

ADHEMAR ZERLOTINI NETO

**Integração e uso de dados genômicos de
Schistosoma mansoni na descoberta de
candidatos à alvos terapêuticos**

Belo Horizonte
2009



Universidade Federal de Minas Gerais
Instituto de Ciências Biológicas
Programa de Pós-graduação em Bioinformática



ADHEMAR ZERLOTINI NETO

**Integração e uso de dados genômicos de
Schistosoma mansoni na descoberta de
candidatos à alvos terapêuticos**

Projeto de tese apresentado ao Curso de Pós-graduação em Bioinformática da Universidade Federal de Minas Gerais em 10 de julho de 2009 como requisito parcial para a obtenção do grau de Doutor em Bioinformática.

Orientador: **Dr. Guilherme Oliveira**
Laboratório de Parasitologia Celular e Molecular
Centro de Pesquisas René Rachou – CPqRR
Fundação Oswaldo Cruz – FIOCRUZ
Belo Horizonte – MG – Brasil

Co-orientadora: **Dra. Jessica Kissinger**
Center for Tropical and Emerging Global Disease - CTEGD
Department of Genetics
University of Georgia - UGA
Athens – GA – EUA

Belo Horizonte
2009

Esse trabalho foi iniciado no Centro de Pesquisas René Rachou – FIOCRUZ (CPqRR-FIOCRUZ) em 2005 em colaboração com o *Center for Tropical and Emerging Global Diseases* da *University of Georgia* (CTEGD-UGA-EUA). Parte do trabalho foi desenvolvido na *University of Georgia*, durante estágio de doutoramento sanduíche realizado no período de agosto de 2007 a agosto de 2008, em colaboração com o grupo da Dra. Jessica Kissinger (CTEGD-UGA-EUA). Posteriormente, foi concluído no laboratório de Parasitologia Celular e Molecular (CPqRR-FIOCRUZ), sob a orientação do Dr. Guilherme Oliveira. O projeto conta com o suporte financeiro do *Fogarty-NIH #5D43TW007012-03*.

Agradecimentos

Ao meu amigo e orientador Dr. Guilherme Oliveira, pela feliz oportunidade ofertada para meu ingresso no meio acadêmico; por toda orientação, incentivo, cobranças e paciência;

À minha co-orientadora Dra. Jessica Kissinger, pelo apoio, incentivo e *insights*.

Aos professores que compõem o corpo docente do Programa de Pós-graduação em Bioinformática da UFMG;

Aos integrantes das secretarias da UFMG e do CPqRR, Ângela, Carlos e Maureen pela atenção constante e proatividade na busca soluções;

À Paola pela companhia, apoio e incentivo desde o início do doutorado, além da revisão do texto;

Ao Rômulo e Anderson pela companhia, amizade e por toda apoio;

Aos Amigos Fernanda, Flávio, Lívia, Luiza e Mariana, pelas várias horas de bate papo de alto nível, científico ou não;

A todos os demais integrantes do LPCM pela salutar convivência e importantes discussões científicas;

À Regina pela ajuda e parceria nos experimentos para testes *in vivo* de drogas.

Ao suporte técnico-científico provido pelo Laboratório de Malacologia e Biotério do CPqRR;

Ao grupo americano da *University of Georgia*, particularmente, Dra. Jessica Kissinger, Mark Heiges, Haiming Wang, Cristina Aurrecochea, Cary Pennington e Alan Gingle pelo apoio e parceria nesse projeto;

Aos amigos, pelas horas de descontração em meio a tantas responsabilidades;

E, principalmente, à minha família e minha filhota pelo amor, apoio, incentivo e compreensão.

E a todos aqueles que direta ou indiretamente me apoiaram neste processo, os quais infelizmente não foram citados pois renderiam uma lista infindável.

Resumo

Dentre as doenças parasitárias existentes, a esquistossomose, causada pelo parasito *Schistosoma mansoni*, é a segunda principal causa de morbidade no mundo depois da malária. No presente trabalho, foi demonstrada a criação de uma base de dados genômica de *S. mansoni* (<http://SchistoDB.net>) para integração de diferentes tipos de dados e análises. O principal objetivo para construção de tal base de dados foi selecionar candidatos a alvos terapêuticos. Dentre as diversas análises efetuadas, a predição de vias metabólicas foi utilizada como principal fator na busca de candidatos. Estes resultados serviram como incentivo para que fossem testadas *in vivo* quatro drogas com mecanismo de ação descrito para outros organismos. Dentre as quatro drogas testadas, Alopurinol, Metotrexado, Pravastatina e Trimetoprima, duas apresentaram redução no número de vermes adultos, quais sejam Alopurinol (30,5%) e Metotrexato (35,7%). Dessa forma, foi demonstrada efetividade na seleção de candidatos a alvos terapêuticos utilizando a técnica de integração de dados, além de serem disponibilizadas na *Internet* uma lista com os candidatos encontrados e uma base de dados genômica de *S. mansoni*, as quais poderão ser utilizadas como fonte de informações para diferentes pesquisas em esquistossomose.

Abstract

Schistosomiasis, caused by the parasite *Schistosoma mansoni*, is the second main cause of morbidity in the world among parasitic diseases, right after malaria. In the present research work, we describe the construction of a genomic database for *S. mansoni* (<http://SchistoDB.net>) aiming to integrate different data types and analyses. The main purpose for this database implementation was to enable data mining of candidate therapeutic targets. Among several different analyses, prediction of metabolic pathways was central towards finding candidates. These findings propelled us to conduct *in vivo* tests for drugs with known mechanisms of action. Among four tested drugs, Alopurinol, Metotrexado, Pravastatin and Trimethoprim, two resulted in decreased adult worm rates, Alopurinol 30.5% and Methotrexate 35.7%. In conclusion, we demonstrate that selecting candidate therapeutic targets by using data integration and mining was effective. In addition, a candidate list and a *S. mansoni* genome database, a rich and useful resource for the international research community, were made available online.

Sumário

I – INTRODUÇÃO.....	1
II – FUNDAMENTAÇÃO TEÓRICA.....	3
2.1 – <i>Schistosoma mansoni</i>	5
2.2 – Terapêutica da esquistossomose.....	8
2.3 – O genoma de <i>Schistosoma mansoni</i>	10
2.4 – Integração de bancos de dados na busca de candidatos a alvos terapêuticos.....	12
III – OBJETIVOS.....	14
3.1 – Objetivo geral.....	14
3.2 – Objetivos específicos.....	14
IV – MATERIAIS & MÉTODOS.....	15
4.1 – Construção do banco de dados genômico de <i>S. mansoni</i>	15
4.1.1 – Estrutura do banco de dados.....	16
4.1.2 – A camada objeto-relacional PERL	19
4.1.3 – Carga dos dados	21
4.1.4 – Análises computacionais dos dados.....	24
4.1.5 – Integração dos diferentes tipos de dados.....	27
4.1.6 – Publicação dos dados.....	29
4.2 – Predição das vias metabólicas com base na anotação dos genes.....	32
4.3 – Seleção dos candidatos a alvos terapêuticos com base na anotação dos genes.....	34
4.4 – Busca na literatura sobre informações dos genes selecionados.....	35
4.5 – Validação experimental dos melhores candidatos.....	36
V – RESULTADOS & DISCUSSÃO.....	37
5.1 – Banco de dados genômicos de <i>S. mansoni</i>	37
5.2 – Predição das vias metabólicas de <i>S. mansoni</i>	44
5.3 – Seleção de candidatos a alvos terapêuticos.....	48
5.4 – Busca na literatura por informações sobre os genes selecionados.....	52
5.5 – Validação experimental dos melhores candidatos.....	56
VI – CONCLUSÕES.....	59
VII – LIMITAÇÕES & PERSPECTIVAS.....	60
VIII – REFERÊNCIAS BIBLIOGRÁFICAS	61
IX – ANEXOS.....	70
9.1: Relação de enzimas que catalizam reações <i>chokepoint</i>	70
9.2: Tabela de frequência da co-ocorrência dos termos pesquisados na literatura.....	73
9.3: Tabela contendo os artigos analisados.....	76
9.4: Artigo – SchistoDB: a <i>Schistosoma mansoni</i> genome resource.....	78
9.5: Artigo – The genome of the blood fluke <i>Schistosoma mansoni</i>	84

Lista de tabelas

Tabela 1: Vocabulários controlados e ontologias.....	21
Tabela 2: Bases de dados de sequências e anotações de <i>S. mansoni</i>	22
Tabela 3: Bases de dados utilizados para análises computacionais	24
Tabela 4: Relação de enzimas que catalizam reações <i>chokepoint</i> e que são alvos de droga com base no Kegg Drugs.....	49
Tabela 5: Relação de drogas que atuam em enzimas que catalizam reações <i>chokepoint</i> , nome de identificação no FDA e ação farmacológica.....	50
Tabela 6: Máxima dosagem encontrada na literatura científica.....	56
Tabela 7: Resultado do experimento de eficácia das drogas em <i>S. mansoni</i>	57

Lista de Figuras

Figura 1: Distribuição global da esquistossomose.....	3
Figura 2: Ciclo biológico de <i>S. mansoni</i>	7
Figura 3: GUS - Diagrama entidade-relacional do dogma central da biologia	16
Figura 4: Seção do arquivo XML de carga dos dados.....	23
Figura 5: Seção do arquivo de execução e carregamento de análises	26
Figura 6: Seção do arquivo XML de configuração do WDK.....	30
Figura 7: Página inicial do banco de dados SchistoDB	37
Figura 8: Exemplo de relatório dos genes	39
Figura 9: Capturas de tela do SchistoDB mostrando consultas combinadas.....	41
Figura 10: Estatísticas de acesso ao SchistoDB a partir de 2008.....	43
Figura 11: Visão geral das vias metabólicas contidas no SchistoCyc.....	45
Figura 12: Exemplo da predição da via glicolítica pelo PathwayTools.....	46
Figura 13: Exemplo de reações <i>chokepoint</i>	48
Figura 14: Diagrama da via de mevalonato que contém a enzima HMG-CoA redutase (1.1.1.34).....	53
Figura 15: Diagrama da via de biossíntese de catecolaminas que contém a enzima LDOPA descarboxilase (4.1.1.28).....	54
Figura 16: Diagrama da via de degradação de triptofano que contém a enzima quinureninase (3.7.1.3).....	55

Lista de Abreviaturas

AA: Aminoácidos
API: *Application Programming Interface*
C: Celcius
CDS: *CoDing Sequence*
CNPq: Conselho Nacional de Desenvolvimento Científico e Tecnológico
DNA: Ácido Desoxirribonucleico
EST: *Expressed Sequence Tag*
EC Number: *Enzyme Commission Number*
FAPEMIG: Fundação de Amparo à Pesquisa do Estado de Minas Gerais
FAPESP: Fundação de Amparo à Pesquisa do Estado de São Paulo
FDA: *US Food and Drug Administration*
GAO: *US Government Accountability Office*
GC: guanina e citosina
GO: *Gene Ontology*
GUS: *Genomics Unified Schema*
HTML: *HyperText Markup Language*
JSP: *Java Server Pages*
h: hora
kbp: *kilo base pairs* (mil de pares de base)
mbp: *million base pairs* (milhões de pares de base)
MCT: Ministério da Ciência e Tecnologia
NA: Nucleotídeos
NCBI : *National Center for Biotechnology Information*
NIH: *National Institutes of Health*
NR: *Non Redundant Protein Sequence Database*
OMS: Organização Mundial de Saúde
ORESTES: *Open Reading Frames ESTs* (ESTs de Janelas Abertas de Leitura)
PZQ: Praziquantel
RCSB PDB: *Research Collaboratory for Structural Bioinformatics – Protein Data Bank*
RNA: Ácido ribonucleico
rRNA: RNA ribossomal
SQL: *Structured Query Language*
TI: Tecnologia da Informação
TIGR: *The Institute for Genomic Research*
tRNA: RNA transportador
XML: *eXtensible Markup Language*
WDK: *Web Development Kit*
WGS: *Whole Genome Shotgun*
WTSI: *Welcome Trust Sanger Institute*

I – INTRODUÇÃO

A esquistossomose é uma doença parasitária que infecta principalmente indivíduos de países em desenvolvimento tropicais e subtropicais. O parasito causador da doença pertence ao gênero *Schistosoma*, sendo *S. mansoni* a única espécie causadora da esquistossomose no Brasil (Bergquist, 2002). Atualmente, as estratégias de controle da esquistossomose dependem quase que totalmente da utilização de uma única droga disponível para tratamento em massa, o praziquantel (PZQ) (Chitsulo et al., 2000). O tratamento é eficaz, administrado em dose única por via oral e tem resultado satisfatório na diminuição da morbidade. Porém, a prevalência da doença permanece praticamente inalterada em muitas regiões endêmicas devido, na maioria dos casos, aos altos níveis de reinfecção de indivíduos já tratados (WHO | Schistosomiasis. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs115/en/index.html>>. Acesso em 1 de abril de 2009). Adicione-se a isso, o fato de o PZQ não ser eficaz contra todos os estágios de desenvolvimento presentes no hospedeiro humano, cujas evidências demonstram diferentes graus de resistência (Pica-Mattoccia, 2004b). A Organização Mundial de Saúde tem incentivado o estudo da genômica como uma nova abordagem para o desenvolvimento de medidas de controle do parasito. Tal abordagem possibilita um entendimento mais profundo da biologia destes organismos, de forma a auxiliar no processo de busca por novos candidatos para diagnóstico da doença, vacinas e alvos de drogas (Degraeve et al., 2001).

O processo de descoberta de novos compostos é complexo e multidisciplinar, o qual requer conhecimento sobre a biologia e genoma do organismo, química e metabolismo de drogas e aspectos clínicos da doença. A abordagem genômica é uma das principais

responsáveis pela explosão no volume de dados gerados por pesquisadores, evidenciada pelos bancos de dados e literatura científicos. A expectativa é que tamanha quantidade de dados pudesse levar a um aumento significativo no número de drogas descobertas. Não obstante, em uma revisão do processo de descoberta de drogas produzido pelo GAO – *US Government Accountability Office* (US Government Accountability Office, 2006), foi relatado que novas tecnologias como genômica e *high-throughput screening* apresentaram aumento significativo dos custos de investigação sem aumento correspondente no número de drogas disponíveis.

Nesse contexto, a técnica de integração de dados pode otimizar a exploração desses recursos no intuito de melhor selecionar candidatos a alvos terapêuticos (Loging et al., 2007). Historicamente, a tecnologia de integração de dados tem sido considerada um problema de Tecnologia da Informação - TI, o que subestima o desafio científico e gerencial que realmente representa (Searls, 2005).

Esse projeto busca demonstrar tanto a eficácia da utilização da tecnologia de integração de dados quanto possíveis novas drogas para o combate à esquistossomose. São descritas abordagens computacionais, direcionadas para a identificação de hipóteses passíveis de comprovação em laboratório, especificamente, a busca de candidatos a alvos terapêuticos contra *S. mansoni*.

II – FUNDAMENTAÇÃO TEÓRICA

Dentre as doenças parasitárias existentes, a esquistossomose é a segunda principal causa de morbidade no mundo depois da malária (Chitsulo et al., 2000). De acordo com a Organização Mundial de Saúde, a esquistossomose é endêmica em 74 países sendo que 80% dos indivíduos infectados vivem na África (Figura 1) (TDR, 2005). Estimativas sugerem que mais de 600 milhões de indivíduos estejam sob risco de infecção e que existam 200 milhões de infectados. Por ano, ocorrem aproximadamente 11 mil mortes diretamente relacionadas à esquistossomose. A doença atinge principalmente crianças e adolescentes com faixa etária entre 10 e 19 anos. Quando infectadas, crianças têm seu potencial cognitivo consideravelmente diminuído, podendo haver comprometimento da memória e da capacidade de processar dados, com efeito ainda mais grave naquelas subnutridas (Ezeamama et al.,

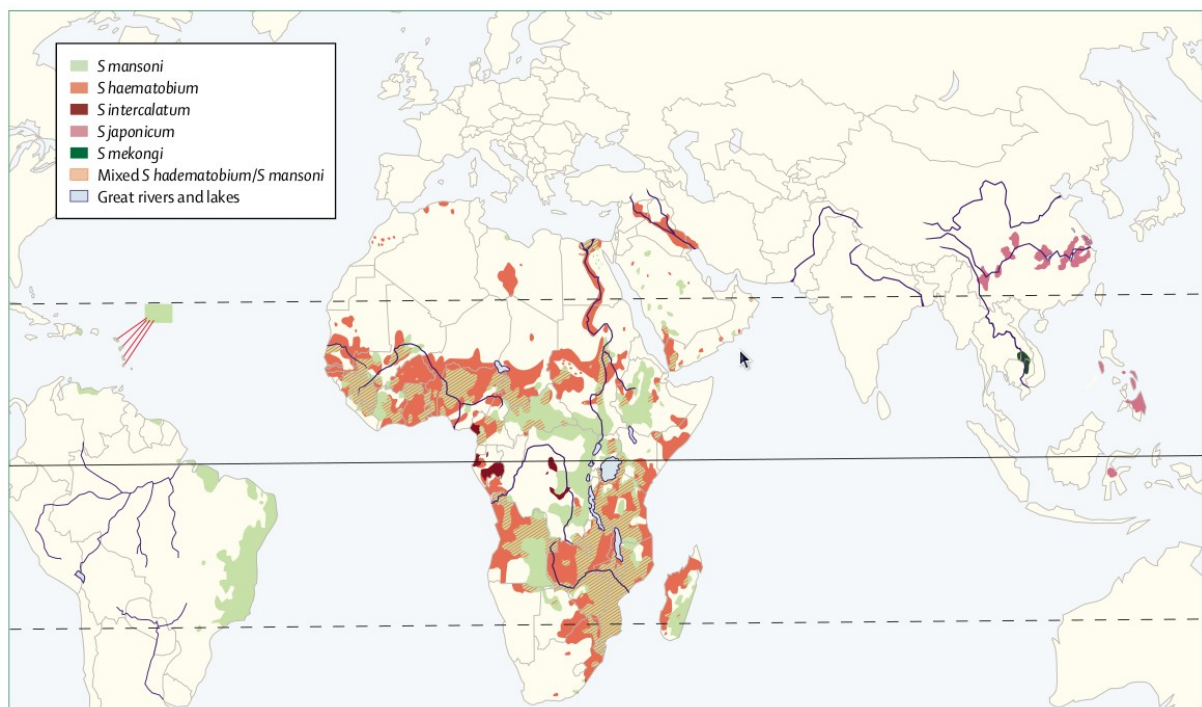


Figura 1: Distribuição global da esquistossomose

S. mansoni: África subsaariana, Brasil, Suriname, Venezuela, Caribe, Egito e Península Árabe; *S. haematobium*: África subsaariana, Egito, Sudão, Magrebe e Península Árabe; *S. japonicum*: China; Mindanau, província de Leyte e algumas ilhas das Filipinas e Indonésia; *S. mekongi*: Laos e Camboja; *S. intercalatum*: África ocidental e central (Gryseels et al., 2006).

2005; Jukes et al., 2002).

A esquistossomose acomete populações pobres que vivem perto de coleções de água contaminada pelo parasito, particularmente lagos, rios e canais de irrigação (Steinmann et al., 2006). Para essas populações, a água é necessária para atividades diárias como cozinhar e lavar ou para fins de geração de renda como agricultura e pesca. As doenças tropicais acometem, principalmente, países subdesenvolvidos, onde os recursos de saúde pública são muito escassos e voltados primariamente para outras doenças como a malária, síndrome da imunodeficiência adquirida -AIDS e tuberculose (TDR, 2005).

Dentre as espécies do gênero *Schistosoma* que causam esquistossomose no homem, as principais são *S. mansoni*, *S. japonicum* e *S. haematobium*. Outras espécies, tais como *S. intercalatum*, *S. mekongi*, *S. mattheei* e *S. malayensis*, também tem sido relatadas como infectantes (McManus et al., 1993).

A espécie *S. mansoni* ocorre principalmente na África e América Latina (Figura 1) (Gryseels et al., 2006). A infecção causada por essa espécie é denominada esquistossomose mansônica ou intestinal, devido à localização dos parasitos nas vênulas da parede do intestino grosso, sigmóide e reto, com sintomas predominantemente intestinais (TDR, 2005). Estima-se que existam mais de 6 milhões de pessoas infectadas pela esquistossomose mansônica no Brasil, o que o torna o país mais afetado na América Latina (Katz et al., 2000).

Sendo *S. mansoni* foco do nosso estudo, faz-se necessário tecer algumas considerações acerca de sua biologia, aspectos da terapêutica da doença e genoma do parasito para que, finalmente, a integração de banco de dados na busca de alvos terapêuticos seja demonstrada.

2.1 – *Schistosoma mansoni*

O trematódeo *S. mansoni* possui um ciclo biológico que alterna entre hospedeiros intermediários e definitivos. O seu hospedeiro intermediário é o caramujo do gênero *Biomphalaria* (Davis, 1984). Nele, milhares de cercárias são produzidas por um único esporocisto durante a reprodução assexuada, as quais são liberadas de forma intermitente. A saída das cercárias do hospedeiro intermediário é induzida pela luz, ocorrendo, preferencialmente, nas horas mais claras do dia. Quando madura, a cercária sai do caramujo para a água apta a infectar hospedeiros vertebrados. Uma vez que não se alimentam, as cercárias precisam encontrar um hospedeiro vertebrado dentro de um período de 6 a 8 horas (Davis, 1984), e quando o encontram, penetram pela sua pele ou mucosa, liberando a cauda, onde se inicia, então, o processo de transformação em esquistossômulos.

Após permanecerem na pele por aproximadamente 72h, os esquistossômulos iniciam o processo de migração através do corpo de seu novo hospedeiro, caso não sejam destruídos pelos mecanismos de defesa do mesmo. Por meio da circulação, os esquistossômulos chegam ao coração direito e em seguida aos pulmões em torno de 4 dias, tornando-se mais longos e delgados, o que facilita a sua migração através da rede vascular pulmonar. Do pulmão, os esquistossômulos voltam ao coração e são enviados pela circulação geral a todas as partes do corpo do hospedeiro. Somente quando alcançam o sistema porta intra-hepático, são capazes de completar seu desenvolvimento (Rollinson, 1988). Quatro semanas após a infecção, a maioria dos vermes encontra-se adulta e pronta para se acasalar. Acasalados, deslocam-se ativamente contra a corrente circulatória do sistema porta e migram para as veias mesentéricas pélvicas. Os vermes adultos tem, aproximadamente, um centímetro em comprimento, são delgados, longos e caracterizados por terem duas ventosas (Rollinson, 1988). A fêmea é mais

fina, cilíndrica e longa que o macho, existindo também diferenças fisiológicas e antigênicas entre macho e fêmea (Aronstein et al., 1984). A fêmea depende do contato com o macho para completar a sua maturação (Loverde et al., 1991). O par permanece em constante associação e a fêmea se localiza no canal ginecóforo do macho. Cerca de 340 ovos de *S. mansoni* são liberados por casal de vermes por dia em ratos infectados, sendo este número maior em primatas (Cheever et al., 1994). Os vermes alimentam-se de plasma e células do sangue venoso do hospedeiro definitivo (Rollinson, 1988).

A produção de ovos começa 30 a 40 dias após a infecção. No momento da oviposição, o desenvolvimento embrionário é ainda incompleto, requerendo mais seis ou sete dias para completar-se. A expectativa de vida dos ovos maduros é de, aproximadamente, 20 dias. Os ovos são eliminados no ambiente por meio das fezes de indivíduos infectados (Rollinson, 1988). Os ovos eliminados nas fezes do hospedeiro eclodirão, se encontrarem condições adequadas como água fresca, temperatura morna, baixa hipotonicidade e iluminação adequada, liberando assim o miracídio de dentro da casca do ovo (Rollinson, 1988). Uma vez que o ovo é rompido em água fresca, o miracídio emerge e começa a nadar ativamente. Os miracídios morrem caso a expulsão não se complete dentro de três semanas após a oviposição. O miracídio penetra no hospedeiro intermediário específico por movimentos rotatórios e ação lítica (Rollinson, 1988). Por volta do oitavo dia, o miracídio apresenta-se como um tubo enovelado, imóvel, repleto de células germinativas em multiplicação, processo pelo qual se transforma em esporocisto. Por volta da segunda semana de existência, rompem-se pra liberar esporocistos filhos, em número de 20 a 40. A transformação dos esporocistos em cercárias ocorre apenas quando atingem sua localização permanente nas glândulas digestivas do caramujo. A cercária deixa o hospedeiro e, caindo em água fresca, fecha o ciclo de vida do parasito (Rollinson, 1988) (Figura 2).

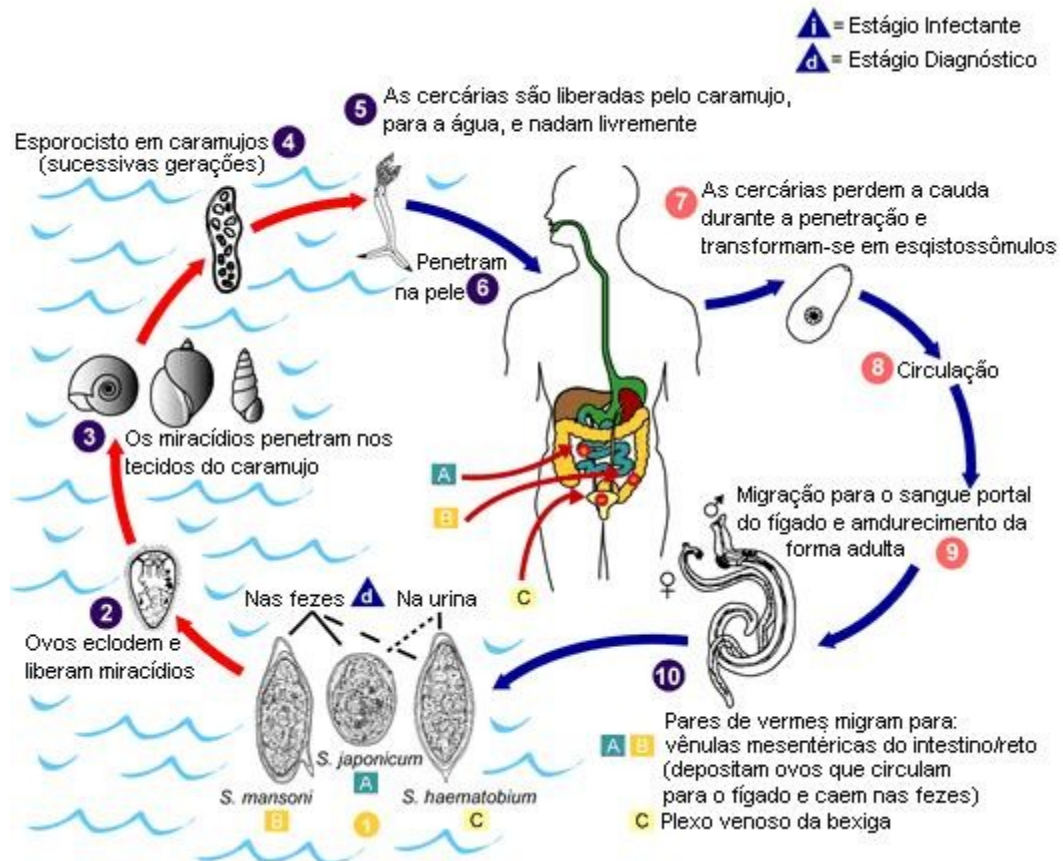


Figura 2: Ciclo biológico de *S. mansoni*

(1) Sob condições ótimas, os ovos eclodem e liberam os miracídeos (2) que nadam e penetram no caramujo, hospedeiro intermediário específico (3). Os estágios no caramujo incluem duas gerações de esporocistos (4) e a produção de cercárias (5). Ao abandonar o caramujo, as cercárias infectantes nadam, penetram na pele do hospedeiro humano (6), e perdem sua cauda bifurcada e tornam-se esquistossômulos (7). Os esquistossômulos migram através de diversos tecidos e desenvolvem-se até atingirem a veia porta onde se acasalam (8, 9). Vermes adultos, nos humanos, residem nas vênulas mesentéricas em várias localizações (10). *S. mansoni* ocorre mais frequentemente nas veias mesentéricas superiores que drenam o intestino grosso (B). As fêmeas depositam ovos nas pequenas vênulas dos sistemas porta e perivesical. Os ovos são movidos progressivamente para o lúmen do intestino e são eliminados com as (d) fezes (1) (INFORME-NET DTA *Schistosoma mansoni*/Esquistossomíase. Disponível em: <http://www.cve.saude.sp.gov.br/htm/hidrica/IF_ESQUII05.htm>. Acessado em: 6 de abril de 2009).

Muitos dos ovos de *S. mansoni* não ultrapassam a parede intestinal do hospedeiro definitivo, sendo levados a órgãos e tecidos pela corrente sanguínea. A presença destes ovos nos tecidos resulta na formação de granulomas, resultantes da resposta do sistema imune do hospedeiro. A resposta de hipersensibilidade retém antígenos citotóxicos ao redor dos ovos, mas causa danos ao tecido, sendo o principal responsável pela manifestação clínica da esquistossomose crônica (Katz et al., 1986).

2.2 – Terapêutica da esquistossomose

Atualmente, o controle da esquistossomose é feito por meio da utilização de drogas e tem sido considerado uma medida de controle eficaz, embora existam altos índices de reinfeção em áreas endêmicas. A importância das drogas não se limita ao tratamento da infecção e controle da morbidade; são também utilizadas para controle da transmissão da doença. A OMS recomenda a utilização do PZQ e oxamniquina para controle da esquistossomose, sendo o praziquantel a droga de escolha (WHO | Schistosomiasis. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs115/en/index.html>>. Acesso em 1 de abril de 2009).

A opção da Organização Mundial de Saúde pelo PZQ se deve à sua eficácia contra todas as espécies de *Schistosoma* que infectam o homem, sua alta taxa de cura, baixa toxicidade e baixo custo. O PZQ é um pó de cor branca, sabor amargo e sua temperatura de fusão varia entre 136° e 140° C. É um composto estável em condições normais e praticamente insolúvel em água, porém solúvel em etanol e em alguns solventes orgânicos como o clorofórmio (Cioli et al., 2003). O pico de ação do PZQ ocorre entre 1 e 2 horas após administração e sua eliminação ocorre essencialmente por meio da urina e fezes, sendo que mais de 80% da droga são eliminados após 24 h (Cioli et al., 2004).

O tratamento com PZQ é feito em dose única por via oral, com dosagem recomendada entre 40 e 60 mg/Kg. O percentual de cura da doença causada por *S. mansoni* é de 60% a 90%, causando redução do número de ovos de 90% a 95%, dependendo do nível de infecção (de Silva et al., 1997). A eficácia do PZQ depende da idade e sexo do parasito, sendo os vermes fêmeas menos sensíveis ao PZQ (Pica-Mattoccia et al., 2004a). O tratamento com PZQ pode causar efeitos colaterais transitórios como vômito, cólicas e diarreia. A

frequência e a gravidade dos efeitos colaterais produzidos pelo PZQ estão diretamente relacionadas com a intensidade da infecção (Stelma et al., 1995).

A maior desvantagem do PZQ é a sua baixa eficácia contra as formas jovens de *Schistosoma* (Gryseels et al., 2001). Algumas alternativas tem sido propostas para solucionar tal problema, como a associação de PZQ com artemeter, uma droga efetivamente utilizada no tratamento da malária e ativa contra as formas juvenis de schistossomas (Utzinger et al., 2001).

Evidências acerca da resistência do parasito ao PZQ tem levantado questões sobre o uso da droga, bem como de outras para o tratamento da esquistossomose, o que poderia torná-las inúteis ao longo do tempo. Investigações recentes demonstraram casos de parasitas resistentes ao PZQ no Senegal e Egito, bem como de resistência induzida em condições de laboratório (Doenhoff et al., 2008) incluindo evidências relatadas pelo nosso grupo (dados não publicados).

Existem, ainda, grupos de pesquisa envolvidos no desenvolvimento de vacinas contra a esquistossomose. A possibilidade de desenvolvimento de uma vacina contra a esquistossomose já foi demonstrada em camundongos (Aitken et al., 1987) e em babuínos (Kariuki et al., 2004). Na tentativa de repetir esses resultados com antígenos recombinantes, pesquisadores desenvolveram vários candidatos à vacina, mas nenhum produziu um nível alto ou completo de proteção (Wilson et al., 2006). Ainda que muito tenha se aprendido sobre o a resposta imune contra esquistossomose, métodos tradicionais não geraram os resultados esperados e novas abordagens serão necessárias para o desenvolvimento de vacinas contra o *S. mansoni* (Hokke et al., 2007; Loukas et al., 2007; Wilson et al., 2004).

Tendo em vista os novos dados gerados por pesquisas sobre o genoma deste parasito e investigações realizadas à luz de abordagens computacionais, novas medidas de controle poderão ser encontradas.

2.3 – O genoma de *Schistosoma mansoni*

O projeto genoma de *S. mansoni* foi iniciado pela OMS em 1994 com a premissa de que a descoberta de genes, desenvolvimento de mapas cromossômicos, sequenciamento genômico completo e análise genômica seriam as abordagens mais promissoras para a identificação de novos alvos para drogas, vacinas e ferramentas diagnósticas.

O parasito *S. mansoni* tem um genoma haploide de aproximadamente 363 mbp contidos em 7 pares de cromossomos autossômicos e um par de cromossomos sexuais Z e W (Berriman et al., 2009).

O sequenciamento de transcritos de *S. mansoni* iniciou-se em 1992 por meio de um projeto colaborativo entre instituições brasileiras, financiado por agências nacionais (Franco et al., 2000). Posteriormente, dois grandes projetos brasileiros de sequenciamento do transcriptoma de *S. mansoni* foram realizados. A partir de 1994, a OMS financiou a Rede Genoma de *Schistosoma* para descoberta de novos genes com o objetivo de identificar novos alvos para o desenvolvimento de drogas e vacinas. Durante este período, a comunidade científica mundial produziu aproximadamente 16.000 ESTs (Oliveira et al., 2001)

No que concerne aos projetos de sequenciamento de transcritos de *S. mansoni*, o primeiro, que foi financiado pela FAPESP/MCT/CNPq, utilizou uma biblioteca normalizada de verme adulto e minibibliotecas de ORESTES de seis estágios do ciclo de vida do parasito (Verjovski-Almeida et al., 2003). O projeto gerou 124.681 ORESTES e ESTs de *S. mansoni*. As sequências foram agrupadas resultando em 30.988 SmAE (*Schistosoma mansoni assembled ESTs*). O segundo projeto, financiado pela FAPEMIG/MCT/CNPq, consistiu de uma rede genômica formada por instituições do Estado de Minas Gerais que caracterizou o transcriptoma de diferentes estágios de desenvolvimento do parasito a partir da geração de,

aproximadamente, 42.500 ESTs convencionais (Oliveira, 2007).

A iniciativa internacional para o sequenciamento do genoma completo do organismo foi conduzida pelo instituto *The Institute for Genomic Research* - TIGR em associação com o *Welcome Trust Sanger Institute* – WTSI, por meio de financiamento do *National Institutes of Health* - NIH e da *Welcome Trust*, respectivamente (LoVerde et al., 2004). As últimas versões do sequenciamento genômico e todas as análises efetuadas estão disponíveis nos bancos de dados *online* GeneDB (GeneDB. Disponível em: <<http://genedb.org>>. Acessado em: 10 de maio de 2009) e SchistoDB (Zerlotini et al., 2009), cujas informações foram recentemente aceitas para publicação pelo periódico *Nature* (Berriman et al., 2009).

A sequência do genoma nuclear de *S. mansoni* foi obtida por meio da metodologia WGS (*Whole Genome Shotgun*). Essa técnica consiste na quebra de DNA genômico em pequenos fragmentos que foram selecionados por tamanho em gel de agarose. Os fragmentos foram agrupados em 5745 *scaffolds* com tamanho superior a 2kbp, totalizando 363mbp. Além disso, foram identificados 11812 genes que codificam 13162 transcritos. Apesar de 45% do genoma ser composto de elementos repetitivos, 50% das bases estão presentes em contigs de tamanho maior que 16,3 kbp e em *scaffolds* maiores que 824,5kbp. Ainda, a localização cromossomal de 43% da montagem genômica utilizando-se hibridização *in situ* foi identificada (Berriman et al., 2009)

De posse de dados gerados por diferentes técnicas de biologia molecular, abordagens computacionais podem ser utilizadas com o intuito de se integrarem diferentes tipos de dados e, assim, possibilitar análises mais aprofundadas.

2.4 – Integração de bancos de dados na busca de candidatos a alvos terapêuticos

A descoberta de drogas é um processo complexo dividido em muitas etapas que englobam diferentes áreas como Biologia e Química, além de exigir conhecimento sobre metabolismo e observação clínica na relação parasito-hospedeiro. Todas essas áreas tem produzido grandes quantidades de dados, particularmente, a área genômica. A grande variedade de recursos *in silico* disponíveis para a pesquisa na área biológica é vista como promissora para promover o processo de descoberta de drogas. A expectativa era de que essa explosão de dados fosse levar a um aumento proporcional no número de drogas disponíveis; entretanto, tal correlação não foi observada (US Government Accountability Office, 2006).

Para atingir esse objetivo, cientistas da computação têm utilizado a técnica de integração de dados. Estão em foco técnicas de análise em alta escala que podem ser iniciadas independentemente de experimentos em laboratório e aplicadas para várias doenças. A utilização dessas abordagens no processo de descoberta de drogas tem evidenciado a contribuição das técnicas *in silico* e servem também para destacar a necessidade de colaboração com as áreas biológicas (Loging et al., 2007).

Visto que o processo de descoberta de drogas é multidisciplinar, podemos exemplificar diferentes tipos de dados que se inter-relacionam e podem ser integrados por meio da utilização do gene como chave principal (Searls, 2005). Diferentes resultados de análise de expressão de genes como microarranjo, SAGE, gel 2D ou espectrometria de massa são frequentemente relacionados utilizando-se identificadores de genes. Outro exemplo, seria o produto gênico que pode ser relacionado em cascata com uma enzima, a enzima com uma reação e a reação com uma droga inibidora.

A falta de padronização é um dos fatores que mais dificultam a integração de dados

provenientes de diferentes experimentos e grupos de pesquisa. Grande parte do trabalho de um bioinformata consiste em converter os diferentes formatos de dados para que sejam efetivamente integrados. A criação de vocabulários controlados (Eilbeck et al., 2005) e ontologias (Harris et al., 2004) tem contribuído consideravelmente para a integração de dados. A terminologia biológica é frequentemente ambígua. Uma mesma palavra pode ser utilizada para descrever mais de um objeto. Ontologias possibilitam conceitualização de informações e facilitam tanto a comunicação entre pesquisadores quanto a utilização dessas informações por computadores.

A integração de dados é um elemento-chave na condução de investigações científicas das atuais plataformas de pesquisa, que possibilita processar análises complexas. Contudo, considerar a técnica de integração de dados como exclusiva da área de tecnologia da informação subestima o desafio científico e gerencial que realmente representa, uma vez que requer mudanças metodológicas e culturais significativas nas abordagens de análise de dados (Searls, 2005).

No presente trabalho, foi efetuada a seleção de candidatos a alvos terapêuticos de *S. mansoni* utilizando-se a técnica de integração de dados. Para verificação da eficácia dessa metodologia, foram efetuadas buscas na literatura científica sobre os alvos encontrados e testes de drogas. Os achados nos incentivaram a testar quatro drogas, cujos mecanismos de ação eram conhecidos, para que sua atividade esquistossomicida fosse avaliada *in vivo*. Foram disponibilizadas na *Internet* uma lista com os candidatos encontrados e uma base de dados genômica de *S. mansoni*, as quais poderão ser utilizadas como fonte de informações para diferentes pesquisas em esquistossomose.

III – OBJETIVOS

3.1 – Objetivo geral

Identificar genes de *S. mansoni* candidatos a alvos terapêuticos e novos compostos por meio da integração e mineração dos dados genômicos do parasito.

3.2 – Objetivos específicos

1. Criar um banco de dados com as informações do genoma do parasito;
2. Predizer vias metabólicas com base no proteoma predito;
3. Selecionar candidatos a alvos terapêuticos com base na anotação dos genes;
4. Validar experimentalmente os melhores genes candidatos a alvos terapêuticos.

IV – MATERIAIS & MÉTODOS

4.1 – Construção do banco de dados genômico de *S. mansoni*

Para construir a base de dados, foi escolhido o esquema de integração 'Genomics Unified Schema' (GUS – <http://gusdb.org>) (Davidson et al., 2001). O GUS é um esquema de tabelas que contém um *framework* de aplicação integrado que foi desenvolvido para armazenar, integrar, analisar e publicar dados genômicos. Nesse esquema de tabelas podem ser incluídos dados genômicos, de expressão gênica, de agrupamentos de transcritos, proteômicos, entre outros. Além disso, o esquema GUS utiliza vocabulário controlado e ontologias para garantir integridade de referencial dos tipos de dados. O *framework* de aplicação foi desenvolvido em PERL (About Perl – perl.org. Disponível em: <<http://www.perl.org/about.html>>. Acessado em: 6 de abril de 2009) e oferece mapeamento objeto-relacional e uma API (*Application Programming Interface*), utilizada para facilitar a criação de programas para carga dos dados. O GUS conta ainda com um *kit* de desenvolvimento *web* (WDK – *Web Development Kit*) desenvolvido utilizando a tecnologia JSP (JavaServer Pages Technology. Disponível em: <<http://java.sun.com/products/jsp/>>. Acessado em: 6 de abril de 2009) para publicação dos dados.

A documentação para o GUS, a API para carga dos dados e o WDK estão disponíveis na internet (GUS: The Genomics Unified Schema. Disponível em: <<http://gusdb.org>>. Acessado em: 7 de maio de 2009) e nas próximas sessões serão descritos tópicos chave na implementação de um banco de dados genômico.

O GUS possui uma grande variedade de tabelas relacionadas ao armazenamento de sequências e suas características. Dentre essas tabelas, duas são utilizadas para o armazenamento de sequências (DoTS::NASequence e DoTS::AASequence) e duas para características das sequências (DoTS::NAFeature e DoTS::AAFeature). Como o esquema NA (nucleotídeos) e AA (aminoácidos) são conceitualmente similares, será descrito apenas o NA.

As tabelas pertencentes ao esquema NA são usadas para armazenar sequências nucleicas. Sequências de nucleotídeos podem atingir tamanhos consideravelmente grandes, portanto foram adotadas medidas especiais para manipulação deste tipo de dado tanto na camada objeto-relacional quanto na estrutura das tabelas do banco de dados.

As características são ligadas às suas sequências através do identificador (*na_sequence_id*). Cada sequência pode ter várias características, como por exemplo, os exons. O GUS contém uma tabela adicional para a especificação da localização de características dentro da sequência (DoTS::NALocation). Uma determinada característica pode estar ligada a múltiplas localizações, o que indica que ela pode ser encontrada em todas as localizações especificadas.

Em um banco de dados genômico completo existirão ligações representando o dogma central da biologia. As principais tabelas que representam esse relacionamento são DoTS::GeneFeature, DoTS::ExonFeature, DoTS::RNAFeature e DoTS::TranslatedAAFeature, para armazenamento de características de genes, exons, RNA e proteínas, respectivamente.

As características 'exons' pertencem diretamente a uma única característica 'gene'. Se existem múltiplas formas de *splicing*, então alguns exons podem se sobrepor. Porém não devem existir exons que possuem posições iniciais e finais idênticas associados ao mesmo gene. Podemos então notar que a posição inicial do gene é a mesma posição inicial do seu

primeiro exon e a posição final do gene, a mesma posição final do seu último exon. Se existirem informações adicionais sobre promotores, sítios de iniciação de transcrição, etc., então a posição inicial do gene pode ser expandida para também abranger essas características.

Cada RNA pertence a um único gene. O conjunto específico de exons que compõem um RNA são ligados através da tabela `DoTS::RnaExonFeature` e sua localização é determinada através da posição inicial do primeiro exon e da posição final do último exon. A camada objeto-relacional permite a construção de uma sequência de RNA (ou proteína) dinamicamente a partir da sequência genômica, porém é recomendado o armazenamento da cópia da sequência.

Por último, seguem algumas informações sobre a tradução de RNA para proteína. Muitas proteínas são formadas a partir da tradução direta de RNA para proteína. O GUS suporta esse processo no qual a tabela `DoTS::TranslatedAAFeature` é ligada à tabela `DoTS::RNAFeature` para indicar qual RNA originou aquela sequência.

A tabela `DoTS::NALocation` contém várias colunas que são usadas para armazenar informações presentes no formato utilizado pelo GenBank (Benson et al., 2000). Nessa tabela, as colunas mais importantes são: `start_min`, `start_max`, `end_min` e `end_max`. Esses campos são utilizados para indicar a posição de uma determinada característica em uma sequência e dessa forma é permitido uma certa 'incerteza'. Uma posição só é considerada exata se o campo `start_min` e `start_max` contém exatamente o mesmo valor. Quando um dos campos é deixado em branco ou contém um valor divergente do outro, a posição exata pode estar em qualquer lugar entre 'min' e 'max'. As posições são sempre especificadas na fita positiva, portanto a posição inicial de uma determinada característica sempre será menor que a sua posição final. A direção da fita é determinada através da utilização do campo 'is_reversed'.

4.1.2 – A camada objeto-relacional PERL

A camada de objetos PERL oferece mapeamento objeto-relacional do banco de dados. Dessa forma, o acesso ao banco de dados é feito através da utilização de classes que mapeiam todas as tabelas e, principalmente, obedecem aos relacionamentos garantido de certa forma a integridade da base. A eficiência dessa camada para a extração dos dados é limitada, por isso foi criada uma segunda camada para publicação dos dados.

Uma das grandes vantagens da utilização dessa camada objeto-relacional é um sistema embutido de proveniência e controle de versões. Toda vez que um *plugin* PERL é utilizado para o armazenamento de um determinado conjunto de dados, registros adicionais são também armazenados em tabelas específicas que identificam qual o *plugin* e sua versão, os parâmetros utilizados, a data e a hora da execução, qual usuário executou, entre outras informações.

Ao longo deste projeto, vários *plugins* para carga de dados foram desenvolvidos. Existem dois repositórios principais de *plugins*: o *Supported* que contém *plugins* desenvolvidos pela equipe do GUS e o *Community* que contém *plugins* desenvolvidos pela comunidade de usuários. Existe ainda um repositório adicional de *plugins* chamado ApiDB (ApiDB Repository. Disponível em: <<https://www.cbil.upenn.edu/svn/apidb>>. Acessado em: 7 de abril de 2009) que é mantido pelo grupo que desenvolve o EuPathDB (EuPathDB : The Eukaryotic Pathogen genome resource. Disponível em: [.http://eupathdb.org](http://eupathdb.org)>. Acessado em: 7 de abril de 2009). Nesses repositórios podem ser encontrados *plugins* capazes de carregar dados contidos em arquivos que utilizam os principais formatos conhecidos, como FASTA, EMBL e Genbank, além de resultados de análises computacionais como BLAST, BLAT e InterProScan.

Após a fase inicial de planejamento dos dados que serão armazenados, normalmente se constrói um arquivo no formato XML que especifica a localização dos dados, o *plugin* que será utilizado para cada tipo de dado e seus respectivos parâmetros. Uma rotina que interpreta esse arquivo XML é executada e durante esse processo são gerados *logs* individuais para cada um dos *plugins*, contendo as saídas padrão e eventuais mensagens de erro na execução de cada um deles.

Todos os arquivos de dados são replicados em um repositório local que permanece intocável e também são gerados *logs* durante o processo de criação desse repositório. Dessa forma, tem-se uma cópia local de cada conjunto de dados carregado que pode ser utilizada para fins diversos como replicação do banco e auditoria. O repositório criado pode também ser útil na eventual criação de um banco de dados de outro organismo, já que os arquivos de dados referentes a ontologias e bases de dados externas de proteínas, enzimas, estruturas, entre outros estarão disponíveis localmente.

4.1.3 – Carga dos dados

A primeira etapa da construção da base de dados SchistoDB foi a importação de vocabulários controlados e ontologias (Tabela 1). Os vocabulários controlados são utilizados para garantir integridade referencial. Por exemplo, ao invés de armazenar o nome da espécie de um determinado organismo, é armazenado o código da espécie obtido da base de taxonomia. As bases de dados de ontologia são utilizadas com o mesmo propósito; por exemplo, não são armazenados os tipos de sequência, mas os códigos dos tipos encontrados na base de ontologia de sequências.

Tabela 1: Vocabulários controlados e ontologias

Tipo de dado / Plugin	Origem
Taxonomia / GUS::Supported::Plugin::LoadTaxon	NCBI Taxonomy (Wheeler <i>et al.</i> , 2000)
Ontologia de sequências / GUS::Supported::Plugin::InsertSequenceOntologyOBO	Sequence Ontology (Karen Eilbeck <i>et al.</i> , 2005)
Ontologia de Genes / ApiCommonData::Load::Plugin::InsertGOEvidenceCodesFromObo ApiCommonData::Load::Plugin::InsertGOTermsFromObo	Gene Ontology (Harris <i>et al.</i> , 2004)
Base de dados de nomenclatura de enzimas / GUS::Community::Plugin::LoadEnzymeDatabase	Expasy (Bairoch, 2000)

O SchistoDB provê acesso à versão 4.0 da montagem do genoma de *S. mansoni*, cepa Puerto Rico (El-Sayed *et al.*, 2004; Haas *et al.*, 2007), obtida do WTSI – *Wellcome Trust Sanger Institute* e ao genoma mitocondrial, cepa NMRI (Le *et al.*, 2000), obtida do NCBI – *National Center for Biotechnology Information*. A atual versão do SchistoDB contém oligonucleotídeos (Verjovski-Almeida *et al.*, 2007) utilizados no microarranjo de 44 mil

oligos da Agilent e todas as *ESTs* de *S. mansoni* disponíveis no NCBI - DBEST mapeados no genoma (Tabela 2).

Tabela 2: Bases de dados de sequências e anotações de *S. mansoni*

Tipo de dado / Plugin	Quantidade	Origem
Genoma completo / GUS::Supported::Plugin::InsertSequenceFeatures	19.022 scaffolds	WTSI (El-Sayed <i>et al.</i> , 2004; Haas <i>et al.</i> , 2007)
Genoma mitocondrial / GUS::Supported::Plugin::InsertSequenceFeatures	1 scaffold	NCBI (Le <i>et al.</i> , 2000)
ESTs (<i>Expressed Sequence Tags</i>) / ApiCommonData::Load::Plugin::InsertDbESTFiles	205892 ESTs	NCBI DBEST
Microarranjo da Agilent / GUS::Supported::Plugin::LoadFastaSequences	41.967 oligos	(Verjovski-Almeida <i>et al.</i> , 2007)

Nas Tabelas 1 e 2, foram listados não somente os dados que foram carregados, mas também o nome do *plugin* utilizado. A execução desses *plugins* ocorreu de forma sistematizada através da utilização de um arquivo XML (Material Suplementar online. Disponível em: <<http://bioinfo.cpqrr.fiocruz.br/content/2009/06/10/material-suplementar-tese-de-adhemar-zerlotini-neto>>. Acessado em: 10 de junho de 2009) que define, dentre outras diretivas, a localização do arquivo de dados, o *plugin*, sua localização e parâmetros. A Figura 4 apresenta uma seção do arquivo XML, onde podem ser visualizadas todas as diretivas para carga de dados. O arquivo XML é também utilizado para exibição dos dados carregados na página *Data Sources* do SchistoDB (<http://SchistoDB.net>).


```

<resource resource="InterproscanData" version="14.1"
  plugin="ApiCommonData::Load::Plugin::InsertInterproDomainDbs"
  displayName="InterPro Database"
  publicUrl="http://www.ebi.ac.uk/interpro"
  category="Proteome"
  url="ftp://ftp.ebi.ac.uk/pub/databases/interpro/iprscan/DATA/"
  dbCommit="@dbcommit@">
  <wgetArgs>--tries=5 --no-parent --no-directories --no-host-directories --cut-dirs=5 --accept=iprscan_DATA_14.1.tar.gz</wgetArgs>
  <unpack>tar xzf @downloadDir@/InterproscanData/iprscan_DATA_14.1.tar.gz</unpack>
  <unpack>rm -f @downloadDir@/InterproscanData/iprscan_DATA_14.1.tar.gz</unpack>
  <unpack>cp /scratch2/adhemar/schistodb20/tmpFiles/insertInterpro-config.xml @downloadDir@/InterproscanData/iprscan/data/</unpack>
  <unpack>fixSuperfamily @downloadDir@/InterproscanData/iprscan/data/</unpack>
  <pluginArgs> --inPath @downloadDir@/InterproscanData/iprscan/data/ </pluginArgs>
  <description>
    <![CDATA[
      InterPro, from the <a href="http://www.ebi.ac.uk/">European
      Bioinformatics Institute</a>, is a curated database of
      protein families, domains and functional sites in
      which identifiable features found in known proteins
      can be applied to unknown protein sequences.
      <p>InterPro is formed by collating protein domains
      from a set of approximately a dozen public protein
      domain databases into functionally equivalent groups.
      The groups are given InterPro names and possibly
      assigned one or more Gene Ontology term.
    ]]>
  </description>
  <publication pmid="15608177"/>
</resource>

```

Figura 4: Seção do arquivo XML de carga dos dados

Essa figura mostra a entrada para carregamento do banco de dados InterProScan. Em cada entrada desse arquivo XML são especificados o nome (*resource*), a categoria (*category*) e a localização do arquivo de dados (*url*), os parâmetros para o comando de download desse arquivo (*wgetArgs*), o *plugin* para carga do arquivo e sua localização (*plugin*), os parâmetros do *plugin* (*wgetArgs*), comandos para eventual pós processamento do arquivo de dados (*unpack*), o identificador da publicação do arquivo de dados no PubMed (*pmid*) e a descrição do arquivo de dados (*description*).

4.1.4 – Análises computacionais dos dados

O SchistoDB contém diversas análises computacionais dos dados de *S. mansoni*, com o intuito de mais detalhadamente caracterizar as sequências e confirmar as anotações dos genes preditos. Bases de dados adicionais foram importadas para possibilitar algumas das análises (Tabela 3).

Tabela 3: Bases de dados utilizados para análises computacionais

Tipo de dado / Plugin	Origem
Domínios de proteínas / <code>ApiCommonData::Load::Plugin::InsertInterproDomainDbs</code>	InterPro (Mulder & Rolf Apweiler, 2008)
Proteínas não-redundantes / <code>ApiCommonData::Load::Plugin::LoadNothing</code>	NCBI (Benson <i>et al.</i> , 2000)
Drogas / <code>SchistoDBData::Load::Plugin::LoadKeggDrugDatabase</code>	Kegg Drug (Kanehisa <i>et al.</i> , 2008)
Genes Ortólogos / <code>ApiCommonData::Load::Plugin::InsertOrthologousGroups</code> <code>ApiCommonData::Load::Plugin::InsertPhylogeneticProfile</code>	OrthoMCL (Chen <i>et al.</i> , 2006)
Estruturas de proteínas / <code>GUS::Supported::Plugin::LoadFastaSequences</code>	RCSB PDB (Kouranov <i>et al.</i> , 2006)

Foram efetuadas análises de similaridade utilizando o software BLAST (Altschul *et al.*, 1997) de sequências do genoma de *S. mansoni* com a base de dados de proteínas não redundantes do NCBI (NR); do proteoma predito com a base de dados (NR) e com a base de dados de estruturas de proteínas (RCSB PDB). Adicionalmente, foi realizada análise para identificação de janelas abertas de leitura (ORFs) no genoma e calculado o conteúdo ACGT.

Utilizando o proteoma predito, foram localizados os domínios e famílias de proteínas, presença do peptídeo sinal, os domínios transmembrana, os termos da ontologia de genes

(GO) e também calculado o peso molecular e o ponto isoelétrico de cada uma das proteínas. Após a predição das vias metabólicas que será detalhada na sessão 4.2, foram atribuídos *EC numbers* (*Enzyme Comission Number*) às proteínas e essa informação foi posteriormente carregada no SchistoDB.

As ESTs foram clusterizadas utilizando o software cap3 (Huang et al., 1999) para minimizar a redundância desse conjunto de dados.

Para execução e carga sistematizada de todas as análises acima descritas foram criados vários programas PERL que são executados a partir de um arquivo principal, chamado schistoAnalysis (Material Suplementar online. Disponível em: <<http://bioinfo.cpqrr.fiocruz.br/content/2009/06/10/material-suplementar-tese-de-adhemar-zerlotini-neto>>. Acessado em: 10 de junho de 2009). A Figura 5 apresenta uma seção deste arquivo que mostra os comandos para análise de proteínas e carregamento dos resultados do *software* InterProScan que foi utilizado para predição de famílias, domínios e termos do *Gene Ontology*. Para cada linha de comando executada são gerados dois arquivos de log que contêm a saída padrão e a saída de erro da execução do comando.

```

##### InterPro #####
&createIprscanDir($mgr, "SmansoniAnnotatedProteins.fsa");
&copyFilesToComputeCluster($mgr, "iprscan");
&startIprScanOnComputeCluster($mgr, "SmansoniAnnotatedProteins.fsa", "apidb", 1);
##### Load InterPro #####
$mgr->waitForCluster("Iprscan of SmansoniAnnotatedProteins", "waitSmansoniAnnotatedProteins-iprscan");
&copyFilesFromComputeCluster($mgr, "SmansoniAnnotatedProteins.fsa", "iprscan");
my $insertInterproConfig = "/scratch2/adhemar/schistodb20/pipeline/Downloads/InterproscanData/iprscan/data/insertInterpro-config.xml";
my $interpro_version = "14.1";
&loadIprscanResults($mgr, "SmansoniAnnotatedProteins.fsa", "INTERPRO", $interpro_version, $insertInterproConfig);

```

Figura 5: Seção do arquivo de execução e carregamento de análises
createIprscanDir: cria um diretório contendo a sequência de proteínas que será analisada
copyFilesToComputeCluster: copia o diretório criado para o *cluster* de computadores
startIprScanOnComputeCluster: inicia a análise do InterProScan no *cluster* de computadores
waitForCluster: finaliza o programa e pede para reiniciá-lo quando terminada a análise
copyFilesFromComputeCluster: copia o arquivo de resultados para a máquina local
loadIprscanResults: carrega o resultado no banco de dados

4.1.5 – Integração dos diferentes tipos de dados

A maior vantagem da utilização de bancos de dados relacionais é justamente a possibilidade de integrar diferentes tipos de dados. A integração de dados é o processo de se combinar dados provenientes de diferentes fontes e, dessa forma, oferecer ao usuário uma visão unificada do conjunto de dados. Na construção do SchistoDB foi utilizada uma abordagem de integração de dados conhecida como *local-as-view*. Nessa abordagem é necessário que o esquema tabelas e seus relacionamentos estejam especificados no banco de dados local para que os dados sejam extraídos do arquivo de dados original para as tabelas adequadas (Lenzerini, 2002).

O principal conjunto de dados armazenado é o genoma de *S. mansoni*, dessa forma, todos os outros conjuntos de dados foram relacionados/integrados ao genoma ou com os genes preditos. Os demais tipos de dados puderam ser integrados ao genoma por meio da similaridade entre as sequências ou por meio do código identificador do dado como chave primária do relacionamento.

A maioria dos relacionamentos foram efetuados com base no resultado da análise de similaridade das sequências. Utilizando-se o software BLAST (Altschul et al., 1997), foram integrados ao genoma os seguintes conjuntos de dados: proteínas não-redundantes (NCBI NR), estruturas de proteínas (RCSB PDB) e *44K Agilent Oligoarray*. Já as *ESTs* e os agrupamentos de *ESTs* foram alinhadas ao genoma utilizando o software BLAT (Kent, 2002). Ao navegar pelo banco de dados o usuário pode obter informações estatísticas sobre a similaridade (*score*) e a probabilidade do alinhamento ter ocorrido ao acaso (*evaluate*) de cada um dos alinhamentos. Dessa forma o usuário pode avaliar o nível de identidade entre as sequências para inferir se as moléculas são homólogas.

Os demais dados foram integrados através de códigos de identificação, o que garante total integridade relacional. Os genes preditos de *S. mansoni* foram caracterizados como pertencentes a grupos de genes ortólogos pelos desenvolvedores do OrthoMCL (Chen et al., 2006) e a análise foi importada para o SchistoDB, utilizando-se o identificador do gene como chave primária do relacionamento.

A base de dados de drogas Kegg Drug (Masoudi-Nejad et al., 2007) foi carregada e os alvos dessas drogas foram associados aos genes preditos de *S. mansoni* que o têm mesmo *EC number* do alvo original. Dessa forma foi possível identificar enzimas ortólogas de *S. mansoni* que são potenciais alvos de drogas descritos para outros organismos.

4.1.6 – Publicação dos dados

O GUS WDK (*GUS Web Development Kit*) foi utilizado na criação da interface para a extração e publicação do conteúdo do banco de dados. O GUS WDK foi desenvolvido com o objetivo de acelerar a produção de *websites* para mineração de dados genômicos através da criação de uma camada objeto-relacional que utiliza a tecnologia JSP.

A utilização de um *website* desenvolvido com esse WDK ocorre basicamente em três etapas: 1) o usuário executa uma questão fornecendo os devidos parâmetros (*Question page*); 2) uma tela com os relatórios dos resultados é apresentada (*Summary page*); 3) ao clicar em algum dos resultados é exibida uma página com todas as informações sobre o determinado tipo de dado (*Record page*).

A descrição de todo o conteúdo de cada uma das páginas se concentra em um arquivo no formato XML (Material Suplementar online. Disponível em: <<http://bioinfo.cpqrr.fiocruz.br/content/2009/06/10/material-suplementar-tese-de-adhemar-zerlotini-neto>>. Acessado em: 10 de junho de 2009). Esse arquivo é dividido em sessões que descrevem e controlam o funcionamento de cada uma das páginas. Para a *Question page* há uma sessão que contém todas as perguntas (*Questions*) e consultas (*Queries*) que serão disponibilizadas, bem como especificações dos campos de entrada e colunas que serão exibidas na página de resultados (*Summary page*) (Figura 6). As informações contidas tanto na *Summary page* quanto na *Record page* são obtidas em uma parte específica do arquivo XML (*Attributes*). Todo acesso ao banco de dados é feito através desse arquivo, porém as partes que compõem a *Record page* são descritas em um arquivo auxiliar para cada um dos tipos de dados.

O GUS WDK possibilita que todas as consultas (*queries*) efetuadas sejam armazenadas em tabelas cache de forma a diminuir a execução no banco de dados e consequentemente o tempo de resposta para o usuário. Quanto mais o banco é utilizado, mais rápido será o seu desempenho já que o resultado das consultas efetuadas estará armazenado em tabelas cache.

```

<!-- GeneQuestions.GenesBySingleLocusTag -->
<question name="GeneBySingleLocusTag"
  category="Other Attributes"
  displayName="ID"
  queryRef="GeneFeatureIds.GeneBySingleLocusTag"
  summaryAttributesList="location_text,product,gene_type,contig"
  recordClassRef="GeneRecordClasses.GeneRecordClass">
  <summary>
    Find a gene by its ID.
  </summary>
  <description>
    <![CDATA[
      Find a gene by its EMBL locus tag ID.
      The search is case insensitive.
      Wild card searches, e.g. 'Smp_0001*', are permitted.
    ]]>
  </description>
</question>

<!-- GeneFeatureIds.GeneBySingleLocusTag -->
<sqlQuery name="GeneBySingleLocusTag" displayName="By Gene ID" isCacheable='true'>
  <paramRef ref="params.locus_tag"/>
  <column name="source_id" />
  <sql>
    <![CDATA[
      SELECT g.source_id
      FROM dots.genefeature g,
            dots.transcript t
      WHERE g.na_feature_id = t.parent_id (+)
      AND (
        LOWER(g.source_id) LIKE TRIM(LOWER(REPLACE('$$locus_tag$$', '*', '%')))
        OR LOWER(t.protein_id) LIKE TRIM(LOWER(REPLACE('$$locus_tag$$', '*', '%'))) || '%'
      ) and g.source_id is not null
    ]]>
  </sql>
</sqlQuery>

```

Figura 6: Seção do arquivo XML de configuração do WDK

Essa seção apresenta a entrada para a busca 'Find genes by Gene ID' e sua respectiva *query* SQL. A entrada que define os atributos da pergunta (*Question*) contém os parâmetros para sua identificação e caracterização (name, category, displayName, queryRef, recordClassRef, summary e description), bem como a lista de colunas que serão exibidas na página de resultados (summaryAttributeList). A entrada que define os atributos da *sqlQuery* que será executada no banco de dados também contém parâmetros para sua identificação e caracterização, além da *query* SQL (name, displayName, isCacheable, paramRef, column e sql).

Outra funcionalidade contida no GUS WDK é o histórico de consultas efetuadas pelos usuários. Essas consultas podem ser visualizadas na página de resultados ou baixadas em formato 'texto separado por vírgula' enquanto a sessão estiver ativa. A página de histórico é uma das ferramentas mais importantes do banco de dados. Através dela torna-se possível combinar diferentes consultas para obtenção de resultados mais refinados.

4.2 – Predição das vias metabólicas com base na anotação dos genes

Utilizando a anotação dos genes foi possível efetuar a predição computacional das vias metabólicas de *S. mansoni*. Para tal, foi utilizado o software PathwayTools desenvolvido pelo SRI International (Karp et al., 2002).

O PathwayTools utiliza como referência para suas predições uma base de dados de vias metabólicas e enzimas, conhecida por MetaCyc (Caspi et al., 2008). As vias metabólicas contidas no MetaCyc são extraídas da literatura e elucidadas experimentalmente.

Para predição das vias metabólicas, o PathwayTools oferece um componente conhecido como Pathologic. Para utilizá-lo, foi necessário desenvolver um *script* PERL que convertesse o arquivo de anotação obtido no formato EMBL para o formato Genbank, de forma que fossem atendidas as especificações do Pathologic. Os campos utilizados na predição são: CDS (genes que codificam uma proteína), tRNA (genes que codificam um RNA transportador), rRNA (genes que codificam um RNA ribossomal) e misc_RNA (genes que codificam outros RNAs), sendo que os atributos para esses campos são: gene (nome do gene), EC_number (identificador da reação), product_comment (comentários sobre o produto do gene), locus_tag (identificador da localização genômica), gene_comment (comentários sobre o gene), note (notas adicionais), alt_name (nomes alternativos do gene), product (nome da proteína que o gene codifica), pseudo (informação se o gene é um pseudogene) e db_xref (identificador para um banco de dados externo).

Na primeira etapa, conhecida como *trial parse*, o arquivo foi preliminarmente analisado no intuito de se validar o formato. Nessa fase, além de serem exibidas mensagens para cada campo que foi localizado durante o processo, é exibido um pequeno relatório que nos indica o número de genes encontrados, de *EC numbers* e de produtos.

Na segunda etapa, conhecida como *automated build*, o banco de dados de vias metabólicas foi efetivamente construído utilizando-se basicamente os atributos *EC number* e produto. O processo de predição das vias metabólicas ocorre através da inferência da presença das reações catalizadas com base no produto dos genes anotados. Se um *EC number* foi atribuído ao produto do gene, então o software infere a presença da reação com base neste identificador. Caso contrário, o software compara o produto do gene com um extenso dicionário de nomes de enzimas para, então, inferir a presença da reação.

Após esse processo, foram submetidas ao Hole Filler (Green et al., 2004) todas vias metabólicas que continham enzimas ausentes (*holes*), ou seja, enzimas que não foram localizadas através de seus atributos. O Hole Filler é um módulo do Pathologic utilizado para localizar enzimas ausentes nas vias metabólicas através de análise similaridade de sequência. Para inferir a presença de tais enzimas, esse módulo gera um arquivo contendo todas as sequências de proteínas obtidas do UNIPROT (UniProt Consortium, 2008) que se referem ao *EC number* das enzimas ausentes, independentemente do organismo a que elas pertencem. Utilizando o software BLAST (Altschul et al., 1997), as sequências de enzimas do UNIPROT são comparadas com todas as enzimas contidas no arquivo de anotação. Ao final desse processo, é exibida uma janela contendo todas as inferências feitas pelo Hole Filler para que sejam manualmente confirmadas.

Relatórios da inferência da presença das enzimas por *EC number*, nome ou similaridade de sequência são automaticamente gerados. Contudo essa informação pode ser obtida individualmente para cada uma das enzimas durante a navegação pelo banco de dados.

4.3 – Seleção dos candidatos a alvos terapêuticos com base na anotação dos genes

A construção da base de dados de vias metabólicas nos permitiu efetuar, entre outras análises, a busca por enzimas que catalizam reações *chokepoints*. O termo *chokepoint* foi cunhado por Yeh e colaboradores (Yeh et al., 2004) ao se referir a reações que unicamente consomem ou produzem um específico substrato ou produto. Espera-se que a inibição de uma enzima que consome um substrato único resulte na acumulação desse substrato, potencialmente tóxico para a célula e que a inibição de uma enzima que produz um composto único resulte em inanição, o que potencialmente desequilibraria funções celulares. Essa análise pôde ser efetuada diretamente no software PathwayTools utilizando-se um *plugin* fornecido pelo próprio SRI International (Chokepoint Analysis Tool. Disponível em: <<http://bioinformatics.ai.sri.com/ptools/chokepoint/chokepoint-analysis.html>>. Acessado em: 4 de junho de 2009).

Inicialmente, o *plugin* agrupa todos os metabólitos a serem examinados, coletando-os a partir de das reações presentes em vias metabólicas e reações isoladas que utilizam apenas pequenas moléculas como metabólitos. São também consideradas reações presentes em vias que utilizam substratos macromoleculares, como proteínas que estão sendo modificadas.

O algoritmo busca localizar dois tipos de reações *chokepoint*. (a) Reações R1 em que apenas uma reação produz determinado metabólito M e ao menos uma reação consome M. (b) Reações R2 em que apenas uma reação consome determinado metabólito M e ao menos uma reação produz M. Dado que uma reação *chokepoint* é definida como sendo a única reação que consome ou produz um determinado metabólito, isso implica que para encontrar uma reação *chokepoint*, todas as reações que envolvem M devem ser unidirecionais. As reações que atendem aos critérios listados são reunidas em um relatório e disponibilizadas para o usuário.

4.4 – Busca na literatura sobre informações dos genes selecionados

Foi realizada uma extensa busca na literatura utilizando-se a base de dados do PubMed (PubMed Overview. Disponível em: <http://www.ncbi.nlm.nih.gov/corehtml/query/static/overview.html>). Acessado em: 6 de maio de 2009) no intuito de encontrar artigos científicos referentes à caracterização e testes de drogas das enzimas que catalizam reações consideradas *chokepoint*. Tal pesquisa foi efetuada utilizando-se o software PubMatrix (Becker et al., 2003) para recuperação dos artigos relacionados com os termos pesquisados. Essa é uma ferramenta de mineração de texto que disponibiliza uma tabela de frequência de coocorrência dos nomes das enzimas versus os termos modificadores: *drugs*, *parasites* e *Schistosoma mansoni*. O resultado da busca é apresentado no formato HTML, onde as frequências são *links* para a lista de artigos.

A partir da tabela de resultados obtida utilizando-se o PubMatrix, foram analisados todos os artigos científicos nos quais foi constatada a co-ocorrência do nome da enzima e os termos *drugs* e *Schistosoma mansoni*, com o objetivo de se identificar trabalhos que relatam atividade esquistossomicida de drogas que atuam nas enzimas selecionadas. Foram ainda analisados todos os artigos nos quais foi constatada a co-ocorrência do nome da enzima e o termo *Schistosoma mansoni*, com o objetivo de se identificar trabalhos que relatam a caracterização dessas enzimas em *S. mansoni*.

4.5 – Validação experimental dos melhores candidatos

Com a finalidade de confirmar os resultados da análise computacional foi efetuado um teste de eficácia de 4 das drogas que atuam em enzimas que catalizam reações *chokepoint* não testadas anteriormente contra *S. mansoni*. As drogas Alopurinol (A8003-5G), Metotrexato (M9929-25MG), Pravastatina Sódica (P4498-25MG) e Trimetoprima (T7883-5G) foram adquiridas da empresa Sigma-Aldrich Brasil Ltda. Através de revisão da literatura sobre experimentos que utilizaram essas drogas em camundongos, foi estabelecida a máxima dosagem. Tal busca foi efetuada utilizando-se o PubMed, onde buscou-se localizar artigos que apresentaram a co-ocorrência dos termos *mouse* e o nome da droga.

Para o teste da eficácia, 60 camundongos suíços albinos fêmeas (~25g) foram infectados com 80 ± 10 cercárias de *S. mansoni* cepa LE, por via subcutânea seguindo as orientações do Comitê de Ética para uso experimental de animais. Após 40 dias da data de infecção, grupos de aproximadamente 10 animais foram tratados com uma dose única via oral de cada uma das drogas e então sacrificados e perfundidos 10 dias após o tratamento. Os camundongos foram sacrificados por fratura cervical. Foi então aberta a cavidade peritoneal e feita a ligadura da veia renal e a secção da veia porta, para permitir a extração dos vermes. Com o auxílio de uma agulha acoplada a um pipetador automático foi injetada solução salina a 0,85% na aorta descendente procedendo-se à perfusão das veias mesentéricas e coleta dos vermes (Pellegrino et al., 1956). O mesmo procedimento foi efetuado com o grupo controle, que é compreendido de camundongos infectados e não tratados, 40 dias após a data de infecção. A comparação do número de vermes entre os grupos tratados e o grupo controle nos permite avaliar a atividade terapêutica. Foi efetuado o teste T-Student para avaliar se a diminuição do número de vermes foi estatisticamente significativa.

V – RESULTADOS & DISCUSSÃO

5.1 – Banco de dados genômicos de *S. mansoni*

Existem diversos repositórios de dados genômicos na Internet. Porém, cada um desses repositórios utiliza diferentes formatos e normalmente não há integração direta com os demais. No presente estudo, o primeiro passo para a investigação em escala genômica do parasito foi criar uma base de dados relacional que proporcionasse integração dos diversos tipos de dados e análises.

O banco de dados criado no decorrer desse projeto, SchistoDB (<http://SchistoDB.net>)

Schistosoma mansoni Database Release 2.0 July 2nd, 2008

Home Query History Sequence Retrieval Genome Browser SchistoCyc

Publications relying on SchistoDB should please acknowledge the database developers and the scientists who have made their data available. SchistoDB is funded by NIH-Fogarty Grant 5D43TW007012-03 Please explore the site and provide feedback and suggestions to schistodb@cpqrr.fiocruz.br

Queries

Find Genes By

Gene ID [Go!](#) [Help?](#)

Keyword [Go!](#) [Help?](#)

Gene Type [Go!](#) [Help?](#)

Genomic Location [Go!](#) [Help?](#)

Exon/Intron Structure [Go!](#) [Help?](#)

EST Evidence by Sex [Go!](#) [Help?](#)

EST Evidence by Stage [Go!](#) [Help?](#)

Orthology Profile [Go!](#) [Help?](#)

Interpro/Pfam Domain [Go!](#) [Help?](#)

Transmembrane Domain [Go!](#) [Help?](#)

Signal Peptide [Go!](#) [Help?](#)

Oligo ID [Go!](#) [Help?](#)

GO Term [Go!](#) [Help?](#)

EC Number [Go!](#) [Help?](#)

Drug [Go!](#) [Help?](#)

PDB Similarity [Go!](#) [Help?](#)

Protein Motif [Go!](#) [Help?](#)

Molecular Weight [Go!](#) [Help?](#)

IsoelectricPoint [Go!](#) [Help?](#)

Find Contigs By

Contig ID [Go!](#) [Help?](#)

Find ORFs By

ORF ID [Go!](#) [Help?](#)

Genomic Location [Go!](#) [Help?](#)

ORF Motif [Go!](#) [Help?](#)

Find ESTs By

EST ID [Go!](#) [Help?](#)

Genomic Location [Go!](#) [Help?](#)

Gene Overlap [Go!](#) [Help?](#)

Library ID [Go!](#) [Help?](#)

Developmental Stage [Go!](#) [Help?](#)

Sex [Go!](#) [Help?](#)

NEWS

SchistoDB 2.0 is released

ACKNOWLEDGMENTS

We would like to acknowledge the genome sequencing consortium, TIGR, and WTSI for the availability of the genome assembly of *S. mansoni*. Preliminary genome sequence was obtained from GeneDB website. For further information go to Acknowledgements.

SchistoDB Date: 07/2008 Schistosoma mansoni genome resource For help, contact SchistoDB Administrator

Figura 7: Página inicial do banco de dados SchistoDB

Na página inicial do SchistoDB, aproximadamente 30 diferentes buscas estão disponíveis para possibilitar buscas nos dados de *S. mansoni*, além de acesso à ferramentas como BLAST, Genome Browser, Sequence Retrieval Tool. As buscas podem ser combinadas através do 'Query History' ampliando consideravelmente a capacidade de filtragem dos dados.

(Zerlotini et al., 2009) (Figura 7), foi publicado no periódico *Nucleic Acids Research* e oferece várias formas de busca e acesso às informações contidas na base, além de um relatório para cada tipo de dado contendo detalhes das anotações e das análises efetuadas.

Estão disponíveis aproximadamente 30 campos para filtragem dos dados, além de ferramentas para busca por similaridade nas sequências depositadas - BLAST, para *download* em larga escala de diferentes conjuntos de dados - *Sequence Retrieval Tool* e para navegação através do genoma de forma gráfica – Genome Browser (Stein et al., 2002).

Para cada um dos tipos de dados cadastrados, uma página contendo detalhes das anotações e análises efetuadas está disponível. Por exemplo, ao acessar a página de um determinado gene, podemos encontrar o nome e o identificador do mesmo, sua posição genômica, a descrição da proteína codificada e uma representação gráfica da sua localização genômica gerada pelo Gbrowse, o qual possibilita ampla navegação na região onde o gene se encontra. Além disso, é possível visualizar outros tipos de dados que se alinham à essa região como ESTs, agrupamentos de ESTs, resultados de BLAST, oligonucleotídeos e ORFs. Também estão disponíveis tabelas contendo resultados de atribuições computacionais de *EC numbers*, termos do *Gene Ontology* e drogas conhecidas por atuar na reação especificada pelo *EC number* em outros organismos. Na página em questão, estão ainda disponíveis análises computacionais, por meio de representação gráfica gerada pelo Gbrowse, efetuadas com o uso da proteína traduzida, como resultados de BLAST contra o banco de dados de proteínas não-redundantes do NCBI e de InterProScan, o qual identifica famílias e domínios proteicos, regiões transmembrana e peptídeos sinais. Ao final da página, *links* para estruturas do PDB e as sequências do gene, mRNA e da proteína estão disponibilizados (Figura 8).

Todas as consultas efetuadas podem ser acessadas por meio de uma página de histórico ou obtidos em formato “texto separado por vírgula” enquanto a sessão estiver ativa.

Home Query History Sequence Retrieval Genome Browser SchistoCyc

1 Overview
 This *Schistosoma mansoni* gene spans positions **261868 - 266439** of contig **Smp_scaff000112**.
 Approximate protein mol. wt. (Daltons): **39147** [computed from raw translation]

Annotation Notes
 predicted by EVM2

Product
 ribokinase, putative

2 Genomic Context

3 EC Numbers

Accession	Description	Source
2.7.1.15	Ribokinase	SchistoCyc
2.7.11.24	Mitogen-activated protein kinase	SchistoCyc

GO Terms

Ontology	GO ID	GO Term Name	Source	Evidence Code
F	GO:0004747	ribokinase activity	Interpro	IEA
P	GO:0006014	D-ribose metabolic process	Interpro	IEA

Kegg Drugs

Entry	Name	Synonyms	Activity	target	CAS	PubChem
D03736	Doramapimod (USAN)		Treatment of rheumatoid arthritis, Crohn's disease and psoriasis	mitogen-activated protein (MAP) kinase inhibitor [KO:K04441] [E.C.2.7.11.24]	285983-48-4	17397825

4 Predicted Protein Features

5 Similarities to Protein Data Bank (PDB) Chains

PDB Structure	PDB Description	Taxon	% Coverage	% Identity	P-value
2lv7_A	Ribokinase	Homo sapiens	92	42	1.3 x 10 ⁻⁵³
2lv7_B	Ribokinase	Homo sapiens	92	42	1.3 x 10 ⁻⁵³
1gqt_C	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻²⁶
1gqt_D	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻²⁶

Predicted Protein Sequence

```

MKREHVVVSGSLNVDLEVFTHIIPVVGIVTISGNFLLYGGKSNQCVASVVLGNTALV
GRVGGVYFGEHIGRLQGLQVYDGLIENQDNTGVAIIIVETSTGGNIIIVPGRANLV
SEKIIFDAKLLIDTRVIVCGFFENPAGLYELNLSNAGKATILNAPAPVAGSSEPI
LGNELMEDILPNCDFVCPNSEECSITSDGSLNLSKNEIGSLNIDAFIPGLAYLLEK
IKCFIVTLGKGVIAIIEGDMANIYKDAISEVARIITFENGEKLVVHISAPNEFDVDTT
GASDCTVGLIATVACREDITLAEQINRNVVAVSISINNKQTGSYLRKGFELDTLFASE
TFQSP
  
```

Genome Sequencing and Annotation by:
 Preliminary sequence data for *S. mansoni* genome was obtained from GeneDB website.

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 Nov; 117(3):225-8.

El-Sayed NM, Bartholomeu D, Ivens A, Johnston DA, LoVerde PT. Advances in schistosome genomics. Trends Parasitol. 2004 Apr; 20(4):154-7.

SchistoDB Schistosoma mansoni genome resource For help, contact SchistoDB Administrator
 Date: 07/2008

Figura 8: Exemplo de relatório dos genes

Cada gene possui uma página que contém 1) nome e identificador; posição genômica; produto; 2) representação gráfica da localização genômica e outros tipos de dados associados como ESTs, resultados de BLAST, oligonucleotídios e ORFs; 3) EC numbers; termos do Gene Ontology; drogas; 4) representação gráfica da proteína traduzida e resultados de BLASTP e InterProScan; 5) estruturas do PDB; e sequências.

Utilizando-se a página de histórico torna-se possível combinar diferentes consultas para obtenção de resultados mais específicos. Por exemplo, após executar três consultas na página inicial, tais como genes que possuem homólogos no banco de dados de estruturas tridimensionais, genes que são alvos de drogas em outros organismos e genes que tem ESTs alinhadas, podemos acessar a página de histórico e, então, combiná-las para que sejam exibidos apenas genes com as três características consultadas (Figura 9).

Para a construção do SchistoDB, vários esquemas de bancos de dados genômicos poderiam ser utilizados como o Chado (Mungall et al., 2007), o Atlas (Shah et al., 2005) ou o BioWareHouse (Lee et al., 2006). O esquema GUS foi escolhido por sua ampla estrutura de armazenamento, normalização, análise e publicação de dados. Foi também considerado o fato de que vários bancos de dados genômicos de organismos parasitas utilizam o GUS, como o PlasmoDB (Aurrecochea et al., 2009), o CryptoDB (Heiges et al., 2006), o ToxoDB (Gajria et al., 2008), o ProtozoaDB (Dávila et al., 2008) e o TcruziDB (Agüero et al., 2006), o que pode proporcionar futura integração por meio de *web services* com o objetivo de efetuar análises comparativas.

O GUS não apresenta documentação suficiente para uma fácil implementação de um novo banco de dados. Toda a estrutura e componentes estão relacionados na documentação *online* (GUS: The Genomics Unified Schema. Disponível em: <<http://gusdb.org>>. Acessado em: 7 de maio de 2009), porém os detalhes de utilização e execução de rotinas somente são descritos utilizando-se os comandos de “ajuda” dos *plugins* durante a fase de implementação. Para se obter conhecimento avançado, faz-se necessário o contato com os idealizadores do esquema e utilização de fóruns *online*. Vale destacar que o conteúdo presente na sessão Materiais e Métodos (4.1) referente à estrutura da base de dados, carga, análise e publicação dos dados não pretende substituir a documentação oficial, mas apenas oferecer detalhes sobre

Schistosoma mansoni Database Release 2.0 July 24, 2008

Home Query History Sequence Retrieval Genome Browser SchistoCyc

Search query history: Please explore the Smp IDs, protein functions and suggestions

Query History

ID	Query	Site
1	ribosom	229
2	kinase	205
3	EST	204

Gene query history

Gene Results

Gene	Gene ID	Location	Protein	Gene type	Contig ID
Smp_00001	Smp_scaffold012_12541-12541	Trypanothione synthase (trypanothione 5-hydroxymethyl-PTC) kinase	coding_gene	Smp_scaffold012	
Smp_00002	Smp_scaffold012_12542-12542	cytochrome P450 reductase	coding_gene	Smp_scaffold012	
Smp_00003	Smp_scaffold012_12543-12543	cytochrome P450 reductase	coding_gene	Smp_scaffold012	

Gene Page

Smp_044850.2 Record

Overview: The *Schistosoma mansoni* gene spans positions 261868 - 266439 of contig Smp_scaff000112.

Approximate protein mol. wt. (Daltons): 26147 (predicted) (see also UniProt)

Annotation Notes

Annotation Notes predicted by EVAM2

Product

ribokinase, putative

Links to Other Web Pages

Schistosoma Pathway Database
Sanger GeneDB

Genomic Context

EC Numbers

Accession	Description	Source
2.7.1.15	ribokinase	SchistoCyc
2.7.11.24	Mitogen-activated protein kinase	SchistoCyc

GO Terms

Ontology	GO ID	GO Term Name	Source	Evidence Code
F	GO:0004747	ribokinase activity	Inexpo	IEA
P	GO:0006014	D-ribose metabolic process	Inexpo	IEA

Kegg Drugs

Entry Name	Synonyms	Activity	target	CAS	PubChem
D00076	Drug treatment of (USAM)	Treatment of rheumatoid arthritis, Crohn's disease and psoriasis	mitogen-activated protein (MAP) kinase inhibitors [EC:00444]	285683-48-4	1759285

Schistosoma mansoni Paralogs (Orthologs)

none

Predicted Protein Features

Kyle-Donnell Hydrophathy Plot

Interpro Domains

Signal Peptide

Transmembrane Domains

Protein Hits

Similarities to Protein Data Bank (PDB) Chains

PDB Structure	PDB Description	Taxon	% Coverage	% Identity	P-value
1R77_A	RIBOKINASE	Homo sapiens	92	42	1.3 x 10 ⁻⁶³
1R77_B	RIBOKINASE	Homo sapiens	92	42	1.3 x 10 ⁻⁶³
1R77_C	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_D	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_E	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_F	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_G	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_H	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_I	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_J	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_K	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_L	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_M	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_N	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_O	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_P	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_Q	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_R	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_S	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_T	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_U	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_V	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_W	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_X	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_Y	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶
1R77_Z	RIBOKINASE	Escherichia coli	71	34	5.2 x 10 ⁻⁷⁶

Predicted Protein Sequence

Sequence Length: 305 aa

Gene Sequencing and Annotation by:
Preliminary sequence data for *S. mansoni* genome was obtained from GenBank website.

Hass B, Brenner M, Hrad H, Cerqueira GS, Laverde P, El-Sayed NA. *Schistosoma mansoni* genome: closing in on a final gene set. *Exp Parasitol* 2007; Nov; 116:225-8.

El-Sayed NA, Barnabani D, Hrad H, Cerqueira GS, Laverde P, El-Sayed NA. *Schistosoma mansoni* genome: closing in on a final gene set. *Exp Parasitol* 2007; Nov; 116:225-8.

El-Sayed NA, Barnabani D, Hrad H, Cerqueira GS, Laverde P, El-Sayed NA. *Schistosoma mansoni* genome: closing in on a final gene set. *Exp Parasitol* 2007; Nov; 116:225-8.

Figura 9: Capturas de tela do SchistoDB mostrando consultas combinadas. A partir da tela inicial, usuários podem executar diferentes consultas e um relatório será exibido para cada uma delas (*Summary Page*). Pode-se fazer download dos resultados ou combiná-los através da tela de histórico (*Query History*).

o processo de implementação e vantagens de utilização dessa estrutura.

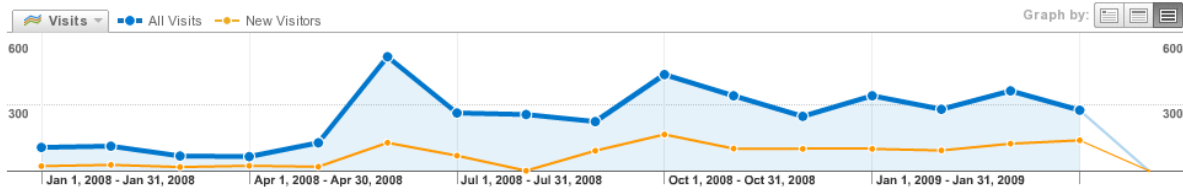
No intuito de possibilitar a reprodutibilidade da base de dados durante este projeto, disponibilizamos os principais arquivos de configuração que especificam os dados carregados, as análises efetuadas e as consultas disponíveis (Material Suplementar online. Disponível em: <<http://bioinfo.cpqrr.fiocruz.br/content/2009/06/10/material-suplementar-tese-de-adhemar-zerlotini-neto>>. Acessado em: 10 de junho de 2009). Após a instalação e configuração do GUS, o usuário terá acesso a um banco de dados de exemplo, conhecido por *wdkToySite*. Pode-se, então, utilizar os arquivos de configuração citados para efetuar o carregamento e análise dos dados, com os mesmo parâmetros definidos neste projeto, bem como disponibilizar as mesmas consultas. É importante destacar que em adição aos pacotes fornecidos pela distribuição padrão do GUS, foram incluídos diversos *plugins* que podem ser encontrados nos repositórios *Community* e *ApiDB*.

A construção da atual versão SchistoDB consumiu aproximadamente 3 meses de processamento computacional utilizando-se um *cluster* de 100 computadores apenas para a análise das sequências. Dentre as análises efetuadas, as que têm maior custo computacional são a análise de similaridade da sequencia genômica contra as proteínas não-redundantes do NCBI, a busca por famílias/domínios de proteínas e a clusterização das ESTs.

O SchistoDB, além de ser utilizado para a busca de genes candidatos a alvos terapêuticos nesse projeto, também tem sido utilizado para auxiliar pesquisas sobre esquistossomose realizadas pelo nosso grupo. Este banco de dados está disponível na Internet para toda a comunidade científica (Figura 10) e tem sido utilizado e referenciado por pesquisadores de diferentes grupos (Bergquist, 2008; DeMarco et al., 2009; Schramm et al., 2009; Bayne, 2009).

Visitors Overview

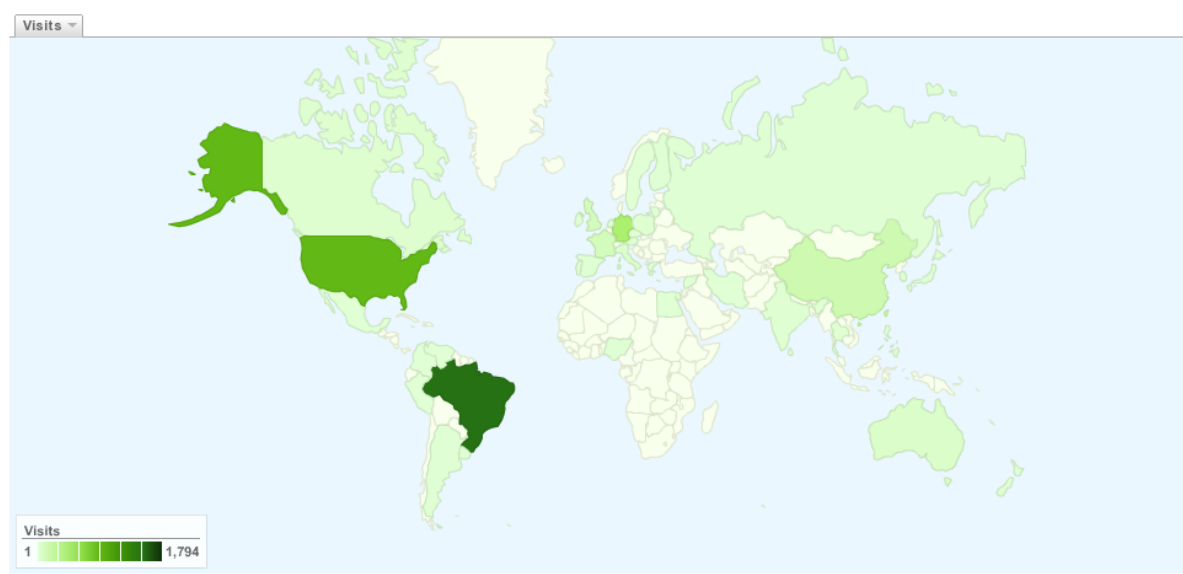
Jan 1, 2008 - May 1, 2009



This site received 4,033 visits in the "All Visits" segment

Map Overlay

Jan 1, 2008 - May 1, 2009



4,033 visits came from 39 countries/territories

Detail Level: City | Country/Territory | Sub Continent Region | Continent Dimension: None

Site Usage Goal Conversion Views: [Grid] [Clock] [List] [Filter]

Visits 4,033 % of Site Total: 100.00%	Pages/Visit 9.04 Site Avg: 9.04 (0.00%)	Avg. Time on Site 00:11:40 Site Avg: 00:11:40 (0.00%)	% New Visits 32.06% Site Avg: 31.99% (0.23%)	Bounce Rate 33.23% Site Avg: 33.23% (0.00%)
--	--	--	---	--

Rank	Country/Territory	Visits	Percentage
1.	Brazil	1,794	44.48%
2.	United States	1,067	26.46%
3.	Germany	525	13.02%
4.	China	187	4.64%
5.	France	128	3.17%
6.	United Kingdom	122	3.03%
7.	Australia	58	1.44%
8.	Italy	34	0.84%
9.	Canada	17	0.42%
10.	Portugal	13	0.32%

Figura 10: Estatísticas de acesso ao SchistoDB a partir de 2008
A figura apresenta um gráfico de número de acessos por mês e por localização dos usuários ("Google Analytics," [Página de Internet]).

5.2 – Predição das vias metabólicas de *S. mansoni*

A anotação dos genes nos permitiu efetuar a predição computacional das vias metabólicas de *S. mansoni*. Foi utilizado o software PathwayTools desenvolvido pelo *SRI International* (Karp et al., 2002), que possibilitou a criação de uma base de dados integrada ao SchistoDB por convenção nomeada SchistoCyc (<http://SchistoCyc.SchistoDB.net>).

Foram identificadas 112 vias metabólicas e 607 reações enzimáticas. Um relatório *online*, onde estão listadas todas as vias metabólicas, reações, enzimas e compostos, bem como informações sobre a inferência da presença de cada uma das enzimas (por *EC number*, nome ou similaridade), pode ser encontrado no SchistoCyc.

A Figura 12 apresenta o diagrama das vias preditas gerado pelo SchistoCyc que pode ser acessado pela *Internet*. Como exemplo, foi selecionado um componente da via glicolítica e, dessa forma, disponibilizada uma janela contendo a miniatura da via em questão, informações adicionais e *links* para a página da via e do composto. Ao selecionar o nome da via, o usuário é redirecionado para uma página que contém informações específicas sobre a predição de cada uma das reações, genes que codificam enzimas, compostos e a direção das reações (Figura 11).

Após a construção do banco de dados de vias metabólicas, foi possível efetuar a análise de identificação de reações *chokepoint*, utilizada no presente projeto para seleção de candidatos a alvos terapêuticos. Adicionalmente, essa base de dados permite comparar o metaboloma predito com o de outros organismos e visualizar resultados de experimentos de expressão gênica, proteômicos ou metabolômicos. A comparação de metabolomas por meio do PathwayTools (*Comparative analysis*) pode ser feita de forma gráfica, nos quais são destacadas as enzimas comuns entre os organismos no diagrama de vias metabólicas ou de forma tabular em nível de vias, enzimas, reações ou compostos. Já a ferramenta de



Overview of the *S. mansoni* Metabolic Map

This diagram provides a schematic of all pathways of *S. mansoni* metabolism in the smcyc database. Nodes represent metabolites, with shape indicating class of metabolite (see key to right). Lines represent reactions. Move the mouse over a metabolite icon to identify it. Click on a metabolite icon to navigate to the metabolite page or a related pathway page.

- [Instructions](#)
- [Pathway Tools query page](#)
- [Omics Viewer: Paint omics data onto this diagram](#)
- [Species Comparison: Highlight reactions shared with other organisms](#)

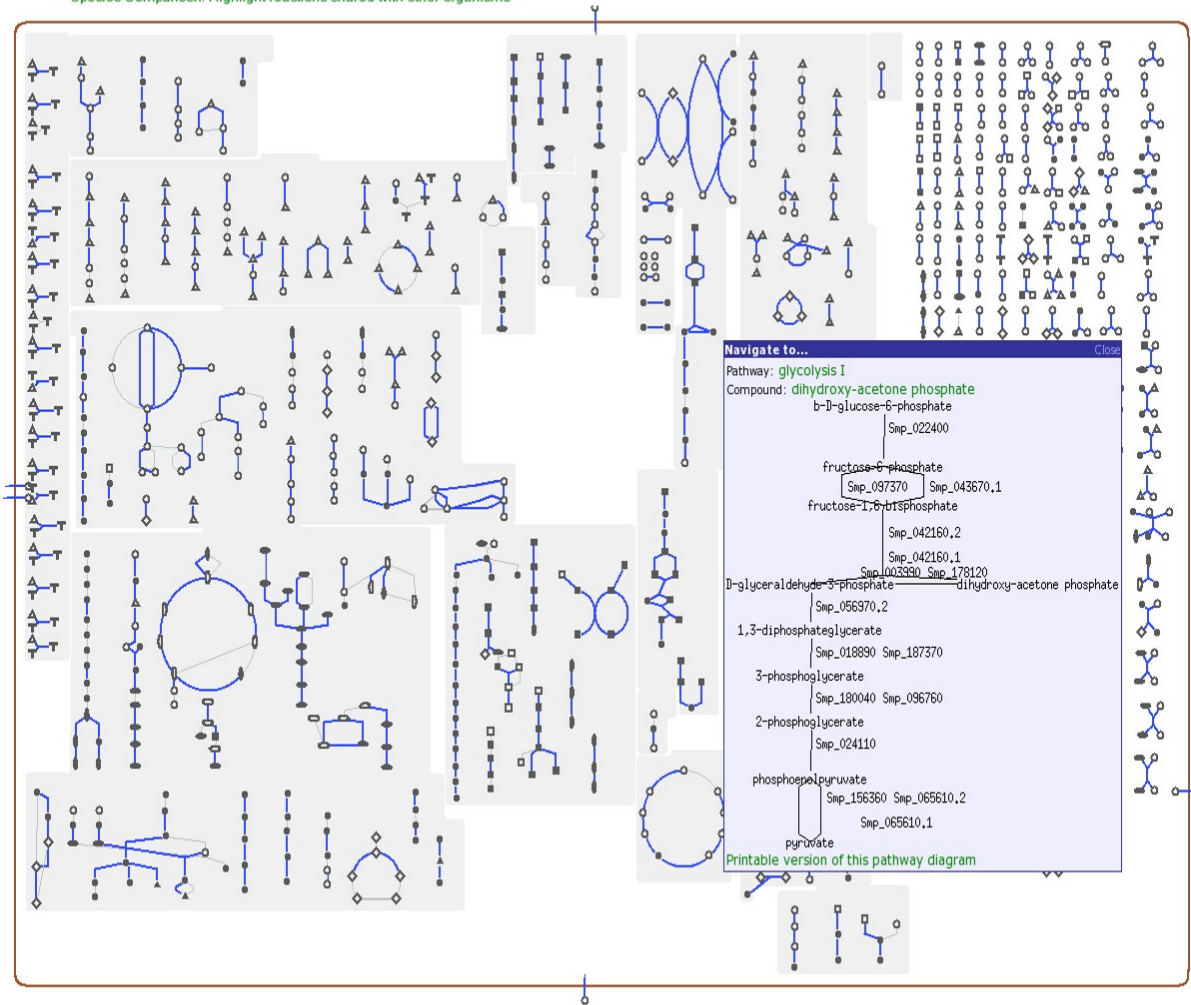
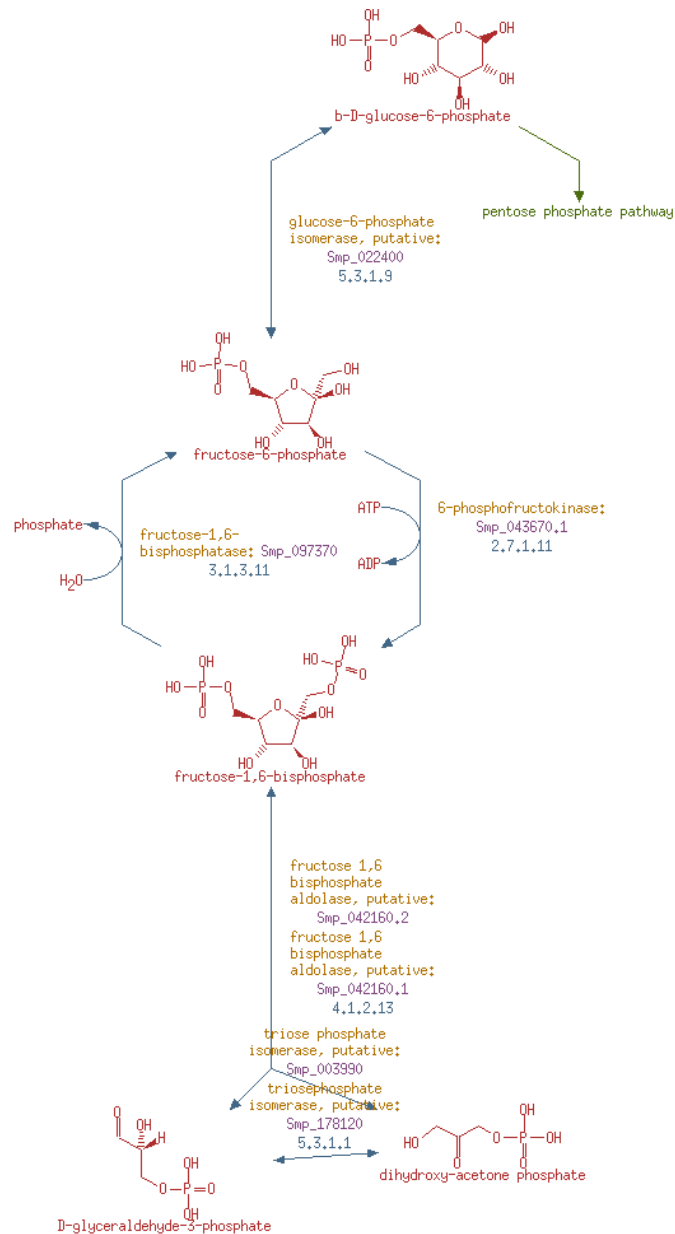


Figura 12: Visão geral das vias metabólicas contidas no SchistoCyc
O diagrama gerado pelo programa PathwayTools contém uma representação esquemática das vias metabólicas previstas, permitindo acesso e visualização das vias, reações, enzimas e compostos. Selecionando-se um determinado composto, é disponibilizada uma janela com informações adicionais e *links* para a página da via em questão, para a página do composto e para a versão formatada para impressão.



Legend for Pathway Diagram

If an enzyme name is shown in bold, there is experimental evidence for this enzymatic activity.

Synonyms: Embden-Meyerhof pathway , glucose degradation , Embden-Meyerhof-Parnas pathway , EMP pathway , glycolysis (plastidic)

Superclasses:

Generation of precursor metabolites and energy -> Glycolysis

Pathway Evidence Glyph:



Key to pathway glyph edge colors:

Green: An enzyme catalyzing this reaction is present in this organism

Blue: An enzyme catalyzing this reaction was identified in this organism by the Pathway Hole Filler

Black: No enzyme catalyzing this reaction has been identified in this organism

Orange: The reaction and any enzyme that catalyzes it (if one has been identified) is unique to this pathway

Figura 11: Exemplo da predição da via glicolítica pelo PathwayTools

Essa figura apresenta parte do diagrama gerado pelo *software* PathwayTools referente a via glicolítica. Estão contidas as reações que compõem a via, os *EC numbers*, os identificadores dos genes, a fórmula do composto e a direção das reações. Todos os elementos desse diagrama são links que levam para uma página que contém informações detalhadas de cada um deles. Logo abaixo do diagrama se localiza uma miniatura que indica se a reação está presente no organismo (verde), se foi predita por similaridade de sequência (azul), se está ausente (preto) e se é uma reação *chokepoint* (laranja).

visualização de resultados de experimentos (*Omic viewer*) utiliza uma escala de cores para destacar a concentração de enzimas em experimentos proteômicos ou a concentração de compostos em experimentos metabolômicos no diagrama de vias metabólicas, enzimas codificadas por genes diferencialmente expressos em experimentos de expressão gênica.

Bader e colaboradores publicaram um metabanco de dados chamado PathGuide, o qual categoriza informações e *links* para diversos programas e bancos de dados relativos à construção de vias metabólicas e de sinalização (Bader et al., 2006). Após análises de algumas ferramentas disponíveis, foi selecionado o *software* PathwayTools produzido pela SRI International, por oferecer de forma unificada ferramentas para predição de vias com base em anotação genômica, publicação de dados na Internet, edição/curadoria dos dados, visualização em forma de diagrama, análise comparativa e análise de balanço de fluxo. Essas características foram cruciais para a escolha do PathwayTools para o nosso estudo, além do amplo canal de comunicação proporcionado por seus desenvolvedores tanto para esclarecimento quanto para implementação de novas funcionalidades e correção de erros.

5.3 – Seleção de candidatos a alvos terapêuticos

A construção de uma base de dados de vias metabólicas nos permitiu efetuar, entre outras análises, buscas por enzimas que catalizam reações *chokepoints*, ou seja, reações que unicamente consomem ou produzem um substrato específico ou produto (Figura 13). Foram identificados 120 alvos potenciais de drogas com base nesse critério (Anexo 9.1).

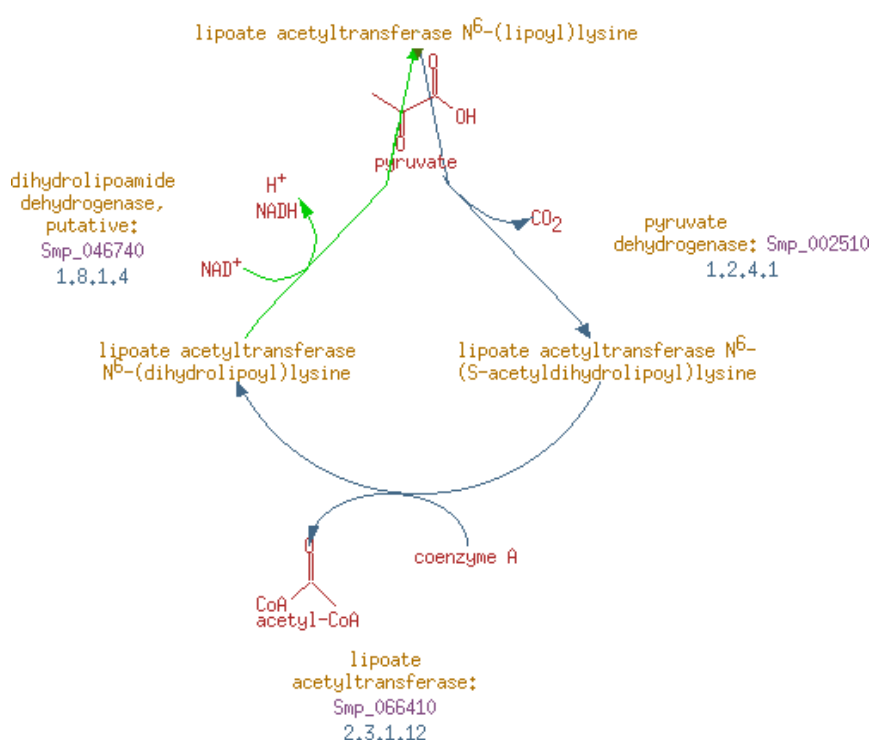


Figura 13: Exemplo de reações *chokepoint*

Essa figura mostra a via de biossíntese de acetil-CoA a partir do piruvato. Todas as três reações são consideradas *chokepoint*. A reação catalizada pela dihidrolipoamida desidrogenase (verde) é considerada *chokepoint* por ser a única reação que consome seu substrato, mas não por produzir piruvato. Já as outras reações (azul) são as únicas que consomem e produzem seus respectivos metabólitos.

Um relatório *online* das enzimas que catalizam reações *chokepoints* foi gerado, o qual pode ser encontrado na página de Internet do SchistoCyc. Nesse relatório, as reações estão agrupadas por vias metabólicas, e informações sobre os genes que codificam enzimas, ortologia com sequências do organismo hospedeiro, similaridade com modelos 3D de proteínas e drogas que atuam nessas enzimas em outros organismos estão integradas.

Por meio do identificador *EC number*, foram selecionadas todas as enzimas que catalizam reações *chokepoint* e que são alvos de droga em outros organismos presentes na base de dados Kegg Drugs (Tabela 4). No total, foram encontradas 35 drogas que atuam em oito diferentes enzimas, sendo que apenas 19 dessas drogas foram aprovadas pelo FDA (Tabela 5).

Tabela 4: Relação de enzimas que catalizam reações *chokepoint* e que são alvos de droga com base no Kegg Drugs.

Enzima	EC Number	Gene	Kegg ID - Droga
Dihidrofolato redutase	1.5.1.3	Smp_175230	D01142 - Sulfadimetoxina D02472 - Atovaquone - proguanil hidrocloreto D02487 - Cloroguanido hidrocloreto D02115 - Metotrexato sódico D00285 - Cotrimoxazol D00145 - Trimetoprima D00142 - Metotrexato D02448 - Pirimetamina - sulfadoxina D00447 - Sulfametoxazol D00488 - Pirimetamina D03771 - Diaveridina D03942 - Edatrexato D06238 - Trimetrexato
Hydroximetilglutaril-CoA redutase (NADPH)	1.1.1.34	Smp_138590	D02258 - Atorvastatina cálcica hidrato D01915 - Rosuvastatina cálcica D01862 - Pitavastatina cálcica D00887 - Atorvastatina cálcica D00892 - Fluvastatina sódica D00889 - Cerivastatina sódica D00893 - Pravastatina sódica D00359 - Lovastatina D00434 - Sinvastatina D03601 - Crilvastatina D03643 - Dalvastatina D03816 - Tenvastatina cálcica
L-aminoácido-aromático descarboxilase	4.1.1.28	Smp_135230	D00253 - Carbidopa hidrato - levodopa D01653 - Benserazida hidrocloreto D00558 - Carbidopa D03082 - Benserazida
Dopamina- β -monooxigenase	1.14.17.1	Smp_163900	D03787 - Nepicastato hidrocloreto
Diidroorotato oxidase	1.3.3.1	Smp_078730	D02472 - Atovaquone - proguanil hidrocloreto D03154 - Brequinar sódico
Orotidina-5'-fosfato descarboxilase	4.1.1.23	Smp_050540	D00224 - Alopurinol
Isoleucina--tRNA ligase	6.1.1.5	Smp_041600 Smp_129650	D01076 - Mupirocina D02195 - Mupirocina cálcica
Álcool desidrogenase	1.1.1.1	Smp_171580	D00707 - Fomepizol

Tabela 5: Relação de drogas que atuam em enzimas que catalizam reações *chokepoint*, nome de identificação no FDA e ação farmacológica.

Droga	Nome FDA	Ação Farmacológica
D01142 - Sulfadimetoxina		Agente anti- infectante
D02472 - Atovaquone	Malarone	Agente anti malária
D02487 - Cloroguanido hidrocloreto	Malarone	Agente anti malária
D00142 - Metotrexato	Metotrexato sódico	Agente antineoplásico - anti metabólico -
D02115 - Metotrexato sódico		Agente imunossupressor
D00285 - Cotrimoxazol	Septra, Bactrim	Agente antibacteriano sintético
D00145 - Trimetoprima	Trimetoprima	Agente antibacteriano sintético
D02448 - Pirimetamina - sulfadoxina	Fansidar	Agente anti malária
D00447 - Sulfametoxazol		Agente anti- infectante
D00488 - Pirimetamina	Daraprima	Agente antiparasita
D03771 - Diaveridina		Agente antibacteriano
D03942 - Edatrexato		Agente Antineoplásico
D06238 - Trimetrexato		Agente antineoplásico - anti metabólico
D02258 - Atorvastatina cálcica	Lipitor	Anti hiperlipidêmico
D01915 - Rosuvastatina cálcica	Crestor	Anti hiperlipidêmico
D01862 - Pitavastatina cálcica		Anti hiperlipidêmico
D00887 - Atorvastatina cálcica	Lipitor	Anti hiperlipidêmico
D00892 - Fluvastatina cálcica	Lescol	Anti hiperlipidêmico
D00889 - Cerivastatina cálcica		Anti hiperlipidêmico
D00893 - Pravastatina cálcica		Anti hiperlipidêmico
D00359 - Lovastatina	Lovastatina, Mevacor	Altprev, Anti hiperlipidêmico
D00434 - Sinvastatina	Sinvastatina, Zocor	Anti hiperlipidêmico
D03601 - Crilvastatina		Anti hiperlipidêmico
D03643 - Dalvastatina		Anti hiperlipidêmico
D03816 - Tenvastatina cálcica		Anti hiperlipidêmico
D00253 - Carbidopa hidrato - levodopa		Agente antiparkinsoniano
D01653 - Benserazida hidrocloreto		Agente antiparkinsoniano
D00558 - Carbidopa	Lodosin	Agente antiparkinsoniano
D03082 - Benserazida		Agente antiparkinsoniano
D03787 - Nepicastato hidrocloreto		Tratamento de insuficiência cardíaca
D03154 - Brequinar sódico		Agente antineoplásico - anti metabólico - Agente imunossupressor
D00224 - Alopurinol	Alopurinol, Ziloprim	Agentes que afetam o metabolismo
D01076 - Mupirocina	Mupirocina, Bactroban,	Agente antibacteriano
D02195 - Mupirocina cálcica	Centany	
D00707 - Fomepizol	Antizol	Inibidor da álcool desidrogenase

No presente trabalho, os principais critérios de seleção de candidatos a alvos terapêuticos foram a identificação de genes que codificam enzimas que catalizam reações *chokepoint* e a identificação de drogas aprovadas pelo FDA que atuam no gene homólogo de outros organismos. Critérios como identificação da presença do gene no organismo hospedeiro, identificação da estrutura tridimensional homóloga, evidências de transcrição ou essencialidade do gene podem ser utilizados no intuito de se selecionar candidatos a alvos terapêuticos em organismos parasitas.

Os dados sobre ortologia provenientes do OrthoMCL, por exemplo, são normalmente utilizados para inferência de função com base na anotação em outras espécies e como fonte de dados para análises de distância evolutiva. Tais dados poderiam ser utilizados para se selecionar genes que atendessem a alguns dos critérios supra citados, como a presença no organismo hospedeiro ou a inferência de sua essencialidade com base no fenótipo identificado em organismos evolutivamente próximos. Em um trabalho recente, do qual nosso grupo participou, Caffrey e colaboradores selecionaram candidatos a alvos terapêuticos contra *S. mansoni* por meio da identificação de ortólogos que possuíam fenótipos deletérios em organismos evolutivamente próximos como *Caenorhabditis elegans* (Caffrey et al., 2009). Neste trabalho, análises por similaridade de sequência para inferência de homologia foram realizadas, sendo que os dados contidos no OrthoMCL poderiam ter sido utilizados para poupar tal processamento computacional.

5.4 – Busca na literatura por informações sobre os genes selecionados

O Anexo 9.2 apresenta resultados da busca na literatura fornecidos pelo programa PubMatrix. Esta tabela contém a frequência de co-ocorrência das enzimas que catalizam reações *chokepoint* e os termos modificadores. A utilização do *software* PubMatrix limitou-se à obtenção automatizada de uma lista contendo os *links* para todos artigos que atendessem aos critérios da busca. A ferramenta utilizada para pesquisa na literatura facilitou, consideravelmente, a localização e o acesso aos artigos. A ausência de ferramentas de análise estatística para avaliação de relevância e correlação dos resultados não foi de grande importância já que os artigos encontrados foram analisados. O Anexo 9.3 apresenta o resultado da inspeção dos artigos encontrados, no qual foram listados os artigos que relatam atividade esquistossomicida de determinadas drogas e artigos que relatam a caracterização das enzimas selecionadas em *S. mansoni*.

A verificação da literatura selecionada por meio do PubMatrix mostrou que algumas das 120 enzimas que catalizam reações *chokepoint* foram caracterizadas experimentalmente e são alvos de droga tanto em *S. mansoni* quanto em outros organismos, indicando a eficácia da metodologia utilizada. Nos parágrafos seguintes, listamos artigos que demonstram experimentos, nos quais foi relatada a inibição de enzimas que catalizam reações *chokepoint* em *S. mansoni*.

Na via metabólica do Mevalonato (Figura 14), quase todas as reações são *chokepoint*. A via do Mevalonato está presente em todos os eucariotos superiores e em diversas bactérias, além de ser parte integrante da supervia de biossíntese de colesterol. Desta via, a enzima 3-hidroxiacetilglutaril-coenzima A redutase (HMG-CoA redutase, 1.1.1.34) foi amplamente estudada (Rajkovic et al., 1989; Vandewaa et al., 1989). Foi demonstrado que a inibição dessa

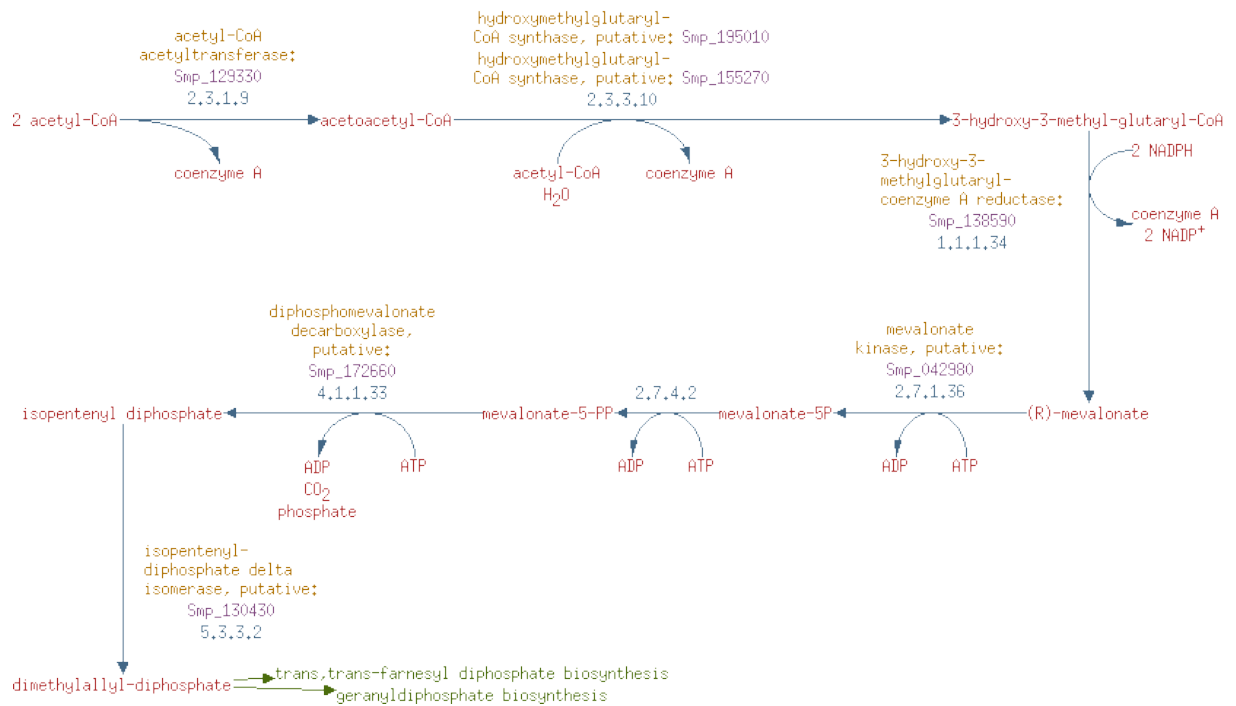


Figura 14: Diagrama da via de mevalonato que contém a enzima HMG-CoA redutase (1.1.1.34)

enzima é acompanhada pela interrupção da produção de ovos pela fêmea e pela redução da eficiência de glicosilação de proteínas (Chen et al., 1991; Chen et al., 1990). Chen e colaboradores também demonstraram que essa enzima é significativamente diferente da homóloga em mamíferos e que sua inibição utilizando a droga Mevinolin tem efeito esquistossomicida. Em outros trabalhos, Araújo e colaboradores demonstraram significativa diminuição do número de ovos nos intestinos de animais tratados com a droga Lovastatina (Araújo et al., 2002), conhecida por produzir um inibidor competitivo da HMG-CoA redutase, além de demonstrarem a atividade dessa droga associada com praziquantel e oxaminiquine (Araújo et al., 2008a). De fato, pode-se notar que as drogas que atuam nessa enzima são anti hiperlipidêmicas (Tabela 4), ou seja, drogas que inibem a biossíntese de colesterol. Em um trabalho recente, Stanley e colaboradores demonstraram os efeitos da presença de *S. mansoni* e seus ovos no colesterol circulante de camundongos (Stanley et al., 2009). Foi demonstrado que a diminuição do colesterol circulante é mediada por fatores liberados dos ovos de *S.*

mansoni por meio de injeção intraperitoneal de ovos e fatores derivados.

A enzima L-DOPA descarboxilase (4.1.1.28) tem um papel importante na via metabólica de biossíntese de catecolaminas (hormônios liberados pela glândula adrenal) (Figura 15). Essa enzima é conhecida por descarboxilar 5-hidroxitriptofano e L-DOPA em *S. mansoni* e sua inibição pode apresentar efetividade terapêutica contra esquistossomose (Catto, 1981). Metildopa, uma droga utilizada no tratamento de hipertensão, responsável pela inibição da enzima homóloga em mamíferos, foi testada e os resultados demonstraram efeitos inibidores em *S. mansoni*. Catto e colaboradores infectaram camundongos com 50 cercárias cada e os trataram com 200 mg/kg de Metildopa. Um número reduzido de vermes adultos foi quantificado no momento da perfusão (7-8 semanas).

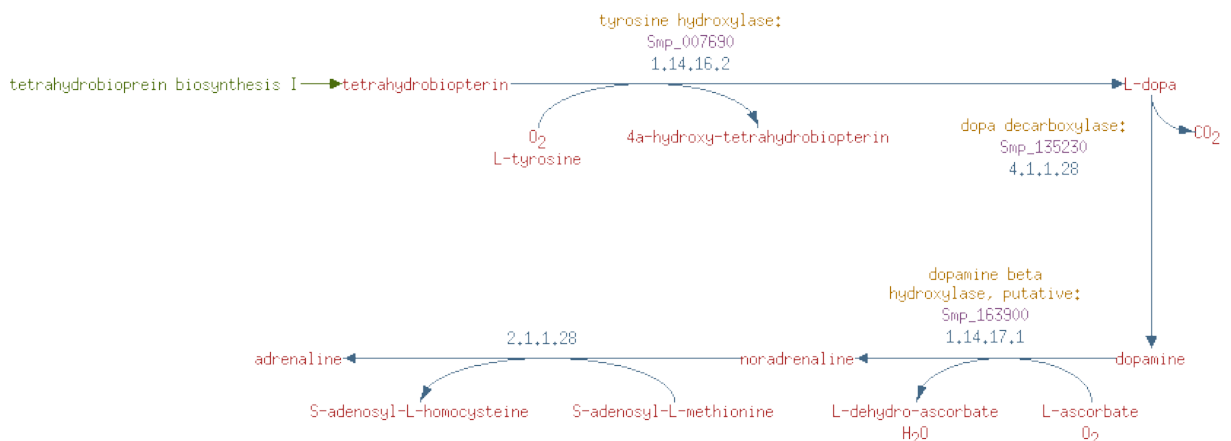


Figura 15: Diagrama da via de biossíntese de catecolaminas que contém a enzima LDOPA descarboxilase (4.1.1.28)

A via glicolítica (Figura 11) contém 4 enzimas que catalizam reações *chokepoint*. Uma delas, a fosfofrutoquinase (2.7.1.11), conhecida por catalizar a transferência de um fosfato para a frutose-6-fosfato, foi bastante estudada (Mansour et al., 2000; Ding et al., 1994). Foi descrita sua inibição por antimoniais, como tartarato de potássio e antimônio e o stibofen (Su et al., 1996; Bueding et al., 1957).

A via de degradação de triptofano (Figura 16) contém duas enzimas que catalizam reações *chokepoint*. Uma delas, a quinureninase (3.7.1.3), foi estudada por Amer e

colaboradores (Amer et al., 1968) e seus achados indicam que o tratamento com tártaro emético inibe essa reação.

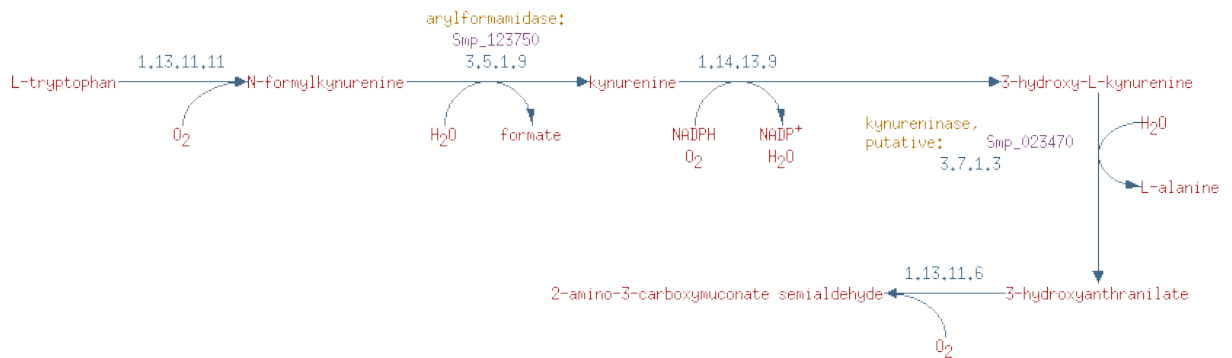


Figura 16: Diagrama da via de degradação de triptofano que contém a enzima quinureninase (3.7.1.3)

Em conclusão, a sobreposição de dados da literatura com nossos achados de ensaios *in silico* demonstraram a viabilidade da abordagem utilizada.

5.5 – Validação experimental dos melhores candidatos

Drogas conhecidas por atuar em enzimas de outros organismos identificadas como candidatas a alvos terapêuticos em *S. mansoni*, utilizando a técnica de integração de dados, foram testadas no presente trabalho com o objetivo de se avaliar a existência de atividade esquistossomicida. A dosagem de cada uma das drogas foi determinada através de revisão da literatura científica na qual foram selecionados artigos que demonstravam a utilização das drogas em questão em testes *in vivo* em camundongos (Tabela 6).

Tabela 6: Máxima dosagem encontrada na literatura científica

Droga	Dosagem	Referência
Alopurinol	300mg/kg	(Horiuchi et al., 2000)
Metotrexato	20mg/kg	(Horie et al. 1998)
Pravastatina	100mg/kg	(Thelen et al. 2006)
Trimetoprima	25mg/kg	(Keith et al. 1989)

A Tabela 7 mostra o resultado do experimento *in vivo* para avaliação de atividade esquistossomicida. Merecem destaque as drogas Alopurinol e Metotrexato que apresentaram respectivamente 30,5% e 35,7% de redução no número de vermes encontrados em relação ao grupo controle ($P < 0,05$). A taxa de redução não foi estatisticamente significativa para as demais drogas. A taxa de redução foi calculada de acordo com a equação utilizada por Cioli (Cioli et al., 2004):

$$\% \text{ redução} = \frac{\text{no. médio de vermes no grupo controle} - \text{no. médio de vermes no grupo tratado}}{\text{no. médio de vermes no grupo controle}} \times 100$$

Tabela 7: Resultado do experimento de eficácia das drogas em *S. mansoni*

Droga	Dosagem (mg/kg)	Camundongos		Média de Vermes	Taxa de Redução (%)
		Tratados	Examinados		
Alopurinol	300	10	4	25,25	30,5
Metotrexato	20	10	8	23,38	35,67
Pravastatina	65	8	7	32,57	10,35
Trimetoprima	25	20	14	36,25	0,23
Controle	0	9	9	36,33	0

Este é um resultado muito interessante quando levadas em consideração as taxas de redução de vermes adultos alcançadas com PZQ, droga mundialmente usada no combate à esquistossomose nos dias atuais. Em dois testes de drogas recentemente publicados, foi observado que 200mg/kg de PZQ reduziu o número de vermes em aproximadamente 30,3% (Araújo et al., 2008a) e 23,0% (Araújo et al., 2008b). Dados anteriores a estes, provenientes de um experimento em maior escala demonstrou que o PZQ, apresentou aproximadamente 50% de redução no número de vermes com dosagens de aproximadamente 100 mg/kg (Cioli et al., 2004). Essas informações são apenas de caráter ilustrativo. Para se definitivamente obter uma comparação com PZQ seria necessária a implementação de um grupo controle PZQ no experimento.

Uma característica importante do PZQ é sua baixa toxicidade para o hospedeiro vertebrado. As drogas selecionadas no presente estudo precisam ser investigadas mais detalhadamente para inferência de tal propriedade. Apesar do experimento nos mostrar indícios de atividade esquistossomicida, devemos levar em consideração que tais drogas foram originalmente desenvolvidas para outros propósitos.

O Alopurinol reduz a produção de ácido úrico por meio da inibição de reações bioquímicas que precedem sua formação. Sua principal utilização é no tratamento de hiperuricemia (DailyMed: About DailyMed. Disponível em:

<<http://dailymed.nlm.nih.gov/dailymed/about.cfm>>. Acessado em: 10 de maio de 2009).

O Metotrexato é um antimetabólito usado no tratamento de câncer e doenças autoimunes, inibindo a enzima ácido fólico redutase (1.5.1.3), integrante da via de metabolismo de ácido fólico (DailyMed: About DailyMed. Disponível em: <<http://dailymed.nlm.nih.gov/dailymed/about.cfm>>. Acessado em: 10 de maio de 2009).

É importante destacar que esse experimento tem o objetivo exclusivo de, preliminarmente, demonstrar ação esquistossomicida dos compostos escolhidos e não pretende sugerir novas abordagens para a terapêutica da esquistossomose. Esse projeto se limita à primeira fase do processo de descoberta de drogas, conhecida como descoberta de alvos.

Para se obter informações concretas sobre a eficácia de compostos são necessários testes *in vivo* para se determinar a correta dosagem e a toxicidade dos mesmos. Adicionalmente, as técnicas computacionais de modelagem molecular e *docking* de pequenas moléculas podem ser utilizadas. Tais técnicas permitem manipular a estrutura química do composto para aumentar sua afinidade com a enzima do parasito e, conseqüentemente, diminuir afinidade com a enzima do hospedeiro vertebrado, no intuito de minimizar a toxicidade.

O processo de busca de novas indicações para drogas existentes, como demonstrado no presente trabalho, é conhecido como reposicionamento de drogas. A grande vantagem da utilização de tal abordagem se deve ao fato de as drogas em questão já serem amplamente estudadas e caracterizadas, além de terem sido aprovadas para uso clínico pelos órgãos regulamentadores (Ashburn et al., 2004). Considerando-se que o custo de desenvolvimento de novas drogas é consideravelmente alto e não tem atingido os resultados esperados, tal processo tem sido amplamente praticado pela indústria farmacêutica.

VI – CONCLUSÕES

- O banco de dados construído (SchistoDB) tem sido utilizado por pesquisadores de todo o mundo e temos recebido contatos solicitando a implementação de alterações e novas funcionalidades.
- Além das informações genômicas, também estão disponíveis informações sobre o metabolismo do parasito, obtidas através da predição computacional das vias metabólicas.
- O banco de dados construído (SchistoDB) facilita execução de diversas análises além da busca de candidatos a alvos terapêuticos, como: filogenia, genômica comparativa, caracterização de proteínas, proteômica e expressão gênica.
- A metodologia descrita demonstrou que a técnica de seleção de genes candidatos a alvos terapêuticos por meio da integração de dados relativos ao genoma pode muito contribuir no processo de direcionamento de recursos para experimentos laboratoriais.
- Além de obtermos indícios de atividade esquistossomicida de duas drogas por meio de experimentos *in vivo*, demonstramos que a metodologia descrita apresenta eficácia na diminuição do número de genes candidatos a alvos terapêuticos para um número considerável de experimentação laboratorial.

VII – LIMITAÇÕES & PERSPECTIVAS

A principal limitação da utilização da técnica de integração de dados na pesquisa está no fato dessa técnica ser totalmente dependente da existência de conhecimento prévio dos dados que compõem a base. Em outras palavras, não é possível descobrir uma nova droga utilizando tal técnica, mas sim obter indicações sobre potenciais alvos e também possíveis novos usos para drogas que tem seu mecanismo de ação descrito. Não é possível descobrir uma nova via metabólica, apenas inferir a existência de vias metabólicas já caracterizadas.

Por outro lado, as vantagens da técnica de integração de dados são o baixo custo de implementação e, especialmente, a possibilidade de re-execução de todo o processo com regularidade, com objetivo de se integrarem as informações disponíveis mais atualizadas. Vale destacar que a metodologia da presente investigação pode ser muito útil para análise de dados experimentais, visto que a publicação de tais resultados tem aumentado exponencialmente e de forma sistemática e estruturada.

A metodologia apresentada demonstra ainda grande potencial para diminuição dos custos de experimentos em laboratório. Com base em informações sobre o genoma e metabolismo do parasito, sobre o genoma do hospedeiro e sobre drogas, torna-se possível priorizar os candidatos à alvos terapêuticos aumentando as chances de sucesso nos primeiros experimentos.

Esperamos em breve oferecer novas versões do SchistoDB integrando diferentes tipos de dados, como resultados de expressão gênica, marcadores populacionais e genômica comparativa com outras espécies do gênero *Schistosoma*.

VIII – REFERÊNCIAS BIBLIOGRÁFICAS

- About Perl - perl.org. Recuperado Abril 6, 2009, de <http://www.perl.org/about.html>.
- Agüero, F., Zheng, W., Weatherly, D. B., Mendes, P. & Kissinger, J. C. (2006). TcruziDB: an integrated, post-genomics community resource for *Trypanosoma cruzi*. *Nucleic Acids Research*, 34(Database issue), D428-431. doi: 10.1093/nar/gkj108.
- Aitken, R., Coulson, P. S., Dixon, B. & Wilson, R. A. (1987). Radiation-resistant acquired immunity of vaccinated mice to *Schistosoma mansoni*. *The American Journal of Tropical Medicine and Hygiene*, 37(3), 570-577.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), 3389–402.
- Amer, M. S., Abdel-Daim, M. H. & Abdel-Tawab, G. A. (1968). Studies with tryptophan metabolites in vitro. Kynurenine metabolism in kidneys of mice infested with *Schistosoma mansoni*. *The Biochemical Journal*, 109(4), 613–5. doi: 5683509.
- ApiDB Repository. Recuperado Abril 7, 2009, de <https://www.cbil.upenn.edu/svn/apidb/>.
- Araújo, N., Kohn, A., Oliveira, Á. A. D. & Katz, N. (2002). *Schistosoma mansoni*: ação da lovastatina no modelo murino *Schistosoma mansoni*: the action of lovastatin on the murine model. *Revista da Sociedade Brasileira de Medicina Tropical*, 35, 35-38.
- Araújo, N., Mattos, A. C. A. D., Coelho, P. M. Z. & Katz, N. (2008). Association of oxamniquine praziquantel and clonazepam in experimental Schistosomiasis mansoni. *Memórias Do Instituto Oswaldo Cruz*, 103(8), 781-785.
- Araújo, N., Mattos, A. C. A. D., Sarvel, A. K., Coelho, P. M. Z. & Katz, N. (2008). Oxamniquine, praziquantel and lovastatin association in the experimental Schistosomiasis mansoni. *Memórias do Instituto Oswaldo Cruz*, 103, 450-454.
- Aronstein, W. S. & Strand, M. (1984). Gender-specific and pair-dependent glycoprotein antigens of *Schistosoma mansoni*. *The Journal of Parasitology*, 70(4), 545-557.
- Ashburn, T. T. & Thor, K. B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nature Reviews. Drug Discovery*, 3(8), 673-683. doi: 10.1038/nrd1468.
- Aurrecochea, C., Brestelli, J., Brunk, B. P., Dommer, J., Fischer, S., Gajria, B. et al. (2009).

- PlasmoDB: a functional genomic database for malaria parasites. *Nucleic Acids Research*, 37(Database issue), D539-543. doi: 10.1093/nar/gkn814.
- Bader, G. D., Cary, M. P. & Sander, C. (2006). Pathguide: a pathway resource list. *Nucleic acids research*, 34(Database issue), D504–6.
- Bayne, C. J. (2009). Successful parasitism of vector snail *Biomphalaria glabrata* by the human blood fluke (trematode) *Schistosoma mansoni*: A 2009 assessment. *Molecular and Biochemical Parasitology*, 165(1), 8-18. doi: 10.1016/j.molbiopara.2009.01.005.
- Becker, K. G., Hosack, D. A., Dennis, G., Lempicki, R. A., Bright, T. J., Cheadle, C. et al. (2003). PubMatrix: a tool for multiplex literature mining. *BMC Bioinformatics*, 4, 61. doi: 10.1186/1471-2105-4-61.
- Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., Rapp, B. A. & Wheeler, D. L. (2000). GenBank. *Nucleic Acids Research*, 28(1), 15–8. doi: PMC102453.
- Bergquist, N. R. (2002). Schistosomiasis: from risk assessment to control. *Trends in Parasitology*, 18(7), 309-314. doi: 10.1016/S1471-4922(02)02301-2.
- Bergquist, R. (2008). A century of schistosomiasis research. *Acta Tropica*, 108(2-3), 65-68. doi: 10.1016/j.actatropica.2008.11.002.
- Berriman, M., Haas, B. J., LoVerde, P. T., Wilson, R. A. et al. (2009). The genome of the blood fluke *Schistosoma mansoni*. *Nature*, 460(7253), 352-358. doi: 10.1038/nature08160.
- Bueding, E. & Mansour, J. M. (1957). The relationship between inhibition of phosphofructokinase activity and the mode of action of trivalent organic antimonials on *Schistosoma mansoni*. *British Journal of Pharmacology and Chemotherapy*, 12(2), 159–65. doi: 13446367.
- Caffrey, C. R., Rohwer, A., Oellien, F., Marhöfer, R. J., Braschi, S., Oliveira, G. et al. (2009). A comparative chemogenomics strategy to predict potential drug targets in the metazoan pathogen, *Schistosoma mansoni*. *PloS One*, 4(2), e4413. doi: 10.1371/journal.pone.0004413.
- Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P., Krummenacker, M., Latendresse, M. et al. (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Research*, 36(Database issue), D623–31. doi: PMC2238876.
- Catto, B. A. (1981). *Schistosoma mansoni*: decarboxylation of 5-hydroxytryptophan, L-dopa,

- and L-histidine in adult and larval schistosomes. *Experimental Parasitology*, 51(1), 152–7. doi: 6970143.
- Cheever, A. W., Macedonia, J. G., Mosimann, J. E. & Cheever, E. A. (1994). Kinetics of egg production and egg excretion by *Schistosoma mansoni* and *S. japonicum* in mice infected with a single pair of worms. *The American Journal of Tropical Medicine and Hygiene*, 50(3), 281-95.
- Chen, F., Mackey, A. J., Stoeckert, C. J. & Roos, D. S. (2006). OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. *Nucleic Acids Research*, 34(Database issue), D363–8. doi: 34/suppl_1/D363.
- Chen, G. Z., Foster, L. & Bennett, J. L. (1990). Antischistosomal action of mevinolin: evidence that 3-hydroxy-methylglutaryl-coenzyme a reductase activity in *Schistosoma mansoni* is vital for parasite survival. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 342(4), 477–82. doi: 2123968.
- Chen, G. Z., Foster, L. & Bennett, J. L. (1991). Purification and characterization of 3-hydroxymethylglutaryl-coenzyme A reductase of *Schistosoma mansoni*: regulation of parasite enzyme activity differs from mammalian host. *Experimental Parasitology*, 73(1), 82–92. doi: 1905241.
- Chitsulo, L., Engels, D., Montresor, A. & Savioli, L. (2000). The global status of schistosomiasis and its control. *Acta Tropica*, 77(1), 41-51.
- Chokepoint Analysis Tool. Recuperado Junho 4, 2009, de <http://bioinformatics.ai.sri.com/ptools/chokepoint/chokepoint-analysis.html>.
- Cioli, D., Botros, S. S., Wheatcroft-Francklow, K., Mbaye, A., Southgate, V., Tchuenté, L. T. et al. (2004). Determination of ED50 values for praziquantel in praziquantel-resistant and -susceptible *Schistosoma mansoni* isolates. *International Journal for Parasitology*, 34(8), 979-987. doi: 10.1016/j.ijpara.2004.05.001.
- Cioli, D. & Pica-Mattoccia, L. (2003). Praziquantel. *Parasitology Research*, 90 Supp 1, S3-9. doi: 10.1007/s00436-002-0751-z.
- DailyMed: About DailyMed. Recuperado Maio 10, 2009, de <http://dailymed.nlm.nih.gov/dailymed/about.cfm>.
- Davidson, S. B., Crabtree, J., Brunk, B. P., Schug, J., Tannen, V., Overton, G. C. et al. (2001). K2/Kleisli and GUS: experiments in integrated access to genomic data sources. *IBM Syst. J.*, 40(2), 512–531.

- Dávila, A. M. R., Mendes, P. N., Wagner, G., Tschoeke, D. A., Cuadrat, R. R. C., Liberman, F. et al. (2008). ProtozoaDB: dynamic visualization and exploration of protozoan genomes. *Nucleic Acids Research*, 36(Database issue), D547–52. doi: gkm820.
- Davis, A. H. (1984). Schistosomiasis. In *Epidemiology and the Community Control of Disease in Warm Climate Countries* (2nd ed., pp. 389-412). Churchill Livingstone.
- Degrave, W. M., Melville, S., Ivens, A. & Aslett, M. (2001). Parasite genome initiatives. *International Journal for Parasitology*, 31(5-6), 531-535. doi: 10.1016/S0020-7519(01)00144-8.
- DeMarco, R. & Verjovski-Almeida, S. (2009). Schistosomes—proteomics studies for potential novel vaccines and drug targets. *Drug Discovery Today*, 14(9-10), 472-478. doi: 10.1016/j.drudis.2009.01.011.
- Ding, J., Su, J. G. & Mansour, T. E. (1994). Cloning and characterization of a cDNA encoding phosphofructokinase from *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 66(1), 105–10. doi: 7984171.
- Doenhoff, M. J., Cioli, D. & Utzinger, J. (2008). Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Current Opinion in Infectious Diseases*, 21(6), 659-667. doi: 10.1097/QCO.0b013e328318978f.
- Eilbeck, K., Lewis, S., Mungall, C., Yandell, M., Stein, L., Durbin, R. et al. (2005). The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biology*, 6(5), R44. doi: 10.1186/gb-2005-6-5-r44.
- El-Sayed, N. M. A., Bartholomeu, D., Ivens, A., Johnston, D. A. & LoVerde, P. T. (2004). Advances in schistosome genomics. *Trends in Parasitology*, 20(4), 154–7. doi: 15099549.
- EuPathDB : The Eukaryotic Pathogen genome resource. Recuperado Abril 7, 2009, de <http://eupathdb.org>.
- Ezeamama, A. E., Friedman, J. F., Acosta, L. P., Bellinger, D. C., Langdon, G. C., Manalo, D. L. et al. (2005). Helminth infection and cognitive impairment among Filipino children. *The American Journal of Tropical Medicine and Hygiene*, 72(5), 540-8.
- Franco, G. R., Valadão, A. F., Azevedo, V. & Rabelo, M. L. (2000). The Schistosoma gene discovery program: state of the art. *International Journal for Parasitology*, 30(4), 453–463. doi: 10.1016/S0020-7519(00)00020-5.
- Gajria, B., Bahl, A., Brestelli, J., Dommer, J., Fischer, S., Gao, X. et al. (2008). ToxoDB: an

- integrated *Toxoplasma gondii* database resource. *Nucleic Acids Research*, 36(Database issue), D553-6. doi: PMC2238934.
- GeneDB. Recuperado Maio 15, 2009, de <http://www.genedb.org/>.
- Green, M. L. & Karp, P. D. (2004). A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC bioinformatics*, 5, 76.
- Gryseels, B., Mbaye, A., De Vlas, S. J., Stelma, F. F., Guissé, F., Van Lieshout, L. et al. (2001). Are poor responses to praziquantel for the treatment of *Schistosoma mansoni* infections in Senegal due to resistance? An overview of the evidence. *Tropical Medicine & International Health: TM & IH*, 6(11), 864-873.
- Gryseels, B., Polman, K., Clerinx, J. & Kestens, L. (2006). Human schistosomiasis. *Lancet*, 368(9541), 1106-18. doi: 10.1016/S0140-6736(06)69440-3.
- gusdb.org | Schema Browser. Recuperado Maio 7, 2009, de <http://www.gusdb.org/SchemaBrowser>.
- GUS: The Genomics Unified Schema. Recuperado Maio 7, 2009, de <http://www.gusdb.org/>.
- Haas, B. J., Berriman, M., Hirai, H., Cerqueira, G. G., Loverde, P. T. & El-Sayed, N. M. (2007). *Schistosoma mansoni* genome: closing in on a final gene set. *Experimental Parasitology*, 117(3), 225–8. doi: S0014-4894(07)00168-3.
- Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R. et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, 32(Database issue), D258–61. doi: 14681407.
- Heiges, M., Wang, H., Robinson, E., Aurrecoechea, C., Gao, X., Kaluskar, N. et al. (2006). CryptoDB: a Cryptosporidium bioinformatics resource update. *Nucleic Acids Research*, 34(Database issue), D419-22. doi: PMC1347441.
- Hokke, C. H., Deelder, A. M., Hoffmann, K. F. & Wuhrer, M. (2007). Glycomics-driven discoveries in schistosome research. *Experimental Parasitology*, 117(3), 275-283. doi: 10.1016/j.exppara.2007.06.003.
- Huang, X. & Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Research*, 9(9), 868–77. doi: PMC310812.
- JavaServer Pages Technology. Recuperado Abril 6, 2009, de <http://java.sun.com/products/jsp/>.
- Jukes, M., Nokes, C. A., Alcock, K. J., Lambo, J. K., Kihamia, C., Ngorosho, N. et al. (2002). Heavy schistosomiasis associated with poor short-term memory and slower reaction times in Tanzanian schoolchildren. *Tropical Medicine & International Health: TM &*

- IH*, 7(2), 104-17.
- Kariuki, T. M., Farah, I. O., Yole, D. S., Mwenda, J. M., Van Dam, G. J., Deelder, A. M. et al. (2004). Parameters of the attenuated schistosome vaccine evaluated in the olive baboon. *Infection and Immunity*, 72(9), 5526-5529. doi: 10.1128/IAI.72.9.5526-5529.2004.
- Karp, P. D., Paley, S. & Romero, P. (2002). The Pathway Tools software. *Bioinformatics (Oxford, England)*, 18 Suppl 1, S225–32.
- Katz, N., Faria, I. & Reis, F. A. (1986). Modernos conhecimentos sobre esquistossomose mansônica. *Academia Mineira de Medicina*.
- Katz, N. & Peixoto, S. V. (2000). Análise crítica da estimativa do número de portadores de esquistossomose mansoni no Brasil Critical analysis of the estimated number of schistosomiasis mansoni carriers in Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, 33, 303-308.
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Research*, 12(4), 656–64. doi: PMC187518.
- Lee, T. J., Pouliot, Y., Wagner, V., Gupta, P., Stringer-Calvert, D. W. J., Tenenbaum, J. D. et al. (2006). BioWarehouse: a bioinformatics database warehouse toolkit. *BMC Bioinformatics*, 7, 170. doi: 10.1186/1471-2105-7-170.
- Lenzerini, M. (2002). Data integration: a theoretical perspective. In *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (pp. 233-246). Madison, Wisconsin: ACM. doi: 10.1145/543613.543644.
- Le, T. H., Blair, D. & McManus, D. P. (2000). Mitochondrial DNA sequences of human schistosomes: the current status. *International Journal for Parasitology*, 30(3), 283–90. doi: 10719121.
- Loging, W., Harland, L. & Williams-Jones, B. (2007). High-throughput electronic biology: mining information for drug discovery. *Nature reviews. Drug discovery*, 6(3), 220–30.
- Loukas, A., Tran, M. & Pearson, M. S. (2007). Schistosome membrane proteins as vaccines. *International Journal for Parasitology*, 37(3-4), 257-263. doi: 10.1016/j.ijpara.2006.12.001.
- LoVerde, P. T., Hirai, H., Merrick, J. M., Lee, N. H. & El-Sayed, N. (2004). *Schistosoma mansoni* genome project: an update. *Parasitology International*, 53(2), 183-92. doi: 15081950.

- Loverde, P. T. & Chen, L. (1991). Schistosome female reproductive development. *Parasitology Today (Personal Ed.)*, 7(11), 303-8.
- Mansour, J. M., McCrossan, M. V., Bickle, Q. D. & Mansour, T. E. (2000). *Schistosoma mansoni* phosphofructokinase: immunolocalization in the tegument and immunogenicity. *Parasitology*, 120 (Pt 5), 501–11. doi: 10840980.
- Masoudi-Nejad, A., Goto, S., Endo, T. R. & Kanehisa, M. (2007). KEGG Bioinformatics Resource for Plant Genomics Research. *Methods in molecular biology (Clifton, N.J.)*, 406, 437–58.
- Material Suplementar online. Recuperado Junho 10, 2009, de <http://bioinfo.cpqrr.fiocruz.br/content/2009/06/10/material-suplementar-tese-de-adhemar-zerlotini-neto>.
- McManus, D. P. & Hope, M. (1993). Molecular variation in the human schistosomes. *Acta Tropica*, 53(3-4), 255-76.
- Mungall, C. J. & Emmert, D. B. (2007). A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics (Oxford, England)*, 23(13), i337-346. doi: 10.1093/bioinformatics/btm189.
- Oliveira, G. & Johnston, D. A. (2001). Mining the schistosome DNA sequence database. *Trends in Parasitology*, 17(10), 501-3.
- Oliveira, G. (2007). The *Schistosoma mansoni* transcriptome: an update. *Experimental Parasitology*, 117(3), 229–35. doi: S0014-4894(07)00167-1.
- Pellegrino, J. & Siqueira, A. F. (1956). [A perfusion technic for recovery of *Schistosoma mansoni* from experimentally infected guinea pigs.]. *Revista Brasileira De Malariologia E Doenças Tropicais. Publicações Avulsas*, 8(4), 589-97.
- Pica-Mattoccia, L. & Cioli, D. (2004a). Sex- and stage-related sensitivity of *Schistosoma mansoni* to in vivo and in vitro praziquantel treatment. *International Journal for Parasitology*, 34(4), 527-533. doi: 10.1016/j.ijpara.2003.12.003.
- Pica-Mattoccia, L. & Cioli, D. (2004b). Sex- and stage-related sensitivity of *Schistosoma mansoni* to in vivo and in vitro praziquantel treatment. *International Journal for Parasitology*, 34(4), 527-533. doi: 10.1016/j.ijpara.2003.12.003.
- PubMed Overview. Recuperado Maio 6, 2009, de <http://www.ncbi.nlm.nih.gov/corehtml/query/static/overview.html>.
- Rajkovic, A., Simonsen, J. N., Davis, R. E. & Rottman, F. M. (1989). Molecular cloning and

- sequence analysis of 3-hydroxy-3-methylglutaryl-coenzyme A reductase from the human parasite *Schistosoma mansoni*. *Proceedings of the National Academy of Sciences of the United States of America*, 86(21), 8217–21. doi: 2813388.
- Rollinson, D. (1988). *The Biology of Schistosomes: From Genes to Latrines*. Academic Press.
- Schramm, G., Hamilton, J., Balog, C., Wuhrer, M., Gronow, A., Beckmann, S. et al. (2009). Molecular characterisation of kappa-5, a major antigenic glycoprotein from *Schistosoma mansoni* eggs. *Molecular and Biochemical Parasitology*, 166(1), 4-14. doi: 10.1016/j.molbiopara.2009.02.003.
- Searls, D. B. (2005). Data integration: challenges for drug discovery. *Nature reviews. Drug discovery*, 4(1), 45–58.
- Shah, S. P., Huang, Y., Xu, T., Yuen, M. M. S., Ling, J. & Ouellette, B. F. F. (2005). Atlas - a data warehouse for integrative bioinformatics. *BMC Bioinformatics*, 6, 34. doi: 10.1186/1471-2105-6-34.
- de Silva, N., Guyatt, H. & Bundy, D. (1997). Anthelmintics. A comparative review of their clinical pharmacology. *Drugs*, 53(5), 769-788.
- Stanley, R. G., Jackson, C. L., Griffiths, K. & Doenhoff, M. J. (2009). Effects of *Schistosoma mansoni* worms and eggs on circulating cholesterol and liver lipids in mice. *Atherosclerosis*. doi: 10.1016/j.atherosclerosis.2009.04.037.
- Stein, L. D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A. et al. (2002). The generic genome browser: a building block for a model organism system database. *Genome Research*, 12(10), 1599–610. doi: 12368253.
- Steinmann, P., Keiser, J., Bos, R., Tanner, M. & Utzinger, J. (2006). Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases*, 6(7), 411-25. doi: 10.1016/S1473-3099(06)70521-7.
- Stelma, F. F., Talla, I., Sow, S., Kongs, A., Niang, M., Polman, K. et al. (1995). Efficacy and side effects of praziquantel in an epidemic focus of *Schistosoma mansoni*. *The American Journal of Tropical Medicine and Hygiene*, 53(2), 167-170.
- Su, J. G., Mansour, J. M. & Mansour, T. E. (1996). Purification, kinetics and inhibition by antimonials of recombinant phosphofructokinase from *Schistosoma mansoni*. *Molecular and Biochemical Parasitology*, 81(2), 171–8. doi: 8898332.
- TDR. (2005). *Tropical disease research: progress 2005-2006*. Geneva: World Health

Organization.

- UniProt Consortium. (2008). The universal protein resource (UniProt). *Nucleic Acids Research*, 36(Database issue), D190–5. doi: gkm895.
- US Government Accountability Office. (2006). *New Drug Development: Science, Business, Regulatory, and Intellectual Property Issues Cited as Hampering Drug Development Efforts*. Recuperado de <http://www.gao.gov/new.items/d0749.pdf>.
- Utzinger, J., Xiao, S., N'Goran, E. K., Bergquist, R. & Tanner, M. (2001). The potential of artemether for the control of schistosomiasis. *International Journal for Parasitology*, 31(14), 1549-1562.
- Vandewaa, E. A., Mills, G., Chen, G. Z., Foster, L. A. & Bennett, J. L. (1989). Physiological role of HMG-CoA reductase in regulating egg production by *Schistosoma mansoni*. *The American Journal of Physiology*, 257(3 Pt 2), R618–25. doi: 2782464.
- Verjovski-Almeida, S., DeMarco, R., Martins, E. A. L., Guimarães, P. E. M., Ojopi, E. P. B., Paquola, A. C. M. et al. (2003). Transcriptome analysis of the acoelomate human parasite *Schistosoma mansoni*. *Nature Genetics*, 35(2), 148-57. doi: 12973350.
- Verjovski-Almeida, S., Venancio, T. M., Oliveira, K. C. P., Almeida, G. T. & DeMarco, R. (2007). Use of a 44k oligoarray to explore the transcriptome of *Schistosoma mansoni* adult worms. *Experimental Parasitology*, 117(3), 236–45. doi: S0014-4894(07)00098-7.
- WHO | Schistosomiasis. Recuperado Abril 1, 2009, de <http://www.who.int/mediacentre/factsheets/fs115/en/index.html>.
- Wilson, R. A. & Coulson, P. S. (2006). Schistosome vaccines: a critical appraisal. *Memórias Do Instituto Oswaldo Cruz*, 101 Suppl 1, 13-20.
- Wilson, R. A., Curwen, R. S., Braschi, S., Hall, S. L., Coulson, P. S. & Ashton, P. D. (2004). From genomes to vaccines via the proteome. *Memórias Do Instituto Oswaldo Cruz*, 99(5 Suppl 1), 45-50. doi: /S0074-02762004000900008.
- Yeh, I., Hanekamp, T., Tsoka, S., Karp, P. D. & Altman, R. B. (2004). Computational analysis of *Plasmodium falciparum* metabolism: organizing genomic information to facilitate drug discovery. *Genome research*, 14(5), 917–24.
- Zerlotini, A., Heiges, M., Wang, H., Moraes, R. L. V., Daminini, A. J., Ruiz, J. C. et al. (2009). SchistoDB: a *Schistosoma mansoni* genome resource. *Nucleic Acids Research*, 37(Database issue), D579-82. doi: 10.1093/nar/gkn681.

IX – ANEXOS

9.1: Relação de enzimas que catalizam reações *chokepoint*

Enzima	EC Number	Genes	Drogas
RXN66-221	4.1.1.28	1	4
Isoleucine--tRNA ligase*	6.1.1.5	2	2
Dihydroorotate oxidase	1.3.3.1	1	2
Dihydrofolate reductase	1.5.1.3	1	13
Hydroxymethylglutaryl-CoA reductase (NADPH)	1.1.1.34	1	12
Dopamine-beta-monoxygenase	1.14.17.1	1	1
Orotidine-5'-phosphate decarboxylase	4.1.1.23	1	1
RXN-7700	1.1.1.1	1	1
6-phosphogluconolactonase*	3.1.1.31	5	0
DTDPKIN-RXN*	2.7.4.6	5	0
[Acyl-carrier protein] S-malonyltransferase*	2.3.1.39	4	0
Acetyl-CoA carboxylase*	6.4.1.2	4	0
Cytochrome-c oxidase*	1.9.3.1	4	0
Phenylalanine decarboxylase*	4.1.1.53	4	0
carbamoyl-phosphate synthetase (glutamine-hydrolysing)*	6.3.5.5	3	0
Ethanolaminephosphotransferase*	2.7.8.1	3	0
Valine--tRNA ligase*	6.1.1.9	3	0
2OXOGLUTARATEDEH-RXN*		2	0
3-oxoacyl-[acyl-carrier protein] synthase*	2.3.1.41	2	0
Acetolactate synthase*	2.2.1.6	2	0
Cytidylate kinase*	2.7.4.14	2	0
folylpolyglutamate synthetase*	6.3.2.17	2	0
FORMYLTHFGLUSYNTH-RXN*	6.3.2.17	2	0
Glutamate--cysteine ligase*	6.3.2.2	2	0
Hydroxymethylglutaryl-CoA synthase*	2.3.3.10	2	0
Lactoylglutathione lyase*	4.4.1.5	2	0
phosphatidate cytidyltransferase*	2.7.7.41	2	0
Phosphoglycerate kinase*	2.7.2.3	2	0
PNKIN-RXN*	2.7.1.-	2	0
PNPOXI-RXN*	1.4.3.5	2	0
Porphobilinogen synthase*	4.2.1.24	2	0
RXN-8488*	1.14.18.-	2	0
RXN-8489*	1.10.3.-	2	0
RXN-8623*	1.10.3.-	2	0
RXN0-2161*	6.1.1.11	2	0
Thymidylate kinase*	2.7.4.9	2	0

Transketolase*	2.2.1.1	2	0
long-chain-fatty-acid--CoA ligase*	6.2.1.3	12	0
3-dehydroquinate dehydratase	4.2.1.10	1	0
3-dehydroquinate synthase	4.2.3.4	1	0
3-HYDROXY-KYNURENINASE-RXN	3.7.1.3	1	0
3-oxoacyl-[acyl-carrier protein] reductase	1.1.1.100	1	0
3-phosphoshikimate 1-carboxyvinyltransferase	2.5.1.19	1	0
5-aminolevulinic acid synthase	2.3.1.37	1	0
6-phosphofructokinase	2.7.1.11	1	0
Acetyl-CoA C-acetyltransferase	2.3.1.9	1	0
Acetyl-CoA C-acyltransferase	2.3.1.16	1	0
Aconitate hydratase	4.2.1.3	1	0
ACONITATEHYDR-RXN	4.2.1.3	1	0
Adenylosuccinate lyase	4.3.2.2	1	0
adenylosuccinate synthetase	6.3.4.4	1	0
Aminomethyltransferase	2.1.2.10	1	0
Arylformamidase	3.5.1.9	1	0
Aspartate carbamoyltransferase	2.1.3.2	1	0
Branched-chain amino acid aminotransferase	2.6.1.42	1	0
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	2.7.8.5	1	0
Chitobiosyldiphosphodolichol alpha-mannosyltransferase	2.4.1.142	1	0
Cholinephosphate cytidylyltransferase	2.7.7.15	1	0
Coproporphyrinogen oxidase	1.3.3.3	1	0
Cytidine deaminase	3.5.4.5	1	0
dCTP deaminase	3.5.4.13	1	0
Deoxyribose-phosphate aldolase	4.1.2.4	1	0
Dihydrolipoamide S-acetyltransferase	2.3.1.12	1	0
Dihydroorotase	3.5.2.3	1	0
Dimethylallyltransferase	2.5.1.1	1	0
Diphosphomevalonate decarboxylase	4.1.1.33	1	0
Dolichyl-phosphate beta-D-mannosyltransferase	2.4.1.83	1	0
dUTP pyrophosphatase	3.6.1.23	1	0
Ethanolamine-phosphate cytidylyltransferase	2.7.7.14	1	0
Ferrochelatae	4.99.1.1	1	0
Fumarate hydratase	4.2.1.2	1	0
Galactokinase	2.7.1.6	1	0
GDP-mannose 4,6-dehydratase	4.2.1.47	1	0
geranyltranstransferase	2.5.1.10	1	0
Glucosamine-1-phosphate N-acetyltransferase	2.3.1.157	1	0
Glucose-6-phosphate 1-dehydrogenase	1.1.1.49	1	0
glutathione synthetase	6.3.2.3	1	0
Glyceraldehyde 3-phosphate dehydrogenase (phosphorylating)	1.2.1.12	1	0
Glycine amidinotransferase	2.1.4.1	1	0
Glycine dehydrogenase (decarboxylating)	1.4.4.2	1	0
Glycogen (starch) synthase	2.4.1.11	1	0

GMP synthetase (glutamine-hydrolysing)	6.3.5.2	1	0
GTP cyclohydrolase I	3.5.4.16	1	0
Hydroxymethylbilane synthase	2.5.1.61	1	0
Isopentenyl-diphosphate delta-isomerase	5.3.3.2	1	0
Malate dehydrogenase	1.1.1.37	1	0
Mannose-1-phosphate guanylyltransferase (GDP)	2.7.7.22	1	0
Methylcrotonyl-CoA carboxylase	6.4.1.4	1	0
Mevalonate kinase	2.7.1.36	1	0
Orotate phosphoribosyltransferase	2.4.2.10	1	0
Phosphatidylserine decarboxylase	4.1.1.65	1	0
Phosphomannomutase	5.4.2.8	1	0
Phosphopyruvate hydratase	4.2.1.11	1	0
phosphoribosylformylglycinamidine synthetase	6.3.5.3	1	0
Pyrroline-5-carboxylate reductase	1.5.1.2	1	0
Pyruvate dehydrogenase (lipoamide)	1.2.4.1	1	0
ribulose phosphate 3-epimerase	5.1.3.1	1	0
RXN-5462	2.4.1.132	1	0
RXN-5463**		1	0
RXN-5464**		1	0
RXN-5465**		1	0
RXN-5466**		1	0
RXN-5470**		1	0
RXN-7908	4.2.1.96	1	0
RXN-7909	1.5.1.34	1	0
RXN-8629	1.8.1.4	1	0
RXN0-1132	1.8.1.4	1	0
RXN0-2921	6.3.2.17	1	0
RXN0-947**		1	0
RXN0-949**		1	0
Selenide,water dikinase	2.7.9.3	1	0
shikimate-kinase	2.7.1.71	1	0
Threonine dehydratase	4.3.1.19	1	0
Transaldolase	2.2.1.2	1	0
Ubiquinol--cytochrome c reductase	1.10.2.2	1	0
UDP-glucose 4-epimerase	5.1.3.2	1	0
UDP-N-acetylglucosamine pyrophosphorylase	2.7.7.23	1	0
UDP-N-acetylglucosamine--dolichyl-phosphate N-acetylglucosaminephosphotransferase	2.7.8.15	1	0
Uroporphyrinogen decarboxylase	4.1.1.37	1	0
uroporphyrinogen-III synthase	4.2.1.75	1	0

* A existência de mais de um gene que codifica para determinada enzima se deve ao fato de terem ocorrido duplicações gênicas ou erro de anotação/montagem.

** Existem algumas reações que não continham EC Number até a data de publicação da base de dados de vias metabólicas de referência, MetaCyc. Informação obtida no manual do software PathwayTools.

9.2: Tabela de frequência da co-ocorrência dos termos pesquisados na literatura

Enzimas	Drugs	Parasites	<i>Schistosoma mansoni</i>
hmg-CoA reductase	1181	13	7
L-DOPA decarboxylase	292	1	0
dihydrofolate reductase	636	411	1
Orotidine 5'-phosphate decarboxylase	10	9	2
6-phosphofructokinase (Phosphofructokinase) (Phosphohexokinase)	64	22	8
glutathione synthetase	340	64	4
phosphoglycerate kinase	48	27	3
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	72	17	6
adenylosuccinate lyase	4	3	1
ubiquinol-cytochrome C reductase	112	35	1
cytochrome C oxidase	487	341	15
carbamoyl-phosphate synthase	17	19	3
Orotidine 5'-phosphate decarboxylase	10	9	2
carbamoyl-phosphate synthase	17	19	3
Aspartate transcarbamylase	35	13	3
glutamate cysteine ligase	80	17	1
glucose-6-phosphate 1-dehydrogenase	0	0	0
glycogen synthase	255	15	5
dolichol-phosphate mannosyltransferase	1	4	1
Enolase (2-phosphoglycerate dehydratase) (2- phospho-D-glycerate	0	3	0
l-lactate dehydrogenase	1635	262	26
Isocitrate dehydrogenase [NAD] subunit gamma	1	0	0
kynureninase	7	1	1
Isoleucyl tRNA synthetase	3	0	0
dopamine beta hydroxylase	225	1	0
dihydroorotate dehydrogenase	38	23	0
aromatic amino acid decarboxylase	205	6	0
aspartate aminotransferase	1813	242	47
l-serine dehydratase	7	1	0
GTP cyclohydrolase I	20	6	0
5-aminolevulinic acid synthase	56	4	0
GDP mannose-4,6-dehydratase	0	0	0
aspartate ammonia lyase	6	1	0
L-phenylalanine decarboxylase	321	1	0
Cytidine aminohydrolase	93	9	0
phosphatidylglycerophosphate synthase	1	0	0
adenylosuccinate synthetase	6	2	0

selenophosphate synthase	0	0	0
choline-phosphate cytidyltransferase	29	1	0
glucosaminophosphotransferase	0	0	0
galactokinase	6	0	0
pyridoxine kinase	17	8	1
2-oxoglutarate dehydrogenase	62	24	1
pyrroline-5-carboxylate reductase	3	2	0
acetyl-CoA C-acyltransferase	8	1	0
transketolase	35	5	0
acetolactate synthase	4	0	0
glycine amidinotransferase	4	2	0
seryl-tRNA synthetase	3	0	0
3-oxoacyl-[acyl-carrier-protein] synthase	14	6	0
cytidylate kinase	4	0	0
acetyl-CoA carboxylase	56	7	0
malonyl CoA-acyl carrier protein transacylase	1	0	0
GMP-synthase	2	1	0
folylpolyglutamate synthase	46	3	0
acetyl-CoA carboxylase	56	7	0
lactoylglutathione lyase	27	8	0
folylpolyglutamate synthase	46	3	0
phosphomannomutase	0	3	0
long-chain-fatty-acid CoA ligase	27	3	0
gdp-mannose pyrophosphorylase	2	4	0
udp-n-acteylglucosamine pyrophosphorylase	0	0	0
deoxyribose-phosphate aldolase	1	0	0
mannose-1-phosphate guanyltransferase	0	0	0
farnesyl-pyrophosphate synthetase	26	13	0
udp-glucose 4-epimerase	3	3	0
ferrochelatae	41	6	0
acetolactate synthase	4	0	0
3-dehydroquinate dehydratase	3	1	0
3-dehydroquinate synthase	3	1	0
uroporphyrinogen decarboxylase	33	1	0
uroporphyrinogen-III synthase	3	1	0
dihydropteridine reductase	13	3	0
pterin-4-alpha-carbinolamine dehydratase	0	0	0
porphobilinogen deaminase	63	4	0
porphobilinogen synthase	71	8	0
aconitase	27	27	0
coproporphyrinogen III oxidase	2	0	0
phosphatidylserine decarboxylase	9	3	0

isopentenyl-diphosphate delta isomerase	7	2	0
diphosphomevalonate decarboxylase	2	0	0
deoxyuridine 5'-triphosphate nucleotidohydrolase	3	3	0
cytidine and deoxycytidylate deaminase	13	0	0
ribulose-5-phosphate-3-epimerase	0	1	0
kynurenine aminotransferase	12	2	0
6-phosphogluconolactonase	3	4	0
hormone-sensitive lipase	102	0	0
ethanolaminephosphotransferase	6	4	0
ethanolamine-phosphate cytidyltransferase	1	0	0
nucleoside diphosphate kinase	50	10	0
thymidylate kinase	27	3	0
3-dehydroquinate synthase	3	1	0
Shikimate kinase	7	2	0
pyridoxamine 5'-phosphate oxidase	0	0	0
mevalonate kinase	8	1	0
valyl-tRNA synthetase	1	0	0
farnesyl-pyrophosphate synthetase	26	13	0
alpha-1,3-mannosyltransferase	0	0	0
beta1,4 mannosyltransferase	0	0	0
hydroxymethylglutaryl-CoA synthase	98	1	0

9.3: Tabela contendo os artigos analisados

Enzima	Referência Bibliográfica
hmg-CoA reductase	<p>PMID: 2813388 - Molecular cloning and sequence analysis of 3-hydroxy-3-methylglutaryl-coenzyme A reductase from the human parasite <i>Schistosoma mansoni</i>.</p> <p>PMID: 1905241 - Purification and characterization of 3-hydroxymethylglutaryl-coenzyme A reductase of <i>Schistosoma mansoni</i>: regulation of parasite enzyme activity differs from mammalian host.</p> <p>PMID: 2123968 - Antischistosomal action of mevinolin: evidence that 3-hydroxy-methylglutaryl-coenzyme a reductase activity in <i>Schistosoma mansoni</i> is vital for parasite survival.</p> <p>PMID: 16512462 - Recombinant <i>Schistosoma mansoni</i> 3-hydroxymethylglutaryl-coenzyme-A reductase has different inhibitor kinetics compared to the mammalian host enzyme.</p>
L-DOPA decarboxylase	<p>PMID: 8501588 - Characteristics of phenol oxidase of <i>Schistosoma mansoni</i> and its functional implications in eggshell synthesis</p> <p>PMID: 7669109 - Electrophoretical and histochemical characterization of <i>Schistosoma mansoni</i> phenol oxidases</p> <p>PMID: 3100949 - Possible eggshell protein gene from <i>Schistosoma mansoni</i></p> <p>PMID: 6970143 - <i>Schistosoma mansoni</i>: decarboxylation of 5-hydroxytryptophan, L-dopa, and L-histidine in adult and larval schistosomes</p> <p>PMID: 417162 - Phenol oxidase activity: induction in female schistosomes by in vitro incubation</p> <p>PMID: 8759145 - Biogenic monoamines in the freshwater snail, <i>Biomphalaria glabrata</i>: influence of infection by the human blood fluke, <i>Schistosoma mansoni</i></p>
dihydrofolate reductase	<p>PMID: 4401788 - Comparative properties of schistosomal and filarial dihydrofolate reductases</p>
Orotidine 5'-phosphate decarboxylase	<p>PMID: 6090897 - Enzymes of uridine 5'-monophosphate biosynthesis in <i>Schistosoma mansoni</i></p> <p>PMID: 6111750 - The enzymes of pyrimidine biosynthesis in a range of parasitic protozoa and helminths</p>
6-phosphofructokinase (Phosphofructokinase) (Phosphohexokinase)	<p>PMID: 7984171 - Cloning and characterization of a cDNA encoding phosphofructokinase from <i>Schistosoma mansoni</i>.</p> <p>PMID: 10840980 - <i>Schistosoma mansoni</i> phosphofructokinase: immunolocalization in the tegument and immunogenicity.</p> <p>PMID: 8898332 - Purification, kinetics and inhibition by antimonials of recombinant phosphofructokinase from <i>Schistosoma mansoni</i>.</p> <p>PMID: 13446367 - The relationship between inhibition of phosphofructokinase activity and the mode of action of trivalent organic antimonials on <i>Schistosoma mansoni</i>.</p>
phosphoglycerate kinase	<p>PMID: 7477104 - Cloning of the gene for phosphoglycerate kinase from <i>Schistosoma mansoni</i> and characterization of its gene product</p> <p>PMID: 7591063 - Immune response to <i>Schistosoma mansoni</i> phosphoglycerate kinase during natural and experimental infection: identification of a schistosome-specific B-cell epitope.</p> <p>PMID: 14516923 - In vitro cellular and humoral responses to <i>Schistosoma mansoni</i> vaccine candidate antigens</p>
Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	<p>PMID: 1484556 - Complete sequence of the <i>Schistosoma mansoni</i> glyceraldehyde-3-phosphate dehydrogenase gene encoding a major surface antigen</p> <p>PMID: 10712729 - Production of Sm37-GAPDH, a major therapeutic target in human schistosomiasis.</p>
adenylosuccinate lyase	<p>PMID: 12392914 - Adenylosuccinate lyase of <i>Schistosoma mansoni</i>: gene structure, mRNA expression, and analysis of the predicted peptide structure of a potential chemotherapeutic target.</p>
ubiquinol-cytochrome C reductase	<p>PMID: 9297698 - Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in <i>Schistosoma mansoni</i></p>
cytochrome C oxidase	<p>PMID: 8522764 - Cytochemical localization of cytochrome c oxidase activity in mitochondria</p>

	in the tegument and tegumental and parenchymal cells of the trematodes <i>Echinostoma trivolvis</i> , <i>Zygotocotyle lunata</i> , <i>Schistosoma mansoni</i> , <i>Fasciola gigantica</i> and <i>Paragonimus ohirai</i>
carbamoyl-phosphate synthase	PMID: 45544 - Glutamine-dependent carbamoyl-phosphate synthetase and control of pyrimidine biosynthesis in the parasitic helminth <i>Schistosoma mansoni</i> . PMID: 6111750 - The enzymes of pyrimidine biosynthesis in a range of parasitic protozoa and helminths.
Orotidine 5'-phosphate decarboxylase	PMID: 2312224 - Pyrimidine salvage pathways in adult <i>Schistosoma mansoni</i> PMID: 6090897 - Enzymes of uridine 5'-monophosphate biosynthesis in <i>Schistosoma mansoni</i>
carbamoyl-phosphate synthase	PMID: 45544 - Glutamine-dependent carbamoyl-phosphate synthetase and control of pyrimidine biosynthesis in the parasitic helminth <i>Schistosoma mansoni</i> .
Aspartate transcarbamylase	PMID: 15469699 - Activity of some hepatic enzymes in schistosomiasis and concomitant alteration of arylsulfatase B. PMID: 6111750 - The enzymes of pyrimidine biosynthesis in a range of parasitic protozoa and helminths.
glutamate cysteine ligase	PMID: 14966568 - Vanin-1(-/-) mice show decreased NSAID- and Schistosoma-induced intestinal inflammation associated with higher glutathione stores.
glucose-6-phosphate 1-dehydrogenase	PMID: 2533355 - Hybrids between <i>Schistosoma mansoni</i> and <i>S. rodhaini</i> : characterization by isoelectric focusing of six enzymes PMID: 8595093 - Effect of schistosomal infection and its treatment on some key enzymes of glucose metabolism in mice livers.
glycogen synthase	PMID: 6821318 - Glycogen synthase of <i>Schistosoma mansoni</i> : a preliminary report PMID: 6789597 - Glycogen synthase of <i>Schistosoma mansoni</i> : a preliminary report PMID: 6804614 - Effect of <i>Schistosoma mansoni</i> on glycogen synthase and phosphorylase from <i>Biomphalaria glabrata</i> (mollusca) PMID: 2304487 - Substrate cycling between glucose 6-phosphate and glycogen occurs in <i>Schistosoma mansoni</i>
dolichol-phosphate mannosyltransferase	PMID: 11959028 - Dolichol phosphate mannose synthase is differentially expressed in male and female worms of <i>Schistosoma mansoni</i>
Enolase (2-phosphoglycerate dehydratase) (2-phospho-D-glycerate hydro-lyase)	PMID: 17606306 - Excretory-secretory proteome of larval <i>Schistosoma mansoni</i> and <i>Echinostoma caproni</i> , two parasites of <i>Biomphalaria glabrata</i> PMID: 9297698 - Gene linkage and steady state RNAs suggest trans-splicing may be associated with a polycistronic transcript in <i>Schistosoma mansoni</i> PMID: 7665603 - RNA trans-splicing in flatworms. Analysis of trans-spliced mRNAs and genes in the human parasite, <i>Schistosoma mansoni</i> .
L-lactate dehydrogenase	PMID: 738760 - Purification, immunochemical and biological characterization of malate dehydrogenase of <i>Schistosoma mansoni</i> PMID: 365558 - <i>Schistosoma mansoni</i> : antigenic characterization of malate dehydrogenase isoenzymes and use in the defined antigen substrate spheres (DASS) system PMID: 729691 - <i>Schistosoma mansoni</i> : purification and characterization of malate dehydrogenases PMID: 5542938 - Variations in malate dehydrogenase isoenzymes of <i>Schistosoma mansoni</i> PMID: 5137636 - Alteration of <i>Schistosoma mansoni</i> malate dehydrogenase isoenzymes on passage in the laboratory
Isocitrate dehydrogenase [NAD] subunit gamma	PMID: 4199254 - Enzyme levels in cercariae and adult <i>Schistosoma mansoni</i> PMID: 3150324 - Effect of immature <i>Schistosoma mansoni</i> worms on hamsters' plasma enzymes.
kynureninase	PMID: 5683509 - Studies with tryptophan metabolites in vitro. Kynurenine metabolism in kidneys of mice infested with <i>Schistosoma mansoni</i>
carbamoyl-phosphate synthase	PMID: 6111750 - The enzymes of pyrimidine biosynthesis in a range of parasitic protozoa and helminths.

9.4: Artigo – SchistoDB: a *Schistosoma mansoni* genome resource

9.5: Artigo – The genome of the blood fluke *Schistosoma mansoni*.