fenol:Clorofórmio:Álcool isoamílico (IAA) na proporção 25:24:1 (Invitrogen), e novamente a solução foi precipitada em etanol absoluto e ressuspensa em água livre de RNase (Invitrogen).

Adaptadores contendo oligonucleotídeos de RNA/DNA (Dharmacon Research, Boulder, CO) foram ligados às extremidades 5' e 3' dos pequenos RNAs. O adaptador 3' (5'-pUUUaaccgcatccttctcx-3'; maiúsculo - RNA; minúsculo - DNA; p - fosfato; x - deoxitimidina invertida) possui na extremidade 5' um monofosfato e na extremidade 3' uma deoxitimidina invertida com o objetivo de evitar sua auto-ligação quando ligado ao RNA desfosforilado. Desta forma, o produto foi primeiramente ligado ao adaptador 3' e, em seguida, separado novamente em gel de poliacrilamida desnaturante 15% adicionado de urea 8M (Sequel™ NE, American Bioanalytical). Os fragmentos de RNA contendo aproximadamente 40-60 nucleotídeos foram cortados do gel e eluídos em NaCl 0,3M por 16 horas a 4°C, sob constante agitação. A solução foi precipitada e, em seguida, a extremidade 5' foi fosforilada utilizando a enzima T4 polinucleotídeo Kinase (1U, NEB). O produto foi novamente extraído utilizando Fenol:Clorofórmio:Álcool isoamílico (IAA) na proporção

25:24:1 (Invitrogen), seguido de precipitação com etanol absoluto. O produto foi ressuspendido em água livre de RNase (Invitrogen). Em seguida, o adaptador 5' (5' tactaatacgaactcactAAA3'; maiúsculo - RNA; minúsculo - DNA) foi adicionado à amostra fosforilada e os fragmentos de RNA contendo aproximadamente 60-80 nucleotídeos foram cortados do gel e eluídos em NaCl 0,3M por 16 horas a 4°C, sob constante agitação. Foi realizada a reação de transcrição reversa, seguida da reação de PCR utilizando os iniciadores RT reverso (5' TTTT**CTGCAGA**AGGATGCGGTTAAA3', negrito representa o sítio para endonuclease *PstI*) e RT *forward* (5' AAAC**CATGGT**ACTAATACGACTCACTAAA3', negrito representa o sítio para endonuclease *NcoI*). O produto de PCR foi purificado através de extração utilizando Fenol:Clorofórmio (1:1), precipitado com etanol absoluto e digerido com as endonucleases *PstI* e *NcoI* (NEB). Foi realizada nova extração utilizando Fenol:Clorofórmio (1:1), seguida da precipitação das amostras. Os produtos digeridos contendo os sítios complementares as endonucleases foram ligados uns aos outros utilizando a enzima T4 DNA ligase (NEB) formando concatâmeros com tamanho de aproximadamente 300-600 pb. Foi realizado o preenchimento das extremidades finais pela ação da enzima polimerase Klenow (NEB). Os amplicons gerados foram ligados ao vetor pCR 2.1-TOPO (Invitrogen) e utilizados na transformação de células quimicamente competentes (TOP10; Invitrogen). Clones foram plaqueados em meio seletivo contendo X-Gal, e após crescimento por 16 horas as colônias brancas foram selecionadas e submetidas ao seqüenciamento, direto e reverso, no *Joint Technology Center* (JTC), laboratório responsável por prover ao TIGR serviços de mini-prep e seqüenciamento de DNA. Todos os procedimentos foram realizados de acordo com protocolo recomendado pelo fabricante. Mini-prep e reações de seqüenciamento foram executados de acordo com protocolos utilizados no JTC.

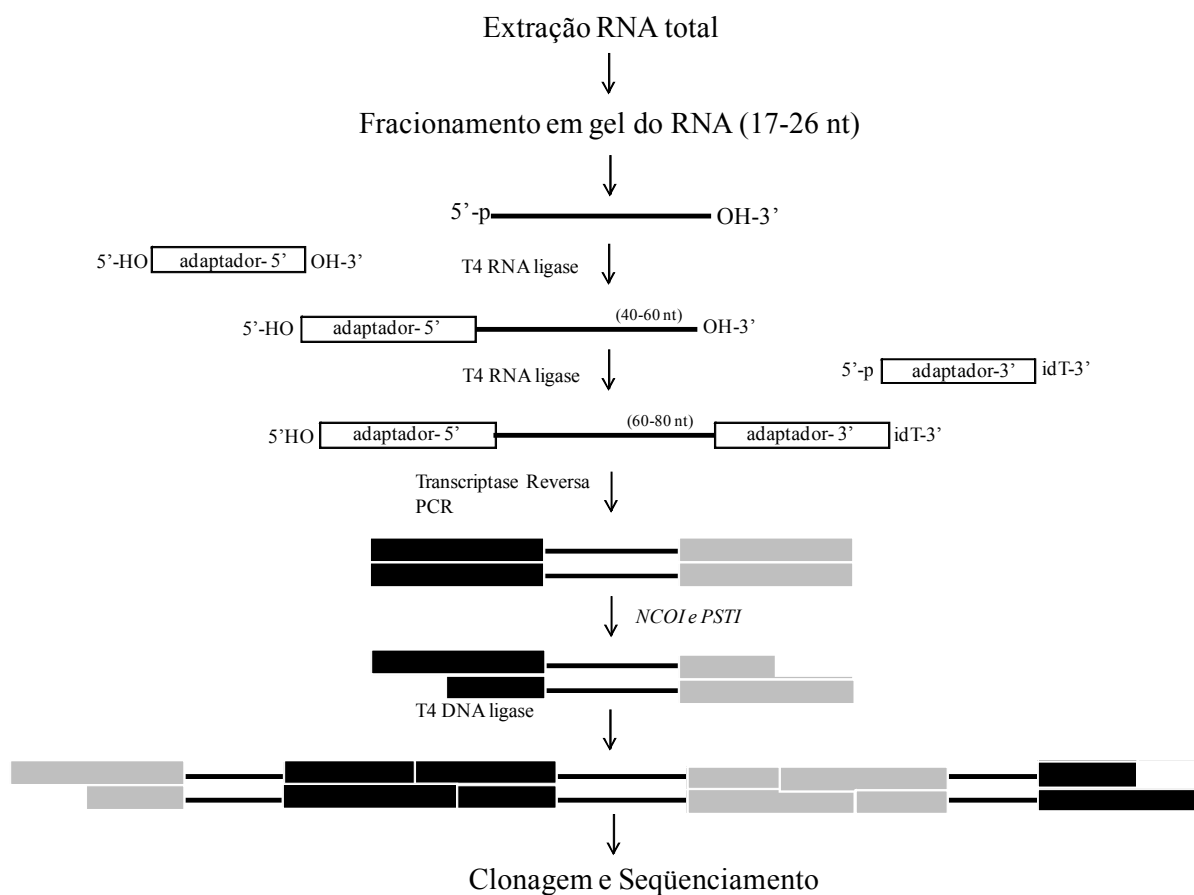


Figura 3 – Representação esquemática da metodologia utilizada para a realização da biblioteca de cDNA contendo os pequenos RNAs de *S. mansoni*.

3.4 - Análise e identificação dos candidatos a miRNA

O produto resultante do seqüenciamento foi pré-processado *in silico* através de *pipeline* semi-automático desenvolvido através de *scripts* baseados nas linguagens de programação awk, sed e grep, fazendo uso extensivo de programas do pacote EMBOSS (Rice *et al.*, 2000). Além de facilitar a análise, o uso de programas durante esta etapa assegurou que todos os procedimentos fossem realizados de forma idêntica para todo o conjunto de amostras.

A primeira tarefa do *pipeline* foi identificar os sítios de restrição presentes nos concatâmeros. Cada concatâmero continha fragmentos contendo um inserto candidato a miRNA e adaptadores nas extremidades 5' e 3'. Para a identificação dos sítios das endonucleases foi utilizado o programa pertencente ao pacote EMBOSS chamado EMBOSS-restrict (<http://bioweb2.pasteur.fr/docs/EMBOSS/restrict.html>). Após a identificação, foi usado um *script* na linguagem awk com a finalidade de separar os concatâmeros em unidades menores (fragmentos), contendo cada unidade aproximadamente 60 nt. Os fragmentos que apresentaram menos que 45 nt foram eliminados das análises seguintes, devido a grande chance de representarem contaminações da técnica.

Em seguida, vetores e adaptadores foram removidos com o auxílio do programa Cross_Match que utiliza a implementação do algoritmo de Smith-Waterman-Gotoh (Smith & Waterman, 1981; Gotoh, 1982), substituindo a base original por um x. As seqüências mascaradas foram alinhadas utilizando o programa ClustalW (Thompson *et al.*, 1994) com o objetivo de evitar seqüências redundantes. As seqüências distintas foram submetidas a análises utilizando o programa BLAST (www.ncbi.nlm.nih.gov/BLAST) contra o banco de dados contendo o genoma de *S. mansoni* disponível no TIGR (<http://www.tigr.org>; versão 3), e Rfam (<http://www.sanger.ac.uk/Software/Rfam>), eliminando possíveis contaminações de mRNAs, tRNAs, rRNAs, e outros pequenos RNAs. Foi também realizado BLASTn

contra o banco de dados de miRNAs conhecidos de diferentes organismos, miRBase (<http://microrna.sanger.ac.uk>, versão 13.0).

3.5 – Validação da expressão de miRNAs

Para confirmar a expressão e o tamanho dos candidatos a miRNA, a técnica de *northern blot* foi utilizada. Foram escolhidos para validação os candidatos encontrados o maior número de vezes nas bibliotecas. Foram sintetizadas 30 sondas específicas complementares a seqüência do miRNA. As amostras de RNA total das fases de verme adulto e esquistossômulo foram obtidas pelo método do tiocianato de guanidina-fenol-clorofórmio (Invitrogen) de tecidos congelados (-70°C) de acordo com as instruções do fabricante. Sessenta microgramas de RNA total foram separados em gel de poliacrilamida 15% desnaturante (8M urea) a 200 V por 5 horas e fotografado em foto-documentador modelo *Eagle Eye II* (Stratagene). Em seguida, o gel foi eletrotransferido para membrana Hybond-N+ (Amersham Biosciences, Buckinghamshire, UK), em 1x TBE através do Mini Trans-Blot Cell Aparatus (Bio-Rad), seguindo-se as recomendações do fabricante. A fixação do RNA ocorreu pela colocação das membranas em um transluminador (125 mJ, GS GENE LINKER, Bio-Rad). As sondas complementares foram marcadas utilizando o kit de marcação DIG Oligonucleotide 3'-End Labeling Kit, *2nd generation* (Roche). Oligonucleotídeos correspondendo a seqüências do miRNA foram também sintetizados e marcados com o mesmo kit, e usados como controle para garantir a especificidade e sensibilidade da hibridização. As membranas foram pré-hibridizadas com um excesso (20 ml/100cm² de membrana a hibridizar) de solução DIG Easy Hyb (Roche) na temperatura ideal de *melting* para cada sonda ($T_m = 4^\circ\text{C} (G/C) + 2^\circ\text{C} (T/A) - 10^\circ\text{C}$) durante 30 minutos. Em seguida, a sonda marcada com biotina (4.5 pmol/μl) foi adicionada à solução DIG Easy Hyb (Roche) pré-aquecida. A hibridação para cada sonda ocorreu na T_m ideal durante toda a noite. Em seguida, as membranas foram lavadas em condições de alta estringência, duas

vezes em 2X SSC contendo 0,1% SDS em temperatura ambiente por 10 minutos e 0,5X SSC contendo 0,1% SDS a 70°C duas vezes por 30 minutos. Após esta primeira etapa, as membranas foram incubadas em solução de lavagem, e solução bloqueadora por 30 minutos e 1 hora, respectivamente, a 23°C (DIG Wash and Block Buffer Set, Roche). Em seguida, as membranas foram incubadas em solução bloqueadora contendo anticorpo anti-digoxigenina 1:10.000 (anti-DIG-AP Fab fragment, Roche) por 30 minutos à TA, e foram lavadas por mais 4 vezes por 15 minutos a TA. Finalizando, as membranas foram incubadas por 5 minutos a 23°C em solução de detecção (DIG Wash and Block Buffer Set, Roche). O substrato quimioluminescente foi revelado pela fosfatase alcalina (CDP-star, Roche) e os sinais foram visualizados em filme raio-X (Hyperfilm – ECL) após 20 minutos.

3.6 – Identificação da estrutura secundária

Para a identificação da estrutura secundária foram desenvolvidos *scripts* na linguagem de programação Perl (*Practical Extracting and Reporting Language*). Primeiramente, os miRNAs foram mapeados em todas as posições possíveis no genoma do parasito *S. mansoni* depositado no banco de dados SchistoDB (Zerlotini *et al.*, 2009, versão 4). Foi utilizado o programa BLAST (Altschul *et al.*, 1990) para identificar as regiões com similaridade aos miRNAs e foi gerado um arquivo contendo as posições genômicas referente a cada candidato. Em seguida, o mesmo arquivo foi utilizado para gerar mais dois novos arquivos no formato Fasta contendo aproximadamente 500 seqüências para cada miRNA. Um arquivo contendo a seqüência do miRNA mais 40, 50, 60 e 70 nucleotídeos na extremidade 3' e o outro arquivo contendo os candidatos a miRNA e mais 40, 50, 60 e 70 nucleotídeos na extremidade 5'. Estes arquivos serviram de entrada para o programa RNAfold que integra o pacote ViennaRNA (Hofacker, 2003) para a predição das estruturas secundárias com formato de grampo. Cada imagem gerada pelo programa RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>) foi analisada manualmente e foi escolhida a

estrutura contendo a seqüência do miRNA maduro em um dos braços da estrutura e apresentando um índice de MFE baixo (*Minimum Free Energy*). A versão final da imagem foi gerada através do programa VARNA (<http://varna.lri.fr/index.html>).

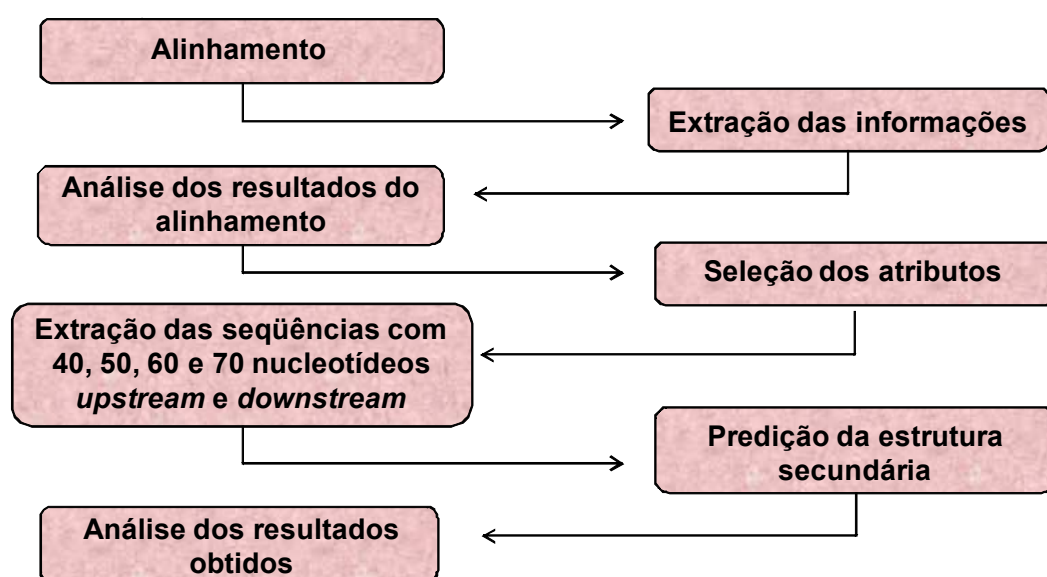


Figura 4 – Representação esquemática da metodologia utilizada para a predição da estrutura secundária em forma de grampo.

IV – RESULTADOS & DISCUSSÕES

4.1 – Clonagem de pequenos RNAs de *S. mansoni*

Na última década, uma atenção especial tem sido dada aos RNAs não-codificadores, visto que são considerados importantes elementos relacionados ao sistema de regulação da expressão gênica (Mattick, 2009). Os miRNAs compõem um pequeno grupo destes RNAs endógenos, formados por uma seqüência composta por 17 a 25 nucleotídeos, e exibem diversas funções biológicas. Em plantas e animais, sabe-se que o miRNA é o tipo mais abundante de pequenos RNAs (Ambros, 2003; Ambros, 2004). Baseado neste contexto, o principal objetivo do presente trabalho consistiu na tentativa de seqüenciar em massa bibliotecas geradas pelo fracionamento de gel de pequenos RNAs para identificação de miRNAs de *S. mansoni*. Devido à grande quantidade de material biológico necessário para realização da técnica, o protocolo foi primeiramente otimizado utilizando RNA de *Trypanosoma brucei*, uma vez que este parasito é facilmente cultivado em laboratório e não apresenta um ciclo de vida tão complexo como de *S. mansoni*. Além disso, esta metodologia já havia sido utilizada em *T. brucei* e tínhamos conhecimento do produto esperado para servir como controle da técnica (Djikeng *et al.*, 2004). Estudos anteriores tentaram identificar miRNAs em *T. brucei*, e não conseguiram (Djikeng *et al.*, 2001). Acredita-se que *T. brucei* possua apenas siRNAs, e não miRNAs. Além disso, existem evidências de que o organismo não possui a enzima Drosha, enzima nuclear importante, parte da maquinaria de miRNAs (Berriman *et al.*, 2005).

Uma vez que a técnica foi otimizada, a mesma metodologia foi aplicada utilizando-se RNA de *S. mansoni*. Métodos computacionais normalmente identificam miRNAs que são conservados entre espécies (Lim *et al.*, 2003; Grad *et al.*, 2003). Já a metodologia de clonagem também permite identificar aqueles miRNAs que não sejam tão conservados, e que sejam tecido-específico (Berezikov *et al.*, 2006a; Wang *et al.*, 2007; Beuvink *et al.*, 2007). Contudo, uma limitação comum na descoberta de miRNAs através de clonagem é a

dificuldade em identificar miRNAs pouco expressos, em um estágio específico ou em tipos de células raras. Neste contexto, seis bibliotecas independentes de pequenos RNAs de vermes adultos de *S. mansoni* foram geradas utilizando a técnica de fracionamento de gel (Chappell *et al.*, 2005). Em seguida, os fragmentos foram ligados aos adaptadores nas suas extremidades 5' e 3', clonados, e um total de 582 clones contendo o inserto específico foram selecionados aleatoriamente das bibliotecas e seqüenciados. Foram geradas aproximadamente 1.500 seqüências, das quais 1.200 representavam seqüências pequenas de cDNA com tamanho entre 17 e 26 nucleotídeos, sendo que a maioria apresentou 21 nt (Figura 5). Através deste resultado foi também analisado o sucesso da concatenação dos fragmentos indicando que cada clone continha aproximadamente 2 a 3 unidades de seqüências de pequenos RNAs (Figura 6). O restante destas seqüências apresentava um tamanho muito pequeno para genes de miRNA ou apenas representavam ligação entre dois adaptadores, e foram eliminadas das análises posteriores. Com o objetivo de identificar a origem genômica de cada seqüência gerada, análises utilizando o programa Blastn (<http://blast.ncbi.nlm.nih.gov/Blast>) revelaram que 97% destas seqüências apresentaram pelo menos 1 *hit* no genoma de *S. mansoni*, versão 3 disponível no TIGR (<http://www.tigr.org>). Os 3% restantes que não apresentaram nenhuma similaridade na busca foram descartados das análises seguintes. A falta de similaridade destas seqüências deve-se possivelmente às regiões de *gaps* ainda não concluídas no seqüenciamento do genoma do parasito, erros de seqüenciamento, entre outras possibilidades. Foram também eliminados das próximas análises os fragmentos dos abundantes RNAs não-codificadores (rRNAs, tRNA e pequenos RNAs nucleares), determinados por Blastn contra o banco de dados contendo uma coleção de alinhamentos múltiplos de várias famílias de ncRNA, Rfam (<http://www.sanger.ac.uk/Software/Rfam>; versão 8.1) e contra o banco de dados contendo seqüências de *S. mansoni* (<http://www.tigr.org>; versão 3). As 584 seqüências restantes foram

classificadas como possíveis miRNAs de *S. mansoni* (Figura 7), e foram utilizadas nas análises seguintes.

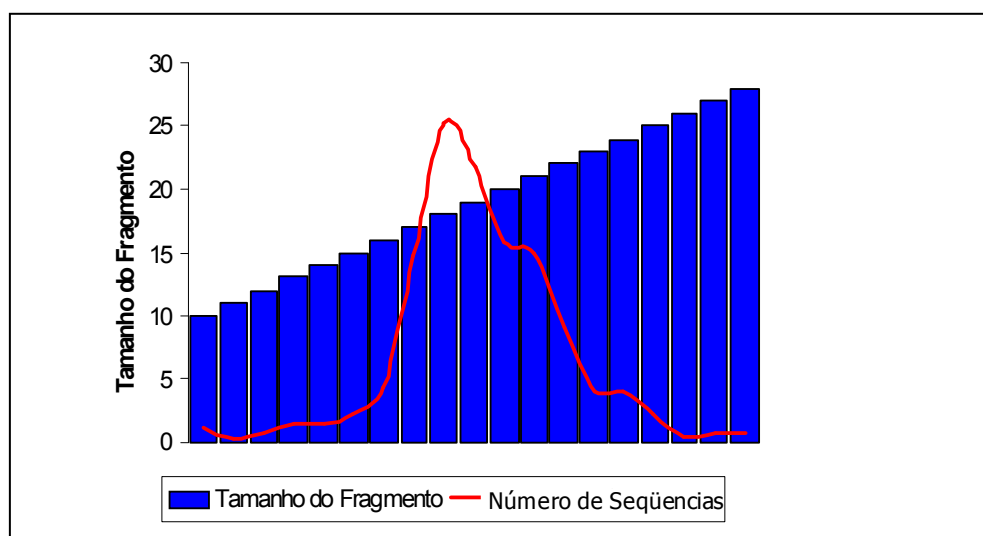


Figura 5 – Resultado da clonagem dos pequenos RNAs. Como esperado, o tamanho da maioria dos fragmentos clonados variou aproximadamente entre 17 e 26 nucleotídeos.

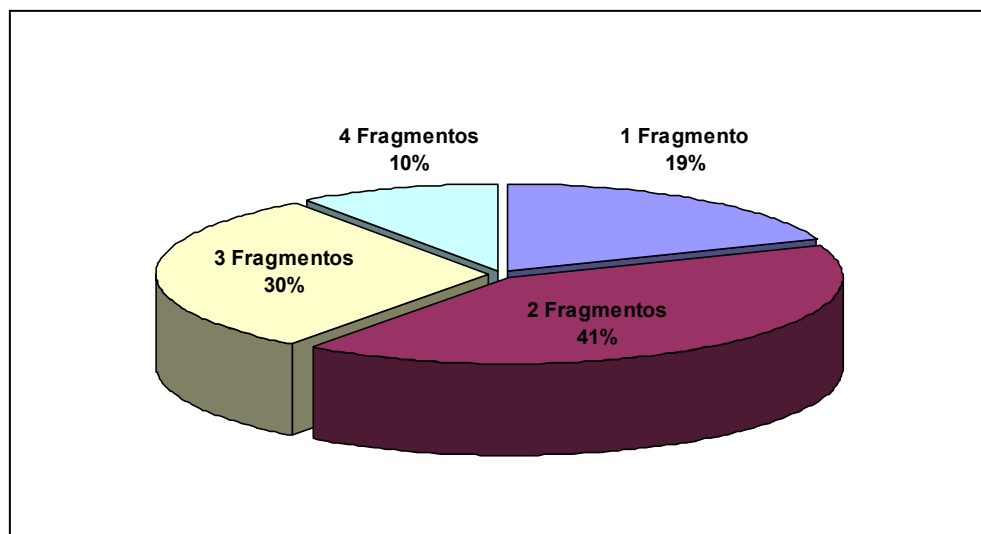


Figura 6– Resultado da concatenação dos pequenos RNAs no experimento de clonagem. A maioria dos clones analisados apresentou pelo menos 2 fragmentos (41%).

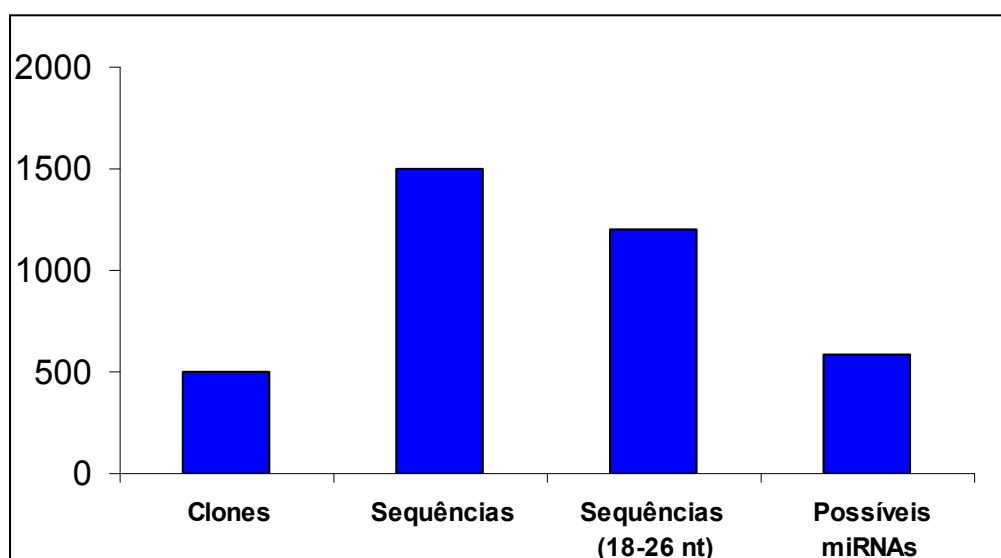


Figura 7 – Resultado das análises após a clonagem dos pequenos RNAs de *S. mansoni*.

Em seguida, com o objetivo de evitar seqüências redundantes, as 584 seqüências candidatas a miRNA foram agrupadas por identidade em 211 agrupamentos, representando potenciais candidatos a miRNA de *S. mansoni* (Apêndice 1). O número de seqüências por agrupamento variou de 1 a 32 seqüências, sendo 161 seqüências únicas representadas por somente um clone e 50 seqüências representadas na biblioteca por dois ou mais clones. Acredita-se que miRNAs representem aproximadamente 0.5-1.5% do total de genes de um determinado genoma (Carthew, 2006). Foram preditos ~12.000 genes no genoma de *S. mansoni* (Berriman *et al.*, 2009), dos quais aproximadamente 100-300 representem genes de miRNA. Desta forma, o número de possíveis miRNA identificados no presente trabalho comparado à predição do número esperado de genes de miRNAs no genoma de *S. mansoni* foi dentro do esperado e acredita-se que os miRNAs clonados de *S. mansoni* representem provavelmente os RNAs mais abundantes do parasito.

O número de novos miRNAs descobertos vem crescendo a cada ano, e espera-se que esse crescimento continue aumentando devido a grande quantidade de organismos que ainda não possuem esse tipo de estrutura identificada. Neste contexto, Griffiths-Jones e colaboradores (2006) estabeleceram algumas regras de nomenclatura de miRNAs para o registro no miRBase. O miRBase é um banco de dados mundial de armazenamento de seqüências de miRNAs conhecidos (<http://microrna.sanger.ac.uk>). A base de dados do miRBase é um repositório das seqüências publicadas de miRNA e anotações associadas com a funcionalidade, fornecida previamente pelo registro do miRNA, determinando a nomenclatura de registro no miRBase. As regras consistem basicamente em: 1- Estabelecer um prefixo contendo de no máximo 3-4 letras para identificação da espécie; 2- A seqüência madura do miRNA é identificada no banco de dados como “miR” e seqüência precursora “mir”; 3- Deve-se adicionar um número de referência. No caso do miRNA novo ser similar a um miRNA de outra espécie presente no miRBase, o mesmo número deve ser mantido,

significando que aqueles miRNAs representam genes ortólogos. No caso de seqüências parálogas que apresentam apenas pequenas diferenças mantém-se o mesmo número e atribuem-se sufixos. Desta forma, a nomenclatura estabelecida para os miRNAs de *S. mansoni* de acordo com as regras do miRBase é: “sma-miR-xx”. Onde “xx” será um número estabelecido de acordo com as normas exigidas. No presente estudo a nomenclatura foi ordenada de forma crescente (Apêndice 1). Como requisito para registro no miRBase é necessário um artigo aceito para publicação (Griffiths-Jones *et al.*, 2008). O presente trabalho encontra-se submetido e após a decisão da revista a nomenclatura será alterada e estabelecida de forma oficial. Isso significa que apenas os miRNAs que tiveram sua expressão confirmada pela técnica de *northern blot* serão depositados no banco de dados.

Com o objetivo de comparar os miRNAs de *S. mansoni* com miRNAs de diferentes espécies, as seqüências candidatas a miRNA foram analisadas utilizando o banco de dados miRBase (versão 13). Neste contexto, foi realizada busca por similaridade entre as seqüências de miRNA de *S. mansoni* e miRNAs de outros organismos. Nossos resultados mostraram que apenas 41 dos 211 candidatos a miRNA de *S. mansoni* apresentaram identidade com ao menos uma das famílias de miRNA depositadas no banco de dados, dentre elas miR-832, miR-71, miR-297 e let-7 (Tabela I). Sma-miR-36 apresentou 100% de homologia com a família de miRNA miR-87, demonstrando alto grau de conservação filogenética deste miRNA entre mais de 10 espécies. Estudos anteriores em *C. elegans* demonstraram que este miRNA é muito expresso durante todos os estágios de desenvolvimento deste nematódeo (Lim *et al.*, 2003), sugerindo que este miRNA também possa exercer papel importante nas diferentes fases de *S. mansoni*. Alta similaridade também foi encontrada com miRNAs do platelminto *Schmidtea mediterrânea*, espécie de planária de água doce também pertencente ao Superfilo Lophotrochozoa, confirmando a idéia da possível existência de miRNAs filogeneticamente conservados entre espécies

relativamente próximas. Zayas e colaboradores (2005) compararam ESTs de *S. mediterrânea* e *S. mansoni* e observaram 47% de similaridade entre os transcritos, o que explica a presença de possíveis miRNAs conservados entre ambas as espécies. Outro miRNA presente em diferentes espécies pertence à família miR-739, observado em *Danio rerio* (dre-miR-739). Este gene tem origem intergênica e foi seqüenciado sete vezes no genoma do parasito *S. mansoni*. Com base na literatura, este miRNA tem como alvo o gene que codifica para a proteína de interação com Tax 1, que é também homólogo em *C. elegans* (Kloosterman *et al.*, 2006). Sabe-se que este gene está possivelmente envolvido em vias de sinalização celular e na expressão de proteínas de membrana. Contudo, o grau de conservação dos miRNAs entre as espécies utilizando *S. mansoni* como modelo é bem menor se fossemos realizar a mesma análise utilizando outro organismo modelo, uma vez que *S. mansoni* e organismos do seu filo possuem pouca ou nenhuma informação disponível no banco de dados de miRNA. Mesmo sabendo que os primeiros miRNAs foram identificados em nematódeos (Lee *et al.*, 1993), atualmente a grande maioria da informação disponível no miRBase (Griffiths-Jones, 2004; Griffiths-Jones *et al.*, 2006) representa dados relacionados a plantas ou humanos (Griffiths-Jones *et al.*, 2008).

Tabela I – Candidatos a miRNA de *S. mansoni* que apresentaram similaridade com miRNAs de diferentes espécies presentes no banco de dados, miRBase.

miRNAs de <i>S. mansoni</i>	Número de <i>Hits</i>	Melhor <i>Hit</i>	Família miRNA
sma-miR-7	4	gma-miR171b	miR-171
sma-miR-9	2	ptr-miR-1303	miR-1303
sma-miR-16	1	gga-miR-1465	miR-1465
sma-miR-17	10	dre-miR-739	miR-739/miR-197
sma-miR-21	1	ebv-miR-BART2	miR-BART2
sma-miR-24	10	ptr-miR-451	miR-451
sma-miR-31	10	hsa-miR-513b	miR-513
sma-miR-32	1	ath-miR832-5p	miR-832
sma-miR-33	3	tca-miR-71	miR-71
sma-miR-36	10	tca-miR-87	miR-87
sma-miR-37	1	cbr-miR-240	miR-240
sma-miR-40	4	ptr-miR-432	miR-432
sma-miR-47	1	sme-let-7c	let-7
sma-miR-77	1	ptc-miR472b	miR472
sma-miR-78	1	ath-miR832-5p	miR-832
sma-miR-86	1	hsa-miR-1268	miR-1268
sma-miR-96	10	ptr-let-7b	let-7
sma-miR-103	1	kshv-miR-K12-3	miR-K12
sma-miR-105	8	dya-miR-289	miR-289
sma-miR-122	1	gga-miR-1465	miR-1465
sma-miR-129	8	dya-miR-289	miR-289
sma-miR-141	2	osa-miR166i	miR-166
sma-miR-149	1	gga-miR-1810	miR-1810
sma-miR-150	1	mmu-miR-297b-3p	miR-297
sma-miR-156	3	tni-miR-101b	miR-101
sma-miR-170	1	cel-miR-78	miR-78
sma-miR-182	10	dre-miR-739	miR-739
sma-miR-187	1	mmu-miR-297b-3p	miR-297
sma-miR-205	1	ppt-miR896	miR-896
sma-miR-209	1	mmu-miR-297b-3p	miR-297
sma-miR-210	1	gga-miR-1810	miR-1810
sma-miR-213	1	ath-miR832-5p	miR-832
sma-miR-216	1	kshv-miR-K12-3	miR-K12
sma-miR-226	1	mmu-miR-297b-3p	miR-297
sma-miR-234	1	gga-miR-1810	miR-1810
sma-miR-241	1	odi-miR-1500	miR-1500
sma-miR-247	4	ptr-miR-432	miR-432
sma-miR-262	1	sme-let-7c	let-7
sma-miR-276	1	mmu-miR-297b-3p	miR-297
sma-miR-278	1	gga-miR-1810	miR-1810
sma-miR-283	2	ptr-miR-1299	miR-1299
sma-miR-284	1	mmu-miR-297b-3p	miR-297

* Nomenclatura seguida pelo miRBase

4.2 – Análise da expressão dos miRNAs de *S. mansoni*

O método para identificação de miRNAs mais utilizado atualmente consiste na clonagem dos pequenos RNAs fracionados em gel de poliacrilamida e a confirmação da expressão e do tamanho através da técnica de *northern blot* (Kloosterman *et al.*, 2006; Xue *et al.*, 2008; Subramanian *et al.*, 2008; Shao *et al.*, 2008). Apesar de novas técnicas estarem surgindo para análises em larga escala, *northern blot* ainda é a técnica padrão para a validação de potenciais candidatos a miRNA. O número de seqüências em uma biblioteca de pequenos RNAs na qual um miRNA está representado, normalmente, correlaciona-se com o nível de expressão daquele miRNA (Landgraf *et al.*, 2007; Fahlgren *et al.*, 2007). Por exemplo, miRNAs que tiveram suas seqüências representadas muitas vezes na biblioteca são provavelmente expressos em altos níveis quando comparado a miRNAs que foram seqüenciados com baixa freqüência. No presente trabalho foram analisados os níveis de expressão dos 30 candidatos a miRNA de *S. mansoni* que tiveram a sua seqüência representada o maior número de vezes na biblioteca de cDNA, denominados como os miRNAs mais expressos. Experimento de *northern blot* foi realizado utilizando RNA total de *pool* de verme adulto de *S. mansoni* hibridizado com sonda marcada com biotina. A expressão de miRNAs é altamente regulada no tempo e espaço (Válóczi *et al.*, 2006; Winter *et al.*, 2009). miRNAs que são estágio-específico, ou que apresentam uma expressão diferenciada em determinado estágio do organismo, indicam possível papel funcional durante o estágio em que são expressos. Desta forma, nós analisamos a expressão dos miRNAs no estágio larval presente no hospedeiro vertebrado, o esquistossômulo, e no verme adulto. Outros estudos analisaram o padrão de expressão da ribonuclease Dicer nas diferentes fases do parasito e revelaram uma alta expressão de Dicer, enzima fundamental presente na via de RNAi, durante a fase de esquistossômulo, sendo 3 vezes maior comparado a verme adulto (Krautz-Peterson & Skelly, 2008). É possível que a alta

expressão de Dicer durante esta fase seja importante para controlar a ativação de retrotransposons, cuja probabilidade de ocorrer neste período é, provavelmente, maior devido aos mecanismos de divisão celular e crescimento. Estas observações indicam que o maior padrão de expressão de Dicer é um indicativo de uma maior expressão de todos os componentes envolvidos na maquinaria de RNAi (Krautz-Peterson *et al.*, 2007). Neste contexto, nossos resultados indicaram que 11 dos 30 miRNAs analisados provavelmente desempenhem papel importante em pelo menos umas das duas fases analisadas, ou em ambas (Figura 8). Baseando no padrão de expressão de ambas as fases analisadas, os miRNAs representaram 3 diferentes grupos: miRNAs expressos em ambas as fases (sma-miR-2 e sma-miR-7), miRNAs específicos de verme adulto (sma-miR-4, sma-miR-6, sma-miR-9, sma-miR-32, sma-miR-125, sma-miR-5 e sma-miR-3), e miRNAs específicos de esquistossômulo (sma-miR-20, sma-miR-18, sma-miR-22, sma-miR-26 e sma-Bantam). Nós também analisamos a expressão dos 5 miRNAs (Sja-Bantam, Sja-miR-125, Sja-miR-71, sja-miR-let7 e Sja-miR-new1) identificados em *S. japonicum*, espécie de *Schistosoma* que causa esquistossomose intestinal na China, Filipinas e Indonésia (Xue *et al.*, 2008). Foram marcadas sondas complementares as seqüências dos miRNAs de *S. japonicum* e hibridizadas em RNA de *S. mansoni*. Três dos cinco miRNAs analisados (miR-71, miR-125 e Bantam) foram detectados por *northern blot*, demonstrando que estes miRNAs são conservados entre espécies próximas evolutivamente (Figura 8). Contudo, foi observada uma diferença no padrão de expressão de sja-miR-71 e sja-bantam em *S. mansoni* quando comparado com os resultados encontrados em *S. japonicum*. Xue e colaboradores (2008) detectaram uma baixa na expressão de ambos miRNAs utilizando a técnica de RT-PCR em tempo real na fase de esquistossômulo. Nossos resultados indicam que ambos os miRNAs são expressos durante esta fase de transformação, sendo sma-miR-71 também expresso durante a fase de verme adulto. Ainda não se sabe claramente a função dos miRNAs da

família miR-71 nos organismos, porém sabe-se que os miRNAs da família Bantam estimulam a proliferação celular e reduzem morte celular programada (Brennecke *et al.*, 2003), importantes processos para o desenvolvimento do parasito. Supomos que outros miRNAs cuja expressão não foi detectada em *S. mansoni* (*new-1* e *let-7*), nas fases de verme adulto e esquistossômulo, sejam autênticos, porém expressos em alguma outra fase do ciclo de vida do parasito, ou possivelmente apresentam baixo grau de expressão nas fases analisadas não sendo possível detectá-los pela técnica de *northern blot*. Em particular *let-7* cuja seqüência e padrão de expressão mostraram-se conservados em um grande número de organismos, incluindo vertebrados, moluscos, anelídeos, artrópodes, entre outros (Pasquinelli *et al.*, 2000), sugerindo que *let-7* possui papel importante nos processos biológicos em animais. Em *C. elegans*, *let-7* regula o desenvolvimento larval inibindo a expressão de *lin-14* (Reinhart *et al.*, 2000; Slack *et al.*, 2000). Estas informações são importantes para futuros estudos que visem esclarecer o papel dos miRNAs no parasito e o seu provável envolvimento em eventos celulares relacionados ao desenvolvimento do parasito, defesa contra patógenos e resposta ao estresse.

O gene de miRNA *sma-miR-1* foi seqüenciado o maior número de vezes. Apesar do método de seqüenciamento ser um indicativo do grau da expressão de determinado miRNA, a expressão de *sma-miR-1* não foi detectada pela técnica de *northern blot* nos estágios de verme adulto e esquistossômulo. Este resultado sugere que *sma-miR-1* possa talvez representar algum artefato de clonagem, ou que as análises de expressão dos pequenos RNAs utilizando *northern blot* nem sempre reflitam a freqüência da clonagem. Como por exemplo, *sma-miR-32* foi representado por somente 2 clones na biblioteca de pequenos RNAs, o que indicaria um nível de expressão aproximadamente 12 vezes menor quando comparado ao *sma-miR-3*, que foi representado por 24 seqüências, levando em consideração a freqüência da clonagem como medida do nível de expressão. Contudo, nossos resultados

demonstraram uma maior expressão de sma-miR-32 comparado ao sma-miR-3 em verme adulto (Figura 8). A mesma situação já foi observada em outros organismos e não se sabe claramente a razão. Acredita-se que uma possível explicação possa ser atribuída a uma diferença nos nucleotídeos das extremidades 5' e 3' prejudicando a ligação com os adaptadores 5' e 3', respectivamente, e conseqüentemente levando a uma clonagem parcial dos fragmentos (Sunkar & Zhu, 2004; Reddy *et al.*, 2009). Esta diferença deve-se as propriedades físicas de determinados miRNAs, incluindo a composição da seqüência, ou até mesmo modificações pós-transcricionais, dificultando a clonagem de miRNAs específicos (Luciano *et al.*, 2004; Yang *et al.*, 2006).

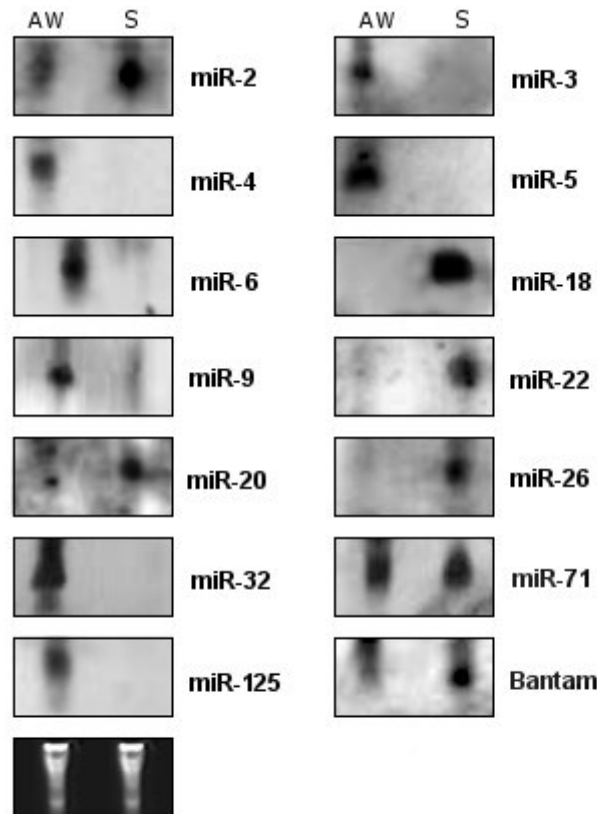


Figura 8 – Análises da expressão de miRNAs pela técnica de *northern blot*. Os miRNAs presentes foram detectados através do uso de sondas não-radioativas marcadas com dioxigenina (DIG), acrescentado anticorpo antidigoxigenina conjugado com fosfatase alcalina (AP-alkaline phosphatase, $150\text{mU}\cdot\text{ml}^{-1}$). $60\ \mu\text{g}$ de RNA total das fases de verme adulto (AW) e esquistossômulo (S) foram utilizados para cada amostra. miR-71, miR-125 e Bantam representam os miRNAs identificados primeiramente em *S. japonicum*, e confirmados em *S. mansoni*. Ao final da figura está o controle do gel corado com brometo de etídeo.

4.3 – Estrutura secundária dos miRNAs

De uma forma geral, miRNAs e siRNAs são muito similares em termos de suas características moleculares, função e biogênese. São moléculas pequenas de RNA contendo de 21-25 nucleotídeos, ambos participam do complexo de silenciamento RISC e se ligam a mRNAs alvo para repressão gênica pós-transcricional. Contudo, miRNAs e siRNAs são diferentes em vários aspectos, particularmente ligados à origem evolutiva, conservação, e o tipo de genes que silenciam (Bartel, 2004). Estas características são imprescindíveis para classificar estas moléculas. Contudo, o melhor método para diferenciar miRNAs de outros pequenos RNAs endógenos é a capacidade que eles possuem em adquirir a estrutura secundária em forma de grampo (Kim, 2005; Berezikov *et al.*, 2006b). Entretanto, o número de possíveis estruturas secundárias de um RNA cresce enormemente com o tamanho da molécula. Neste contexto, foram selecionados para a realização da predição da estrutura secundária os mesmos candidatos analisados pela técnica de *northern blot*. Os miRNAs foram mapeados em diferentes regiões genômicas totalizando aproximadamente 500 localizações para cada seqüência. Com o objetivo de identificar a localização do possível gene de miRNA, as estruturas secundárias de todos os locais onde os miRNAs foram mapeados foram analisadas através do programa RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>), e cada imagem manualmente inspecionada. Um total de 26 seqüências foi capaz de adquirir a estrutura secundária em grampo característica de precursores de miRNAs (Tabela II). Os miRNAs foram divididos em dois grupos: miRNAs capazes de adotar estrutura secundária com forma de grampo que tiveram sua expressão confirmada pela técnica de *northern blot* (Figura 9), e miRNAs capazes de adotar estrutura secundária com forma de grampo que não tiveram sua expressão confirmada pela técnica de *northern blot* (Figura 10). Em seguida, a localização genômica de cada estrutura foi analisada utilizando o programa Blast contra o banco de dados do genoma de *S. mansoni*

(http://www.sanger.ac.uk/Projects/S_mansoni, versão 3.1). Os miRNAs analisados foram localizados em regiões intergênicas. Esta observação está de acordo com resultados já descritos na literatura, na qual a maioria dos miRNAs descritos em outros organismos possui seus precursores presentes em regiões de introns ou regiões intergênicas (Lau *et al.*, 2001; Bartel, 2004; Rodriguez, *et al.*, 2004; Martinez *et al.*, 2008). Devido ao fato dos miRNAs não estarem em sua maioria em regiões codificantes, acredita-se que miRNAs intergênicos representem unidades transcricionais independentes, com seus próprios elementos reguladores transcricionais. Acredita-se que os miRNAs possuam metilguanossina (m7G) à extremidade 5' (5'-cap), e cauda poli (A) à extremidade 3', sugerindo que a transcrição é realizada pela enzima RNA polimerase II (Smalheiser *et al.*, 2003; Cai *et al.*, 2004; Bracht *et al.*, 2004). Muitos miRNAs intergênicos encontram-se agrupados no genoma, sugerindo que um pri-miRNA “policistrônico” seja capaz de gerar vários miRNAs distintos após a finalização do processamento (Lagos-Quintana *et al.*, 2001; Altuvia *et al.*, 2005). Por outro lado, trabalhos anteriores sugerem que os miRNAs presentes em exons ou introns de transcritos que codificam proteínas possuem sua expressão coincidente com a do gene na qual estão presentes (Rodriguez *et al.*, 2004).

Tabela II – miRNAs de *S. mansoni* capazes de adquirir estrutura secundária, e seu respectivo tamanho, posição genômica e resultado da validação por *northern blot*.

miRNA	Seqüência (5'-3')	Tamanho (nt)	Localização genômica*	Validação por <i>northern blot</i>
sma-miR-2	TGATCCGGCGGGTGCGGTGCCAGTT	25	Smp_scaff013386:1052..1077	+
sma-miR-3	ACGTGAGTATAGGGGCGAAAGACTA	25	Smp_scaff 015051:1455..1482	+
sma-miR-4	TTGAATCTGGGTCGTGCGGTCGCAT	25	Smp_scaff 015219:199..228	+
sma-miR-5	AATGTAGGTAAGGGAAGTCGGCA	23	Smp_scaff 015642:1137..1159	+
sma-miR-6	TTCGATTCCGGAGAGGGAGCCTG	23	Smp_scaff 015219:463..486	+
sma-miR-7	TTGATTCGGTGGGTGGTGGTGC	22	Smp_scaff 015609:101..123	-
sma-miR-9	TAAATCGGAACAGAGACGGGGTGC	24	Smp_scaff 007962:435..461	+
sma-miR-10	TAAACCAGGGGTCGGGAGTTC	21	Smp_scaff 003291:1015..1035	-
sma-miR-11	AGGTCGTTGGATTGGCGTCGTTGT	24	Smp_scaff 015609:591..615	-
sma-miR-13	TACATCCATCATGTCTCCTCACAA	24	Smp_scaff 15230:1826..1852	-
sma-miR-15	CGGGTGTCCGGAGCGGACGGC	21	Smp_scaff 003242:75..98	-
sma-miR-16	ATAGGTGGGAGCTGGGTGCTTC	22	Smp_scaff 015642:1783..1804	-
sma-miR-17	CGAGGTGGAGAAGGGTTCCA	20	Smp_scaff 014853:236..255	-
sma-miR-18	TTGGGGTACGGGGTTCGAC	20	Smp_scaff 015270:461..481	+
sma-miR-19	TGGTTGGGTCGGTAGGTGG	19	Smp_scaff 017601:2493..2511	-
sma-miR-20	AGTCGTGGTGCGAATTGCGGTGT	23	Smp_scaff 015642:1952..1974	+
sma-miR-21	GTTGGTGGAGCGATTTGTCTG	21	Smp_scaff 015122:970..990	-
sma-miR-22	TGATCGTAAGGCAATGTGGTGT	24	Smp_scaff 013172:1207..1230	+
sma-miR-23	GCGGTGTGTCGGAGATGGCGGC	22	Smp_scaff 015318:1301..1323	-
sma-miR-24	ATGGTAACCGGGTGTCTATG	22	Smp_scaff 013300:510..531	-
sma-miR-25	ACGTCGGCGAGATTCTGGGA	21	Smp_scaff 006489:1115..1135	-
sma-miR-26	TTGGTGTGAGGGTCTGCTTG	20	Smp_scaff 015318:1119..1138	+
sma-miR-27	TTGAATCTGGGTCGTGCG	18	Smp_scaff 015219:210..227	-
sma-miR-29	AAGTCGTAACAAGGTTTCCGTAGGT	25	Smp_scaff 013172:104..128	-
sma-miR-31	CGGAGTTGAAGCAAGAGGTGCA	24	Smp_scaff 015642:2198..2221	-
sma-miR-32	CGTACTGGGATCGGGGCTTGCA	23	Smp_scaff 003570:35..57	+

+ : *northern blot* positivo;

- : *northern blot* negativo;

*Smp_scaffxxxxx:x.y; indica o *scaffold* na versão x do genoma; x...y, indica a posição no *scaffold*.

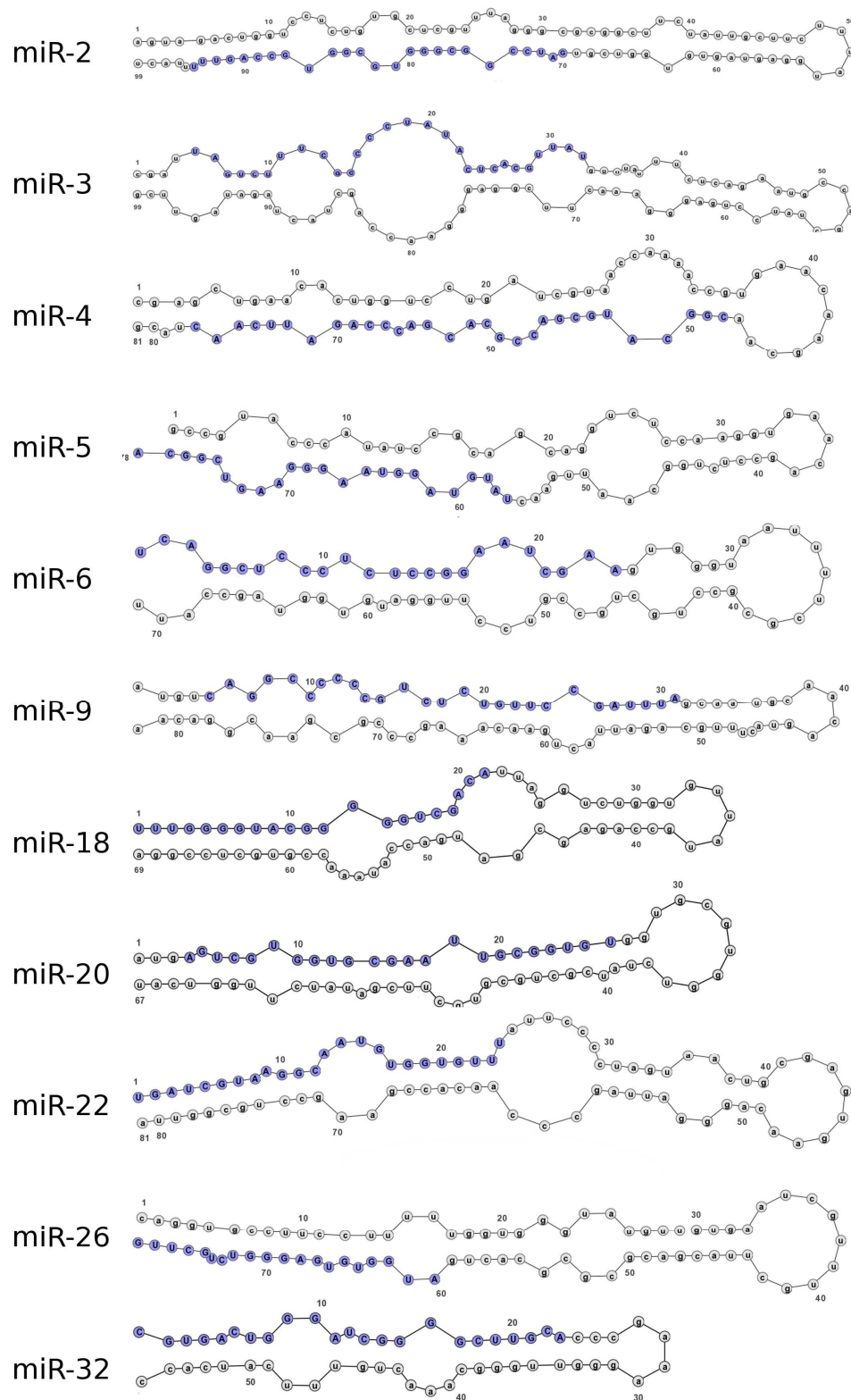


Figura 9 – Estrutura secundária dos miRNAs que tiveram sua expressão confirmada pela técnica de *northern blot*. Em azul está identificada a seqüência do miRNA maduro.

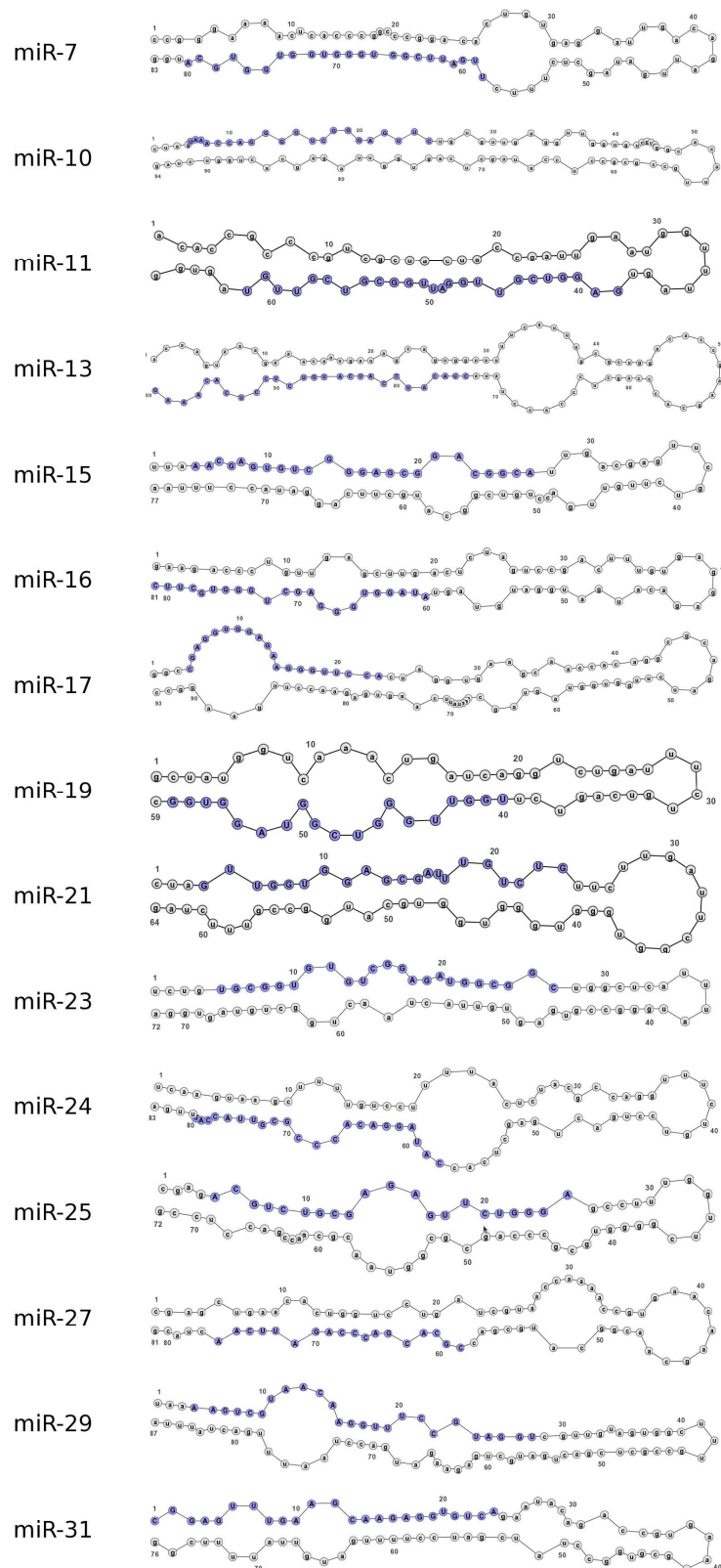


Figura 10 – Estrutura secundária dos miRNAs cuja expressão não foi detectada pela técnica de *northern blot*. Em azul está identificada a seqüência do miRNA maduro.

O próximo passo para o presente estudo visa o emprego do mecanismo de RNAi para estudos funcionais dos miRNAs identificados em *S. mansoni*. O mecanismo de RNAi, além de ser uma descoberta que modificou a forma de ver as células, o seu uso como ferramenta para inibir a expressão de genes específicos também revolucionou a forma de determinar o papel de genes nas células de vários organismos. Para isso diferentes estratégias podem ser empregadas que permitam a redução da expressão de genes específicos (conhecido como mecanismo de *knock-down*), através de moléculas complementares. A estratégia é basicamente transfectar pequenas moléculas de RNA de fita dupla (18 a 26 pb) diretamente nos parasitas em cultura, visando a degradação específica de um gene. Contudo, o primeiro desafio a ser vencido para o desenvolvimento desse tipo de terapia é a identificação de genes alvo. A identificação de genes alvo não é um processo trivial, principalmente em animais no qual o pareamento entre o miRNA e o mRNA alvo ocorre de forma imperfeita impedindo a tradução. Contudo, existem atualmente várias ferramentas computacionais que podem ser usadas para obter as possíveis seqüências alvo para cada gene de miRNA. Um dos primeiros trabalhos para identificação de genes alvo foi em *Drosophila*, mostrando que o miRNA *bantam* regula negativamente o gene pró-apoptótico *hid* (Brenneck *et al.*, 2003; Stark *et al.*, 2003). Em seguida, outros trabalhos foram publicados demonstrando a predição de genes alvo em vertebrados (John *et al.*, 2004; Kiriakidou *et al.*, 2004), e hoje existem várias publicações sobre o assunto. A maioria das ferramentas computacionais desenvolvidas para identificar seqüências de mRNA alvo baseia-se na complementaridade entre a seqüência *seed* do miRNA, que corresponde aos 6-8 nucleotídeos da extremidade 5' da seqüência do miRNA maduro, e a seqüência alvo (Enright *et al.*, 2003). Os métodos também abordam aspectos relacionados à conservação evolutiva entre as espécies e estabilidade termodinâmica para determinar a formação dos duplexes miRNA-mRNA (Lewis *et al.*,

2003; John *et al.*, 2004; Krek *et al.*, 2005). Este estudo será importante para elucidar as funções dos miRNAs identificados em *S. mansoni*. Uma variedade de fatores influencia o desenvolvimento do parasito, a interação parasito-hospedeiro, sua diferenciação, entre outros mecanismos ilustrando a complexidade deste organismo. Uma maior compreensão do papel dos miRNAs para o crescimento do parasito, desenvolvimento, resistência ao estresse, e habilidade de penetração no hospedeiro vertebrado serão particularmente importantes devido ao ciclo de vida complexo deste organismo, envolvendo vários estágios de desenvolvimento e um repertório único de expressão de genes durante as diferentes fases. A elucidação dos mecanismos moleculares de controle de expressão gênica em *Schistosoma* e um melhor entendimento dos mecanismos que controlam o crescimento e morte celular, e sua relação com a transformação e adaptação do parasito podem, além do aumento do conhecimento básico do parasito, trazer novas perspectivas para a manipulação terapêutica e descoberta de novas drogas.

V – CONCLUSÕES

- Verificamos neste trabalho que *S. mansoni* expressa de modo regulado miRNAs, alguns dos quais pertencem a famílias de miRNAs conhecidos em outras espécies e outros representam novos miRNAs.

- 1 Foram geradas seis bibliotecas independentes de pequenos RNAs de vermes adultos de *S. mansoni* utilizando a técnica de fracionamento de gel;
- 2 Através de clonagem e seqüenciamento de 1200 seqüências provenientes de bibliotecas de pequenos RNAs, foram identificados 211 potenciais candidatos a miRNA utilizando ferramentas de bioinformática;
- 3 41 dos 211 candidatos a miRNA de *S. mansoni* apresentaram homologia com ao menos uma das famílias de miRNA depositadas no banco de dados miRBase;
- 4 11 de 30 candidatos a miRNA testados foram detectados pela técnica de *northern blot* nas fases de verme adulto e/ou esquistossômulo;
- 5 3 de 5 miRNA identificados em *S. japonicum* foram detectados pela técnica de *northern blot* em *S. mansoni* nas fases de verme adulto e/ou esquistossômulo;
- 6 26 de 30 candidatos a miRNA analisados foram capazes de adquirir a estrutura secundária típica de um miRNA;
- 7 Os 26 miRNAs analisados estão localizados em regiões intergênicas.

VI – CONSIDERAÇÕES FINAIS

O número de novos microRNAs descobertos em várias espécies está crescendo a cada ano devido à caracterização deste tipo de molécula em novos genomas seqüenciados e pelo importante papel desempenhado como regulador de importantes processos biológicos. O presente trabalho demonstrou pela primeira vez a presença de miRNAs no genoma do parasito *S. mansoni* seguindo a combinação dos critérios de biogênese e expressão. Apesar do mecanismo de ação e do verdadeiro papel do miRNAs na regulação gênica ser ainda foco de bastante estudo, acredita-se que os miRNAs sejam essenciais para a diferenciação e desenvolvimento dos metazoários, possivelmente também para o *S. mansoni* (Reinhart *et al.*, 2000; Bartel, 2004; Giraldez *et al.*, 2005). miRNAs podem também permitir uma melhor compreensão em relação à grande quantidade de elementos transponíveis e elementos repetitivos presente no genoma do parasito, cuja expressão possa ser regulada através da maquinaria de RNAi. Como ainda pouco se conhece dessa nova categoria de RNA, os miRNAs, estes estudos abrem um vasto campo de pesquisa sobre o genoma deste organismo.

O uso da bioinformática tornou-se uma ferramenta imprescindível para atender aos desafios científicos da era genômica. A bioinformática abrange um campo multidisciplinar desde a biologia, matemática, ciência da computação, até a tecnologia da informação. O termo “bioinformática” foi introduzido em 1980 por Hwa Lim, do *Supercomputer Computations Research Institute* (Universidade do Estado da Flórida) visando englobar estudos computacionais nos problemas biológicos. A bioinformática possibilitou a análise e a extração de conhecimento do grande volume de dados provenientes dos atuais projetos da biologia. Assim, o desenvolvimento de novas metodologias e ferramentas computacionais é essencial para analisar toda a informação disponível.

A identificação de miRNAs para todos os tipos de organismos será um instrumento não somente para estudos de regulação gênica como também poderá auxiliar na organização do genoma (ex. atividade de transposons e organização da cromatina), em

estudos de comparação filogenética, estudos envolvendo comparação de desenvolvimento entre espécies, dentre outras análises evolucionárias. No futuro espera-se analisar o papel dos miRNAs em um cenário no qual estes RNAs sejam integrantes de um circuito de regulação gênica mais complexo, na qual sua compreensão será requisito para entender todo o sistema em eucariotas.

VII – ENDEREÇOS ELETRÔNICOS

<http://apps.who.int/tdr/svc/diseases/schistosomiasis> - Página da TDR (*Tropical Disease Research*) que fornece informações gerais sobre a esquistossomose.

<http://www.schistodb.org> – SchistoDB, banco de dados genômico desenvolvido pelo nosso grupo para incorporar todos os dados disponíveis sobre o *S. mansoni*.

<http://www.genedb.org/genedb/smansoni> - GeneDB, desenvolvido pelo Sanger Institute, o portal contém informações sobre o genoma de *S. mansoni*.

<http://www.geneontology.org> – Página do Gene Ontology: Variedade de informações sobre produtos de genes e anotações de diferentes organismos.

<http://www.sanger.ac.uk/Software/Rfam> - Banco de dados contendo famílias de RNAs não-codificadores.

<http://bioweb2.pasteur.fr/docs/EMBOSS/restrict.html> - Permite identificação de sítios de endonucleases.

<http://www.ncbi.nlm.nih.gov/BLAST> - Página da ferramenta Blast, utilizada para busca de homologia de seqüências de nucleotídeos e/ou aminoácidos.

<http://www.tigr.org> - Página do Instituto de Pesquisa Genômica contendo informações sobre o genoma de diferentes organismos.

<http://microrna.sanger.ac.uk> - Banco de dados miRBase: Possui seqüências de microRNAs identificados em diferentes organismos.

<http://varna.lri.fr/index.html> - Programa para edição de estruturas secundárias.

<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi> - Página do programa RNAfold. Programa para predição de estrutura secundária utilizando seqüências de RNA ou DNA.

http://www.sanger.ac.uk/Projects/S_mansoni - Página sobre o projeto genoma do parasito *S. mansoni*.

VIII - REFERÊNCIAS BIBLIOGRÁFICAS

- Al-Sherbiny M, Osman A, Barakat R, El Morshedy H, Bergquist R, Olds R. In vitro cellular and humoral responses to *Schistosoma mansoni* vaccine candidate antigens. *Acta Tropica*, 88:117-130, 2003.
- Amaral PP, Mattick JS. Noncoding RNA in development. *Mamm Genome*, 19:454-492, 2008.
- Ambros V. microRNAs: tiny regulators with great potential. *Cell*, 107:823-826, 2001.
- Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. *Cell*, 113:673-676, 2003.
- Ambros V. The functions of animal microRNAs. *Nature*, 431:350-355, 2004.
- Al-Anouti F, Ananvoranich S. Comparative analysis of antisense RNA, double-stranded RNA, and delta ribozyme-mediated gene regulation in *Toxoplasma gondii*. *Antisense and Nucleic Acid Drug Dev.*, 12:275-281, 2002.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool.. *J Mol Biol*, 215:403-410, 1990.
- Altuvia Y, Landgraf P, Lithwick G, Elefant N, Pfeffer S, Aravin A, Brownstein MJ, Tuschl T, Margalit H. Clustering and conservation patterns of human microRNAs. *Nucleic Acids Res.*, 33:2697-2706, 2005.
- Baev V, Daskalova E, Minkov I. Computational identification of novel microRNA homologs in the chimpanzee genome. *Comput Biol Chem.*, 33:62-70, 2009.
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, 116:281-297, 2004.
- Baulcombe DC. Unwinding RNA silencing. *Science*, 290:1108-1109, 2000.
- Begemann G. MicroRNAs and RNA interference in zebrafish development. *Zebrafish*, 5:111-119, 2008.
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature*, 409:363-366, 2001.
- Berezikov E, Guryev V, van de Belt J, Wienholds E, Plasterk RH, Cuppen E. Phylogenetic shadowing and computational identification of human microRNA genes. *Cell*, 120:21-24, 2005.
- Berezikov E, Thuemmler F, van Laake LW, Kondova I, Bontrop R, Cuppen E, Plasterk RH. Diversity of microRNAs in human and chimpanzee brain. *Nature genetics*, 38:1375-1377, 2006 a.
- Berezikov E, Cuppen E, Plasterk RH. Approaches to microRNA discovery. *Nat Genet.*, 38:S2-7, 2006 b.
- Bergquist NR. Schistosomiasis: from risk assessment to control. *Trends in Parasitology*, 18:309-314, 2002.
- Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African *Trypanosoma brucei*. *Science*, 309:416-422, 2005.

- Berriman M, Haas B, LoVerde P, Wilson A, Dillon GP, Cerqueira G, et al. The genome of the blood fluke *Schistosoma mansoni*. *Nature*, 460:352-358, 2009.
- Beuvink I, Kolb FA, Budach W, Garnier A, Lange J, Natt F, Dengler U, Hall J, Filipowicz W, Weiler J. A novel microarray approach reveals new tissue-specific signatures of known and predicted mammalian microRNAs. *Nucleic acids research*, 35:e52, 2007.
- Bracht J, Hunter S, Eachus R, Weeks P, Pasquinelli AE. Trans-splicing and polyadenylation of let-7 microRNA primary transcripts. *RNA*, 10:1586-1594, 2004.
- Brennecke J, Hipfner DR, Stark A, Russell RB, Cohen SM. Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene *hid* in *Drosophila*. *Cell*, 113:25-36, 2003.
- Brennecke J, Stark A, Russell RB, Cohen SM. Principles of microRNA-target recognition. *PLoS Biol*, 3:e85, 2005.
- Boyle JP, Wu XJ, Shoemaker CB, Yoshino TP. Using RNA interference to manipulate endogenous gene expression in *Schistosoma mansoni* sporocysts. *Mol Biochem Parasitol.*, 128:205-215, 2003.
- Brindley PJ, Kalinna BH, Dalton JP, Day SR, Wong JY, Smythe ML, McManus DP. Proteolytic degradation of host hemoglobin by schistosomes. *Mol Biochem Parasitol.*, 89:1-9, 1997.
- Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA*, 10:1957-1966, 2004.
- Cerutti H, Casas-Mollano JA. On the origin and functions of RNA-mediated silencing: from protists to man. *Curr Genet*, 50:81-99, 2006.
- Chappell L, Baulcombe D, Molnar A. Isolation and cloning of small RNAs from virus-infected plants. In *Current Protocols in Microbiology*. Jon Wiley & Sons Inc. in press. Chapter 16:Unit 16H.2., v.1, 2005.
- Cheng GF, Lin JJ, Shi Y, Jin YX, Fu ZQ, Jin YM, Zhou YC, Cai YM. Dose-dependent inhibition of gynecophoral canal protein gene expression in vitro in the schistosome (*Schistosoma japonicum*) by RNA interference. *Acta Biochim Bioph Sin*, 37:386-390, 2005.
- Chitsulo L, Engels D, Montresor A, Saviolli L. The global status of schistosomiasis and its control. *Acta Tropica*, 77:41-51, 2000.
- Cioli D, Pica-Mattocia L. Praziquantel. *Parasitology Research*, 90:3-9, 2003.
- Coelho MV. O parasito *Schistosoma mansoni*. In: CUNHA, A.S. *Esquistossomose mansoni*. São Paulo: Editora Savier, 1:1-12, 1970.
- Correnti JM, Brindley PJ, Pearce EJ. Long-term suppression of cathepsin B levels by RNA interference retards schistosome growth. *Mol Biochem Parasitol.*, 143:209-215, 2005.
- Costa FF. Non-coding RNAs: lost in translation? *Gene*, 386:1-10, 2007.

- Cunha AS. Avaliação Terapêutica da oxaminiquine na esquistossomose humana pelo método do oograma por biópsia retal. *Revista do Instituto de Medicina Tropical de São Paulo*, 24:88-94, 1982.
- Criscione C, Valentim C, Hirai H, LoVerde P, Anderson T. Genomic linkage map of the human blood fluke *Schistosoma mansoni*. *Genome Biology*, 10:R71, 2009.
- Crooke A, Diez A, Mason PJ, Bautista JM. Transient silencing of *Plasmodium falciparum* bifunctional glucose-6-phosphate dehydrogenase-6-phosphogluconolactonase. *Febs Journal*, 273:1537-1546, 2006.
- Da'dara AA, Skelly PJ, Fatakawala M, Visovatti S, Eriksson E, Harn DA. Comparative efficacy of the *Schistosoma mansoni* nucleic acid vaccine, Sm23, following microseeding or gene gun delivery. *Parasite Immunology*, 24:179-187, 2002.
- DaRocha WD, Otsu K, Teixeira SM, Donelson JE. Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasitol.*, 133:175-186, 2004.
- Degrave WM, Melville S, Ivens A, Aslett M. Parasite genome initiatives. *International Journal of Parasitology*, 31:532-536, 2001.
- Denli AM, Hannon GJ. RNAi: An ever-growing puzzle. *Trends in Biochemical Sciences*, 28:196-201, 2003.
- Dingirard N, Yoshino TP. Potential role of a CD36-like class B scavenger receptor in the binding of modified low-density lipoprotein (acLDL) to the tegumental surface of *Schistosoma mansoni* sporocysts. *Mol Biochem Parasitol.*, 146:219-230, 2006.
- Djikeng A, Shi H, Tschudi C, Ullu E. RNA interference in *Trypanosoma brucei*: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA*, 7:1522-1530, 2001.
- Djikeng A, Shen S, Tschudi C, Ullu E. Analysis of gene function in *Trypanosoma brucei* using RNA interference. *Methods Mol Biol*, 270:287-298, 2004.
- Doenhoff MJ, Pica-Mattoccia L. Praziquantel for the treatment of schistosomiasis: its use for control in areas with endemic disease and prospects for drug resistance. *Expert review of anti-infective therapy*, 4:199-210, 2006.
- Engels D, Chitsulo L, Montresor A, Savioli L. The global epidemiological situation of schistosomiasis and new approaches to control and research. *Acta Tropica*, 82: 139-146, 2002.
- Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. microRNA Target in *Drosophila*. *Genome Biology*, 5:R1, 2003.
- Esquela-Kerscher A, Slack FJ. Oncomirs — microRNAs with a role in cancer. *Nat Rev Cancer*, 6:259-269, 2006.
- Fabbri M. MicroRNAs and cancer epigenetics. *Curr Opin Investig Drugs*, 9:583-590, 2008.
- Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangel JL, Carrington JC. High-throughput sequencing of Arabidopsis

- microRNAs: evidence for frequent birth and death of miRNA genes. *Plos One*, 2:e219, 2007.
- Fantappie MR, Gimba ER, Rumjanek FD. Lack of DNA methylation in *Schistosoma mansoni*. *Experimental Parasitology*, 98:162-166, 2001.
- Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, Burge CB, Bartel DP. The widespread impact of mammalian microRNAs on mRNA repression and evolution. *Science*, 310:1817-21, 2005.
- Ferrari ML, Coelho PM, Antunes CM, Tavares CA, da Cunha AS. Efficacy of oxamniquine and praziquantel in the treatment of *Schistosoma mansoni* infection: a controlled trial. *Bull World Health Organization*, 81:190-196, 2003.
- Files VS. The study of the vector-parasite relationships in *Schistosoma mansoni*. *Parasitology*, 41:264-269, 1951.
- Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391:806-811, 1998.
- Fonseca CT, Cunha-Neto E, Goldberg AC, Kalil J, et al. Human T cell epitope mapping of the *Schistosoma mansoni* 14-kDa fatty acid-binding protein using cells from patients living in areas endemic for schistosomiasis. *Microbes and Infection*, 7:204-212, 2005.
- Ginger MR, Shore AN, Contreras A, Rijnkels M, Miller J, Gonzalez-Rimbau MF, Rosen JM. A noncoding RNA is a potential marker of cell fate during mammary gland development. *Proc Natl Acad Sci U S A*, 103:5781-5786, 2006.
- Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, et al. MicroRNAs (miRNAs) are small RNAs that regulate gene expression posttranscriptionally. *Science*, 308:833-838, 2005.
- Geldhof P, Vissera A, Clarka D, Saundersa G, Brittona C, GilleardaJ, Berrimana M, Knox D. RNA interference in parasitic helminths: current situation, potential pitfalls and future prospects. *Parasitology*, 134:609-619, 2007.
- Gomes M, Cabral F, Jannotti-Passos LK, Carvalho O, Rodrigues V, Baba EH, Sá RG. Preliminary analysis of miRNA pathway in *Schistosoma mansoni*. *Parasitology International*, 58:61-68, 2009.
- Gotoh O. An improved algorithm for matching biological sequences. *Journal of Molecular Biology*, 162:705-708, 1982.
- Grad Y, Aach J, Hayes GD, Reinhart BJ, Church GM, Ruvkun G, Kim J. Computational and experimental identification of *C. elegans* microRNAs. *Molecular Cell*, 5:1253-1263, 2003.
- Gregory RI, Chendrimada TP, Shiekhattar R. MicroRNA biogenesis: isolation and characterization of the microprocessor complex. *Methods Mol Biol.*, 342:33-47, 2006.
- Grevelding, CG, Sommer, G, and Kunz, W. Female-specific gene expression in *Schistosoma mansoni* is regulated by pairing. *Parasitology*, 115:635-640, 1997.
- Griffiths-Jones S. The miRNA Registry, *Nucleic Acids Res.*, 32:109-111, 2004.

- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.*, 34:140-144, 2006.
- Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, 36:154-158, 2008.
- Gryseels B, Polman K, Clerinx J, Kestens L. Human schistosomiasis. *Lancet*, 368:1106-1118, 2006.
- Hannon GJ. RNA interference. *Nature*, 418:244-251, 2002.
- He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet.*, 5:522-531, 2004.
- Heikkinen L, Asikainen S, Wong G. Identification of phylogenetically conserved sequence motifs in microRNA 5' flanking sites from *C. elegans* and *C. briggsae*. *BMC Mol Biol.*, 9:105, 2008.
- Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, et al. The expansion of the metazoan microRNA repertoire. *BMC Genomics*, 7:25, 2006.
- Hillman GR, Senft AW. Anticholinergic properties of the antischistosomal drug hytacone. *American Journal of Tropical Medicine and Hygiene*, 24:827-834, 1975.
- Hirai H, LoVerde PT. Identification of telomeros on *Schistosoma mansoni* chromosomes by FISH. *Journal of Parasitology*, 82:511-512, 1996.
- Hofacker IL: Vienna RNA secondary structure server. *Nucleic Acids Res.*, 31:3429-3431, 2003.
- Hutvagner G, Zamore PD. A microRNA in a multiple-turnover RNAi enzyme complex. *Science*, 297:2056-2060, 2002.
- Jinek M, Doudna JA. A three-dimensional view of the molecular machinery of RNA interference. *Nature*. 457:405-412, 2009.
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. *Plos Biology*, 2:e363, 2004.
- Jorgensen RA, Cluster PD, English J, Que Q, Napoli CA. Chalcone synthase cosuppression phenotypes in petunia flowers: comparison of sense vs. antisense constructs and single-copy vs. complex T-DNA sequences. *Plant Mol Biol.*, 3:957-973, 1996.
- Kasashima K, Nakamura Y, Kozu T. Altered expression profiles of microRNAs during TPA-induced differentiation of HL-60 cells. *Biochem Biophys Res Commun.*, 322:403-410, 2004.
- Katz N, Peixoto SV. Análise crítica da estimativa do número de portadores de esquistossomose no Brasil. *Revista da Sociedade Brasileira de Medicina Tropical*, 33:303-308, 2000.
- Kawasaki H, Taira K. Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells. *Nature*, 423:838-842, 2003.

- Kawasaki H, Wadhwa R, Taira K. World of small RNAs: from ribozymes to siRNA and miRNA. *Differentiation*, 72:58-64, 2004.
- Kheir MM, Baraka OZ, El-Tom IA, Mukhtar MM, Homieda MM. Effects of single-dose praziquantel on morbidity and mortality resulting from intestinal schistosomiasis. *Eastern Mediterranean Health Journal*, 6:926-931, 2000.
- Kim VN. MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol.*, 6:376-385, 2005.
- King CH, Dickman K, Tisch DJ. Reassessment of the cost of the chronic helminthic infection: A meta-analysis of disability-related outcomes in endemic schistosomiasis. *Lancet*, 365:1561-1569, 2005.
- King CH, Dangerfield-Cha M. The unacknowledged impact of chronic schistosomiasis. *Chronic Illn.*, 4:65-79, 2008.
- Kiriakidou M, Nelson PT, Kouranov A, Fitziev P, Bouyioukos C, Mourelatos Z, Hatzigeorgiou A. A combined computational-experimental approach predicts human microRNA targets. *Genes Dev.*, 18:1165-1178, 2004.
- Kloosterman W, Steiner F, Berezikov E, Bruijn, et al: Cloning and expression of new microRNAs from zebrafish. *Nucleic Acids Res.*, 34:2558-2569, 2006.
- Krautz-Peterson G, Radwanska M, Ndegwa D, Shoemaker CB, Skelly PJ: Optimizing gene suppression in schistosomes using RNA interference. *Mol Biochem Parasitol.*, 153:194-202, 2007.
- Krautz-Peterson G, Skelly PJ. *Schistosoma mansoni*: The dicer gene and its expression. *Exp Parasitol.*, 118:122-128, 2008.
- Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, MacMenamin P, da Piedade I, Gunsalus KC, Stoffel M, Rajewsky N. Combinatorial microRNA target predictions. *Nat Genet.*, 37:495-500, 2005.
- Kunz W. Schistosome male-female interaction: induction of germ-cell differentiation. *Trends Parasitol.*, 17:227-231, 2001.
- Lagos-Quintana M, Rauhut R, Lendeckel W, Tuschl T. Identification of novel genes coding for small expressed RNAs. *Science*, 294:853-858, 2001.
- Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T. New microRNAs from mouse and human. *RNA*, 9:175-179, 2003.
- Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*, 129:1401-1414, 2007.
- Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*, 294:858-862, 2001.
- Lau NC, Lai EC. Diverse roles for RNA in gene regulation. *Genome Biol.*, 6:315, 2005.

- Lee RC, Ambros V. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science*, 294:862-864, 2001.
- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, 75:843-854, 1993.
- Lee Y, An C, Han J, Choi H, Kim J, Yim Y, Lee J, Provost P, Radmark O, Kim S, et al. The nuclear RNase III Drosha initiates microRNA processing. *Nature*, 425:415-419, 2003.
- Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. *Embo Journal*, 23:4051-4060, 2004.
- Lewin B. The mystique of epigenetics. *Cell* 93:301-303, 1998.
- Lewis BP, Shi I, Jones-Rhoades MW, Bartel DP, Burge CB. Prediction of mammalian microRNA targets. *Cell*, 115:787-798, 2003.
- Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. The microRNAs of *C. elegans*. *Genes Dev.*, 17:991-1008, 2003.
- Lim LP, Lau NC, Garrett-Engel P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature*, 433:769-773, 2005.
- Lindbo JA, Dougherty WG. Untranslatable transcripts of the tobacco etch virus coat protein gene sequence can interfere with tobacco etch virus replication in transgenic plants and protoplasts. *Virology*, 189:725-733, 1992.
- Lindbo JA, Silva-Rosales L, Proebsting WM, Dougherty WG. Induction of a highly specific antiviral state in transgenic plants: Implications for regulation of gene expression and virus resistance. *Plant Cell*, 5:1749-1759, 1993.
- Liu CG, et al. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc Natl Acad Sci USA*, 101:9740-9744, 2004.
- Llave, C., Kasschau, K. D., Rector, M. A. and Carrington, J. C. Endogenous and silencing-associated small RNAs in plants. *Plant Cell*, 14:1605 -1619, 2002.
- Loverde PT, Andrade LF, Oliveira G. Signal transduction regulates schistosome reproductive biology. *Curr Opin Microbiol.*, 2009.
- Luciano DJ, Mirsky H, Vendetti NJ, Maas S. RNA editing of a miRNA precursor. *RNA*, 10:1174-1177, 2004.
- Lund E, Guttinge S, Calado A, Dálberg JE, Kutay U. Nuclear export of microRNA precursors. *Science* 303:95-98, 2003.
- Malhotra P, Dasaradhi PV, Kumar A, Mohammed A, Agrawal N, Bhatnagar RK, Chauhan VS. Double-stranded RNA-mediated gene silencing of cysteine proteases (falcipain-1 and -2) of *Plasmodium falciparum*. *Mol Microbiol.*, 45:1245-1254, 2002.
- Martinez NJ, Ow MC, Reece-Hoyes JS, Barrasa MI, Ambros VR, Walhout AJ. Genome-scale spatio temporal analysis of *Caenorhabditis elegans* microRNA promoter activity. *Genome Research*, 18:2005-2015, 2008.

- Marx K A, Bizzaro JW, Blake R D, Tsai MH, Tao LF. Experimental DNA melting behavior of the three major *Schistosoma* species. *Molecular and Biochemical Parasitology*, 107:303-307, 2000.
- Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet.*, 15:17-29, 2006.
- Mattick JS. The genetic signatures of noncoding RNAs. *PLoS Genet.*, 5(4):e1000459, 2009.
- Matzke M, Matzke AJ, Kooter JM. RNA: Guiding gene silencing. *Science*, 293:1080-1083, 2001.
- McManus DP, Hope M. Molecular variation in the human schistosomes. *Acta Tropica*, 53:255-276, 1993.
- McRobert L, McConkey GA. RNA interference (RNAi) inhibits growth of *Plasmodium falciparum*. *Mol Biochem Parasitol.*, 119:273-278, 2002.
- Navarro P, Chantalat S, Foglio M, Chureau C, Vigneau S, Clerc P, Avner P, Rougeulle C. A role for non-coding Tsix transcription in partitioning chromatin domains within the mouse X-inactivation centre. *Epigenetics Chromatin.*, 2(1):8, 2009.
- Ngo H, Tschudi C, Gull K, Ullu E. Double-stranded RNA induces RNA degradation in *Trypanosoma brucei*. *PNAS*, 95:14687-14692, 1998.
- Noya O, De Noya BA, Ballen DE, Bermudez H, Bout D, Hoebeke J. Immunogenicity of synthetic peptides from the Sm31 antigen (cathepsin B) of the *Schistosoma mansoni* adult worms. *Parasite Immunology*, 23:567-573, 2001.
- Oliveira GC, Rodrigues NB, Romanha A, Bahia D. Genome and genomics of Schistosomes. *Canadian Journal of Zoology-Revue Canadienne de Zoologie*, 82:375-390, 2004.
- Pasquinelli AE, Reinhart BJ, Slack F, Martindale MQ, Kuroda MI, Maller B, Hayward DC, Ball EE, Degan B, Muller P, et al. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature*, 408:86-89, 2000.
- Pasquinelli AE, Ruvkun G. Control of developmental timing by microRNAs and their targets. *Ann. Rev. Cell Dev. Biol.*, 18:495-513, 2002.
- Pellegrino J, Siqueira A F Técnica de perfusão para colheita de *Schistosoma mansoni* em cobaias experimentalmente infectadas. *Rev. bras. Malar.*, 8:589-597, 1956.
- Pellegrino J, Coelho PMZ. *Schistosoma mansoni*, wandering capacity of a worm couple. *Journal of Parasitology*, 64:181, 1978.
- Pereira TC, Pascoal VD, Marchesini RB, Maia IG, Magalhães LA, Zanotti-Magalhães EM, Lopes-Cendes I. *Schistosoma mansoni*: evaluation of an RNAi-based treatment targeting HGPRase gene. *Exp Parasitol.*, 118:619-623, 2008.
- Petrocca F, Lieberman J. Micromanipulating cancer: microRNA-based therapeutics? *RNA Biol.*, 6, 2009.

- Pica-Mattoccia L, Dias LCS, Archer S. Binding of oxaminiquine to DNA of schistosomes. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 83:89-96, 1989.
- Ponting CP, Oliver PL, Reik W. Evolution and functions of long noncoding RNAs. *Cell*, 136:629-641, 2009.
- Ramalho-Pinto FJ, Gazinelli G, Howells RE, Mota Santos TA, Figueiredo EA, Pelegrino J. *Schistosoma mansoni*, Defined system for stepwise transformation of cercarie to schistosomule *in vitro*. *Exper. Parasitol.*, 36:360-362, 1974.
- Rao KV, He YX, Kalyanasundaram R. Expression of a 28-kilodalton glutathione S-transferase antigen of *Schistosoma mansoni* on the surface of filamentous phages and evaluation of its vaccine potential. *Clinical and Diagnostic Laboratory Immunology*, 10:536-41, 2003.
- Rathjen T, Nicol C, McConkey G, Dalmay T. Analysis of short RNAs in the malaria parasite and its red blood cell host. *FEBS Letter*, 580:5185-5188, 2006.
- Ravasi T, Suzuki H, Pang KC, Katayama S, Furuno M, et al. Experimental validation of the regulated expression of large numbers of non-coding RNAs from the mouse genome. *Genome Res.*, 16:11-19, 2006.
- Reddy AM, Zheng Y, Jagadeeswaran G, Macmil SL, Graham WB, Roe BA, Desilva U, Zhang W, Sunkar R. Cloning, characterization and expression analysis of porcine microRNAs. *BMC Genomics*, 10:65, 2009.
- Redman C, Robertson A, Fallon PG, Modena J, Kusel GR, Doenhoff MJ. Praziquantel, an urgent and exciting challeng. *Parasitology Today*, 12:14-20, 1996.
- Reinhart BJ, Slack FJ, Basson M, Bettinger JC, Pasquinelli AE, Rougvie AE, Horvitz HR, Ruvkun G. The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature*, 403:901-906, 2000.
- Reinhart BJ, Weinstein EG, Rhoades MW, Bartel B, Bartel DP. MicroRNAs in plants. *Genes & Development*, 16:1616-1626, 2002.
- Rey L. Parasitos e doenças parasitárias do homem nas Américas e na África. *Parasitologia*. Rio de Janeiro: Guanabara Koogan, 2 ed., 731p., 1991.
- Reynolds SR, Dahl CE, Harn DA. T and B epitope determination and analysis of multiple antigenic peptides for the *Schistosoma mansoni* experimental vaccine triose-phosphate isomerase. *The Journal of Immunology*, 152:193-200, 1994.
- Ribeiro F, Coelho PMZ, Vieira LQ, Watson DG, Kusel GR. The effect of Praziquantel treatment on glutathione concentration in *Schistosoma mansoni*. *Parasitology*, 116:229-236, 1998.
- Ribeiro de Jesus A, Araújo I, Bacellar O, Magalhães A, Pearce E, Harn D, Strand M, Carvalho EM. Human immune responses to *Schistosoma mansoni* vaccine candidate antigens. *Infect Immun.*, 68:2797-2803, 2000.
- Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.*, 6:276-277, 2000.

- Rinaldi G, Morales ME, Alrefaei YN, Cancela M, Castillo E, Dalton JP, Tort JF, Brindley PJ. RNA interference targeting leucine aminopeptidase blocks hatching of *Schistosoma mansoni* eggs. *Mol Biochem Parasitol.*, 167:118-126, 2009.
- Robinson KA, Beverley SM. Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasitol.*, 128:217-228, 2003.
- Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. *Genome Res.*, 14:1902-1910, 2004.
- Rollinson D, Walker TK, Simpson AJG. New approaches to schistosome identification. *Parasitology Today*, 2:24-25, 1986.
- Rollinson D, Southgate, VR. The genus *Schistosoma*: A Taxonomic Appraisal. Edition London Academic Press 1:1-49, 1987.
- Saini HK, Griffiths-Jones S, Enright AJ. Genomic analysis of human microRNA transcripts. *Proc. Natl Acad. Sci. USA*, 104:17719-17724, 2007.
- Schratt GM, Tuebing F, Nigh EA, Kane CG, Sabatini ME, Kiebler M, Greenberg ME. A brain-specific microRNA regulates dendritic spine development. *Nature*, 439:283-289, 2006.
- Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD. Asymmetry in the assembly of the RNAi enzyme complex. *Cell*, 115:199-208, 2003.
- Shabalina SA, Koonin EV. Origins and evolution of eukaryotic RNA interference. *Trends Ecol Evol.*, 23:578-587, 2008.
- Shao P, Zhou H, Xiao ZD, He JH, Huang MB, Chen YQ, Qu LH. Identification of novel chicken microRNAs and analysis of their genomic organization. *Gene*, 418:34-40, 2008.
- Shi H, Djikeng A, Mark T, Wirtz E, Tschudi C, Ullu E. Genetic interference in *Trypanosoma brucei* by heritable and inducible double-stranded RNA. *RNA*, 6:1069-1076, 2000.
- Shi H, Djikeng A, Tschudi C, Ullu E. Argonaute protein in the early divergent eukaryote *Trypanosoma brucei*: control of small interfering RNA accumulation and retroposon transcript abundance. *Mol Cell Biol.*, 24:420-427, 2004.
- Shi H, Tschudi C, Ullu E. An unusual Dicer-like1 protein fuels the RNA interference pathway in *Trypanosoma brucei*. *RNA*, 12:1-10, 2006.
- Shi H, Tschudi C, Ullu E. Depletion of newly synthesized Argonaute1 impairs the RNAi response in *Trypanosoma brucei*. *RNA*, 13:1132-1139, 2007.
- Short RB, Menzel MY. Chromosomes of nine species of schistosomes. *Journal of Parasitology*, 46:273-287, 1960.
- Short RB, Menzel MY, Pathak S. Somatic chromosomes of *Schistosoma mansoni*. *Journal of Parasitology*, 65:471-473, 1979.

- Short RB, Grossman AI. Conventional geimsa and C-banded karyotypes of *Schistosoma mansoni* and *S. rodhaini*. *Journal of Parasitology*, 67:661-667, 1981.
- Simpson AJG, Sher A, Mccutchan TF. The genome of *Schistosoma mansoni*, isolation of DNA, its size and repetitive sequences. *Molecular and Biochemical Parasitology*, 6:125-137, 1982.
- Slack FJ, Basson M, Liu Z, Ambros V, Horvitz HR, Ruvkun G. The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell*, 5:659-669, 2000.
- Smalheiser NR. EST analyses predict the existence of a population of chimeric microRNA precursor-mRNA transcripts expressed in normal human and mouse tissues. *Genome Biol.*, 4:403, 2003.
- Smith TF, Waterman MS. Overlapping genes and information theory. *Journal of Theoretical Biology*, 91:379-380, 1981.
- Stark A, Brennecke J, Russel RB, Cohen SM. Identification of *Drosophila* microRNA targets. *Plos Biology*, 1:E60, 2003.
- Storz G. An expanding universe of noncoding RNAs. *Science*, 296:1260-1263, 2002.
- Subramanian S, Fu Y, Sunkar R, Barbazuk WB, Zhu JK, Yu O. Novel and nodulation-regulated microRNAs in soybean roots. *BMC Genomics*, 9:160, 2008.
- Sunkar R, Zhu JK. Novel and stress-regulated microRNAs and other small RNAs from *Arabidopsis*. *Plant Cell*, 16:2001-2019, 2004.
- Sunkar R, Girke T, Zhu JK. Identification and characterization of endogenous small interfering RNAs from rice. *Nucleic Acids Res.*, 33:4443-4454, 2005.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W:improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22:4673-4680, 1994.
- Ullu E, Tschudi C, Chakraborty T. RNA interference in protozoan parasites. *Cell Microbiol.*, 6:509-519, 2004.
- Valadares TE, Coelho PMZ, Pellegrino J, Sampaio IBM. *S. mansoni*: Aspectos da oviposição da cepa LE em camundongos infectados com um casal de vermes. *Revista do Instituto de Medicina Tropical de São Paulo*, 23:6, 1981.
- Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev.*, 20:515-524, 2006.
- Válóczy A, Várallyay E, Kauppinen S, Burgyán J, Havelda Z. Spatio-temporal accumulation of microRNAs is highly coordinated in developing plant tissues. *Plant Journal*, 47:140-151, 2006.
- Van der Werf MJ, et al. Quantification of clinical morbidity associated with schistosome infection in sub-Saharan Africa. *Acta Tropica*, 86:125-139, 2003.
- Vaucheret H. Post-transcriptional small RNA pathways in plants: mechanisms and regulations. *Genes Dev.*, 20:759-771, 2006.

- Vidal L, Blagden S, Attard G, de Bono J. Making sense of antisense. *Eur J Cancer*, 41:2812-2818, 2005.
- Xue X, Sun J, Zhang Q, Wang Z, Huang Y, Pan W: Identification and characterization of novel microRNAs from *Schistosoma japonicum*. *Plos One*, 3:e4034, 2008.
- Zayas R, Herna'nde A, Habermann B, Wang Y, Sary JM, Newmark P. The planarian *Schmidtea mediterranea* as a model for epigenetic germ cell specification: Analysis of ESTs from the hermaphroditic strain. *PNAS*, 102:18491-18496, 2005.
- Zeng Y, Yi R, Cullen B. MiRNAs and siRNAs can inhibit mRNA expression by similar mechanisms. *PNAS*, 100:9779-9784, 2003.
- Zerlotini A, Heiges M, Wang H, Moraes RL, Dominitini AJ, Ruiz JC, Kissinger JC, Oliveira G. SchistoDB: a *Schistosoma mansoni* genome resource. *Nucleic Acids Res.*, 37(Database issue):D579-582, 2009.
- Zhang WW, Matlashewski G. Analysis of antisense and double stranded RNA downregulation of A2 protein expression in *Leishmania donovani*. *Mol Biochem Parasitol.*, 107:315-319, 2000.
- Zhang B, Pan X, Wang Q, Cobb GP, Anderson TA. Computational identification of microRNAs and their targets. *Comput Biol Chem.*, 30:395-407, 2006.
- Zhang X, Cairns M, Rose B, O'Brien C, Shannon K, Clark J, Gamble J, Tran N. Alterations in miRNA processing and expression in pleomorphic adenomas of the salivary gland. *Int J Cancer.*, 124:2855-63, 2009.
- Zhao ZR, Lei L, Liu M, Zhu SC, Ren CP, Wang XN, Shen JJ. *Schistosoma japonicum*: inhibition of Mago nashi gene expression by shRNA-mediated RNA interference. *Exp Parasitol.*, 119:379-384, 2008.
- Wang L, Wang MB, Tu JX, Helliwell CA, Waterhouse PM, Dennis ES, Fu TD, Fan YL. Cloning and characterization of microRNAs from *Brassica napus*. *FEBS Letter*, 58:3848-3856, 2007.
- Waterhouse PM, Wang MB, Lough T. Gene silencing as an adaptive defence against viruses. *Nature*, 411:834-842, 2001.
- Winter J, Jung S, Keller S, Gregory RI, Diederichs S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Bio.*, 11:228-234, 2009.
- Wienholds E, et al. MicroRNA expression in zebrafish embryonic development. *Science*, 309: 310-311, 2005.
- World Health Organization (WHO): Preventive chemotherapy in human helminthiasis: Coordinated use of anthelmintic drugs in control interventions: a manual for health professionals and programme managers, Geneva, 2008.
- Yang Z, Ebright YW, Yu B, Chen X. HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. *Nucleic Acids Res.*, 34:667-675, 2006.

IX - ARTIGOS

Cloning and characterization of *Schistosoma mansoni* microRNAs

Mariana Simões^{1,4*}, Appolinaire Djikeng², Gustavo Cerqueira³, Adhemar Zerlotini^{1,4}, Rosiane A. da Silva-Pereira⁴, Philip LoVerde⁵, Najib El-Sayed^{3*}, Guilherme Oliveira^{4*}

¹Dept. of Bioinformatics, Universidade Federal de Minas Gerais, Av. Antonio Carlos 6627, Belo Horizonte, MG, Brazil

²J Craig Venter Institute (JCVI), 9704 Medical Center Drive, Rockville, MD 20850, USA

³ Biomolecular Sciences Bldg University of Maryland College Park, MD 20742, USA

⁴Laboratory of Cellular and Molecular Parasitology, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Av. Augusto de Lima 1715, Belo Horizonte, 30190-002, Brazil

⁵University of Texas Health Science Center, 7703 Floyd Curl Dr. Mail Code 7760, San Antonio, Texas 78229-3900, USA

*Corresponding authors.

Email addresses:

MS: msimoes@cpqrr.fiocruz.br

AD: adjikeng@jcv.org

GC: cerca@umd.edu

AZ: adhemar@cpqrr.fiocruz.br

RSP: rosiane@cpqrr.fiocruz.br

PLV: loverde@uthscsa.edu

NES: elsayed@umd.edu

GO: oliveira@cpqrr.fiocruz.br

Abstract

Background

Small RNAs play a crucial role in regulating development and controlling gene expression by targeting mRNAs and triggering either translation repression or RNA degradation. microRNAs (miRNAs) are widespread in eukaryotes and several hundred of them have been identified in a wide range of organisms. In schistosomes, the full set of miRNAs and their expression patterns during development are unknown. We sought to identify the majority of miRNAs genes in *S. mansoni* through cDNA cloning and sequencing of size-fractionated RNA samples.

Results

In this study, we report on the cloning and characterization of novel miRNAs from *Schistosoma mansoni*. By sequencing small-RNA cDNA libraries from adult worm pairs, we identified 211 miRNA candidates from the *S. mansoni* genome. Northern blot analysis was used to detect the expression of selected miRNAs and to compare the expression level of these miRNAs between the lung stage schistosomula and adult worm stages. Analysis of the expression pattern of these miRNAs showed that some are present ubiquitously, while others are stage-specific. Out of the 30 most abundant cDNA sequences, expression of 11 novel miRNAs was confirmed by northern blot analysis. Three miRNAs previously identified from *S. japonicum* were also present in *S. mansoni*.

Conclusion

Construction and analysis of a small RNA library led to the identification of 14 novel miRNAs from *S. mansoni* through a combination of molecular cloning, DNA sequencing and expression studies. Our results significantly expand the set of known miRNAs in multicellular parasites and provide a basis for understanding the structural and functional evolution of miRNAs in these metazoan parasites.

Background

Small non-coding RNAs are increasingly providing insights into important aspects of the biology of many organisms [1-2]. They include small interfering RNAs (siRNAs) and microRNAs (miRNAs), which are hallmarks of two important processes involved in RNA silencing [3]. RNA silencing is a general process in which small RNA molecules derived from precursor dsRNA molecules trigger sequence-specific repression of gene expression [4-6].

miRNAs comprise a family of non-coding RNAs with approximately 21-25 nucleotides that down-regulate gene expression at the post-transcriptional level. miRNAs are generated from endogenous hairpin structures in the nucleus and play an important role in controlling diverse cellular functions in eukaryotes, including cell differentiation, development, apoptosis, and genome integrity [7-9]. *In vivo* experiments indicate a crucial role in cell proliferation and cell death processes for some miRNAs, including *lin-4* and *let-7* in *C. elegans*; *bantam* and *mir-14* in *Drosophila*; and *mir-23* in humans [13].

According to our current understanding, biogenesis of miRNA involves a series of coordinated processes. Briefly, primary transcripts of miRNAs are processed in the nucleus by Drosha (an RNase III-like enzyme) into pre-miRNA, which are first exported into the cytoplasm by exportin-5 and then processed into miRNAs by Dicer, another type III RNase [10-12].

The primary method of identifying miRNA genes has been to isolate, reverse transcribe, clone, and sequence small cellular RNAs [15-16]. However, molecular cloning is biased toward finding miRNAs that are relatively abundant. In animals, miRNA gene discovery by molecular cloning has been supplemented by systematic computational approaches that identify evolutionarily conserved miRNA genes by searching for patterns of sequence and secondary structure conservation that are characteristic of metazoan miRNA hairpin precursors [17-19]. The 5' extremity of miRNA has been reported to have a perfect base alignment of at least 7 consecutive nucleotides [14], which enables its identification. The most sensitive of these methods indicate that miRNAs constitute nearly 1% of all predicted genes in nematodes, flies, and mammals [19-21].

Although the first miRNA was identified in 1993 [22], it was not until 2001 that the breadth of the miRNA gene class was recognized with cloning and sequencing of more

than one hundred miRNAs from worms, humans, mice, and other species [23]. Nevertheless, up to now, such information has not been obtained from a large-scale identification of miRNAs in *Schistosoma mansoni*.

S. mansoni infects approximately two hundred million people worldwide, causing an important parasitic disease mostly in tropical regions, and is endemic in 74 developing nations [24]. Although schistosomicidal drugs and other control measures exist, the development of new strategies is of special interest. In recent years increasing attention has emerged over siRNAs as therapeutical agents [25-26]. The emergence of gene ablation technologies based on the RNAi phenomenon has opened up new experimental opportunities. However, there are only a few reports applying siRNAs against parasite infections [26-29], and none with miRNAs in schistosomes. In this context, we attempted to identify potential miRNAs in *S. mansoni*, and verify experimentally the expression levels and patterns of some of them. We further provide evidence for 211 potential miRNAs candidates in *S. mansoni* by sequencing small-RNA cDNA libraries. The identification of miRNA in the parasite genome presents another factor likely to be involved in parasite development and sexual maturation.

Results and Discussion

Cloning of short RNAs from *S. mansoni* adult worm pairs

One of the goals of this study was to identify and characterize miRNAs from the *S. mansoni* parasite. An adult worm cDNA library of small RNAs was constructed using an established method based on a sequential ligation of oligonucleotide adapters to a size-fractionated sample of small RNAs. Concatenated DNA fragments (each from one putative miRNA) were cloned into the plasmid 2.1 TOPO TA vector (Invitrogen, Carlsbad, CA) to generate a library. A total of 582 recombinant clones randomly selected from the library were sequenced. Twelve hundred sequences were analyzed, which indicated that each concatemer contained ~2-3 small RNA sequences. Size distribution of the non-redundant miRNA set ranged from 17 to 25 nt, although the majority contained 20-24 nt, and 21 nt being the most abundant. To identify the putative origin of the cloned sequences, a FASTA search [30] was performed against GenBank and The Wellcome Sanger Institute databases

containing the *S. mansoni* genome, removing the sequences that had significant homology to breakdown products of abundant non-coding (nc) RNA such as rRNA, tRNA, and other abundant RNA species. A total of 584 ncRNAs were grouped into 211 families, and were identified as candidate miRNAs (see Additional file 1). One hundred and sixty-one miRNAs were represented in the library by only one read and 50 were represented by multiple clones. The number of sequences in each cluster was between 1 to 32. Since miRNAs are believed to occur at a frequency of approximately 0.5-1.5% of the total genes in the genome [31], the 13,200 genes predicted for *S. mansoni* [32] should have generated 100 of the 300 miRNAs. Thus, the number of miRNAs from this cloning identification is within the expected number.

We further screened the candidate sequences against a database of known miRNAs, miRBase (<http://microrna.sanger.ac.uk>; release 13.0) to compare our candidate *S. mansoni* miRNAs to miRNAs from different species. Forty-one sequences had at least one match with mature miRNAs from different metazoan miRNA families, i.e. miR-832, miR-71, miR-297, and let-7 (Table 1). Sma-miR-36 perfectly matched miRNA family miR-87 from different species demonstrating miRNA phylogenetic conservation among more than 10 species. Previous studies in *C. elegans* showed that this miRNA family is expressed throughout development [20]. The yield of probable miRNA candidates was much lower for this analysis with *S. mansoni* than analyses of species that contain closer relatives in miRBase. The closest relative to *S. mansoni* in miRBase is *S. mediterranea*. These two organisms belong to the same phylum, a relatively broad classification. The sequences that did not match any of the known miRNAs (170 sequences) were considered to be putative members of novel families of schistosome miRNAs.

Expression analysis of miRNAs in *S. mansoni*

The expression of miRNAs is tightly regulated both in time and space. Stage-specific miRNAs or a high level of expression of miRNAs only in certain stages implies that these miRNAs play critical roles in the developmental stages where they are expressed [20-21]. While newer techniques, such as microarray methods with higher throughputs are emerging, northern blot still remains the consensus method for validating miRNAs [33]. The frequency of reads of a specific miRNA in a non normalized library can correlate with the expression level of that miRNA [34-35]. Based on that, the most abundant candidate miRNAs were further examined by northern blot to test for expression and confirm their

size. Northern blots of total RNA from a mixture of male and female *S. mansoni* adult worms were hybridized with biotin-labeled probes. In addition, we also tested the larval stage that exists in the mammalian host, schistosomula. Previous studies analyzed the expression pattern of the Dicer gene during different life stages of *S. mansoni* and showed the highest relative expression is detected in 7-d schistosomula which showed over 3 fold the level of expression as the adult worm pairs [36]. It is possible that higher Dicer gene expression at this time has been selected for the control of retrotransposon activation that may be more prone to occur during this period of active larval cell division and growth [37]. We detected the expression for 11 of 30 miRNAs in at least one of the 2 analyzed stages (Figure 1). We also analyzed the expression of the five novel miRNAs recently identified in *Schistosoma japonicum* [38]. Three of the five probes gave a hybridization signal that was characteristic of miRNAs, demonstrating evolutionary conservation. Although the expression of sja-miR-71 and sja-bantam dropped quickly in *S. japonicum* lung-stage schistosomulum [38], we observed the hybridization signal for both miRNAs in *S. mansoni* (Figure 1). The remaining 2 candidates (new-1 and let-7) may be expressed in other life cycle stages or in undetectable amounts in *S. mansoni*. Based upon the two-stage expression profiles, these new miRNAs have been grouped into 3 categories: nonspecific (2 miRNAs), *S. mansoni* adult worm specific (7 miRNAs), and schistosomula specific (5 miRNAs). These results suggest a very important role for these miRNAs over the life cycle stages of *S. mansoni* mediating important processes in the parasite growth and development.

Our sequence analysis in this study indicated that miR-1 is most abundant (32 reads). Although sequencing-based miRNAs expression profiling is a tool for measuring the relative abundance of miRNAs, the expression of the miR-1 was not detected by northern blot. This result suggests that miR-1 might be a cloning artifact or that small RNA blot analysis does not always reflect cloning frequency. For instance, miR-32 is represented only by 2 clones in our sequences, which indicates a 12-fold lower expression level compared to that of miR-3 (24 reads). However, our small RNA blot analysis indicated that miR-32 was much more abundantly expressed than miR-3 in the adult worm stage (Figure 1). The discrepancies between the cloning frequency and small RNA blot results for miR-3 and miR-32 could not be attributed to variations in RNA content because the same RNA

samples were used for both experiments and could be due to bias in cloning efficiencies. Another possible explanation is the differential turnover rates of these miRNAs [39]. The best method to differentiate miRNA from other endogenous small RNA is the ability of flanking sequences to adopt a pre-miRNA fold-back structure with the mature miRNA properly positioned within one of its strands enabling Dicer processing [17]. We analyzed the structure of the potential miRNAs detected by northern blot mapped to different genomic regions, leading to a total of ~500 distinct genomic locations for each sequence. To assess which of the regions corresponded to the real location of the possible miRNA gene, their secondary structures were studied using Vienna RNAfold (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>), and each image visually inspected. A non-redundant set of 26 potential miRNA sequences were predicted to be capable of forming stem-loop structures characteristic of miRNA precursors (Table 1; see Additional file 2), 11 of them were also confirmed by northern blot (Figure 2). Our results also show that multiple hairpin precursors for the same miRNA were observed in more than one location in the parasite genome (data not shown), indicating that the same mature miRNA may be transcribed from more than one miRNA gene. Next, the miRNA genomic location was analyzed using the program BLAST against the *S. mansoni* genome (http://www.sanger.ac.uk/Projects/S_mansoni/). The selected miRNAs genes are located on intergenic regions, in agreement with published results [40-42] (Table 2).

Conclusions

The discovery of small regulatory RNA molecules, miRNAs, is undoubtedly one of the most important recent findings in biological research. This study demonstrates for the first time the presence of miRNAs in *S. mansoni* identified by cloning and sequencing of 1200 sequences from a small RNA library. Further bioinformatics analysis identified 211 potential miRNA candidates, of which 26 were predicted to form stem-loop structures characteristic of miRNA precursors. The expression of 14 of them was confirmed by northern blot analysis. miRNAs likely play a crucial role in parasite development and also in genome dynamics related to mobile elements via RNA silencing mechanisms.

Methods

Parasites and nucleic acid extraction

Total RNA was extracted from adult worm pairs and lung-stage schistosomula of *S. mansoni* with use of Trizol® (Invitrogen). Cercariae were obtained from infected *Biomphalaria glabrata* and isolated parasite bodies were prepared as previously described [43]. Parasites were cultured for 7 days in complete RPMI medium supplemented with 10 mM HEPES, 2 mM glutamate, 5% fetal calf serum and antibiotics (100 U/ml penicillin and 100 µg/ml streptomycin) at 37°C, in a 5% CO₂ atmosphere.

RNA isolation and miRNA cloning

A 200 µg aliquot of total RNA isolated from adult worms by guanidine thiocyanate phenol-chloroform extraction was pooled [44]. The short RNA fraction with a size from 17 to 26 nt was purified and cloned as described in Chappell *et al.* [45]. The concentration was quantified using NanoDrop Spectrophotometer (NanoDrop Technologies, USA). Total RNA (1 mg) was resolved by electrophoresis on denaturing 15% polyacrylamide gel (8 M urea, 1 x TBE buffer), and short RNAs (17 to 26 nucleotides in length) were excised and eluted in 3M NaCl solution at 4°C for 16 h. The gel purified small RNAs were dephosphorylated and ligated directly to a 5'-phosphorylated 3'-adapter oligonucleotide with a blocked 3'-hydroxyl terminus to prevent self-ligation (Dharmacon Research, Boulder, CO). The ligation products were separated from the excess of 3'-adapter (Dharmacon Research, Boulder, CO) on a 15% denaturing polyacrylamide gel and were subsequently ligated to a non-phosphorylated 5'-adapter oligonucleotide using T4 RNA ligase (Invitrogen). The final products were again gel purified by size fractionation and submitted to reverse transcription and PCR amplification. The PCR products were digested with *Pst*I and *Nco*I and subsequently concatenated using T4 DNA ligase. The ends of the concatamers were filled in with Klenow/AT-tailing and ligated into a 2.1 TOPO TA vector (Invitrogen, Carlsbad, CA). Ligated plasmids were transformed into TOP10 cells (Invitrogen). The libraries were plated on Luria-Bertani (LB) amp plates and individual colonies were picked and put into 96-well plates containing 75 µl LB[Amp] and grown overnight at 37°C with continuous shaking. The recombinant clones were selected, sequenced and the data was analyzed as described below.

Computational analysis

Base calling and quality trimming of sequence chromatograms were conducted by PHRED software [46]. After masking of vector and adapter sequences using EMBOSS-restrict (<http://bioweb2.pasteur.fr/docs/EMBOSS/restrict.html>), small RNA sequences ranging between 17-25 nt in length obtained from the libraries were multi-aligned by ClustalW2 program [47] to exclude redundant sequences. The distinct sequences were used in BLAST searches against the *S. mansoni* genome and miRBase database (<http://microrna.sanger.ac.uk>) to identify sequences from other species that closely match candidate *S. mansoni* miRNAs, and removal of contaminating mRNAs, tRNAs, rRNAs, and other small RNAs. To predict the secondary structure of the remaining small RNA, Perl scripts were implemented to align the sequences to the genome of the parasite *S. mansoni* (http://www.sanger.ac.uk/Projects/S_mansoni/) aiming at retrieving all possible genomic locations. In brief, the script executes BLAST [48] to perform the sequence similarity analysis and the result is parsed to retrieve the genomic positions to which each miRNA aligns. In the next step, the script builds a FASTA file containing two sets of approximately 500 entries for each miRNA: one set of genomic sequences plus 40, 50, 60 and 70 nucleotides upstream and another with 40, 50, 60 and 70 nucleotides downstream. After the FASTA file creation for each miRNA, the secondary structures were predicted using RNAfold from the Vienna RNA package [49]. Each image was further visually inspected to confirm the presence of a typical stem-loop conformation of pre-miRNAs. Among all structures created for each sequence, the one containing the mature miRNA in one arm of the hairpin precursor and with lower folding free energy was selected. The images were created using the VARNA (<http://varna.lri.fr/index.html>) to insert subtitles and highlight the mature miRNA sequence.

miRNA expression analysis

For the northern blot analysis, total RNA from adult worm pairs and 7-d schistosomula were extracted by guanidinium thiocyanate–phenol–chloroform method from frozen tissues (snap freezing in liquid nitrogen and storage at -80°C). Sixty micrograms of total RNA were separated on 15% denaturing polyacrylamide gels, and then were electrotransferred to Hybond N+ membranes (GE Healthcare) in 1x Tris Borate EDTA using the Mini Trans-Blot Cell apparatus (Bio-Rad) according to the manufacturer's instructions. Membranes

were UV cross-linked in *UV Stratalinker*[®] (Stratagene) and pre-hybridized in DIG Easy Hyb solution (Roche) at 37°C for 30 min. DNA oligonucleotides complementary to the miRNA sequences were labeled with DIG Oligo 3'-End Lab. Kit, 2nd generation (Roche). Hybridization was performed overnight at 37°C with 3' digoxigenin-labeled RNA probes at 4.5 pmol/ul. Probe detection was performed using the DIG Wash and Block Buffer Set (Roche). In brief, blots were incubated in blocking solution for 1 hour and then in antibody solution (anti-DIG, alkaline phosphatase conjugated antibody) for 30 min, followed by washing twice in washing buffer. After equilibration in detection buffer, blots were incubated with chemiluminescent substrate CSPD (Roche). Membranes were exposed to X-ray film for 20 minutes and the films were digitized using a transmission scanner GS-800 Calibrated Densitometer (Bio-Rad).

Author's contributions

MS performed all experiments and wrote the paper. AD directed the miRNA library construction and MS, GC and AZ carried out the prediction and computational analysis of miRNAs. RP contributed to the northern blot experiments. PLV, GO and NES designed and directed the project. All authors read and approved the final manuscript.

Acknowledgements

This work was funded by NIH-NIAID Grant 5D43TW007012-04, NIH Training Grant D43TW006580 (PTL), FAPEMIG (CBB-1181/0 and 5323-4.01/07) Capes and Fiocruz. The parasites were kindly provided by Fred Lewis (Biomedical Research Institute through NIH-NIAID, MD) and Laboratory of Mollusks, Centro de Pesquisas René Rachou-Fiocruz, Belo Horizonte, Brazil.

References

1. Kim VN, Nam JW: **Genomics of microRNA**. *Trends Genet* 2006, **22**:165-173.
2. Huttenhofer A, Vogel J: **Experimental approaches to identify non-coding RNAs**. *Nucleic Acids Res*. **2006**, **34**(2):635-646.
3. He L, Hannon GJ: **MicroRNAs: small RNAs with a big role in gene regulation**. *Nat Rev Genet* 2004, **5**(7):522-31.

4. Lagos-Quintana M, Rauhut R, Meyer J, Borkhardt A, Tuschl T: **New microRNAs from mouse and human.** *RNA*, 2003, **9**: 175-179.
5. Wiznerowicz M, Szulc J, Trono D: **Tuning silence: conditional systems for RNA interference.** *Nature Methods* 2006, **3**:682 – 688.
6. Valencia-Sanchez MA, Liu J, Hannon GJ, Parker R: **Control of translation and mRNA degradation by miRNAs and siRNAs.** *Genes Dev.* 2006, **20**(5):515-524.
7. Lau NC, Lai EC: **Diverse roles for RNA in gene regulation.** *Genome Biol.* **2005**, **6**(4):315.
8. Ambros V: **The functions of animal's microRNAs.** *Nature* 2004, **431**:350-355.
9. Miska A: **How microRNAs control cell division, differentiation and death.** *Curr. Opin. Genet. Dev* 2005, **15**:563–568.
10. Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S, Kim VN: **The nuclear RNase III Drosha initiates microRNA processing.** *Nature* 2003, **425**:415–419.
11. Winter J, Jung S, Keller S, Gregory RI, Diederichs S: **Many roads to maturity: microRNA biogenesis pathways and their regulation.** *Nat Cell Biol.* 2009, **11**:228-34.
12. Lund E, Guttinger S, Calado A, Dálberg JE, Kutay U: **Nuclear export of microRNA precursors.** *Science* 2003, **303**:95–98.
13. Bushati N, Cohen SM: **microRNA functions.** *Annual Review of Cell and Developmental Biology* 2007, **23**(I): 175-205.
14. Lewis BP, Shi I, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** *Cell* 2003, **115**:787-798.
15. Lee RC, Ambros V, **An extensive class of small RNAs in *Caenorhabditis elegans*.** *Science* 2001, **294**:862-864.
16. Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T: **Identification of tissue-specific microRNAs from mouse.** *Curr Biol.* 2002, **12**:735-739.
17. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, Dreyfuss G, Eddy SR, Griffiths-Jones S, Marshall M, et al: **A uniform system for microRNA annotation.** *RNA* 2003, **9**:277-279.

18. Kloosterman W, Steiner F, Berezikov E, Bruijn, et al: **Cloning and expression of new microRNAs from zebrafish.** *Nucleic Acids Res* 2006, **34**: 2558-2569.
19. Lai EC, Tomancak P, Williams RW, Rubin GM: **Computational identification of *Drosophila* microRNA genes.** *Genome Biol* 2003, **4(7)**: R42.
20. Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP: **The microRNAs of *Caenorhabditis elegans*.** *Genes Dev.* 2003, **17(8)**:991-1008.
21. Lim LP, Glasner ME, Yekta S, Burge CB, Bartel DP: **Vertebrate microRNAs genes.** *Science* 2003, **299**:1540.
22. Lee, R.C., Feinbaum, R.L., Ambros, V: **The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*.** *Cell* 1993, **75**: 843–854.
23. Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, et al: **The expansion of the metazoan microRNA repertoire.** *BMC genomics* 2006, **7**:25.
24. Savioli L, Stansfield S, Bundy DA, Mitchell A, Bhatia R, Engels D, Montresor A, Neira M, Shein AM: **Schistosomiasis and soil-transmitted helminth infections: forging control efforts.** *Trans R Soc Trop Med Hyg* 2002, **96(6)**:577-579.
25. Bumcrot D, Manoharan M, Koteliansky V, Sah DW: **RNAi therapeutics: a potential new class of pharmaceutical drugs.** *Nat Chem Biol.* 2006, **2(12)**:711-719.
26. Pereira TC, Pascoal VD, Marchesini RB, Maia IG, Magalhães LA, Zanotti-Magalhães EM, Lopes-Cendes I. ***Schistosoma mansoni*: evaluation of an RNAi-based treatment targeting HGPRTase gene.** *Exp Parasitol* 2008, **118(4)**:619-623.
27. Mohammed A, Dasaradhi PV, Bhatnagar RK, Chauhan VS, Malhotra P: **In vivo gene silencing in *Plasmodium berghei*--a mouse malaria model.** *Biochem Biophys Res Commun.* 2003, **309(3)**:506-511.

-
28. Cheng G, Fu Z, Lin J, Shi Y, Zhou Y, Jin Y, Cai Y: **In vitro and in vivo evaluation of small interference RNA-mediated gynaecophoral canal protein silencing in *Schistosoma japonicum*.** *J Gene Med* 2009, **11**(5):412-421.
 29. Chandramohanadas R, Davis PH, Beiting DP, Harbut MB, Darling C, Velmourougane G, Lee MY, Greer PA, Roos DS, Greenbaum DC: **Apicomplexan Parasites Co-Opt Host Calpains to Facilitate Their Escape from Infected Cells.** *Science*, 2009.
 30. Pearson WR: **Flexible sequence similarity searching with the FASTA3 program package.** *Methods Mol. Biol* 2000, **132**:185-219.
 31. Carthew RW: **Molecular biology. A new RNA dimension to genome control.** *Science* 2006, **313**(5785):305-306.
 32. Haas BJ, Berriman M, Hirai H, Cerqueira GG, Loverde PT, El-Sayed NM. ***Schistosoma mansoni* genome: closing in on a final gene set.** *Exp Parasitol* 2007, **117**(3):225-228.
 33. Chaudhuri K, Chatterjee R: **MicroRNA detection and target prediction: integration of computational and experimental approaches.** *DNA Cell Biol* 2007, **26**(5):321-337.
 34. Landgraf P, et al. **A mammalian microRNA expression atlas based on small RNA library sequencing.** *Cell* 2007, **129**(7):1401-1414.
 35. Fahlgren N, Howell MD, Kasschau KD, Chapman EJ, Sullivan CM, Cumbie JS, Givan SA, Law TF, Grant SR, Dangl JL, Carrington JC: **High-throughput sequencing of *Arabidopsis* microRNAs: evidence for frequent birth and death of miRNA genes.** *PLoS ONE* 2007, **2**(2):e219.
 36. Krautz-Peterson G, Skelly PJ: ***Schistosoma mansoni*: the dicer gene and its expression.** *Exp Parasitol* 2008, **118**:122-128.
 37. Krautz-Peterson G, Radwanska M, Ndegwa D, Shoemaker CB, Skelly PJ: **Optimizing gene suppression in schistosomes using RNA interference.** *Mol Biochem Parasitol* 2007, **153**(2):194-202.

-
38. Xue X, Sun J, Zhang Q, Wang Z, Huang Y, Pan W: **Identification and characterization of novel microRNAs from *Schistosoma japonicum***. *PLoS ONE* 2008, **3(12)**:e4034.
39. Ding X, Weiller J, Großhans H: **Regulating the regulators: mechanisms controlling the maturation of microRNAs**. *Trends in Biotech* 2009, **27**:27-36.
40. Grishok A, Pasquinelli AE, Conte D, et al: **Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing**. *Cell* 2001, **106**:23-34.
41. Ro S, Song R, Park C, Zheng H, Sanders KM, Yan W: **Cloning and expression profiling of small RNAs expressed in the mouse ovary**. *RNA* 2007, **13**:2366-2380.
42. Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**:281-297.
43. Skelly PJ, Da'dara A, Harn DA: **Suppression of cathepsin B expression in *Schistosoma mansoni* by RNA interference**. *Int J Parasitol* 2003, **33(4)**:363-369.
44. Chomczynski, P, Sacchi, N: **Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction**. *Anal Biochem*, **162(1)**:156-159.
45. Chappell L, Baulcombe D, Molnar A: **Isolation and cloning of small RNAs from virus-infected plants**. In *Current Protocols in Microbiology* 2005, **1**: Jon Wiley & Sons Inc. in press.
46. Ewing B, Hillier L, Wendl MC, Green P: **Base-calling of automated sequencer traces using phred. I. Accuracy assessment**. *Genome Res* 1998, **8(3)**:175-185.
47. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Res* 1994, **22**:4673-4680.

48. Altschul et al., 1990 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucl Acids Res* 1997, **25**:3389–3402.
49. Hofacker IL: **Vienna RNA secondary structure server.** *Nuc. Acids Res* 2003, **31**:3429-3431.

Figures

Figure 1 - Northern blot analysis of selected miRNAs across two different developmental stages in *S. mansoni*.

Sixty micrograms of total RNA for each sample were separated on a 15% denaturing polyacrilamide gel, blotted and probed using DIG-labeled probes. *Lanes from left to right: S. mansoni* adult worm pairs (AW) and schistosomula (S) (7 days after mechanical transformation). miR-71, miR-125, and Bantam are the miRNAs identified in *S. mansoni* homolog to miRNAs of *S. japonicum*. The tRNA and 5S rRNA bands were visualized by ethidium bromide staining of polyacrylamide gels and served as loading controls and are shown at the bottom.

Figure 2 - Pre-miRNA secondary structure of selected novel *S. mansoni* miRNAs identified in small RNA libraries from adult worms.

S. mansoni genomic sequences upstream and downstream of the novel miRNAs were predicted using the computer program RNAfold from ViennaRNA package. Dark blue text represents the mature miRNA sequence. These miRNAs also had their expression confirmed by northern blot.

Tables

Table 1 – Small *S. mansoni* RNAs with matches to known microRNAs present in the miRBase.

Table 2 – Most frequently sequenced miRNAs in the *S. mansoni* small RNA libraries and their locations within the published chromosomal sequences.

Additional files

Additional file 1 – Distinct sequences identified by cloning from *Schistosoma mansoni* adult worm

File format: .xls

Description: 211 families of *S. mansoni* miRNA candidates identified by cloning and sequencing of small RNA library. The file is available at <http://www.cebio.org/content/2009/04/08/schistosoma-mansoni-micromnas>.

Additional file 2 – Predicted precursor structures of newly *S. mansoni* miRNAs

File format: .doc

Description: The predicted structures are separated into two groups: miRNA detected by northern blot (11 miRNAs), and miRNA undetected by northern blot (15 miRNAs). The RNA secondary structure of the precursors was predicted using using the software RNAfold from Vienna RNA package (<http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi>). The file is available at <http://www.cebio.org/content/2009/04/08/schistosoma-mansoni-micromnas>.

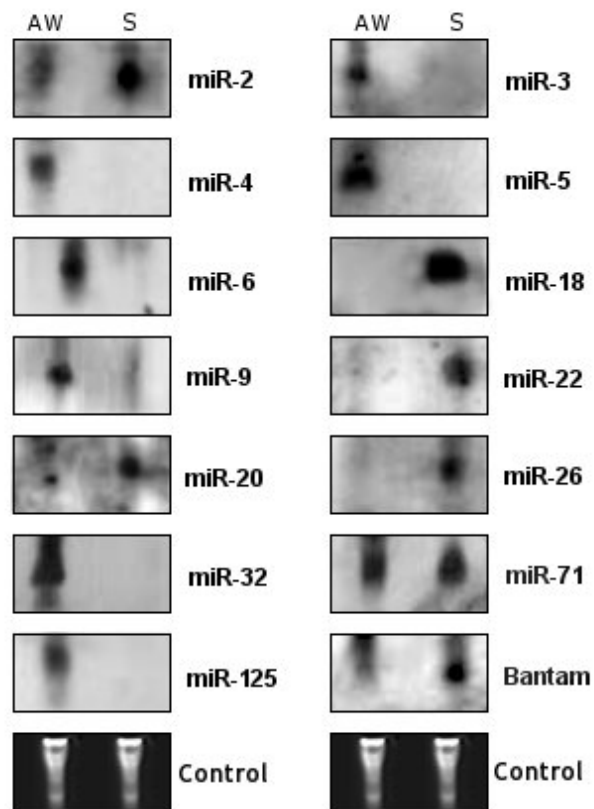


Figure 1

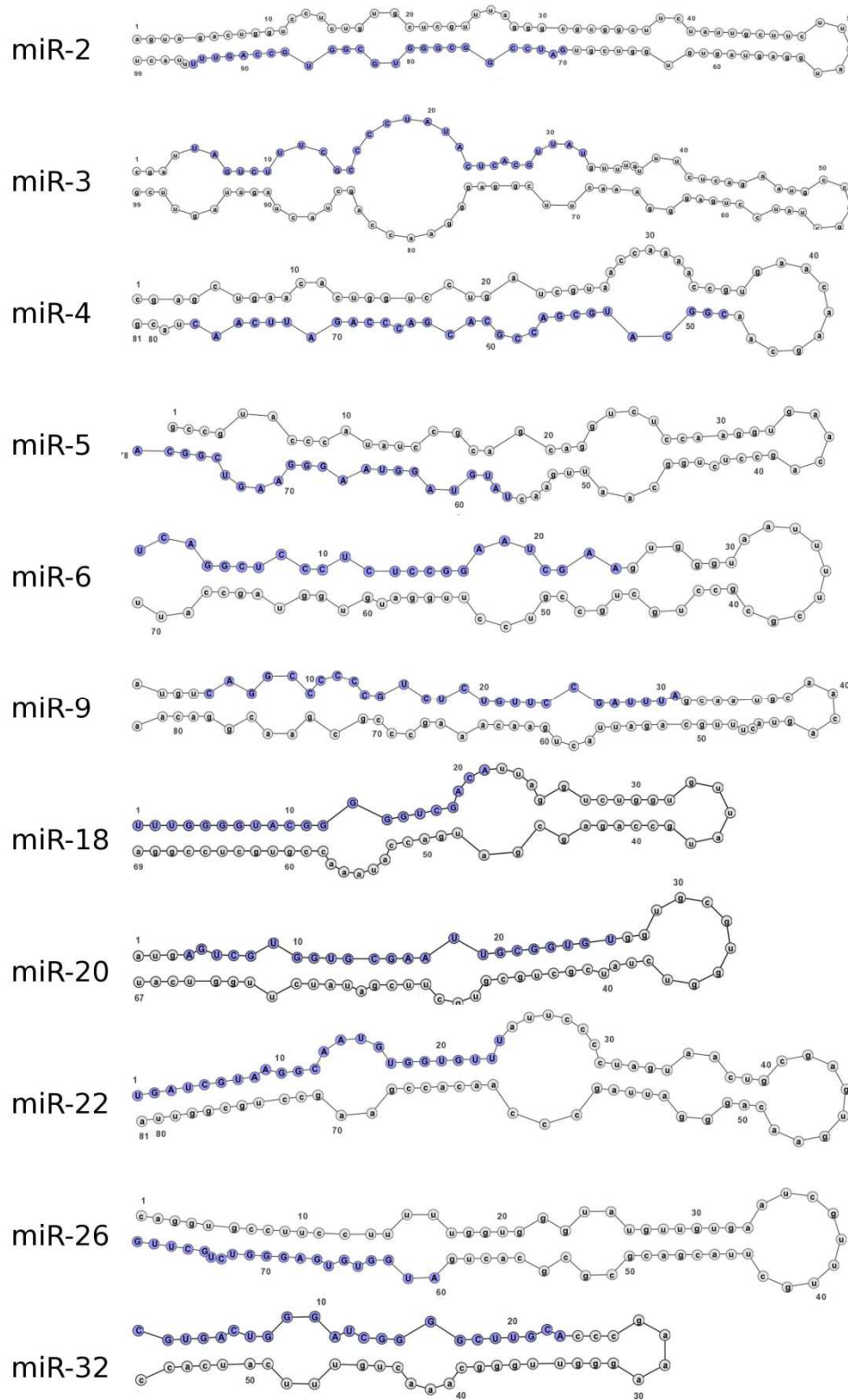


Figure 2

<i>S. mansoni</i> miRNAs	Number of Hits	Best Hit Name	miRNA family
sma-miR-7	4	gma-miR171b	miR-171
sma-miR-9	2	ptr-miR-1303	miR-1303
sma-miR-16	1	gga-miR-1465	miR-1465
sma-miR-17	10	dre-miR-739	miR-197
sma-miR-21	1	ebv-miR-BART2	miR-BART2
sma-miR-24	10	ptr-miR-451	miR-451
sma-miR-31	10	hsa-miR-513b	miR-513
sma-miR-32	1	ath-miR832-5p	miR-832
sma-miR-33	3	tca-miR-71	miR-71
sma-miR-36	10	tca-miR-87	miR-87
sma-miR-37	1	cbr-miR-240	miR-240
sma-miR-40	4	ptr-miR-432	miR-432
sma-miR-47	1	sme-let-7c	let-7
sma-miR-77	1	ptc-miR472b	miR472
sma-miR-78	1	ath-miR832-5p	miR-832
sma-miR-86	1	hsa-miR-1268	miR-1268
sma-miR-96	10	ptr-let-7b	let-7
sma-miR-103	1	kshv-miR-K12-3	miR-K12
sma-miR-105	8	dya-miR-289	miR-289
sma-miR-122	1	gga-miR-1465	miR-1465
sma-miR-129	8	dya-miR-289	miR-289
sma-miR-141	2	osa-miR166i	miR-166
sma-miR-149	1	gga-miR-1810	miR-1810
sma-miR-150	1	mmu-miR-297b-3p	miR-297
sma-miR-156	3	tni-miR-101b	miR-101
sma-miR-170	1	cel-miR-78	miR-78
sma-miR-182	10	dre-miR-739	miR-739
sma-miR-187	1	mmu-miR-297b-3p	miR-297
sma-miR-205	1	ppt-miR896	miR-896
sma-miR-209	1	mmu-miR-297b-3p	miR-297
sma-miR-210	1	gga-miR-1810	miR-1810
sma-miR-213	1	ath-miR832-5p	miR-832
sma-miR-216	1	kshv-miR-K12-3	miR-K12
sma-miR-226	1	mmu-miR-297b-3p	miR-297
sma-miR-234	1	gga-miR-1810	miR-1810
sma-miR-241	1	odi-miR-1500	miR-1500
sma-miR-247	4	ptr-miR-432	miR-432
sma-miR-262	1	sme-let-7c	let-7
sma-miR-276	1	mmu-miR-297b-3p	miR-297
sma-miR-278	1	gga-miR-1810	miR-1810
sma-miR-283	2	ptr-miR-1299	miR-1299
sma-miR-284	1	mmu-miR-297b-3p	miR-297

Table 1

miRNA	Sequence (5' – 3')	Size (nt)	<i>S. mansoni</i> shotgun read (Sanger)	Northern blot validation
sma-miR-2	TGATCCGGCGGGTGCGGTGCCAGTT	25	Smp_scaff013386:1052..1077	+
sma-miR-3	ACGTGAGTATAGGGGCGAAAGACTA	25	Smp_scaff015051:1455..1482	+
sma-miR-4	TTGAATCTGGGTCGTGCGGTCGCAT	25	Smp_scaff015219:199..228	+
sma-miR-5	AATGTAGGTAAGGGAAGTCGGCA	23	Smp_scaff015642:1137..1159	+
sma-miR-6	TTCGATTCCGGAGAGGGAGCCTG	23	Smp_scaff015219:463..486	+
sma-miR-7	TTGATTCGGTGGGTGGTGGTGC	22	Smp_scaff015609:101..123	-
sma-miR-9	TAAATCGGAACAGAGACGGGGTGC	24	Smp_scaff007962:435..461	+
sma-miR-10	TAAACCAGGGTTCGGGAGTTC	21	Smp_scaff003291:1015..1035	-
sma-miR-11	AGGTCGTTGGATTGGCCTCGTTGT	24	Smp_scaff015609:591..615	-
sma-miR-13	TACATCCATCATGTCTCCTCACAA	24	Smp_scaff15230:1826..1852	-
sma-miR-15	CGGGTGTCCGGAGCGGACGGC	21	Smp_scaff003242:75..98	-
sma-miR-16	ATAGGTGGGAGCTGGGTGCTTC	22	Smp_scaff015642:1783..1804	-
sma-miR-17	CGAGGTGAGAAAGGGTTCCA	20	Smp_scaff014853:236..255	-
sma-miR-18	TTGGGGTACGGGGTTCGAC	20	Smp_scaff015270:461..481	+
sma-miR-19	TGGTTGGGTCGGTAGGTGG	19	Smp_scaff017601:2493..2511	-
sma-miR-20	AGTCGTGGTGCGAATTGCGGTGT	23	Smp_scaff015642:1952..1974	+
sma-miR-21	GTTGGTGGAGCGATTTGTCTG	21	Smp_scaff015122:970..990	-
sma-miR-22	TGATCGTAAGGCAATGTGGTGTTT	24	Smp_scaff013172:1207..1230	+
sma-miR-23	GCGGTGTGTCGGAGATGGCGGC	22	Smp_scaff015318:1301..1323	-
sma-miR-24	ATGGTAACCGGGTGTCTATG	22	Smp_scaff013300:510..531	-
sma-miR-25	ACGTCGGCGAGAGTTCTGGGA	21	Smp_scaff006489:1115..1135	-
sma-miR-26	TTGGTGTGAGGGTCTGCTTG	20	Smp_scaff015318:1119..1138	+
sma-miR-27	TTGAATCTGGGTCGTGCG	18	Smp_scaff015219:210..227	-
sma-miR-29	AAGTCGTAACAAGGTTTCCGTAGGT	25	Smp_scaff013172:104..128	-
sma-miR-31	CGGAGTTGAAGCAAGAGGTGCA	24	Smp_scaff015642:2198..2221	-
sma-miR-32	CGTACTGGGATCGGGGCTTGCA	23	Smp_scaff003570:35..57	+

Table 2

Identification of novel miRNA genes in the *Schistosoma mansoni* genome

Jonathan Lee^{1*}, Mariana Simões^{2,3*}, Guilherme Oliveira^{3§}, Andrew Dalby¹

¹Department of Statistics, Peter Medawar Building, South Parks Road, Oxford, UK

²Graduate Program in Bioinformatics, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

³Laboratory of Cellular and Molecular Parasitology and CEBio, Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, MG, Brazil

*These authors contributed equally to this work

§Corresponding author

Email addresses:

JL: jonlee@gmail.com

MS: msimoes@cpqrr.fiocruz.br

GO: oliveira@cpqrr.fiocruz.br

AD: dalby@stats.ox.ac.uk

Abstract

Background

MicroRNAs (miRNAs) are a recently discovered class of single-stranded RNAs involved in the regulation of gene expression. Computational and experimental techniques have identified over 6,000 miRNAs to date, and have demonstrated high conservation of miRNAs across species. In *Schistosoma*, the full set of miRNAs and their expression patterns during development are still poorly understood when compared to the other species. In this manuscript we describe the development and implementation of a homology-based detection strategy to search for miRNA genes in the parasite, *Schistosoma mansoni*.

Results

Strategies adapted from existing literature were applied to the detection of novel miRNAs in *Schistosoma mansoni*. The homology search for the high-throughput pipeline development was performed with all known miRNAs in miRBase. In total, 6,211 mature miRNAs were used as reference sequences. As a result, 110 unique *S. mansoni* sequences were returned from the BLASTn search followed by filtering steps. The existing mature miRNAs that produced these hits are reported, as well as the locations of the homologous sequences in the *S. mansoni* genome. All hits aligned with at least 95% of the miRNA sequence, resulting in alignment lengths of 19-24 nt. Fifteen potential miRNA candidates were detected with this approach.

Conclusions

Evidence for the presence of miRNAs in *S. mansoni* has been presented. Comparison of the number of miRNAs detected in this analysis to other work suggests that homology-based computational techniques in *S. mansoni* may be limited in their detection due to the lack of close relatives in the miRNA repository. The discovery pipeline described here can likely be applied to the identification of pre-miRNA hairpins in other organisms.

Background

MicroRNAs (miRNAs) are a recently discovered class of single-stranded RNAs involved in the regulation of gene expression. They are short in length (typically ~22nt), and belong to a group of non-protein coding RNAs that also includes short-interfering RNAs (siRNAs) [1]. Although the functions of most miRNAs are still unknown, the small amounts of experimental data currently available suggest that they play important roles in areas such as embryonic stem cell development, neuronal differentiation and cancer development [1].

To date, over 6,000 miRNAs have been discovered in organisms ranging from plants to worms and mammals. In humans alone, 678 miRNA genes were identified [2]. These genes were discovered through a combination of molecular cloning techniques and computational approaches.

Most known miRNA genes were discovered experimentally, and all computationally-predicted miRNAs must be verified experimentally [3]. The most common approach for experimental detection is through cloning of miRNAs and pre-miRNAs from size-fractionated RNA samples, then confirmation of the candidates by northern blot analysis [4]. While other techniques, such as microarray methods, with higher throughputs are emerging, northern blot still remains the consensus method for validating miRNAs [5].

However, cloning and northern blot techniques have two main drawbacks. First, these techniques are low throughput and are resource intensive. Secondly, since they are based on the amount of RNA present, they have limited ability to detect rare miRNAs [5]. As such, computational prediction of miRNA genes provides a useful supplement to experimental techniques because of the speed at which miRNA candidates can be evaluated and the absence of a bias toward abundant miRNAs.

In this study, we report the development and implementation of a homology-based detection strategy to search for miRNA genes in the parasite, *Schistosoma mansoni*. *S. mansoni* is a parasitic flatworm and one of the most important species that cause schistosomiasis [6]. Schistosomiasis is a debilitating disease affecting the liver and gastrointestinal tract, and is typically found in Southeast Asia, Africa and South America. *S. mansoni* is endemic in Africa and South America, and through its life cycle inhabits two hosts: *Biomphalaria* snails, where the parasite reproduce asexually, and humans, where the parasite lives in the blood stream and reproduces sexually. An estimated 200 million people have schistosomiasis, with over 600 million at risk due to their proximity to

transmission areas. Schistosomiasis is considered to rank second to malaria with respect to global health and economic impact of endemic countries [6]. Recently, five novel miRNAs were identified in *Schistosoma japonicum*, a close relative of *S. mansoni* [7]

Our strategy is based on the observations that miRNAs are highly conserved across species. However, considerable filtering must be performed to elucidate likely miRNA candidates. The homology-based search is a powerful approach to miRNA detection due to the high conservation of miRNAs across different species [8]. This method involves the use of a sequence alignment tool, such as BLASTN, to search the database of the organism of interest against known and experimentally verified miRNAs. In the initial stages of miRNA discovery, it was homology searches that first revealed orthologs of miRNAs across species and provided evidence for conservation [9].

The development of this homology-filtering approach was used to create a high-throughput pipeline in which all known miRNA genes were used as reference miRNAs. Fifteen potential miRNA candidates were discovered in *S. mansoni* in this analysis. These findings may provide some of the first evidence of the presence of miRNAs in *S. mansoni*. The pipeline can also be applied to other genomes, and requires only moderate computational resources. While the pipeline would likely benefit from further development and additional parameters, we demonstrate that portions of computational miRNA detection that are often performed manually and subjectively can be automated. In particular, a rules-based filtering approach to extracting the candidate pre-miRNA sequence appears to have merit.

Results

Homology search

The homology search for the high-throughput pipeline development was performed with all known miRNAs in miRBase (<http://microrna.sanger.ac.uk>). In total, 6211 mature miRNAs were used as reference sequences. The e-value cutoff for this analysis was set at 0.01. 180 total hits were observed with e-values less than 0.01. Of these hits, 15 *S. mansoni* sequences were represented multiple times. As a result, 110 unique *S. mansoni* sequences were returned from the BLASTn search. The BLASTn results with e-values of less than 0.01 are shown in the Additional file 1. The existing mature miRNAs that

produced these hits are reported, as well as the locations of the homologous sequences in the *S. mansoni* genome. All hits aligned with at least 95% of the miRNA sequence, resulting in alignment lengths of 19-24 nt.

RNA folding

All of the 110 unique mature miRNA candidates returned by the BLASTn search were assigned an analysis identifier with prefix, 'SMan'. The extended sequences (proximally 50nt flanking sequence around mature sequence) were folded with RNASHAPES [10]. A Table of the complete results for the extended sequence folding can be found in the Additional file 2.

Mean free energy – Mean free energy is a widely used criterion for filtering RNA folding results, and was observed to be an important filtering step in this analysis as well. The rationale for how MFE thresholds are derived, however, is not obvious when examining the literature. In fact, the guidelines proposed for uniform determination and annotation of miRNAs given by Ambros *et al.* [11] do not mention MFE thresholds. Instead, the guidelines suggest that to be considered a miRNA, a candidate's lowest MFE fold should be a hairpin. The MFE threshold is therefore a better criterion to rule out less likely candidates. Clearly, this method of filtering is necessary when experimental validation is required as resources are always limited. However, it may be a better practice to assign candidates probability scores based on their MFE, instead of having strict cutoffs. For those candidates that have been experimentally confirmed or not to be miRNAs, it would be useful to observe the relationship between MFE and the probability of being confirmed experimentally.

From the 110 unique *S. mansoni* sequences returned from the BLASTn search, 66 displayed MFE values of -15 kcal/mol or less when folded with RNASHAPES. 43 hairpins had MFE values of greater than -15 kcal/mol.

Hairpin structure – All 66 of the extended sequences with $MFE \leq -15$ kcal/mol displayed hairpins in at least one portion of the sequence when folded by RNASHAPES. At ~122nt in length, the sequence is considerably longer than a typical miRNA hairpin, and as a result, only the ~70nt surrounding the mature sequence are of interest. In each of the 66 hairpins detected, the region surrounding the mature sequence was within a hairpin.

Mature sequence location - Of the 66 hairpins with $MFE \leq -15$ kcal/mol, 36 structures contained the mature sequence entirely within the stem. The other 30 sequences contained the mature sequence partially or completely in a loop.

Extraction and folding of pre-miRNA sequence – Candidate pre-miRNA sequences were extracted from the 36 remaining hairpins and were refolded with RNAshapes. After refolding, 15 candidate pre-miRNAs had $MFE \leq -15$ kcal/mol. These 15 sequences were considered probable pre-miRNA sequences. A summary of the results and the structures for the pre-miRNAs candidates are shown in Figure 1.

The yield of probable miRNA candidates was much lower for this analysis with *S. mansoni* than analyses of species that contain closer relatives in miRBase. The findings suggest that it may have been difficult to find a large number of miRNAs in this analysis, as the closest relative to *S. mansoni* in miRBase is *S. mediterranea*. These two organisms merely belong to the same phylum, a relatively broad classification. On the contrary, if one was interested in studying miRNAs in a mammal not found in miRBase, one would find 22 members from the same class and over 2,500 miRNA sequences in miRBase. If one is interested in finding miRNAs, but not necessarily for a particular organism, one could take a similar approach as Sunkar and Jagadeeswaran [12]. Instead of localizing their search to one organism, they searched all publicly available nucleotide databases to find 682 miRNAs in 155 plant species. This approach, by continuing to populate miRBase, may in turn help searches for less represented phylogenetic branches. However, in the near term, organisms such as *S. mansoni* must continue to have a mix of computational and experimental approaches, with likely an emphasis on experimental discovery.

Comparison of results to existing work

The high-throughput pipeline yielded fifteen probable pre-miRNA candidates. This number is comparable to the number of miRNAs found by Palokodeti *et al* [13], who identified ten miRNA candidates in *S. mediterranea* by using all known human, *Drosophila* and *C. elegans* miRNAs as reference sequences. However, the yields for the *S. mediterranea* analysis and for this analysis with *S. mansoni* are considerably less than other studies that have used homology methods with multiple genomes. Luo *et al.* [14] identified 118 miRNAs in *Tribolium castaneum* (red flour beetle) using all available metazoan miRNAs as reference sequences. Zhou *et al.* [15] using a homology-based computational approach, found 300 human miRNA homologs in the domestic dog using

only human miRNAs as the reference miRNAs. Furthermore, Baev *et al.* [16] identified 639 chimpanzee miRNAs with a homology-based approach, also using only human pre-miRNA sequences as a reference set.

These trends are not necessarily surprising, as one would expect more hits when using reference miRNAs from organisms more closely related to the model organism. It is noteworthy, however, the extent to which the yield decreases as the distance between species grows. As a result, these findings suggest that the yield of a homology-based analysis is very dependent on the available miRNAs in miRBase. For example, Chatterjee and Chaudhuri [17] detected 489 homologous miRNA sequences in the *Anopheles gambia* genome, using only *Drosophila* miRNA as the query sequence. *Drosophila* is a well represented group of organisms in miRBase, but it is also in the same order as *A. gambia*. As a result, this close relationship produced a large number of hits. Vertebrates, especially mammals, are currently the most represented organisms in miRBase, with respect to number of species and number of miRNA sequences. Artzi *et al.* [18] used their recently released homology search web-server, miRNAMiner, to increase the number of miRBase miRNAs for seven mammals by 50%, identifying 790 new miRNAs. The strategy and filtering steps used by miRNAMiner are very similar to those used in this paper, but moderately more comprehensive. Clearly, the relationship between the model organism and the existing miRNAs plays a large role in the yield of the analysis.

Analysis of homology search hits by species

As shown in Figure 2, when the homology search for the high-throughput pipeline was performed, seventeen species had at least three or more hits with e-values < 0.01. An additional eleven species had two hits, and another seven species had one hit. *Mus musculus* displayed the highest number of hits with 56, over four times the number of hits for the next most represented species, *Triticum aestivum*. Although *M. musculus* has a high number (488) of miRNA sequences in miRBase, it does not appear that the sheer number of sequences is solely responsible for the higher frequency of hits. Other organisms are also well represented in miRBase, but displayed relatively few hits. For example, 695 human miRNAs are recorded in miRBase, but only five hits were observed here. Including the eight hits from *Rattus norvegicus*, over one-third of all hits observed were from the order, Rodentia. The cause of this observation is unclear. Interestingly, all 22 new human miRNAs identified by Artzi *et al.* [18] with miRNAMiner were homologs

of *M. musculus* or *R. norvegicus*. These findings may be merely coincidental or it may be possible that the miRNAs contained in these Rodentia species are particularly well conserved across species.

Both metazoan and non-metazoan miRNAs were used as reference miRNAs in this analysis. It was assumed that metazoan miRNAs would be more likely to produce hits in *S. mansoni*, and that the non-metazoan sequences would provide somewhat of a negative control for the method, i.e. few hits should be observed with non-metazoan sequences. Interestingly, four of the top eight represented species are plants: *T. aestivum* (13 hits), *Physcomitrella patens* (11 hits), *Arabidopsis thaliana* (7 hits) and *Oryza sativa* (6 hits). The number of hits with e-values < 0.01 for each major taxon listed in miRBase (subphylum, phylum or kingdom) is shown in Figure 3. With 37 hits, plants (Viridiplantae) represent 20% of the total number of hits with e-values < 0.01. The most represented taxon is the subphylum, Vertebrata, with 122 hits or 68% of the total hits. This finding is not surprising as Vertebrata is also the taxon with the most number of miRNAs in miRBase (5157), more than three times the number of miRNAs in the next largest taxa, Viridiplantae (1638) and Arthropoda (1194).

The percentages of miRNAs from each major taxon in miRBase that returned a hit with e-value < 0.01 are shown in Figure 4. The kingdom Protistae (6.1%) and phylum Platyhelminthes (4.8%) display the highest percentages of hits. However, both of these taxa contain only one organism in miRBase, with each organism containing less than 65 miRNAs. Of the three taxa that have the most representatives in miRBase, Vertebrata and Viridiplantae display similar percentages (1.6-1.7%), and are both higher than Arthropoda (0.8%). It is interesting that a higher frequency of hits is observed for Viridiplantae than for Arthropoda, considering *S. mansoni* and Arthropoda would be more closely related as both are metazoans.

Observed miRNA families

As shown in Figure 5, 36 different miRNA families were observed in the homology search. Of these, 22 families were observed multiple times, either from different species or within the same species. The miRNA family observed most frequently was miR-19, with 22 hits. Also shown in Figure 5, above the bar for each miRNA family, is the number of probable miRNA candidates that were observed in each family. Five of the six families with the most hits displayed at least one probable miRNA candidate. Ten of the thirteen families

that displayed probable miRNA candidates rank in the top sixteen families with respect to number of hits. These results suggest that miRNA families that are highly conserved, appearing in the most number of species, may be most likely to yield probable miRNA candidates.

Conclusions

The study and discovery of miRNA genes and targets have grown immensely in just several years. The potential therapeutic benefits that could result from further understanding miRNAs and their function have motivated numerous experimental and computational methods to detect these genes. In this analysis, using limited computational resources, fifteen putative miRNAs were detected in the parasitic organism *S. mansoni*. These results are encouraging, suggesting even laboratories without access to experimental resources can contribute to the detection of miRNA genes.

As the amount of miRNA data increases and research spreads into new organisms, the need for optimized computational techniques and a concerted effort between experimental and computational approaches grows. The impact of miRNAs is only beginning to be understood, and will be more fully realized as excitement surrounding miRNA research grows and discoveries continue.

Methods

Selection of reference miRNAs

All mature miRNA sequences in miRBase: (<http://microrna.sanger.ac.uk>, Sept 2008) were downloaded as reference miRNAs. The search set included both metazoan and non-metazoan sequences, e.g. plants and viruses. In total, 6,211 mature miRNA sequences were selected for searching against in the *S. mansoni* genome. A flowchart describing the high-throughput pipeline is shown in Figure 6.

Homology search with all known miRNA sequences

A BLASTn search was performed with the mature miRNA sequences downloaded from miRBase on the *S. mansoni* database. The expectation value cutoff for the pipeline development was set at 0.01.

Extraction of extended sequences from *S. mansoni* genome

MATLAB Bioinformatics toolbox was used to extract the candidate miRNA sequences from the *S. mansoni* genome. An additional 50nt on each side of the candidate mature miRNA sequence were taken. These extended sequences were then used for further analysis with the understanding that the sequence contained the candidate mature sequence, candidate hairpin, and extra nucleotides.

Folding of extended miRNA sequences using RNASHAPES

Extended candidate miRNA sequences were folded using the standalone version of RNASHAPES. The Shape Folding mode was used with Shape Type 3, allowing intermediate abstraction level of the shape. RNASHAPES generates multiple folds for each sequence, ranking them by MFE. For these analyses, the fold with the lowest MFE was considered. During method development, a mix of rules-based filters (e.g. MFE) and manual/visual inspection of the folded extended miRNA were used to determine probable pre-miRNA candidates. In the development of the pipeline, with an emphasis on automation, the following rules-based filters were developed and implemented:

- Folded extended miRNA sequences with MFE greater than -15 kcal/mol were removed.
- RNASHAPES outputs a dot-bracket representation of the secondary structure. An example of the dot-bracket output is shown here:

... (((.....)))... ((((((((((.....(((.....)))..))))))))))..

Opposing sets of parentheses represent individual hairpins. Each parenthesis represents a paired base within a hairpin. Dots represent unpaired bases. In the dot-bracket output above, the underlined portion represents one hairpin, with five paired bases on either stem and four bases in the loop. The entire dot-bracket output represents two hairpins with three unpaired bases on either end of the sequence.

- Structures with the mature miRNA contained or partially contained in the loop of the hairpin were excluded, i.e. no bases in the mature miRNA can be represented by dots that are between opposing parentheses. For example, the structure below was filtered out of the analysis. The mature miRNA sequence is bolded in the sequence and in the dot-bracket diagram.

3. Lindow M, Gorodkin J: **Principles and limitations of computational microRNA gene and target finding.** *DNA and Cell Biology* 2007, **26**(5):339-351.
4. Lim L, Lau N, Weinstein E, Abdelhakim A, Yekta S, Rhoades M, Burge C, Bartel D: **The microRNAs of *Caenorhabditis elegans*.** *Genes and Development* 2003, **17**:991-1008.
5. Chaudhuri K, Chatterjee R: **MicroRNA detection and target prediction: integration of computational and experimental approaches.** *DNA and Cell Biology* 2007, **26**(5), 321-337.
6. WHO/TDR - Special Programme for Research & Training in Tropical Diseases (TDR), **World Health Organization** ,(<http://www.who.int/tdr/>), 2008.
7. Xue X, Sun J, Zhang Q, Wang Z, Huang Y, Pan W: **Identification and Characterization of Novel MicroRNAs from *Schistosoma japonicum*.** *PLoS One* 2008, **3**(12): e4034.
8. Freghult EK, Bollback JP, Gardner PP: **Exploring genomic dark matter: a critical assessment of the performance of homology search methods on noncoding RNA.** *Genome Research* 2007, **17**:117-125.
9. Zhang B, Pan X, Wang Q, Cobb G, Anderson T: **Computational identification of microRNAs and their targets.** *Computational Biology and Chemistry* 2006, **30**:395-407.
10. Steffen B, Voß B, Rehmsmeier M, Reeder J, Giegerich R: **RNASHapes: an integrated RNA analysis package based on abstract shapes.** *Bioinformatics* 2006, **22**(4).
11. Ambros V, Bartel B, Bartel D, Burge C, Carrington J, Chen X, Dreyfuss G, Eddy S, Griffiths-Jones S, Marshall M, Matze M, Ruvkun G, Tuschl T: **A uniform system for microRNA annotation.** *RNA* 2003, **9**:277-279.
12. Sunkar R, Jagadeeswaran G: **In silico identification of conserved microRNAs in large number of diverse plant species.** *BMC Bioinformatics* 2008, **8**(37).
13. Palakodeti D, Smielewski M, Graveley B: **MicroRNAs from the planarian *Schmidtea mediterranea*: a model system for cell biology.** *RNA* 2006, **12**:1640-1649.
14. Luo Q, Zhou Q, Yu X, Lin H, Hu S, Yu, J: **Genome-wide mapping of conserved microRNAs and their host transcripts in *Tribolium castaneum*.** *J. Genetics and Genomics* 2008, **35**:349-355.

15. Zhou D, Li S, Wen J, Gong X, Xu L, Luo Y: **Genome wide computational analyses of microRNAs and their targets from *Canis familiaris***. *Comput. Biol. Chem.* 2008, **32**: 61-66.
16. Baev V, Daskalova E, Minkov: **Computational identification of novel microRNA homologs in the chimpanzee genome**. *Comput. Biol. Chem.* 2008, article in press.
17. Chatterjee R, Chaudhuri K: **An approach for the identification of microRNA with an application to *Anopheles gambiae***. *Acta Biochimica Polonica* 2006, **53(2)**:303-309.
18. Artzi S, Kiezun A, Shomron N: **miRNAMiner: A tool for homologous microRNA gene search**. *BMC Bioinformatics* 2008, **9(39)**.

Figures

Figure 1 - Probable miRNA Candidates

For each structure, the location in the *S. mansoni* genome (version 4.0), pre-miRNA sequence, miRNA family, sequence length and MFE are given. The start and end of the mature sequence are circled in the structure. The mature sequence is also bolded in the pre-miRNA sequence.

Figure 2 - Number of microRNAs hits per species with three or more hits with e-value < 0.01

A homology search for the high-throughput pipeline was performed. Seventeen species had at least three hits with e-values < 0.01. An additional eighteen species had one or two hits. *M. musculus* displayed the highest number of hits with 56, over four times the number of hits for the next most represented species, *T. aestivum*.

Figure 3 - Homology search hits by taxon

The number of hits with e-values < 0.01 for each major taxon listed in miRBase (subphylum, phylum or kingdom) is shown. The two most represented taxa in miRBase, Vertebrata and Viridiplantae display the highest number of hits.

Figure 4 - Homology search hits as a percentage of miRBase entries

The percentages of miRNAs from each major taxon in miRBase that returned a hit with e-value < 0.01 are shown. The kingdom Protistae and phylum Platyhelminthes display the highest percentages of hits, possibly due to low representation in miRBase. These taxa each contain only one organism and less than 65 miRNAs in miRBase.

Figure 5 - Frequency of homology search hits by miRNA family

36 different miRNA families were observed in the homology search, with 22 families observed multiple times. Families with higher number of hits were more likely to display probable miRNA candidates.

Figure 6 -Flowchart of high-throughput pipeline

A pipeline has been developed to quickly screen for probable miRNA candidates in *S. Mansoni*. The discovery pipeline can likely be applied to the identification of pre-miRNA hairpins in other organisms.

Additional files

Additional file 1 – High-throughput Pipeline Homology Search Results

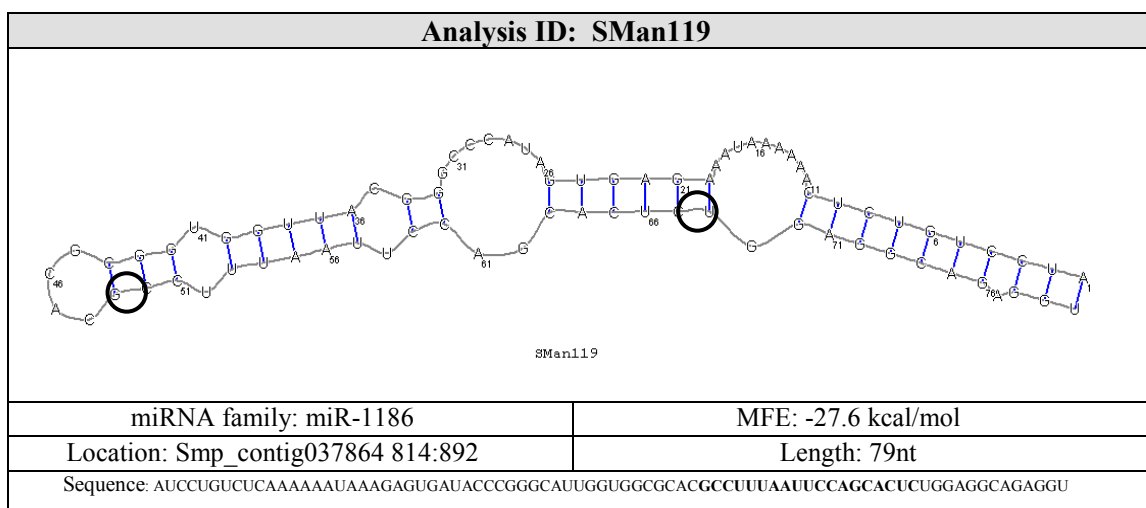
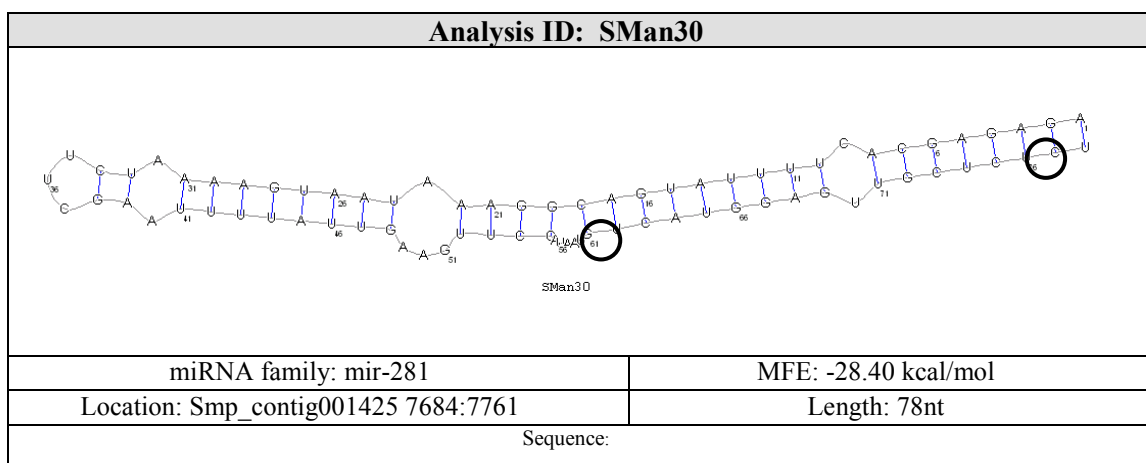
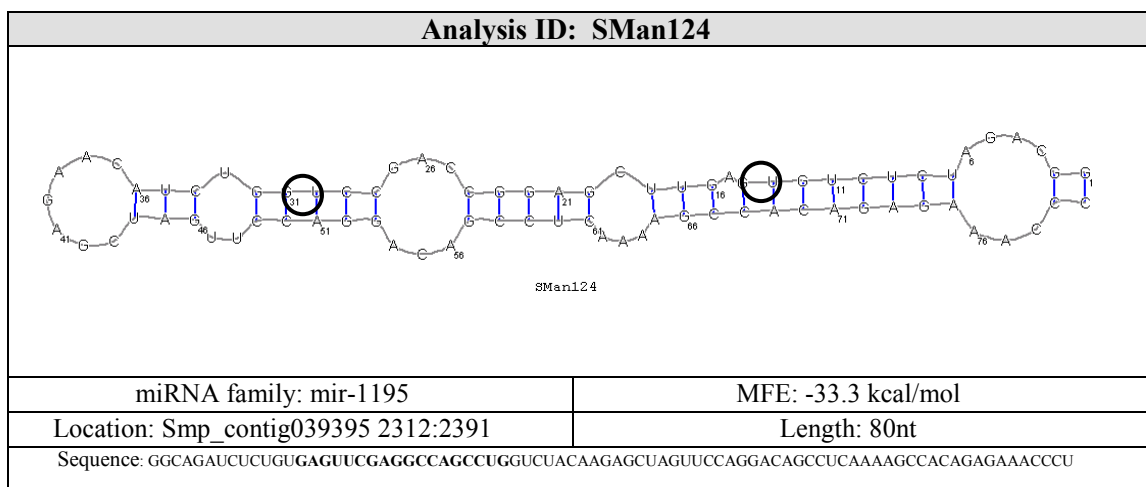
File format: .doc

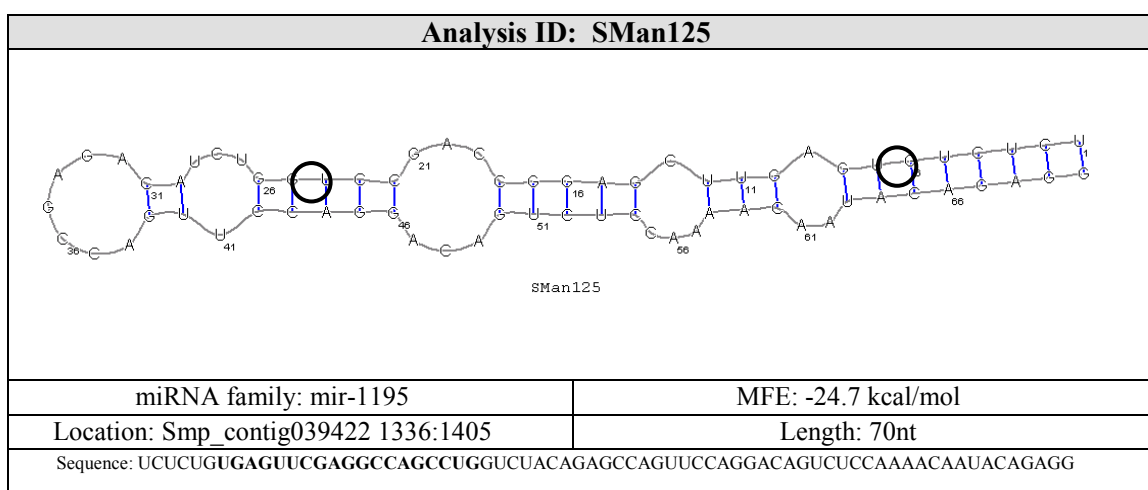
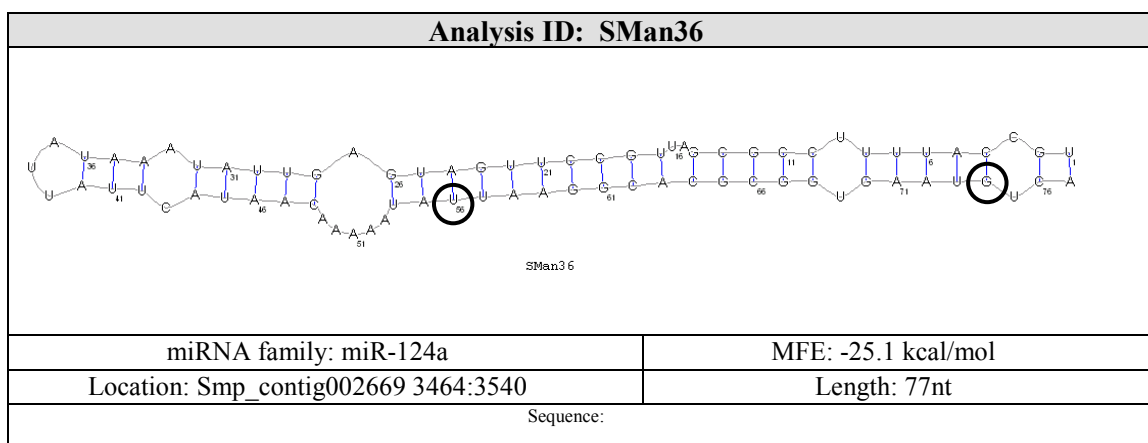
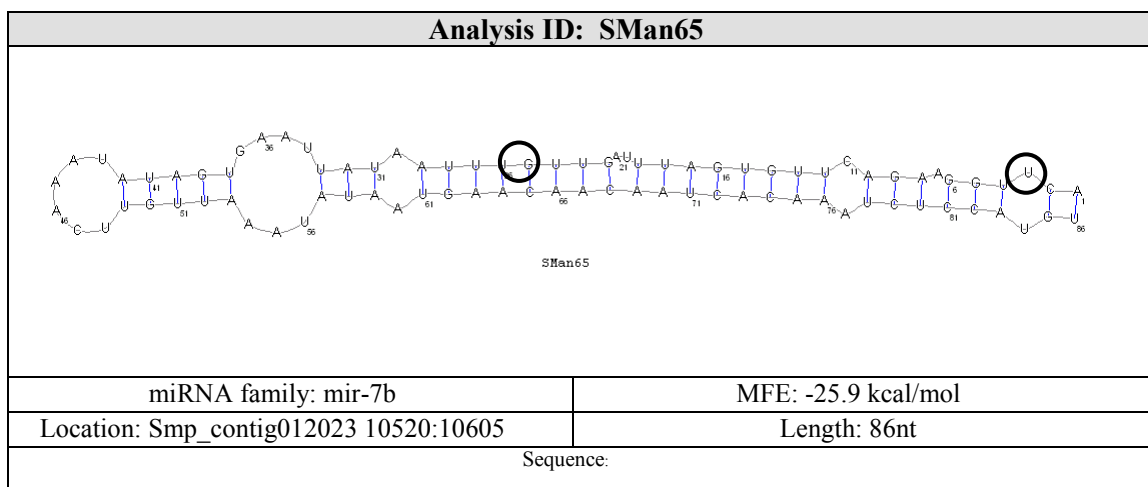
Description: Results from *Blastn* homology search using all known mature miRNA sequences to search the *S. Mansoni* miRNA database (e-value < 0.01). The file is available at <http://www.cebio.org/content/2009/04/08/schistosoma-mansoni-micromas>.

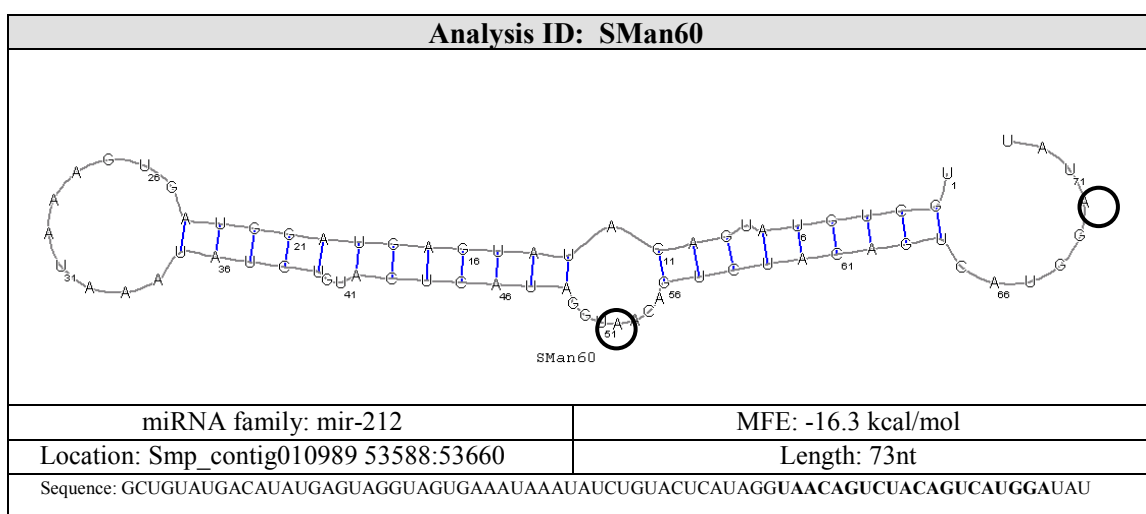
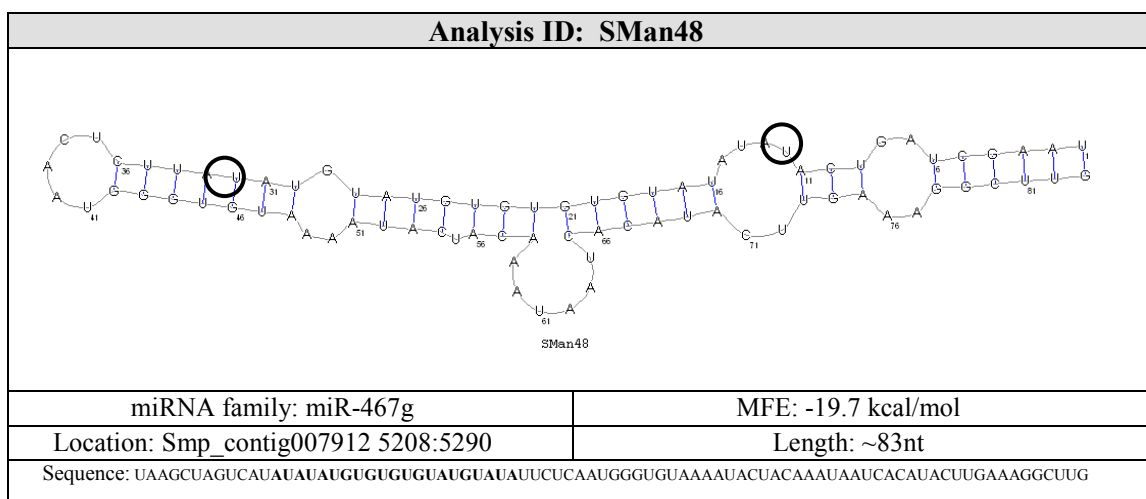
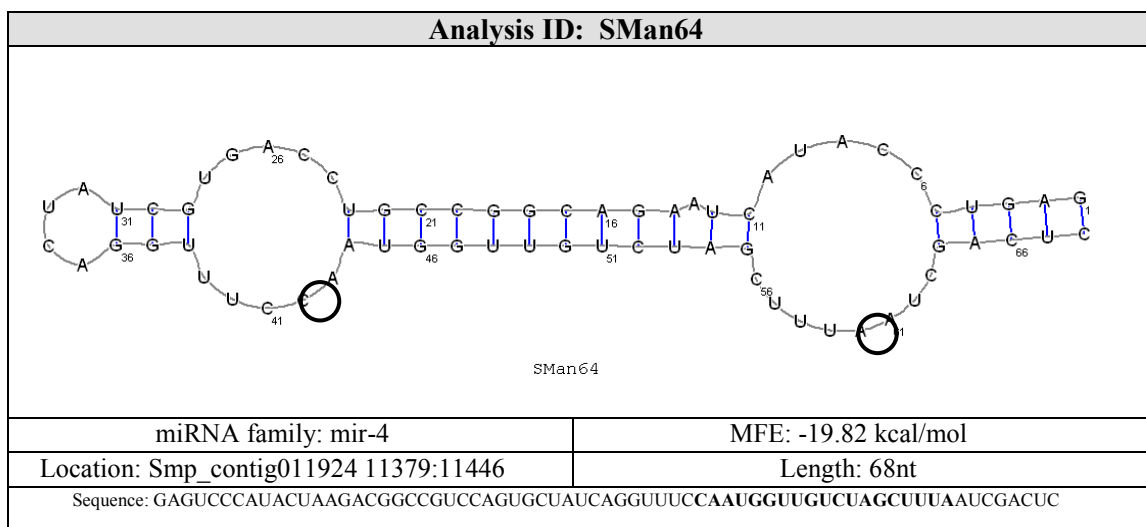
Additional file 2 – High-throughput Pipeline Extended Sequence Folding Results

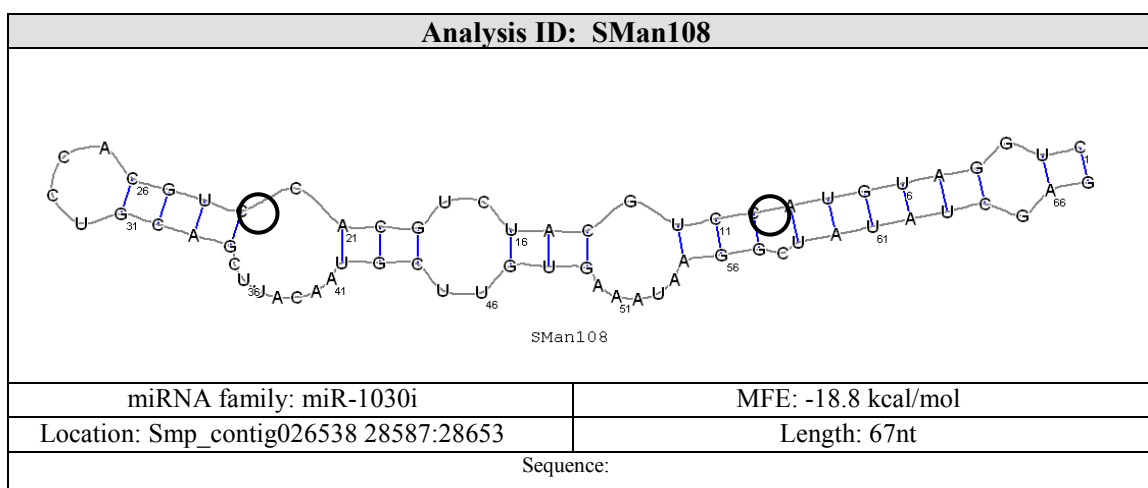
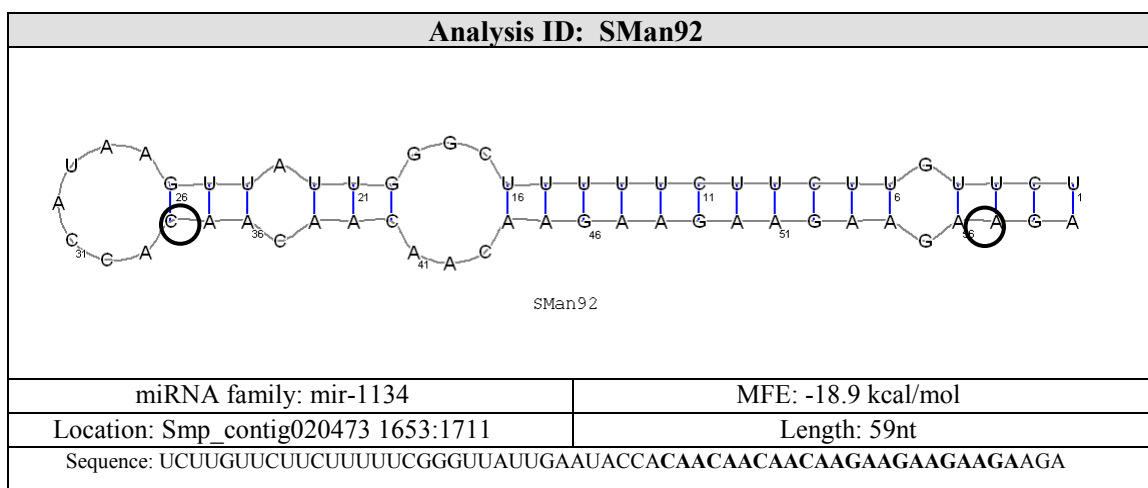
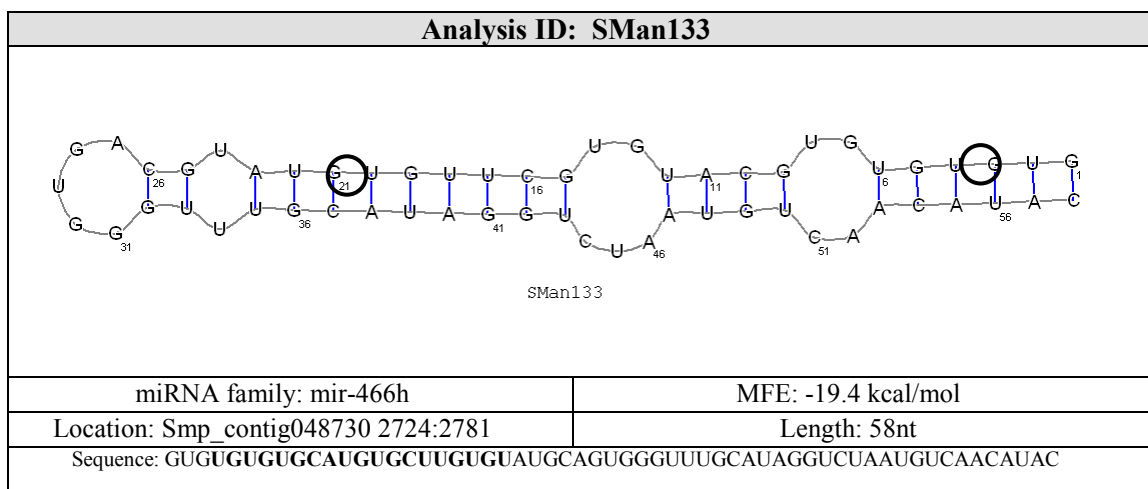
File format: .doc

Description: Results from *RNAShapes* folding of 110 extended miRNA sequences. The file is available at <http://www.cebio.org/content/2009/04/08/schistosoma-mansoni-micromas>.









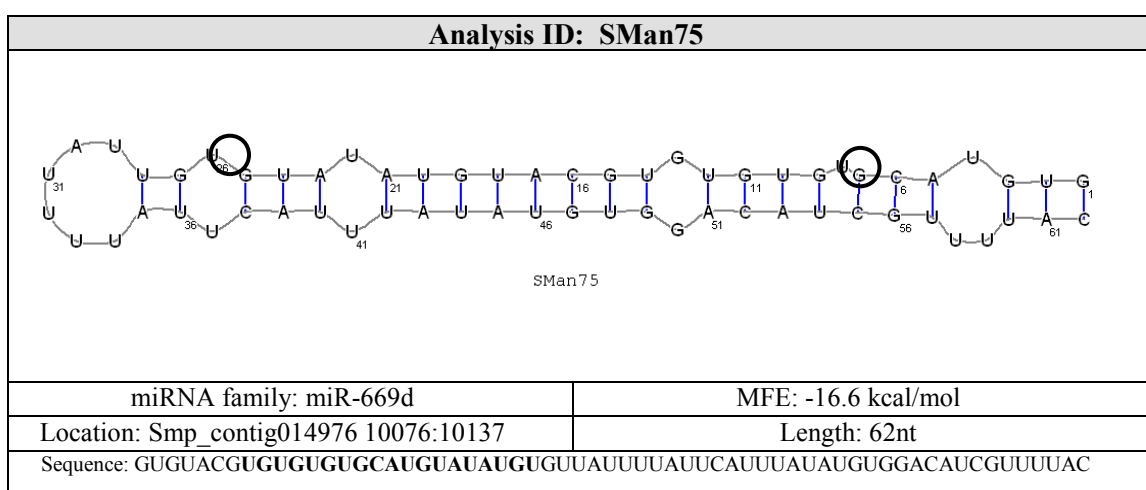
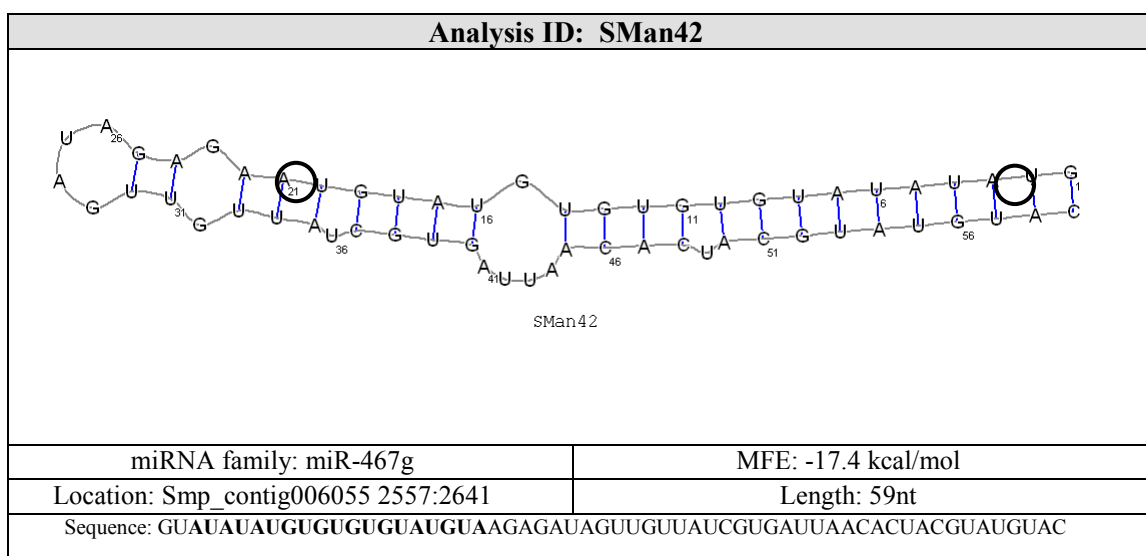
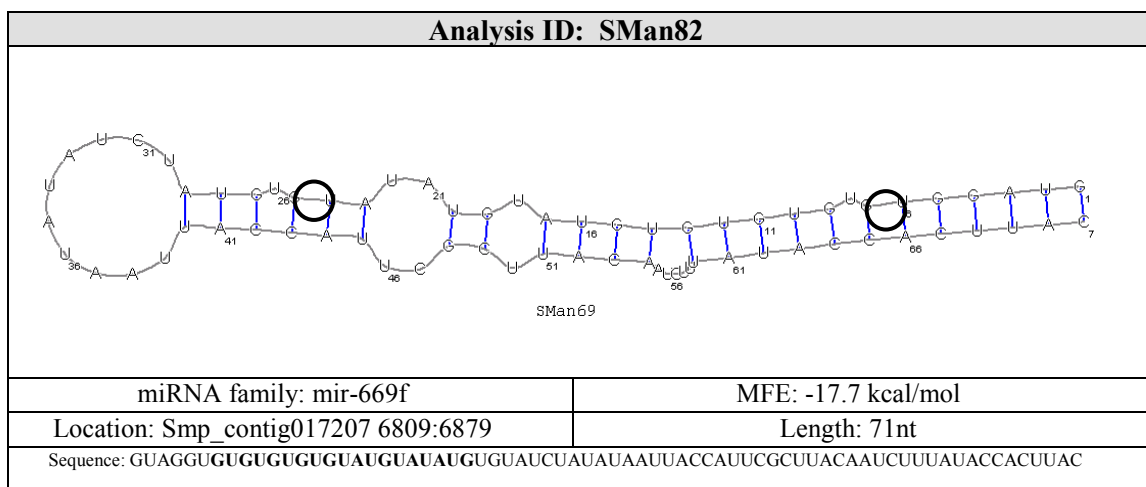


Figure 1

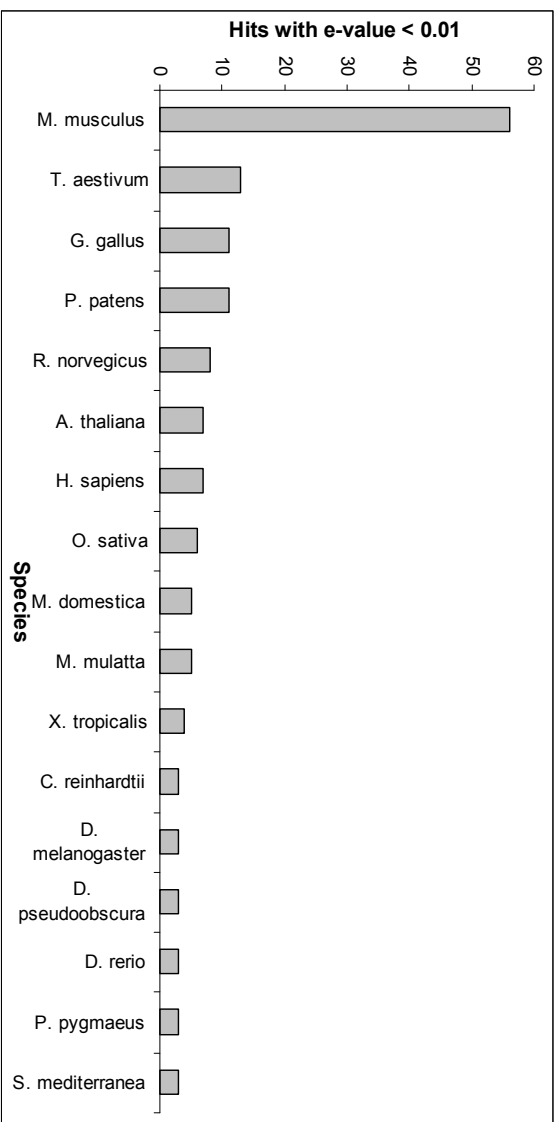


Figure 2

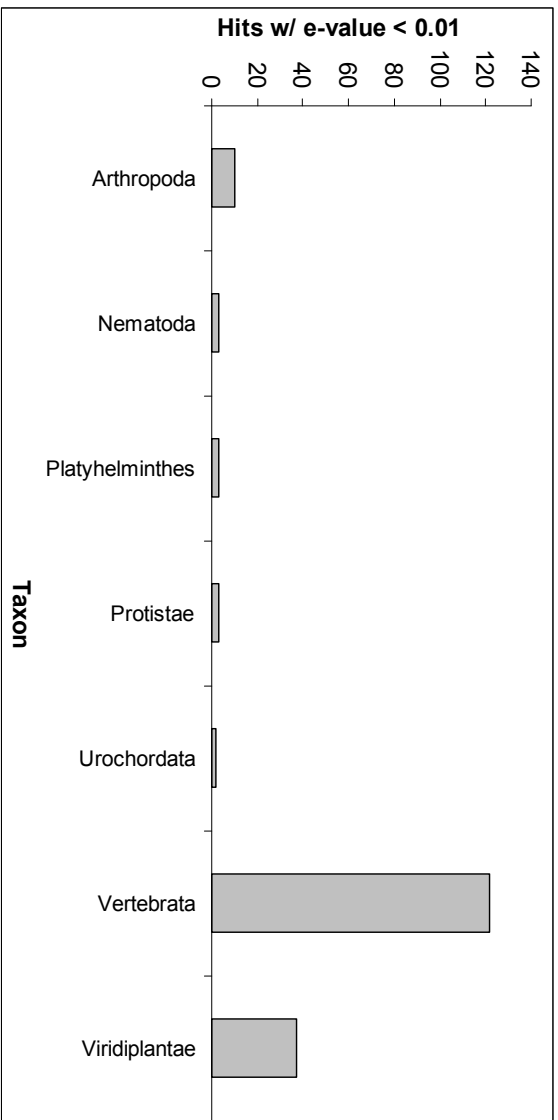


Figure 3.

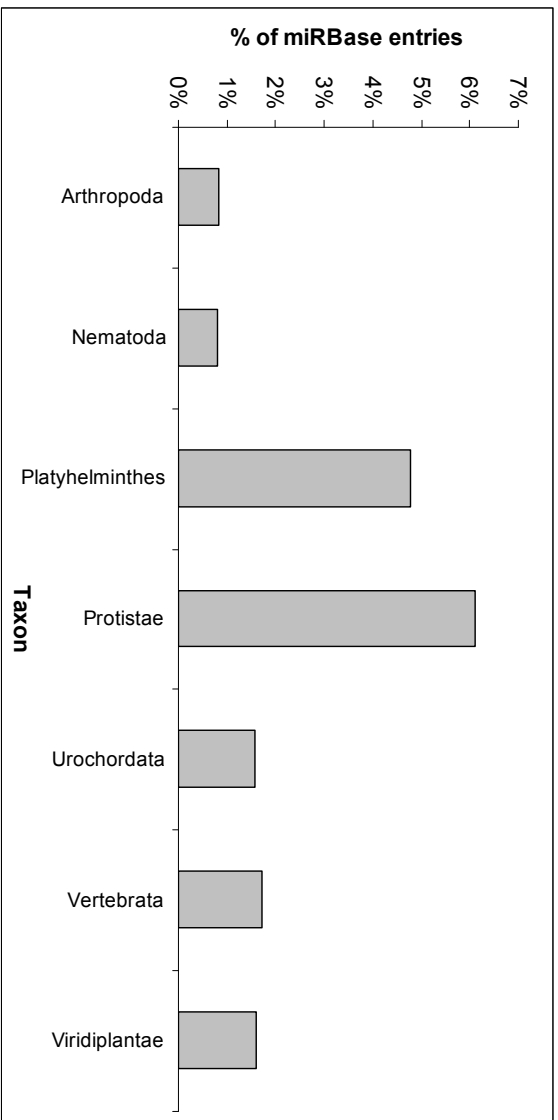


Figure 4.

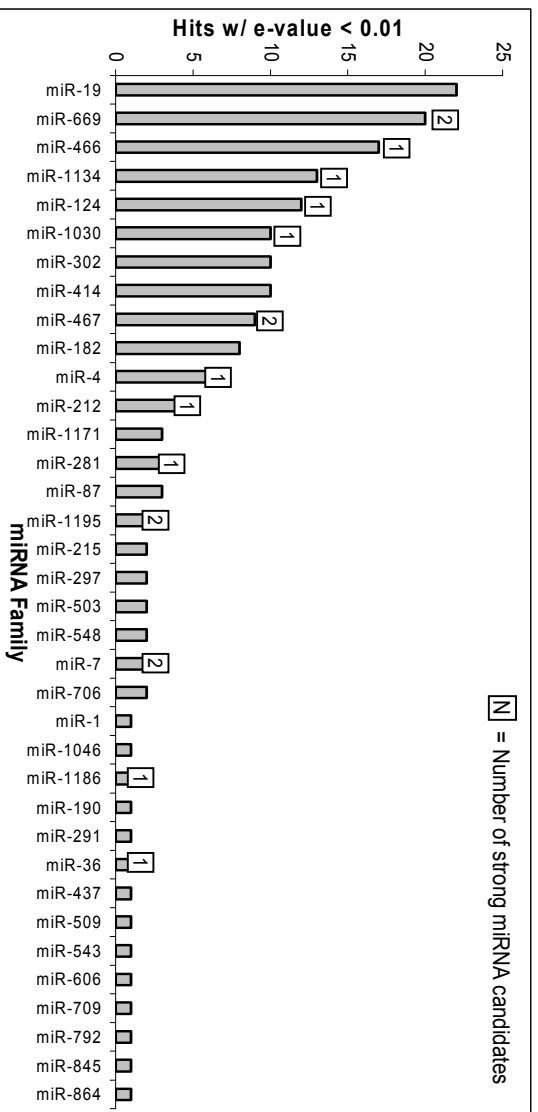


Figure 5

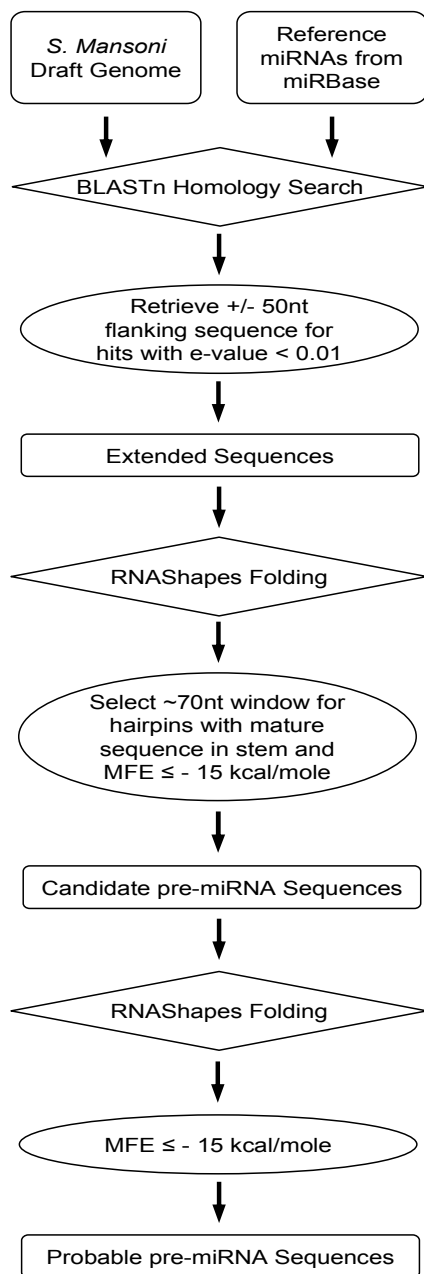


Figure 6

Glucose transporters are critical for schistosome survival in the vertebrate host

Greice Krautz-Peterson, Mariana Simoes*, Zahra Faghiri, Guilherme Oliveira* and Patrick J. Skelly.

Molecular Helminthology Laboratory, Division of Infectious Diseases, Department of Biomedical Sciences, Tufts University, Cummings School of Veterinary Medicine, Grafton, MA 01536, USA

*Laboratory of Cellular and Molecular Parasitology, Instituto René Rachou - FIOCRUZ, Belo Horizonte, MG, Brazil.

Mammalian-stage schistosomes absorb glucose and other nutrients from the host's bloodstream through their unique syncytial double lipid bilayer surface, referred as tegument. Two facilitated glucose diffusion transporters have been identified in the tegument i.e., schistosome glucose transporter 1 (SGTP1) and 4 (SGTP4). SGTP4 is present mostly or exclusively within the apical membranes, while SGTP1 is found only within the basal membranes. This asymmetric localization in addition to their distinct kinetic properties facilitates the directional movement of glucose into the worm. A similar mechanism of glucose diffusion is found in other eukaryotes as well. Given that schistosomes are highly dependent on host's glucose, we use RNAi to knockdown expression of both SGTPs and to monitor phenotypic changes and/or survival of parasites in culture and in vivo. When schistosomula were treated in vitro for 14 days with siRNA specific to SGTP1 or SGTP4, individually or mixed, targeted transcript levels were reduced by 50% for SGTP1 and by 80 to 90% for SGTP4 as determined by quantitative real time-PCR, while non-targeted transcripts were unaffected. Parasite exposure to SGTP1 or SGTP4 siRNA reduced schistosomula glucose-uptake capacity by 50%. The reduction

was even more pronounced (70%) when parasites were treated with both siRNAs together, resulting in 40% of parasite death in culture. Injection of mice with schistosomula, pre-electroporated with combined SGTP1 and 4 siRNAs, significantly decreased the number of worms recovered from mice 28 days after parasite-injection, compared to mice injected with control siRNA-treated parasites. Our findings demonstrate that glucose transporters are important for schistosome development in the mammalian host and could be a potential target for RNAi-based treatment of schistosomiasis.

X - APÊNDICE

Apêndice 1 – Candidatos a miRNA de *Schistosoma mansoni*

S. mansoni miRNA	Seqüência (5' – 3')	pb	Freqüência
sma-miR-1	GGTCATGTCAGATGAGTCGTGGTGC	25	32
sma-miR-2	TGATCCGGCGGGTGCGGTGCCAGTT	25	24
sma-miR-3	ACGTGAGTATAGGGGCGAAAGACTA	25	24
sma-miR-4	TTGAATCTGGGTCGTGCGGTCGCAT	25	17
sma-miR-5	AATGTAGGTAAGGGAAGTCGGCA	23	13
sma-miR-6	TTCGATTCCGGAGAGGGAGCCTG	23	10
sma-miR-7	TTGATTCCGGTGGGTGGTGGTGC	22	10
sma-miR-8	ACTGGGGCGGTACATCCGTCAA	23	8
sma-miR-9	TAAATCGGAACAGAGACGGGGTGC	24	9
sma-miR-10	TAAACCAGGGGTCGGGAGTTC	21	6
sma-miR-11	AGGTCGTTGGATTGGCGTCGTTGT	24	8
sma-miR-12	TTATGGAGTAGTGTGGTCGTG	21	8
sma-miR-13	TACATCCATCATGTCTCCTCACA	24	9
sma-miR-14	ATTTCCGGGGTGCCTGGTCT	19	8
sma-miR-15	CGGGTGTCGGGAGCGGACGGC	21	6
sma-miR-16	ATAGTGGGAGCTGGGTGCTTC	22	6
sma-miR-17	CGAGGTGGAGAAGGGTTCCA	20	6
sma-miR-18	TTTGGGGTACGGGGTTCGAC	20	6
sma-miR-19	TGGTTGGGTCGGTAGGTGG	19	6
sma-miR-20	AGTCGTGGTGCGAATTGCGGTGT	23	5
sma-miR-21	GTTGGTGGAGCGATTTGTCTG	21	4
sma-miR-22	TGATCGTAAGGCAATGTGGTGT	24	3
sma-miR-23	GCGGTGTGTCGGAGATGGCGGC	22	3
sma-miR-24	ATGGTAACGCGGGTGTCTATG	22	3
sma-miR-25	ACGTCGGCGAGAGTTCTGGGA	21	3
sma-miR-26	TTGGTGTGAGGGTCTGCTTG	20	3
sma-miR-27	TTGAATCTGGGTCGTGCG	18	3
sma-miR-29	AAGTCGTAACAAGGTTTCCGTAGGT	25	2
sma-miR-30	CCCGGCCCGGACACTGTGAGGATT	24	2
sma-miR-31	CGGAGTTTGAAGCAAGAGGTGTCA	24	2
sma-miR-32	CGTGACTGGGATCGGGGCTTGCA	23	2
sma-miR-33	TGAAAGACGATGGTAGTGAGATG	23	2
sma-miR-34	CCACGCCTGTCCGAGGGTCCGGC	22	2
sma-miR-35	ACGCGGGTGTCTATGGTGAGC	22	2
sma-miR-36	GTGAGCAAAGTTTCAGGTGTGT	22	2
sma-miR-37	CTTGAGGCCAGGGTGCGGGAT	22	3
sma-miR-39	GTA CTGGGCATACTGGCCGC	21	2
sma-miR-40	TGTGCGCGAGTCATTGGGTGT	21	2
sma-miR-41	ATGAGCGTGAGGATACTAAAT	21	2
sma-miR-42	CTGGTGCCAGCAGCCGCGGTA	21	2
sma-miR-43	GTTAGTTGGTGGAGTTGGATA	21	2
sma-miR-44	GTGTCGGCGACGGATCTTTCA	21	2
sma-miR-45	CAACGAGTAGGAGGGTCCGCT	20	2
sma-miR-46	CTCTGTGGGCGTGGTTCCGA	20	1
sma-miR-47	ACGGGGAGGTAGTGACGAAA	20	2
sma-miR-48	TGTGGTGTAGTGGTAGACG	19	2
sma-miR-49	ATGGGCTGTGAGTGTTACT	19	2
sma-miR-50	ATCGAGAGGTCCCCGGTTC	19	2
sma-miR-51	ATGGCGGTGTTAGAGGTG	18	2
sma-miR-52	GCCCAGCCACGTGTCGGA	18	2

sma-miR-65	TGACCTGTCGGCATGCTTCCGGATGC	26	1
sma-miR-66	TGGAGTGACGGCAACGAGTAGGAGG	25	1
sma-miR-68	ACAGTGCGTCAAACCGCTCAAAGGT	25	1
sma-miR-74	AGTCGGAAGCCGCTAAGGAGTGT	23	1
sma-miR-75	ACGTCGGGCCCGGTCAGTACTTG	23	1
sma-miR-77	ACGGGTGGAGTTGAACTGCAAG	22	1
sma-miR-78	CACCGTCGTGACTGGGATCGGG	22	1
sma-miR-79	CAGCCCCCTGTAGCGAAAAGGGA	22	1
sma-miR-80	TTGGCCAATAGTCTGTGGTGT	22	1
sma-miR-83	CTTTTGGTGTCTGGACTGGACC	22	1
sma-miR-86	GCTGGGCGTGGTGGCGTGCGCC	22	1
sma-miR-89	TCGGTCCAGTTGGGGGCTGGT	21	1
sma-miR-90	CTACGAAGGTTACGGACGTAT	21	1
sma-miR-91	GTGTCGGAGATGGCGGCTTCG	21	1
sma-miR-92	ACGAGAGGAACCGCAGGTTCA	21	1
sma-miR-94	TCGGCGAGAGTTCTGGGAAGA	21	1
sma-miR-95	CGTCTAAGGAGGGGCGGACCG	21	1
sma-miR-96	GGAGGTAGTTCGTTGTGTGGT	21	1
sma-miR-98	CGGGGGTCGACACGATATCTA	21	1
sma-miR-99	CAGCTGCCACCATTCCCAA	21	1
sma-miR-100	GTCAGTCTTGTTGGGTCGGT	21	1
sma-miR-102	ACTCACCCGGCCCGGACACTG	21	1
sma-miR-103	CCAAGGACGGCAGCAGGCGCG	21	1
sma-miR-105	AGGAGTGGAGCCTGCGGTTTA	21	1
sma-miR-106	TTGGTATATGTGCCTGGTCGA	21	1
sma-miR-108	ATCGTGTCTGCCGAAAGAGC	21	1
sma-miR-112	ACTTACACGAGTCCGATAGC	20	1
sma-miR-114	TAGTCTGTGGTGTAGTGGTA	20	1
sma-miR-115	TGTGTGCACCGTCGGGGCGG	20	1
sma-miR-116	CCGTCGTGAGACAGGTTAGT	20	1
sma-miR-121	TATCTGAAGTAACTGCGAA	20	1
sma-miR-122	AGCTGGGTGCTTCGGTGTCC	20	1
sma-miR-124	ATGTGAGTGCCAAGTGGGCC	20	1
sma-miR-125	ATCTGGGTCGTGCGGGCGCA	20	1
sma-miR-126	TCCGAGAGGGAGCCTGAGA	20	1
sma-miR-128	TAAGAGGGTCGCTGTGGCTA	20	1
sma-miR-129	ACCAGGAGTGGAGCCTGCGG	20	1
sma-miR-133	TTAGAGGCATTGGGGAGGAA	20	1
sma-miR-135	TGACTGGGGCGGTACATCCG	20	1
sma-miR-136	ATTGAATGTGAGTGCCAAGT	20	1
sma-miR-138	TTAAAAGTGGGTGAGATTGA	20	1
sma-miR-139	AACTGGCGCTGTGGGATGA	19	1
sma-miR-140	CTTGGTGGTAGTAGCAAAT	19	1
sma-miR-141	CGACCTCGGATCAGGCGTG	19	1
sma-miR-142	ATCGGAGCTTGGGGTCATG	19	1
sma-miR-143	GTCTGTGGTGTAGTGGTAG	19	1
sma-miR-145	AGACATGATGGATGTAGTA	19	1
sma-miR-146	TTGAAGCAAGAGGTGTCAG	19	1
sma-miR-147	AACGGGTGCTGGTGGGTTG	19	1
sma-miR-148	CGTCGGCGGTTGCGTGTGT	19	1
sma-miR-149	AGGGAACGTGAGCTGGGTT	19	1
sma-miR-150	TTTTGGTGGGTATGTGTGA	19	1
sma-miR-151	GGTGAGCTCAGTCAGGACA	19	1
sma-miR-152	TTGTGGCGGCAAGCGTTC	19	1
sma-miR-153	CTTGACGTAATGATTTGAG	19	1

sma-miR-154	ACGGCAACGAGTAGGAGGG	19	1
sma-miR-155	ATGTGGGCGTGGGTTTCGTA	19	1
sma-miR-156	TCTTCATCATAGTACTCGG	19	1
sma-miR-157	AAGTCGCACGAAATTGAGC	19	1
sma-miR-159	TCACGTCGGCGGTTGCGTG	19	1
sma-miR-160	CTTGGTCTAGGCATGGAGT	19	1
sma-miR-161	AGTGAAGCGTGCTGGGCC	19	1
sma-miR-163	TTGAACAATGTAGGTAAGG	19	1
sma-miR-164	TGCCGCTCGACTGATGCTG	19	1
sma-miR-165	GCTTCGGTGCGCGTTGTGG	19	1
sma-miR-168	CTGGTGGAGGACCGTAGCG	19	1
sma-miR-169	TAACAGGTCTGTGATGCC	19	1
sma-miR-170	TGGAGTCGGGTTGTTGTG	19	1
sma-miR-171	TTCCCCTAGTAACTGCGAG	19	1
sma-miR-172	TAACGTGAGTATGGGGCG	19	1
sma-miR-176	GTTGGTTCGGTAGGTGGC	18	1
sma-miR-178	ACTCTGGTGGAGGACCGT	18	1
sma-miR-180	CATATCACTAAGCGGAGG	18	1
sma-miR-181	TTCAGTGAAACGGGAACG	18	1
sma-miR-182	CCGAGGTGGAGAAGGGTT	18	1
sma-miR-183	TCCATGTGACAGCCCGGT	18	1
sma-miR-184	GTGCTGGTGGGTTGACGA	18	1
sma-miR-185	ATTCAGCTGATGAGTGTG	18	1
sma-miR-186	AAGTCCCCGCGTGGTGCG	18	1
sma-miR-187	TTTTGGTGGTATGTGTG	18	1
sma-miR-189	ACCAGGGTTCGGGGTTC	18	1
sma-miR-190	GTAAAAGTCGTAACAAGG	18	1
sma-miR-191	AGGTCCAACGCTCGAGCG	18	1
sma-miR-192	CTGGGGGTTTTGGTGGGC	18	1
sma-miR-194	ACGGGGAATCAGGGTTCCG	18	1
sma-miR-195	CCCTTGGAGTCGGGTTGT	18	1
sma-miR-197	AACTGGCGCTGTGGGATG	18	1
sma-miR-198	GCTGGTGGGTTGACGAGT	18	1
sma-miR-199	ACTTACACGAGTCCGATA	18	1
sma-miR-200	ATGAGCGTGAGGATACTA	18	1
sma-miR-202	ATTTTCGGAGTTTGAAGC	18	1
sma-miR-203	ATACTTACACGAGTCCGA	18	1
sma-miR-204	AACTGCGAGTGAACAGGG	18	1
sma-miR-205	GTCAGGATGGCCGAGTGG	18	1
sma-miR-206	TCGCTCCACCAACTAAGA	18	1
sma-miR-207	AACTTCGGGAGAAGGATT	18	1
sma-miR-208	TGTGAGTGCCAAGTGGGC	18	1
sma-miR-209	TTTGGTGGGATGTGTGA	18	1
sma-miR-210	ATAGGGAACGTGAGCTGG	18	1
sma-miR-211	GTCTGTGGTGTAGTGGTA	18	1
sma-miR-212	TCTCAGGGTCGTGGGTTTC	18	1
sma-miR-213	ACTGGGATCGGGGCTTGC	18	1
sma-miR-214	AGGACTCGGAAGTAGGGG	18	1
sma-miR-215	AGTCGTGGTGC GAATTGC	18	1
sma-miR-216	AAGGACGGCAGCAGGCCG	18	1
sma-miR-218	TTGTGAGGAGACATGATG	18	1
sma-miR-219	ATGAGCGTGAGGATACT	17	1
sma-miR-221	TCGGTGGGTTGGTGGTGC	17	1
sma-miR-222	ATTGTGGTATGAGCGTG	17	1
sma-miR-223	GTCAGATGAGTCGTGGT	17	1

sma-miR-225	CGTGAGTATAGGGGCGA	17	1
sma-miR-226	TTTGGTGGGTATGTGTG	17	1
sma-miR-227	AGATGAGTCGTGGTGCG	17	1
sma-miR-228	TGGGTCAGGCGGTCCTA	17	1
sma-miR-229	AGTAACTGCGAGTGAAC	17	1
sma-miR-231	AACGTGAGTATAGGGGC	17	1
sma-miR-233	AGGTCTCCAAGGTGAAC	17	1
sma-miR-234	AATAGGGAACGTGAGCT	17	1
sma-miR-235	ATGAGCGTGAGGATACT	17	1
sma-miR-236	ACGTGAGTATAGGGGCG	17	1
sma-miR-238	ACGGGTGTCGGGAGCGG	17	1
sma-miR-241	CTGAGGGATTAAGACTG	17	1
sma-miR-242	ATGAGCGTGAGGATACT	17	1
sma-miR-243	TTCGGGGTGCGTGGTCT	17	1
sma-miR-244	TTCGGGGTGCGTGGTCT	17	1
sma-miR-245	TTCGGGGTGCGTGGTCT	17	1
sma-miR-246	AGATGAGTCGTGGTGCG	17	1
sma-miR-247	GCGAGTCATTGGGTGTT	17	1
sma-miR-248	GTGCTGGTGGTTGACG	17	1
sma-miR-250	ACGTGAGTATAGGGGCG	17	1
sma-miR-251	ACGGGTGTCGGGAGCGG	17	1
sma-miR-252	ACGGGTGTCGGGAGCGG	17	1
sma-miR-253	CCGCTCGACTGATGCTG	17	1
sma-miR-255	AGGGACAGACGGGGGCA	17	1
sma-miR-256	ATCTGTGCGGTGTGTCC	17	1
sma-miR-257	AACGTGAGTATAGGGGC	17	1
sma-miR-258	AGAGGCATTGGGGAGGA	17	1
sma-miR-259	CCAGGGGTCGGGAGTTC	17	1
sma-miR-260	AGATGAGTCGTGGTGCG	17	1
sma-miR-261	ATTGTGGTATGAGCGTG	17	1
sma-miR-262	CGGGGAGGTAGTGACGA	17	1
sma-miR-263	TTTTGTCACGTCGGCGG	17	1
sma-miR-265	AAGGCCAAGTGAAGGTA	17	1
sma-miR-266	AGAGGCATTGGGGAGGA	17	1
sma-miR-269	TTCGGGGTGCGTGGTCT	17	1
sma-miR-270	ACGTGAGTATAGGGGCG	17	1
sma-miR-271	ACGTGAGTATAGGGGCG	17	1
sma-miR-272	GGGTGTCGGGAGCGGAC	17	1
sma-miR-273	AGGTCGTTGGATTGGCG	17	1
sma-miR-275	ACGTGAGTATAGGGGCG	17	1
sma-miR-276	TTGGTGGGTATGTGTGA	17	1
sma-miR-277	GATGTGGGCGTGGGTTC	17	1
sma-miR-278	AGGGAACGTGAGCTGGG	17	1
sma-miR-279	AGATGAGTCGTGGTGCG	17	1
sma-miR-280	ACGTGAGTATAGGGGCG	17	1
sma-miR-282	ACGTGAGCTGGGTTTAG	17	1
sma-miR-283	CTATGCTGTGTGAGGGC	17	1
sma-miR-284	TTTGGTGGGTATGTGTG	17	1
sma-miR-285	ATGCTTCCGGATGCCTT	17	1
sma-miR-286	ATCCGGCGGGTGCGGTG	17	1
sma-miR-287	TGGGTCGTGCGGTCGCA	17	1
