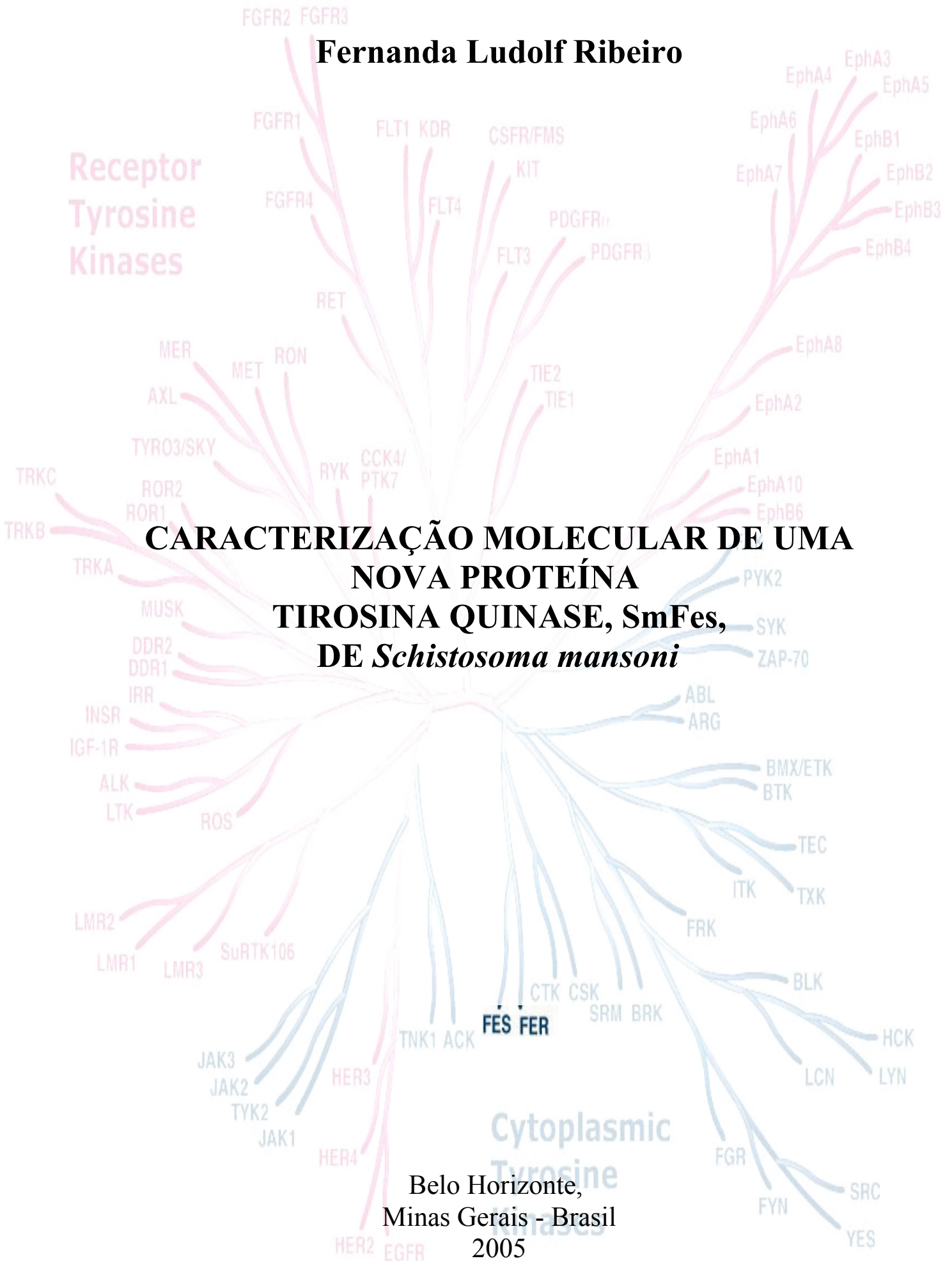


Fernanda Ludolf Ribeiro

Receptor Tyrosine Kinases

**CARACTERIZAÇÃO MOLECULAR DE UMA NOVA PROTEÍNA TIROSINA QUINASE, SmFes, DE *Schistosoma mansoni***



Belo Horizonte,  
Minas Gerais - Brasil  
2005

Ministério da Saúde  
FUNDAÇÃO OSWALDO CRUZ  
**Centro de Pesquisa René Rachou**  
Programa de Pós-graduação em Ciências da Saúde

**CARACTERIZAÇÃO MOLECULAR DE UMA  
NOVA PROTEÍNA  
TIROSINA QUINASE, SmFes,  
DE *Schistosoma mansoni***

por

**Fernanda Ludolf Ribeiro**

Dissertação apresentada a Pós-graduação em Ciências da Saúde do CPqRR  
como requisito parcial para obtenção do grau de Mestre em Ciências da Saúde.  
Área de Concentração: Biologia Celular e Molecular.

Orientador: Dr. Guilherme Corrêa Oliveira  
Co-orientadora: Dra. Diana Bahia

CPqRR  
Belo Horizonte  
2005

Ministério da Saúde  
FUNDAÇÃO OSWALDO CRUZ  
**Centro de Pesquisa René Rachou**  
Programa de Pós-graduação em Ciências da Saúde

**CARACTERIZAÇÃO MOLECULAR DE UMA  
NOVA PROTEÍNA  
TIROSINA QUINASE, SmFes,  
DE *Schistosoma mansoni***

por

**Fernanda Ludolf Ribeiro**

Este trabalho foi realizado no Laboratório de Parasitologia Celular e Molecular do Centro de Pesquisas René Rachou/FIOCRUZ sob a orientação do Dr. Guilherme Corrêa Oliveira<sup>1</sup> e co-orientação da Dra. Diana Bahia<sup>1</sup>. O projeto contou com o suporte financeiro da FIOCRUZ, FAPEMIG, Fundep e do convênio FIOCRUZ/INSERM. Parte deste projeto foi realizado em colaboração com os Drs. Raymond Pierce<sup>2</sup> e Dra. Colette Dissous<sup>2</sup>.

1 – Laboratório de Parasitologia Celular e Molecular, Centro de Pesquisas René Rachou – FIOCRUZ, Belo Horizonte, MG, Brasil.

2 - Instituto Pasteur de Lille, Unidade INSERM U 547 – Lille, França

Ministério da Saúde  
**Centro de Pesquisa René Rachou**  
Programa de Pós-graduação em Ciências da Saúde

**CARACTERIZAÇÃO MOLECULAR DE UMA  
NOVA PROTEÍNA  
TIROSINA QUINASE, SmFes,  
DE *Schistosoma mansoni***

por

**Fernanda Ludolf Ribeiro**

foi avaliada pela banca examinadora composta pelos seguintes membros:

Dr. Marcelo Rosado Fantappié  
Dra. Cristiana Ferreira Alves de Brito

Dissertação defendida e aprovada em 11 de julho de 2005

*A mente que se abre a uma idéia  
jamais voltará ao  
seu tamanho original.  
(Albert Einstein)*

*Dedico esse trabalho  
àqueles que me inspiraram:  
força, coragem e determinação.*

## AGRADECIMENTOS

Em primeiro lugar agradeço ao Dr. Guilherme Corrêa de Oliveira pela oportunidade, confiança, orientação e aprendizado ao longo de anos de iniciação científica e mestrado, e também pelo exemplo de sabedoria e conhecimento;

À Dra. Diana Bahia pela amizade, incentivo, dedicação, vibração e conhecimentos, assim como também pela ajuda na elaboração e elucidação de experimentos;

Ao Dr. Álvaro Romanha pelas palavras sábias e experientes, e por disponibilizar aos estudantes um laboratório com excelente infra-estrutura;

Ao grupo francês do Instituto Pasteur de Lille, Dra. Colette Dissous, Dr. Raymond Pierce e Dr. Naji Khayath pela parceria nesse projeto;

Ao CPqRR e ao Dr. Roberto Sena Rocha, atual diretor, pela oportunidade de crescimento científico;

Ao grupo do Dr. Guilherme: Regina, Flávio, Talita, Adhemar, Anderson, Rômulo, Elisângela, Kleider e Solange, pela amizade e trabalho em equipe;

Às garotas Quinase, Luíza e Livia, pela companhia, amizade e disponibilidade em ajudar;

À Maricota por ter tornado uma grande amiga, imprescindível nas horas de tristeza e alegria, aflição e alívio e por todos os momentos de conversa;

À Dra. Silvane Murta pela sua atenção e disponibilidade em ajudar;

Ao todos os demais do LPCM: Maureen, Héliida, Rosana, Fernanda B., Fernanda F., Nilton, Bernardo, Daniel, Juciane, Marcilene, Carol, Maíra, Sara, Paula, Thaís, Núbia, Andréa, Kênya e Silvia pela convivência e pela ajuda na realização e compreensão dos experimentos;

Ao Laboratório de Malacologia pelo fornecimento dos parasitos;

Ao Laboratório de Imunologia pela utilização de sua estrutura;

Ao Laboratório de Entomologia Médica pela receptividade;

À informática pela disponibilidade em ajudar;

À minha família pelo apoio, incentivo e paciência. Em especial minha mãe pela amizade e exemplo de garra. À minha avó e mãe Di pelos cuidados. Ao meu pai, Vuru, Hick e Sofia pela torcida, mesmo à distância;

Ao Neto pelo amor, companheirismo, incentivo, atenção e paciência. Pelo exemplo de força, coragem e determinação;

Ao Janjão pelos primeiros passos na ciência;

À Eliane por me acolher com todo carinho;

À Amanda pela companhia e amor, e à Pit pela alegria;

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

# ÍNDICE

<b>ÍNDICE DE QUADROS.....</b>	<b>XI</b>
<b>ÍNDICE DE FIGURAS.....</b>	<b>XII</b>
<b>LISTA DE ABREVIATURAS.....</b>	<b>XIII</b>
<b>RESUMO.....</b>	<b>XVI</b>
<b>ABSTRACT.....</b>	<b>XVII</b>
<b>I INTRODUÇÃO .....</b>	<b>1</b>
<b>I.1 ESQUISTOSSOMOSE.....</b>	<b>2</b>
<i>I.1.1 Aspectos Gerais e Epidemiológicos.....</i>	<i>2</i>
<i>I.1.2 Genômica do S. mansoni.....</i>	<i>7</i>
<b>I.2 RELAÇÃO PARASITO E MEIO AMBIENTE.....</b>	<b>9</b>
<b>I.3 MECANISMOS DE TRANSDUÇÃO DE SINAL .....</b>	<b>10</b>
<b>I.4 AS PROTEÍNAS QUINASE.....</b>	<b>14</b>
<i>I.4.1 Função e Atividade.....</i>	<i>14</i>
<i>I.4.2 Presença de Proteínas Quinase em Genomas Seqüenciados.....</i>	<i>17</i>
<i>I.4.3 Classificação de Proteínas Quinase.....</i>	<i>17</i>
<i>I.4.4 As Proteínas Tirocinas Quinase (PTKs).....</i>	<i>18</i>
<i>I.4.5 Importância das Proteínas Quinase para a Integridade da Vida.....</i>	<i>21</i>
<b>I.5 PROTEÍNAS DE S. MANSONI ENVOLVIDAS NA TRANSDUÇÃO DE SINAL ...</b>	<b>21</b>
<i>I.5.1 Proteínas Tirocinas Quinase Identificadas em S.mansoni.....</i>	<i>23</i>
<b>II JUSTIFICATIVA .....</b>	<b>25</b>
<b>III OBJETIVOS .....</b>	<b>27</b>
<b>III.1 OBJETIVO GERAL.....</b>	<b>28</b>
<b>III.2 OBJETIVOS ESPECÍFICOS.....</b>	<b>28</b>
<b>IV MATERIAIS E MÉTODOS .....</b>	<b>29</b>
<b>IV.1 COLETA DE PARASITOS.....</b>	<b>30</b>
<i>IV.1.1 Vermes Adultos.....</i>	<i>30</i>
<i>IV.1.2 Ovos.....</i>	<i>30</i>
<i>IV.1.3 Miracídeos.....</i>	<i>30</i>
<i>IV.1.4 Esporocistos.....</i>	<i>30</i>
<i>IV.1.5 Cercárias.....</i>	<i>31</i>
<i>IV.1.6 Esquistossômulos.....</i>	<i>31</i>
<b>IV.2 PREPARO DE CÉLULAS COMPETENTES TRATADAS COM CLORETO DE CÁLCIO.....</b>	<b>32</b>
<b>IV.3 TRANSFORMAÇÃO QUÍMICA .....</b>	<b>32</b>
<b>IV.4 EXTRAÇÃO DE PLASMÍDEO.....</b>	<b>32</b>
<b>IV.5 EXTRAÇÃO DE DNA GENÔMICO.....</b>	<b>33</b>
<i>IV.5.1 Extração DNA Genômico Kit Wizard.....</i>	<i>33</i>
<i>IV.5.2 Extração DNA genômico.....</i>	<i>33</i>
<b>IV.6 PREPARAÇÃO DE EXTRATO DE PROTEÍNAS.....</b>	<b>34</b>
<b>IV.7 EXTRAÇÃO DE RNA.....</b>	<b>34</b>
<i>IV.7.1 Rneasy.....</i>	<i>34</i>



IV.7.2	<i>Tri-Reagente</i> .....	35
IV.7.3	<i>Ultracentrifugação em Cloreto de Césio</i> .....	35
IV.8	SÍNTESE DE CDNA.....	35
IV.8.1	<i>SuperScript</i> .....	35
IV.8.2	<i>ThermoScript</i> .....	36
IV.9	MÉTODOS DE ELETROFORESE.....	36
IV.9.1	<i>Eletroforese de DNA em Gel de Poliacrilamida</i> .....	36
IV.9.2	<i>Eletroforese de DNA em Gel de Agarose</i> .....	37
IV.9.3	<i>Eletroforese de Proteínas em Gel de Poliacrilamida-SDS</i> .....	37
IV.9.4	<i>Eletroforese de Proteína em Gel de Poliacrilamida Bis-Tris Nu-Page</i> .....	37
IV.9.5	<i>Eletroforese de RNA em Gel de Agarose Desnaturante</i> .....	38
IV.10	EXTRAÇÃO DE DNA DE GEL DE AGAROSE.....	38
IV.11	SEQUENCIAMENTO DE DNA.....	38
IV.12	AMPLIFICAÇÃO DE DNA POR PCR.....	39
IV.13	DIGESTÃO DE DNA POR ENZIMA DE RESTRIÇÃO.....	39
IV.14	HIBRIDIZAÇÃO DE ÁCIDOS NUCLÉICOS.....	39
IV.14.1	<i>Preparo da sonda marcada com Digoxigenina</i> .....	39
IV.14.2	<i>Preparo da sonda marcada por Fósforo Radioativo</i> .....	39
IV.14.3	<i>Preparo do DNA para Southern-blot</i> .....	44
IV.14.4	<i>Preparo do RNA para Northern-blot</i> .....	44
IV.14.5	<i>Transferência do Ácido Nucléico</i> .....	44
IV.14.6	<i>Hibridização com sonda não radioativa</i> .....	45
IV.14.7	<i>Hibridização por sonda radioativa</i> .....	45
IV.15	WESTERN BLOT.....	46
IV.16	PCR QUANTITATIVO.....	47
IV.17	EXPRESSÃO DE PROTEÍNA RECOMBINANTE.....	47
IV.17.1	<i>Amplificação do Domínio SH2 e PK</i> .....	47
IV.17.2	<i>Clonagem em Vetor pCR 2.1</i> .....	47
IV.17.3	<i>Clonagem em vetor pGEX</i> .....	48
IV.17.4	<i>Expressão da Proteína Recombinante de SmFes</i> .....	48
IV.18	ANÁLISE IN SILICO DA SEQUÊNCIA CODIFICANTE COMPLETA DE SMFES.....	49
IV.18.1	<i>Determinação de Domínios</i> .....	49
IV.18.2	<i>Determinação da Organização Gênica</i> .....	49
IV.18.3	<i>Análise Filogenética</i> .....	49
IV.18.4	<i>Identificação de Parceiros de SmFes</i> .....	50
<b>V</b>	<b>RESULTADOS</b> .....	<b>51</b>
V.1	ANÁLISE E CARACTERIZAÇÃO DA SEQUÊNCIA IN SILICO.....	52
V.1.1	<i>Seqüências de SmFes</i> .....	52
V.1.2	<i>Identidade de SmFes com Proteínas Fes/Fps/Fer Ortólogas</i> .....	52
V.1.3	<i>Domínios Protéicos em SmFes</i> .....	52
V.2	ANÁLISE FILOGENÉTICA DE SMFES.....	58
V.3	ORGANIZAÇÃO GENÔMICA DE SMFES.....	62
V.4	SOUTHERN-BLOT.....	68
V.5	PADRÃO DE EXPRESSÃO DA PROTEÍNA SMFES.....	73
V.6	PADRÃO DE TRANSCRIÇÃO DO RNA DE SMFES.....	73
V.6.1	<i>PCR em Tempo Real</i> .....	73
V.6.2	<i>Northern-blot</i> .....	73
V.7	POLIMORFISMO EM SMFES.....	77
V.7.1	<i>Inserção de 9 pb no cDNA de SmFes</i> .....	77
V.7.2	<i>Inserção de 15 pb no DNA genômico contendo o gene SmFes</i> .....	77

<i>V.7.3 Polimorfismos do tipo SNPs no gene SmFes.....</i>	<i>78</i>
<b><u>V.8 IDENTIFICAÇÃO DE ORTÓLOGOS DE SMFES EM DIVERSAS ESPÉCIES DO GÊNERO SCHISTOSOMA.....</u></b>	<b><u>78</u></b>
<b><u>V.9 VIA DE SINALIZAÇÃO.....</u></b>	<b><u>78</u></b>
<i>V.9.1 Proteínas de sinalização que participam da via de sinalização de SmFes.....</i>	<i>78</i>
<i>V.9.2 Clonagem e Expressão de Proteína para Ensaio de Pull-Down.....</i>	<i>84</i>
<b><u>VI DISCUSSÃO .....</u></b>	<b><u>91</u></b>
<i>VI.1 CLASSIFICAÇÃO ESTRUTURAL.....</i>	<i>92</i>
<i>VI.2 ESTRUTURA GENÔMICA.....</i>	<i>95</i>
<i>VI.3 POLIMORFISMOS.....</i>	<i>96</i>
<i>VI.4 NÍVEL DE EXPRESSÃO E FUNÇÃO DEDUZIDA.....</i>	<i>97</i>
<i>VI.5 VIA DE SINALIZAÇÃO.....</i>	<i>98</i>
<i>VI.6 SMFES COMO POSSÍVEL ALVO DE DROGAS.....</i>	<i>101</i>
<b><u>VII CONCLUSÕES.....</u></b>	<b><u>102</u></b>
<b><u>VIII REFERÊNCIA BIBLIOGRÁFICA .....</u></b>	<b><u>104</u></b>
<b><u>IX ANEXOS .....</u></b>	<b><u>106</u></b>
<i>IX.1 ANEXO I.....</i>	<i>107</i>
<i>IX.2 ANEXO II.....</i>	<i>117</i>
<i>IX.3 ANEXO III.....</i>	<i>132</i>

## ÍNDICE DE QUADROS

Quadro 1: Classificação das proteínas quinases, segundo Hanks e Quinn (1991)._____	19
Quadro 2: Protocolo de Reação de PCR._____	41
Quadro 3: Programa de termociclagem da PCR._____	42
Quadro 4: Lista de iniciadores._____	43
Quadro 5: Digestão de DNA por enzima de restrição._____	44
Quadro 6: BLAST-p da proteína SmFes contra banco de dados nr._____	58
Quadro 7: Quadro comparativo do tamanho encontrado para os exons e introns na montagem do Instituto Sanger e na nossa montagem._____	65
Quadro 8: Possíveis SNPs encontrados nas seqüências genômicas quando comparadas ao cDNA de SmFes depositado no banco de dados._____	81
Quadro 9: Relação de proteínas descritas para a via de sinalização de Fes/Fps/Fer encontradas nos bancos de dados de <i>Schistosoma mansoni</i> ._____	86

## ÍNDICE DE FIGURAS

Figura 1: Ciclo biológico do <i>S. mansoni</i> .	04
Figura 2: Via de sinalização simples.	11
Figura 3: Famílias de receptores de superfície celular.	13
Figura 4: Fosforilação reversível de proteínas sinalizadas.	15
Figura 5: Cascata de sinalização.	16
Figura 6: Seqüência nucleotídica de SmFes.	54
Figura 7: Seqüência de aminoácidos da proteína SmFes.	56
Figura 8: Gráfico de hidropaticidade da seqüência protéica de SmFes.	57
Figura 9: Esquema representativo dos domínios de SmFes e sua localização na seqüência de aminoácido.	60
Figura 10: Alinhamento múltiplo do domínio catalítico de SmFes e de proteínas ortólogas de outros organismos.	61
Figura 11: Árvore filogenética sem raiz de SmFes construída com outras proteínas membros da subfamília Fps/Fes/Fer.	62
Figura 12: Organização genômica de SmFes e localização dos iniciadores na seqüência	64
Figura 13: Análise do DNA genômico para comprovação dos introns de SmFes.	66
Figura 14: <i>Southern-blot</i> não radioativo do gene SmFes.	69
Figura 15: Esquema representativo de <i>Southern-blot</i> não radioativo.	70
Figura 16: <i>Southern-blot</i> radioativo do gene SmFes e esquema representativo.	71
Figura 17: <i>Western-blot</i> de SmFes.	73
Figura 18: Perfil de transcrição de SmFes no diferentes estágios do parasito por PCR em tempo real.	74
Figura 19: <i>Northern-blot</i> radioativo do gene SmFes.	75
Figura 20: Polimorfismo de 9 pb encontrado em SmFes	78
Figura 21: Digestão parcial da amplificação de cDNAs da região contendo os 9pb.	79
Figura 22: Inserção de 15 pb encontrados em clones de seqüências genômicas e ausente em seqüência de cDNA e plasmidial.	80
Figura 23: Amplificação do DNA genômico de diversas espécies do gênero <i>Schistosoma</i> com pares de iniciadores desenhados especificamente para SmFes.	82
Figure 24: Via de sinalização proposta por Greer (2002) para a proteína Fer.	84
Figura 25: Via de sinalização proposta por Greer (2002) para a proteína Fes/Fps.	85
Figura 26: Expressão da proteína recombinante de SmFes clonada em pGEX-4T3.	88

## **Lista de abreviaturas**

aa - aminoácido.

AMPc - adenosina monofosfato cíclica.

APS - persulfato de amônio.

ATP - adenosina trifosfato.

BLAST - *Basic local alignment search tool*, ferramenta de busca de alinhamento local básico.

cDNA - DNA complementar.

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico.

CO<sub>2</sub> - dióxido de carbono.

CPqRR - Centro de Pesquisa René Rachou.

DALYs - *Disability Adjusted Life Years*, anos de vida ajustados por incapacidade.

DEPC - dietilpirocarbonato.

DNA - ácido desoxirribonucléico.

DNase - desoxirribonuclease.

dNTP - desoxirribonucleosídeo trifosfato.

DTT - dithiotheitol.

E - exon.

EDTA - etileno diamino tetra-acetato.

EST - *expressed sequence tags*, etiquetas de seqüência expressa.

FAPEMG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais.

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo.

FCH - motivo de homologia a *Fps/Fes/Fer/CIP4*.

Fer - proteína relacionada a *Fes*.

Fes - sarcoma de felinos.

FIOCRUZ - Fundação Oswaldo Cruz.

Fps - *sarcoma* aviário.

GDP - guanosina difosfato.

GO - *gene ontology*, ontologia gênica.

GTP - guanosina trifosfato.

HF- *high fidelity*, alta fidelidade.

I - intron.

Il - interleucina.

IPTG - Isopropil-β.D. galactosídeo.

Kb - kilobase, 10<sup>3</sup> bases.

kDa - kilodalton.

LB - Meio Luria Bertani.

LE - cepa de *S. mansoni* Luiz Evangelista.

Mb - megabase,  $10^6$  bases.

MCT - Ministério da Ciência e Tecnologia.

MOPS – Ácido 3-[n-morfolino] profano sulfônico.

mRNA - RNA mensageiro.

NaAc - ácido sódico.

NCBI - *National Center of Biotechnology Information*, Centro Nacional de Informação Biotecnológica.

NRTK - tirosina quinase não receptora.

OD- densidade ótica.

ONSA - Organização para Análise e Seqüenciamento de Nucleotídeos.

ORESTES - “ORF ESTs”.

ORF - *open read frame*, janela de leitura aberta.

pb - pares de base.

PCR - *polimerase chain reaction*, reação em cadeia da polimerase.

PK - proteína quinase.

PM - peso molecular.

PTK - proteína tirosina quinase.

q.s.p - quantidade suficiente para.

r.p.m - rotação por minuto.

RACE-PCR - *5’-3’ rapid amplification of cDNA ends*, amplificação rápida dos finais 3’ e 5’ do cDNA.

RAP-PCR - *arbitrarily primed polymerase chain reaction*, reação em cadeia da polimerase com iniciação arbitrária.

RFLP - *restriction fragment length polymorphism*, polimorfismo de tamanho de fragmentos de restrição.

RNA - ácido ribonucleico.

RNAse - ribonuclease.

RPS-BLAST - BLAST de posição reversa específica.

RTK - receptor tirosina quinase.

RT-PCR - PCR de um produto de transcrição reversa.

SAGE - *serial analysis of gene expression*, análise seriada de expressão gênica.

SDS - sódio dodecil sulfato.

SH - homologia a *Src*.

SmAE - *Shistosoma mansoni assembled ESTs*, ESTs agrupadas de *S. mansoni*.

SNP - *single nucleotide polymorphism*, polimorfismo de base única.

T.A - temperatura ambiente.

TBE - tris borato EDTA.

TEMED - N, N, N', N'-tetra metiletilenodiamina.

TGF $\beta$  - fator de crescimento tumoral beta.

TIGR - *The Institute for Genomic Research*, Instituto para Pesquisa Genômica.

TNF $\alpha$  - fator de necrose tumoral alfa.

Tris - trishidroximetilaminometano.

U - unidade.

UV - ultravioleta.

## **RESUMO**

A identificação de moléculas sinalizadoras e a elucidação de processos de transdução de sinal de *Schistosoma mansoni* são necessárias para o entendimento da interação hospedeiro-parasito e de processos fisiológicos do parasito. Proteínas tirosinas quinase (PTKs) são moléculas importantes na comunicação intra e intercelular, tendo um papel crucial nos mecanismos de transdução de sinal. Estas proteínas estão envolvidas nos processos de desenvolvimento, diferenciação e comunicação celular entre outros. Uma nova PTKs foi identificada em *S. mansoni* e nomeada, SmFes. SmFes é um gene de cópia única e o seu cDNA completo contém uma ORF de 3.780 pb. SmFes exibe características da família Fes/Fps/Fer das PTKs, com assinaturas de domínio proteína quinase, SH2 e *coiled-coil* característicos. SmFes é o primeiro gene da família Fes/Fps/Fer identificado em *S. mansoni*. Análise filogenética revelou que SmFes está relacionada a proteínas da família Fes/Fps/Fer, agrupadas mais próximas às proteínas ortólogas de organismos invertebrados. O alinhamento entre o cDNA de SmFes e seqüências genômicas depositadas pelos bancos de dados de *S. mansoni* indicaram a presença de 17 introns, alguns demonstrados experimentalmente. Análises parciais de clones de cDNA indicaram a inserção de 9 pb na posição 3' do exon 10, formando duas populações de cDNA, comprovadas por RFLP. A análise da seqüência genômica indicou que existem dois possíveis sítios de edição do RNA na proximidade 5' do intron, indicando um evento de edição alternativa. A existência de um alelo contendo uma inserção de 15 pb na seqüência genômica foi observada. As amplificações de diferentes espécies de *Schistosoma* com pares de iniciadores desenhados para SmFes indicaram a presença de ortólogos de SmFes em várias espécies do gênero. SmFes é transcrita em baixos níveis nas diversas fases de desenvolvimento do parasito, com uma maior taxa de transcrição em vermes machos, como detectado por PCR em tempo real e *northern-blot*. A expressão da proteína SmFes foi verificada em esporocisto, miracídio, verme adulto e em nível mais elevado em cercária. Uma proteína recombinante de SmFes foi expressa para ser usada em experimentos de identificação de parceiros de SmFes em futuros ensaios de co-precipitação e também em experimentos de duplo híbrido. Vários possíveis parceiros de SmFes foram identificados por análise computacional. A compreensão de genes envolvidos nos mecanismos de transdução de sinal pode fornecer novas estratégias de desenvolvimento de drogas.



## **ABSTRACT**

The identification and elucidation of signal transduction molecules and mechanisms are necessary for the understanding of the *Schistosoma mansoni* host-parasite interaction and of physiological process of the parasite. Protein Tyrosine kinases (PTKs) are important molecules for intra and intercellular communication, playing a crucial role in signal transduction processes. These proteins are known to be involved in developmental, differentiation and communication processes of cells, among others. A novel member of PTK was identified in *S. mansoni* and designated SmFes. SmFes is a single copy gene and the complete cDNA contains a 3.780 bp ORF. SmFes exhibits the characteristic features of Fes/Fps/Fer protein tyrosine kinases family, containing three coiled-coil regions, an SH2 and a protein tyrosine kinase catalytic domain signatures. SmFes is the first gene from Fes/Fps/Fer family identified in *S. mansoni*. Phylogenetic analyses revealed that SmFes is more closely related to invertebrate proteins orthologues of the Fes/Fps/Fer family. The assembly of the SmFes cDNA and genomic sequences indicated the presence of 17 introns, some experimentally demonstrated. Analysis of partial cDNA clones indicated the presence of a 9 pb insertion at the 3' end of exon 10, producing two different populations of cDNA. The analysis of the genomic sequence indicated that there are two possible splice sites at the 5' end of the intron, pointing to an alternative splicing event in the SmFes pre-mRNA. An allele of the SmFes containing a 15 pb insertion was observed in genomic sequence. The amplification of different *Schistosoma* species with SmFes specific primers indicated that orthologues of SmFes are present in several species of the genus. SmFes is transcribed at low levels in the different developmental stages of the parasite, with higher level in adult male worms as detected by real time PCR and *northern-blot*. The expression of SmFes protein was verified in sporocyst, miracidium, adult worm and in higher levels in cercariae. A recombinant protein was expressed to be used in experiments to identify SmFes partner proteins in future pull-down assays and also in two-hybrid systems. Various possible partners of SmFes were identified by computational analysis. The comprehension of signal transduction processes could also contribute to the identification of possible new targets for drugs.



# **I INTRODUÇÃO**

## I.1 ESQUISTOSSOMOSE

### I.1.1 Aspectos Gerais e Epidemiológicos

O termo parasitismo refere-se à relação entre dois organismos, o parasito e o hospedeiro, que é benéfica apenas ao parasito. O hospedeiro fornece um habitat, nutrientes e estímulos que regulam o desenvolvimento do parasito. A esquistossomose é uma doença parasitária que tem como agente infectante um parasito pertencente ao filo Platelmino, classe Trematoda, ordem Digenea, subordem Strigeidida, superfamília Schistosomatoidea, família Schistosomatidae, e gênero *Schistosoma*. *Schistosoma* é um metazoário acelomado, de simetria bilateral e dimorfismo sexual durante a fase adulta .

O agente da esquistossomose foi descoberto em 1851. Em 1913 o hospedeiro intermediário foi identificado e o ciclo de vida do parasito descrito e reproduzido em laboratório . Várias espécies de *Schistosoma* já foram identificadas: *Schistosoma bovis*, *S. curassoni*, *S. edwardiense*, *S. haematobium*, *S. hippopotami*, *S. incognitum*, *S. indicum*, *S. intercalatum*, *S. japonicum*, *S. leiperi*, *S. malayensis*, *S. mansoni*, *S. margrebowiei*, *S. mattheei*, *S. mekongi*, *S. nasale*, *S. ovuncatum*, *S. rodhaini*, *S. sinensium* e *S. spindale*. As três espécies mais importantes para a saúde humana, *S. mansoni*, *S. haematobium* e *S. japonicum*, têm distribuição geográfica diferentes, ou localizações topográficas distintas no organismo do hospedeiro definitivo, além de características morfológicas e fisiológicas peculiares . A espécie *S. mansoni* ocorre na África, América Latina e Antilhas. A infecção causada por esta espécie é denominada esquistossomose mansônica ou intestinal, pela localização dos parasitos nas vênulas da parede do intestino grosso, sigmóide e reto, com sintomas predominantemente intestinais. Nos casos mais graves, há envolvimento hepatoesplênico e hipertensão no sistema porta . A esquistossomose mansônica foi introduzida no Brasil pelo tráfico de escravos, que foi confirmada por estudos de diferenciação gênica e isoenzimáticos entre parasitos do velho e novo mundo . A espécie *S. haematobium* ocorre predominantemente na África, sua presença estende-se para a Bacia do Mediterrâneo, Oriente Próximo e Médio. A doença causada por esta espécie é conhecida como esquistossomose hematóbica, vesical ou urinária e também por genito-urinária. Os parasitos localizam-se preferencialmente no plexo vesical e a doença produz um quadro clínico com sintomas urinários . A espécie *S. japonicum* é responsável por outra morbidade intestinal da doença, porém circunscrita ao extremo oriente e Pacífico ocidental, e é conhecida por esquistossomose japônica .

O ciclo de vida do *S. mansoni* envolve duas gerações: a primeira no hospedeiro definitivo vertebrado em que ocorre a reprodução sexuada e maturação; e a segunda no hospedeiro

intermediário invertebrado em que somente ocorre a reprodução assexuada. O hospedeiro intermediário para *S. mansoni* é o caramujo do gênero *Biomphalaria* (Figura 1). Milhares de cercárias são produzidas por um único esporocisto durante a reprodução assexuada no hospedeiro intermediário, que as liberam de forma intermitente. A saída das cercárias do hospedeiro intermediário é induzida pela luz, portanto a liberação ocorre preferencialmente nas horas mais claras do dia. Quando madura, a cercária sai do caramujo para a água, estando apta a infectar hospedeiros vertebrados. A cercária mede cerca de 0,5 mm em comprimento e tem como característica uma cauda bifurcada. Na água, acusam geotropismo negativo, fototropismo positivo e tendem a acumular-se sob a superfície líquida. Uma vez que não se alimentam, possuem uma grande reserva de glicogênio e precisam encontrar um hospedeiro vertebrado num período de 36-48 horas. Quando encontram um hospedeiro apropriado, penetram pela sua pele ou mucosa liberando a cauda. A capacidade invasora das larvas depende de um esforço mecânico e da ação química exercida pela secreção das glândulas cefálicas de penetração. A ação mecânica ocorre por auxílio de uma ventosa oral. A ação química ocorre por proteases, do tipo collagenases e elastases, ou por enzimas que são ativas contra as glicoproteínas da pele. Após penetrar, a cercária começa a passar por mudanças, na conformação da membrana que se torna pentalamínada e no metabolismo que se torna principalmente anaeróbico. Inicia-se o processo de transformação em esquistossômulo. Após permanecer na pele por um tempo variável, de minutos a 24 horas, os esquistossômulos iniciam a migração através do corpo de seu novo hospedeiro, caso não sejam destruídos pelos mecanismos de defesa do hospedeiro. Os esquistossômulos utilizam as secreções líticas das glândulas cefálicas posteriores para penetrar nos vasos cutâneos. Através da circulação, eles chegam ao coração direito e aos pulmões em 5 dias, tornando-se mais longos e delgados, o que facilita a sua migração através da rede vascular pulmonar (Miller & Wilson 1978, Rollinson & Simpson 1987). Do pulmão, os esquistossômulos voltam ao coração esquerdo e são enviados pela circulação geral a todas as partes do corpo do hospedeiro. Somente quando alcançam o sistema porta intra-hepático podem completar seu desenvolvimento (Miller & Wilson 1980, Rollinson & Simpson 1987). Quatro semanas após a infecção, a maioria dos vermes encontram-se maduros e prontos para se acasalarem. Os vermes acasalados deslocam-se ativamente contra a corrente circulatória do sistema porta e migram para as veias mesentéricas pélvicas. As localizações habituais são as vênulas da parede do reto, sigmóide e intestino

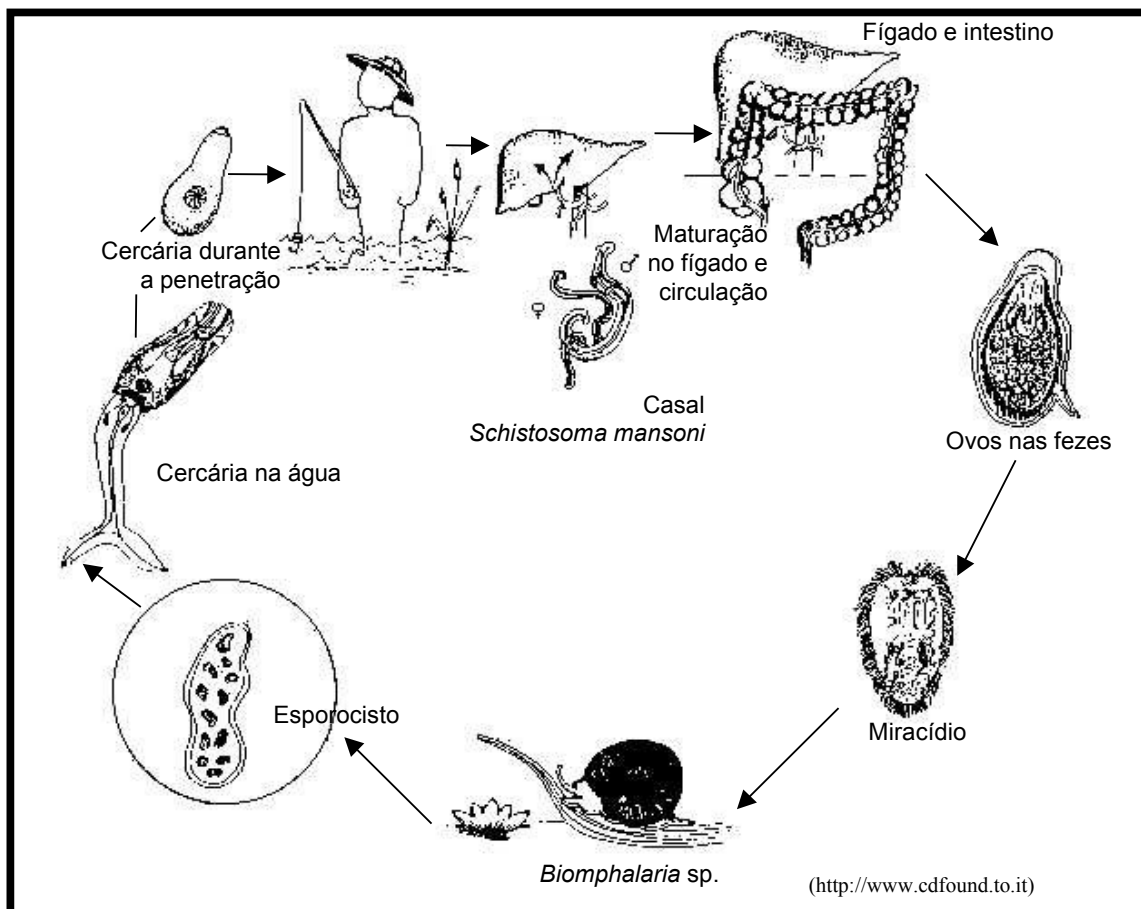


Figura 1: Ciclo biológico do *S. mansoni*. As cercárias ao saírem do caramujo nadam em busca de um hospedeiro, penetram pela pele do hospedeiro definitivo, o homem, e perdem sua cauda, iniciando-se o processo de transformação em esquistossômulos. Os esquistossômulos migram através da pele até as veias. Vermes adultos de *S. mansoni* residem mais freqüentemente nas veias mesentéricas superiores que drenam o intestino grosso. As fêmeas depositam seus ovos nas pequenas vênulas dos sistemas porta e perivesical, de onde são ativamente movidos para o lúmen do intestino sendo eliminados com as fezes. Em água fresca os ovos eclodem e liberam os mirácidos que nadam e penetram no hospedeiro intermediário, a *Biomphalaria*. Os estágios no caramujo incluem duas gerações de esporocistos e a produção e liberação de cercárias. (adaptado de <http://www.cdfound.to.it/HTML/sch1.htm>)

grosso (Bloch 1980, Rollinson & Simpson 1987). Os vermes adultos têm, aproximadamente, um centímetro em comprimento, são delgados e longos, e caracterizados por terem duas ventosas, uma oral e uma ventral . A fêmea é mais fina, cilíndrica e longa que o macho, existindo também diferenças fisiológicas e antigênicas entre macho e fêmea. A fêmea depende do contato com o macho para completar a sua maturação . O par está em constante associação, encontrando-se a fêmea no canal ginecóforo do macho. Os vermes alimentam-se de plasma e células do sangue venoso . A maior parte da energia das fêmeas é gasta na produção de ovos. Cerca de 340 ovos de *S. mansoni* são liberados por casal de vermes em ratos infectados a cada dia, sendo que este número aumenta em primatas . A produção de ovos começa 30 a 40 dias após a infecção. Medem, aproximadamente, 142 µm em comprimento e 60 µm em largura. A forma do ovo é característica para a esquistossomose mansônica por possuir uma espícula lateral. Os ovos são eliminados para o ambiente através das fezes de indivíduos infectados . Muitos dos ovos não ultrapassam a parede intestinal e são levados a órgãos e tecidos do hospedeiro pela corrente sanguínea. A presença destes ovos nos tecidos resulta na formação de granuloma, resultante da resposta do sistema imune do hospedeiro. Esta resposta de hipersensibilidade retém antígenos citotóxicos e mata o ovo, mas também causa danos ao tecido, sendo responsável pela manifestação clínica da esquistossomose crônica. O órgão mais atingido é o fígado, sendo o baço também um órgão bastante acometido . No momento da ovoposição, 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 miracídios morrem caso a expulsão não se complete dentro de três a quatro semanas após a ovoposição. Os ovos eliminados pelo hospedeiro através das fezes eclodirão se encontrarem condições adequadas como: água fresca com temperatura morna, baixa hipotonicidade e iluminação adequada, liberando assim o miracídio de dentro da casca do ovo . Uma vez que o ovo é rompido em água fresca, o miracídio emerge e começa a nadar ativamente. Os miracídios apresentam uma forma piriforme e medem, aproximadamente, 70 µm em largura por 160 µm em comprimento. Seu corpo é coberto por células anucleadas ciliadas e possui na sua extremidade anterior organelas sensoriais. O miracídio possui geotropismo negativo e fototropismo positivo, que o auxilia na localização do hospedeiro intermediário, além de ser atraído por substâncias químicas eliminadas pelo caramujo. O encontro deve ocorrer em um período de 24 horas. O miracídio penetra no hospedeiro intermediário específico por movimentos rotatórios e ação lítica. Após a penetração o miracídio perde seu revestimento epitelial e seus órgãos de penetração, atrofiando sua musculatura . Por volta do oitavo dia o miracídio apresenta-se como um tubo enovelado, imóvel, repleto de células germinativas em multiplicação, se transformado em

esporocisto. O esporocisto primário apresenta grandes células germinativas isoladas ou agrupadas. Por volta da segunda semana de existência rompem-se pra liberar esporocistos filhos, em número de 20 a 40. Os esporocistos secundários apresentam células germinativas, em constante multiplicação. Pouco a pouco, aglomerados celulares vão se diferenciando para formar cercárias. Os esporocistos podem formar várias gerações de cercárias. A transformação dos esporocistos em cercárias ocorre apenas quando atingem sua localização permanente, nas glândulas digestivas do caramujo. A cercária deixa o hospedeiro caindo em água fresca, fechando o ciclo de vida do parasito .

A esquistossomose mansônica tem o quadro clínico dividido em fase aguda e crônica. A fase aguda é caracterizada pelo período que precede a ovoposição. Logo no início da fase aguda, há um aumento na resposta de proliferação celular contra antígenos do parasito, diminuindo-se ao longo da fase aguda. A esquistossomose crônica tem seu quadro clínico dividido nas formas intestinal, hepatointestinal e hepatoesplênica. A lesão típica e elemento anatomo-patológico básico do processo esquistossomótico crônico é o granuloma que se forma em torno dos ovos do parasito. Os ovos que não conseguem deixar o organismo do hospedeiro são imobilizados e envolvidos por uma reação inflamatória, o granuloma. Muitos ovos são arrastados pela corrente sanguínea e ficam retidos nos capilares do espaço porta do fígado. Do território periportal, alguns ovos são levados até os pulmões, e são retidos nos capilares pulmonares. As lesões do sistema nervoso central e de diferentes órgãos são devido à localização ectópica dos ovos .

Segundo a Organização Mundial de Saúde, a esquistossomose é a terceira doença tropical de maior importância sócio-econômica e de saúde pública do mundo. A mortalidade na esquistossomose não é grande, porém a patologia é crônica e debilitante. A esquistossomose é endêmica em mais de 74 países em desenvolvimento, sendo que 80% dos casos encontram-se na África sub-Saariana. Estima-se que mais de 600 milhões de indivíduos estão sob o risco de infecção e que existam 200 milhões de indivíduos infectados, dos quais 120 milhões são sintomáticos e 20 milhões têm conseqüências severas da infecção . Por ano ocorrem aproximadamente 11 mil mortes diretamente relacionadas à esquistossomose. Seqüelas clínicas tardias e a mortalidade/morbidade indireta causadas pela esquistossomose estão relacionadas a 200 mil mortes por ano . O número de DALYs (*Disability Adjusted Life Years*) provenientes da esquistossomose atinge 1,7 milhões (valor considerado subestimado pelo *The Expert Committees on the Control of Schistosomiasis and Soil-transmitted Helminths*), correspondendo a 13% do número de DALYs entre as doenças tropicais. O valor de DALYs, em 2001, para as doenças tropicais correspondeu a 12,9 milhões, sendo 1,5 milhões para tripanossomíase, 649 mil para doença de Chagas, 2,3 milhões para leishmaniose,



5,6 milhões para filariose linfática e 987 mil para oncocercariose . As doenças tropicais acometem principalmente países subdesenvolvidos em que os recursos de saúde pública são muito pequenos e voltados principalmente para outras doenças como malária, HIV/AIDS e tuberculose .

A esquistossomose atinge principalmente crianças, com faixa etária entre 10 e 19 anos . O potencial cognitivo de crianças diminui consideravelmente quando infectadas por *Schistosoma*, diminuindo a sua capacidade de memória e velocidade em processar dados. Esse efeito é ainda maior entre crianças subnutridas .

Para um programa de controle visando interromper o ciclo biológico em um dos pontos do ciclo de transmissão da parasitose é importante o conhecimento epidemiológico, da biologia do verme, e também de fatores ligados ao hospedeiro intermediário e definitivo . O controle da doença vem sendo realizado através de processos químicos e/ou biológicos de eliminação do hospedeiro intermediário, através do saneamento e educação evitando o contato das pessoas com águas contaminadas e, principalmente, através do tratamento com drogas como praziquantel (única droga usada em larga escala no Brasil) e oxaminiquine .

### **I.1.2 Genômica do *S. mansoni***

Atualmente a abordagem mais promissora para a identificação de novos alvos de drogas, vacinas e reagentes para diagnóstico, assim como para a compreensão da resistência a drogas, diversidade antigênica, infectibilidade e patologia, consiste em compreender e decifrar as informações nos genomas dos parasitos .

*S. mansoni* tem um genoma diplóide com sete pares de cromossomos autossômicos e um par de cromossomo sexual. Os cromossomos variam em tamanho de 18 a 73 Mb e podem ser distinguidos pelo tamanho, forma e padrão de bandas C. O tamanho do genoma haplóide é de aproximadamente 270 Mb com um conteúdo CG de 29,4% . A metilação do DNA genômico não é observada no parasito adulto . O genoma do parasito é constituído de 4% a 8% de seqüências de DNA altamente repetitivas (>1.000 cópias), 35% a 40% de seqüências de média repetitividade (~100 cópias) e 60% de seqüências correspondentes a famílias de genes ou regiões de cópia única . *Schistosoma* corresponde a um dos primeiros organismos a desenvolverem dimorfismo sexual e cromossomos sexuais heteromórficos. A fêmea é heterogamética, possuindo o par de cromossomos sexuais ZW e o macho o par ZZ .

O estudo do transcriptoma de *S. mansoni* foi uma iniciativa brasileira, financiada por agências locais que, teve início em 1992 . A partir de 1994, a Organização Mundial de Saúde iniciou o financiamento da Rede Genoma de *Schistosoma* para a descoberta de novos genes com o objetivo final de identificar novos alvos para o desenvolvimento de drogas e vacinas.

Durante este período a comunidade científica mundial produziu aproximadamente 16.000 ESTs . Recentemente, dois grandes projetos de sequenciamento do transcriptoma do *S. mansoni* foram realizados. Ambos os projetos são brasileiros .

O primeiro, já publicado, 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 . O projeto gerou 124.681 ORESTES e ESTs úteis de *S. mansoni*. As seqüências foram agrupadas resultando em 30.988 SmAE (*Schistosoma mansoni assembled ESTs*), correspondendo a aproximadamente 92% do transcriptoma. Dentre as SmAEs anotadas, 23% encontraram relação com outras seqüências de *S. mansoni* já depositadas, tendo 2% delas identidade com genes conhecidos e 21% com ESTs depositadas no dbest. Os outros 77% não encontraram relação com seqüências de *S. mansoni*, sendo descritas como novos genes relatados de *S. mansoni*. Destas, 1% apresentou identidade com proteínas conhecidas, sendo genes parálogos, 20% com genes de outros organismos, sendo genes ortólogos e 55% sem relação com nenhum outro gene já relatado de outras espécies . A comparação das ESTs de *S. mansoni* com seqüências do bancos de dados permitiu identificar um grande número de genes de interesse. Foi possível classificar por *Gene Ontology* 8.001 ESTs, obtendo-se principalmente proteínas envolvidas em metabolismo e processos biológicos . Na busca por domínios conservados foram encontradas 180 proteínas identificadas como proteínas quinase, sugerindo que *S. mansoni* tem um complemento de quinase menor que qualquer outro metazoário já estudado, apesar de ainda assim ser a família de proteínas mais abundante em *S. mansoni* .

O segundo projeto, financiado pela FAPEMIG/MCT/CNPq, consiste de uma rede genômica formada por instituições do Estado de Minas Gerais que, têm como objetivo caracterizar o transcriptoma de diferentes estágios de desenvolvimento do parasito, a partir da geração de 70 mil ESTs convencionais. Esse projeto contribuirá na complementação do projeto desenvolvido em São Paulo e também, será indispensável em estudos de proteoma e SAGE . Até o final de 2004, foram geradas 74 mil ESTs das diferentes fases do parasito, ainda não disponíveis nos bancos de dados públicos (rgmg.cpqrr.fiocruz.br).

Até o momento, mais de 150.000 ESTs foram depositadas pela comunidade científica na divisão dbest do GenBank. O baixo nível de redundância indica que ainda existam muitos novos genes para serem gerados pelas bibliotecas disponíveis .

Os institutos TIGR (The Institute for Genomic Research) e Sanger foram os responsáveis pelo sequenciamento em larga escala do genoma de *S. mansoni*, gerando juntos uma cobertura de 9 vezes do genoma completo, estimando-se que menos de 0,5% do genoma não tenha sido seqüenciado. A estratégia de seqüenciamento utilizada foi de *Whole Genome*

*Shotgun* (WGS), por seleção aleatória de clones e com sequenciamento de ambas as extremidades. Estas seqüências estão atualmente sendo montadas e anotadas .

## **I.2 RELAÇÃO PARASITO E MEIO AMBIENTE**

O parasito *S. mansoni* tem um ciclo de vida complexo, em que sua sensibilidade a fatores do ambiente estimula mudanças fisiológicas, morfológicas e bioquímicas. A capacidade de infectar ativamente ambos os hospedeiros, assim como a migração pelos sistemas dos hospedeiros demonstram uma sofisticada coordenação dos seus sistemas fisiológicos em diferentes fases de desenvolvimento em resposta ao meio em que se encontra . O crescimento e desenvolvimento dos vermes adultos de *Schistosoma* requerem uma comunicação permanente do parasito com o ambiente em que se encontram, com os hospedeiros, definitivo e intermediário, e também na relação entre macho e fêmea .

Na relação parasito-hospedeiro, o verme é capaz de sobreviver por décadas no organismo parasitado utilizando-se de suas biomoléculas para completar seu ciclo. Já foi visto que o parasito utiliza de TNF- $\alpha$  do hospedeiro com sinal estimulatório para a produção de ovos . A utilização de IL-7 já foi demonstrada para o crescimento e fecundação do verme . A presença de receptor de TGF- $\beta$  expresso na superfície do tegumento sincicial indica a utilização deste fator pelo verme . Fatores de crescimento epitelial humano estimulam a expressão do fator de crescimento epitelial no parasito . Produtos de células T CD4+ modulam o desenvolvimento do parasito . Em hospedeiros imunodeprimidos o desenvolvimento do parasito é retardado. Uma possível explicação para este fenômeno é que o parasito aguarda melhores condições de sobrevivência do hospedeiro para então finalizar o seu processo de crescimento, pareamento, fecundidade e transmissão . Estes resultados indicam que o parasito está tão adaptado ao meio, que utiliza proteínas imunoreguladoras do hospedeiro como sinal de replicação e transmissão .

A fêmea necessita do contato com o macho para sua maturação. Já se sabe que a maturação da fêmea não depende da transferência de esperma, mas sim do contato direto com o macho, com possíveis transferências de moléculas sinalizadoras, como hormônios esteróides, que são captadas por receptores . Fêmeas virgens que crescem na ausência de machos são significativamente menores, além de não colocarem ovos. Quando as fêmeas são separadas dos seus parceiros cessam a ovoposição e começam a regredir para seu estado reprodutivo imaturo. Assim que repareadas, voltam à sua atividade reprodutiva normal . Foi observado que a separação do casal implica na interrupção da transcrição de alguns genes responsáveis pela diferenciação dos vitelocistos . O desenvolvimento reprodutivo do macho não depende da interação com a fêmea. No entanto, o macho necessita da fêmea para regular o mecanismo

de acúmulo de glutathiona e lipídeo, na utilização de lipase e estimulando a incorporação de tirosina . Hoffman e colaboradores estudaram a expressão diferencial de genes entre vermes machos e fêmeas. Foi observado que a transcrição de genes variava em verme adulto, encontrando-se 12 novos transcritos associados à fêmea e 4 novos associados a machos.

Por uma perspectiva biológica, o balanço das trocas moleculares entre o parasito e o hospedeiro dita o sucesso ou a falha no estabelecimento da infecção dentro do hospedeiro . Estes fenômenos requerem comunicações moleculares, sendo estas interações mediadas por processos de transdução de sinal que determinam uma expressão diferencial de genes .

### **I.3 MECANISMOS DE TRANSDUÇÃO DE SINAL**

Os complexos comportamentos celulares, como a manutenção da homeostase ou proliferação, são em grande parte, estimulados por sinais extracelulares específicos. A célula tem que integrar a informação proveniente destes sinais a fim de construir uma resposta apropriada: viver ou morrer, proliferar ou permanecer em repouso, entre outras . É necessário um sistema elaborado de sinalização para que a célula responda de maneira específica ao sinal recebido. Esse sistema é formado por um conjunto de proteínas que inclui proteínas receptoras que são estimuladas pelo sinal externo e mais uma variedade de proteínas de sinalização intracelulares que distribuem o sinal pela célula, até atingirem seus alvos específicos (Figura 2). A transdução de sinal demanda mais que trocas moleculares para ativar ou inativar processos celulares. Conexões lineares simples não são suficientes. O sucesso da transdução de sinal requer a criação de ramos que interajam

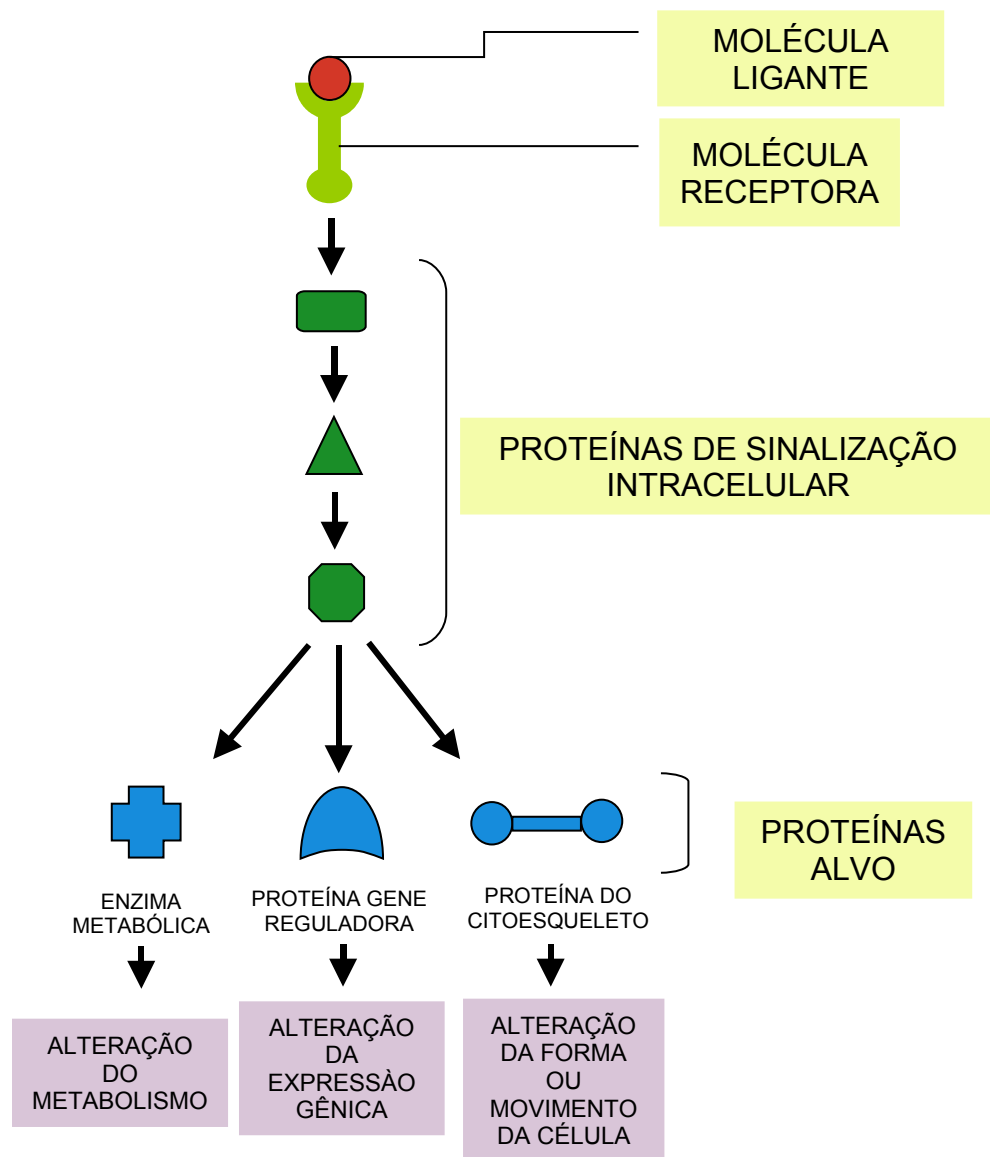


Figura 2: Exemplo de uma via de sinalização intracelular simples ativada por molécula sinalizadora extracelular. A molécula sinalizadora se liga à proteína receptora, ativando uma via de sinalização intracelular mediada por uma série de proteínas intracelulares. Este sinal intracelular interage com uma proteína alvo, alterando-a, por exemplo, por fosforilação, e gerando uma resposta celular ao sinal (adaptada de Albert *et al* 2002).

ativamente e influenciem um ao outro. A simplicidade da comunicação molecular é transformada em sofisticação pela maneira como as vias são interligadas .

Cada célula está programada para responder a combinações específicas de moléculas sinalizadoras. Qualquer célula de um organismo está exposta a muitos sinais provenientes do seu ambiente. A célula deve responder seletivamente, de acordo com sua característica específica, adquirida através da especialização celular progressiva durante o desenvolvimento. Células diferentes podem responder de formas distintas ao mesmo sinal químico em função do grupo de proteínas receptoras, ou de acordo com a maquinaria celular que integra e interpreta a informação recebida. Assim, a mesma molécula ligante tem, freqüentemente, efeitos diversos sobre diferentes células-alvo . Moléculas ligantes incluem proteínas, peptídeos, aminoácidos, nucleotídeos, esteróides, retinóides, ácido graxos e gases dissolvidos como o óxido nítrico e monóxido de carbono. As moléculas ligantes geralmente atuam em baixa concentração e os receptores específicos as reconhecem com alta afinidade .

Moléculas ligantes são reconhecidas por receptores específicos localizados na célula alvo. Na maioria dos casos os receptores são proteínas de transmembrana na superfície da célula-alvo. Quando os receptores se ligam a uma molécula ligante, tornam-se ativados e geram uma cascata de sinais intracelulares que alteram o comportamento da célula . Existem três famílias principais de receptores de superfície celular, cada uma das quais transmitem os sinais extracelulares de uma maneira diferente (Figura 3):

- A. receptores associados a canais iônicos que abrem e fecham canais rapidamente em resposta à ligação de um neurotransmissor, por exemplo, o receptor de acetilcolina ;
- B. receptores associados a proteínas G que ativam ou inativam indiretamente enzimas associadas à membrana plasmática ou canais iônicos via proteínas ligantes de GTP, por exemplo, a rodopsina ;
- C. receptores associados a enzimas, que atuam como enzimas diretamente ou associadas a estas, por exemplo, os receptores de fatores de crescimento. As enzimas são geralmente proteínas quinase que fosforilam proteínas específicas na célula-alvo .

As moléculas sinalizadoras intracelulares são classificadas em segundo mensageiro ou proteínas de sinalização intracelulares. Os segundos mensageiros são gerados em grande número frente à ativação do receptor e se difundem pela célula, podendo ser solúveis no citosol como o AMP cíclico e o  $\text{Ca}^{2+}$ , ou solúveis na membrana como o diacilglicerol . As proteínas de sinalização intracelulares transmitem seu sinal através da

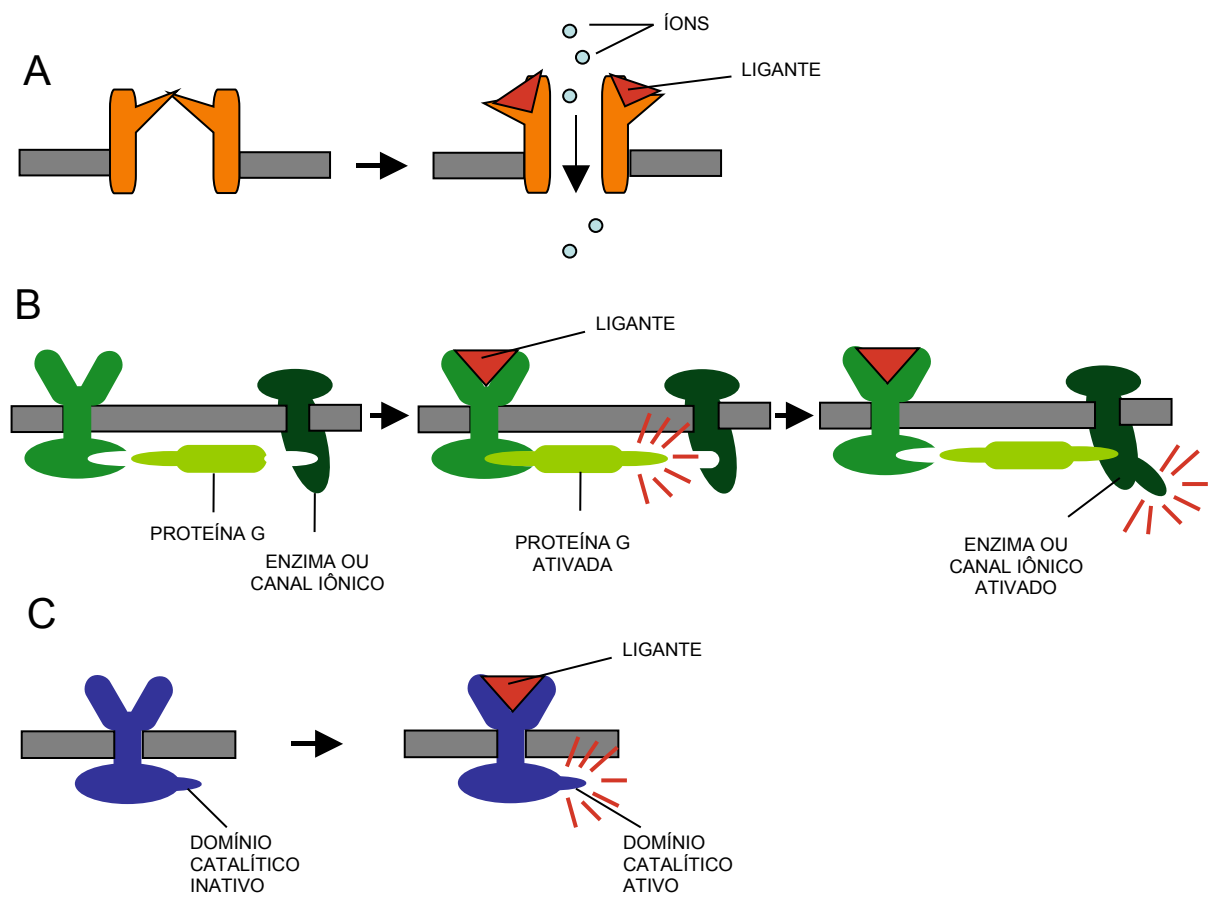


Figura 3: Famílias de receptores de superfície celular. A- receptor associado a canal iônico. B- receptor associado a proteína G. C- receptor associado a enzima (adaptada de Albert *et al* 1994).

ativação de uma outra proteína de sinalização ou gerando um segundo mensageiro. As proteínas de sinalização são classificadas de acordo com sua função em particular . Muitas proteínas de sinalização agem por troca molecular, que ao receberem um estímulo passam de um estado inativo para um estado ativo, ou vice e versa. Essa troca é realizada pelo ganho ou perda de grupo fosfato, ativando e inativando a proteína, respectivamente . Existem duas classes de proteínas de sinalização:

1- As proteínas que são fosforiladas por proteínas quinase e desfosforiladas por proteínas fosfatases. Grande parte das proteínas sinalizadoras são proteínas quinase que irão fosforilar outras proteínas sinalizadoras, o que leva a uma cascata de sinalização ;

2- As proteínas que se encontram ativadas pela presença de GTP e desativadas pela presença de GDP. A troca da molécula GTP por GDP é realizada por GTPases .

## **I.4 AS PROTEÍNAS QUINASE**

### **I.4.1 Função e Atividade**

As proteínas quinase (PKs) pertencem a uma grande família de enzimas, muitas das quais mediam respostas a estímulos externos em células eucarióticas . As proteínas quinase são proteínas de fosforilação reversível que desempenham um papel central na regulação de funções básicas de todos os eucariotas, tais como: replicação, controle do ciclo celular, transcrição gênica, rearranjo do citoesqueleto, apoptose, metabolismo energético e outras. As PKs são também requeridas em funções mais finas, em eucariontes mais complexos, como: diferenciação celular de tecidos e órgãos, comunicação entre células, interação da célula com o substrato e mediação de interações complexas com o ambiente externo . As proteínas quinase agem na fosforilação do substrato, transferindo um grupo fosfato (P) do ATP para a cadeia de aminoácido de uma proteína substrato (Figura 4). O fosfato é carregado negativamente o que leva a mudança na conformação da proteína, causando assim modificações na sua estrutura e como consequência na sua atividade. Muitas PKs são moduladas por autofosforilação ou fosforiladas por outras proteínas quinase. Outros domínios dentro das PKs regulam sua atividade quinase, ligando-a a outras moléculas sinalizadoras ou relocando a proteína (Figura 5).



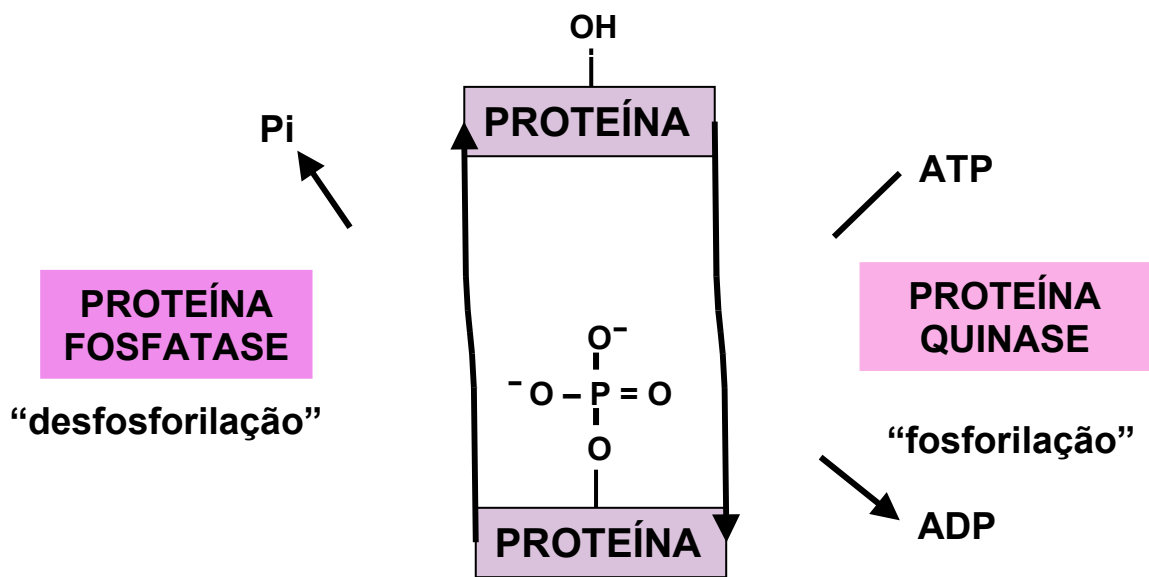


Figura 4: Fosforilação reversível de proteínas sinalizadas. As proteínas quinase catalisam a transferência de um grupo fosfato (P) do ATP [ADP(P)] a uma proteína substrato específica. A proteína fosfatase remove o grupo fosfato, permitindo que a proteína substrato volte ao seu estado basal.

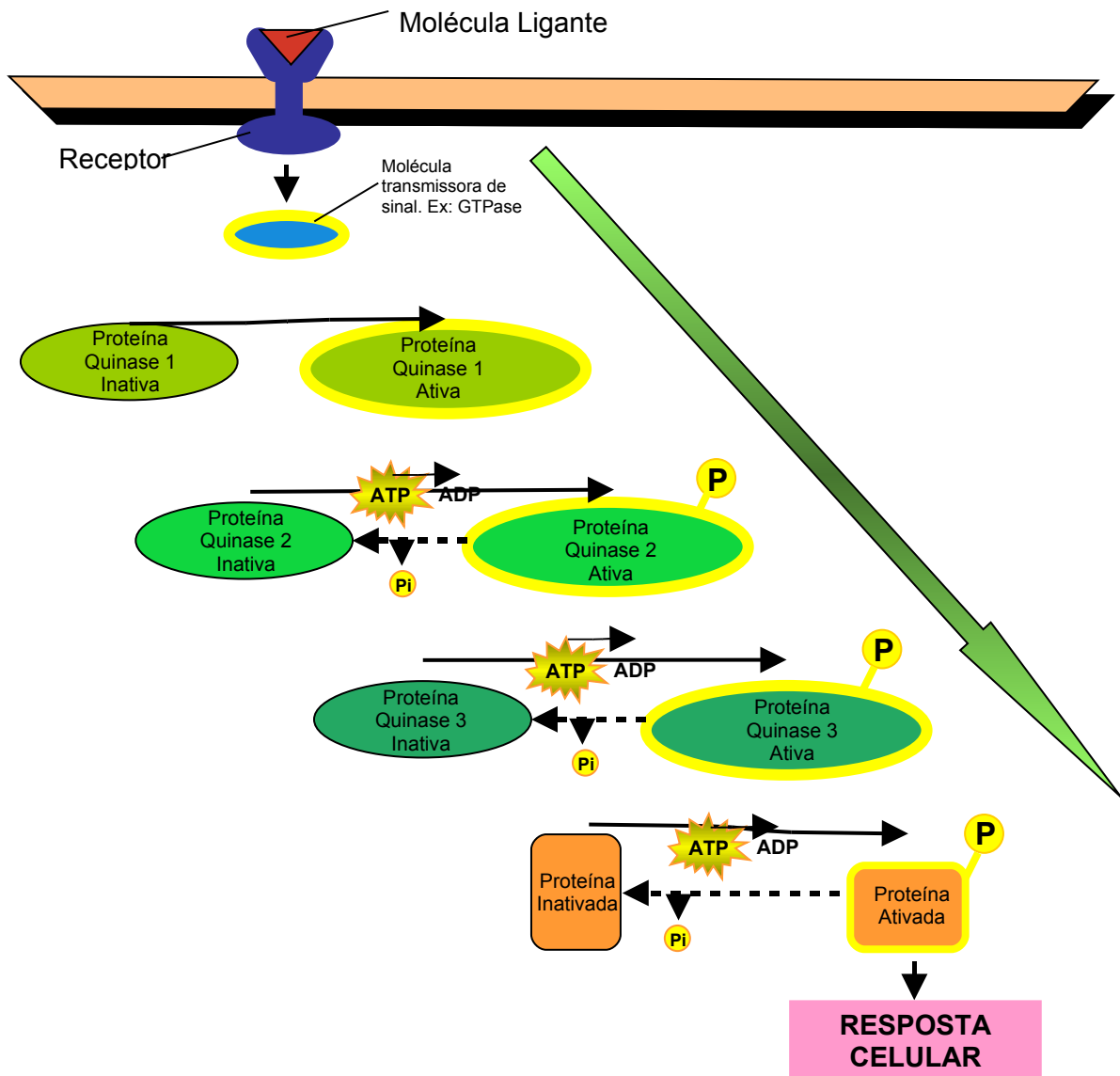


Figura 5: Cascata de sinalização. Uma proteína chave atua em uma segunda proteína alvo, esta proteína alvo irá atuar como proteína chave em uma terceira proteína. Essa parceria continuará até que uma proteína chave tenha efeito sobre alguma maquinaria celular criando uma resposta celular ao estímulo inicial. Moléculas sinalizadoras extracelulares são reconhecidas por receptores específicos localizados na células alvo, que ao serem ativado geram uma cascata de sinais intracelulares. As proteínas de sinalização intracelulares transmitem seu sinal através da ativação de uma outra proteína de sinalização. Muitas proteínas de sinalização agem por troca molecular, com ganho ou perda de grupo fosfato, ativando e inativando, respectivamente. Na grande maioria, as proteínas sinalizadoras são

proteínas quinase que irão ativar outras proteínas sinalizadoras, o que leva a uma cascata de sinalização.

### **I.4.2 Presença de Proteínas Quinase em Genomas Seqüenciados**

As PKs têm sido caracterizadas não somente por tradicionais técnicas bioquímicas, mas também com a utilização de análises de domínios catalíticos a partir de seqüência de aminoácidos de sua estrutura primária . PKs compreendem a maior família de proteínas, correspondendo de 1,5 a 2,5% de todos os genes eucarióticos . Usando *Caenorhabditis elegans* como modelo para estudo de transdução de sinal, por ser o primeiro organismo multicelular totalmente seqüenciado, as proteínas quinase foram compreendidas na segunda maior família de domínios protéicos nestes vermes, com um total de 411 seqüências completas de PKs . Com a elucidação do projeto genoma humano foram identificados 518 genes codificantes para PKs, consistindo 1,7% do total de genes humanos. Dentre 258 PKs analisadas foram identificados 83 tipos de domínios, dos quais a maioria encontram-se relacionados com a interação entre proteínas sinalizadoras, por exemplo SH2, que reconhece e se liga a resíduos de tirosina fosforilados . O kinoma de rato foi elucidado, tendo sido identificados 540 genes de PKs, havendo uma correspondência de 510 genes ortólogos entre homem e rato . Comparando-se os kinomas de *S. cerevisiae* (levedura), *C. elegans* (verme), *D. melanogaster* (inseto) e *H. sapiens* (mamífero), de 209 subfamílias analisadas, 51 estão presentes nos quatro genomas, 7 são únicas de levedura (*S. cerevisiae* e *S. pombe*), 15 de *C. elegans*, 13 de *H. sapiens* e nenhuma de *D. melanogaster* .

A identificação de genes similares às proteínas quinase em procariotas sugere a possibilidade de que o gene do qual se originam as proteínas quinase tenha surgido antes da divergência de procariotas e eucariotas . Pkn1 é uma proteína quinase-like de bactéria primeiramente descrita em *Myxococcus xanthus*. Proteínas relacionadas com Pkn estão presentes em outros procariontes incluindo: *Streptomyces*, *Bacillus*, *Mycobacterium*, *Pseudomonas*, *Chlamydia*, e *Synechocystis*. Estas proteínas estão envolvidas com virulência, metabolismo secundário, esporulação, ciclos complexos de crescimento. No entanto, não se encontram Pkn em bactérias com ciclo de desenvolvimento mais simples como *Escherichia coli* e *Haemophilus influenza*, sugerindo que as proteínas quinase poderiam ter sido adquiridas por transferência lateral, com transferência gênica entre táxons não relacionados .

### **I.4.3 Classificação de Proteínas Quinase**

As proteínas quinase são classificadas de acordo com similaridades na estrutura primaria, baseando-se na filogenia dos domínios conservados. O domínio catalítico característico da superfamília PK consiste de aproximadamente 250 a 300 aminoácidos e

divide-se em onze subdomínios. Estes subdomínios são reconhecidos por serem invariantes (conservados em mais de 95% de um total de 370 seqüências estudadas de proteínas quinase) . Eles estão fortemente implicados no desempenho de importante papel na função catalítica das enzimas, tendo ação direta no sítio ativo ou contribuindo para a formação do sítio ativo em uma estrutura secundária. A região não conservada parece ocorrer em voltas (*loops*), permitindo que as regiões catalíticas se encontrem. Aminoácidos individuais, altamente conservados dentro dos subdomínios, participam da ligação ao ATP e da transferência do fósforo .

As proteínas quinase são classificadas não apenas pelo seu domínio catalítico, mas também pelo conjunto e ordem de diversos outros domínios e regiões presentes na proteína completa. Existem duas subdivisões dentro da superfamília das proteínas quinase: as serinas/treoninas quinase, que fosforilam proteínas nas serinas e nas treoninas; e as tirosinas quinase, que fosforilam proteínas somente nas tirosinas . As PKs usam o fosfato- $\gamma$  do ATP (ou GTP) para gerar fosfatos monoésteres usando serina/treonina ou tirosina como aceptores .

As proteínas serinas/treoninas quinase contêm os principais grupos, separados de acordo com sua estrutura básica e seu modo de regulação em: AGC (grupo contendo PKA, PKG e PKC), CaMK (grupo contendo as proteínas quinase dependentes de cálcio/calmodulina) e CMGC (grupo contendo CDK, MAPK, GSK3 e CLK). Dentre as proteínas tirosinas quinase encontra-se o grupo PTK (grupo contendo as famílias de proteínas tirosinas quinase), que se divide em diversas famílias podendo ou não estar ancoradas a membrana celular (Quadro 1).

#### **I.4.4 As Proteínas Tirosinas Quinase (PTKs)**

PTKs são proteínas encontradas em todos os organismos eucariotas multicelulares . As proteínas tirosinas quinase podem estar presentes nas células ancoradas à membrana, atuando como receptoras; livres no citoplasma, fazendo parte da cascata de sinalização; ou no núcleo, participando diretamente da ativação gênica. A ativação do domínio catalítico PTK resulta da interação com outras proteínas de sinalização propagando o sinal de forma específica. Não foi relatada nenhuma PTK em leveduras, não obstante 49 de 239 PKs de *D. melanogaster* e 105 de 454 PKs de *C. elegans* foram classificadas como PTKs .

##### **I.4.4.1 As Proteínas Tirosinas Quinase Receptoras (RTKs)**

As RTKs são compostas de três regiões distintas: um domínio de ligação extracelular, uma hélice transmembrana e uma região citoplasmática que realiza a atividade quinase

Quadro 1: Classificação das proteínas quinases, divididas em grupos e famílias segundo Hanks e Quinn (1991). A classificação teve como critério primário a similaridade da sequência de aminoácidos do domínio catalítico.

SUBDIVISÕES	GRUPOS	FAMÍLIAS		
<b>A- Serina/Treoninas</b>	<b>1- ACG</b>	PKA		
		PKC		
		RAC		
		Fosforila proteína G acoplada a receptor		
		ACG-relatadas		
		Fosforila proteína S6 ribossomal		
		Dbf		
		PVPK1		
		Outras ACG relacionadas		
		<b>2- CaMK</b>	Reguladas por Ca <sup>2+</sup> / Calmodulina	
			KIN1/SNF1/Nim1	
		<b>3- CMGC</b>	Outras CaMK relacionadas	
			CDKs	
			Erk (MAP)	
			GSK3	
Casein kinase II				
Clk				
<b>B - Tirosinas</b>	<b>4.a-PTK não receptoras (não ancoradas a membrana)</b>	Src		
		Tec/Atk		
		Brk		
		Csc		
		<b>Fes/Fps/Fer</b>		
		Abl		
		Syk/ Zap70		
		Tyk2/Jak1		
		Ack		
		Fak		
		<b>4.b-PTK receptoras (ancoradas a membrana)</b>	EGFR (receptor de fator de crescimento da epiderme)	
	Receptor Eph/ Elk/ EcK			
	Axl			
	Tie/ Tec			
	PDGFR (receptor de fator de crescimento derivado de plaqueta)			
	FGFR (receptor de fator de crescimento de fibroblasto)			
	Receptor de Insulina			
	Ltk/ Alk			
	Ros/ Sev			
	Tkr/ Ror			
	Ddr/ Tkt			
	HGFR (receptor de fator de crescimento de hepatócito)			
	Kin15/ 16			
	Outras PTK relacionadas			
	<b>C- Outras PKs</b>		<b>OPK (outras PKs não relacionadas aos grupos acima)</b>	Polo
				MEK/STEE7
		PAK/STE7		
PAK/STE20				
MEKK/STE11				
NimA				
Wee1/mik1				
Quinases envolvidas em controle transcricional				
Activin/ TGF-β (receptor de fator de crescimento tumoral)				
Rat				
PSK/PTK com domínio de linhagem mista zipper leucina				
caseinal				
Resceptor putativo de plantas florindo				
PKN				
Outras PK relacionadas				

. A ativação de RTKs é tipicamente iniciada pela ligação de um ligante ao sítio específico no domínio extracelular do receptor. A presença de um estímulo na porção extracelular do RTK ativa o domínio catalítico do outro lado da membrana através da dimerização do receptor, fazendo com que ocorra autofosforilação . Esse evento de autofosforilação ativa a quinase produzindo um novo sítio de ligação para moléculas adaptadoras intracelulares.

As famílias de RTKS apresentam diversificados domínios extracelulares . O primeiro receptor protéico reconhecido como sendo uma proteína quinase específica de tirosina foi o receptor do fator de crescimento da epiderme (EGFr). Existem, porém vários receptores para fatores de crescimento e diferenciação que também pertencem aos receptores proteínas tirosinas quinase, tais como: receptor de fator de crescimento derivado de plaquetas (PDGFr), receptor de fator de crescimento de fibroblasto (FGFr), receptor de fator de crescimento do hepatócito (HGF), receptor de fator de crescimento similar à insulina-1 (IGFr-1), receptor de fator de crescimento da célula nervosa (NGFr), receptor de fator de crescimento endotelial vascular (VEGFr), receptor de fator estimulante de colônia de macrófago (M-CSFr), receptor de insulina (Ir), dentre outros .

#### **I.4.4.2 Proteínas Tirosinas Quinase Citoplasmáticas (NRTK)**

Um conjunto de proteínas sinalizadoras intracelulares foi identificado como sendo capaz de ligar-se as fosfotirosinas de PTKs . Embora as proteínas sinalizadoras intracelulares que se ligam aos resíduos de fosfotirosina nas proteínas tirosinas quinase ativadas tenham estruturas e funções variadas, elas geralmente compartilham dois domínios não catalíticos altamente conservados, denominados SH2 e SH3. Esses domínios SH não possuem atividade catalítica intrínseca, sua função é acoplar a ligação entre proteínas fosforiladas em tirosina, bem como proteínas tirosinas quinase ativadas, a outras proteínas. Muitas dessas proteínas sinalizadoras são também proteínas tirosinas quinase que, além de serem fosforiladas estariam também fosforilando outras proteínas substrato, ocorrendo assim uma cascata de sinalização .

#### **I.4.4.3 Proteínas Tirosina Quinase “Nucleares”**

Proteínas tirosinas quinase solúveis podem ser encontradas no citoplasma, e também no núcleo. Enquanto as proteínas tirosinas quinase citoplasmáticas são reguladas por sinais extracelulares, as proteínas tirosinas quinase nucleares parecem ser reguladas por sinais intracelulares . PTKs já foram identificadas no núcleo por ensaios de imunolocalização com anticorpos específicos para fosfotirosina. Porém, a fosforilação pode ter ocorrido tanto no núcleo quanto no citoplasma. STAT (transdutor de sinal e ativador de transcrição) é fosforilada na membrana plasmática pelo receptor ativado, Jak (janus quinase), e então transferida para o núcleo. MAPK (proteína quinase mitógeno-ativada) é ativada por uma variedade de sinais extracelulares, não se conhecendo o mecanismo através do qual é

translocada para o núcleo. Wee1, Rak, Abl (tirosina quinase Abelson), Fes/Fer (sarcoma de felino/relacionada a Fes), FerT, Fgr (Gardner-Rasheed de felino) e Src são outras proteínas quinase encontradas no núcleo. A descoberta da translocação para o núcleo de proteína tirosina quinase citoplasmática fosforilada explica, em parte, o aumento de fosfotirosina no núcleo após estímulo com hormônio de crescimento e fator de crescimento insulina tipo I. O domínio SH2 é frequentemente observado em proteínas que são translocadas para o núcleo, como na STAT e nas proteínas tirosinas quinase encontradas no núcleo. As proteínas tirosinas quinase nucleares participam da regulação de transcrição, no ciclo celular e em outros processos nucleares

#### **I.4.5 Importância das Proteínas Quinase para a Integridade da Vida**

As células animais normalmente se dividem quando estimuladas por fatores de crescimento, que são normalmente produzidos por outras células e geralmente atuam via RTKs. Os exemplos melhor estudados das vias de sinalização são em células cancerosas. As células cancerosas proliferam em excesso por se tornarem capazes de se dividir sem o estímulo de outras células, não estando sujeitas ao controle normal. Isto ocorre por haver mutações nos genes que codificam as proteínas sinalizadoras. Sendo assim, qualquer mutação que resulte na produção de uma proteína ativa anormal que atue nas vias de sinalização da transmissão do sinal de um fator de crescimento para o núcleo, pode contribuir para a promoção do câncer através da estimulação da proliferação da célula na ausência dos sinais extracelulares apropriados. Pelo fato de erros na codificação de proteínas quinase serem uma causa comum de câncer, diabetes e outras doenças, uma compreensão de como estas enzimas regulam tantas funções, pode possibilitar a elaboração de estratégias de intervenção terapêutica. Muitas estratégias no combate ao câncer têm sido baseadas no desenvolvimento racional de drogas a partir do estudo de PTKs.

### **I.5 PROTEÍNAS DE *S. MANSONI* ENVOLVIDAS NA TRANSDUÇÃO DE SINAL**

Recentemente, uma grande variedade de moléculas sinalizadoras foram identificadas e descritas em *S. mansoni*, sendo algumas delas: PKC, SmSmad1, SmSmad2, SmSmad-4, SmMAK16, SmRK1/SmT $\beta$ RI, SmRK2/SmT $\beta$ RII, SmRXR1, SmRXR2, SmFTZ-F1, SmRhoI, SmFKBP12, schP2X, Sm14-3-3 $\epsilon$ , eIF2 $\alpha$ , SMA3, SIP, SmRas1, CaBPs, SmTOR, HSF, GAP e MAP quinase. No entanto, poucas PTKs foram identificadas e descritas.

Algumas das proteínas sinalizadoras parecem estar envolvidas na via de sinalização de SmRK. SmRK é um membro divergente da família de receptor serina/treonina quinase TGF- $\beta$

, possivelmente envolvido na resposta a fatores de crescimento do hospedeiro como: migração celular, diferenciação, adesão e apoptose. SmRK1 é um membro TGF- $\beta$ , enquanto SmRK2 é um membro receptor Activina tipo 2, ambos da família TGF- $\beta$  de *S. mansoni*. Ambas as proteínas estão localizadas na superfície do tegumento do parasito, sendo possíveis parceiras. SmRK1 está também localizada nas gônadas das fêmeas . As SmRK parecem fosforilar eIF2 $\alpha$  . Sm14-3-3 $\epsilon$  é uma proteína citoplasmática associada à TGF- $\beta$ . A superexpressão de Sm14-3-3 $\epsilon$  aumenta a sinalização de TGF- $\beta$ , enquanto a superexpressão de eIF2 $\alpha$  inibe-a . Smads são encontrados nos mesmos estágios de desenvolvimento de SmRK1 e são também preferencialmente expressos nas gônadas e tegumento. Smads são capazes de interagir com moléculas receptoras levando a mensagem ao núcleo. SmSmad2 interage com SmRK1, enquanto SmSmad4 interage com SmSmad1 e SmSmad2, além de fosforilar Erk1/2 (quinase regulada por sinal extracelular) . FKBP12 influencia em eucariotas uma variedade de respostas envolvidas na regulação da divisão celular, diferenciação e homeostase iônica e está envolvida na via de sinalização de TGF- $\beta$  e da fosfatase calcineurina. Foi demonstrado que SmFKBP12 é uma parceira direta de SmRK1 estando ambas presentes no tegumento e na gônada da fêmea .

As demais proteínas sinalizadoras identificadas em *S. mansoni* participam de vias e funções mais diversificadas. PKC é uma proteína quinase relacionada com a transdução de sinal na superfície do parasito, com maior expressão em vermes adulto do que nas formas larvais . SmMAK16 apresenta uma porção sinalizadora nuclear e sítio para fosforilação de CK2 (caseína quinase 2), estando relacionada na biogênese da subunidade 60S do ribossomo e no ciclo celular, apresentando maior nível de expressão em vermes fêmea . Proteínas SmRXR são receptores nucleares ativadores de transcrição gênica. Eles são proteínas constitutivas que participam da ativação do precursor do gene p14 da casca do ovo, estando localizados nas células vitelínicas. Os SmRXR podem também participar da ativação de sistemas de levedura em híbrido simples . SmFTZF1 é um outro receptor nuclear com domínio de ligação a DNA e de atividade bem conservados, estando relacionado com o desenvolvimento e diferenciação sexual . SmTOR é uma proteína transmembrana da superfície do tegumento sem homologia a nenhuma outra proteína já relatada, sendo capaz de se ligar a C2 humano formando CP C3 convertase . CaBP é uma proteína de ligação a cálcio-calmodulina ou gelsolina . SchP2X está relacionada a abertura do canal iônico de ATP . SmRhoI é uma GTPase com possível participação na organização do citoesqueleto, transcrição gênica, ciclo celular e tráfico de membrana e está mais presente em vermes fêmea . SmRas é uma proteína G presente em todos os estágios, mas mais expressa em vermes



fêmeas . SMA3 é um ortólogo de ca-ATPase, sugerindo a participação no controle da homeostase por cálcio, estando presente no tegumento de adultos .

### **I.5.1 Proteínas Tirosinas Quinase Identificadas em *S.mansoni***

Em *S. mansoni*, três RTKs e quatro NRTKs foram identificadas e caracterizadas: SmRTK-1, SmRTK-2, SER, TK5, TK4, TK3 e SmFes, que serão descritas a seguir:

SmRTK1 é uma proteína de membrana com domínio de ligação extracelular (similar ao domínio de várias outras proteínas que compartilham a estrutura Vênus *Flytap*-VFT) e domínio TK citoplasmático (similar ao domínio catalítico do receptor de insulina-IR). O gene SmRTK1 é expresso em todos os estágios de desenvolvimento. Nos machos, encontra-se preferencialmente nas células parenquimais. Nas fêmeas, encontra-se principalmente nos ovócitos e ductos vitelínicos. Especula-se que SmRTK1 constitua um original RTK GABA, que esteja envolvido no reconhecimento de ferormônio, necessário para o desenvolvimento dos ovários da fêmea .

SmRTK-2 é uma tirosina quinase similar aos membros da família de receptores de insulina. Tem possivelmente um papel na maturação de células vitelíneas de fêmeas e sua expressão é influenciada pela presença do verme macho .

SER é um receptor de fator de crescimento epitelial que tem homologia ao domínio TK da família erbB. O gene é traduzido em uma proteína de 170 kDa dividida em peptídeo sinal, domínio extracelular rico em cisteína, seqüência hidrofóbica transmembrana e domínio TK intracelular. A proteína SER encontra-se presente em cercárias e preferencialmente nos músculos de vermes adulto . O gene produz três transcritos variantes, decorrentes de edição alternativa do gene SER . SER parece ser ativada por ligantes de EGF de vertebrados, além de ativar a via de sinalização ERK, indicando uma conservação da função do EGFR em *Schistosoma* .

TK5 é uma tirosina quinase ortóloga a família Src. Tem suas extremidades N-terminal e especialmente a C-terminal alongadas, possui domínio SH4, seguido de região única pouco conservada, domínio SH3, domínio SH2 e domínio catalítico com atividade tirosina quinase. Foi a primeira tirosina da sub-família Fyn identificada em invertebrados .

TK3 é uma outra tirosina quinase ortóloga a família Src. É um gene de cópia única e codifica para uma proteína de 71 kDa expressa em vermes adultos de ambos os sexos, predominantemente nos órgãos reprodutivos. Sua atividade enzimática foi comprovada experimentalmente em sistema de cultura de células eucarióticas, sendo capaz de fosforilar p130Cas, que está relacionada à via de adesão focal e na organização do citoesqueleto .

TK4 é uma tirosina quinase ortóloga a família Syk. Contém dois domínios SH2 e um domínio tirosina quinase. Está presente na fase larvária e em vermes adultos, machos e fêmeas. A presença de TK4 em oocistos e espermatócitos sugerem que esta proteína possua alguma função no desenvolvimento das células germinativas. Não obstante, essa informação é inesperada, por não haver outra descrição de tirosina do tipo Syk envolvidas em diferenciação de gônadas .

SmFes é uma tirosina quinase previamente identificada pelo nosso grupo. O gene SmFes apresentou-se diferencialmente transcrito em ovos maduros quando comparado a ovos imaturos por RAP-PCR, uma técnica que utiliza iniciadores aleatórios a baixa estringência. O sequenciamento do gene SmFes ocorreu pela identificação das extremidades 5' e 3' utilizando a técnica de RACE-PCR. A caracterização de SmFes será descrita ao longo deste trabalho.

## **II JUSTIFICATIVA**

Os parasitos platelmintos do gênero *Schistosoma* são causadores da esquistossomose, uma doença crônica e debilitante. A esquistossomose é a segunda doença parasitária em relação a problemas de saúde pública no mundo (WHO, 2002). A extrema pobreza, as más condições de saneamento básico e a saúde pública inadequada são condições diárias em que vivem milhões de pessoas, facilitando o risco destas comunidades de contraírem a esquistossomose. O comprometimento da doença no desenvolvimento dos jovens e na produtividade dos adultos tem efeito na economia doméstica e da coletividade, sendo este ainda mais negativo para comunidades com nível sócio-econômico baixo, o que justifica os esforços para o melhor conhecimento da doença e para a busca de soluções adequadas .

O controle da esquistossomose é realizado, hoje, basicamente com o uso do praziquantel. A sua utilização resultou em uma marcada diminuição da morbidade das populações. Porém, existem relatos de resistência/tolerância a droga em zonas endêmicas . Pouquíssimas drogas foram desenvolvidas contra agentes causadores de doenças tropicais, apenas 1% entre 1975-99. Este fato deve-se principalmente ao alto custo de desenvolvimento de drogas contra o baixo poder aquisitivo de países em desenvolvimento . É, portanto, essencial que esforços sejam feitos para que novos alvos de drogas sejam identificados.

A proliferação, diferenciação e interação com o meio ambiente são essenciais para o complexo ciclo de vida do *Schistosoma*. Estes fenômenos requerem comunicações moleculares, sendo estas interações mediadas por processo de transdução de sinal. Por esta razão a identificação de moléculas sinalizadoras e a elucidação da transdução de sinal tornaram-se ferramentas na compreensão do desenvolvimento do parasito, bem como no desenvolvimento de novas drogas que interfiram na sinalização celular, impedindo que os processos essenciais para o desenvolvimento e sobrevivência do parasito se completem.

O conhecimento das proteínas quinase reguladoras das funções celulares forneceram novas estratégias de desenvolvimento de agentes quimioterapêuticos . Existem sistemas, principalmente em câncer, que inibidores de TKs vêm sendo utilizados, tendo vários já atingido o estágio de estudo clínico em população humana .

A identificação de moléculas sinalizadoras e a elucidação de processos de transdução de sinal são alguns dos maiores desafios no entendimento da biologia do *S. mansoni*. PTKs podem ser consideradas alvos em potencial na geração de agentes anti-parasitos. Deste modo, tendo em vista a necessidade de se desenvolver novos alvos terapêuticos e a viabilidade destes alvos serem proteínas quinase, o projeto proposto justifica-se.

### **III OBJETIVOS**

### **III.1 OBJETIVO GERAL**

Caracterizar uma nova proteína tirosina quinase de *S. mansoni*, denominada SmFes.

### **III.2 OBJETIVOS ESPECÍFICOS**

1. Analisar e caracterizar a seqüência nucleotídica e protéica de SmFes *in silico*;
2. Realizar análise filogenética da proteína SmFes;
3. Identificar e caracterizar a seqüência e estrutura genômica do gene SmFes;
4. Identificar o número de cópias do gene de SmFes;
5. Identificar o nível de expressão e transcrição de SmFes nos diferentes estágios de desenvolvimento do parasito;
6. Identificar possíveis polimorfismos na seqüência de SmFes;
7. Identificar genes ortólogos de SmFes em diferentes espécies de *Schistosoma*;
8. Selecionar possíveis participantes da via de sinalização de SmFes *in silico* pela identificação de cDNAs seqüenciados;
9. Expressar e purificar proteína recombinante de SmFes.

## **IV MATERIAIS E MÉTODOS**

## **IV.1 COLETA DE PARASITOS**

Neste trabalho foram utilizados parasitos da espécie *S. mansoni* cepa LE, sendo que os vermes, cercárias, ovos e miracídios foram provenientes de infecções rotineiras e gentilmente cedidos pelo Moluscário do CPqRR/FIOCRUZ. Os vermes adultos de outras espécies do gênero *Schistosoma* foram gentilmente cedidos pelo Dr. David A. Johnston (The Natural History Museum, London, England, UK). Os parasitos fornecidos foram das espécies: *S. margrebowiei*, *S. bovis*, *S. intercalatum*, *S. haematobium*, *S. japonicum*, *S. rodhaini*, *S. curassoni* e *S. mattheei*.

### **IV.1.1 Vermes Adultos**

Os vermes provenientes de perfusão de camundongos infectados, realizada segundo Pellegrino e Siqueira, foram lavados em solução salina a 1,7% e mantidos a 4°C para a separação dos casais. Vermes individuais, grupo de casais, de vermes machos e de vermes fêmeas foram armazenados em tubos de 1,5 ml a -70°C para posterior utilização.

### **IV.1.2 Ovos**

Os ovos foram obtidos através do tritamento do fígado de camundongos infectados que foram em seguida, filtrados em peneiras de 190, 120 e 19  $\mu\text{m}^2$  sucessivamente. Os ovos foram lavados em solução salina a 1,7% e centrifugados a 2.000 r.p.m em centrífuga Sorvall RT7 (DuPont) por 1 min. O sobrenadante foi descartado e os ovos armazenados em tubo de 1,5 ml a -70°C para posterior utilização.

### **IV.1.3 Miracídios**

Os miracídios foram obtidos pela eclosão dos ovos em água milliQ em balão volumétrico escuro apenas com a extremidade superior sob foco de luz. Os miracídios, por fototropismo positivo e geotropismo negativo, migraram para o topo do balão volumétrico e foram coletados em tubos de 50 ml com auxílio de uma pipeta Pasteur. Os miracídios foram mantidos em gelo por aproximadamente 30 min para que seus movimentos diminuíssem e se depositassem ao fundo do tubo. O sobrenadante foi descartado com a utilização de bomba de vácuo e os miracídios armazenados em tubo de 1,5 ml a -70°C para posterior utilização.

### **IV.1.4 Esporocistos**

Os esporocistos foram obtidos através do cultivo de miracídios (ver IV.1.3) em 15 ml de meio RPMI 1640 (Gibco) suplementado com 5% de soro bovino (Gibco), 100 U/ml de penicilina (Gibco) e 100  $\mu\text{l/ml}$  de estreptomicina (Gibco), e incubados em estufa BODMOD



347CD (Fanem) a 28°C, com ambiente de 5% de CO<sub>2</sub> por 24 hr. Os esporocistos sedimentados foram coletados e lavados em RPMI não suplementado. O sobrenadante foi descartado e os esporocistos armazenados em tubo de 1,5 ml a -70°C para posterior utilização.

#### **IV.1.5Cercárias**

As cercárias liberadas em água pelo hospedeiro intermediário foram mantidas no gelo por 2 hr e, em seguida, centrifugadas em tubos de 50 ml a 2.000 r.p.m a 4°C por 5 min em centrífuga Sorvall RT7. O sobrenadante foi descartado e o sedimento armazenado em tubos de 1,5 ml a -70°C para posterior utilização.

#### **IV.1.6Esquistossômulos**

Os esquistossômulos equivalentes ao estágio pulmonar foram obtidos através do cultivo de cercárias (ver IV.1.5) por 7 dias. Cercárias foram ressuspensas em 6 ml de meio ELAC {9 ml solução [1,12M NaCl, 60mM KCl, 8mM MgSO<sub>4</sub>, 12mM NaH<sub>2</sub>PO<sub>4</sub>], 1 ml 0,2 M CaCl<sub>2</sub>, 0,178 g glicose, 0,5 g hidrolisado de lactoalbumina, 2,0 mg fenol vermelho, 400 µg/ml PS (penicilina e estreptomicina), 2,0 ml 1M HEPES, pH=7,4} e agitadas a velocidade máxima por aproximadamente 3 min para que sua cauda se separasse do corpo. Em capela de fluxo laminar cada 1 ml de cercárias em meio ELAC foi transferido para novo tubo. As cercárias foram incubadas com 5 ml de meio ELAC por 5 min por 5 vezes, sendo dois tubos agrupados a cada lavagem. O sobrenadante foi retirado com auxílio de pipeta Pasteur e os parasitos foram ressuspensos em 12 ml de meio M169 (10g talco MEM, 0,1% glicose, 5x10<sup>-7</sup>M hipoxantina, 1x10<sup>-6</sup>M serotonina, 1x10<sup>-6</sup>M hidrocortisona, 2x10<sup>-7</sup>M triiodotironina, 0,1% hidrolisado de lactoalbumina, 0,5% vitaminas MEM, 5% meio Schneiders estéril, 10mM HEPES, 26mM NaHCO<sub>3</sub>) suplementado com 5% de soro bovino (Gibco), 1% de gentamicina (Sigma) e 1% de glutamicina (Sigma). Os parasitos foram distribuídos em placa de cultura de 24 poços e mantidos em estufa (Forma Scientific) a 37°C, com ambiente de 5% de CO<sub>2</sub> e 95% de umidade por 7 dias, com troca de 300 µl de meio M169 suplementado a cada dia de cultura. No sétimo dia de cultivo, todos os poços foram agrupados, o sobrenadante foi descartado com auxílio de bomba de vácuo e os esquistossômulos foram armazenados em tubos de 1,5 ml a -70° para posterior utilização.

## **IV.2 PREPARO DE CÉLULAS COMPETENTES TRATADAS COM CLORETO DE CÁLCIO**

Pré-inóculo de colônia de bactéria *E. coli* cepa BL21 ou XL1-blue foi realizado em 5 ml de meio LB líquido (10 mg NaCl, 5 mg extrato de levedura, 10 mg peptona, H<sub>2</sub>O q.s.p 1 L, pH=7,0) com ampicilina (100 µg/ml) e incubado em agitador orbital (Forma scientific), a 300 r.p.m a 37°C por 16 hr. 1 ml do pré-inóculo foi transferido para 100 ml de meio líquido LB que foi incubado a 37°C sob agitação constante de 300 r.p.m, até atingir O.D. 600nm entre 0,4 e 0,6. Em seguida, o inóculo foi resfriado no gelo por 15 min e então transferido para tubos de 50 ml. Os tubos foram centrifugados a 7.100 r.p.m em centrífuga GS-6R (Beckman) por 7 min a 4°C e o sobrenadante foi descartado. Os sedimentos de cada tubo foram ressuspensos em 25 ml de solução de cloreto de cálcio (100mM CaCl<sub>2</sub>, 10mM Hepes, pH=7,0), gelada e estéril e mantidos no gelo por 20 min. A suspensão foi então submetida a recentrifugação e o sobrenadante foi descartado. Os sedimentos foram ressuspensos em 1 ml de solução de cloreto de cálcio-glicerina (100mM CaCl<sub>2</sub>, 10mM Hepes, 10% glicerol, pH=7,0), gelado e estéril. As células foram mantidas por 1 hr no gelo antes de serem usadas ou armazenadas a -70°C.

## **IV.3 TRANSFORMAÇÃO QUÍMICA**

As transformações dos plasmídeos foram realizadas em células de *E. coli* cepa XL1-blue ou BL21 quimicamente competentes tratadas com cloreto de cálcio (ver IV.2). 65 ng de plasmídeo foram adicionados à alíquota de 100 µl de célula competente. A mistura foi incubada em gelo por 30 min seguido de choque térmico a 42°C por 40 seg, o tubo foi então imediatamente colocado em gelo. 250 µl de meio SOC (100 µl 1M MgSO<sub>4</sub>, 1 ml 0,1M MgCl<sub>2</sub>, 200 µl 1M glicose, meio SOB q.s.p 10 ml) foram adicionados ao tubo que foi, em seguida, incubado a 37°C por 1 hr. A cultura foi plaqueada em meio LB-agar (10 mg NaCl, 5 mg extrato de levedura, 10 mg peptona, 1,5% ágar, H<sub>2</sub>O q.s.p 1L, pH=7,0) com ampicilina (100 µg/ml) ou kanamicina (50 µg/ml) e incubada a 37°C por 16 hr.

## **IV.4 EXTRAÇÃO DE PLASMÍDEO**

Para a extração do plasmídeo foi utilizado o kit de extração plasmidial Quiaprep *spin miniprep* (QIAGEN). Colônias de bactérias contendo o plasmídeo de interesse foram cultivadas em 5 ml de meio LB com ampicilina (100 µg/ml) ou kanamicina (50 µg/ml) em agitador orbital a 300 r.p.m, a 37°C por 16 hr. A cultura foi centrifugada em micro-centrífuga a 13.000 r.p.m por 5 min e o sedimento ressuspensado em 250 µl de tampão P1 de ressuspensão (50mM Tris-HCl, pH=8,0; 10mM EDTA, 100 µg/ml RNase A). Em seguida,

250 µl do tampão de lise P2 (200mM NaOH, 1% SDS) foram acrescentados ao tubo, que foi invertido por 6 vezes. 350 µl do tampão N3 de neutralização (fórmula proprietária do fabricante) foram acrescentados e, após a mistura das soluções, o tubo foi centrifugado por 10 min. O sobrenadante foi aplicado nas micro-colunas, que foram centrifugadas por 2 min. A coluna foi equilibrada com 750 µl de tampão PB de equilíbrio (fórmula proprietária do fabricante) e centrifugada por 2 min. A coluna foi em seguida lavada com 750 µl de tampão PE de lavagem (fórmula proprietária do fabricante) e novamente centrifugada por 2 min. O DNA foi eluído com 50 µl de tampão EB de eluição (10mM Tris-HCl, pH=8,5) pré-aquecido a 60°C. A quantificação foi realizada no BioFotômetro (Eppendorf) e o plasmídeo armazenado a -20°C para posterior utilização.

## **IV.5 EXTRAÇÃO DE DNA GENÔMICO**

### **IV.5.1 Extração DNA Genômico Kit Wizard**

A extração de DNA de *S. mansoni* foi realizada de acordo com o procedimento do kit de purificação de DNA genômico, Wizard (Promega). Um grupo de vermes foi homogeneizado com auxílio de pistilo em 600 µl de tampão de lise nucleica (fórmula proprietária do fabricante) e incubado a 65°C por 30 min. 3 µl de solução de RNase A (4 mg/ml) foram acrescentados ao lisado, invertendo-se o tubo por 25 vezes. Os tubos foram incubados a 37°C por 30 min e resfriados por 5 min à T.A. 200 µl de solução de precipitação de proteína (fórmula proprietária do fabricante) foram adicionados ao tubo, que foi agitado por 20 seg e, em seguida, centrifugado a 13.000 r.p.m por 3 min. O sobrenadante foi transferido para um novo tubo. 600 µl de isopropanol foram acrescentado ao tubo, que foi invertido por 20 min e, em seguida, centrifugado a 10.000 r.p.m por 2 min. O sobrenadante foi então descartado e o sedimento lavado duas vezes, adicionando-se 600 µl de etanol 70% ao tubo, que foi invertido por 5 min e, em seguida, centrifugado a 10.000 r.p.m por 1 min. O etanol foi em seguida descartado. O sedimento de DNA foi ressuscitado em 50 µl de solução de reidratação de DNA (10mM Tris-HCl, pH=7,4; 1mM EDTA, pH=8,0) e o tubo foi incubado a 65°C por 1 hr. A amostra foi quantificada no BioFotometro e armazenada a -20°C para posterior utilização.

### **IV.5.2 Extração DNA genômico**

Para cada espécie de *Schistosoma* (ver IV.1) foram separados grupos de vermes em tubos de 1,5 ml. Os tubos foram mergulhados em nitrogênio líquido por 10 min, depois por 5 min em água fervendo e novamente por 10 min em nitrogênio líquido. 50 µl de solução

Proteinase K (60 µg/ml Proteinase K, 1% SDS) foram adicionadas ao tubo que foram misturados e incubados a 37°C por 16 hr. A solução contendo os vermes foi precipitada com a adição de 1/10 do volume de 3M NaAc e 3 vezes o volume de etanol absoluto a -20°C por 16 hr. Os tubos foram centrifugado a 10.000 r.p.m e o sobrenadante descartado. 200 µl de etanol 70% foram adicionados ao sedimento de DNA, o tubo foi centrifugado a 13.000 r.p.m e o sobrenadante descartado, repetindo-se esse passo por mais uma vez. O sedimento foi então ressuspendido em 30 µl TE (10mM Tris, 0.1mM EDTA). Os DNAs foram diluídos 1:10 em água esterilizada e armazenados a -20°C para posterior utilização.

## **IV.6 PREPARAÇÃO DE EXTRATO DE PROTEÍNAS**

Proteínas dos parasitos foram extraídas diretamente em tampão de amostra. 25 µg de parasitos foram acrescentados a 15 µl de tampão de amostra (62,5mM Tris-base, pH=6,8; 3% SDS; 3% sacarose; 5% mercaptoetanol; azul de bromofenol). As amostras foram mantidas em água fervente por 3 min e, em seguida, foram agitadas por 1 min.

## **IV.7 EXTRAÇÃO DE RNA**

### **IV.7.1 Rneasy**

Para as extrações de RNA de verme individual ou de grupos de vermes de até 30 mg foi utilizado o Kit de mini extração de RNA total para tecido animal, Rneasy (QIAGEN) de acordo com instruções do fabricante. O tecido foi rompido em 350 µl de tampão RLT de lise acrescido de 1% de β-Mercaptoetanol (MERK) com auxílio de pistilos. O lisado foi centrifugado a 13.000 r.p.m por 3 min. Ao sobrenadante foram acrescentados 350 µl de etanol 70% que foram, então, aplicados na coluna Rneasy. A coluna foi centrifugada a 13.000 r.p.m por 15 seg. 700 µl de tampão RW1 de lavagem (fórmula proprietária do fabricante) foram aplicados na coluna, que foi centrifugada a 13.000 r.p.m por 15 seg. A coluna foi transferida para novo tubo coletor e 500 µl de tampão RPE de lavagem (fórmula proprietária do fabricante) (acrescido de etanol 100%) foram aplicados na coluna, que foi centrifugada a 13.000 r.p.m por 15 seg. Novamente, 500 µl de tampão de lavagem RPE (fórmula proprietária do fabricante) foram aplicados na coluna e centrifugados a 13.000 r.p.m por 2 min. O RNA foi eluído em 50 µl de água livre de RNase previamente tratada com 1% DEPC (Gibco) e a coluna foi centrifugada a 13.000 r.p.m por 1 min. Os RNAs foram quantificados no BioFotômetro (Eppendorf). A pureza foi determinada pela razão da absorbância 260nm/280nm e considerada boa entre 1,8 e 2,0. As amostras foram armazenadas a -70°C para posterior utilização.

### **IV.7.2 Tri-Reagente**

Para as extrações de RNA de grupos de parasitos de peso superior a 30 mg das diversas fases de desenvolvimento foram realizados procedimentos de acordo com o protocolo de utilização do Tri-Reagente (Sigma). Os parasitos foram homogeneizados em Tri Reagente com auxílio de pistilo, sendo 1 ml de reagente utilizado para cada 50 a 100 mg de tecido. As amostras foram deixadas por 5 min à T.A. 0,2 ml de clorofórmio foram adicionados às amostras, agitando-as vigorosamente por 15 seg e, em seguida, mantendo-as por 15 min à T.A com posterior centrifugação das amostras a 15.000 r.p.m a 4°C por 15 min. A fase aquosa, na qual se encontra o RNA, foi transferida para tubo novo. 0,5 ml de isopropanol foram misturados à fase aquosa e o tubo incubado por 10 min à T.A. O tubo foi então centrifugado a 15.000 r.p.m a 4°C por 1 min. O sobrenadante foi removido e o sedimento lavado com 1 ml de etanol 75% incubando-o em seguida por 5 min. O tubo foi centrifugado a 15.000 r.p.m a 4°C por 1 min. O excesso de etanol foi retirado e as amostras foram mantidas a 37°C por 10 min. O sedimento foi ressuspendido em H<sub>2</sub>O DEPC. O RNA foi quantificado no BioFotômetro e armazenado a -70°C para posterior utilização.

### **IV.7.3 Ultracentrifugação em Cloreto de Césio**

Extração de RNA de grupos de miracídios, esporocistos, cercárias, e vermes adultos e fêmeas foram realizados segundo a técnica de Chirgwin . Os tecidos foram homogeneizados em solução de guanidina (4M tiocinato de guanidina; 1mM EDTA, pH=7,4; 25 mM acetato de sódio, pH=5,5; 5% mercaptoetanol; 2% lauril sarcosinado) e centrifugados em ultracentrifuga L80 (Beckman, rotor SW55ti) a 6.000 t/min por 30 min. Ao sobrenadante foi adicionado a solução de CsCl (densidade de 1,77) e a amostra foi centrifugada a 8.000 t/min a 18°C por 20 hr. O sobrenadante foi desprezado e o sedimento ressuspendido com água DEPC e armazenado a -70°C para posterior utilização.

## **IV.8 SÍNTESE DE CDNA**

### **IV.8.1 SuperScript**

Para a síntese de cDNA foi utilizado o Kit de transcriptase reversa SuperScript II Rnase H (Invitrogen). Com a finalidade de se eliminar contaminação com DNA genômico, aproximadamente 1 µg de RNA (ver IV.7.1) foi tratado com 2 µl de DNase-RQ1 (1 U/µl, Promega) e 1 µl de tampão de DNaseI (10X, Invitrogen) para um volume final de 10 µl, incubando-se a amostra a 37°C por 20 min e, logo em seguida, a 75°C por 5 min. Aproximadamente 500 ng de RNA tratado foram misturados a 1 µl de oligo(dt) (500 µg/ml) e

1 µl de dNTPs (10mM) em um volume final de 10 µl. A amostra foi incubada a 65°C por 5 min e colocada imediatamente no gelo. Ao tubo foram acrescentados 4 µl de tampão de síntese de primeira fita 5X (250mM Tris-HCl, pH=8,3; 375mM KCl; 15mM MgCl<sub>2</sub>), 2 µl de DTT (0,1mM) e 1 µl de Rnase-out (40 µg/ml, Invitrogen). O tubo foi incubado a 42°C por 2 min e foi, em seguida, acrescentado 1 µl de SuperScript (200 U/µl). A reação foi então incubada a 42°C por 50 min e, em seguida, incubada a 70°C por 15 min. As amostras foram armazenadas a -20°C para posterior utilização.

#### **IV.8.2ThermoScript**

Para a síntese de cDNA foi utilizado o Kit de transcriptase reversa ThermoScript RT-PCR System (Invitrogen). Com a finalidade de se eliminar contaminação com DNA genômico, os RNA (ver IV.7.3) foram tratados com DNase como descrito em IV.8.1. Aproximadamente 5 µg de RNA tratado com DNase foram misturados a 1 µl de oligo(dt)<sub>20</sub> (500 µg/ml) em um volume final de 10 µl. A amostra foi incubada a 65°C por 5 min e colocada imediatamente no gelo. Ao tubo foram acrescentados 2 µl de dNTPs (10mM), 4 µl de tampão de síntese de primeira fita 5X (250mM Tris-acetato, pH=8,4; 375mM acetato de potássio; 40mM acetato de magnésio), 1 µl de DTT (0,1mM), 1 µl de Rnase-out (40 U/µl), 1 µl H<sub>2</sub>O DEPC, 1 µl de ThermoScript RT (15 U/µl). A amostra foi incubada a 50°C por 60 min, e seguida de incubação a 85°C por 5 min. 1 µl de RNase H (2 U/µl) foi acrescentado e a reação foi incubada a 37°C por 20 min. A amostra foi diluída 10 vezes em H<sub>2</sub>O DEPC e armazenada a -70°C para posterior utilização.

### **IV.9 MÉTODOS DE ELETROFORESE**

#### **IV.9.1Eletroforese de DNA em Gel de Poliacrilamida**

O gel de poliacrilamida foi preparado a 8% utilizando-se o aparelho Mini-Protean (BIO-RAD). Para o gel foram misturados 10 ml de acrilamida [2,67 ml Bis-acrilamida 30%, 2 ml TBE 5X (0,45 M Tris-Borate, 0,01 M EDTA, pH=8,3), 5,33 ml H<sub>2</sub>O bidestilada], 125 µl de APS e 12,5 µl de TEMED. A eletroforese do gel foi realizada a 100V em fonte 2301 Macrodrive 1 (Pharmacia LKB) em tampão de corrida TBE 1X (89mM Tris-borato, 2mM EDTA, pH=8,0). O gel foi corado em brometo de etídeo (0,5 µg/ml em TBE 1X, Sigma) e a imagem digitalizada com o aparelho Eagle Eye II (Stratagene). Alternativamente o gel foi corado por prata em solução corante (0,2% nitrato de prata), revelado em solução de

revelação (0,75M NaOH) e fixado em solução fixadora (10% álcool etílico e 0,5% álcool acético).

#### **IV.9.2 Eletroforese de DNA em Gel de Agarose**

O gel de agarose foi preparado entre 0,7 e 1,5% (peso/volume). A agarose (Promega) foi fundida em tampão de corrida TBE 1X por aquecimento em forno de microondas. A solução foi resfriada a 60°C e aplicada à cama de eletroforese BRL Horizontal Gel Electrophoresis Horizon 11.14 (Gibco - Life Technologies) até sua polimerização. A eletroforese do gel foi realizada a 70V em fonte GPS200/400 (Pharmacia LKB) em tampão de corrida TBE 1X. O gel foi corado em brometo de etídeo e a imagem digitalizada com o aparelho Eagle Eye II.

#### **IV.9.3 Eletroforese de Proteínas em Gel de Poliacrilamida-SDS**

Cuba de eletroforese Mini-Protean foi montada na vertical para que as fases de separação e concentração fossem formadas. Gel separador foi preparado a 12%: 6 ml de acrilamida 40% (40 g acrilamida, 1,33 g bisacrilamida H<sub>2</sub>O q.s.p 100 ml, pH=7,0), 5 ml de tampão separador 4X (181,7 g tris-hidroxi-metilaminoetano, 4 g SDS, H<sub>2</sub>O q.sp. 1L, pH=8,8), 134 µl APS 10%, 13,4 µl TEMED, H<sub>2</sub>O q.s.p 20 ml. O gel separador foi aplicado até 1 cm de distância do pente, sendo colocado em seguida uma camada de álcool isobutílico para que o gel ficasse reto na extremidade superior. Após a polimerização, o álcool foi descartado e o gel concentrador preparado com: 2,5 ml de tampão concentrador 4X (78,8 g Tris-Cl, 4 g SDS, H<sub>2</sub>O q.s.p 1 L, pH=6,8), 1,3 ml de acrilamida 40%, 50 µl APS 10%, 25 µl de TEMED, H<sub>2</sub>O q.s.p. 10 ml. O pente foi colocado para que canaletas fossem formadas. O gel foi submetido à eletroforese em tampão de corrida SDS 1X (3 g tris, 14,4 g glicina, 0,5 g SDS H<sub>2</sub>O q.s.p 1 L, pH=8,8) a 50V iniciais e 100V no restante da corrida. O gel foi então retirado, corado em solução de Coomassie (50% metanol, 0,05% azul Coomassie R Brilhante, 10% ácido acético, 40% H<sub>2</sub>O MilliQ) por 30 min e descorado em 10% ácido acético.

#### **IV.9.4 Eletroforese de Proteína em Gel de Poliacrilamida Bis-Tris Nu-Page**

O gel pronto para uso, Nu-PAGE 4-12% Bis-Tris gel, 1,5 mm e 10 canaletas (Invitrogen), foi utilizado na eletroforese de proteínas para *western-blot* segundo orientações do fabricante. A eletroforese foi realizada em tampão de corrida NuPAGE-MES (Invitrogen) a 200V por 2 hr.

#### **IV.9.5 Eletroforese de RNA em Gel de Agarose Desnaturante**

O gel de RNA foi preparado a 1,2%. Todo o material e soluções foram cuidadosamente preparados a fim de eliminar as RNases, preparando-se as soluções com água DEPC, lavando-se a cuba com detergente e álcool etílico, utilizando-se ponteiras com barreiras e plásticos livres de RNase. 1 g de agarose (Promega) foi fundida em 49 ml de água DEPC por aquecimento em forno de microondas. A solução de agarose foi resfriada a 60°C e foram adicionados 15,4 ml de tampão MOPS 5X (0,2M MOPS, 5mM EDTA, 25mM acetato de sódio) e 14 ml de formaldeído. O gel foi preparado em aparelho de eletroforese BRL Horizontal Gel Electrophoresis Horizon 11.14. A eletroforese foi realizada em tampão MOPS 1X a 3-4V/cm. O gel foi corado em brometo de etídeo e a imagem digitalizada com o aparelho Eagle Eye II.

#### **IV.10 EXTRAÇÃO DE DNA DE GEL DE AGAROSE**

Para a extração de banda, contendo o DNA de interesse, do gel de agarose (ver IV.9.2) foi utilizando Kit de extração de gel QIAquick (QIAGEN), segundo as orientações do fabricante. A banda foi cortada do gel de agarose e transferida para um tubo novo. 3 volumes ( $\mu\text{l}/\text{mg}$  de gel) de tampão QG de solubilização (fórmula proprietária do fabricante) e ligação foram acrescentados ao tubo, que foi então incubado a 50°C por 10 min e misturado a cada 2 min. 1 volume ( $\mu\text{l}/\text{mg}$  de gel) de isopropanol foi acrescentado à amostra. A amostra foi aplicada na coluna de purificação e centrifugada a 13.000 r.p.m por 1 min. Em seguida, a coluna foi lavada com 0,7 ml de tampão PE de lavagem (fórmula proprietária do fabricante) e centrifugada a 13.000 r.p.m por 1 min. O DNA foi eluído em 20  $\mu\text{l}$  de H<sub>2</sub>O MilliQ e armazenado a -20°C para posterior utilização.

#### **IV.11 SEQUENCIAMENTO DE DNA**

O sequenciamento foi realizado com o kit de sequenciamento *DYEnamic ET Dye terminator* para análise de DNA em sistema MegaBACE (Amershan Bioscience), segundo orientações do fabricante. Na placa de PCR para sequenciamento foram colocados 2  $\mu\text{l}$  do DNA plasmidial recombinante (100 ng/ $\mu\text{l}$ ) ou 5  $\mu\text{l}$  de DNA amplificado, 4  $\mu\text{l}$  pré-mix de sequenciamento *DYEnamic terminator*, 1  $\mu\text{l}$  de iniciador (3 $\mu\text{M}$ ) (Quadro 4), para volume total de 10  $\mu\text{l}$ . Para a reação de sequenciamento foi utilizado o programa: 95°C por 20 seg (desnaturação), 50°C por 15 seg (anelamento) e 60°C por 1 min (extensão) por 25 ciclos consecutivos. A placa foi centrifugada em centrífuga 5804R (Eppendorf) e, em seguida, foram adicionados 1  $\mu\text{l}$  de 7,5M acetato de amônia e 27  $\mu\text{l}$  de etanol 100%. Após misturar os reagentes por inversão, a placa foi centrifugada a 800 r.p.m por 1 min, incubando-a por 10



min a T.A e ao abrigo de luz. A placa foi centrifugada a 3.700 r.p.m por 45 min e, em seguida, o sobrenadante foi descartado. 100 µl de etanol 70% foram adicionados à placa centrifugado-a a 3.700 r.p.m por 10 min. O sobrenadante foi descartado e a placa foi centrifugada de cabeça para baixo a 800 r.p.m por 1 min. Após 10 min, a reação de sequenciamento foi ressuspensa em 10 µl de tampão de amostra do kit (70% formamida, 1mM EDTA) e a placa foi centrifugada a 3.700 r.p.m por 2 min. A placa foi então sequenciada no aparelho de sequenciamento de DNA MegaBace. Os cromatogramas das seqüências fornecidas pelo MegaBace foram analisados utilizando-se os programas Phred, Phrap e Consed, verificando-se a qualidade do sequenciamento.

## **IV.12 AMPLIFICAÇÃO DE DNA POR PCR**

As amplificações dos DNAs foram realizadas por PCR, em diferentes condições, como descrito nos Quadro 2 e 3. As amplificações foram analisadas em gel de poliacrilamida (ver IV.9.1) ou agarose (ver IV.9.2).

## **IV.13 DIGESTÃO DE DNA POR ENZIMA DE RESTRIÇÃO**

As digestões dos DNAs foram realizadas, em diferentes condições, como descrito no Quadro 5. As digestões foram analisadas em gel de poliacrilamida (ver IV.9.1) ou agarose (ver IV.9.2).

## **IV.14 HIBRIDIZAÇÃO DE ÁCIDOS NUCLÉICOS**

### **IV.14.1 Preparo da sonda marcada com Digoxigenina**

Para a realização do *Southern-blot* não radioativo foram preparadas sondas através da amplificação por PCR de fragmentos de SmFes com incorporação de nucleotídeos marcados com digoxigenina (ver IV.12). A amplificação da sonda foi verificada em gel de agarose 1% e a banda foi extraída do gel para posterior utilização (ver IV.10).

### **IV.14.2 Preparo da sonda marcada por Fósforo Radioativo**

Para a realização do *blots* radioativos foram preparadas sondas através da amplificação por PCR de fragmentos do gene SmFes (ver IV.12) e da marcação com fósforo radioativo pelo kit *Nick Translation* (Invitrogen). A amplificação da sonda foi verificada em gel de agarose 1% e a banda foi extraída do gel (ver IV.10). Para a adição do fósforo radioativo, aproximadamente, 150 ng de produto de PCR H<sub>2</sub>O q.s.p 38 µl, foram desnaturados a 96°C por 5 min e imediatamente incubados no gelo. Ao produto de PCR desnaturado foram

Quadro 2: Protocolo de Reação de PCR

PCR	DNA			Iniciadores			Tampão		Magnésio		dNTP		Enzima Taq DNA polimerase				H <sub>2</sub> O q.s.p.	
	P	C	G	[ ] final	For	Rev	PCR	PCR-HF	MgCl <sub>2</sub>	MgSO <sub>4</sub>	Não marcado	Marcado com DIG	T	T.G	T. HF	T. pfu		
Espécies de Schistosoma e Polimorfismo de 15 pb *1			1,5 µl	1µM	TK4	“4”	1X		2,5mM		1mM			0,75 U			10 µl	
					P2for	“B”												
Sonda Radioativa	3 µl			0,5µM	“E”	P2rev	1X		5mM		1mM		2,5 U				20 µl	
Sonda não Radioativa	3 µl			0,5µM	SmGOTK2	SmGOTK1	1X		2,5mM			1X		2,5 U			20 µl	
					TK4	PK1												
Proteína Recombinante	2 µl*2			0,2µM	Fes9for	Fessh2rev	1X			3mM	0,2mM						2,5 U	50 µl
					Ferfor													
					Fes9for	Fespkrev												
					Fesfor													
Polimorfismo 9 pb		1 µl		1µM	SmGOTK2	SMGOTK1	1X		2,5mM		1mM			0,75 U			20 µl	
					TK6													
					TK7													
Estrutura Genômica	I-01			1,5 µl	0,4µM	“A”	“B”	1X			2mM	0,2mM				0,75 U		25 µl
	I-06					“3”	P2rev											
	I-11					SmGOTK4	PK1											
	I-12					ST7	Fessh2rev											
	I-13					TK5T	TK3											
	I-14					TK9	“C”											
	I-16					PK10	“C”											
	I-17																	

P- DNA plasmidial contendo toda a sequência codificante do gene SmFes (1,3 ng/µl); \*2- DNA plasmidial contendo toda a sequência codificante do gene SmFes (50 ng/µl); C- cDNA de vermes individuais e grupos de cercárias, ovos, machos e fêmeas de *S. mansoni* (ver IV.8.1); G- DNA genômico de *Schistosoma* (480 ng/µl) (ver IV.5.2); For - iniciadores forward (Quadro 4); Rev - iniciadores reverse (Quadro 4); T- enzima *Taq* polimerase (5 U/µl, Gibco); T.g- Taq Gold (5 U/µl, Perkin-Elmer); T.pfu - Taq DNA polimerase Pfu Turbo (2.5 U/µl, invitrogen); T.HF- Taq Platinum HF (5 U/µl, invitrogen); DIG- mistura de PCR dig 10X (2mM dCTP, 2mM dGTP, 2mM, dATP, 1,9mM dTTP, 0,1mM Dig-11-dUTP, Boehringer Mannheim); I- introns; \*1- Para o polimorfismo de 15 pb a reação foi realizada também com cDNA no mesmo volume.

Quadro 3: Programa de termociclagem da PCR. A PCR foi realizada em programa de 3 etapas em termociclador (Thermo Hybaid)

PCR	ETAPA I	ETAPA II				ETAPA III
	DESNATURAÇÃO INICIAL	DESNATURAÇÃO	ANELAMENTO	EXTENSÃO	NÚMERO DE CICLOS	EXTENSÃO FINAL
Espécies de Schistosoma	95°C por 10 min	95°C por 1 min	50°C por 1 min	72°C por 1 min	35	72°C por 10 min
Sonda Radioativa	94°C por 1 min	94°C por 1 min	56°C por 1 min	72°C por 3 min	35	72°C por 10 min
Sonda não radioativa	94°C por 1 min	94°C por 1 min	56°C por 1 min	72°C por 3 min	35	72°C por 10 min
SH2 e PK para clonagem	95°C por 2 min	95°C por 30 seg	50°C por 30 seg	72°C por 1 min	35	72°C por 10 min
Polimorfismo 9pb	95°C por 10 min	95°C por 1 min	52°C por 1 min	72°C por 3 min	40	72°C por 10 min
Estrutura Genômica-Introns	94°C por 2 min	94°C por 30 seg	55°C por 30 seg	68°C por 6 min	40	68°C por 10 min

Quadro 4: Lista de Iniciadores

NOMEAÇÃO	SEQUÊNCIA (5' → 3')	DIREÇÃO	POSIÇÃO NA SEQUÊNCIA de cDNA
“3”	ATA AGA GCT CCC TGT ATA TCA TAC CTT C	forward	1180- 1197
“4”	AAT ACC CGG GAC ATA GAT TGT CGT ACA C	reverse	2084- 2101
“A”	ATG GGA AAT CAT ACA GTT AAT G	forward	0001-0023
“B”	AAC ATA TAC TGA CGC GCT TGT G	reverse	746-776
“C”	TTA ATC TTC TCT AGT TCG TCT ATC	reverse	3757-3780
“D”	ATA TAT ATG TTC TTC TCT TCA C	reverse	3309-3330
“E”	AAT CGT AAA ATT ATT GAC AG	forward	868-887
Fes9for	GAG GAG GAG GGA TCC GTG AGT GAG TGT GAT ATG	forward	2433-2442
Fesfor	GAG GA GGA GGG ATC CTT CCT ATA TGT GAT ATG	forward	2426-2442
fespkrev	GAG GAG CTC GAG ATT TAA ATT TGG TGT AAA	reverse	3654-3672
Fesh2rev	GAG GAG CTC GAG TTC CCA ATC AGG TCT TGA	reverse	2848-2865
M13 rev	CAG GAA ACA GCT ATG AC	reverse	Universal
M13-40	GTT TTC CCA GTC ACG AC	forward	Universal
P2for	ACG TGC ATT ATT TCA TAA A	forward	639-667
P2rev	AAT AGGTAC CCA CAG CAC GCC ATT GAG AT	reverse	1380-1398
PK1	GAC AGT TGG ATT ATG TAA AGG	reverse	2275-2295
PK10	CGA GGT CAA ATC CCT ATC AAA TGG	forward	3346-3369
SmFesRTF	TCA ATA CAA AAC GAT ATA CTG ATT CTG TGA	forward	233-262
SmFesRTR	TGG ACA GTT TTT TCA GCT TTG ACT	reverse	363-386
SmGOTK1	ATG CCA AAA GAC AGA TAA AAC CAT ACG	reverse	2662- 2688
SmGOTK2	AAT AAG CCT TAC ATA ATC CAA CTG TCA ATG	forward	2269- 2299
SmGOTK4	GTG TTT TGC CTA GAG CAG AAG TGG A	forward	2471-2495
ST6	TCA TTG GAA TGG AAC TAA TGG TGA TCT A	forward	2848-2875
ST7	ATC AAT GGT AAT CAT AAT GAA ACA ATA T	forward	2599-2626
TK3	ACA TCA TTC CAT TAG CAG C	reverse	3196- 3214
TK4	TCA TAT TCT ACT GAT CCT CC	forward	1942-1961
TK5T	CAC ATG AAA ACT CAA ACT CC	forward	2788-2807
TK6	CCT TTA CAT AAT CCA ACT GTC	forward	2275- 2295
TK9	GGT CAA TTA AAA ATA GCT GAT TTT GGC	forward	3280-3307

Quadro 5: Digestão por enzima de restrição

Digestão	DNA			Enzima de Restrição						Tampão de Reação			Incubação °C/ hr
	PCR *1	G *2	P	<i>Sfc I</i> (2 U/μl, New England Biolabs)	<i>EcoR V</i> (10 U/μl, Promega)	<i>EcoR I</i> (10 U/μl, Gibco)	<i>BamH I</i> (10 U/μl, Gibco)	<i>Hind III</i> (10 U/μl, Gibco)	<i>Xho I</i> (10 U/μl, Gibco)	NEBuffer 4 (10X, New England Biolabs) *3	React 2 (10X, Gibco) *4	React 3 (10X, Gibco) *5	
Polimorfismo 9pb	8 μl			2 U						1X			25/16
Estrutura Genômica	I-06	15 μl			10 U						1X		37/16
	I-12 e I-13-14	15 μl				10 U						1X	37/16
	I-16-17	15 μl					10 U					1X	37/16
Southern- blot	DNA	10 μg						10 U			1X		37/16
					10U						11X	37/16	
	Controle positivo			1 μl *6					2 U		1X		37/16
Proteína recombinante - Clonagem em vetor pGEX			1 μl *7				10 U		10 U			1X	37/2
			1 μl *8										

\*1- ver quadro 2 \*2- G- DNA genômico (ver IV.5.1); \*3- (500mM acetato de potássio, 200mM Tris-acetato, 100mM acetato de magnésio 10mM DTT, pH=7,9), \*4- (50mM Tris-HCl, pH=8,0; 10mM MgCl<sub>2</sub>; 100mM NaCl), \*5- (50mM Tris-HCl, pH=8,0; 10mM MgCl<sub>2</sub>; 50mM NaCl), \*6- P- DNA plasmidial contendo toda a seqüência codificante do gene SmFes (1,3 ng/μl); \*7- P- pCR 2.1 clonado a fragmentos de SmFes (300 ng/μl) (ver IV.17.2), \*8- P- vetor pGEX-4T3 (300 ng/μl)

acrescentados 5 µl de dNTP menos dCTP (0,2mM dATP, dGTP, dTTP; 500mM Tris-HCl, pH=7,8; 50mM MgCl<sub>2</sub>; 100mM 2-mercaptoetanol), 5 µl de mistura de Poli/DnaseI (0,5 U/µl DNA polimerase I; 0,4 mU/µl DNaseI, 50mM Tris-Hcl, pH=7,5; 5mM acetato de magnésio; 0,1mM PMSF; 50% (v/v) glicerol; 100 µg/ml BSA sem nuclease) e 2 µl de α-<sup>32</sup>P dCTP (10 mCi/ml, Amershan). A reação foi incubada a 16°C por 1 hr. 5 µl de solução para parar a reação (0,5M EDTA, pH=8,0) foram adicionados à sonda. A sonda foi purificada em colunas G-50 (Pharmacia), segundo orientação do fabricante. A sonda marcada foi aplicada à coluna, que foi centrifugada a 3.000 r.p.m. O material que passou pela coluna foi coletado e utilizado em seguida.

#### **IV.14.3Preparo do DNA para *Southern-blot***

Duas alíquotas de 10 µg de DNA genômico (ver IV.5.1) foram digeridas (ver IV.13). Plasmídeo contendo toda a região codificante do gene SmFes foi digerido (ver IV.13) para ser usado como controle positivo do *Southern-blot*. Antes da eletroforese as amostras foram aquecidas a 56°C por 2 min, e então, aplicadas com tampão de amostra 2X (2mM EDTA, 10% glicerol, azul de bromofenol) em gel de agarose 0,7% (ver IV.9.2). Após eletroforese o gel foi desnaturado por duas vezes em solução de desnaturação (NaOH 0,5M; NaCl 1,5M) por 15 min à T.A, e em seguida, o gel foi neutralizado por duas vezes em solução de neutralização (Tris-HCl 0,5M, pH=7,5; NaCl 3M) por 15 min em T.A.

#### **IV.14.4Preparo do RNA para *Northern-blot***

Solução de mistura (66,5% formamida deionizada, 20% formaldeído e 13,5% tampão MOPS 5X) a 2,5 vezes o volume do RNA foram adicionados às alíquotas de 20 µg de RNA total extraído de grupo de: casais de vermes, vermes machos, vermes fêmeas, ovos, miracídios, cercárias ou esquistossômulos (ver IV.7.2). As amostras foram centrifugadas por 5 seg, incubadas a 56°C por 15 min e imediatamente levadas ao gelo. As amostras de RNA foram aplicadas com tampão de amostra 10X (50% glicerol, 1mM EDTA pH=8,0, 0,25 azul de bromofenol e 0,25% xileno-cianol FF) em gel de agarose desnaturante 1,2% (ver IV.9.5). A eletroforese das amostras foi finalizada após atingirem uma distância de 2/3 do comprimento do gel. Após eletroforese o gel de foi lavado por 3 vezes em H<sub>2</sub>O MilliQ por 20 min.

#### **IV.14.5Transferência do Ácido Nucléico**

A membrana de nylon carregada positivamente (Boehringer Mannheim) foi cortada do tamanho exato do gel e, em seguida, foi mergulhada em solução de transferência SSC 20X

(3M NaCl, 300mM citrato de sódio, pH=7,0). A estante de transferência foi montada utilizando-se o processo de transferência por capilaridade. A transferência do DNA (ver IV.14.3) ou RNA (ver IV.14.4) contido no gel para a membrana de nylon foi realizada por 16 hr em tampão de transferência SSC 20X. O ácido nucléico foi fixado à membrana em forno de UV *cross-linker* Stratalinker 1800 (Stratagene) a 120.000  $\mu$ Joules.

#### **IV.14.6Hibridização com sonda não radioativa**

A membrana foi incubada em 10 ml de solução de pré-hibridização *High-SDS* (7% SDS; 50% formamida deionizada; 50mM fosfato de sódio, pH=7,0; 2% solução de bloqueio, 0,1% N-laurilsarcosina, SSC 5X) a 42°C por, aproximadamente, 5 hr em forno de hibridização Personal Hyb (Stratagene). As sondas marcadas por digoxigenina (ver IV.14.1) foram desnaturadas por fervura a 5 min e colocadas imediatamente no gelo. Aproximadamente 75 ng de sonda foram acrescentadas a 5 ml de solução *High-SDS* de pré-hibridização. A membrana foi incubada a 42°C para a sonda TK1-TK2 (construída a partir dos iniciadores SmGOTK1 e SmGOTK2) e 37°C para a sonda TK4-PK1 (construída a partir dos iniciadores TK4 e PK1) por 16 hr. A membrana foi então lavada duas vezes a 25°C por 5 min em 10 ml de solução de lavagem 2X (SSC 2X, 0,1% SDS) e duas vezes a 68°C por 15 min em 10 ml de solução de lavagem 0,5X (SSC 0,5X, 0,1% SDS). Logo em seguida, a membrana foi equilibrada em 20 ml de tampão de lavagem (0,3% Tween 20, 0,1M ácido maléico, 0,15M NaCl, pH=7,5) por 1 min. A membrana foi bloqueada com 15 ml de solução bloqueadora 2,5% [2,5% reagente de bloqueio (Roche), 0,1M ácido maléico, 0,15M NaCl, pH=7,5] por 1 hr. 1,5  $\mu$ l de anticorpo secundário anti-Digoxigenina (750 U/ml) foram adicionados à solução bloqueadora, deixando-o atuar por exatos 30 min. A membrana foi lavada duas vezes em 30 ml de tampão de lavagem por 15 min. Para a detecção, 5 gotas de substrato químico CSPD pronto para uso (Boehringer Mannheim) foram espalhadas pela membrana por 5 min ao abrigo de luz. O excesso de CSPD foi retirado e a membrana revelada. A membrana foi exposta por 16 hrs ao filme Kodak X-Omat. O filme foi revelado por 2 min em solução de revelação Kodak, lavando em água e fixando por 2 min em solução fixadora Kodak. Mesma membrana foi reutilizada após a lavagem em água MilliQ seguida de lavagem em solução para retirar a sonda (0,2M NaOH, 1%SDS) a 37°C por 2 vezes de 15 min. Para a reutilização, a membrana foi então lavada em SSC 2X, estando apta para ser pré-hibridizada novamente.

#### **IV.14.7Hibridização por sonda radioativa**

A membrana foi incubada a 50°C em forno de hibridização HB100 (UVP) por aproximadamente 5 hr em garrafa de hibridização contendo 10 ml de solução de pré-

hibridização (7% SDS; 250mM fosfato de sódio, pH=6,8; 1% BSA; 1mM EDTA, H<sub>2</sub>O q.s.p 70 ml). A sonda radioativa (ver IV.14.2) foi desnaturada, por fervura por 5 min e levada imediatamente ao gelo. A sonda foi acrescentada à 15 ml de solução de pré-hibridização nova. A temperatura de hibridização para o *northern-blot* foi de 50°C e para o *Southern-blot* foi de 65°C, por 16 hr. A membrana foi então lavada duas vezes em 75 ml de solução de lavagem (2X SSC, 0,1 % SDS) a 28°C por 10 min para o *northern-blot* e por 20 min para o *Southern-blot*. Para detecção da hibridização a membrana foi exposta ao filme Hiperfilme (Amershan) a -70°C por 20 dias. O filme foi revelado por 2 min em solução de revelação Kodak, lavando em água e fixando por 2 min em solução fixadora Kodak.

#### **IV.15WESTERN BLOT**

5 µl de tampão de amostra Nu-PAGE (Invitrogen) foram adicionados às amostras de proteína (ver IV.6), que foram em seguida fervidas por 5 min. Os extratos protéicos dos parasitos foram corridos em gel de proteína (ver IV.9.4). O gel contendo o extrato protéico foi transferido para membrana invitrolon PVDF (difluoreto de polivinilideno) 2 µm (Invitrogen) em cuba de transferência (Invitrogen) e tampão de transferência (5% tampão de transferência 20X NuPAGE, 12% metanol, H<sub>2</sub>O q.s.p 1 L), segundo orientações do fabricante (Invitrogen). A transferência foi realizada por 1 hr a 30V e 170mA. A membrana foi corada em Ponceau-S (Sigma) para verificar a transferência das proteínas. A membrana foi bloqueada em solução PBST [PBS 1X (14mM NaCl, 2mM fosfato de potássio pH=7,4), 0,05% Tween] e 5% leite em pó desnatado (Molico) por 1 hr. A membrana foi então incubada em solução PBST com 5% de leite em pó contendo o anticorpo anti-SmFes (IgG total não purificado), obtido a partir de imunização de coelho com peptídeo sintético correspondente aos resíduos 583 a 597 da seqüência protéica de SmFes (região sem identidade com proteínas já descritas), SYPNTPITFSRDPC, e conjugado à proteína ovalbumina bovina (New England Peptide), na concentração de 1:1.000 por 1 hr à T.A. O soro pré-imune do coelho foi usado nas mesmas condições como controle negativo do ensaio. A membrana foi lavada por três vezes por 10 min em solução PBST com 1% de leite em pó. O anticorpo secundário anti-coelho conjugado à peroxidase (Sigma) foi adicionado a uma nova solução PBST com 1% de leite em pó na concentração 1:10.000 por 1 hr em T.A. A membrana foi novamente lavada por três vezes em PBST com 1% de leite em pó e três vezes em solução PBST por 10 min cada. Para a detecção, a membrana foi incubada com a solução ECL para *western-blot* (Amersham Biosciences) por 5 min. O excesso de ECL foi retirado e a membrana exposta por 16 hrs ao filme Kodak X-Omat. O filme foi revelado por 2 min em solução de revelação Kodak, lavando em água e fixando por 2 min em solução fixadora Kodak.



## IV.16 PCR QUANTITATIVO

RT-PCR quantitativo foi realizado através do sistema de detecção de sequência ABI PRISM 7.000 (Applied Biosystems) e o kit SYBR® Green PCR Master Mix (Applied Biosystems). Iniciadores de tubulina de *S. mansoni* (número de acesso GenBank: 80214, posição na sequência: for 851-873 e rev 925-904) e de SmFes (SmRTF, SmRTR) foram desenhados pelo programa Primer Express (Applied Biosystems). A amplificação por PCR foi realizada em placa de 96 poços e em triplicata, utilizando-se: 2 µl de cDNA (ver IV.8.2), 3 µl de cada iniciador (300nmol) (Quadro 4), 12,5 µl 2X master mix SYBR® Green, para um volume total de 25 µl. A PCR foi realizada utilizando-se o programa de duas etapas. As amostras foram mantidas na etapa inicial a 95°C por 10 min (desnaturação inicial), na etapa seguinte, 40 ciclos, a 95°C por 15 seg (desnaturação) e 60°C por 1 min (anelamento e extensão). O protocolo de dissociação foi realizado aumentando progressivamente de 60°C a 95°C. A representação gráfica foi realizada segundo os valores de *threshold* (valores Ct) obtidos de verme fêmea, miracídio, esporocisto e cercária deduzidos a partir do valor Ct obtido para os transcritos de verme macho usando o método  $2^{-\Delta\Delta Ct}$  com o nível de tubulina como controle interno.

## IV.17 EXPRESSÃO DE PROTEÍNA RECOMBINANTE

### IV.17.1 Amplificação do Domínio SH2 e PK

Iniciadores foram desenhados para clonagem em vetor pGEX-4T3 (Amersham Bioscience) do domínio SH2 e dos domínios SH2-PK, ambos contendo ou não o inserto de 9 pb. Os iniciadores forward Fes9for (contendo os 9 pb) e Fesfor (sem os 9 pb) foram desenhados com sítio de corte para enzima BamH I, ambos formando pares com os iniciadores reverso Fessh2rev (incluindo o domínio SH2) e Fespkrev (incluindo os domínios SH2 e PK) que foram desenhados com sítio para enzima *Xho I* (Quadro 4). A amplificação foi realizada segundo protocolo descrito nos Quadros 2 e 3. Os produtos das amplificações foram analisados em gel de poliacrilamida corado pela prata (ver IV.9.1).

### IV.17.2 Clonagem em Vetor pCR 2.1

Uma pré-clonagem foi realizada utilizando-se o kit de clonagem TOPO TA (Invitrogen), de acordo com as instruções do fabricante. 6,5 µl da reação de PCR (ver IV.17.1), 1 µl de tampão de PCR (10X), 1 µl de dNTPs (10mM), 0,5 µl de MgCl<sub>2</sub> (50mM) e 1 µl de enzima *Taq* polimerase (5 U/µl, Invitrogen) foram incubados a 70°C por 15 min para a adição adenina extra requerida na clonagem em Topo. A ligação foi realizada misturando-se 6

μl do produto de PCR, 1 μl de sal (1,2M NaCl; 0,06M MgCl<sub>2</sub>), 1 μl de vetor pCR 2.1 Topo e o tubo foi incubando por 5 min à T.A. A transformação foi realizada por choque térmico com seleção de clones por kanamicina (ver IV.3). O plasmídeo foi extraído para reclonagem em pGEX-4T3 (ver IV.4). O sequenciamento dos clones foi realizado com o uso dos iniciadores M13-40, M13rev, TK-6, ST-7, “D”, TK-9, PK-10, TK-3, ST-6 (ver IV.11). A seqüência fornecida pelo seqüenciador MegaBace foi conferida utilizando-se o programa *GeneTools Lite* (Bio Tools) e comparada à seqüência do cDNA de SmFes depositada no GenBank.

### **IV.17.3 Clonagem em vetor pGEX**

O plasmídeo clonado em pCR 2.1 (ver IV.17.2) e o vetor pGEX-4T3 foram digeridos nas mesmas condições segundo o protocolo descrito no Quadro 5 em 10 μl de reação final. Para clonagem em vetor pGEX-4T3, 10 μl do fragmento digerido e 2 μl do vetor pGEX-4T3 digerido foram separados por eletroforese em gel de agarose a 1%, (ver IV.9.2). A banda do vetor linearizado pGEX-4T3 e a banda do inserto foram purificadas juntas em um único tubo (ver IV.10). O DNA dos fragmentos foram concentrados no Speed Vac DNAmimi (Heto) e ressuspendidos em 10 μl de H<sub>2</sub>O MilliQ. Em seguida, uma reação de ligação do vetor com o inserto foi realizada incubando-se 10 μl do DNA eluído (vetor e fragmento), 3 μl de tampão T4 DNA ligase 5X (250mM Tris-HCl, pH=7,6; 50mM MgCl<sub>2</sub>; 5mM ATP; 5mM DTT; 25% (p/v) polietileno glicol-8000) (Gibco) e 2μl de ligase a 16°C por 16 hr. Após a reação de ligação, 2 μl dos produtos ligados foram transformados em células *E. coli* BL21 com seleção por ampicilina (ver IV.3).

### **IV.17.4 Expressão da Proteína Recombinante de SmFes**

Pré-inóculo de bactéria *E. coli* BL21 contendo o clone de SmFes em pGEX-4T3 (ver IV.17.3) foi realizado em 5 ml de LB-Ampicilina (100 μg/ml) com agitação em agitador orbital a 300 r.p.m 37°C por 16 hr. O inóculo foi realizado com 300 μl de pré-inóculo em 10 ml de LB-ampicilina e mantidos sob agitação de 300 r.p.m até atingir O.D. 600nm de 0,6. Ao atingir a O.D. esperada uma alíquota de 3 ml foi retirada do meio não induzido. Os 7 ml restantes foram induzidos com IPTG (Promega) em concentrações finais de 1mM, 5mM e 10mM, retirando-se uma alíquota de 1,5 ml após 4 hr da indução. Todas as alíquotas foram centrifugadas por 20 min a 5.500 r.p.m, o sobrenadante foi descartado e o sedimento foi ressuspendido em 40 μl de tampão de amostra SDS (1 ml tampão concentrador 4X, 1 ml 10% SDS, 250 μl β-Mecaptoetanol, 1 ml glicerina, 52 μl azul de bromofenol, H<sub>2</sub>O q.s.p 5 ml). As amostras foram congeladas a -20°C para posterior análise em gel de eletroforese de proteína a 12% (ver IV.9.3).

## **IV.18 ANÁLISE *IN SILICO* DA SEQUÊNCIA CODIFICANTE**

### **COMPLETA DE SMFES**

A sequência do gene *SmFes* foi traduzida em sua ORF correta utilizando o programa *Translate tool* (<http://us.expasy.org/tools/dna.html>). Neste mesmo sítio da internet, os parâmetros físico-químicos da sequência protéica, como massa molecular e ponto isoelétrico teórico, foram feitos com o uso do programa *ProtParam*. A análise de hidropaticidade foi conduzida pelo programa *ProScale*, segundo Kyte e Dolittle. As buscas por identidade de *SmFes* com outras sequências já previamente depositadas em bancos de dados foram realizadas usando programas da família BLAST, no site do NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>).

#### **IV.18.1 Determinação de Domínios**

A análise dos domínios conservados foi realizada utilizando RPS-BLAST (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) (banco de dados: CDDv2.03) e Interproscan (<http://www.ebi.ac.uk/InterProScan/>) (bancos de dados: PROSITE, PRINTS, Pfam, ProDom, SMART e TIGRFAMs). A probabilidade de *SmFes* adotar uma região de conformação coloidal foi conduzida pelo programa *Coils* versão 2.1 usando a matriz MTIDK ([http://www.ch.embnet.org/software/COILS\\_form.html](http://www.ch.embnet.org/software/COILS_form.html)) comparando-se a sequência de *SmFes* a sequências coloidais previamente conhecidas.

#### **IV.18.2 Determinação da Organização Gênica**

A sequência de cDNA de *SmFes* foi manualmente dividida em sequências de 200 pb e estes fragmentos foram comparados por BLAST com sequências depositadas no banco de dados genômico de *S. mansoni*: TIGR (<http://www.tigr.org/tdb/e2k1/sma1/>) e Sanger ([http://www.sanger.ac.uk/Projects/S\\_mansoni/](http://www.sanger.ac.uk/Projects/S_mansoni/)). Para determinar a localização dos introns na sequência do cDNA de *SmFes*, foram realizados alinhamentos desta sequência às sequências genômicas de *S. mansoni* selecionadas no banco de dados. O alinhamento foi realizado pelo programa *GeneTools Lite* que utiliza o algoritmo de alinhamento XALING. Os exons e introns foram identificados nas sequências genômicas pelo alinhamento com o cDNA.

#### **IV.18.3 Análise Filogenética**

Para a análise filogenética de *SmFes*, sequências protéicas pertencentes à família de proteínas Fes/Fps/Fer foram selecionadas do banco do GenBank e alinhadas pelo ClustalW no BioEdit (Hall, 1999). As sequências usadas e o respectivo número de acesso no banco de dados foram: FBS de Fujinami sarcoma virus (NP\_955606); Fert2 protein de *Mus musculus*,

(AAH58100); Fps/Fes de *Mus musculus* (AAN33122); PTK de *Ephydatia fluvialis* (BAA81721); Fer de *Canis familiaris* (AAF00543); Fer de *Homo sapiens* (NP\_005237); Fer de *Mus musculus* (AAB18988); c-FES de *Homo sapiens* (P07332); Fps85D(dFer) de *Drosophila melanogaster* (P18106); Fer-frk-1 de *Caenorhabditis elegans* (NP\_501818); Fes/Fps de *Felis catus* (TVCTFF) e PTK de *Sycon raphanus* (CAA76605). A análise filogenética foi conduzida com o programa PHYLIP 3.6 com interfase PIE (Phylogenetic Interface Environment) no sítio da internet HGMP ([www.menu.hgmp.mrc.ac.uk](http://www.menu.hgmp.mrc.ac.uk)). As seqüências foram alinhadas inteiras e também parcialmente, contendo apenas o domínio SH2 e PK, utilizando-se o método de parcimônia e *bootstrap* de 1.000 replicas. A árvore foi construída usando o “TreeView 1.6.6” .

#### **IV.18.4 Identificação de Parceiros de SmFes**

A seleção de possíveis parceiros de SmFes foi realizada pela busca de SmAes no *site* do ONSA (<http://cancer.lbi.ic.unicamp.br/schisto6/>), de ESTs no *site* do GeneDB (<http://www.genedb.org/genedb/smanson/>) e de proteínas do PUBMED no *site* do NCBI, anotadas e denominadas como ortólogas as pertencentes a via de sinalização proposta por Greer (2002). A correta anotação destas proteínas não foi verificada e comprovada, exceto pelas proteínas encontradas no PUBMED, fornecendo-se desta maneira apenas evidências de que as proteínas existam.

## **V RESULTADOS**

## V.1 ANÁLISE E CARACTERIZAÇÃO DA SEQUÊNCIA *IN SILICO*

### V.1.1 Sequências de SmFes

O cDNA completo de SmFes apresentou uma ORF de 3.780 pb, (Figura 6) codificando para uma proteína de 1.260 aa (Figura 7). O cDNA de SmFes apresentou um conteúdo CG de 32% correspondendo ao valor do conteúdo CG descrito para outros genes de *S. mansoni*. A massa molecular calculada da proteína foi de 143,5 kDa e seu ponto isoelétrico teórico de 6,60. A análise de hidropaticidade da sequência protéica de SmFes foi realizada segundo Kyte e Doolittle (1982) em janela de 7 e 19 resíduos de aminoácidos. O resultado obtido indicou uma maior concentração de picos de aminoácidos em posição hidrofílica, indicando que SmFes é uma proteína que não apresenta regiões transmembrana, de acordo com o descrito às proteínas da família Fes/Fps/Fer (Figura 8).

### V.1.2 Identidade de SmFes com Proteínas Fes/Fps/Fer Ortólogas

A proteína SmFes apresentou, em análise por BLAST-p contra o banco de dados nr, uma maior identidade com proteínas classificadas como membros da família Fes/Fps/Fer, possuindo uma identidade de 40 a 50% para os aminoácidos entre 777 e 1.216, correspondentes aos domínios SH2 e PK, e em torno de 20% para os aminoácidos entre 91 e 452 da sequência protéica de SmFes. Os organismos identificados por identidade com os valores descritos acima foram proteínas pertencentes à família Fes/Fps/Fer de: *Drosophila melanogaster* (P18106), *Rattus norvegicus* (XP\_217492, XP\_341877), *Canis familiaris* (AAF00543), *Mus musculus* (AAH58100, AAN33122, AAB18988), *Felis catus* (TVCTFF), *Homo sapiens* (P07332, NP\_005237), *Ephydatia fluviatilis* (BAA817213), *Sycon raphanus* (CAA76605), *Gallus gallus* (CAA26155) e também de alguns oncogenes (P00541). Além das proteínas denominadas como Fes/Fps/Fer, foram identificados, por identidade, proteínas de outros organismos ainda não anotadas de: *Drosophila pseudoobscura* (EAL26982), *Anopheles gambiae* (XP\_313423), *Xenopus laevis* (AAH73445) (GenBank, 28/05/2005) (Quadro 6).

### V.1.3 Domínios Protéicos em SmFes

A proteína SmFes apresentou, por busca de domínios de aa, algumas características das proteínas da família Fes/Fps/Fer: três regiões helicoidais, uma assinatura de domínio SH2

# A

gaa -121

cacggttaacggattatgaggttaattcaacaaaatattcagcatactcctcaaggactct -061  
 attttgtaaaccctttaaacactttattattcatgtattttaatttggtatcacaaaat -001  
 ATGggaaatcatacagttaatgaagtcagtaactattttgaacgtggcagtgatgtcgaa 0060  
 gatgatatacaatacttagcatctaataatacacatattacagaaaataatcaatccgctc 0120  
 ttggtatcttctacatcggacacatttttagttcctcctttgctccgcatagtctagga 0180  
 catgataaatcattatcttcttcatccatttcaacacctaagtggacaatatcaataca 0240  
 aaacgatatactgattctgtgatgcttcgaagtgatcataaagagtcttctcaggtaatc 0300  
 gtttcagccatcgatgccagatgaaatctatcaatgaaatgctatcctacttttctaaa 0360  
 ttagtcaaagctgaaaaaactgtccaacaagctttgtccgatttacttgatgatcaaaag 0420  
 gaatcatccactgttgtagtggtgatggtggccagcgttcgaaataccaccgctatgcgt 0480  
 aatttattcaaatcaaattattctcgtggacgctcgtcgtcgtttaaccatgaatatttg 0540  
 aataaaaacttatcgaattcacctactgataatgaattaaataaaacttttacaaaatctt 0600  
 gttatttcaacaaatcatcaagtacgtatgttttcaacacgtgcattatttcataaatta 0660  
 attactgctggacaattagatagtttaataaattgtgaacattgtttaagacgaacttat 0720  
 ttagaatgtgaagaagaattaaatcacaagcgcgtcagtatatgttaagcttaaagca 0780  
 tatgaaaagaaatatttcaaatatattggtcaattatcaaatatacaatcgaaattaaca 0840  
 aatttatgtgataaacgtttaattaataatcgtaaaattattgacagtctaataataat 0900  
 ttagaatatttctactaaaaaattgcatttattacataatgaatattgtttgccattgatc 0960  
 actgctattcattatcaccaatggttatatacacatttacgaccatgtttactgaaaggt 1020  
 gttgaacgtactatgcaattagcatcggagttggttgccatcttctagatattcgggtggt 1080  
 gaaaatattcatcattttcaatcaaattggcttaaagaatttcaatctgaattagattta 0140  
 acaagtttattagaacaaaaatggagtaaaatcttctggacctgtatatcataccttcaat 1200  
 caagcactaaaataataatacaacattaggtgaaatcttctcctggttccttaattgtc 1260  
 aatgatttaactggacaatcattgattgaattaacacaattgcataaaaaataaggaatta 1320  
 ttgcatagagaacggttaatcaactgactaaagaatgtttacaatgggaagaggcgcta 1380  
 tctcaatggcgtgctgtgttttgtaatccaggtcctaataccatggctcatcaaatttattc 1440  
 aatatacatctcatctgggtgttccaccaatagtcaatactccacaagataataattcaaat 1500  
 ggatttcatcatgtagataattcaagtcacttagcgttgctgatataccaagatctttgg 1560  
 catgatgatcgagctcgacctgttaatctttgtgaagtatcatttcgtttagcaatatgt 1620  
 gaatttgaattatgttttgccaatgtcttctgctgattcagaagctcaattagtcaatgta 1680  
 ttaactgatgcacaatgtcgtctcaattggctttcattaaaattacttacaatgattca 1740  
 actacaagttatccaaatactacaccaattacatttccacgtgatccttgtgatattact 1800  
 accaccactactactactaataatgggaatagtaacaataataatcatgaagatatagat 1860  
 atacaaaatgggactgatgataatttcagtcagctttgaaatggctcatgataattcaatt 1920  
 ccattgtgtaatctcagtgattcatatttctactgatcctcctaaattcattgaacaaata 1980  
 tcaaattctccatttgcatatctacatcgaatacatctgatatgtcatctagtatagaa 2040  
 caccagcaacaacaacgtccagaacgtaatcgtctttggcaaagtgtagacaatctatg 2100  
 tcaaaatctcgttcacatttcatcgttaataataataatatacgtcttcgacaagttctata 2160  
 ccaaatacacgtacatatcaatgtcgtcttctgtaaaactattattactactataccaata 2220  
 ttaaataccgtaaatgataataataacaatagtaaatagtaataataaagccttta 2280  
 cataatccaactgtcaatggactaaaacagcaaaaccatttaggtatggatactccattt 2340  
 tatcatcaaaaatagaaaattcaatcatcaaaaatcatattcaatctaatacgagaatctcaa 2400  
 gattcaacatcaattttaactagttttcctatatgtgatattggatatcaataaggaaccg 2460  
 tggtttcatggtgttttgctagagcagaagtggaacgtcttttacagaatcaaggcgat 2520  
 ttccttgttcgacaaaacaagtaaacgttccagtggtcgtgataccataggtcattggaat 2580

ggaactaatggtgatttaataatcaatggtaatcataatgaaacaatattaaataaagatcat 2640  
 aatatggatggaactgttttacgtatggttttatctgtcttttggcatggtcatagacat 2700  
 ttcatctatatggtggaccagagacaggtgaagggtggcatttagaagacggtcatttc 2760  
 tctacaataagggaaactgattgaatatcacatgaaaactcaaactcccgtgacagcgaaa 2820  
 tcaggagcatgtcttgtgacaccaatttcaagacctgattgggaactagacaatcgtgat 2880  
 gtacaactactacaaaaaattggtcagggtaattttgggtgatgtttatcgagggtatata 2940  
 aatggatgtgaagttgcagtgaaaacttgtcgtggtgatgactgcatcggatttacgt 3000  
 aggaaatttctacaagggtgaaacaactgcattaaatttcaatcatccaaatattgtcaaa 3060  
 ttagttggtattgcagttcaatcttatccaattatgattgtaatggagtatgttccaggt 3120  
 ggttctctgttaaatacatttacgtaaatcaaaaaatgcattacctggttatgaaactactt 3180  
 caaatgagtttagatgctgctaataatggaatgatgtatttagaagcgagaaattgtattcat 3240  
 cgtgatttagcagcaagaaattgtttaatttctgatgatggtcaattaaaaatagctgat 3300  
 tttggcatgtcaagagaagaacatatataatgaattaagtataaacgaggtcaaatccct 3360  
 atcaaatggacggctcctgaagcacttctgactggctggttataactattaaatgtgatgta 3420  
 tggctcttatggtgtacttttatgggaaatatttacttttcggtgatgtaccttatcgaaat 3480  
 tggctcaatcaacaaactagggatgatgattgaaatctggttatagattacctgcaccagat 3540  
 ttaatgccagtttgggtgctacattaatgaatcattggtggcatgatgaaccaatgaat 3600  
 cggccaagtttttcaaaaatttctaataatgaaatacaaacctaatgtcaattcatttaca 3660  
 ccaaatttaaatagatctatagacaatgctcaattgaagaaatctgctattgacaaccta 3720  
 tcaacaactccattacatttatccgtttctgttaaagatagacgaactagagaagatTAA 3780  
 tatggatctttttttgttactatgacctatacacatacacatatacacatacacaca 3840  
 tacatacaatagatcatagagatctatatacacatcatgtaattgtgtatattgcaa 3900  
 tctttttctgttttttttttttttggcaagcgctttatttactggtatttactatggca 3960  
 actaacttacagttccgagtgacaaacctttctatttattattatcactaatataattga 4020  
 ttcaattcatggatataatggtatacatagttgtactgtaacctgtatatcatataatcc 4080  
 actacctatgatcattgtaattgtcattatatacatttcattttaataattttatgattt 4140  
 caatttatactacttatatacttatttaaccttctttcattggtttcttatctttactta 4200  
 tagactgtgataaaaagtagatttaatttgtatattctcttcagattacattacatttac 4260  
 cataaacaaaaataatacacatcaatgaaccattctcaaagagaaagggagtatgtgtgtg 4320  
 tgtgtgtgtaagtgttcaatatgacatacattcacaatgtaatttatgatttcgattgtg 4380  
 taagtgcctgcttaccctgttaataacttatataatataatgacgtgatgatccccctg 4440  
 gcaattagtgaggagttatcattcattaatctccatggtatataacatagttcatccaat 4500  
 atttcattattgaatatagacacttaattacttatgtatacttgaactaacttacctact 4560  
 tacttatacctgttacccttgaacctatacgacatatcttctgtagttatttctcata 4620  
 ggatgattctttacataccttaataaggaataattggctcaatgggtttcttttctttttg 4680  
 gatattccattgtatacttatgaatgtaggggttttgattttgaggggtattaggtaat 4740  
 tattacaaattgtaacttataacttactttttatagaagtaataaccaaaaaaaaaaaaaa 4800

## B

**2401** gattcaacatcaattttaactagttttcctata**GTGAGTGAG**gtgatatggata

Figura 6: Sequência nucleotídica de SmFes. A- O cDNA completo de SmFes contém uma ORF de 3.780 pb. O codón inicial, ATG, está representado por letras maiúsculas nas bases 1-3. O stop códon, TAA, está representado por letras maiúsculas nas bases 3.778-3.780. A região 5' UTR está representada pelas bases -123 a -1 e a 3' UTR está representada a partir da base 3.780 B- O polimorfismo de 9 pb na base 2.433 está representado em negrito e letras maiúsculas.



MGNHTVNEVSNYFERGSDVEDDINTLASNNTHITENNQSVLVSSTSDFL 0050  
 VPPLPPHSLGHDKSLSSSSISTPNVDNINTKRYTDSVMLRSDHKESSOVI 0100  
VSAIDAOMKSINEMLSYFSKLVKAEKTVOOALSDDLDDOKESSTVGSVDG 0150  
 GQRSNTTAMRNLFKSNYSRGRRRRLTMNILNKNLSNSPTDNELNKLQNL 0200  
 VISTNHQVRMFSTRALFHKLITAGQLDSLNNCEHCLRRTYLECEEELKSQ 0250  
AROYMLSLKAYEKKYFKYIGOLSNIOSKLTNLCDKRLINNRKIIDSLMNN 0300  
LEYSTKKLHLLHNEYCLALITAIHYHQWLYTHLRPCLLKGVERTMQLASE 0350  
 VVHLLLVFGGENIHHFQSNWLKEFQSELDLTSLLEQNGVKFPGPVYHTFN 0400  
QALNNNTTLGGNLSPGSLIVNDLTGOSLIELTQLHKNKELLHRETVNQLT 0450  
KECLOWEEALSQWRAVFANPGPNPWSNLFNIHSSGVSPIVNTPQDNNSN 0500  
 GFHHVDNSSHLALPDIQDLWHDDRARPVNLCEVSFRLAICEFELCFANCL 0550  
 ADSEAQLVNVLTDAQCRNLNWSLKLTTNDSTT**SYPNTTPITFSRDP**CDIT 0600  
 TTTTTTTNGNSNNNNHEDIDIQNGTDDNFSQALNGHDNSIPLCNLSDSYS 0650  
 TDPKFIEQISNSPFALSTSNTSDMSSSIEHQQQQRPERNRLWQSVRQSM 0700  
 SKFRSHFIRNNNIRSSTSSIPNTRTYQCRLRETIITTIPIILNTVNDNNNN 0750  
 SNSNSNNKPLHNPTVNGLKQQNHLGMDTPFYHQNRKFNHQNHQSNRESQ 0800  
 DSTSILTSFPiVseCDMDINKE**PWFHGVLPRAEVERLLQ****NGD****FLVRQTS** 0850  
**KRSSGRDTIGHWNGTNGDLINGNHNETILNKDHNMDGTVLRMVL****S****VFWHG** 0900  
**HRHFILYGGPETGEGWHLEDGHFSTIRELIEYHMKTQTPVTAKSGAC****LVT** 0950  
 PISRPDWELDNRDVQLLOKIGOGNFGDVYRGVYNGCEVA**V****K****T****CR****VD****M****T****A****S** 1000  
 DLRRKFLOGE**T****T****A****L****N****F****N****H****P****N****I****V****K****L****V****G****I****A****V****O****S****Y****P****I****M****I****V****M****E****Y****V****P****G****G****S****L****L****N****H****L** 1050  
 RKSKNALPVMKLLQMSLDAANG**M****M****Y****L****E****A****R****N****C****I****H****R****D****L****A****A****R****N****C****L****I****S****D****D****G****O****L****K** 1100  
**I****A****D****F****G****M****S****R****E****E****H****I****Y****E****L****S****D****K****R****G****O****I****P****I****K****W****T****A****P****E****A****L****R****T****G****R****Y****T****I****K****C****D****V****W****S****Y****G****V****L****L** 1150  
WEIFTFGDVPYRNWSNOQTRDMIESGYRLPAPDLMPVWLRTLMNHCWHDE 1200  
PMNRPSFSKISNEIQIPNVNSFTPNLNRSIDNAQLKKSAIDNLSTPLHL 1250  
 SVSVKDRRTRED- 1263

Figura 7: Seqüência de aminoácidos da proteína SmFes. A proteína codificada pelo gene *SmFes* está representada pela seqüência de 1.263 aa, contendo o inserto de 3 aa, <sup>812</sup>v-s-e <sup>814</sup> representado por letras minúsculas (ver Figura 20). Os prováveis aminoácidos correspondentes à região *coiled-coil* estão sublinhados e em itálico. O aminoácidos correspondentes ao domínio SH2 estão em negrito. Os aminoácidos correspondentes ao domínio PK estão sublinhados. Os 11 subdomínios do domínio PK estão representados por seus aminoácidos mais conservados, correspondendo às posições: SDI 971-977, SDII 989-991, SDIII 1.010, SDIV 1.021, SDV 1.038-1.041, SDVIA 1.072-1.083, SDVIB 1.085-1.091, SDVII 1.101-1.105, SDVIII 1.123-1.130, SDIX 1.142-1.148, SDX (subdomínio pouco conservado localizado entre o SDIX-XI) e SDXI 1.204 em sublinhados e em negrito. O resíduo de tirosina correspondente ao sítio de autofosforilação proposto está sublinhado em itálico e negrito na posição 1.113. O peptídeo desenhado para obtenção do anticorpo utilizado no ensaio de *western-blot* está representado em itálico e negrito.

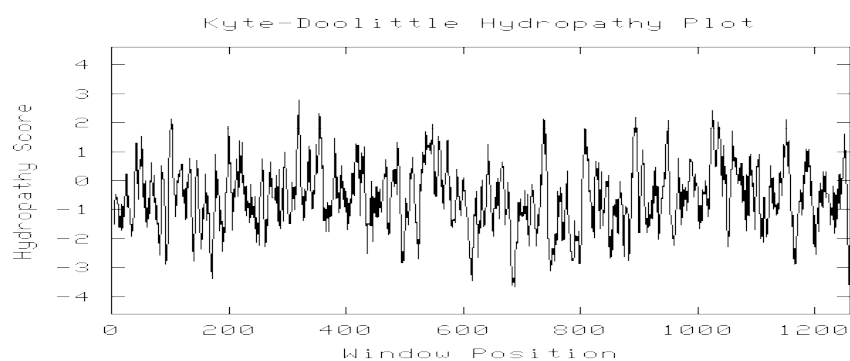
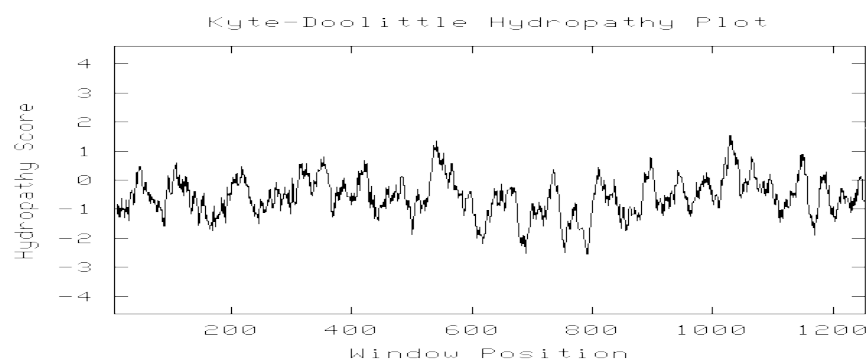
**A****B**

Figura 8: Gráfico de hidropaticidade da seqüência protéica de SmFes, segundo Kyte e Doolittle (1982). A- em janela de aminoácidos de 7 resíduos, B- em janela de aminoácido de 19 resíduos. O eixo X indica as posições referentes aos aminoácidos da seqüência protéica. O eixo Y indica a pontuação referente à hidropaticidade. Valores positivos indicam regiões hidrofóbicas e valores negativos regiões hidrofílicas. A linha horizontal, correspondente ao valor 1,8, indica o linear de hidrofobicidade correspondente a regiões de transmembrana.

Quadro 6: BLAST-p da proteína SmFes contra banco de dados nr.

ORGANISMO	DENOMINAÇÃO	REGIÃO DE IDENTIDADE COM SMFES	SCORE (BITS)	IDENTIDADE	VALOR-E
<i>Drosophila melanogaster</i>	Tyrosine-protein kinase Fps85D (dFer) (P18106)	777-1.214	372	45%	9e-99
		193-379	46,2	22%	0,007
<i>Rattus norvegicus</i>	Similar to Fer (XP_217492) similar to tyrosine kinase Fps/Fes (XP_341877)	811-1.214	320	42%	2e-85
		91-346	45,4	21%	0,012
<i>Canis familiaris</i>	Protein tyrosine kinase fer (AAF00543)	793-1.216	319	41%	4e-85
		91-346	42,0	21%	0,13
<i>Mus musculus</i>	Fert2 protein (AAH58100)	811-1.216	318	42%	6e-85
		245-346	38,9	26%	1,1
	Tyrosine kinase Fps/Fes (AAN33122)	804-1.214	317	43%	2e-84
	Fer (AAB18988)	811-1.216	318	42%	6e-85
		91-346	43,9	21%	0,034
<i>Felis catus</i>	Protein-tyrosine kinase fes/fps (TVCTFF)	811-1.215	316	43%	3e-84
<i>Homo sapiens</i>	Proto-oncogene tyrosine-protein kinase Fes/Fps C-Fes (P07332)	811-1.215	315	43%	8e-84
		811-1.216	316	42%	3e-84
		91-346	39,7	19%	0,65
<i>Gallus gallus</i>	c-fps proto oncogene (CAA26155)	810-1.214	314	41%	1e-83
<i>Ephydatia fluviatilis</i>	Protein tyrosine kinase (BAA817213)	819-1.215	313	(41%)	2e-83
		186-347	38,1	20%	1,9
<i>Sycon raphanus</i>	Tyrosine kinase (CAA76605)	814-1.214	306	41%	4e-81
		94-452	40,4	21%	0,38
<i>Drosophila pseudoobscura</i>	GA21383-PA (EAL26982)	790-1.214	364	46%	1e-98
		193-423	45,8	21%	0,009
<i>Anopheles gambiae</i>	ENSANGP00000020257 (XP_313423)	819-1.214	362	48%	4e-98
		205-379	38,5	20%	1,4
<i>Xenopus laevis</i>	MGC80946 protein (AAH73445)	788-1.208	311	42%	9e-83

(Src-homology-2) ([IPR000980](#), [gnl|CDD|16538](#)) e uma de domínio catalítico Proteína Tirosina Quinase ([IPR001245](#), [gnl|CDD|5392](#)). No entanto, nenhum domínio FCH (Fes/CIP4 homology domain), usualmente encontrado na família Fes/Fps/Fer, foi identificado em SmFes. Os domínios encontrados se relacionam às regiões de maior homologia da seqüência protéica de SmFes. (Figura 7 e 9)

#### **V.1.3.1 Subdomínios do Domínio PK**

Onze subdomínios característicos do domínio catalítico PK, relatado por Hanks (1988), foram encontrados em SmFes em posição regular quando comparado pelo alinhamento de seqüências da família Fes/Fps/Fer de outros organismos (Figura 7 e 10). Esses onze subdomínios principais evidentes estão separados por regiões de baixa conservação. Os subdomínios VI e VIII encontrados em SmFes são característicos de Proteínas Tirosinas Quinase, confirmando a anotação funcional de SmFes.

#### **V.1.3.2 Sítio de autofosforilação em resíduo de tirosina**

No alinhamento múltiplo do domínio catalítico de SmFes e de proteínas ortólogas de outros organismos foi observado que o sítio de autofosforilação no resíduo de tirosina relacionado ao aa 713 proposto para c-Fes humana permanece presente para todas as proteínas relacionadas, com exceção de *C. elegans* (Figura 10). Este resíduo foi encontrado na posição 1.113 da proteína de SmFes (Figura 7). Outro sítio de autofosforilação em resíduo de tirosina relacionado ao aa 811 proposto para c-Fes humana não foi observado em SmFes e também na maioria das proteínas alinhadas, estando presente apenas em dFer de *D. melanogaster*, Fes de *F. catus* e PTK de *S. raphanus* (Figura 10).

## **V.2 ANÁLISE FILOGENÉTICA DE SmFES**

Uma reconstrução filogenética foi realizada para confirmar a proximidade de SmFes às proteínas da família Fes/Fps/Fer. A análise filogenética foi realizada baseando-se na seqüência de aminoácido dos domínios SH2-PK ou na seqüência completa de aminoácidos. O dendograma resultante das seqüências de aminoácidos dos domínios SH2-PK indicou uma relação evolutiva mais próxima aos organismos invertebrados *C. elegans*, *D. melanogaster*, *E. fluviialis* e *S. raphanus*. com valor de *bootstrap* significativo (Figura 11 B). O dendograma resultante das seqüências de aminoácidos completas indicou uma relação evolutiva com as seqüências Fes-like dos organismos invertebrados *D. melanogaster*; *E. fluviialis* e *S. raphanus* com valor de *bootstrap* significativo (Figura 11 A). Ambos dendogramas indicaram resultado similar, com exceção de CeFerFk1, Fer de *C. elegans*, que foi agrupada no ramo das seqüências de invertebrados na árvore dos domínios SH2-PK, enquanto que na árvore com as

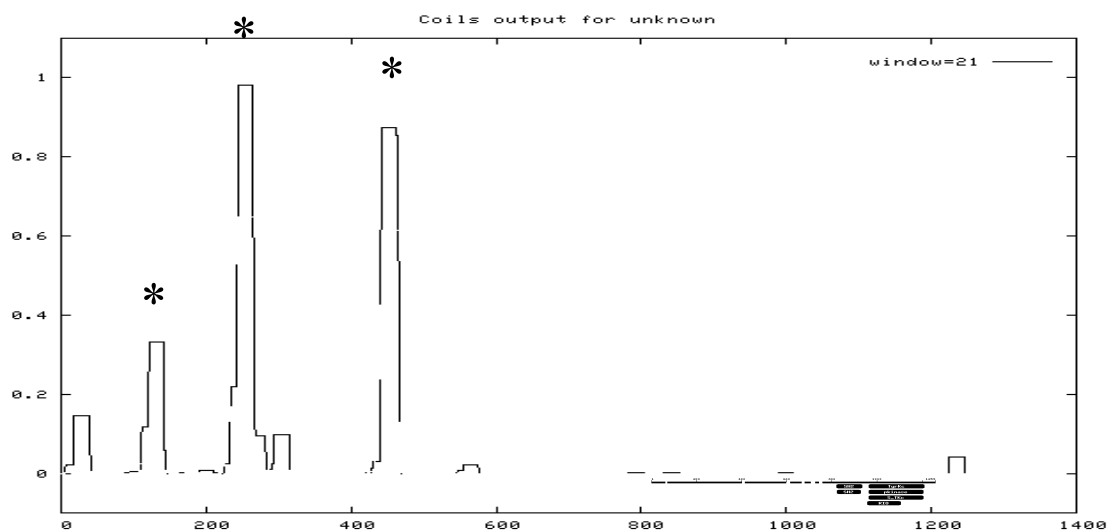


Figura 9: Esquema representativo dos domínios de SmFes e sua localização na seqüência primária de aminoácidos. Análises dos domínios da seqüência de aminoácidos de SmFes foram realizadas pelo programa RPS-BLAST e Coils 2.1. No eixo X estão representadas as posições dos aminoácidos da seqüência primária de SmFes. As caixas representam as localizações dos domínios SH2 e catalítico PK. No eixo Y está representada a chance da seqüência ter a conformação *coiled-coil*. Os picos do gráfico representam as regiões com maior probabilidade de adquirir a conformação *coiled-coil*, observando-se três picos mais prováveis entre os aminoácidos 100 e 500, indicados por um asterisco.

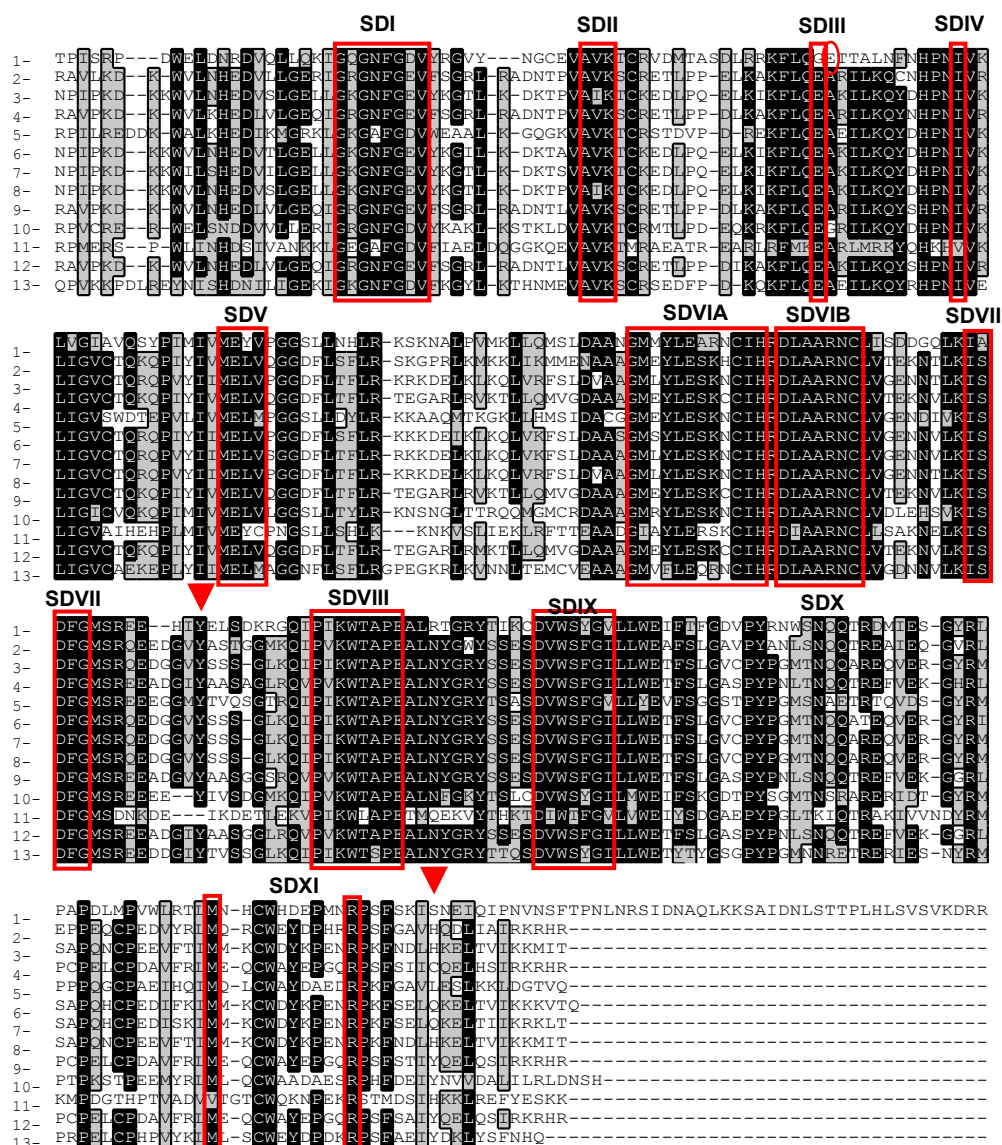
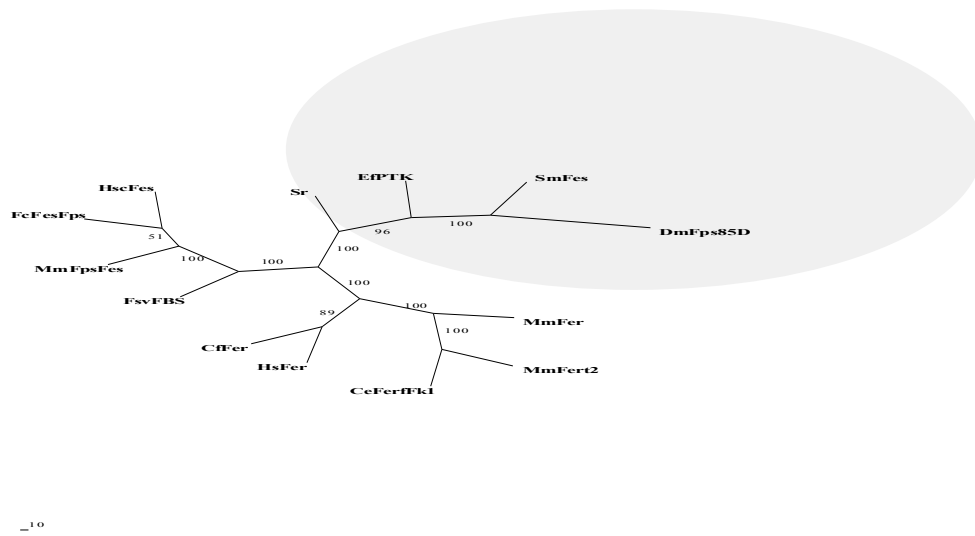


Figura 10: Alinhamento múltiplo do domínio catalítico de SmFes e de proteínas ortólogas de outros organismos. As seqüências usadas e o respectivo número de acesso no banco de dados foram: 1- Fes like de *Schistosoma mansoni*, (AAP68901); 2- FBS de Fujinami sarcoma vírus (NP\_955606); 3- Fert2 proteína de *Mus musculus*, (AAH58100); 4- Fps/Fes de *Mus musculus* (AAN33122); 5- PTK de *Ephydatia fluvialis* (BAA81721); 6- Fer de *Canis familiaris* (AAF00543); 7- Fer de *Homo sapiens* (NP\_005237); 8- Fer de *Mus musculus* (AAB18988); 9- c-FES de *Homo sapiens* (P07332); 10- Fps85D(dFer) de *Drosophila melanogaster* (P18106); 11- Fer-frk-1 de *Caenorhabditis elegans* (NP\_501818); 12- Fes/Fps de *Felis catus* (TVCTFF); 13- PTK de *Sycon raphanus* (CAA76605). Os subdomínios conservados estão indicados no topo do alinhamento em caixas vermelhas. Em cada subdomínio encontram-se os aminoácidos mais conservados: SD-I GxGxxGxV; SDII AxK; SDIII E ; SDIV I; SDV MEYV; SDVIA GxxYL6xH; SDVIB DLAARNC; SDVII IxDFG; SDVIII PxxWxAPE; SDIX DxWSxGV; SDX; SDXI M11xR. As setas vermelhas indicam sítios de autofosforilação em resíduo de tirosina determinados em c-Fes de *Homo sapiens* (P07332). Os aminoácidos em preto são idênticos na maioria das seqüências e os aminoácidos em cinza são similares (polaridade, hidropaticidade). Para alinhamento das seqüências completas ver anexo 3.

A



B

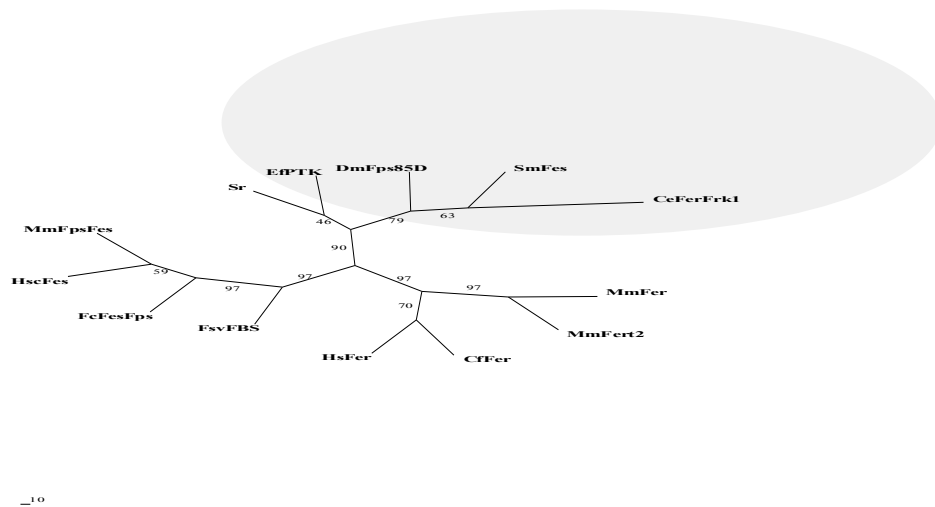


Figura 11: Árvore filogenética sem raiz de SmFes construída com outras proteínas da família Fps/Fes/Fer. As duas primeiras letras correspondem ao organismo: Sm, *Schistosoma mansoni*; Hs, *Homo sapiens*; Cf, *Canis familiaris*; Mm, *Mus musculus*; Ce, *Caenorhabditis elegans*; Fsv, *Fujinami sarcoma virus*; Ef, *Ephydatia fluvialis*; Dm, *Drosophila melanogaster*; Fc, *Felis catus*; Sr, *Sycon raphanus*; Rn, *Rattus norvegicus*. As letras subsequentes correspondem à denominação recebida para cada ortólogo da família Fes/Fps/Fer. A - Alinhamento das seqüências completas das proteínas. B - Alinhamento das regiões SH2 e PK. As árvores foram construídas a partir de seqüências alinhadas com ClustalW, utilizando-se o método de

parcimônia com 1.000 replicações. A árvore e o valor de *bootstrap* foram visualizados no programa TreeView. seqüências completas foi agrupada no ramo das seqüências caracterizadas como da subfamília Fer (Figura 11).

### V.3 ORGANIZAÇÃO GENÔMICA DE SMFES

A análise resultante do alinhamento do cDNA de SmFes às seqüências genômicas de *S. mansoni* depositadas nos bancos de dados dos Institutos Sanger e TIGR, indicou a presença de 17 introns. Os tamanhos dos exons variaram de 65 a 575 pb. Apenas 11 introns tiveram sua seqüência completa finalizada. Entre os introns que foram completamente seqüenciados, o tamanho variou de 686 a 6.777 pb. Os elementos doadores e aceptores esperados foram identificados em todos os introns (Anexo I e II). Quatro introns foram localizados no domínio PK e dois no domínio SH2 (Figura 12).

Um resultado similar foi obtido entre nossa montagem e a realizada pelo Instituto Sanger, indicando a mesma localização e número de introns. Os resultados referentes ao tamanho e a seqüência dos introns foram também, semelhantes, com exceção do intron I-17, que teve na nossa montagem tamanho de 686 pb, enquanto que, na montagem do Instituto Sanger teve tamanhos de 359 pb (contig 333748 c.006012707) (Anexo II) ou 566 pb (contig d000017208) (Anexo II). No alinhamento do Instituto Sanger, para o I-16, foram encontrados dois contigs distintos: um deles foi semelhante ao da nossa montagem com 1.290 pb (contig 333748 c.004513974) (Anexo II) e outro diferente, com 2.010 pb (contig 333748 c.002729397) (ver Anexo II) (Quadro 7).

Os introns I-01, I-06, I-11, I-12, I-13, I-14, I-16 e I-17 tiveram seus tamanhos e seqüências verificadas por amplificação e digestão do produto de PCR do DNA genômico (Figura 13 A e B).

A amplificação do intron I-01 foi realizada utilizando um par de iniciadores localizados, respectivamente, nos exons E-01 e E-02, esperando-se uma banda de 1.822 pb. No gel de agarose, foi observada uma banda de peso molecular de acordo com o esperado (Figura 13 A). A amplificação do intron I-06 foi realizada utilizando um par de iniciadores localizados nos exons E-06 e E-07, esperando-se uma banda de 2.010 pb. No gel de agarose observou-se uma banda de peso molecular de acordo com o esperado (Figura 13 A). Para comprovar a especificidade do produto especificado, foi realizada a digestão desta amplificação com enzima de restrição *EcoR V* esperando-se obter, segundo os sítios da enzima na seqüência nucleotídica, bandas de 339 pb, 663 pb e 1.012 pb. As três bandas foram observadas no gel de agarose em seu tamanho esperado (Figura 13 B). A amplificação do intron I-11 foi realizada



utilizando um par de iniciadores localizados nos exons E-11 e E-12, esperando-se uma banda de 3.701 pb. No gel de agarose foi observada uma banda de peso molecular de acordo com o

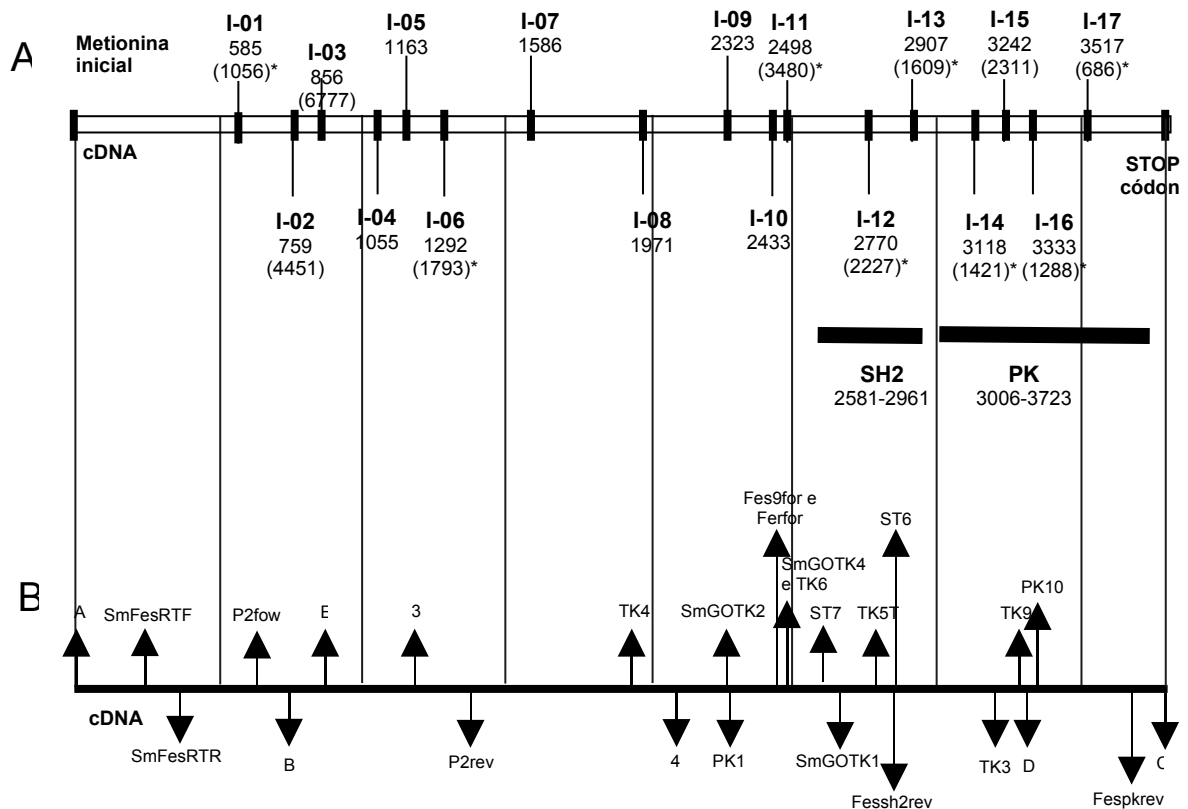


Figura 12: Organização genômica do gene *SmFes* e localização dos iniciadores na seqüência. A- Os 17 introns (I-01 a I-17) estão indicados na seqüência de cDNA de 3.780 pb. As posições dos introns na seqüência de cDNA estão indicadas pelo número abaixo de cada intron correspondente. Os introns I-01-03, I-06 e I-11 a I-17 tiveram suas seqüências completas obtidas a partir do banco de dados do Instituto Sanger e TIGR e seus tamanhos estão representados em parêntesis abaixo da sua posição. Os demais introns não foram completamente seqüenciados e montados. Os asteriscos indicam os introns que foram experimentalmente verificados por PCR. Os domínios SH2 e PK, em suas respectivas posições, estão representados por caixas. B- A linha representa a posição em que os iniciadores usados nesse trabalho encontram-se na seqüência de cDNA. Os iniciadores estão indicados por setas de modo que: setas apontadas para cima indicam os iniciadores *forward* e as setas apontadas para baixo, os iniciadores *reverse* (Quadro 4). Todos os elementos desta figura encontram-se em escala, representada pelas linhas verticais, indicada a cada 500 pb.

Quadro 7: Quadro comparativo do tamanho encontrado para os exons e introns na montagem realizada por nós e pela montagem do Instituto Sanger

EXONS	TAMANHO EM PB	INTRONS	TAMANHO EM PB	
			CPqRR* <sup>1</sup>	INSTITUTO SANGER
E-01	585			
		I-01	1.056	1.058
E-02	174			
		I-02	-	4.451
E-03	97			
		I-03	-	6.777
E-04	199			
		I-04	-	-
E-05	108			
		I-05	-	-
E-06	129			
		I-06	-	1.791
E-07	294			
		I-07	-	-
E-08	331			
		I-08	-	-
E-09	406			
		I-09	-	-
E-10	110/119* <sup>2</sup>			
		I-10	-	-
E-11	65			
		I-11	-	3.480
E-12	273			
		I-12	2.227	2.245
E-13	136			
		I-13	1.609	1.611
E-14	211			
		I-14	1.421	1.420
E-15	124			
		I-15	2.311	2.310
E-16	91			
		I-16	1.294	2.010/1.290
E-17	184			
		I-17	686	566/359
E-18	263			
TOTAL	3.780/3.789* <sup>2</sup>		10,6 Kb	27,7 Kb

\*1- a montagem do nosso grupo foi realizada através de alinhamento de seqüências genômicas, depositadas nos bancos de dados de *S. mansoni*, através do programa *GeneTool*; \*2- Os dois valores encontrados no Exon 9 correspondem as duas populações com ou sem os 9 pb inseridos.

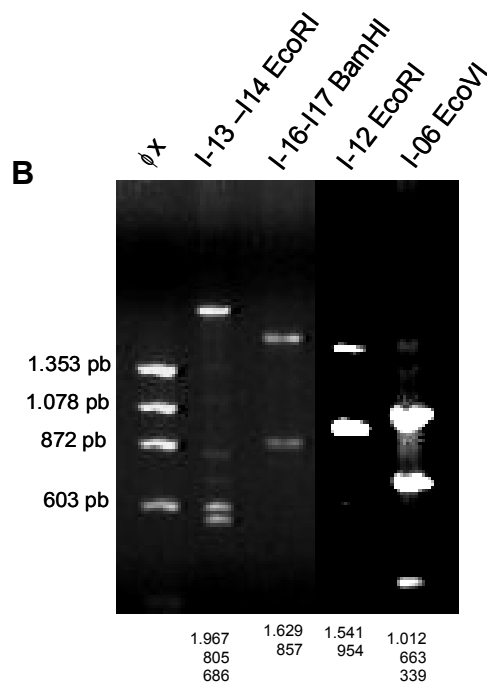
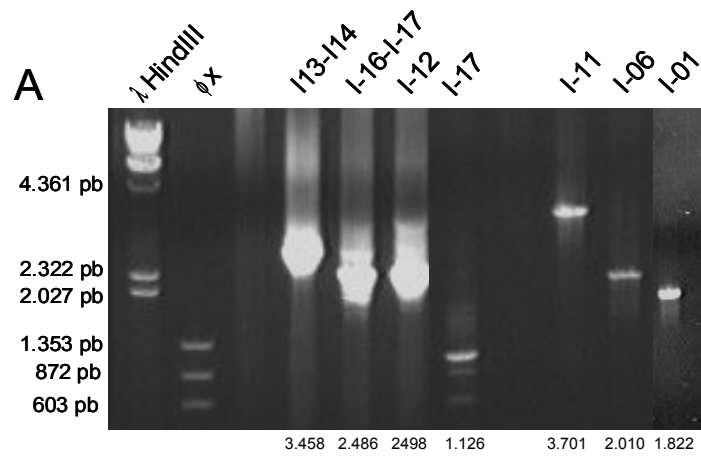


Figura 13: Análise do DNA genômico para comprovação dos introns de SmFes. Gel de agarose 0,8% corados por brometo de etídeo. A- Amplificação dos introns por PCR (ver Quadro 2). B- Digestão com enzima de restrição dos introns amplificados em A (ver Quadro 5). Abaixo de cada canaleta estão representados, em pb, os tamanhos das bandas esperadas.

esperado (Figura 13 A). A amplificação do intron I-12 foi realizada utilizando um par de iniciadores localizados nos exons E-12 e E-13, esperando-se uma banda de 2.495 pb. No gel de agarose observou-se uma banda de peso molecular de acordo com o esperado (Figura 13 A). Para comprovar a especificidade do produto especificado, foi realizada a digestão desta amplificação com enzima de restrição *EcoR I* esperando-se obter, segundo os sítios da enzima na seqüência nucleotídica, bandas de 954 pb e 1.541 pb. As duas bandas foram observadas no gel de agarose em seu tamanho esperado (Figura 13 B). A amplificação do intron I-13 e I-14 foi realizada utilizando um par de iniciadores localizados nos exons E-13 e E-15, esperando-se uma banda de 3.458 pb. No gel de agarose observou-se uma banda de peso molecular menor que o esperado (Figura 13 A). O produto desta amplificação foi digerida com enzima *EcoR I* esperando-se obter, segundo os sítios da enzima localizados na seqüência nucleotídica, bandas de 686 pb, 805 pb e 1.967 pb. A banda de tamanho 1.967, correspondente ao intron I-13, foi observada no gel de agarose, enquanto que as demais bandas tiveram um tamanho menor ao esperado (Figura 13 B). A amplificação do intron I-16 e I-17 foi realizada utilizando um par de iniciadores localizados nos exons E-16 e E-18, esperando-se uma banda de 2.486 pb, segundo a nossa montagem. No gel de agarose foi observado uma banda de peso molecular de acordo com o esperado pela montagem do nosso grupo (Figura 13 A). Para comprovar a especificidade do produto especificado e a montagem do intron foi realizada digestão desta amplificação com enzima de restrição *BamH I* esperando-se obter, segundo os sítios da enzima na seqüência nucleotídica, bandas de 857 pb e 1.629 pb. As duas bandas foram observadas no gel de agarose em seu tamanho esperado (Figura 13 B). A amplificação do intron I-17 isoladamente foi realizada utilizando um par de iniciadores localizados nos exons E-17 e E-18, esperando-se uma banda de 1.126 pb, segundo a montagem realizada pelo nosso grupo (Figura 13 A). No gel de agarose foi observada uma banda de peso molecular de acordo com esperado, confirmando o resultado da nossa montagem e um valor aproximado para a montagem do I-16 do Instituto Sanger correspondente a 566 pb.

## V.4 SOUTHERN-BLOT

A hibridização da sonda TK4-PK1, complementar ao exon E-09, revelou apenas uma banda de aproximadamente 4,5 Kb no DNA genômico digerido com *Hind III* e uma banda de aproximadamente 1,6 Kb para o DNA genômico digerido com *EcoR I* (Figura 14 A). A confirmação do tamanho exato correspondente a estas bandas não pôde ser realizada, uma vez que a seqüência nucleotídica do intron I-09 não foi concluída. Não existindo um sítio de corte para ambas as enzimas no exon E-09 é esperado a revelação de uma única banda para cada enzima, o que é uma forte indicação que SmFes existe em cópia única no genoma de *S. mansoni* (Figura 15).

A hibridização da sonda SmGOTK1-SmGOTK2, que hibridiza com os exons E-09, E-10, E-11 e E-12, revelou duas bandas de aproximadamente 5,2 Kb e 4,5 Kb para DNA genômico digerido com *Hind III* e três bandas de aproximadamente 1,6 Kb, 3,5 Kb e 5,0 Kb para o DNA genômico digerido com *EcoR I* (Figura 14 B). A confirmação do tamanho exato correspondente a estas bandas não pode ser determinado, uma vez que as seqüências nucleotídicas dos introns I-09 e I-10 não estão completas. As hibridizações dos exons E-11 e E-12 puderam ter seus pesos confirmados, uma vez que sítios para ambas as enzimas *Hind III* e *EcoR I* foram identificados na seqüência parcial do intron I-10 e na seqüência completa do intron I-11. A enzima *EcoR I* libera os exons E-11 e E-12, por não haver sítio para a enzima no intron I-11, com um fragmento de tamanho esperado de 5.164 pb (Figura 15). Para *EcoR I*, foi confirmada a presença da banda de 5,0 Kb, correspondente ao E-11 e E-12, e de mais duas bandas, correspondentes aos exons E-09 e E-10 (1,6 e 3,5 Kb) (Figura 14 B). A enzima *Hind III* libera o exon E-11 com um fragmento de tamanho de 280 pb e o E-12 com um fragmento de tamanho de 5.267 pb (Figura 15). Para *Hind III*, foi confirmada a presença da banda de 5,2 Kb, correspondente ao E-12 (Figura 14 B). No entanto, a banda do fragmento de 280 pb, correspondente ao E-11, não foi observada. Levando em consideração o fato de que esta banda apresenta um tamanho pequeno, a quantidade em que ela esta presente no gel pode ser insuficiente para sua detecção em *Southern-blot* não radioativo. Esperava-se encontrar outras duas bandas na digestão com *Hind III*, correspondentes aos exons E-09 e E-10, no entanto, uma única banda foi observada. A ausência da outra banda poderia ser explicada devido à sobreposição de fragmentos de tamanho similar digeridos no gel de agarose.

A hibridização da sonda radioativa RP2rev-“E”, que hibridiza com os exons E-04, E-05, E-06 e E-07, revelou duas banda de, aproximadamente, 4,8 Kb e 3,5 Kb para DNA genômico digerido com *EcoR I* e três banda de, aproximadamente, 8,0 Kb, 5,5 Kb e 3,0 Kb para o DNA genômico digerido com *Hind III*. Para a hibridização com a sonda radioativa era esperado encontrar, no máximo, 3 bandas para *Hind III* e *EcoR I*. Não se pôde determinar o

tamanho das bandas que hibridizam com os E5 e E6-E7, já que as seqüências encontram-se incompletas nos I-04, I-05 e I-07. A hibridização do exon E-04 teve seu peso confirmado, uma vez que sítios para ambas as enzimas *Hind III* e *EcoR I* foram identificados na seqüência do intron I-03 e I-04, prevendo para DNA genômico digerido com *Hind III* banda correspondente a 3.151 pb e para *EcoR I* banda correspondente a 4.792 pb (Figura 16).

Ambos os *Southern-blot* radioativo e não radioativo indicaram que SmFes é um gene de cópia única.

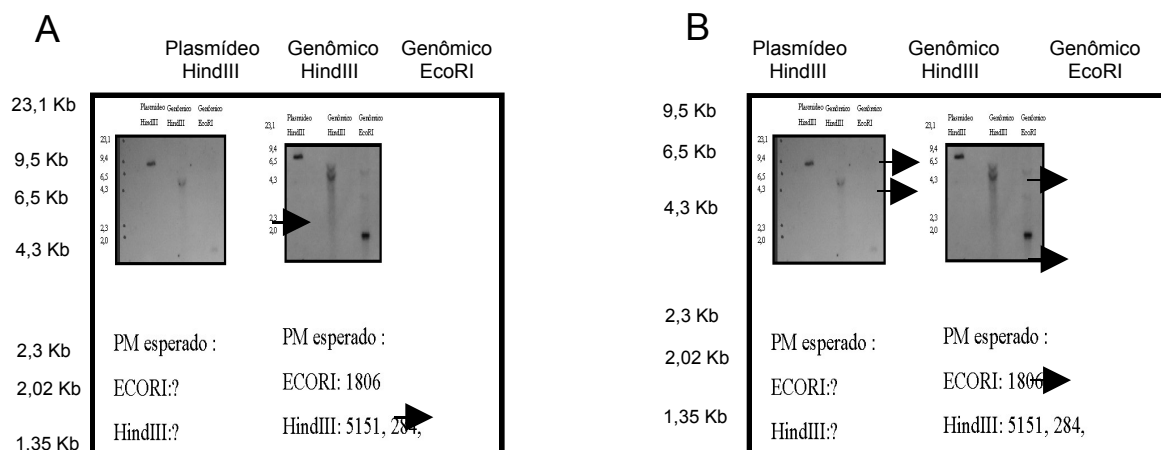


Figura 14: *Southern-blot* do gene SmFes. Nos painéis A e B o controle positivo é indicado pelo plasmídeo, cujo inserto contém o cDNA completo de SmFes, digerido com *Hind III* apresentando uma banda esperada de 7,8 Kb. A- Imagem correspondente à hibridização com a sonda TK4-PK1. O DNA genômico digerido com *Hind III* revelou uma banda de 4,5 Kb. A digestão com *EcoR I* revelou uma banda de 1,6 Kb. Os fragmentos revelados estão indicados por setas. B- Imagem correspondente à hibridização com a sonda SmGOTK1-SmGOTK2. O DNA genômico digerido com *Hind III* revelou duas bandas de, aproximadamente, 5,2 Kb e 4,5 Kb. A digestão com *EcoR I* revelou três bandas de, aproximadamente, 1,6 Kb, 3,5 Kb e 5,0 Kb



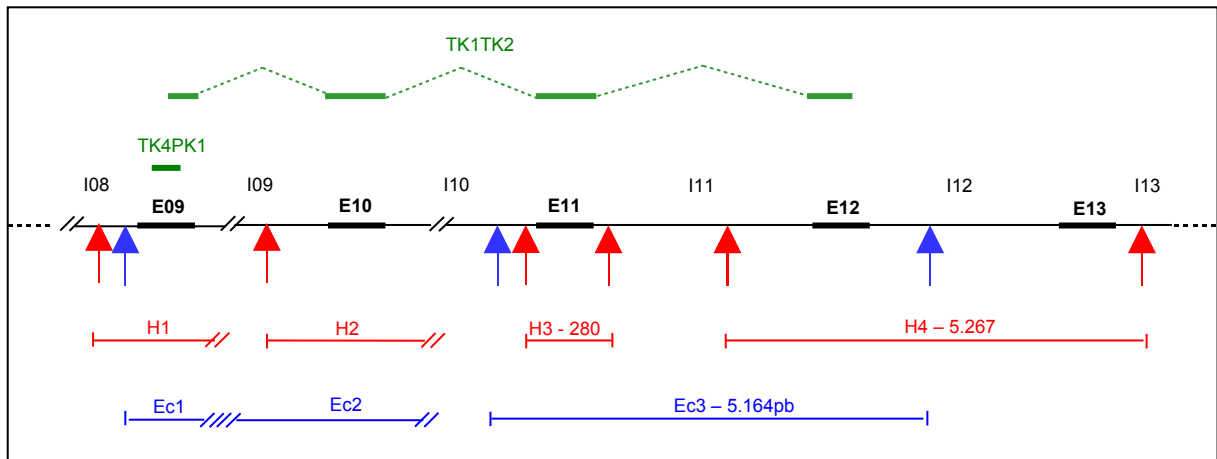
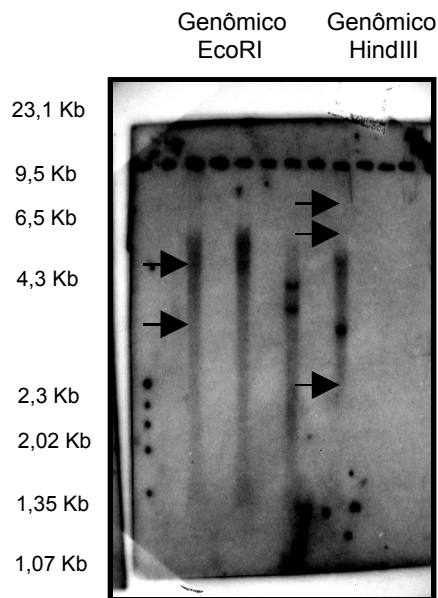


Figura 15: Esquema representativo de *Southern-blot* não radioativo. A linha preta corresponde à sequência genômica parcial de SmFes com os introns (linha mais fina) e exons (linha mais larga) representados por I e E, respectivamente. As seqüências dos introns que se encontram incompletas estão indicadas por barras paralelas. As linhas verdes correspondem às sondas utilizadas na hibridização, que contêm somente os exons, pois foram produzidas a partir de cDNA. As setas correspondem aos sítios de enzimas de restrição encontrados na seqüência genômica, em vermelho sítios para *Hind* III e em azul para *EcoR* I. H1-H4 correspondem aos fragmentos esperados para a digestão com *Hind* III. Ec1-3 correspondem aos fragmentos esperados pra a digestão com *EcoR* I.



B

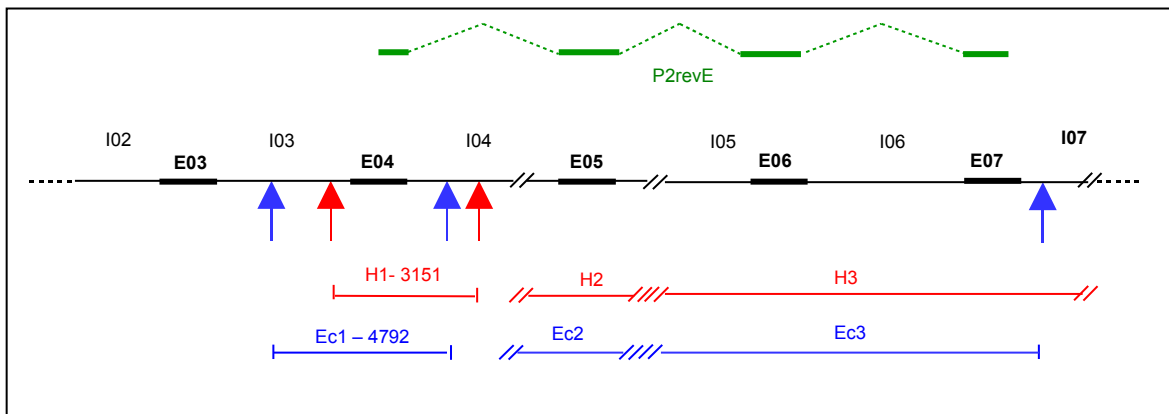


Figura 16: *Southern-blot* radioativo do gene SmFes e esquema representativo. A- Imagem correspondente à hibridização com a sonda P2rev-“E”. O DNA genômico digerido com *EcoR I* revelou duas banda de, aproximadamente, 4,8 Kb e 3,5 Kb. A digestão com *Hind III* revelou três banda de, aproximadamente, 8,0 Kb, 5,5 Kb e 3,0 Kb. Os fragmentos revelados estão indicados por setas. B- Esquema representativo de *Southern-blot* radioativo. A linha preta corresponde a seqüência genômica de SmFes com os introns (linha mais fina) e exons (linha mais larga) representados por I e E, respectivamente. As seqüências dos introns que se encontram incompletas estão indicadas por barras paralelas. As linhas verdes correspondem à sonda utilizada na hibridização, que contém somente os exons, pois foram produzidas a partir de cDNA. As setas correspondem aos sítios de enzimas de restrição encontrados na seqüência genômica, em vermelho sítios para *Hind III* e em azul para *EcoR I*. H1-H3 correspondem aos fragmentos esperados para a digestão com *Hind III*. Ec1-3 correspondem aos fragmentos esperados pra a digestão com *EcoR I*.

## V.5 PADRÃO DE EXPRESSÃO DA PROTEÍNA SmFES

O perfil de expressão de SmFes foi verificado em diversas fases de desenvolvimento do parasito pelo método de *western-blot* com anticorpo produzido a partir de peptídeo sintético. SmFes apresentou maior expressão em cercárias, no entanto, miracídios, esporocistos e vermes adultos também apresentaram algum nível de expressão da proteína. Uma banda com peso molecular de 143 kDa foi observada, de acordo com o esperado (Figura 17).

## V.6 PADRÃO DE TRANSCRIÇÃO DO RNA DE SmFES

### V.6.1 PCR em Tempo Real

A quantificação relativa de SmFes por PCR em tempo real indicou um perfil de transcrição distinto para as diferentes fases de desenvolvimento do parasito. Pelo método de  $2^{-\Delta\Delta Ct}$ , foram encontrados valores correspondentes a uma transcrição de 69% em miracídio, 37% em verme fêmea, 65% em esporocisto e 27% em cercária quando comparado a verme adulto macho (Figura 18). O método de análise dos resultados utilizado normaliza a quantidade relativa de RNA entre as amostras através da equação  $2^{-\Delta\Delta Ct}$ . Os valores Ct correspondem ao número de ciclos que cada amostra leva para atingir um determinado número de cópias amplificadas, os valores de  $\Delta Ct$  correspondem ao Ct de cada amostra menos o Ct do controle interno (tubulina) e os valores de  $\Delta\Delta Ct$  correspondem ao  $\Delta Ct$  encontrado para cada amostra menos o  $\Delta Ct$  encontrado para a amostra usada como normalizador (verme adulto macho).

### V.6.2 Northern-blot

A hibridização da sonda radioativa RP2rev-“E” à membrana contendo 20  $\mu\text{g}$  de cada fase de desenvolvimento do parasito indicou a presença de transcritos em vermes machos e esquistossômulos, pela presença de banda de tamanho esperado para o RNA de SmFes, correspondente a 5 Kb. No entanto, não se pode fazer uma análise quantitativa deste *Northern-blot*, uma vez que não foi também utilizada uma sonda correspondente a um gene constitutivo de *S. mansoni* (Figura 19).

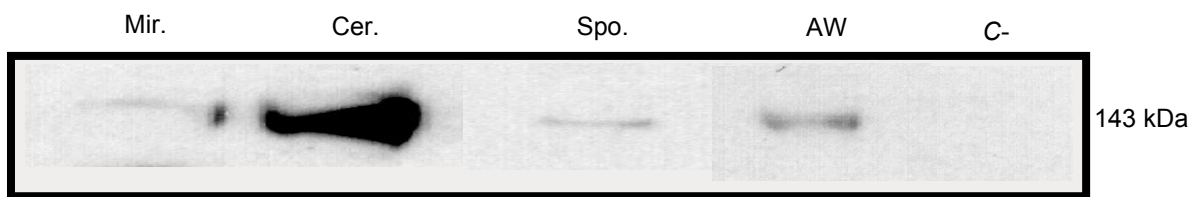


Figura 17: *Western-blot* de SmFes realizado com soro de coelho anti-SmFes na diluição 1:1.000 de diferentes estágios de desenvolvimento do parasito de *S. mansoni*, verme adulto casal (AW.), cercária (Cer.), miracídio (Mir.) e esporocisto (Spo.); C- representa o *western-blot* realizado com soro pré-imune em cercárias.

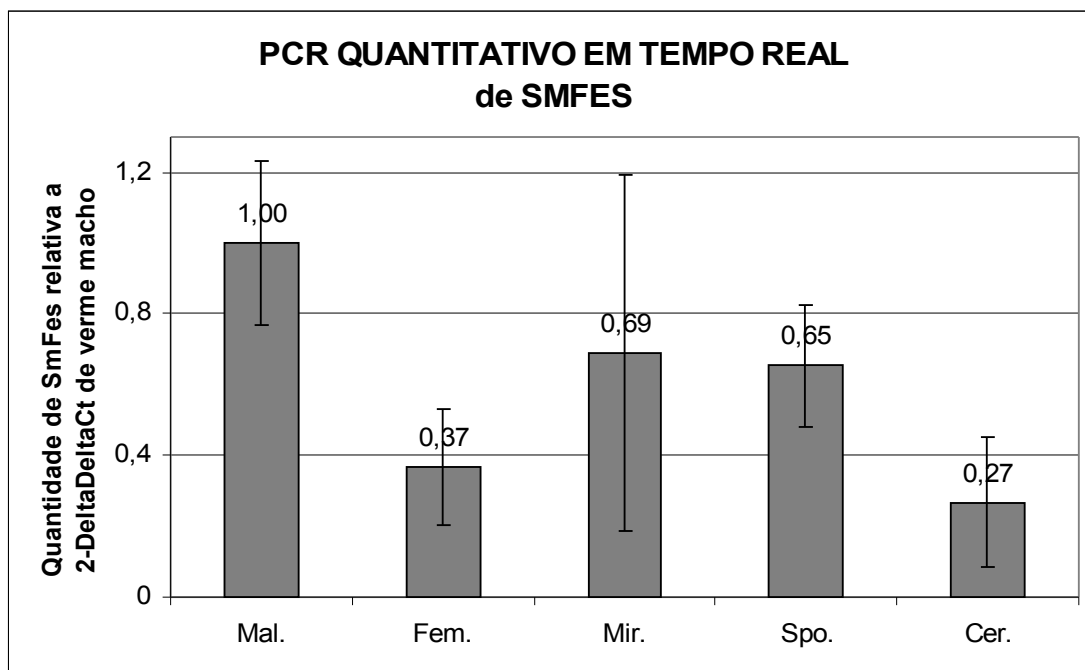


Figura 18: Gráfico representativo do perfil de transcrição de SmFes no diferentes estágios de *S. mansoni* por PCR em tempo real. O gráfico indica a quantidade relativa de transcritos de SmFes entre verme macho (Mal.), verme fêmea (Fem.), miracídio (Mir.), esporocisto (Spo.) e cercária (Cer.). O cálculo foi baseado no  $2^{-\Delta\Delta Ct}$  de verme macho.

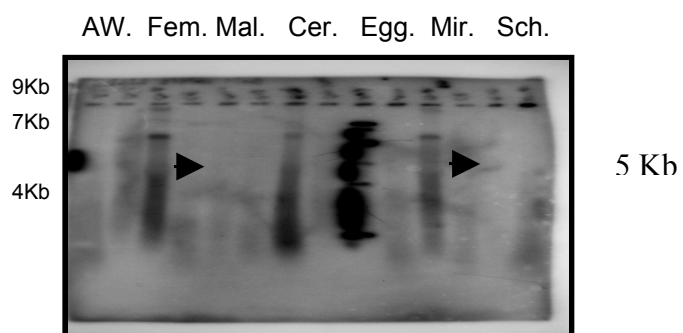


Figura 19: *Northern-blot* do gene SmFes. 20 µg de RNA de cada fase de desenvolvimento de *S. mansoni*, verme adulto casal (AW.), verme adulto fêmea (Fem.), verme adulto macho (Mal.), cercária (Cer.), ovo (Egg), miracídio (Mir.), esquistossômulo (Sch.), foram corridos em gel de agarose desnaturante e transferidos por capilaridade para membrana de nylon. A hibridização foi realizada por sonda de DNA, amplificada a partir do par de iniciadores P2rev e “E”, marcada por  $^{32}\text{P}$ -dCTP. Banda de tamanho correspondente ao do RNA de SmFes foram observadas em verme macho e esquistossômulo e estão indicadas por setas.

## V.7 POLIMORFISMO EM SMFES

### V.7.1 Inserção de 9 pb no cDNA de SmFes

A análise de clones parciais de cDNA apontou para a possibilidade de ocorrerem duas populações de RNA, contendo ou não 9 pb inseridos na posição 2.433 do cDNA, codificando para os aminoácidos V-S-E (valina-serina-glutamato) após o aminoácido 811 e localizado a 5' da região codificante para o domínio SH2 (Figura 20 A). A comparação da seqüência de cDNA com seqüências genômicas da região correspondente indicaram a existência de um sítio de edição alternativa na região 5' do intron I-10, o que poderia resultar em um evento de edição alternativa (Figura 20 B). A verificação da existência destas duas populações ocorreu pela amplificação e digestão do cDNA com a enzima de restrição *Sfc I* que corta exatamente no sítio de polimorfismo (Figura 20 B). Três diferentes amplificações da região do inserto foram realizadas com cDNA de vermes adulto. Todos os fragmentos foram digeridos parcialmente, indicando a presença das duas populações de cDNA. A soma dos pesos moleculares dos fragmentos gerados por *Sfc I* corresponde ao peso molecular do fragmento amplificado de cDNA não digerido (Figura 20 C). A mesma observação foi feita para cDNAs de grupos de macho, fêmea e cercária indicando a presença das duas populações em diferentes estágios de desenvolvimento do parasito. A amplificação de verme individual mostrou que estas duas populações de cDNA podem coexistir em um mesmo organismo. (Figura 21)

### V.7.2 Inserção de 15 pb no DNA genômico contendo o gene SmFes

A comparação entre seqüências de DNA genômico depositadas no banco de dados de *S. mansoni* e seqüências de cDNA indicou a possibilidade de existir um polimorfismo entre as bases CACCAG na posição 2.041 da seqüência do cDNA de SmFes ou CAACAACACTACAACATCAGCAA na seqüência genômica correspondente ao exon E-09 (Figura 22). Esse polimorfismo poderia levar a uma troca de aa entre HQ (histidina-glutamina) por QQLQHQQ (2Xglutamina-leucina-glutamina-histidina-2Xglutamina), correspondendo a um acréscimo de 5 aa. A presença destes 5 aa correspondem a um inserto de 15 pb no DNA genômico em relação ao cDNA. Uma primeira análise foi realizada comparando-se 11 clones de cDNA com 3 seqüências de DNA genômico depositadas nos bancos de dados dos Institutos Sanger e TIGR (shisto5723h04.q1k, shisto3533e02.p1k, SNABH02TF) e com a seqüência do plasmídeo contendo a seqüência codificante completa de SmFes. Os 15 pb estavam presentes em todas as seqüências genômicas depositadas e ausentes nas demais. Uma amplificação desta região do DNA genômico de *S. mansoni* com os

iniciadores TK4 e “4” gerou bandas de 172 pb e 187 pb, que correspondem ao tamanho do fragmento com ou sem os 15 pb (Figura 23 B, canaleta 9). O sequenciamento dos fragmentos de DNA genômico amplificados com os iniciadores descritos acima identificou a existência das duas populações (Figura 22).

### **V.7.3 Polimorfismos do tipo SNPs no gene SmFes**

A análise de clones parciais do cDNA indicou a possibilidade de existirem polimorfismos de base única (SNPs), presentes nas bases: 1.698 (A-T), 2.055 (T-C), 2.293 (G-A), 2.596 (T-C) e 2.994 (A-G) do cDNA.

A análise dos clones de seqüências genômicas depositadas nos bancos de dados de *S. mansoni* indicou a presença de polimorfismos na região transcrita das seqüências, podendo ou não resultar em mudanças de aminoácido (Quadro 8).

## **V.8 IDENTIFICAÇÃO DE ORTÓLOGOS DE SMFES EM DIVERSAS ESPÉCIES DO GÊNERO *SCHISTOSOMA***

A amplificação de DNA genômico de diversas espécies de *Schistosoma* foi realizada utilizando pares de iniciadores desenhados para SmFes. Em todas as espécies, foi possível identificar a presença da banda com o tamanho esperado para cada par de iniciador, indicando que genes ortólogos a SmFes podem estar presentes nas diversas espécies de *Schistosoma* (Figura 23). A presença de duas bandas na amplificação com o par de iniciadores TK4 e “4” está correlacionada à presença do polimorfismo de 15 pb encontrado nesta região.

## **V.9 VIA DE SINALIZAÇÃO**

### **V.9.1 Proteínas de sinalização que participam da via de sinalização de SmFes**

Possíveis proteínas ortólogas às que Greer (2002) aponta na via de sinalização de Fes/Fps/Fer foram identificadas. A busca foi realizada a partir de SmAEs com anotações feitas pelo ONSA e ESTs com anotações feitas pelo geneDB de *S. mansoni*. As ESTs de *S. mansoni* depositadas pela rede ONSA e no GeneDB foram automaticamente anotadas. A correta classificação destas proteínas não foi verificada e comprovada, de modo que existem, desta maneira, possibilidades de existirem anotações incorretas. Algumas proteínas relatadas por Greer (2002) não foram encontradas nos bancos de dados. Apesar disso, estas proteínas podem estar presentes em *S. mansoni*, pois



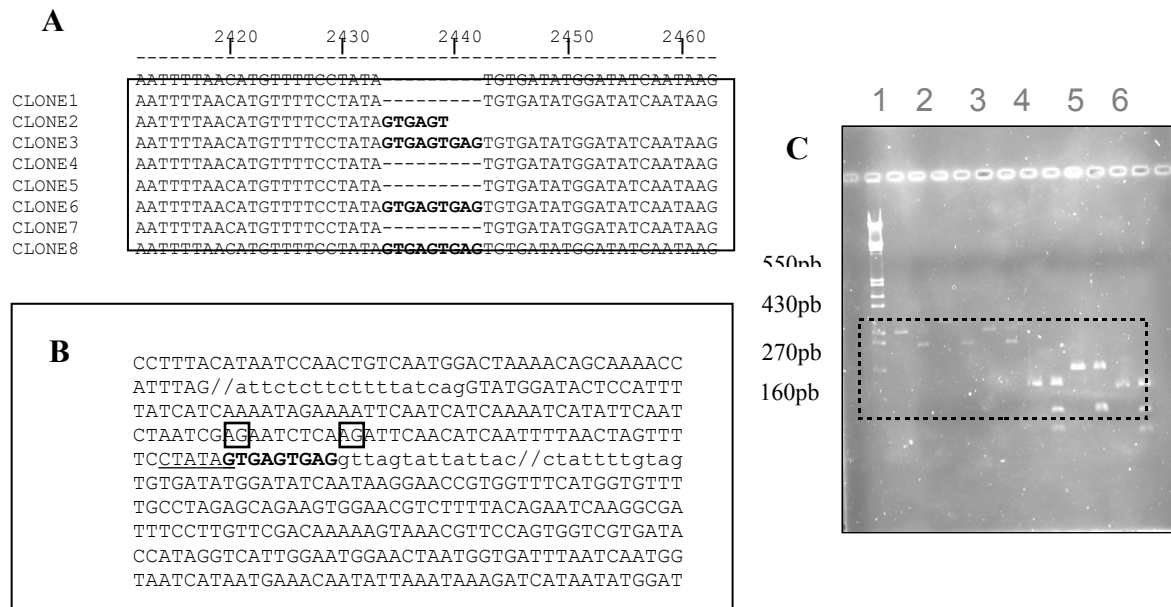


Figura 20: Edição Alternativa de SmFes. A- Sequência de 8 clones de cDNA de SmFes indicando a inserção dos 9 pb na posição 2.433. B- Sequência genômica da região onde ocorre o inserto de 9 pb. O inserto está representado em negrito. Os exons E-09, E-10 e E-11 estão representados em letras maiúsculas. Os introns I-10 e I-11 estão representados por letras minúsculas, sendo que as barras paralelas substituem a parte dos introns não representada. Os sítios doadores da edição alternativa estão indicados dentro de caixas. O sítio de corte da enzima *Sfc I* está sublinhado. C- Digestão de produtos de PCR obtidos de cDNA de verme adulto. As bandas visualizadas nas canaletas 1-2, 3-4 e 5-6 foram obtidas, respectivamente, da amplificação com os pares de iniciadores SmGOTK1-SmGOTK2, SmGOTK1-TK7 e SmGOTK1-TK6. Os cDNAs das canaletas 2, 4 e 6 foram digeridos com *Sfc I*. Os tamanhos dos cDNAs não digeridos são de 428 pb (canaleta 1), 547 pb (canaleta 3) e 423 pb (canaleta 5). Após a digestão, os fragmentos gerados foram de 159 e 269 pb (canaleta 2), 278 e 269 pb (canaleta 4), 154 e 269 pb (canaleta 6), indicados pela caixa. Os números à esquerda correspondem aos pesos moleculares correspondentes no gel.

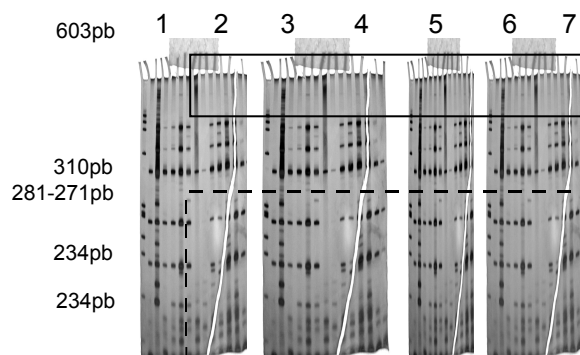


Figura 21: Digestão de produtos de PCR da região contendo os 9pb inseridos no cDNA. A amplificação foi realizada com o par de iniciadores SmGOTK1-SmGOTK2 e a digestão conduzida com a enzima de restrição *Sfc I*. 1- padrão de peso molecular  $\phi$ X174HaeIII, 2- grupo de vermes adultos, 3- grupo de vermes machos, 4- grupo de vermes machos, 5- grupo de cercárias, 6-verme fêmea individual, 7- verme macho individual. Os tamanhos dos cDNAs não digeridos são de 428 pb, indicados na caixa inteira. Os fragmentos gerados pela digestão com *Sfc I* são de 159 e 269 pb, indicados na caixa pontilhada. Os números à esquerda correspondem aos pesos moleculares em pares de base esperados.

2.041

<p>CONSENSO          PLASMIDEO de SMFES          cDNA (11 clones)          GENÔMICO (3 clones, GenBank)          GENÔMICO (seq)          GENÔMICO (seq)</p>	<pre> ----- CTGATATGTCATCTAGTATAGAACA<b>mCa</b>r-----CAACAACAACGTCAGAA CTGATATGTCATCTAGTATAGAACA<b>CCAG</b>-----CAACAACAACGTCAGAA CTGATATGTCATCTAGTATAGAACA<b>CCAG</b>-----CAACAACAACGTCAGAA CTGATATGTCATCTAGTATAGAACA<b>ACA</b>ACTACAACATCAGCAACAACAACGTCAGAA CTGATATGTCATCTAGTATAGAACA<b>ACA</b>ACTACAACATCAGCAACAACAACGTCAGAA CTGATATGTCATCTAGTATAGAACA<b>CCAG</b>-----CAACAACAACGTCAGAA </pre>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

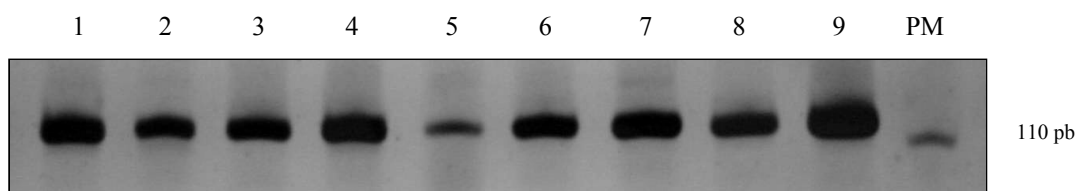
Figura 22: Inserção de 15 pb encontrados em clones de seqüências genômicas. Os 15 pb encontram-se inseridos após a base 2.040 do cDNA, no exon E-09. 11 clones de cDNA foram analisados, e nenhum continha a inserção de 15 pb. 3 clones genômicos depositados no GeneBank foram analisados e todos continham a inserção de 15 pb. Porém, outros clones seqüenciados por nós (seq) apresentaram ou não a inserção de 15 pb.

Quadro 8: Possíveis SNPs encontrados nas seqüências genômicas quando comparadas ao cDNA de SmFes.

Exon	Número de Seqüências Analisadas	Localização na Seqüência de cDNA	Códon cDNA	Códon Seqüência genômica	Aminoácidos codificados no cDNA/seqüência genômica	% de Seqüências genômicas com o alelo variante
1	2	552	TTA	TTG	L/L	100
2	6	-	-	-	-	0
3	3	-	-	-	-	0
4	5	968	ATT	AAT	I/N	100
		978	CAC	CAA	H/Q	100
5	3	-	-	-	-	0
6	2	-	-	-	-	0
7	6	-	-	-	-	0
8	6	1817	ACT	ATT	T/I	100
		1886	TTC	TGC	F/S	100
9	4	2043	CAC	CAG	H/Q	100
		2046	CAG	CAA	Q/Q	100
		2121	TTC	TTT	F/F	100
		2145	TCT	TCA	S/S	100
		2218	ATA	TTA	I/L	100
		2230	GTA	ATA	V/I	100
		2258	AGT	AAT	S/N	100
		2293*	GTC	ATC	V/I	100
10	6	-	-	-	-	0
11	3	-	-	-	-	0
12	26	2556	GGT	GGG	R/G	61
		2570	GGT	GCT	R/A	61
		2731	GAA	TAA	E/STOP	50
		2763	TCT	TCC	S/S	18
13	15	2791	CAC	CAT	H/H	33
		2831	TGT	TTT	C/F	26
14	40	2991	TCG	TCA	S/S	67
		3075	GCA	GCT	A/A	70
		3078	GTT	GTG	V/V	72
15	22	-	-	-	-	0
16	21	-	-	-	-	0
17	14	3360	CCT	CCA	P/P	64
		3372	ACG	ACA	T/T	100
		3408	ATT	ATC	I/I	28
		3420	GTA	GTT	V/V	35
		3468	GTA	GTT	V/V	42
		3516	TCT	TCA	S/S	28
18	10	3558	TTG	TTA	L/L	60

\* representa o SNP com correspondência não somente entre as seqüências genômicas como também entre os clones de cDNA.

A



B

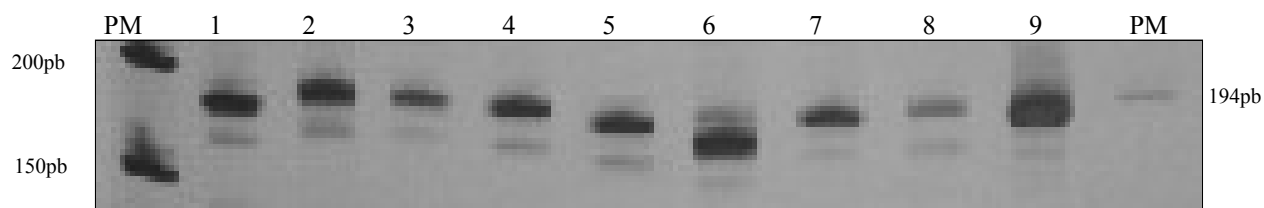


Figura 23: Amplificação do DNA genômico de diversas espécies do gênero *Schistosoma* com pares de iniciadores desenhados especificamente para SmFes. A- Amplificação das espécies com o par de iniciador P2for e “B”, nas canaletas: 1- *S. rodhaini*, 2- *S. mattheei*, 3- *S. margrebowiei*, 4- *S. mansoni*, 5- *S. japonicum*, 6- *S. intercalatum*, 7- *S. haematobium*, 8- *S. curassoni* e 9- *S. bovis*; B- Amplificação das espécies com o par de iniciador TK4 e “4”, nas canaletas: 1- *S. magrebowiei*, 2- *S. bovis*, 3- *S. intercalatum*, 4- *S. haematobium*, 5- *S. japonicum*, 6- *S. rodhaini*, 7- *S. curassoni*, 8- *S. mattheei* e 9- *S. mansoni*.

as ESTs até então depositadas podem não ter sido anotadas ou mesmo seqüenciadas. Como grande parte das proteínas relatadas na via proposta por Greer (2002) foram descritas para *Schistosoma*, pode-se considerar esta como uma provável via de sinalização para a proteína SmFes (Figura 24 e 25) (Quadro 9).

Além da via proposta, Greer (2002) também relata, algumas possíveis parceiras diretas de Fes/Fps/Fer. As parceiras diretas estariam interagindo com os domínios PK, SH2 e *coiled-coil* de SmFes. Dentre as proteínas propostas como parceiras de SmFes foram encontradas em *S. mansoni*: cortactina, EGFR, PDGFR, PI3K, BCR,  $\beta$ -catenin, PTPs assim como também outras PKs como SRC e JAK (Figura 24 e 25) (Quadro 9).

### **V.9.2 Clonagem e Expressão de Proteína para Ensaio de Pull-Down**

A proteína recombinante contendo os domínios SH2 e SH2PK, com ou sem o inserto de 9 pb, foram clonadas em plasmídeo pGEX-4T3, que foram transformados em células BL21 e expressas em fusão com a proteína GST. As proteínas de fusão expressas apresentaram o tamanho esperado de 45 kDa para fragmento SH2 e 76 kDa para SH2PK. As proteínas recombinantes foram expressas em diluição de 1mM, 5mM e 10mM de IPTG. As seqüências dos clones foram verificadas por sequenciamento, confirmando que a proteína recombinante não apresenta nenhuma troca, deleção ou inserção de aminoácidos, mantendo, desta maneira, os mesmos aminoácidos presentes na proteína SmFes (Figura 26).

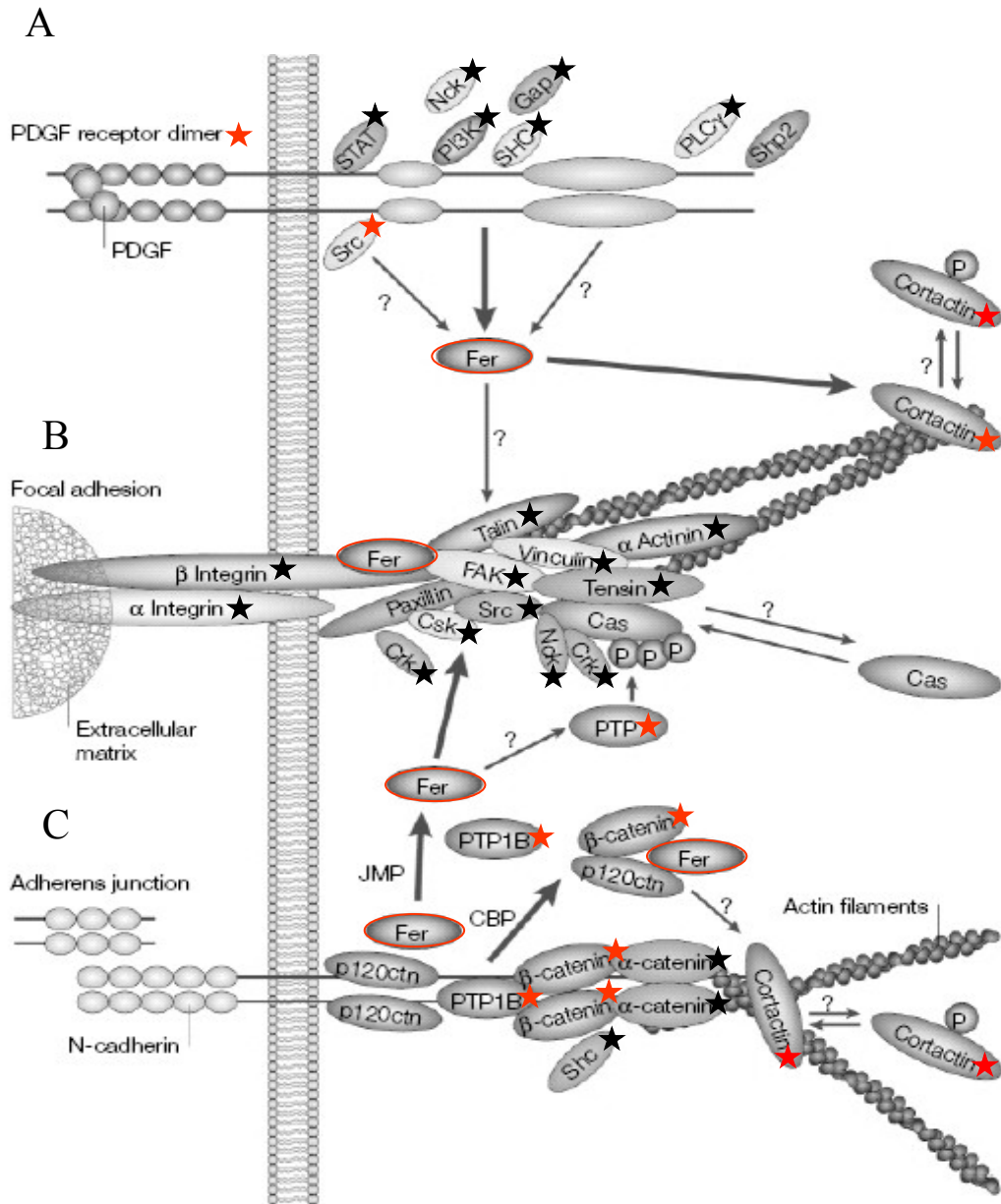


Figure 24: Via de sinalização proposta por Greer (2002) para a proteína Fer (circulada em vermelho). A estimulação dos receptores PDGF leva à ativação de Fer, diretamente ou através de uma Src. A ativação de Fer leva à fosforilação da cortactina que induz mudança no citoesqueleto e na regulação de integrinas e caderinas. A dissociação de Fer de n-caderina leva o direcionamento de Fer para o complexo FAK-p130Cas (FAK, p130Cas, Crk, Nck e tensina), em que Fer estaria diretamente relacionada com a ativação de uma PTP (proteína tirosina fosfatase). As ESTs encontradas nos bancos de dados de *S. mansoni*, classificadas por *Gene Ontology* e identificadas por similaridade com às correspondentes proteínas da via Fer estão indicadas por uma estrela. As estrelas vermelhas indicam possíveis parceiros diretos de Fer encontrados em *S. mansoni*. (adaptada de Greer 2002).



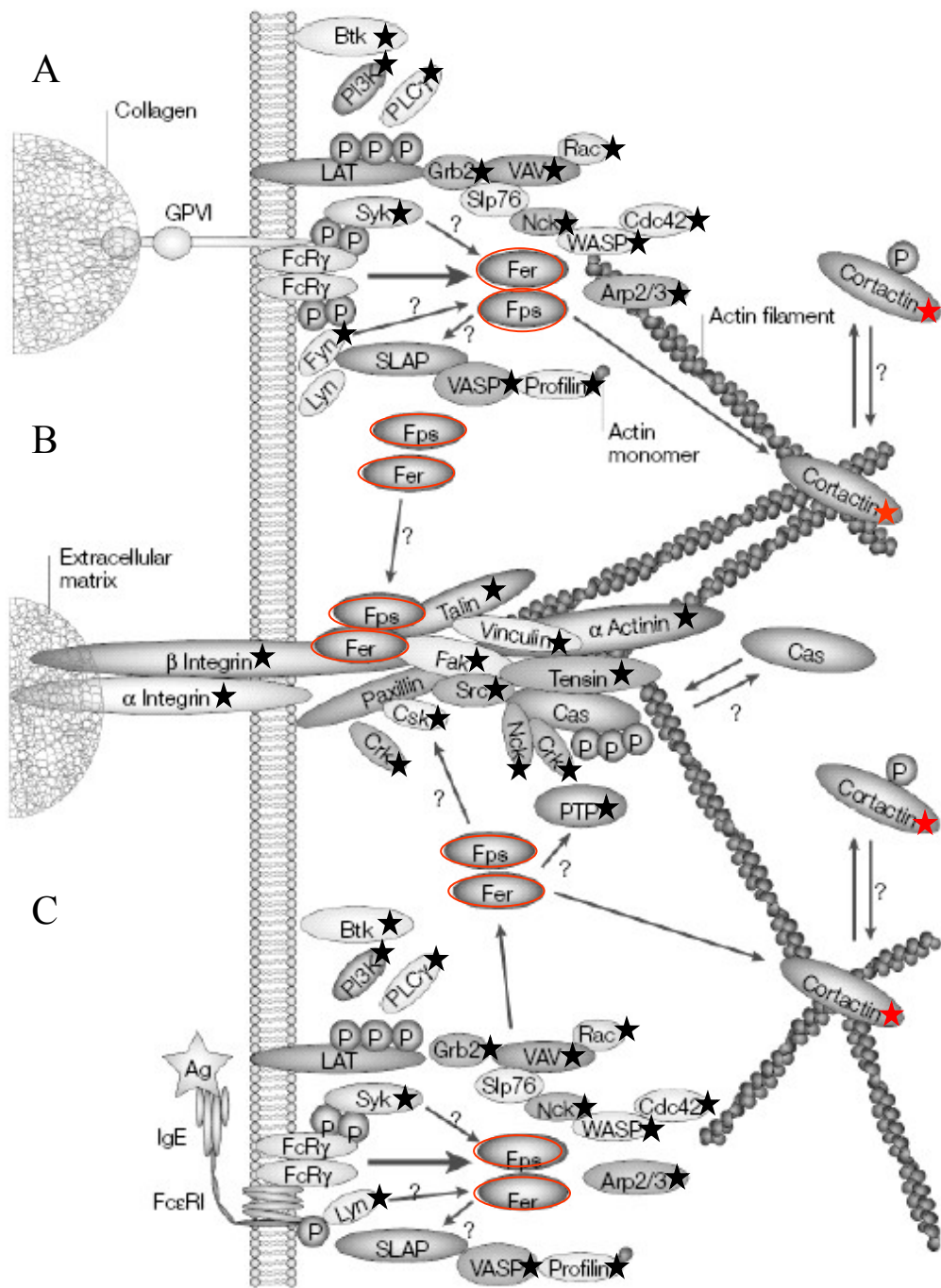
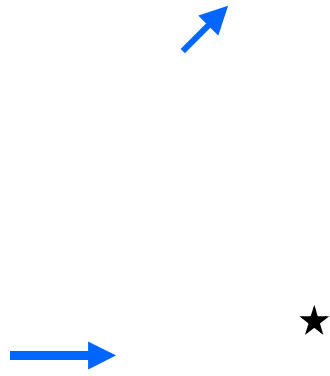


Figura 25: Via de sinalização proposta por Greer (2002) para a proteína Fes/Fps (circulada em vermelho). A estimulação dos receptores de colágeno e de FcεRI leva a ativação de Fes/Fps/Fer podendo fosforilar alguns substratos como cortactina que induz mudança no citoesqueleto e na regulação da integrina. ESTs encontradas nos bancos de dados de *S. mansoni*, classificadas por *Gene Ontology* e identificadas por similaridade com as correspondentes proteínas da via Fes/Fps, estão indicadas por uma estrela. As estrelas vermelhas indicam possíveis parceiros diretos de Fes/Fps encontrados em *S. mansoni*. (adaptada de Greer 2002).





Quadro 9: Relação de proteínas descritas para a via de sinalização de Fes/Fps/Fer encontradas nos bancos de dados de *Schistosoma mansoni*.

<b>Proteínas da via de sinalização de Fps/Fes/Fer</b>	<b>SmAE (ONSA)</b>	<b>ESTs e Phat (GeneDB)</b>	<b>Referência</b>
$\alpha$ -Actinin	600771-602245-602669-604615-604882-604957-605249	Sm 03500-02819-14525 Phat 10029	
<b><math>\beta</math>-Catenin</b>	600257-607405	Sm 08216-02974-10658	
$\alpha$ -Catenin	608838	Sm 13235 Phat01873	
$\beta$ -Integrin	602256-611319-713296	Sm 02377-01202-26282-08335 Phat 00026	
$\alpha$ -Integrin	609598-612087-715651	Sm 02922-24983	
Arp2/3	612401-707765	Phat 10032	
<b>BCR (break point cluster region)</b>		Sm 16385	
Btk (Bruton's tyrosine kinase)		Phat 02986	
Cas (Crk-associated substrate)			
Cdc42	606042-603521-602720	Sm 02565-20398-00523-06284-21384-03100-28791-11782	
<b>Cortactin</b>	703413-715747	Sm 00383	
Crk	606883	Sm 12123	
Csk (carboxy-terminal Src kinase)	604447	Sm 08707	
<b>EGFR (epidermal growth factor receptor)</b>	603132-712966-609547-718789-703390-708456	Sm 25436-05221-12212-19228-02237-17026-27596-03864 Phat 05276	(Shoemaker <i>et al</i> 1992)
FAK (Focal adhesion kinase)	6003559-601801-714481		
Fc $\epsilon$ RI (high affinity IgE receptor)			
Fc $\gamma$ R (high affinity IgG receptor)			
Fps/Fes	714311	Sm 27185-24981 Phat 11720-11350-11782-09892	
<b>FybSLAP130</b>			
Fyn		Sm 00996-11598	
Gap (Gtpase activator)	Múltiplas SmAEs (#43) GO:0005096	Sm 25638-07339-07094-01238-04681-02358-20059-26569	
GPVI (Glycoprotein VI) collagen receptor			
Grb2 (Growth factor receptor bound protein)	605828-608523	Sm 08751-01574-04606	
<b>IRS-1 (insulin receptor substrate 1)</b>			
<b>Jak (janus kinase)</b>		Sm 14275-14224	
Lat			
Lyn		Sm 21063	
n-cadherin		Sm 20267-25451 Phat 12236	
Nck	714096	Sm12060-23699 Phat 10035	
<b>p120ctn (catenin)</b>			
<b>p120RasGap</b>	604153-710788	Sm 25638-13712-07094	
<b>p130Cas</b>			
Paxillin			
<b>PDGF (platelet-derived growth receptor)</b>	-	Sm16530-07195	

<b>PI3K (phosphatidylinositol 3-kinase)</b>	Múltiplas SmAEs (#29) GO: 0016303	Sm27004-07904-09855-04768 Phat 07509	
PLC $\gamma$ (phospholipase C $\gamma$ )	PLC- 601600-609898-605065-704186	Sm 23100	
Profilin	712405	Sm 22625	
PTP (protein tyrosine phosphatase)	GO: 0004725	GO: 0004725	
PTP1B	604311	Sm 22625-06571	
Rac	610098-612269-705063	Sm 01001-06722-19233 Phat 06280	
SHC (SH2 containing)		Sm 13282-21895-02100-13052-07213-04883-11851-06127-04070-12123-04994-03003-20506-23094-11322-05205-03877-15263-08707-26098-23051-08339-25863-19964-07974-05518-27829-27370-11598-11809-10341 Phat 10035-12132-02985-11399-02985-11399-02690	
<b>Shp2 (SH2 contain protein Tyrosine phosphatase)</b>			
SLAP			
Slp76			
Src*	604149	Sm 03422-13275-27829-04488 Phat12731-02012-12730	
STAT (signal transducer and activator of transcription)		Sm 04313	
<b>STAT3</b>			
<b>STAT5A</b>			
Syk			
Talin	604429-604964-605118-605741-610256-707152-711066-711409	Sm 19942-34444-11470-02332-25900 Phat 1371813720	
Tensin	600012-600238-710464-715289		
VASP (Homer)	714496		
VAV	717146-718598-601993	Sm 05205-24424	
Vinculin	702650-708493	Sm 21115-10823	
WASP (Wiscott-Aldrich Syndrome protein)		Sm 18502-28657	

As proteínas descritas por Greer (2002) na via de sinalização proposta e os prováveis substratos de Fes/Fps/Fer estão relatados na primeira coluna. As prováveis proteínas de ligação direta com SmFes estão em negrito. O prefixo Phat indica a predição de genes em um contexto genômico e o prefixo Sm indica ESTs. As proteínas que já descritas na literatura estão indicadas pela referência bibliográfica.

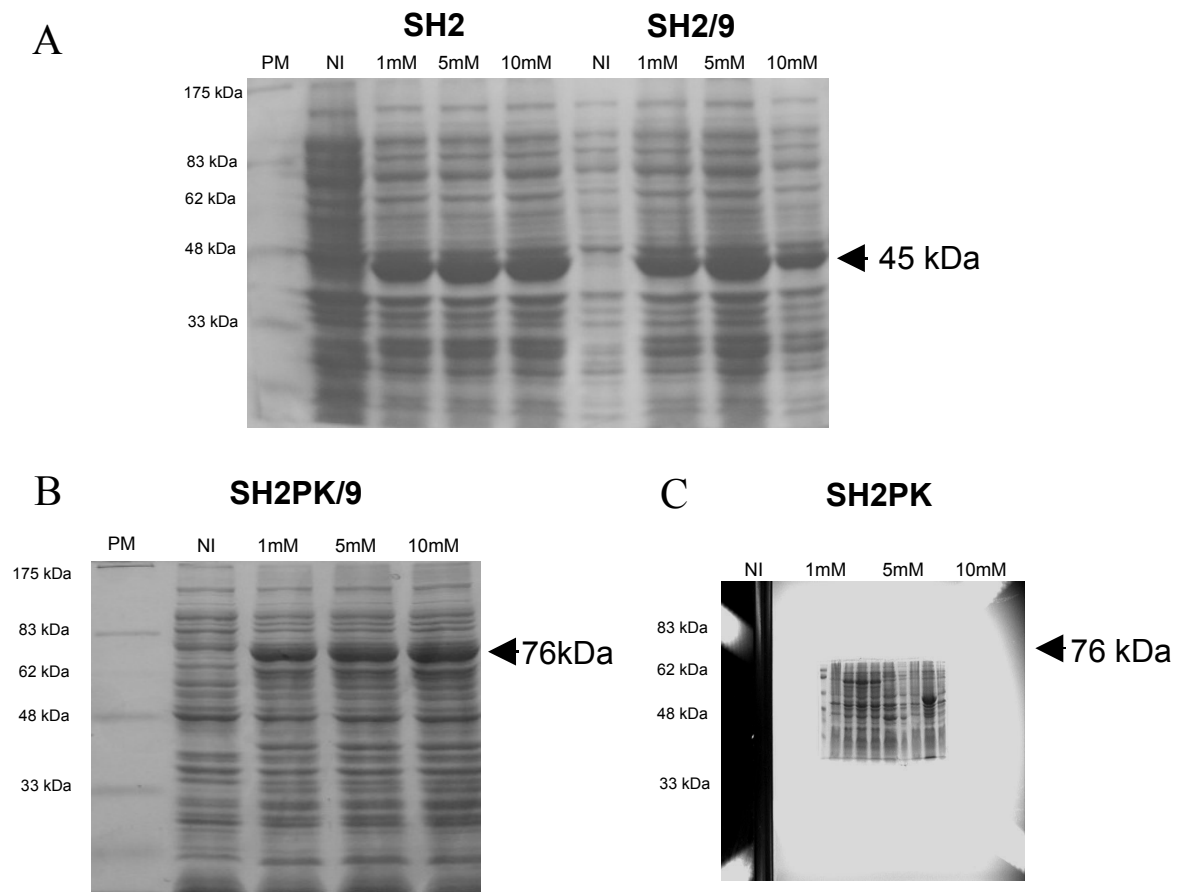


Figura 26: Expressão de proteína de fusão de fragmentos de SmFes dos domínio SH2 (A- peso esperado 45 kDa) ou dos domínios SH2 e PK (B e C peso esperado 76 kDa), contendo (SH2-9 e SH2PK-9) ou não (SH2 e SH2PK) os 9 pb inseridos, clonados em pGEX-4T3 e expressos em células *E. coli* BL21. A expressão foi verificada em gel SDS-PAGE 12% e o gel corado com azul de Coomassie. O peso molecular (PM) é observado na primeira canaleta, em A e B. Os tamanhos em kDa estão indicados ao lado de cada gel. NI corresponde ao extrato bacteriano proveniente de reações em que não houve indução com IPTG (controle negativo). Nas demais canaletas estão relatadas as diferentes concentrações finais de IPTG 1, 5 e 10mM, usadas na indução dos extratos bacterianos.

## **VIDISCOSSÃO**

## VI.1 CLASSIFICAÇÃO ESTRUTURAL

SmFes é o primeiro gene da família Fes/Fps/Fer das PTKs (Proteínas Tirosinas Quinase) identificado em *S. mansoni*. Fps/Fes e Fer são os únicos membros conhecidos de uma subfamília distinta de proteínas não-receptoras/citoplasmática pertencentes à família de PTK. Fps/Fes (Fujinami sarcoma de galinha/sarcoma de felino) foi primeiramente isolada como oncogenes de tumor causado por retrovírus de aves e felinos. Fer foi identificada durante a busca por homólogos celulares da proteína Fps/Fes, já que são estruturalmente e antigenicamente relacionadas. PTKs c-Fps/Fes e c-Fer (Fps/Fes e Fer celular) consistem em uma seqüência N-terminal longa constituída pelo domínio FCH (Fps/Fes/Fer/CIP4), seguida por três regiões atribuídas como *coiled-coils*, um domínio SH2 (Src-homology-2) e um domínio catalítico C-terminal.

Análises estruturais indicam que SmFes exibe características da família Fes/Fps/Fer por conter uma região com três *coiled-coil*, uma assinatura de domínio SH2 e uma de domínio catalítico proteína tirosina quinase C-terminal. No entanto, nenhuma assinatura de domínio FCH N-terminal foi identificada em SmFes.

O domínio catalítico consiste de onze subdomínios separados por regiões pouco conservadas. Os aminoácidos conservados do subdomínio são importantes na função catalítica, tanto como componente direto do sítio de ativação como também indiretamente, contribuindo na formação do sítio da ativação em sua estrutura secundária. O consenso Asp-Leu-Ala-Ala-Arg-Asn do subdomínio VI e o consenso Pro-Ile/Val-Lys/Arg-Trp-Thr/Met-Ala-Pro-Glu do subdomínio VIII são assinaturas de proteínas tirosinas quinase. O consenso de ambos subdomínios presentes em SmFes são consistentes com os observados nas PTKs. Todos os demais subdomínios foram observados em SmFes nas posições esperadas. Estas observações sugerem que o domínio catalítico de SmFes mantém uma estrutura funcional deste domínio. Sítios de autofosforilação em resíduos de tirosina podem regular a atividade quinase do domínio e também recrutar proteínas com domínio SH2. Sítios de autofosforilação em resíduos de tirosina foram determinados no domínio catalítico de c-Fes humana (P07332) nas posições 713 e 811. Neste modelo, a autofosforilação do resíduo de tirosina 713 mostrou influenciar a atividade quinase da proteína, enquanto o resíduo 811 demonstrou estar relacionado com o recrutamento de proteínas por SH2. A autofosforilação destes resíduos parece ser eventos intermoleculares supondo um modelo como ocorre para os RTKs de fatores de crescimento e para as NRTKs associadas a receptores de citocina. No alinhamento múltiplo do domínio catalítico de SmFes e de proteínas ortólogas de outros organismos foi observado que o resíduo de tirosina correspondente ao aa 713 de c-Fes humana encontra-se

presente na posição 1.113 de SmFes, indicando que este resíduo mantém a atividade quinase da proteína na fosforilação do substrato de SmFes. Já o resíduo de tirosina relacionado ao aa 811 de c-Fes humana não foi observado em SmFes e nem em várias outras proteínas ortólogas, indicando que esse resíduo não desempenhe uma função essencial para a atividade quinase mas, como visto apenas recrutando proteínas.

O domínio SH2 reconhece tirosinas fosforiladas e capacita as proteínas que o possui a se ligarem aos receptores proteínas tirosinas quinase ativadas, assim como a outras proteínas sinalizadoras intracelulares que são fosforiladas em tirosinas transitoriamente . O reconhecimento de sítios tirosina fosforilados pelo domínio SH2 é específico, sendo que esta especificidade é conferida pelos três resíduos C-terminais imediatamente após a fosfotirosina . A triagem de biblioteca de fosfopeptídeos, usando o domínio SH2 de Fes/Fps como matriz de afinidade, identificou proteínas com um consenso de ligação YEXV/I. Esse consenso encontra-se presente em muitas proteínas celulares como: FAK (quinase de adesão focal), Tec, Lyn, Abl (homólogo ao oncogene de leucemia viral Abelson), Hck (célula de linhagem hematopoiética), CD72, CD3ε, SHPTP-1 (proteína tirosina fosfatase 1 contendo o domínio SH-2), LAR-PTP (proteína tirosina fosfatase relatada de antígeno de leucócito), ezrin, BCR (*break point cluster region*), 3BP2A e adaptin-γ. Embora muitos parceiros prováveis de ligação do domínio SH2 de Fes/Fps/Fer tenham sido propostos, existe muito pouca evidência bioquímica acerca destas supostas interações . Partindo do pressuposto que o domínio SH2 reconhece tirosina fosforilada e sabendo que SmFes apresenta um possível sítio de fosforilação em tirosina no aa 1.113 com os três aa C-terminais E-L-S (glutamato-leucina-serina) espera-se que SmFes atue no recrutamento de proteínas que contenham o domínio SH2 que reconhece este consenso de ligação especificamente.

As regiões de *coiled-coil* parecem mediar a formação de oligômeros: homotrímeros em Fer e pentâmeros/oligômeros de ordem maior em Fps/Fes . O domínio *coiled-coil* pode ter outras funções além de formar oligômeros como: modular interação entre os domínios SH2 e catalítico, ou interagir com outras proteínas que contenham o domínio *coiled-coil* . Um modelo regulatório foi proposto para Fes/Fps, em que o domínio *coiled-coil* regularia a conversão entre monômeros inativos e oligômero ativos. . Esta regulação envolveria interações intramoleculares mantendo Fes/Fps na forma inativa e relações intermoleculares homotípicas permitindo a ativação em oligômeros por transfosforilação requerida para a atividade quinase. . A oligomerização em Fer parece não ter efeito na atividade da autofosforilação, indicando que há uma diferença na regulação de Fes/Fps e Fer . Foi demonstrado que o domínio N-terminal de Fer pode mediar a ligação com a proteína de junção de aderência p120 catenina que também foi predita como sendo uma proteína com

domínio N-terminal *coiled-coil* pela oligomerização heterotípica entre elas . Visto que SmFes possui três regiões *coiled-coil* N-terminal, é possível que SmFes assuma conformações oligoméricas intramoleculares, homotípicas ou heterotípicas.

O domínio FCH foi encontrado em proteínas que estão envolvidas na regulação do rearranjo do citoesqueleto, do transporte vesicular, da endocitose e da associação de microfilamentos . Fes naturalmente assume uma conformação inativa *in vivo* em *S. cerevisiae*. Nesse modelo, a regulação da ativação requer os domínios *coiled-coil* e SH2, já que clones possuindo apenas estes domínios demonstraram regular a atividade da quinase, mantendo-a inativa ou ativando-a . Esta observação sugere que a ausência do domínio FCH poderia não interferir também na atividade catalítica de SmFes. Em *C. elegans*, 42 dos 52 genes que codificam para NRTKs, foram classificados como pertencentes a subfamília Fer. No entanto, nenhuma destas NRTK preditas contém o domínio FCH ou as regiões *coiled-coil* que caracterizam a família Fes/Fps/Fer. Não é evidente porque estas proteínas foram agrupadas na família Fes/Fps/Fer preferencialmente a outras PKs que possuem o domínio SH2 . Um estudo mais detalhado poderia redistribuir as NRTKs em diferentes famílias. Em camundongos, foi relatada a presença de variante de Fer, FerT, que não possui os domínios FCH e *coiled-coil*. Essa variante está relacionada a espermatogênese, localizando-se especificamente nos testículos . Assim como as proteínas de camundongos e *C. elegans*, SmFes foi preferencialmente agrupada à família Fes/Fps/Fer, mesmo na ausência do domínio FCH que as caracterizam.

Um gene pertencente à família Fes/Fps/Fer também foi relatado em esponjas marinhas (Porífero), o organismo metazoário mais primitivo . Esta observação sugere que *fer* sofreu uma mudança evolucionária de um gene progenitor, que codificava uma PTK contendo o domínio SH2, e que depois foi adquirindo domínios adicionais como os amino-terminais: *coiled-coil* e depois FCH . O gene progenitor *fes/fps* pode ser resultado da subsequente duplicação do gene *fer*, que pode ter evoluído para a formação de um gene parálogo tecido específico que é visto em mamíferos . Por análise comparativa da filogenia gerada para diferentes famílias de proteínas e genes foi mostrado que uma proteína têm taxas de evolução distintas e que uma mesma proteína pode-se evoluir mais rapidamente em alguns organismos que em outros . Uma suposta explicação para a presença de FCH em esponjas e a ausência em *S. mansoni* pode ser derivada da aquisição independente do domínio de FCH pelas esponjas em um momento distinto das demais Fes/Fps/Fer ou pela perda de FCH ao longo da evolução de *S. mansoni* atual.

A reconstrução filogenética foi realizada para confirmar SmFes como um membro da família Fes/Fps/Fer e para propor uma distância evolutiva entre proteínas desta família de



diversos organismos. O resultado de ambas as análises filogenéticas, utilizando as seqüências completas ou somente os domínios SH2-PK, indicaram uma relação de SmFes com proteínas pertencentes a família Fes/Fps/Fer de organismos invertebrados, o que sugere que SmFes está próxima ao possível gene progenitor. A posição de *C. elegans* na árvore filogenética sugere que a proteína possa ter apresentado uma maior taxa de evolução para a região que não representa domínios, porém não existem informações funcionais sobre a molécula que indiquem se houve uma diferenciação na via de sinalização em que atua.

## VI.2 ESTRUTURA GENÔMICA

A presença de SmFes no genoma de *S. mansoni* foi confirmada através da análise de hibridização contra uma biblioteca de BAC arranjada em filtros, utilizando-se uma sonda do fragmento da porção codificadora do domínio SH2 de SmFes. Com o uso da sonda específica para SmFes, foram selecionados 3 clones positivos de BAC (dados não descritos). Considerando que os BACs contidos no filtro inicial cobriam aproximadamente 7,9 vezes o genoma de *S. mansoni*, concluiu-se que uma única cópia de SmFes estaria presente no genoma de *S. mansoni*. A análise de *Southern-blot* realizado com o DNA genômico de *S. mansoni* confirmou que SmFes é um gene de cópia única em *S. mansoni*.

O banco de dados público de *S. mansoni* gerado a partir do projeto genoma de *S. mansoni* forneceu uma fonte de dados importante para estudos genômicos. O número de introns encontrado em SmFes é similar ao número de introns já descrito para o gene humano pertencente a família Fes/Fps/Fer. O gene completo de Fes/Fps humano possui um *locus* de 13 Kb, contendo 18 exons. Em contraste ao arranjo relativamente compacto descrito para Fes/Fps, o *locus* de Fer possui, pelo menos, 500 Kb. SmFes apresenta introns consideravelmente longos, o que também já foi relatado em outros genes de *S. mansoni*. Dos 50 Kb do gene SER caracterizado, menos de 40% do transcrito maduro é codificado. O gene da protease aspártica de *S. mansoni* possui um *locus* de, aproximadamente, 13 Kb que inclui seis introns de tamanhos entre 30 e 5.025 pb. Outro gene que foi identificado e caracterizado pelo nosso grupo, SmPKC1, utilizando a mesma metodologia de identificação de introns de SmFes, possui um *locus* de 15,3 Kb, que inclui doze introns de tamanhos entre 32 e 3.234 pb (dados não publicados). Até o momento não é possível confirmar o tamanho do *locus* de SmFes porque ainda faltam seqüências intrônicas a serem identificadas no projeto genoma; no entanto já se é esperado um *locus* maior do que o caracterizado para Fes/Fps. A seqüência conhecida do *locus*, atualmente, é de 30,8 Kb.

O alinhamento genômico realizado neste trabalho e os contigs formados pelo banco de dados do Instituto Sanger são idênticos na localização e número de introns presentes; no

entanto, a seqüência apresentada pelo Sanger para o intron I-17 foi menor que o relatado por nosso grupo. A comprovação do tamanho do intron por amplificação por PCR demonstrou que a montagem realizada pelo nosso grupo está correta. A montagem realizada pelo Instituto Sanger é realizada de uma forma ampla em que detalhes podem não ser percebidos, enquanto a nossa montagem é manualmente analisada podendo-se corrigir pequenas imperfeições apesar de não terem sido considerados os valores de qualidade das bases. Devido ao grande tamanho e quantidade de seqüências repetitivas presente no genoma de *S. mansoni*, a sua organização e o seu alinhamento tornam-se um processo complexo .

### VI.3 POLIMORFISMOS

A presença da inserção de 9 pb foi observada em diferentes clones de cDNA, sugerindo que fossem alelos distintos, genes distintos ou edição alternativa do transcrito. A análise da estrutura genômica permitiu a classificação da inserção como edição alternativa do tipo “exon extension/truncation at 5' end”, visto que a existência do sítio doador GT alternativo localizado a 5' do intron I-10 permite que existam duas formas possíveis do transcrito. O gene Ser, uma outra PTK de *S. mansoni*, possui três transcritos distintos provenientes de edição alternativa . A proteína SmHSF (heat-shock transcription factor of *S. mansoni*) possui isoformas provenientes de edição alternativa estágio específicas . Na família Fes/Fps/Fer foram identificadas várias proteínas com sua forma variante proveniente de edição alternativa. A proteína Fer de camundongo possui 94 kDa e é amplamente expressa em células somáticas, enquanto sua variante FerT de 51 kDa, não contendo o domínio FCH e *coiled-coil*, é encontrada em células em metáfase testículo-específicas . iFer é uma outra proteína variante que codifica uma proteína sem função quinase mas que mantém o domínio regulatório e possui a habilidade de formar oligômero com outras Fer, possuindo 65 pb deletados em relação à janela de leitura de Fer .

A forma variante da proteína SmFes pode ser tecido específica, ter uma localização celular específica ou mesmo estar presente em fases de desenvolvimento específicas. Alterações no ambiente intracelular pode levar a uma escolha de um sítio de corte ou de outro. A priori não se pode prever qual isoforma será a mais abundante, experimentos adicionais devem ser realizados, o que pode ser muito interessante, porque mesmo uma pequena alteração de 3 resíduos na proteína pode ser importante para sua atividade, regulação, etc. Experimentos de modelagem molecular *in silico* das duas isoformas poderiam prover informações adicionais acerca da possível alteração conformacional em presença dos três resíduos e, conseqüentemente, possível modulação da função. Também pode ser muito interessante elucidar como se regula a seleção de um sítio ou de outro.

Uma inserção de 15 pb foi identificada na seqüência de DNA genômico, e ausente nas seqüências de cDNA. Com a amplificação de DNA genômico, foram observadas duas bandas de peso molecular próximo ao esperado, de 172 e 187 pb. O sequenciamento desta amplificação comprovou a existência da duas populações em DNA genômico. Nenhum elemento de edição alternativa foi encontrado na inserção de 15 pb e sendo SmFes um gene de cópia única é possível que este polimorfismo seja proveniente de alelos distintos. Polimorfismos de alelo específicos foram encontrados na proteína quinase Fyn-like de *S. mansoni*, em que os cDNAs diferem em 6 asparaginas dentro de uma região com 13 asparaginas sem alterar a janela de leitura .

Em *S. mansoni* foram identificados em ESTs uma média de 1 SNP para cada 1.606 pb. SNPs foram identificados e validados em catepsina B de *S. mansoni* através de sequenciamento de vários clones . Muitos SNPs não foram observados nos clones de cDNAs contendo a seqüência de SmFes, embora tenham sido observados em seqüências genômicas. Trabalhos experimentais serão necessários para a validação da real existência destes SNPs.

#### **VI.4 NÍVEL DE EXPRESSÃO E FUNÇÃO DEDUZIDA**

O diferente perfil de expressão de alguns genes ou proteínas tem sido demonstrado nos diversos estágios de desenvolvimento do parasito . Embora SmFes apresente-se expressa em diversos estágios de *Schistosoma*, a proteína apresentou um maior nível de expressão em cercária. Transcritos de SmFes foram identificados por *northern-blot* apenas em vermes machos e esquistossômulos. No entanto, não se pode fazer uma análise quantitativa deste *northern-blot*, uma vez que a membrana não foi hibridizada a uma sonda correspondente a um gene constitutivo de *S. mansoni*. Não obstante, as análises de PCR em tempo real, que é um ensaio mais sensível, mostraram que SmFes é transcrita em cercária, miracídio, esporocisto, verme fêmea e em maior nível em verme macho. A quantificação relativa dada pelo PCR em tempo real através do valor ct obtido de SmFes e do valor ct obtido de tubulina indicou que SmFes tem baixos níveis de transcrição em *S. mansoni*, o que justificaria a difícil análise em *northern-blot*.

Os perfis de transcrição e expressão observados em *S. mansoni* variam de acordo com o ciclo evolutivo do parasito para várias proteínas envolvidas em sinalização. O receptor nuclear FTZ-F1 e SmRXR é transcrito nas diversas fases de desenvolvimento do parasito, mas em maiores níveis nas formas larvais de vida livre. No entanto, a proteína FTZ-F1 é expressa em baixo nível em miracídio e alto em verme adulto macho, e SmRXR é expressa em maiores níveis em esquistossômulo . SmFes apresentou um diferente perfil entre os níveis

de transcrição e expressão, com maior nível de transcrição em macho e de expressão em cercária. No momento não é possível interpretar funcionalmente esta observação.

A única pista para a elucidação da função de qualquer membro da família *Src* é derivada de descrições de tipos celulares em que a expressão é maior. É esperado que membros de uma mesma família, que compartilham domínios catalíticos com divergência limitada um do outro, desempenhem um papel similar na fisiologia celular. Embora nenhuma literatura relate uma função para as proteínas quinase da família *Fes/Fps/Fer* em vermes, há indícios de que os ortólogos de mamíferos, *Fer* e *Fes*, desempenhem um papel na adesão celular, na via de sinalização de receptores de citocinas e que tenham funções como oncogenes. A regulação de uma variante de *Fer*, *FerT*, durante a espermatogênese indica que esta proteína pode estar envolvida na regulação da divisão celular por meiose (WANG, 1994). Embora pareça que *Fes/Fps/Fer* não sejam essenciais para o desenvolvimento, análises genéticas indicam que estas proteínas quinase estariam fortemente implicadas nas vias de sinalização que são importantes para a regulação de vários processos biológicos. Entre os processos, incluem-se interação entre célula-célula e célula-matriz durante a migração celular, ativação de plaquetas, e adesão e extravasamento em células do sistema imune em sítios de inflamação. No momento, não é possível interpretar funcionalmente *SmFes*. Experimento de imunolocalização de *SmFes* no parasito poderia sugerir possíveis funções a proteína.

Mesmo que as proteínas *Fes/Fps/Fer* não demonstrem ser essenciais individualmente, *SmFes* pode participar de uma via de sinalização essencial ou a perturbação da sua ação pode levar a uma desestruturação da homeostase do parasito que facilite a sua eliminação.

## VI.5 VIA DE SINALIZAÇÃO

*Fps/Fes* e *Fer* estão envolvidas nas vias de sinalização de vários receptores de citocinas, fatores de crescimento e imunoglobulinas. No entanto, o envolvimento exato na via de sinalização destes receptores ainda não é bem compreendido.

A presença dos domínios SH2 e PK em *SmFes* possibilitam a interação desta molécula a outras proteínas quinase que contenham estes domínios. Em sistema heterólogo de expressão foi possível demonstrar que PTK da subfamília *Src* fosforila proteínas *Fes/Fps*. A proteína *Src* de *S. mansoni*, TK3, foi isolada e caracterizada, sugerindo-se sua participação na via de sinalização celular em resposta a organização celular. Em experimento de duplo híbrido entre o domínio SH3 e único da proteína TK3, e uma biblioteca de *D. melanogaster*, foram identificados os seguintes parceiros de TK3: *dAbi*, envolvida na arquitetura do citoesqueleto, *vinculina* e *tubulina-β*, ambas envolvidas na adesão focal. Testes funcionais demonstraram que TK3 fosforila o substrato *p130Cas*. *Fer* foi, recentemente, implicada na

regulação de junções de aderência e adesão focal e tem como possíveis substratos p120catenina,  $\beta$ -catenin, p130Cas, Fyb/SLAP130, cortactina, EGFR, PDGFR, PI3K, IRS-1 (substrato 1 do receptor de insulina), BCR, p120RasGAP, STAT3, STA5A, SHP-2, PTP (proteínas fosfatases), assim como também poderiam estar interagindo com outras PKs, como SRC e JAK. . Frente a possível participação de Src na ativação de Fes, substratos similares e a mesma resposta envolvida na ativação de ambas as proteínas é possível que TK3 esteja atuando na ativação de SmFes em *S. mansoni*. A presença da proteína já bem caracterizada, SER, proteína de *S. mansoni* ortóloga ao EGFR, indica uma outra possível parceira direta de SmFes. Essas observações podem ser comprovadas experimentalmente por co-expressão em sistemas heterólogos, co-precipitação ou co-imunoprecipitação ou ainda com o uso de sistemas de duplo híbrido e *pull-down*.. Um clone da proteína TK3 já foi gentilmente cedido ao nosso grupo pelo Dr. Grevelding e será usando em futuros experimentos de interação entre SmFes e TK3. A colaboração com outros grupos que já possuam as proteínas, propostas como possíveis substratos de SmFes, já bem caracterizadas podem ajudar na elucidação da via em que SmFes participa.

Experimentos com camundongos transgênicos contendo a proteína Fer mutante de rápida degradação demonstraram que os mesmos são viáveis e férteis na ausência de Fer. No entanto, graves defeitos envolvendo a sinalização por fator de crescimento, regulação do citoesqueleto e na função das células inflamatórias foram observados nesses camundongos, podendo assim relacionar Fer a algumas importantes vias de sinalização . A participação de Fer na via de PDGF foi relatada, tendo sido observado que o estímulo com PDGF levava à fosforilação de Fer, indicando que Fer é, possivelmente, substrato para PDGF e Src. A ausência de Fer não demonstrou interferir na via de PDGF, mas apresentou um defeito na cinética das moléculas da via. A associação PDGF-Fer parece estar envolvida na regulação do citoesqueleto promovendo a fosforilação da cortactina. Embora a fosforilação da cortactina seja atribuída a Src e Fyn, na ausência de Fer há uma diminuição na fosforilação da cortactina, o que indica que há uma via alternativa para Fer, e Src e Fyn (Figura 24 A) . A participação de Fer na via relacionada à junção de aderência através do receptor n-caderina foi também relatada. O experimento em que a dissociação da junção de aderência é induzida leva a uma dissociação da interação de n-caderina e Fer. A dissociação de Fer com o receptor n-caderina pode ocorrer sozinha ou em um complexo p120ctn (catenina) e  $\beta$ -catenina. A liberação de uma PTP (proteína tirosina fosfatase) também foi relatada, sendo observado o aumento da fosforilação de  $\beta$ -catenina. (Figura 24 C) . A dissociação das proteínas da via de sinalização de aderência focal é correlacionada à um suposto direcionamento de Fer para o complexo FAK-p130Cas (FAK, p130Cas, Crk, Nck e tensina), em que Fer estaria diretamente

relacionada com a ativação de uma PTP, tendo sido observada a diminuição na fosforilação de p130Cas em células com superexpressão de Fer (Figura 24 B e 25 B) . Camundongos transgênicos contendo Fes/Fps/Fer inativas apresentam um fenótipo que suporta a hipótese de que estas proteínas estariam relacionadas na regulação da hematopoiese e funções biológicas. Na via relacionada a células inflamatórias, Fes/Fps são ativadas após a ligação de IgE e colágeno aos seus receptores, sendo que ambos compartilham em comum a cadeia  $\gamma$  do domínio catalítico. Nestas vias, Src parece estar envolvida recrutando e ativando Syk e Lyn que mesmo na ausência de Fer continuam a serem ativadas. Não se pode determinar desta maneira o modo em que Fer é ativada (Figura 25 A e B) . Fes/Fps estariam promovendo a ativação de STAT3 em resposta a GM-CSF (fator estimulante de granulócitos de colônia de macrófagos) em macrófagos. Foi demonstrada a participação de Fes/Fps na via de sinalização de IL-4 cooperando com proteínas JAK, assim como também fosforilando IRS-2 e PI3K . A fosforilação de Fes/Fer foi observada após a estimulação com IL-3, GM-CSF, IL-4, IL-6 e eritropoietina. Fps/Fer também foi relatada em associação com vários receptores de citocinas como: IL-4, IL-3, IL-5, GM-CSF e a cadeia de receptores gp130

Embora *S. mansoni* não possua as células do sistema hematopoiético em que as vias de Fes/Fps/Fer foram propostas, a presença de *S. mansoni* no sistema hematopoiético do hospedeiro vertebrado e a comprovada utilização de moléculas regulatórias do hospedeiro pelo parasito, indicam que as mesmas moléculas sinalizadoras propostas na via de Fes/Fps/Fer poderiam estar estimulando o parasito, mesmo que em outro contexto.

Muitas das proteínas propostas pela via de sinalização de Greer (2002) foram identificadas no banco de dados de *S. mansoni*. As ESTs de *S. mansoni* depositadas pelo grupo ONSA (<http://cancer.lbi.ic.unicamp.br/schisto6/>) e no GeneDB (<http://www.genedb.org/genedb/smansoni/>) foram automaticamente anotadas e denominadas por identidade com a seqüência do banco de dados genético. Aquelas que não foram encontradas podem não ter sido ainda identificadas em ESTs ou corretamente anotadas. A análise do transcriptoma de *S. mansoni* para a presença de genes ou vias específicas foi limitada a busca de genes ortólogos, por similaridade com gene de banco de dados, sem nenhuma evidência experimental . Como que muitas das proteínas da via estão presentes em *S. mansoni*, há a possibilidade de SmFes participar, em *Schistosoma*, de vias de sinalização similares às propostas para as proteínas da família Fes/Fps/Fer.

Muitas proteínas funcionam como parceiras ou como componentes de um conjunto multiprotéico. Compreender estas interações é fundamental para o entendimento de vias biológicas e função celular. A busca por parceiros de SmFes seria o início para a elucidação da real cascata de sinalização em que está envolvida. Para este fim, a proteína recombinante

foi clonada em vetor de expressão para ser usado em experimentos de *pull-down*. O sistema de *pull-down* foi desenhado para detectar a interação entre proteínas, no qual a proteína recombinante estaria capturando proteínas de um extrato protéico que interagissem com ela. Com a identificação de possíveis parceiros na via de SmFes, o experimento de *pull-down* poderá, também, ser realizado em pares, utilizando-se proteínas cujos cDNAs já estejam clonados.

## **VI.6 SMFES COMO POSSÍVEL ALVO DE DROGAS**

Como ocorre em *C. elegans*, mutações em um gene que codifica para PTKs citoplasmática (NRTK) bloqueiam as vias de sinalização de vários receptores proteínas tirosinas quinase e têm efeitos importantes no desenvolvimento do verme . A expressão transgênica do retrovirus ou do alelo mutante ativado *fps/fes* resulta em muitas hiperplasias e tumores malignos de alguns tecidos .

As PTKs têm sido intenso alvo de estudo para o desenvolvimento de novas drogas, especialmente em modelos de câncer . Baseadas neste fato, existem muitas perspectivas para a confecção de drogas que sejam capazes de interferir na via de sinalização relacionada a processos celulares essenciais em *S. mansoni*. Esta interferência seria mais específica que as drogas disponíveis no mercado. As moléculas presentes na cascata de sinalização de *S. mansoni* agem na diferenciação celular, no desenvolvimento da fêmea, na postura de ovos, na formação de granuloma. Interferindo nestes processos essenciais, o ciclo de desenvolvimento do verme seria interrompido e a doença estabilizada. Nosso trabalho visa portanto a elucidação destas vias para uma maior compreensão da biologia deste organismo e também para a identificação de possíveis alvos para o desenvolvimento de novas drogas.

## **VII CONCLUSÕES**



- SmFes é a primeira PTK da família Fes/Fps/Fer identificada em *S. mansoni*, com assinaturas de domínio proteína quinase, SH2 e *coiled-coil* característicos. A ausência do domínio FCH característico da família Fes/Fps/Fer, não é suficiente para excluir SmFes desta família;
- SmFes apresenta um *locus* contendo 18 exons. Os exons variam entre 65 e 575 pb, enquanto os introns variam, até o momento, entre 686 e 6.777 pb;
- SmFes está relacionada a proteínas da família Fes/Fps/Fer, agrupadas mais próximas às proteínas ortólogas de organismos invertebrados;
- SmFes é um gene de cópia única;
- SmFes apresenta duas populações de cDNA, com ou sem 9 pb inseridos na posição 3' do exon 10, que ocorre devido a um evento de edição alternativa. A presença das duas populações foi identificada em vermes machos, fêmeas, cercárias e ovos, assim como também em vermes individuais. A presença das duas populações em vermes individuais indica que elas coexistem em um único indivíduo;
- SmFes apresenta dois alelos, com ou sem 15 pb inseridos no exon 09 correspondente, que possivelmente ocorre pela presença de alelos distintos;
- Genes ortólogos a SmFes foram identificados em diversas espécies de *Schistosoma*;
- SmFes apresenta-se transcrita em baixos níveis. É verificada uma maior transcrição em verme adulto macho quando comparado a verme adulto fêmea, miracídio, cercária e esporocisto. A presença de transcrito de SmFes em esquistossômulo é verificada mas não se pode determinar seu nível de transcrição;
- A proteína SmFes apresenta-se expressa em esporocisto, miracídio, verme adulto e em nível mais elevado em cercária. No entanto, a presença da proteína nas demais fases de desenvolvimento do parasita não foi excluída;
- A presença de ESTs de *S. mansoni* relatadas como as proteínas presentes nas vias de sinalização propostas para a proteína Fes/Fps e Fer indicam que SmFes poderia estar envolvida nesta mesma via.

## **VIII REFERÊNCIA BIBLIOGRÁFICA**

---



## **IX ANEXOS**

## IX.1 ANEXO I

aacgttttaagtataacggcactaaaatcagctagaaggtttcaaagctacaaaattagctaaggtagagatgctcaaa  
attatgcttaacaattagaatattagaattgaactcaggtgattgattttgatggattttcaaactcagctgggttcttt  
ctaacatatcttctgatatgktgtatgattatgaagaaactaccgcaaagctgcacattattgagtaggtcgtgattg  
ttgaaaaactggtgagactggatgtacatccatcgcttgatagggttactaaaatgtcaacagtttattcctgcatta  
gtgaaagatttcgggagatttacggccaatttgagttggcaagtaataacagttggtgatcaggttacactttagtt  
ctttgaagaaccaaataatttaaaacactccatgcaggttgattaacatagtkaaaggtatcgtgaaccacttttcatg  
caccactgaaatgcttagctcgaacaaaataaaaacgactgtccgggtggttagattttttctatctagttattcaacia  
taagtcaaaaataggtacttttcaagtcaggtttcactcagttgtatatacacatcttcattgggtaccttgacccccga  
aaatctcattttcogttttatttccatctgacaGGTAATTCACAAAATATTCAGCATACTCCTCAAGGACTCTATTTT  
GTTAAACCTTTAAACACTTTTATTATTTCATGTATTTTAAATTTGTGTATCACAAAATATGGGAAATCATACAGTTAATGA  
AGTCAGTAACTATTTTGAACGTGGCAGTGATGTGCAAGATGATATCAACTTTTAGCATCTAATAATACACATATTAC  
AGAAAATAATCAATCCGTCTTGGTATCTTCTACATCGGACACATTTTGTAGTTCCTCCTTTGCCTCCGCATAGTCTAGG  
ACATGATAAATCATTATCTTCTTCATCCATTTCAACACCTAATGTGGACAATATCAATACAAAACGATATACTGATTC  
TGTGATGCTTCGAAGTGATCATAAAGAGTCTTCTCAGGTAATCGTTTCAGCCATCGATGCCAGATGAAATCTATCAA  
TGAAATGCTATCCTACTTTTCTAAATTAGTCAAAGCTGAAAAAAGCTGTCCAACAAGCTTTGTCCGATTTACTTGTATGA  
TCAAAAAGGAATCATCCACTGTTGGTAGTGTGTATGTTGGCCAGCGTTTCGAATACCACCGCTATGCGTAATTTATTCAA  
ATCAAATTATTCTCGTGGACGTCGTCGTTTAAACCATGAATATTTTGAATAAAAAACTTRTCGAATTCACCTACTGA  
TAATGAATTAATAAAAgtaagttcaatggtaaccctacctttatcactagattgatagaaaaggggaaatattcagtc  
tttctcaatagaacaagttattataaaaatctaattgtgaaagtttgtaaatattctgggtttacaataacattta  
gacagatgactoctaaaattttagatcogttaaaacttattggcttgggtttattcaatagagtaaatgtgaatatca  
ttttaacctctttactgaacacaagctctaagaactcaatcaatatctgtttatattatcgtagcttcttaatacaaca  
tgcattgtagtcactgttctgaatgattggatcaactgactatgttttaacgtaggttagtttccgataaccgca  
cggaaaccttgaaatgtcttaaaatgaaggtaaaatcatgaattccctcaaaaattagagggtataccttgttgtgcca  
actagtaaaaaaccatgggttgagtggttcttctgtctgactttcaactaagtagcattcaggaagtcctttcaagatg  
tccggtaactaagtcactaaagattaaattgtggagtttttctcattccgaaacataaagtaatttaattaatccctta  
tcagtagataggttttctctctttaaagataataaaggggaattatgtaatcacaattgattgatcactgggtggtgat  
caatgtcatacgtgtcaagtcocytcttagtggtgatcgaatgctatcgatcaagttgcaattgctcagtggttaacgtc  
tctgactgtgaagctcagtgacacgggatcgaatccgctcagggagcatcagttcccccaagattacaggtacacctg  
ctgacgagtgccaagtagcatgaaacccggatccaggggttctctgttgaccacctccaaccocatcttaacctcaata  
taaaggggaactatgtgtatgaagtcataaaaattgttgggtttttctattgtaaccttttctttctgggttaaagtgaag  
taaacttgggttttggtttgcataatgatatgccttttggtttggttatatataactttcagCTTTTACAAAATCTTGTAT  
TTCAACAAATCATCAAGTACGTATGTTTTCAACACGTGCATTATTTTATAAATTAATTACTGCTGGACAATTAGATAG  
TTTTAAATAATTGTGAACATTGTTTAAAGACGAACCTTATTTAGAATGTGAAGAAGAATTAATAATCACAAAGCGCGTCAGgt  
atgctgcaatttattatattccaagtttatctatgatataatagtcacctccggttctctatttcttcttataactggt  
catctcgtagcttcattcgtttggtttatgactgcgttatcaatatagtgggttttgacctattctctctttattgtag  
taagttcttctcattggctatggtgatactcaatgtgttatttaagagtaaaaggatactcataattaccagcactgtgat  
cacagatagctagaaaatatttccagtatctctctgtgatcccggttaagtaattagtagaataatagacagcaggtatt  
agagaatgatgataaaatccgctgatgggggttgtaaatctctcatcogactgagatgggtggggccagctgttacgtatgcc  
tgaccaccgggttaccacgacgtgcatgttgactagtggtggagacgattggaagaaagtaggtgagaccaaaacaaa  
ccggtggtatcagtccttaaagtcactaactctgatctgaacctgttggtagatgcagactacttgggtgggggtccg  
cgtgactatcgtaatcaatggtaggggttctgggtgacatgcagcagaatctatcacaatagcgtagttgtatgcac  
tctcgtcttcccttaaactgtgagattgaaattacttcatatctttctttctacgaatcaattcttctctgtatcatg  
tctgatttttactactattgaagtgactgctcttatgaatttgggtgtttggttttggttgtgctaatgaggtgtcgcaac  
ttggactgatgcataatgtacgtggctcactggttatagctgactgaataactgactgactgtcatctgagacctgta  
catttttgggtacggctcgtagcattgtttactgtggagtggtggggctcaaagattgaaatatttgaatttaagccactaa  
gtcatgaattcgaatcctgatatacttactttgggtttgaggaatcagacgggtatcactagcctcacacttatgggtatg  
aacttgaactggtaaacgagttgtgttcagtaatttgggttataactgagtgactgaatgacaggtaaattagtaaaa  
tcaacagtttaactaaagtatagagtttttgggtgtgttttctctctgttaatacataatgtagaatctagcacaacta  
ataactagtttaagttgtctgatttctcatcogcatatagtaggaaactagtaattgtcttaataaaatttgtcaacaaac  
catttttacttattcatcaatgagtggttatgagatgagttgcattaggagctgtgtaactagtagcgttccactcatgt  
gtgggtaatgacacttaccatctgatgggattagatcattgtagctctatttcaaaaatcgattgaagttggccggtgt  
ggaaaactttgtccaagtcagtggtttaaaatattgtactgacccccaggcaacctctgaccagtgcataggtgt  
agaatgttaaggagcttctgtgttgggatgatacagggcacttgggtatctcttcttcttcaactaatgtttaactaaca

E1

E2

tgccagtcagcattgtaaactgcctttaaaactaaactattcttgataaatctgtgcgactaaatgggagaatatat  
atthaaaatagtaaaataaaaccaatcaaatttaagagtaaaataacaatatgcagacttgagatgatttagaaaaagag  
aggaaataagttgacctatgtcgcgtcgcgtcattgattttgaaccgtatccttcgatcactttgtcgggggttattgcat  
cgatctcaactatgtagagcttaccaattcattcaagtttagattaaaagtcttcgtcttcatttttaacgttctataca  
ttccagtttttacacagatgggttcagtcagtcagtcacagcgtagaaaattcgtacgtacgtacatcagttcaagttg  
ccataccacattaacacagatttagttgtcgcgattcgaatcccatagtggttagaagtagtaagattataagcattaatt  
ggaaagattaaggtttgaagatggttattgaaggagtataatccaatgaaataaatgtgaaaagagaaaaaagataaag  
acatgaagaattccagatgactagaattttgcatccactcctttgtgttctgaattcttcatgcctctatctttttctc  
tttcaaatttag

E3

/.../gattagatataaaactcagatggttgaataacaatgattcatacggcaggggctgttctggacttaactggg  
tgggtctaggcagaatcaggaccgacaagtaacttagattactcggacgggtcccgtatatcattgatccgatcaataa  
ttctctgttctctattcagtggtcgcgtcagttatacattcacagagcattcactcggcccccaagatataacattgt  
atwtstttcatgttttwttaatkttwca**g**TATATGTTAAGCTTAAAAGCATATGAAAAGAAATATTTCAAATATATTG  
GTCAATTATCAAATATACAATCGAAATTAACAAATTTATGTGATAAAC**g**taagttatatttttaatacataracatt  
ggtaacaaggaagcaccagatataatgcccacacaaaatcttatatgatttgtgtgaggggctgtaatactgsttaggt  
gtccaawctgaagcaaatgggttttcttaaggggcccaccccggagcyttcgacstaaggatstgatccacaaggcag  
ggagcatcgttaaggagatgcagtcctcaggttagccgggtgttgtattstattgcaaaaaatattttatattctatttac  
acataaagttatctctaagattcaatgaagccaattaaattgtttttaaagtagtcaataacatttatatttg  
tgttcatcttggtgactatcataaccaatgtttggagactctaggtgatatgggtcaaaaatcgatcacaatgggtcag  
gtatatacatctttattttccmttaaaccctgagaaataaaaattgcttcataaaaattccctccctttccccaaatatnt  
tatccttatatactaccaccaccactaaattaactacttctatgaactcgggtgttgatcttgttgtgtaacgaggta  
tggcaacttgaaccgatgcataaatgagcctgggtcctacgttgtagctgagtgacatttatacgcataatcgtattgag  
cataattttgttgtgcatatctaatgaccatcttctctcttcttcttcttcttcttcttcttcttcttcttcttctg  
ctcaacttttgtcttattcatttgtctttattttattaacatcaaaaatataattacttacaatatttcacataaagtaa  
tgttgttttgttgttttcttctcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttcttct  
gatataatgatactgtttttttaaatttcccattctataactgggtgaaataaatgaagttgattaattatagtagtact  
atgacagatgattgtggaattcg

/.../caagtagaaggaatgttattccaaacagtggtgattatggtaaaaaattgtcaggtatattatagtttttagta  
gaaacaaaatcatcattatagttgtccaatccgcacacagggagttatcagcattgtccagtagggaagctccacgta  
gcaattctagagtattcttccaaaagctgattggatggaatatgtcgggctccttctgggcttcttagatcttctctg  
acttcacctgtggaccatcctcacagaattatttctaataatgatagtcattcaagcataggttaacttgttgtctatggtt  
attccattattttaaattatgatttagattaaagaagatctaaaagtacagggatttcttccaagaaaggagcgggttact  
tagttgttactaggtatcagtggtgatgtgcatgggttgatgagactaggcaatcatcatttttagaaaccaattcatt  
atccattaagttagttaatcacattaacctccagataattcaagcattatgacaaaatttaacttttacaatgaaagt  
gttttagtaccattgaataaataactagcaaatcaacaaagacattaagattccagttagtttaaatgataaagttt  
gaactggacaaaacatctgtaccgtatctcgcactactagtcactatcaaatatttgtataatatttgaacaaaactg  
gtcatgaaataaggatattcagtcagaagaaaagatggatgggtggcagtggaatctaggactcacgtttcgtcc  
catttaggactcttcagctggatgtatctacatccaagagtt**gatatc**caatctaggactcgaacctagtagcgttca  
cttagaacgccatcgcattatccacttaactactgagtcctggtagccacctgcatgtgcaatgggggtgaagttaaa  
ttcacttgggtactttttacttgaatcttcccattgatgtttacaactgcaattgatcagtcctattattggccatgtg  
catatgccaatgacagattgatcaattgcagtcctaaacatcaatgtgaagattcaagtaacaataacctagtgacat  
cagaagaaagttattgagaaactgttataacgtgtgcgtgccagttattatatagaacgttatgagaccgccaatcgg  
agagcagtgattttatcgaccagatccaatgatacacaacaaactgtatgcccagtccttgagtagctcgtcagtcctgctt  
tctgcctagcccaaccagtttaagaccagaacaccaataaaaagcctctgaaatatgaaacattattcacaaacatactg  
ggtcagataccaaacaaacagaccacatcgtaccataaaaatgaaaaataaacatttgtacaagatttggccaaatgtgg  
ctgtgaatagggagaacagtaataatagactgaggataactcaagaatggtaattcgtataataatagttcataggt  
caaaaataagccttaggataagagaaaaatgaatatgaatagtaagcaagatggatagtggttaggagtgaaatccagg  
acgcgcgtttcgtcctatttgggactgggtcaactggatgtacctgcatctcagagttgatgttcgctctgggactcaa  
accactaactttcgtctcaaacgccatcgcgttatccactcgggctactgagtcctctttgcttatcatgcttgtga  
attaaggctatatcgaggcaatacgcacagtatgcacatacgcgaattagagaccagttgtagtcctaacacatcgat  
gggaagattcaaacaaacaataactaaatgaatatgaacagtttagttacttagcaattatacaataggaaacatacat  
atcatattgggtccataaatagtcctcaacagttaccattcataaatctccacgggatataacagaaacsttttattct  
ttgaatgttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt  
tgaatgttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt

E4

AATAATTTAGAATATTCTACTAAAAAATTGCATTTATTACATAATGAATATTGTTTGGCATTGATCACTGCTAATCAT  
TATCAACAATGGTTATATACACATTTACGACCATGTTTACTGAAAGGTGTTGAACGTACTATGCAATTAGCATCGGAA  
GTTGT**g**taagttgttgttttggatgttttgtcagatattttgttttctcaactgatcgatgttgattactttaaatggtt  
tgaagtagctcgattttcaactaaggaatagccgggttcagtataactccatttgcttgatgttgggccaacagatcgt  
attattagccaaaagtagtagtgacctgaagtttattatataataacctgcgcttgaacatgagttatggttag  
catttcaactctactgtgacttkgttaacatttcatgatttagtctggtccataagacttggaaattttcacgtcaac

E4

gtctgtaatcaagaatcgaactttatgtttgaggcaaaagtctcactagtaataactaagaatgtaggattttaactta  
tggttatgaaccattgactatcgtgctcgacctcagcctcttaatacaaaaattttgtctacaatttgaggaataaatat  
tcacctagaaaattatctggggttttataactaagatcactcgcgtaattatgtttagaccatatcatgcttatcttcaag  
gctttcgtcaatatgctattaataacaataatgagtaagtcaattatctttattatctatgagcatggttgatcactta  
ct

/.../atcttttacactaagaataatcatgaacttagtggttaactaatttctataggaagtccaaagtcttagtg  
aaagtctttgaacaatgtagttcacctcatggtgtaatgggaaatttaccogatgtattcattggtttttttcttttc  
tttctcaattttttttaaat**ag**TTGGCATCTTCTAGTATTCGGTGGTGAAAAATATTCATCATTTCATCAAAATTGGC  
TTAAAGAATTTCAATCTGAATTAGATTTAACAAGTTTATTAGAACAAAATGG**g**taagttgataaatgaattccattatg  
atcagaaggggctttttgtggaaatcccagtaattgtatatagttgaaa

E5

/.../tcggtggtggtggtggttatcctttttctctctcgcgaactatttatcttatttttttctttttatctatata**a**  
**g**AGTAAAATTTCTGGACCTGTATATCATACCTTCAATCAAGCACTAAATAATAACACATTAGGTGGAAATCTTT  
CTCCTGGTTCTTAATTGTCAATGATTTAACTGGACAATCATTGATTGAGTT**g**taagttggtttttttctattttggat  
gttaacattactgaattgactatttactgtctttgggttcagcatctatccacaaatgataaatgaaatgatcttca  
aacaggtatcattattaggtgatttcatgtcacttgggtgatgcaaatgatcaacaacaaaagaaacctaaattagca  
cgtaaaaaacagtcgctcgcgaacgtaggtttcaggtcgaatatctatcaatgcaaaactatcacttgttatatcggta  
catataaagtgaaaaattaagataaggagtaacaaattgaaggcaattgtcatagtgcataaaatctggtaaatcgcac  
tcaaaagacacgggtgagattgcagaactaacagttggtagacacactaaactctttaaataaattagtcagagtag  
atggttaacctttgatgcggttagtcagagttagttgtcataatgatttgggttgacttatattttggaaaatatttct  
ttaggcgaacgataagaatataggcaacaaactcatagttcgatggtgaaagtc

E6

/.../ataatgtagcaccgaacttaaggatagagaga**gatatca**attggacatggaaacccccctaaatcgactcaacg  
atgtacgcaaatcgaccatatttccatcagccgtcgtcgaaggggctgatataagacttacaatcattcattcacat  
gtttaaactctgatcatattacagttcaaaaatgtgcttgtttacgatctactgaaaatagaaaagtaacagcaagaa  
gacattgtagatttcaacaaaatgatgaaaaggccaaaagtatatttcagcagcaactagagaagcagttactcaatc  
atcaaagtgatgttcacctctacgtaggttgataggtatccagacagttgtgagaacagcact**gatatct**tttgacta  
tctataatcagaaggttaagaaaagttagtagatttttagtagcgtctactgaatttacagatgctcggaactcacatc  
ggatctgtaacaacaataaataaataaataaataaataaataaataaataaataaataaataaataaataaataaataa  
tagttaaagcaagccaaggaaggacaatgataataggaacaatgtacaacttttcaaacttatcgaagaaatgggt  
attaaaaatgatgagagttgtgtaactgttctataaaaatgtgggacaataatccattttcaacacaagaaacttaag  
cgaagtgtagaacgattgcaaggagagtttaactgctctccatttagccttcagttgcttacttttaccacaacactg  
gatgagatgtcaagatattggcctccgactcttaagtagaaaaagactaacgaacatctttgtcaagttctatggtgat  
gatgattgagcgtctacataagtaaatgtggttcatatttttactggtttaactcatttttacttcttttttaac**ag**  
AACACAATTCATAAAAAATAAGGAATTATTGCATAGAGAAACGGTTAATCAACTGACTAAAGAATGTTTACAATGGGA  
AGAGGCGCT**ATCTCAATGGCGTGCTGT**GTTTGTCTAATCCAGGTCCTAATCCATGGTCATCAAATTTATTCAATATACA  
TTCATCTGGTGTTCACCAATAGTCAATACTCCACAAGATAATAATTCAAATGGATTTTCATCATGTAGATAAATTCAG  
TCATCTAGCGTTGCCT**GATATC**CAAGATCTTTGGCATGATGATCGAGCTCGACCTGTTA**Ag**taagttatattttttt  
gttaatcagttgttttaataaattcatttcatactgactagtcfaatgaactgtaactattgatataatttgtctacgtct  
tt

E7

/.../cgctctacatataacagtttaaagctttttgatcgtattactgtcctttcacgttaatagtatataaactggat  
aatataaactcgtactaatgaacatcataccattgaggcaggattagttgctatctgagaggggactaaactttatgg  
acgagccaatctgaaagctttcgcggttttcaaggtcacggcgtcaagtttggtacctgatgcttacgtacgatcggc  
ttacagcggatttttagagaagacgtgataaattcagcgtctacaagagaagtgattataagaagttcctataacttgtga  
ctgtggtactcaaatgacctttttcgtcfaatcagattaaaggttaagaccctaacctaaaccataccaacatacca  
acmacatgaagttatgtgtccagggtaaaagatttttcagaccttatcatgtctctagaggtcgactactatgrrgcc  
atatgratgagttcctttaccgtatgcactatgatttttagaacctctgaaacctatctgaccttgagaagtttttga  
gtcatgtaaaaggacagtatccgcttttaatttttttagttttgtataatataatttttactgtataactgactgctctg  
attggctaagtttctcagggccatttaagattcgttcaaaggtcgtccataaagtttagtctcgcgctatctgaac  
actgttctcagtttgaagtaatcagttattggattcaagttcagtatata**gatatca**acactagaatgtaaatggatct  
agacatcagatcccgaataaaaatgaaatacacttcttgaatggtgctataatatacattccatataatattctctaat  
cttttatttccatttctttttttgtttcactacttttcttaaaaat**ag**TCTTTGTGAAGTATCATTTCGTTTAGCAATA  
TGTGAATTTGAATTATGTTTTGCCAATTGTCTTGCTGATTGAGAAGCTCAATTAGTCAATGTATTAAGTATGCACAA  
TGTCGTCTCAATTGGCTTTTCAATAAAATTACTTACAAATGATTCAACTACAAGTTATCCAAATACTACACCAATTACA  
TTTTTCAGTGATCCTTGTGATATTACTACCACCCTACTACTATTACTAAANGGGAATAGTAAACAATAATAATCATGA  
AGATATAGATATACAAAATGGGACTGANGATAATTGCAGTCAAGCTTTGAATGGTCATGATAATTCA**g**tgagtggtttt  
gttttcaatgttcaatataccaacgttatgattctgtgtgagattcacagtcctctttctatcttctacagaattgaca  
atgntcntganttatatttontaataaactgntctagtttcatagttattggtttggttctctcattgatggtttt  
attaggactgcaactgggtcagtccttaattggcatatgtgcatactgtgcgtattgcctcgatatagcctaaattca  
caagtatggtaagcaagatggatagtggttagcagtggaatccagttgtgacgcgcgtttcgttctttttgggactcg

E8





/.../atatagccttaattcacaaagcattgtaagcaagatggatagtgggctactagatccaaaatgatacagacaa  
ccatggaaaacctgaaagcacttgacggctgtttcatcctattatgagactcctgaaaagtgcacatacacgatcccc  
cttctgataatttcaatataatggtgaaaaattgaattcagctttatattatcctgtttacttagatgattgtttacc  
tgccacgaagtacatthtaatactgtaataatthtgaattatgtgacattatgaatataaacagtcggacaaaattaa  
agagaaaatatctttcaatthtactaccctcctaataagcaataatgatcataatagggatgaagacaatcggg  
tttcataagttaacatthtaatgcatccgatgataatagtgatgctgcagcaattgcagcaattcgggaagactcct  
ccggactctcacattgtccggacattccatgtgctctaaaaaattgtcgctctgattggttagaaggtgcattggcccc  
tgacttccaaaggaactcggctttcataataagacgctcatagtgattccttctactcccagggcagagtttaaatggg  
ttgaataatthttctgattggcggttttttagcgagttgggtgtctacgggggtggggtggttaaccccatgcccac  
cctcctcctthtaacggggcttgggaccggcagtagccccggaggagctacaagcggagttgtaatagtcagcagat  
aatgataatgaatactthtgtgacattacattctactgattthtgtgatthtcatatthtgcatacagcaaaatcctggg  
cattcaatcctthaaatcagtcagaactgtggagtgaattcagattgatgaaatgaaaattatattaacatgttcagt  
tcttactcaatacataagtgaaaacatgactgaaataggttgagtagcagacataaacatgatgtgaaacttggaaacctc  
tgcaaaccttcaaccgggaacataaacactgacgatgaaatctgattthtattthtaaacatataaatcttggtagata  
ttgatacatatgagccgcacaaatctcattcagattthtgtgtgagggctgtgatactgctcaggtgcccactgaagca  
gggtggthttcttaggggctcacaccacgagcctthtaacctaaaggtctgatccacaagacaatggagcatcgtgagga  
gatgcagtcoccatgatagccggtagccaacaattggthttatactccattthtccctcaggatactggagccagccca  
tgtgcacgattggthttggaatgagthtttccaaatccccttggtagaactthtccgtgttcatcaaatcgattaaagt  
acggatattcgctthttctcctctcaatthtctgtatatacaacatcctcgcacagagaaggcgttgaaataggacttccctg  
acagaggctatatccgcgtggctcatgtgagagcattthttagaaggagagcggacttccccattctcggccgtaccag  
ggthtttggcggcatgaaatctgataactatccataaatagthttgtthtcatthtataatcattcaaaaagthtttaatt  
cgaatccatcaaaataaatgataaacggthttatgtthaaatcaacataccatcatattagatgaaatcgatgaaatgaaa  
aaaagcaggaatcaaagcccatttatacctccataaatatccagtgagtgattthttagacaaagaaatgtttgat  
atagacaacagtcacttaggatgaccgattgtthtattgtgactattactgctattacatgactthttagctcatcgaaa  
tgcataagtgthtaactthttagaccaccgattgtctthtccctattatgattthtattgtthtcaagagggtagtaat  
gataaatthgaaacacatthttctctthtaatgtcacatcattcaaaatagattacagatataaaatgcactacgtggtagg  
taaaacaatcactaagtaaataggacaatataaacatgaatccactthtcaacataatthgaaatthacagaaggag  
gtgtgtggagatthttagtaattctatagattthttagatcatgagtcattgaaactagatcactatggaatacctgga  
agcactagacggctgtthttagtctthttagtgagacttctcatgagtgccatccacgatctcgcctcgcgagatthtcaa  
cctaggacctaccagtgctcgtcaatthtctgtagatthtggthtgaagthtagacattaacaccgttggatgtcggctcagt  
gtctatcggthtaactgctctagcgcgagagcggaggtctcacaataggaagaaacagccgtcaagtgctthttaggthtt  
ccatggtgatctagctacaatthttagctcatgattthtcaactataacacattgaaatataatctcctaagtgataaatacttg  
catagataatgaaatgaaatatacaacaatatacaactaaatgaaatgtgaaatagtgatcaatcacatcactgaaaactatc  
catatcatatggacgaatcattagthttagttgagaggagttcatatataatthttagcattaactatthtcaatactactg  
ccaactcgtctthtcatcacctattacattatgtaactacaatthtctcacttgattagatggaatcaacagataa  
atggaatthtaatgacgagccaatthtgaataatgtgtgtatthtttggthtagthtaatgatcaagthttagctgatatgaa  
gtcagtagtaagtagctthtggataatgacacatgaactctgattgatagthtgcattgtaattgthttagtaggattgt  
thttagatgattctgtgaaatataatcccctattgtthttagctgatatgattthtcatatataatgattctatcatgtctgatagthtga  
tgactatthtgggaggagthtagthttagtaggthttatgtgattcagattactgtthttagtgatgatgtacaacaaaatagtht  
atthttagacttagacgactcactthtcaaatagctcgtctccattctcttacattcagtcagthtaatgctthtatt  
gtcaaacctcacatggacatgttagacagtgctcattgagatactcacaatcgtttacagggcactthtcatthtcaaaatactc  
tgtthtctgtgagatggaatgcaatctgtgtthtgaatgtgtggaatactatcagthttattthtgtgtaaacctthttagctgt  
ctthtcaaatthtctatattaatcaaaataccattgtgtthtgaatgaaatctcgcactgtthtctcattcagggaaacatgca  
tcataaacactthtataatgaaatagggtaatgcaattcactaacattthtgaataagctthtccaatgactagataaacata  
gtaaacgthtctgttaggctthttagaacagatacatagtaaacatgtthtgaatagtggaatagacataatagthtct  
aatatgthttagtagtaactgattthttagtagatthtcaactaaatgaaatctgaaatggtaaatgthtaatgattgtcatct  
gtthttcagaatagthtggccatcgagacgtgtgtaatthtgcacogataatccattgatctthtcaaatcatttagggagct  
ctthttagctcagthattgtaaatatthtgaatgtgtctagcagattatcgatctthttagctcaggtattthttagcaacaaatc  
aagaaatctcactgatgatgaaatgthtattcatgthtthaaactthtgaatcgtgtthtactactagtaacaacaatgatcataa  
taatgctagtaataataatgatataactactaattataatagtaataataacggthattgatattaacggctctgactg  
agctactggataatctggthttagaatcctgtgaaatcatcactcctgtaaaaagcacaggcatagctthtattggthga  
tctgaaaatagcactgaattctthttagaccgctthtcatthtctgaaactthttagaactthtggcatatccatatacaag  
ataatthtctthttagtaaaataaatataatthtttaagthtgggctacattgaggagthttagthttaccggthtctgctgc  
tatcaccgagtagctthttagctthtctgthttagacaaatgthtcatcataagthtggatthtctcgcgatthtctthtgataa  
caaaatcgatataataagcaagatggtagtggtatcagthtgaatccaggatgcccattcgtcctgagthtggataac  
gtgatgactthtthtgaagcgaaggttaatagattcaagthtccagagthtgaacatccactctgagatgagthtgcattccagc  
tgacgagthtccaaataggatgaaacgcgcgtcaaacggattccactgtthttagccactatccattthttagcttacgatgc  
thttagaattatggctatatacgaagcaatacgcacagthtgcacataatgcaaatagagactgaccagthttagctctta  
aaaacatcaatgggaagattcaaacaaacagthtactaaatgaattcaaaaattthttagthttagthtthtaactctagthtthtatt

aagatccaaaacaaacggtttcatccacctgacaagtttcaaatagaatgaaacatgcatgtaatccaccattgtttaa  
ataccctcatgaaacttttacaaaaattaaaagttataaatgaagtcactcccataaacactgtggatgaagaaat  
tggctaactgttttgtgatggacattatcttctattttagaataaatcttcaaaataggaacattgggaaccataa  
cgggatggattcgtatgcatataggaaacaatttgaacctagcacacaacaatatcgtcatgcatggccattgccc  
ttggcagatcagtcattcgaatgtcattagaatgccgtcagggataaagacagggtaggtaccccagtcagaaaaat  
atcaaccgctgaaagcttaaggtaaatgccactaagccggcgtacatgtagtccagaaagaagtcagatgcatgcttcc  
ttcaaaatactacctcattaaaaatattaatcatggacagatgaaatgtataataatctttatttgaagaaaaataaa  
caaagttagatcatcattaaagacctggaaatactgaatggctgtttatcctgggtataagatcctcagcaatggc  
atccacgataccgaaagtggttcaaaaccacgaccttcgggttcgcacgccaacgggtgagtcacgcaatcgtga  
attgatcacgattaaacattatcacctgtggatgcccgaatcaatggctacaagttcagcgttcacgtgtgtgaggtt  
gaaggtccagatgccaacaccgaatccacagctgaaaagttctatactatgatgaaacaacagtcacgtgcttcc  
tcatgccaaaacaaatcgtgcacatgggctccagaatcctgaaggaaacaaatggcgtatgaatcaatcgttggtcaccg  
gctatcatgagactacgtctcctcacgattctccactgcctcgaggatcagatctctagggtcaaaggctccgggtgtg  
gtcccctaagaaaaccacctgcttcagtttgcgcagctgggcaatatcacagccctcacgcaaatcgaatgaaatttg  
tgtggcgcattttatatctgatgctcccttgtaccaatattttagtgtttaaataataataataaatttaattgataggt  
aacaatgaacagcaataaataaataaataaaaaattccataataagttatcatctcattggaaattaaatcaactgtta  
taaagagcagtatgggtgtggtctacttatatcctcataagtagtatatggtaatagtcagacatagaatgtattttgg  
cagaagaccgataaagaagaactgaaatgaagcgaattggtaggaaaatgcatgaaacatgaaatcagagaagatg  
gattgatatttgcagaagaaacggtaagattgaaacgattgattgttattttgcaaatctactgtttgtgtatggt  
tatcagatcttagtaagacagtttgttaatttgtgcttaatacatattgattgtctccaacgggtgttttgttccactac  
agtaatacactaattcagtttcagtttcagtttggaaaggacaggaggcttgatggcctgtgttagtcctggcaatga  
tggctctcaggtgatccaactttcttttgaaggagtcgacggatggagcctcaaccacgtgctgaggtgatgaattc  
cactcgttgattattcagtgaggaaagtcggtagtcagctgacaagtagttcgttcggggcttgtgaacttttttggag  
tgtcctcgtagatttttcaacgaggaaaattagtttaactattaaactttctaatggattcaagatcactaaacttg  
tattttgtgtagcattgtaacctttttcttacttgaatttacaaaagacaatagtaattagcttaactatttataaa  
gtgttcaatagacaatcacattatgggtgggtgctactgatgtcgacagatataagtagcatatcatcaatagaag  
tgaatgtctggtgtgcagaaggttaataagattaaagcaaaaagacgaaaacagagaatgatgggtatgggacga  
agaaacaatgaagctgagatgattgactgatattttgcaaatgacgtgttttttactgtatgggtgctcagacttt  
actaagagattctgtaattgtgtgttcaattacatccagctgtccctgctgtgtcctcttccactacagcattacat  
acggtagctttattttattctaaagtgagatctgtgttgagtgtatagaagagaaagcagacttcataatttgatagcaa  
atatctttcaaatgggttagttgtgataataaccgtttcccttgtcaacttttttgaacatgaagattgaaatcaca  
agtttaactcgtaccgcattttttttcttctgtaatacagTCTTTTACAGAATCAAGGCGATTTTCTTGTTCGACAAAC  
AAGTAAACGTTCCAGTGGGCGTGATACCATAGCTCATTGGAATGGAACATAATGGTGATTTAAACAATGGTAATCATAA  
TGAAACAATATTAATAAAGATCATAATATGGATGGAACGTGTTTTACGTATGGTTTTATCTGTCTTTTGGCATGGTCA  
TAGACATTTTATTCTATATGGTGGACCAGAGACAGGTTAAGGTTGGCATTTAGAAGACGGTCATTTCTCCACAATAAG  
gtaagtagtagtagtcacttttctagctaatgttttttagaacaccatgtataaacgaataatattgttttcatagt  
cgaatcatgagtcfaatgaagctagattaccatggagaacctgaaggtactggacggcgggttcgtcccattatggg  
attcttcagcagtgccatccacgatcccgcactcgcgagattcgcacccaggacctaccagtagcgcgcgggaacac  
ttaaccgatagaccactgagccggcatccgaatctgcacaaaacccttctgaaatattgttttgcattagatcgtca  
tcaggctttactcctaataacagtaataatattatgaatatacaaaattattaaaccaaccaaggcagtttatgagtc  
aattataaccagtggaatagattacataaataaacataataagtgtagtacaatttggtagtgtgatgacaagtgta  
cttgttgtgctatgaataacagaggcttgaattaatgttacataatgagaaatagggtatttgattagaaaaatatag  
acactgaaataacattcataaaaagccataagattacgtaataagaaggaagtggtcagataattggacgggtgaaacgt  
atacttgagaaaaaacaaaatgataggaggggggtgaagagaaacacgtagaaacagataggagaataataatgaata  
aaatcatttataagacctagggagagattagggtttacacaattctgatgtacacaaagacttgaattgttgactc  
gaattcctatagcttctgctattttgtaagggcagatagaaccccataagaaaaactagtaggaatcagatactaca  
tgaccgctatcgatcatctgttattgaaocgagctgctgattgccttgacagtaaccttttctaaactatgaagggaaa  
tgttcaactaactatttggttcatttatctggtagtacgctcagtatagtttatagtgaaacataaaactttgtgatta  
gcaaaaatatactggcactaaactgttagtgtttccagtcgtacattttggtaaatatgacacggttttcttctoga  
aacctttgtttctatttggatttggttatcgacaacctctgaggaatgtgaccgtagtttagcagctgttaccttactgca  
caactaagttgtttatttaggtttttgcatccagttgtataaaaactcatatagaaaatggagacaaaaacttttgtcat  
aatcggtttccaaccagatatacctggctgctttttttaaatttttaagtagcagatggatagtggaacaactgt  
agtagcaatctgagttacggaatgttcttactgggttgcactcctaaatgaaaaacacgtaaatattaaccaattgttc  
accaagttttgacatcttttccactagttcgcactattactttctgggttcaatacacgtaattttcaccogtattttctg  
tttaagaatagtgatcatatctcatagacaaatagtagcaatattgttttctagtagtaagttttatcaagacaatat  
gggtttcataaataatataatcgcataataacccatgctaccggctaccaactagcacacaataaataatatttccagt  
ttagttctaggtagtacttgtgagtagcttagtattttctccgaaatatacatccgtgggtttagtccgataatagtg  
gtccatttattgttcatacacaaattcatcgaaaaacatttcaatcaattgtgtggtattatttagtgaacttgtgtt  
cagagatagattgatcaactaattaactgatcttgagcaagtgtagtcaggccttatgggttttacttttgacattc

E12

E13

cctgtgtaattgacccaatcgcgctcacaagacaaaacctttcatagaatcctgaaggccaaaggaaaagagaagatca  
 aagaacacattacgccgagaaagggaaacagatatgagaaaaatgaacaaaaactggatagaactagaaaggaaggcg  
 cagccaggacagagtggtggtggagaatgctggtcagcggcctatgctccatagggagtaacagggcgtaagtaagtgt  
 attgatgagcattcaatcaatagaaaatctttcgaatattttttcatawgtttatwatcagctctgaaatttctcttg  
 gataatcacgaytgcgtttgcttaataactaaaattgtat**ag**GGAAGTGAATATCACATGAAAACCTCAAACCTC  
 CCGTGACAGCGAAATCAGGAGCATGTCTTGTGACACCAATTTCAAGACCTGATTGGGAACTAGACAATCGTGATGTAC  
 AACTACTACAAAAATTTGGTCAG**gt**atggttttcagacattggtttyacaaaaataataatcgrtttttaataatgcca  
 ttsttatgcttggttatttcgtaatgygtgctcctatgatggaataaaactgccaatcgtccacgggagcaactcac  
 cgamtcrttyrargtaaaagaccgggtgtargcaagggtgcttactctcaccctttctctttcccatgggtgatcgactg  
 gatcatgaagacgtaaacatctggaggaawgcacsggatamagtggtgactggcaagatgcagcttgacgatttaggctt  
 cgcrgatgatctggctcctctatcacaacgcaaaaaacaaatgcaggagaagattaccagtgtagcagtagcctcagc  
 agcagtaggtctcaayataaaacaaakrgaaaagcaagactctccgatacaatacaacatgcaccaatggaattacact  
 tgacrgaga**agc**tttgagggatgtggaaycyttacatatctgggcagcatcatyगतgaacacrgtggtatctgatgc  
 agatgtgaagacgcgatcggaacgagagcagcatalctacaattgaagaacatctggaactcaaaacaattgtc  
 aaccaacaccaaggttagaaatttcaatacaaatgtraagacagttctaytगतggrgaggagacrtggaraactac  
 gaaagcyatyatycagaagatacaartrtttatyacagttgtmtacgaatgatacttcggatytgttggsacagacac  
 tatcgacaacaatctatcgtaagagagaacaaaccagactccatccgggtggatgaagaaatmaggaagaagyrcctgga  
 agtggatagaaccacattgaggaaagcaccacactggtcctgaaggctaaaggggaagaggaaggccaaagaccac  
 attacgccgcgaaatgtagacggatgtaaaagaatgaagaaaaattggatggaactagaaaggaaggccgagggccag  
 agtatgtgggttgagaatggtggtcggcggcctatgctcmattgggagtarsaggcraagtaagtrtgaagttat  
 gtctgtatagtacagaatacttgagaaatgtwacaattgttttggaagtattgcacataatttagcagaaacttcaagc  
 cagaaacattaactcttgcttaagttattatgtaacatagccaagtcagtatcrgagtctgaatttycatcgattwt  
 agttattggsatatgtytatcctccgmtactgtaacgcgtaatgtctgtgttttaayttcgaagyaagattwgttt  
 tttgttgagaactgaatkaarcrmcaggggtctttgtttttccacattacctcataatggaaaaagagcattctc  
 aaccaactatgaaatttactcatgactttcagacctgtttgagtagctacagtcacagatgtaatttttggttgtac  
 tccaacaaaggaaacgcctgattctaactcctcaagtacatgcatgtaggttaccgcagaataattgacagctg

E14

tcagtgtttgtagtattcatggtgtagcgttattattttatattctgcatctcctctttttac**ag**GGTAAT  
 TTTGGTGATGTTTATCGAGGTGTATATAATGGATGTGAAGTTGCAGTGAAAACCTTGTCGTGTTGATATGACTGCATCG  
 GATTTACGTAGGAAATTTCTACAAGGTGAAACAACCTGCATTAATTTCAATCATCCAAATATTGTCAAATTAGTTGGT  
 ATTGCAGTTCAATCTTATCCAATTATGATTGTAATGGAGTATGTTCCAG**gt**tggtgaacttttcatgtttttcat**gaa**  
**ttc**tagcatatttcatgttttattcatgttaatttgttatttttactactactagctataagtagttacattttcaa  
 aggataagtggttagcatgtggattgacagtaatttgtgactgcacgttaacagtggttactctcaaaaaatacagaa  
 aatgtcttttagaggattggtctaataattccgtaaccatcggctgagccacggttgagggttactctcaaaaaatacagaa  
 tggttgatgaagtagatcttcagcaacatgtattatgtatatccaacgaacgctctatcgtaaacgacggttactggcc  
 tgcaagtaagttggagaaagatggagacgataggttagtagcttaacgctcagtcocatgaattaataaaagttattct  
 gagacatattgatagatgcaaacactaccgggttacactcctcaaacggttctcaaccaacgggttaaaaaacattgtgtca  
 tgtttgagaatcgggtcaaaagagcgtaaacatatttggatattgcttacttc**gga**tcctcaattttgtaatcatttt  
 ttcacaccttatcaccctaatacattctgtctacagtcacataatctgtttctaacttttttactatcatggttct  
 aatgtgggtgttatattgggtagttttgcattttgtgtaccgaactatacgttgaatttgtctgggacatcattggaag  
 tggaaacaaatagttttctcagaaaaatgaagataggtgacaggaaaaattattgttccaataaggataaaaaaaaaat  
 ttaattaagtgaatgaacttta**gaattc**ggttgtgaacgctcagtaactgcaatacatacaaatcccaccaactggtcca  
 aatgcaacaaagtgcgttcattaaattccactttagctacagttcagaactgaattaaaagtatgaaactatgggca  
 ataccgaggtttctgttacatctccacgatgctatagctgattgatataatcttgtatatcaataacggggaagtg  
 gaacgcttttcaagtgaactactgtaaaaattcaagtgaactactattgtgattagttgttttctgttgtttcatt  
 ctccgggtttcactttctctctggataatattatcagtaacaaatcaaatagaagtgagttatccagataacaacg  
 ccatgggtcaccattctttctgtccttatctttctgtgtatatcacatcacactacttaataactactcagtagcggag  
 tatcagctctgctacttagttcatataggtatatacaactatacattcatttgattgaaacgatogttttaactctctt  
 cctttgataagtttaataaatcgtattatttctaatttattccggtttgtttttgtattcattataaa**ag**GTGGTTCTCTGT  
 TAAATCATTACGTAAATCAAAAAATGCATTACCTGTTATGAAAACCTACTTCAAATGAGTTTAGATGCTGCTAATGGAA  
 TGATGTATTTAGAAGCGAGAAATTTGATTTCATCG**gt**atggttttattttatgctcatttaatatcatactttactttca  
 tatgattttattttatttatacacataaataattggtacaaggaggtagtaccagatacatatacgcgcacaaatctca  
 ttcgatttgtgtgagggctgtgatactgctcaggtgcccactgaagcagatgggttttcttaagggggccacacacgg  
 agcattttgacctaaaggtctgatccacaaggcagtgagcagcgtgaggagatgcagtcocatggtagccgggtgatca  
 acgattgggttcatacgccttccgttccctcaggatactggagtgcagccatgtgcaccattgggtttggaatgaggggt  
 ttccaactcctctaggtggactttccgtgtccaccaaccgggttaaaagcaccggacattcgttttctgtccctcact  
 ttgggtgaacaacacatccgccacgagaaggcagtgagttaggactccctgaccgagactgtatacgcgtggccatgtg  
 agagcattttggagaggaagagtggtcctctacactctcggccgtaccagggcattttgggggactttcatatgaaa  
 ttcatatttgataaattcaatcctgataatggttgtaaacagattgatggtttcttcttatatcaatgtagtgtcttt  
 taatgacagttggcaggaataactaatccgtaattcccatagaaggaacggttaccactaatctagtctcggcagcaata

E15

tcagtgtttgtagtattcatggtgtagcgttattattttatattctgcatctcctctttttac**ag**GGTAAT  
 TTTGGTGATGTTTATCGAGGTGTATATAATGGATGTGAAGTTGCAGTGAAAACCTTGTCGTGTTGATATGACTGCATCG  
 GATTTACGTAGGAAATTTCTACAAGGTGAAACAACCTGCATTAATTTCAATCATCCAAATATTGTCAAATTAGTTGGT  
 ATTGCAGTTCAATCTTATCCAATTATGATTGTAATGGAGTATGTTCCAG**gt**tggtgaacttttcatgtttttcat**gaa**  
**ttc**tagcatatttcatgttttattcatgttaatttgttatttttactactactagctataagtagttacattttcaa  
 aggataagtggttagcatgtggattgacagtaatttgtgactgcacgttaacagtggttactctcaaaaaatacagaa  
 aatgtcttttagaggattggtctaataattccgtaaccatcggctgagccacggttgagggttactctcaaaaaatacagaa  
 tggttgatgaagtagatcttcagcaacatgtattatgtatatccaacgaacgctctatcgtaaacgacggttactggcc  
 tgcaagtaagttggagaaagatggagacgataggttagtagcttaacgctcagtcocatgaattaataaaagttattct  
 gagacatattgatagatgcaaacactaccgggttacactcctcaaacggttctcaaccaacgggttaaaaaacattgtgtca  
 tgtttgagaatcgggtcaaaagagcgtaaacatatttggatattgcttacttc**gga**tcctcaattttgtaatcatttt  
 ttcacaccttatcaccctaatacattctgtctacagtcacataatctgtttctaacttttttactatcatggttct  
 aatgtgggtgttatattgggtagttttgcattttgtgtaccgaactatacgttgaatttgtctgggacatcattggaag  
 tggaaacaaatagttttctcagaaaaatgaagataggtgacaggaaaaattattgttccaataaggataaaaaaaaaat  
 ttaattaagtgaatgaacttta**gaattc**ggttgtgaacgctcagtaactgcaatacatacaaatcccaccaactggtcca  
 aatgcaacaaagtgcgttcattaaattccactttagctacagttcagaactgaattaaaagtatgaaactatgggca  
 ataccgaggtttctgttacatctccacgatgctatagctgattgatataatcttgtatatcaataacggggaagtg  
 gaacgcttttcaagtgaactactgtaaaaattcaagtgaactactattgtgattagttgttttctgttgtttcatt  
 ctccgggtttcactttctctctggataatattatcagtaacaaatcaaatagaagtgagttatccagataacaacg  
 ccatgggtcaccattctttctgtccttatctttctgtgtatatcacatcacactacttaataactactcagtagcggag  
 tatcagctctgctacttagttcatataggtatatacaactatacattcatttgattgaaacgatogttttaactctctt  
 cctttgataagtttaataaatcgtattatttctaatttattccggtttgtttttgtattcattataaa**ag**GTGGTTCTCTGT  
 TAAATCATTACGTAAATCAAAAAATGCATTACCTGTTATGAAAACCTACTTCAAATGAGTTTAGATGCTGCTAATGGAA  
 TGATGTATTTAGAAGCGAGAAATTTGATTTCATCG**gt**atggttttattttatgctcatttaatatcatactttactttca  
 tatgattttattttatttatacacataaataattggtacaaggaggtagtaccagatacatatacgcgcacaaatctca  
 ttcgatttgtgtgagggctgtgatactgctcaggtgcccactgaagcagatgggttttcttaagggggccacacacgg  
 agcattttgacctaaaggtctgatccacaaggcagtgagcagcgtgaggagatgcagtcocatggtagccgggtgatca  
 acgattgggttcatacgccttccgttccctcaggatactggagtgcagccatgtgcaccattgggtttggaatgaggggt  
 ttccaactcctctaggtggactttccgtgtccaccaaccgggttaaaagcaccggacattcgttttctgtccctcact  
 ttgggtgaacaacacatccgccacgagaaggcagtgagttaggactccctgaccgagactgtatacgcgtggccatgtg  
 agagcattttggagaggaagagtggtcctctacactctcggccgtaccagggcattttgggggactttcatatgaaa  
 ttcatatttgataaattcaatcctgataatggttgtaaacagattgatggtttcttcttatatcaatgtagtgtcttt  
 taatgacagttggcaggaataactaatccgtaattcccatagaaggaacggttaccactaatctagtctcggcagcaata

acttctctcttcttaoctgtagtgagttttccaatatacgttcttccaataactttcgggagggatagaggtcaatatatgt  
aaattgaaatgaactcatctgtcataatTTTTGCTtaataagtaacaaagtaatTTTtagactttatgtaatagtggtgtag  
cttcggtcagatTTCCCTtagtaacatTTTgaatggatggaaatTTTTgttattacactaaacgaaatctgtagtggtg  
gttggtatgTTaagccacatagcttattTTGagcttccgaacagaccgattTgttcaatcttctggagccatactTTa  
actTTgaccgtTgtgtgcacttcatatcctattcatcacatgTTTgtgacttattTTTTatatgactaaaaatattgat  
tcacgctTTggcgaatcgagTTggatcacattgTTTctcaccgcacgtcattTcgcctcgcctctcttccattTccccca  
atagTTTgtctggatcaatgattgTatcaaatatacgtattTcaaacattcattctcgtattTTgacttctTTaatcccg  
ttactcactaacgcgagTTaagTcaataattatataagaaaattTggTgacaagTatattTcaataacaatcattcaca  
tgatctaaatagagcatagggcgcctcaccagcattctccatccaactctgttctgagccttcttccagTgttatt  
catcctcttcatatctgcttcaatcttagTcactatggaatatagagaagagTgactactTgtaactcgattTgaata  
ttgaaactTTattTTtatagTctatgCGTtaatcatcaagTtaaacattTTtaaccaggatcagtaataataaattcga  
ctTTgttagaataaaaactTTacatatctgagcagcatcatcgatgaacacggTggatctgatgcagatgtgaagacg  
cggatcggcaaaccaagagcagcatatctacaaatgaagaatatagggggattcaaaaacagTgccaaccaacccaag  
ttcaaaaTTTTcaatacaaatgtcaagacaattctactgTatggagcgaactTggagaacaacgaaaaacaattatac  
agaagatacagTtTgtTTattTtacagTtTgtctatgcaagattctTcggatcctTTggccagacaatatcagcaacattc  
tactgtgggagagaacaaaccagattccaccggaagaagaatcaggaagaagcgtTggaagTggatagaacacacat  
tcagaaaatcacccaattgcatcacaaggtgagccatcactTggaatcctaaagaccaaaccgggaaggggaaggctaa  
agaacacattgggcccagagaatggagacagacatgagaagaatgaacaaaaattggatagaactaggaaggaagatcc  
aggacagagTgagTgtgagaatgctggTcggctacctatgctcctTTgggagcaacaggcgtaaagTgagTaaagy  
agaataaaaacgataacaaatagTaaatTTcgcgtattctatatctctcattgtatattattcattatatctgtattaa  
aat**agT**GATTTAGCAGCAAGAAATTGTTTAAATTTCTGATGATGGTCAATTAATAAATAGCTGATTTTGGCATGTCAAGA  
GAAGAACATATATATGAAG**gt**atgTtgaagaagtagTTTTatataatgTatagaatgTcggTgTtaggaacgtcattTT  
ttatcaaatacataagTatgTatgcatTTattattTTtaaacacactcaaacagggaacaaatagatatgTcctat  
atgaatcaaatgattTatgTgagggTtgggatactgccccgacagccccaaactgaagcaggtggTtTccttaggaagc  
cattccccgaacctTTgacctacaagTccaatccataaggcagtagatatgTtatatcgtgTcagTgCGTgcctgTt  
taataacacagTgacgatgTcGctaatcggatagcagTggatcagTcggaccattgatacacaaatcCGTgCG  
agcagTctagagTactaatctgTcCGgctatctgTcagccagccagTtaagTccagaacactaaTTcagcctcTg  
cggTatgaatcattgTtctcacacacactgggTttatatacaaaaacaaactgaccacgTcgtaccatgaagTtagagaa  
caacattTgtacagccttagccaaatgtggctgTgaataggggggacagTaatTaatatactgggcataactcagga  
atggTaaatcctacagTaatagTctatgggTcaaaaataaagTcataataaggggaacatgaatatgaatattTgtc  
gataaatcccctcTgactctTTTkgaaagcaatacgtTcatacatctTTTgttccctcaggattctggagacctgTtTc  
acagTtggTtTgaaaccaggattTTTccractcaccagat**ggaT**cctcCGTatccatcaacatggTtaaagcgtcaaaa  
catgCGctTgacgcctTTTcaactcTgTaaacaacagTaatgCGgTgagaagacaatgaataaagTtTtCGTgacagT  
ggctgcataagcgtggTtatgTgagaacattTTTTgaggtcGtactaattctctTcacactTTTggcCGTaccatcgc  
atctggggcaaacagctTgagTtctgagagctatgctatattaaagTgaccgaatccactatctTccattattTytTact  
gaaagattTtatatcayaaggaactTTcGtyatataattcaayTactatgaaggagTtTcacgaaaatatctatYaaa  
ggTcaacacattTctTaaacattacggTTTTtagcCGTactgTaatattctaaTaaacagTgTctTaccgagTTTTc  
TTtatattccttcatatgTcagTtTgtgattcattctatctatctatctatctatYTTTTct**agT**TAAAGTGATAAACG  
AGGTCAAATCCCWATCAAATGGACRGCTCCTGAAGCACTTCGTACTIONTACTATYAAATGTGATGTATGGTC  
TTATGGTGTACTTTTTATGGGAAATATTTACTTTCCGTGATGTWCCTTATCGAAATTTGGTCTAATCAACAACTAGGGA  
TATGATTGAATCNG**gt**taagTtattTaaTTTctcaaatctgTtatgTaaagTtTctTTatgTaaatgaatgaatrgaat  
kaacggaaagTtTggattcacttagTattgTtattTgaatctTcccattgagTgTctaggactgcaactggTcggTc  
tTctattgacatatgTgctTactgTgCGtaatgCctcGatatagcctaaactcacwagcatkgTaaGYaaagatgrat  
agTggstagcartggaatScaggackrcrctTtygTcctattTgggactcatcagatgaatgtacctgTatctcagag  
ttgatgTtcatTctgggactcgaacccaataactgTtCGctTcaaacgcatcGgTtatccactYggcactgagTcG  
tgatagccactTgctagTgCGatgggTgaagTtTaaattcacttagTattgTtTgTtTgaaTctTcccattgagTtT  
ttaggactgcaacatggTaytrggTtygagTcycagagTgaacatcaactcTragatgCaggtayatccagctgacga  
gtccmaartkggacraaacgcgCGTcaaacTggattccactgctagccactatccatctatgCctagaamgtTTTgga  
gtTtagTtctatgTcaactTactTTTtatcagTaatatatacGatacaatcattTctTTTTTTTTcaatattTtctctTt  
**cagT**TATAGATTACCTGCACCAGATTTAATGCCAGTTTGGTTNCGTACATTAATGAATCATTGTTGGCATGATGAAC  
CAATGAATCGCCAAGTTTTTCAAAAATTTCTAATGAAATACAAATACCTAATGTCAATTCATTTACACCAAATTTAA  
ATAGATCTATAGACAATGNTCAATTGAAGAAATCTGCTATTGACAANCTATCAACAACCTCCATTACATTTATCCGTTT  
CTGTTAAAGATAGACGAACCTAGAGAAGATTAATatggatcnnTTTTTTgttactatgcccnnnnnnnnnnnnnnnt  
atacatacatacacatacacaatagwTyatakagatctatatacatacatgTaatTgtgTatagTattgcaatct  
TTTTctgTtktntTTTTgcaagcgcctTattTcmctgTtattcactatggcaactaactTcagTtccgagTga  
caaacctTtctattTattattatcayTaatataattgattcaattcatgatatatggTatacatagTtTgactgTaac  
catgyatatacatataatccactacctatgatcattgTaatTgtcatttatatcattTcattTtaataattTTtatgatt  
tcaattTatactacttatatacttattTaatcctTctTtTcattgTtctTatctTtactTatagactgTgataaaaag  
tagattTaaTTTgtatattctctTcagattacattacattTaccataaaacaaaataatacacatcaatgaaysattct

E16

E17

E18

caaagagatagagagadwggwgdswgwnwnknwnnnkkkkkkwagtgttcaatatgacatacattcayaatgtaatt  
tatgayatsgattgagtaagtgcctgcttaccctgttaatacttatatatatatatatgacgtgatgatccccctg  
gcaatcagtgaggarktaymattcataatctccatggwatataacatagttcatccaatatttcattattgaaatag  
acacttaattacttatgtatacttgaactaactacctacttacttata

Anexo 1: Montagem realizada pelo nosso grupo das seqüências genômicas. Contigs genômicos formados pelo Sanger. Em letras maiúsculas estão representados os exons. Em negrito estão representados os doadores e aceptores de cada intron. Em itálico e sublinhado estão representados os sítios para enzima *EcoR I*, apenas sublinhado os sítio para enzima *Hind III*, apenas em itálico os sítios para enzima *BamH I* e em itálico e negrito os sítios para a enzima *EcoR V*. Em negrito e sublinhado estão representados os iniciadores utilizados em *Southern-blot*. As demais letras, além de A, C ,T e G, indicam regiões em que foram encontradas redundâncias para as bases no

alinhamento, não podendo ser estabelecida a base real, um vez que a seqüência não possuía um valor de qualidade.

## IX.2 ANEXO II

>d000009086.Contig1 d000009086.Contig1:1,23884

ccagaacacctaacaactactgagtgatgaaaatggacaacgttttgtttaatctagctctctttgtgtagctgtgt  
tgaagaaacaaacatggtgtactgataagctcatttattttatgtttttcttgagaaaacaacagagagaacatt  
gtggaaaacatTTTTGCTCTACCGCATGTTATCTTTACGCATATGTGATATGTTTTCTGGCACACCCGTTGTTTTA  
TGCATTTACGAGTGCATTTGACACGCAACATACAACCTTGATTTTTTTGTTCTACAGAACTTTTGGAAAGGATGAAAA  
CGGCATACATTACGATGAATAGATATTTCTATCATATTTTTCTTCAGCAAGGAGACCAGATACAGCTGAACAGAAC  
AAGCTATCAGAAGAGTATTTACAACGACAAGCTCGTTGTGTTGAATCTTACAGCTGACCAGCTCTTGGAGTCACT  
GCTCTACTAGGTGGGAAGTGAAGTGTAGCGGTAAGTGAATCCCCAAAATAAATTGTTGTGTCAGTCTTCGACCCGAT  
GATAATCTAGCTGGTTTCGCCTAGTTGAAAGTTATTGGTCAAACATCTGCACCAGATCATGTTGGGGACGCCTTGC  
TATGCATCTCTTGAACCTGCTAGCCGGCTTAAATAAACGCTTTTCGACATATCGGCAGCCTTCCAAATATACCTTCAA  
ACCTCTCGTTCTCTGGATTTTGATTTGACTGTTGGGTGTACATCGATTATAAACGTTGTTACGTATAGAGGCATTGT  
CAGTCTCGAATTACTGTGACATCTTTGCCTGATAATATCAATGCTTGGTGTGCTGTTGATACGTAAGTACTGCTGCTAT  
CAAGAAATTACTTATTGAACGAAAATGCCGAGATTTTCTGTAATTTCTCTTCTCOCCTAGCTAAACGAAAATATTGA  
GGATTAGTAGTATAAGCTCAAGTGACAACATATGTGATCCCCTGTTAAATCGAAAATCTGTTTGTCTTACTTTAG  
CTCACTTGATCACTTTATGAGAATTGATGATTAATCAACTGAAGTTTACCAGTTTTTGTCTTTATTGTGATGATC  
ACGCCAGTGATATTTGTTGGTTGGATCGTTGGTTGTGAGACTCTAACAGCGCGTAACACGGTTAACGGATTATGAGG  
TTGGTAGTTCTTTGACCATTTCTCGTAATTGTTGGTATCTATTGTTGACACAAAAATTTCAATCCAGTATTGCT  
TTACCAGTGATTATTTCTCGATGACTATTTCTGAAGTTGCCAATTTTTCATGAGTATATTTAGGGAGCGATTTCGG  
GGAACAATATATAAACGTTAGGGTTTCTCTGCTACTTAGCTTTTGTGTTGTTGCAATCAGTACGACTGATGTTAGC  
CTTTTTCACTCAGTTGTCTGTTGCTCGAATTTGAATTGTCATCCTAACAGATAGTGATCATATTTACAGAAAGGTC  
TTTTTTGAATTTCTTAAATAAATACTTCAAAGTTTACCGTTACATTGTGTTACCCTTTATTATATTTCTCTGCTTGC  
GACATTCATGCGCACGTTTTCTGCTTCTGATCTCTGAAACTATGCTTACTTAAAGAGGGTTGTAAGTTAGTTGAGGAAT  
TAGGATTAGAATTGAGAGTCAGAAAATTTGACTTATGACGCTTTAGGGTCAAAAATTAAGGTAGGGTTGAAATTGCAA  
ACTCACATCAGCCATAATGTGGAAATGCTATTTAGCCACATAAATGAACGAATTTAACGTCAGAATGTAAGACTTACT  
GTCTGTGATTCTATTGATTGCTCAATAGTTTTTAGTCTTGTCCGTAGCACCTCGATGTACATTTAATGTCTTCTCTGG  
CTATTTACGACATGGTATTTATAAATGGTTCTTTCAAAAAACGATTACTGGAAATTAAGGGGTGACGTTTTTGATTTT  
TTAGCATTTCTAAGCGTTCTTATCATTGACCGTTCTAATGGTCAAAAAGTATGCGTGAGTATCTTAAATATCATTGT  
AATTAACCGGCTTGACAATGCCGCTGTAAGGATAAATAAGATAAATTGGGTCTCATAGTAGATGGTATCAGTTGTCTATT  
TGTAACCGGTGAGATTGCTGAAACAGTATGATTCTTAATTGTTATTCTGAAATGGAGAATGCAAAATATAACTGAT  
CATACAATAAAAATCAACGCTTGTATTGAAATAGCTTATCTAATGGCTTAAATGAAAGAAATCAAAAGTCTCAAAA  
CCAGCTCGCTAAATCTGAAATAAACAACAACATTATTGACAAGTTTTGGAACAAAATATGCTAGCACCTTATCATT  
ATTAATGTTGTTTTCAACAACGTAGTAAACCCACTGATCTATTAAGGCATAACTAATATTGCGATAGTACTTTGTGA  
AGAATTAATTGTCTTTTTCTGTTTTATTGGTTTTCAATGATCGAACCAAAATCAACGAAAACATTTAAACTGTTTAGTT  
TTCTGTGCTTCCATCGGCCAATGCGCTGTCTAATAATTGCGAATGAAATGATTTTAAGCACGTTGTGTATACATAGA  
ATTTCAAGCGAGACGTGAAACGATGAAGTATTTGAGGAGCTTGACCGTGAAAACCTCGGATAATTTTTTAGGTAATGG  
AAATAGATACGACAAGTGTACATTTTCGGAAATGAGGAGTGTTTGAAATTACGTTATTTTGAATATAGCTTCTGT  
GAATTCAGATGGATACGTAATGTGTAACCTTCGTAATTGAAAAGGTCTTTTACATACATTTATTTACTTTTCAAATA  
TATTATCTTACAAAAATAATAGTATGTTGGCTACAAATTTTCAAGTTGGGTTACCATTATTGAGAATACAACCTTAGT  
GAGTTATGCGTCTTTCTATGAATCAAATATATTTTTGAAGTCTGTGTGCAAAATAGAAATCAGTCAGCTTTTATAGAC  
GTATATTTAGAAATATATCGAACTTCTGATTAGCGTAAACATACGCTTTTTGTAATATAGTCTATGCTATTATCATCG  
TCATTACTGAGCACTGTTGTAACCTATACTATCGTTCTTATATTTAACTGTTTCTACTACCTCTCTTTTCAAAGCATG  
TTTTCATTAATGAATCTGTGATAGTTCATCTGATCAAGTGATCATTGCAAAAATAAACCTTTTTCTGTTTTTAAAGAT  
CATAAACACTGAGGAAACTGTTCTTATATTTCAACGTATATATTTAAATGCATTGAATATAAATAAATTGCATGGTGTAT  
CTCTACAAAATATTTAGAAACTAGGTCCGTCAAAACAACGCAAAACCTGGCAACGTATATATCGACCCAAGTTACCAT  
ACCCTATAAGCACAAAATGAATCAAGCACTCCTGGGAATCGCTTTTTCTAACTTTTCTACTTTTTAATTTAGATAGGA  
CAGTCTGACACTTTATACAAGGCTAATGTTAACAATAATAGAAAACGTTTTAAGTATAACGGCACTAAAATCAGCTAG  
AAGGTTTCAAAGCTACAATAATAGCTAAGGTAGAGATGCTCAAAATATGCTTAAACAATAGAATATTAGAATTTGAAC  
CAGGTGATTGATTTGATGGATTTCAAATCAGCTGGTTCTTTCTAACATATCTTCTGATATGGTGTATGATTATGAA  
GAAACTACGCAAAAGCTGCACATTATTGAGTAGTGGTCTTGTGTTGAAAACGTTGAGACTGGATGTACATCCATCG  
CTTGATAGGGTTACTAAAATGTCAACAGTTTATTCTGCTAGTGAAGATTTCCGGGAGATTTACGGCCAATTTGAG  
TTGGCAAGTAAATAACAGTTGTTGATCAGGTTACACTTTAGTTCTTTGAGAACCAAAATATTTAAACACTCCATGCA  
GGTTGATTAACATAGTGAAGGTATCGTGAACCACTTTTCTGACCACTGAAATGCTTAGCTCGAAAACAAAATAAAC  
GACTGTCCGGTGTTTAGATTTTTTCTATCTAGTTATTCAACAATAAGTCAAAATAGGTACTTTTTCAAGTCATGGTTT  
ACTCAGTTGTATATACACATCTTCAATGGTACCTTGACCCGAAAATCTCATTTTTCTGTTTTTATTTCCATCTGACAGGT

E1

aattcaacaaaatattcagcactactcctcaaggactctatTTTTgTTAAacctTTAAacactTTattattcatgtattt  
 taatttTgtgtatcacaaaatATGGGAAATCATACAGTTAATGAAGTCAGTAACTATTTTTGAACGTGGCAGTGATGTCG  
 AAGAT **GATATCA**ACTACTTTAGCATCTAATAATACACATATTACAGAAAATAATCAATCCGTCTTGGTATCTTCTACAT  
 CGGACACATTTTTAGTTCTCTCTTTGCCTCCGCATAGTCTAGGACATGATAAATCATTATCTTCTTCATCCATTTCAA  
 CACCTAATGTGGACAATATCAATACAAAACGATATACTGATTCTGTGTATGCTTCGAAGTGATCATAAAGAGTCTTCTC  
 AGGTAATCGTTTTAGCCATCGATGCCAGATGAAATCTAATGAAATGCTATCTACTTTTTCTAAATTAGTCAAAG  
 CTGAAAAAAGCTGTCCAACAAGCTTTGTCCGATTTACTTTGATGATCAAAAAGGAATCATCCACTGTTGGTAGTGTGATG  
 GTGGCCAGCGTTCGAATACCACCGCTATGCGTAATTTATTCAAATCAAATTTATTCTCGTGGACGTCGTCGTCGTTTAA  
 CCATGAATATTTTTGAATAAAAACTTGTC **GAATTC**ACCTACTGATAATGAATTAATAAAAgtaagttcaatggtaaccc  
 tacctttatcactagtattgatagaaaagggaaatattcagtcctttctcaatagaaacaagttattataaaaaatctaa  
 ttgtgaaagttttgtaaatattctgggttttacaataacatttagacagatgactcctaaatTTTTtagatccgttaaac  
 ttattggcttTgtggtttattcaatagagtaaatgtgaatatcattttaacctctttactgaacacaagctctaagaac  
 tcaatcaatatctgtttatattatcgtagcttcttaatacaacatgcattgtagtcactgttctgaatgattg **gatatc**  
 acgtggactatgttttaaacgtttagtttagtttccgataaccgcacggaaaacctgaaatgtcttaaaatatgaaggta  
 aaatcat **gaattc** cctcaaaattagaggtataccttgttTgtgccaactagtaaaaaacatgggttagtTgttcttTgt  
 ctgctactttcaactaagtagcattcaggaagtcctttcaagatgtccggtactaagtcactaaagattaaattgt  
 ggagttttcattccgaaacataaagtaatttaattaatcccttatcagtagataggttttTgtctctttaaagataat  
 aaaggaattatgtaatcacaattgattgatcactgggtgggtgatcaatgtcatacgtgtcaagtccttcttagtTgt  
 gatcgaatgctatcgatcaagttgcaattgctcagtggttaacgtctctgactgtgaagctcagtgacacgggatcga  
 tccgtcagggagcatcagttcccccaagattacaggtacacctTgtgacgagtgccaagtagcatgaaaccc **ggatc**  
 cagggtttctctgttgaccacctccaaccacctcttacctcaatataaagggaaactatgtgtatgaagctaaaaatt  
 gttggTTTTctattgtaaccttttcttctgggttaaataagtttaaactTggTTTTgtttTgcatatgatatgcct  
 tttgtttgttatataactttc **ag**CTTTTACAAAATCTTGTATTTCACAAAATCATCAAGTACGTATGTTTTCAAC  
 ACGTGCATTATTTACATAAATTAATTACTGCTGGACAATTAGATAGTTTTAAATAAATTGTGAACATTGTTAAGACGAAC  
 TTATTTAGAAATGTGAAGAAGAATTAATAATCACAAAGCGCGTCAG **gt**atgctgcaatttattatattccaagtttatcta  
 TgatataatagctccccctcggTtTgtctatttctTgtctatactgttcatctcgttagcttccattcgtttgttttatgact  
 gcgTtatcaatatagtggTTTTgacctattctctctttattTgttagtaagttcttcaattggctatgttgatactcaatg  
 TgttatttaagagtaaggatactcataaattaccagTactTgatcacagatatgctagaaatattttcagTatcttc  
 TgtgatcaccgggtaagtaattagTgagaatagacgcagggTattagagaatgatgataaataccgtcgatggggTgt  
 aaatcttcatcgactgagatggTtgggcccagTgttacgtatgcctgaccaccggTtaccacgacgtgcatgttgact  
 agTgtTggagacgattggaagaaagTtaggtgagaccaaaccaaaccgtggTatcagTccttaaagTcactaacttct  
 gatctgaaccatgtTggtagatgcagactactTggTtggggTccgcgtgactatcgtaataatggtaggggtttctg  
 ggtgacatgcagcagaatctatcacaatagcgtagTtTgtatgcactcctcgtcttcccttaaactgtgagattgaaat  
 tacttcatatctttctttctacgaatcaattcttctcgtatcatgtctgatttttactactattgaagTgactgctct  
 tatgaattTggTgtttgtttTgtTgtgctaatgaggtgtcgaactTggactgatgcataatgtacgtggTcctacg  
 ttatagctgactgaataactgactgactgTcatctgagacctgtacattttTgtgtacggTcgtacattgtttactgt  
 ggagTgggggTcTcaagattgaaatattTgaatttaagccactaagTcat **gaattc** gaatcctgatatacttactttg  
 gTttgaggaatcagacggTatcactagcctcacacttatggTatgaactTgtaactggTaaacgagTtTgtgtcagT  
 attTgtttataactgagTgactgaatgacaggtaaattagTaaaatcaacagTtaacttaaagTatagagTttttTgtg  
 TgtgttttctctctgttaatacataatgtagaatctagcacaaactaaactagTttaagTtTgtctgatttctcatcga  
 tcatagtaggaaactagTaatgtcttaataaaattTgtcaacaaaccatttttacttattcatcaatgagTggTtatga  
 gatgagTtgcattaggagTctgtaactagtagcgtTcactcatgtTgtggTaatgacacttaccatctgatgggatta  
 gatcattTgtagctctatttcacaaatcgattgaaagTtggccgtTgtggaaaactTtTgtccaagTccagTggTtaaaaa  
 ttatgTactgaccccaaggcaacctctgaccagTgtcataagTgtagaatgttaaggagTctTgtTgtTgggatgatac  
 aggcactTggTatctctTgtctttcaactaatgttttaactaacaatgccaagTcagcattTgtaaaactgcctttaaact  
 aaactattctTgataaatctgTcgaactaaattTgggagaatatatattTaaaatagTaaataaaacaaactcaaatTta  
 agagTaaataacaatatgCagactTgagatgatttagaaaaagagaggaataagTtgacctatgTcgtcgcTcatt  
 gattttgaaccgtatccttcgatcactttTgtcggggTattTgcatcgatctcaactatgtagTgcttaccattcatt  
 caagTtagattaaaagTcttctgTcttcaatttaacgtTctatacatttccagTttttacacagatggTtcagTcagTc  
 agTcacagcgtagaaattcgtacgtacgtacatcagTtcaagTtgccataccacattaacacagatttagTtTgtcga  
 tCGaatccatagTggTagaagTtagtaagattataagcattaaTtggaaagattaaggTttgaaagTgttattgaaag  
 agTataatccaatgaaataaatTgtgaaaagagaaaaagataaaagacatgaa **gaattc** agatgactagaattTtTgcat  
 ccactctttTgtgttct **gaattc** tTcatgcctctatctttttctcttttcaaatTtatttccctggattaaactcct  
 caataacatcttcaaacctaatctttcacactaatgcttatactcttactacttctaccactatgggattcgaatcg  
 acaactgcatctctgtTtactgtTgtatggcaactcgaactgatgtacgtacgtacacagatggTtagTttgtTtag  
 tataat **gatatc** tTocataacgtattgatacatgatgtcaataaattaggcatctagTgtTtattctagaaatca  
 caccaattacttttacatgggtTgtggTaaacattacttactgaagTtaggattaaatgggaaagaacacccatctgag  
 atgagaagaatagacgtatgtTttaggTcattaaaattagTgggtTgatttaagaacgatcatatcaatgatgattt  
 cagattggacaatggTtTgtttaccatt **gaattc** attat **gaattc** gaattgaaatttatgtagTctactttcataatgt

E2

gattttgaggaatcagacggTatcactagcctcacacttatggTatgaactTgtaactggTaaacgagTtTgtgtcagT  
 attTgtttataactgagTgactgaatgacaggtaaattagTaaaatcaacagTtaacttaaagTatagagTttttTgtg  
 TgtgttttctctctgttaatacataatgtagaatctagcacaaactaaactagTttaagTtTgtctgatttctcatcga  
 tcatagtaggaaactagTaatgtcttaataaaattTgtcaacaaaccatttttacttattcatcaatgagTggTtatga  
 gatgagTtgcattaggagTctgtaactagtagcgtTcactcatgtTgtggTaatgacacttaccatctgatgggatta  
 gatcattTgtagctctatttcacaaatcgattgaaagTtggccgtTgtggaaaactTtTgtccaagTccagTggTtaaaaa  
 ttatgTactgaccccaaggcaacctctgaccagTgtcataagTgtagaatgttaaggagTctTgtTgtTgggatgatac  
 aggcactTggTatctctTgtctttcaactaatgttttaactaacaatgccaagTcagcattTgtaaaactgcctttaaact  
 aaactattctTgataaatctgTcgaactaaattTgggagaatatatattTaaaatagTaaataaaacaaactcaaatTta  
 agagTaaataacaatatgCagactTgagatgatttagaaaaagagaggaataagTtgacctatgTcgtcgcTcatt  
 gattttgaaccgtatccttcgatcactttTgtcggggTattTgcatcgatctcaactatgtagTgcttaccattcatt  
 caagTtagattaaaagTcttctgTcttcaatttaacgtTctatacatttccagTttttacacagatggTtcagTcagTc  
 agTcacagcgtagaaattcgtacgtacgtacatcagTtcaagTtgccataccacattaacacagatttagTtTgtcga  
 tCGaatccatagTggTagaagTtagtaagattataagcattaaTtggaaagattaaggTttgaaagTgttattgaaag  
 agTataatccaatgaaataaatTgtgaaaagagaaaaagataaaagacatgaa **gaattc** agatgactagaattTtTgcat  
 ccactctttTgtgttct **gaattc** tTcatgcctctatctttttctcttttcaaatTtatttccctggattaaactcct  
 caataacatcttcaaacctaatctttcacactaatgcttatactcttactacttctaccactatgggattcgaatcg  
 acaactgcatctctgtTtactgtTgtatggcaactcgaactgatgtacgtacgtacacagatggTtagTttgtTtag  
 tataat **gatatc** tTocataacgtattgatacatgatgtcaataaattaggcatctagTgtTtattctagaaatca  
 caccaattacttttacatgggtTgtggTaaacattacttactgaagTtaggattaaatgggaaagaacacccatctgag  
 atgagaagaatagacgtatgtTttaggTcattaaaattagTgggtTgatttaagaacgatcatatcaatgatgattt  
 cagattggacaatggTtTgtttaccatt **gaattc** attat **gaattc** gaattgaaatttatgtagTctactttcataatgt



E3

cttatattaggacattccatgtgttaaaatgaatatgagtcocacaataggacgaaacggccgctccagtgctttcagg  
ttttccacaatggcttagcttcaattgactcatgatttaaactatataaaaattactaagatctccgcaaaacccctt  
ctgataatgattatagctcatttagtgactggctccatgaggtatctcctggagttctagtgaggaaagcaatgagcagt  
ggagttcaatcaggtctgttggga**gatatca**aactcactgaagacattgggtgaatgggttgctaaacttcgtggattggg  
tgaagttagacactaacaccggttggatgcccggctcagtggtctagtggttaagcgcctcgcgcgcgagagtggttaggt  
ctgggttcagatctcgcgaggtgggatcgtggatgcccactgctgaggagttccacaataggatgagacggccgctcca  
gtgcttccaggtttcccacgggtggcttagcttcaattgactcatgggttgcaactacatagaataaataacttttt  
taggggaacggtttttttgagtgtaggtatgtaccagattttgatcaataatcaattgattttaaaatgtacactaac  
tcgatttaaggaaaaatagcagttatatgacctgagggctcatctattgatcttattgtaggttcgaacctcatcat  
gctcgcggaagttcaaacggccgcttagtagttacttcactggctgtgtgtgaattaaggctgaatacataaattctt  
acttggtaaaatagtttggatgatttaacatcaattatatgtcgaatcaaagattttcgggatgtaacctccaga  
gttattttcgttttaattttattgacagtttcgtttattgcccgtgtatcgttgaatgatcacaatgaacgatattgaag  
taaacataaggtactaggcaagatgttaaatcagttggtaactttgtgggtcttcgtcgattatagtgattagaaaa  
cgtgttgtgtatgcctaatttactgcttactttgactgggtggatctagttgggtgttaaagtcagttaggagaaagtt  
aagcgcctgcgaaaccaaacatcattatatgaagacaataagtggttggacttagttatgcaagaaggttttagactgct  
ggatgggtctactcaataactgcgatcagtgctcgaagacctgggtgacataattcacaatccattacaatgatacag  
atgatttcgaactatattttttgaatattgaatctcagttactccttgctacacatcttttcatactgtcttttcga  
accatattatcaatggtcattctttgtcattatcactactactacaaataacttcogatttcogttactattcaagtt  
attaggctaaaatgatatgacnnnnnnnnnnnttcaattgactcatgatttcaactatgaaaatgttattttccac  
tttatgggtgcgatgttgtctgtttgattagtatataaactcagtatgtttgaaatacaatgattcatacggcaggggc  
tgttctggacttaactgggtggcttaggcagaatcaggaccgacaagtaacttagattactcggacgggtcccgtata  
tcattgatccgatcaataattctctgttctctattcagtggtcagctcaggttatacattcacagagcattcactcgg  
cccaagatataacattgtattttctttcatgttttttaattttac**agTATATGTTAAGCTTAAAAGCATATGAAAAG**  
**AAATATTTCAAATATATTGGTCAATTATCAAATATACAATCGAAATTAACAAATTTATGTGATAAACgta**agttattt  
atftaaatacataaacattgggtacaaggaagcaccagatatatatgcccacacaaaatcttatatgatttgggtgagg  
gctgttaatactgcttaggtgtccaaactgaagcaaatgggtttcttaaggggcccacaccgggagccttcgacctaaag  
atctgatccacaaggcagtgagcagtcgtaaggagatgcagtcoccatgggtagcgggttggttgtattctattgcaaaaa  
tattttatattctatttacacataaagttatctctaagttatcaatgaagccaattaaattgtttttaaaagtatgct  
aaatacatttatatattgggtgttcatcttgggtgactatcataaccaatgtttggagactctaggtgatatgggtcaaa  
atcgatcacaatgggtcaggtatatacattctttatcttcccttaaaccttgagaataaaaattgcttcataaattct  
tccttcctatattatataccttatataactaccaccaccactaaattaactacttctatgaactcgggtgttgatcttgtt  
gtgctaacgaggtatggcaacttgaaccgatgcataaatgagcctggctcctacgttgttagctgagtgacatttatac  
catatcgtattgagcataaattttgttgggtcatalctaatgacctcttctcttcttcttcttcttcttcttcttcttct  
ttatcgtcttttcgctcaacttttgtcttattcatttgtctttattttattaacatcaaaatataattacttacaatatt  
tcacataataagtaattgtttgtttgttttcttctcttcttcttcttcttcttcttcttcttcttcttcttcttctt  
tgaatgaatagaatgatataatgatactgtttttttaaaatttccattctatactgggttgaataaatgaagttgatta  
attatagttacttatatgacagatgattttgttaactgagtggtcattggtaaatggatatacaattttttgttgggtgtt  
gaaaaaataggagaaaaaaattcaattttttggcaagaaatgaattaatcagaagtatactgaaatcccctacatgccc  
tggatggccgagtggtgggagattcagctctccctctcgaataactctcacatagccacgcgtatacagctactgtc  
agggaggtcctactcaatgccttcttgtggcattactgttgtcctaagaaattgagcgggacaaaaagtgatgtccggc  
gctttaactgggttggtagatatggaggattcatctaggggagttggaaaacctcattccaaaccaatgggtgcacat  
gggtccagtatcctgaggggaacaaatggcgtatgaaccaatgggtgggtcactgactatcatgggactgcactcctc  
cagatacttctactgcccattgtggattagatctttagggtcaaaggttcgggaaaatagtttaagttgaattagggaaat  
gtttttgttagattcatalatataataattgtcaattcogattgatatggaagttcatagagcaaaaacaagtcata  
tatcttactatgcaatagtagataatgggtgaggaagtggtttgataagaccgatgtattatattactttcctttt  
attgtccatacaagtttttttgaatcccaccaactccacttgttgggtgatgataaattatagagtggaagaataac  
agtaaccactagtaataaccattgtaactactatagcgtatgtgtctgtcatatctgatgtaccatcagaatgacatct  
tacaatcttgattgttgaacgattactattggattttgttgatgaaagatcaaatagggaaatattcattccacg  
atgccttgatgcatcactaacgacaggaaccagacggttgagaagtcagttgattaagtatacctattttacatgtacta  
gtttattgtactcaaatctcagtcagactttttatgtgagtcctagtgatagttttgtctattttaaaccgaaacag  
atgtcgcgaggtcgaacggttcaaaatgatgaataagagtcgagtggttttagcttcattaactcaacttattcaactt  
ttagacactaatcagctcttaaatttaaacctcaatttgaaaagtttagtttgaccagcagtggttatcattagctgttg  
catgaaaattgctgttcaagttgagtaaaataaaaacctgtacttggatgaatgaacttcataaattacatacgcctga  
tcaccagtgncatcaacatgcagcgtttgctagtgccgaagttgattgacaaatggctaggtacggcttggaaatgaat  
taat**gaattc**cataactgttgggttggatgcttttgataagttcagactacctgtttgagtcogctgggttactata  
atcaataaaaacattacagaatatcgcaatgcttcagatgcataaactttttctcttcttcttcttcttcttcttcttct  
attatgttgtgtttctaatccattttgcattagttgtttgaaatgttcccaataatgtttaagactgcaattgatcag  
ctcttattgacatatgtggattctgtacggacacctcgatattgcttcaagtcataagcattataagctcagatgaaa  
gtgactagcagtggaatccaggacgcgtgtttcgtcctattttgcgattcgtcagtcacttagcttgtgcattgggttat

ctggactcagtagctgaatgggtaacgcgatggattttgagacacacagtagtgggttcgagtcctggagtgaaacatt  
aactcggatgtaggtgaccccggtgacgaatctcaagcagcacgaaacatgcttcctggattcgcactgtagccacc  
attatctctgtttatgttgattttctgtttctgcaaattcattcctattttctcgatttatgctatctgctgtttgatgt  
tttctactactcttctgagatttatctcaagaatttcatctgttggattgatactggataaacgacttcaaccgatg  
cacatatatacctgggttttaagttactaataactgactgactagtgaacaactgctatgtaactggttcattgcttgt  
tacatagtcataataactctccattttctgtcaatttacaaaaataaaaaaaaaaacacgaaagtaaatgatttccgaaa  
acagcaaagaaagtttagtttattgattattcgtgttactaataatagtaataatagtagtggtaatatagtggtggtg  
gtgtgcatgtgtgtagattggcacaatcatcatgctacttactgatcacaacacacacacacatttgcatacagata  
cataactttcattattttcttccattgggtacttatcaataccaatagagtataatattgacttgaattattataattat  
gaactacaatagtgtaaatgattaactatgaatgaatgattggaatgtttattgtttttagaactgctcatttagta  
ttgtaatgaatattacattattgattacgaatgtatacatalataactatttgagataataatgatgaataatggtaaa  
agttattcaaaaaaagaattcattattaagtgtctctggagtgagactggtaggtcctgggtttgaaatctcgcgagtg  
gggatcgtttgatgtgactgctgaggagtcacataataggacgaaacggcctccagtgcttccaggttttcgatgg  
ggctagcttcaattgactcatgatttttaactgtagatataaaattactaaaaatctccacaaaaaaccacttatca  
tatgctcactagtgactggctccatgaggtatttctggagttctagtgagaagcagtgagcagtgagggttaaccag  
gtctgttgtgagctatcaactcaccgaagacattgggtgaaatgggtgcttgacttcatggattgggtgaagttagaca  
ctaaccaccttggatgccggcttagtggtctatcgggttaagtactctggcgcgagactgataggtcctgggttggat  
ctcgcgagacgagatcgtggattgacactgctgatgagtcacataataggacgaaacggcctcagtgcttccaggt  
tttaacaaatttagttcacatcaatgaatagtgaaattcaatgaaccattcatcatgacttctctgttttttaaccaa  
ttcattttatacatgaatacataaataactaacgattgaaattgatgatcacatatagaaattaactattctatctcatt  
atggttgtatcattcaagtacatgtaattttataccttgacctttgaattgatatttagaatgattacgtgtatacaag  
tcataatttaataataatttgattttgtcattgtgtgtagcacaagcaatttgatcatgatggtttatgtnnnnnnnnn  
caatgataactgttatcttgactcattgatcatactcttgtaacacacgatgatcaaatcgtacaaatggttggattga  
acttcttccagttgtaaatattcactaaaggtagaatttgttgtatctcgaacctagatttcttagtctatcgcgttcg  
aaacgtttttcttgactctccagcaaggcttcaaattgtttaaaggtaaatgatatctcgataagctaaccctcgc  
cagttgttgtatctcgaacctagatttcttagtctatcgcgtaacttgaacactcgaacgcgacagaacctctcaag  
tcacaaatcaggagacagaagacaagtagaaggaatgttatttccaaacagtggtgattatggtacaaaattgtcaggt  
atztatagtttttagtagaacaataatcatcattatagttgtccaatcgcacacagggagttatcagcattgtccag  
tagggaagctccacgtagcaattctagagttattcttccaaaagctgattggatggaatagtcgggctcctctcgggc  
ttcttagatcttctcgcacttccctgtggacctcctcacagaattatttctaataatgatagtcattcaagcataggt  
acttgttgtctatgtttatttccattattttaaattatgattagattaaagaagatctaaaagtagcagggtttcttcca  
agaaaggagcgggttacttagttgttactaggtatcagtggtgatgtgcatggttgatgagactaggcaatcatcatt  
ttagaaaccaattcattatccattaagttagttaatcacattaacctccagataattcaagcattatgacaaaatttaa  
acttttacaatgaaagtgtttttagtaccattgaataaataactagcaaatcaacaaagacattaaagattccagttag  
tttaaatgataaagtttgaactggacaaaacatctgtaccgtatctcgtacttagtcaactatcaaatatttgtataat  
attttgaacaaaactgggtcatgaaataaggtatctcagtcagaagaaaagatggatgggtggcgtggcagtggaatcta  
ggactcacgtttctgcccatttaggactctcagctggatgtatctacatccaagagttgatatccataggtc  
gaacctagtagcgttcaacttagaacgccatcgcattatccacttaactactgagtcctggtagccacctgcatgtgca  
atgggggtgaagtttaaattcacttgggtacttttacttgaatcttcccattgatgtttacaactgcaattgatcagtc  
tattattggcacatgtgcatatgccaatgacagattgatcaattgacagtcctaaacatcaatgtgaagattcaagtaa  
acaataccttagtgacatcagaagaaagttatttgagaaactgttataacgtgtgctgcccagttatttatatgaacgtt  
atgagaccgccaatcggagagcagtgattttatcgaccagatccaatgatatacaaaaactgtatgcccagtccttgagt  
actcgtcagtcctgctttctgcttagcccaaccagtttaagaccagaacaccaataaaaagcctctgaaatatagaaacat  
tattcacaacatactgggtcagataccaacaaaacagaccacatcgtaccataaaaatagaaaataaacatttgtacaa  
gatttggccaaatgtggctgtgaatagggagacagtaataatcaatagactgaggataaactcaagaatggtaattcgtat  
aataatagttcataggtcaaaaataaagcttaggataagagaaaaaatgaatatgaatagtaagcaagatggatagtg  
ctaggagtggaatocaggagcgcgcttctgctctatttgggactgggtcaactggatgtaoctgcatctcagagttgat  
gttcgctctgggactcaaacccactaactttcgttcaaacgccatcgcgttatccactcggctactgagtcctctt  
tgcttatcatgcttgtgaattaaggctatatcgaggcaatacgcacagtatgacatacgcacatttagagaccagttg  
tagtcctaacacatcgatgggaagattcaaacaaaacataactaaatgaatatgaacagtttagttacttagcaattat  
acaataggaaacatacatalatcatalattggtccataaataagtcctcaacagttaccattcataaactccacgggatata  
acagaaacgtttttattctttgaaatgttttttttttttttttttttttttttttttttttttttttttttttttttt  
agGTTTAAATTAATAATCGTAAA  
ATTATTGACAGTCTAATGAATAATTTAGAATATTCTACTAAAAAATTGCATTTATTACATAATGAATATTGTTTGGCA  
TTGATCACTGCTAATCATTATCAACAATGGTTATATACACATTTACGACCATGTTTACTGAAAGGTGTTGAACGTACT  
ATGCAATTAGCATCGGAAGTTGTgtaagttgttgggtttgatgttttgcagtatcttctcaactgatcgatgt  
tgattactttaaatgttttgaagtactcgattttcaactaaggaatagccgggttccagtataactccatttgccttga  
tttgggccaacagatcgtattattagccaaaagtatactgatgtgacctgaagttattatataataacctgcgcttg  
taacatgagttatgttagcatttcaactctactgtgactttgttaacatttcatgatttagtctggccataagactt  
tggaaattattcacgtcaacgtctgtaatacaagaatcgaactttatgtttgaggcaaaagtctcactagtaataactaag

E4





gtagtgatcaatTTTTatgTTTatctagtcattatTTTgtgattctattccatgaataagTTgCGAAGCGGTcattagtt  
tatttaagtgaactaagctcgtcgtgtgTgatcctggcccttctatagattgattccaatcactatactTTTctgTTTT  
ctaatttccatattctgatttcaaaaagaaggcttcaTTTctaagtctataatgtaactTTTggCGAAGCAAGGATC  
acattTTTacaattgaaggacatatgggactcaaaaactgtcagccaatatcaaaatcagaatcttcaatacgaac  
gtcaagacagtcctactgtgCGGagctgagactTTggagaaccactacatccatcctcaagaaggtacaagTatttata  
aacagctgtctacgcaaaatactcaacatccattgaccggataccatcagcagcagcctactgtgaaagaggacaac  
cagcttccagttgaagaggaaatcaggaaaagacgTTggaaatggataggacatacattgaggaaatcaccgcactgc  
attacatggatgaacagcatcatcaaaccgcattacgaggcatgccctaactTggaatcctgaaggaaagaggaaaag  
acgaaggtcaaagaacgcactacgcagagaaatagaaggaattattaaaaggatgaataacaactTgaaaggattgcc  
caggcagggTTgaatggagaatgctaatgggtgtcttatgctcctccaagaagggtacaagcgtgaagtatatgtgac  
Ttaaacctatatgtagtctgtactaaaagaagcaagagtgacgatagatgtctcaataacatgtcagccatcacattt  
aatgtatcatggTattTTTTatatacatattgtatcacactacattTTTtagagaagcacatgcaatcatttctTgttac  
cgtgctcTgcaatTTTgtattcatgaccaatatcgagaggattcacaatgattTggTcaaatatgattTgctaaatt  
gaaggggTTgagTatcagggcaagTaatTTTcagagTTgttaccatcttatcaaattaccatatttattctgTTcta  
aagtgatcgtgctcaaaggaaagaatgTTTgtatgaaacatcagcacggTattcaacgTaaagTTgtcgtccaaaaca  
tgagggctcaaaaataaaacacaaatcttacaatgtacataacatcttataaagactaaagaaacaggagagTaaaa  
taaTggataggtTaaatactTTTaaaaaatggaacgccactgTaaatccatcaataattcgcagTgaacctgaga  
aactaagcatattctcgaagagcagagatgTTgaaactgccataattcctaggaaaattctcctaacgTTTgtg  
tctctcgtgattgattagttaccaattatgTtactaattTTTcgtTTgattctaaagcttctttagaatgagTTTTa  
cacgTaaacctataaatttcttatattctgtacctaactaacctTgaaagTgaggactctTTTcatgTtaccctTg  
tagcgaacaaaacatgggCGGGGacaatcgaatTTTaaagcacaattacaactatcttactaagatctgTtaac  
catacagaaaacagTtaattTgTaaaatagcaatcaatcgtctcaatctTgaccgTtctTTTctcaaatatcagTcca  
TTTTctcaattctattgTtcatgcaTTTtctaccaattgcgcttcaTTTcagTtctTTTcttattggTcttctgcc  
aaaatacatcctatactgactatcactatatactactttagaggatataagtagaccataccacacccttattTTTg  
TgtctgTTTgaggtTaaatgtaggggatttctgagaaatgagaaaacagTtaaaaagTTgatgTtaatcattcaac  
tTccaaagTtattagTttaaataatacaaatgTctctTTTaaactTgatacctacgacatatagTtagtccgTtatt  
gTaaaactaaatgTgTcaTTTaaactgTgagTtTgtctTTTcaactcaattgaaatgTattgacaatattattaa  
attatagTaatattTgtTaaactgagTaatcttactTTTcggTcgattTTTgtatggCGGgaatatctTgatagTaca  
TTTcatcgtgTTTgaaaggatacgtcagctgaatactcctgcaccccagTattgatattcacaatcatattTgaat  
gcaatactTgtatcacttcaaaatcagatcattTTTgcggtTgatgTtaagatgTtcaaatggTatattTTTgtcttaa  
atatcaataagaaagTactgTaaatgcgTtTaaagTtTgagTtTgatattTTTcattTTTgtgacaattcattTaaagTca  
agTcaagagattaaaatctctcattTtaggtctTTTgTaaactagatagaatcactTggaacaaaatgggTaaactactc  
TcagTtctTTTTTaaactTaaatcggTaaatacaaaaacgTTTTctataaagTaggTcgtTgattTgaaacaccgctctac  
atataacagTtTaaagcagTtTgatcgtattactgTcctTcagGtTaatagTattaacaactggataatataaactcgt  
actaatgaacatcataccattgaggcaggattagTtTgctatctgagagggactaaactTTTatggacgagccaatctg  
agcttctcgcgggTTTTcaaggTcacggcgtcaagTtTggTacctgatgctTtacgTacgatcggctTcacagcggattt  
tagagaagacgTgataattcagcgtctacaagagaagTgattataagaagTtctatactTgtgactgTggTactcaa  
atgacctTTTTcgtcTaatcagGattaaggtTaaagacctaacctaaacctataccaacatccaacaacattgaaGt  
tatgtggTccagggtaaaagaatTTTcgagaccttatcatgTctctagaggTcgactactatggagccatagggatga  
gttctTTTaccgTatgcaactatgattTtagaacttctgaaacctatatactgacctTgagaagTTTTTgagTcatgTaaa  
aggacagTatccgctTTTaaTTTTTtagTTTTTgtataatataTTTTTactgTataactgacctTctgattggTaat  
gTTTTctcagggccattTaaagattcgtTcaaaggTcgtccataaagTtTtagTctcgcgctatctgaaactgTtctc  
agTTTTgaagTaatcagTattggattcaagTctcagTatagatatcaaacactagaatgTaaatggatctagacatcaga  
tcccgaataaaaatgaaatacacttctTgaatgTgctataatatacattccatataatctcTaatctTTTatttctc  
catttctTTTTTgttTcactactTTTctTaaatagTCTTTGTGAAGTATCATTTCGTTTAGCAATATGTGAATTTG  
AATTATGTTTTGCCAATTGTCTTGCTGATTGAGAAGCTCAATTAGTCAATGTATTAAGTATGATGCACAATGTGCTCTCA  
ATTGGCTTTCATTTAAATTACTTACAAATGATTCAACTACAAGTTATCCAAATACTACACCAATTACATTTTCACGTG  
ATCCTTGTGATATTACTACCACCACTACTACTATTACTAATGGGAATAGTAACAATAAATCATGAAGATATAGATA  
TACAAAATGGGACTGATGATAATTGCAGTCAAGCTTTGAATGGTCATGATAATTCAGTgagTgTTTTgTTTTcaatgT  
tcaatataccaacgTtatgattctgTgtagattcacagTctctTTTctatcttctcTacagaattgacaatgTtctTgat  
ttatatttctTaaTaaactgTtctagTtTcatagTattgTtTgTtTggatcttctcattgagTtTtatttaggactg  
caactggTcagTctcTaatTggcatatgTgcatactgTgcgTattgctcogatatagcctaaattcacaagTatggTga  
agcaagatggatagTggctagcagTggaatccagTgtgacgcgctTtctgTtctTTTTTgggactcgtcagctggatg  
tattTgcatctcagagTtTgatgTtTcactctgagactcgaaccagTacctTtctgctTcaaacaccatcgtgTtatcca  
ctcagTactgagTcctgatagccactTgctTgtgcaaaaaaacatcaatggaactcccccc

E8

>333748.c001016971.Contig1

cgtgtgtgatgatacccttgtcttgtgctttttcaatctgtaggaggtccccccacattctaatttaataaaaagg  
gtcaccaccattgcacaagcaagtggctatcaggactcagtgccgagtgataacgcgatggcggttaaaagcgaat  
ggattaggttccaggtcccagagtgaaacatcaactctgagatgcaggatatactagctgaagagtcctgaatggga  
cgaaacgcgctggattctacttctagccactatccatctttgcttattctagtttcgtaaacccgacgaatcctt  
gtagacagttgttaggaagaaataaaaaaagtctttcgaacaatacagattcaatgattatataatgacaagaaactca  
ttgttatacaataatgtcaacattcogatgaataatcttctgctgagtcattgcaattaaatataaatcatcctaaac  
cttgatttctggttagctggtagctcatgaagaagacgtatgtagatgaatccaattcgaactccacgggtgtaaatc  
agttacgaagaaataaatgactgtatgattgacatcttgatgctcggaaagatgagatattgpcacaggaatttgtt  
tcagtcagtcagccagctacaacgtaggaccaggcacatttatgcatcgggtccaagttgccatacctcgttagcac  
aacaagatctataagtaattagagcgtattgaaaatttagttaggcttagattatataatataatataatataata  
tatttaaaagtaagctttcagtcctttaactgtattctcagtgcttgtgaactcctcacgggtattcatgatcatca  
gacagagaccttagctgacatcagtgaaatgggtttatgatggtgtagatgcatttgttctttatcttctccata  
ccctgagtttcagttattttatacttcttttatttgacgagtagacattacataatctctaattatcattgagactt  
atTTTTcaacttctttgttgcagttttatgcttctaatttgacatgtgtcaaagtgagtcacatacatcttca  
ccaatgttctacaatgggttatatttgactaagtggttgcctcaaatccaagaaattgatcagttttgatatactgat  
gctgtagtgttgaggggtgtcatcgaagacgttgttgatcaacctgtagggtgtttctagttcattctttttgacaa  
tgatgaatgtgtcgtcaacattacgaacccaaattttcgccttgatctttggtaagattgctgggttctagctctg  
aattactgcttctgcgataagtcctgatattagtgatccatcgtctccttttgttgggtcgtgaaatTTTgtt  
ttgaattggaaatatgttttttagtcataaatctatgagttctagtagactttgactttgaagcttcgtatattttg  
caagattttcgtcgttgggtcagaagcaaggataacgtttcttttgcaagcttagttcogattgaagtgaatagcgc  
tgttacgtcgaaggatatacagctcgtcttctgtaacaaacaaagaagaacacctactgttgaggagtcaccatac  
tgggacgaaacgggtcatccgggtgctcgtcctaacattgatcgcactcgtgatttcaatcaaaaacctcaaatcttca  
cgacccatactgatactttaaacaacaaatTTaactcagttgggtatccaattatctattgtactccaatatttcag  
tgtttattttgttaactgttaatgttgacagtagcttgcacaacgggttgacagaatttctaataatccattaatga  
actacccagggatgagatggtaataatacattttaaataaaaacttcttttgcttacataattacataattatctatc  
aaatctcatctataacttgatttcagttatcatatataatcatgaatgggtggagaagatttgtgggtcgggtgggt  
aatttcogatgtaggtttgttctctgaactggatagttgggtcgtggagcttctaccagaagtttgtgcccgatgatg  
tcgttcagaaggaacgatgaaagctccatgatcaaaccatccagctgagagaatgaacctacatcaaaatataatcat  
gaatattgtatctcatctcagtagtatgtttattttgggttaacaaatTTatctcactttctctgtgtatataatgat  
gtatgtatagttattgattatgaatgctaactcagtgataatgagggtaatgaatgtatttctaattactaagataaa  
ggatgatgatctctttgagtggtgtgtgtgtgtgtttgatgtgattaaatgatttggttacgattgggttattttc  
aatcaaaagtggtatagttaatgaatgtttcgttgttaacctatcagttcttatactccagcttcaattcaattgttc  
atacattacctcttgccttgattgtgtaatgatacgtattattgtgtatgagatgatcttgtattttgcccattgtaa  
aaccattcagtcogtattgtgtttgttgaactagtagtctaatgaattcggttctctttcttttttgggtattatctta  
**agATTCCATTGTGTAATCTCAGTGATTCATATTCTACTGATC**CTCCTAAATTCATTGAACAAATATCAAATTCCTCC  
ATTTGCATTATCTACATCGAATACATCTGATATGTCATCTAGTATAGAACAACAACATCAACATCAGCAACAACAA  
CAACGTCCAGAACGTAATCGTCTTTGGCAAAGTGACGACAATCTATGTCAAAATTTTCGTTACATTTTATACGTA  
ATAATAATATACGATCATCGACAAGTTCTATACCAAATACACGTACATATCAATGTCGTCTTCGTGAAACTATTAT  
TACTACTATAACCATTATTAATAACCATAAATGATAATAATAACAATAGTAATAATAATAGTAATAATAAGCCTTTA  
CATAATCCAATCAATGGACTAAAACAGCAAAACCATTTAGgttaagtagtctctattattcttaattggaacatt  
tccattttgttcagataatccttgccttttgagtttatagtttgcctattatttggaaactaacctttattttataga  
tatttaaaaaatgggtgatatacctggttactaagtaattcatgttatatctgaagcgtttcaaatttactagttaat  
gaatggaacaagact

E9

>333748.c006312244.Contig1

gttgctattcctactgttctttcogattcgggtctaaaaatgttatctctttgttctaacgggttatggcaacttgat  
ccgatgtacatataaccaggttctacgttgtgattgactggctgaccgactgcatactgtctacattgatttcatag  
tgtaaacagatatttattgacagcttaaaagagttaaaatgtacttttagctattcacctcctaactattgagaataag  
aaattttgcgaattgtgaataccattttgttatgttctggtagtttgcaaatatatacagtaatacagattgacaatt  
agaacaccgacttattgatgagactaataacgtattaacaagcaagagcagcatgaagtcaaccctgagttcttat  
tgatcctttgatatagttagcttctcacctaaccagcgggttttgcagaaacccaatatacagttctccgctaaat  
gaatcctataattctaggcatacgcagtttatatacaacaaatcagactgcatcataccataaaaatataaattatg  
aactatacaagatcaagccaaaaaagggtgtgggtgtgtaaaagactgaaatttaataaataaggaataactcagg  
aatggaaaaatcgtataataatagtaaatagggtcaaacgggaagctttataataacaggggatataatataaacgtagtt  
acttaatagttatacaataagaacatataaacgtogataaaattgatcctagcagttaccatttacttcttctc  
gtcggacataacgctattttttggaacttttctaattgtatagttttttttcgtccatttccctaaaatctta  
tacgtttgatatacgcttaattgttattctttgcattgttccaagattgtcatcattgattgatctaagctaaataa

E10

E10

ccattttaaattcagttaaaattgaacagtcataatttttggatttccgtagtatttaaatagtgtagacatccataattc  
 cacaaggaaattgaaagtttggctgcgagaacactgtgggtgtagcactactgatgtcgagagatataagtggtagt  
 tatcatcaatcgaaagtgaaatgtctggaggcagaaaagtttagaaaaatcgatgaaaagagagaacgagagaacagagaat  
 gattgggtgtaggaaacgaaggaaaaatgaagtctgagacgattgtttgacatttccgtaaaatgacaatcaaaacta  
 gtaacctacgattgttatctattcttaataagccatgcttccgcatcccaaaagtagaatagagagaatcgctgctaa  
 gtaagctcattctttgggtgaaagacgggtatcccacctatgccacaactaacgtaaattaacaaaagacaatcga  
 gtcttctgaacctgcccgaatttccgtagcagaagagactgtcgggactgatgagattcctaagataagtcgaac  
 gaacgattattgaacaacttactgcttatcattagttcaggcgctgggtacaagccaatcctgaagtaataattta  
 catctggaggagagaacctcgaaccccgaaacatccttgcattgtcgtttaagggtgggtcagatgagctgaatttt  
 gtcagcgtcttcatcaaatagaaatatttactgtatgggtctttagatttttaaattctattgtccacacttgtatt  
 ctggtccaccacaacactacgttttattcgaatagattgtaataataatttattgggggtttttattgggtgact  
 gtttattctctctttttatc**ag**GTATGGATACTCCATTTTATCATCAAAAATAGAAAATTCAATCATCAAAAATCATA  
 TTCAATCTAATCGAGAATCTCAAGATTCAACATCAATTTTAACTAGTTTTCTATA**Ag**tgagtgag**gt**tagtattat  
 tactattactactcttattattatttagtagtattttctgtgatttataaccagtttataaatcctgctgttcatgat  
 gaaattattgaaatcgttccaagaaatacgcgaacggttcatgggtgttggttggttcacgatcaattttctactgaatag  
 agtatcctatcttttagggaacttttaagttcttataaaaaatgtgtcagactaatattctatactattaatagat  
 gaacagtgttttcaacattcagtaatgttgggttgcgattttgaaatcagtatatgatcattaagttaaattttggct  
 accaatctaaacttacgtggagccatactgaggagtcocatataggacaaaacggcgctcagtgcttccagttttcat  
 ggggccaaaccttaatgatcagaa

>333748.c002909677.Contig1

E11

tgtattcogatagattcacttattaaactatatttccagtcctataattcaaccaocttaactataactatgcat  
 cgacctgataaaaagtgaatacatccagttattctgataaaacgaatggatatttccattttgttcatacattttacg  
 ggttccactttttctacaattaaaatagaaaagaaaaaaaatcgaatatcatttagtggtacggttacttgactaat  
 tataaatttgtccacatttcaatgtcattctgattaactctccctctctgtctttttcgattctccattgggttgt  
 ttatgaattgagttgacctgttattttgtttgatcatgtcgtcctcatcatcatatggaatatccaagcaggtatt  
 cataattgttatatacatctagtgatcgacttcataactgggttacaatcttgatctattgagtaactcttatcaac  
 aagttaataaattctctattcaattactttaaactgaattcgggctcatataactgtagtcattttgattgaaatt  
 cgaatactacttacgatcagtaaacgttggatcttattttaaactctactgtaaacagagatgatcttataagcaaaag  
 atggattccactgctagccactatccatcttggcttataatgcttgtgaaattaaggcaatatcaaggtaatacgc  
 tagtatgcacatatgtcaataaaaaattgaccagttgcagtcctaaacatcaataagaa**agcctt**caaacaaacaata  
 ctaagagatcattttgttgcataatcgttaggaaatactgacaatctgggtatttttttgatcaattaaactatat  
 tattagtcctaaatgttatttagatttatcaggaagtcgtaaatgaagaaaaatcatcatttttca**agcctt**ttctgtat  
 atggataagttattgaaacagttcogattgtaccttcttcccttgattccacgctcataatgcataaaacaatacgcctc  
 ctaatgtaattaatgtatttttggaaatacaacgattcttttttctaattgcctattttgt**ag**TGTGATATGGATA  
**TC**AATAAGGAACCGTGGTTTCATGGTGTTCCTTAGAGCAGAAGTGGAAACG**gt**aagtttatcaaaaaatcgtaaa  
 gatggtttattcagtaaatagttgaaatcaatcttga**agcctt**gtgattgatggatagtcactaaccttctgtgtcg  
 aataaataatttatcatataggtctacttttattttgcttatgtgtgactcgccttccgagaactacagtgaggcc  
 tctagagccctgatggagcccaagagtttccatccaatcagaacatgatggagcctctagagattcaatgtggc  
 atgtactattgaacattaggataaatatgtctctatgcggattgggtcaaatgataatgttgaatttaacaaatgcta  
 atactataaataaccatgattttctgtacatttccgattcttaaccaagtgaacacttctccagttttttttctgt  
 ctctctgatttgcggcgggtgaacttcttagtcgtaaacgtattctacgttaatacagtagcgaacttgacattatg  
 atcaactaacgtttaaatgacgcacataatgtaaccaatcaatcgaatagtttaacagatgatgacgttcagggag  
 tgttttccgactaagtggatggatgctggccacgaatattttcacagacgtagaataatcttgaataaggataggtt  
 ggtataatttagagttttaaggatctcattttgtgtgtatttccaaagaacattacctcattaagggtcacaaca  
 taattgtaaaatttggtagataaaacacctcattattaacgtttcattatttttcatagctgaaatcatgaatcaat  
 tgaagctagagttgctatctcacaacagacgtgggttgaactccactcgacactacttccactagaactccaggac  
 atgtatcttgaagtcagtcactagtgagcatatgattataataagaaggggtttttgtggagatttcagtattttt  
 catgaatgaaatgatgcgtcogattga**agcctt**gaccaccaggaaaaacctggaagcattggatgacogtttccgct  
 attacggaactcttggcagtgccatccacgatcccgcacccagaacctactagtctcgcggtagagcatttgac  
 cgatagaccactgagccggcatccaacgggtgtaatgtctaactccaaccaatcaaatctccacaaaaacccttt  
 ggattgtttcattatattatgttaaaataagtcacgtacgatttaattctttgttggtaagggtattttttgtcgttgt  
 ttttcttttgaataaacttgaagtagcagaaaaatgtgatctgattttattgatttatcgattattac**agcctt**ttc  
 tttttgtttttccttaattaaaggatggcttttacttcaactatataatgctactacaacaatataagattttgatt  
 gtttctatgtagatcacttactgggtatagttaccatatactatagatttacttatcaatttattgattattagaa  
 ttataattgtacttgaatagtagacattgacatgtaatgtttgtatgtttgagtaatgaagatcaatactctgtgtg  
 gagttgctctcactcactcacttactccctccctcactccactcactcattccttattactactcttgttcttatt  
 agtttatataatttccagttatcttgttataacatgggtgtcaactgtgtaataattatcacaagatataaacatgact

atgaccacgaacactctgccttcttcttcaactatcttctcgtttatcttcatggtactatggtattctatggagt  
actttctttttgtttcacatagaaactaaatcttctggttagtttttttattaaggacacattgtaaatcctt  
ttgagaaaaaagtattgggaaccttaaaaactaatgaaattatataatcattgtacacagatgccctagtagc  
gccgagagtagggatagtttgctctccctgtcgaaatgctctcacacgggtcacgcgcataatagcctctaccaggg  
agtcctactcactgccttctcgtggcattactggttggttacgaaattgaaaggatgaaaagcaaatgtccgggtgct  
ttaaccagttgggtggacacgggtgagtcacacctcggggagttggaaaactctcattccaaaccaatgggtgcacatg  
ggctgactccagaatcctgaaggaacaaatggcgcataatcaatcaatcgttgggtcaccgactatcatgggactacgct  
tcctcagcattctccactgccttgaggatcagatcttcagggtcaaaggctccgggtggtggtcccctaagaaaacca  
tctgcttcagtttgcgcacctgggcagtatcacagccctcacgcaaatcgaatgaaatctgtggtggcgcataaata  
tctggtgctcccttgatcaatatttatggtgtaaaaataataataataaataaattgataggtaacaatgaacagc  
aataaataaataaataaataaataaataaataaataaataaataaataaataaataaataaataaataaataaataa  
atgggtggtgctacttatatcctcataagtagtatatggtaatagtcagacatagaatgattttggcagaagacc  
gataaagaaagaactgaaatgaagcgcgaattggtaggaaaatgcatgaaacaatgaaatcagagaagatggaaatgat  
atgtgcagaagaaacgggtcaagattgaaacgattgattggttattttgcaaattaactggtgactgtatgggtatca  
gatcttagtaagatagtttgtaatttgcttaaaatacattcgattgtctccaactagtggttcttgatcactacia  
tagtacactaataggttaactattaaactttctaattggattcaagatcactaaacttgattttgtgtagcattg  
taacctttttcttacttgaaatctacaaaagacaataataaattagcttaactatttataaagtgttcaatagacaa  
tcacattatgggtgggtgctactgatgtcgacagatataagtagcatatcatcaatagaaagtgaaatgtctgggt  
gtgcagaaaattagtaagttgaaagcaagagaaacagagaatgattgggtatgggaacgaagaaacaatga  
agtctgagatgattgactgatattttgcaaatgacgtgttttttttactgtatgggtgctcagattttactaagag  
atctctgtaattgtgtgttcaatacatccagctgtccctgcctgtgctccttttactacagcattacatacggta  
cgtttatatttctaaagtgagatctgtgttgagtgatagaagagaaagcgacttcataatttgatagcaaatat  
ctttcaaatggtttagttgggataataaccgtttcccttgtaactttttgaaatgaagattgaaatcacaagt  
ttaatcgtaccgcatttttttttcttacttatac**agTCTTTTACAGAATCAAGGCGATTTCTTGTTCGACAAACA**  
**E12** **AGTAAACGTTCCAGTGGTCGTGATACCATAGGTCATTTGGAATGGAACATAATGGTGATTTAATCAATGGTAAATCATA**  
**ATGAAACAATATTAATAAAGATCATAATATGGATGGAACATGTTTTACGTATGGTTTTATCTGTCTTTTTGGCATTGG**  
**TCATAGACATTTTCATTCTATATGGTGGACCAGAGACAGGTGAAGGTTGGCATTTAGAAGACG****gt**catttctctaca  
ataaggtaagtagtagtcacttttctagctcaatgttttttgaacaccatgtataaacgaacaatattgttttca  
ttgtcgaaatcatgagtcacaatgaagctagattaccatggagaacctgaaggtagtggaggcgtttcgtcccat  
tatgggattcttcagcagtgcgcatccacgatcccgcaactcgcgagattcgaaccaggacctaccagtcgcgcgc  
cggaaacacttaaccgatagaccactgagccggcatctgaaatctccacaaaacccttctgaaatattgttttcgat  
tagatcctcatcaggctttactcctaataacagtaataatttatatgaaatatacaaaattattaaaccaaccaaggc  
agtttatgagtcattataaccagtggaatagattacataaataaacataataaagtgtacagtagcaaatgggtag  
gtgatgacaagtgtagtctgttggctatgaataacagaggcttgaattaatgttacataatgagaaataggttatt  
gattagaaaaaatatagacactgaaataacattcataaaaagccataagattacgtaataagaaggaagtggtcaga  
taattggacgggtgaaacgtatacttgagaaaaaacaaaatgataggaggggtgagagaaacaaacgtagaaacag  
tatggagaataataatgaataaaatcatttattaagacctagggagagattaggggttttacacaattctgatgtac  
acaaaagacttgaattgttgactcgaattccctatagcttctgctatttgtgaagaggcgagatagaaccataaga  
aaaactagtaggaatacagatactacatgaccgctatcgatcatctgttattgaaccgagctgcgtattgccttgac  
agtaccttttttctaactatgaaggggaaatgttcaactaactatttgggttcatttatctggtagtacgctcagtag  
tttatagtgaaacataaaactttgctgatttagcgaaaatatactggcaactaaactgttagtggttcccagtcgtac  
attttggtaaaattatgacacgttttctttcgaaacctttgtttctattgggtattggttatcgacaacctctgaggaa  
tgtgaccgtagtttagcagctgttaccttactgcacaactaagttggtttatttaggtttttgcatccagttgtataaa  
actcatatagaaaaatggagacaaaaacttttgcataaatcggtttccaacc**agata**ctaacctggctgctttttt  
taaaattttaagtagcagatggatattgtgaacaactgttagtagcaatctgagttacggaatgttcttactgggtg  
catcctaaatgaaaaacacgtaaatattaaccaattgttccaagttttgacatcttttcaactagttcogactat  
tactttctgggtcaatacacgtaatttcaaccgatttttctgtttaagaatagtgatcatatctcatagacaaatag  
tacagtaatatgttttctagtagtaacttttatcaagacaatatgggtttcataattaatatcataatcgatata  
accatgctaccggctaccaactagcacaacataaataatttttagtttagttctaggtactacttgtgagtagctt  
agtattttctccgaaatatcatccgtgggttgagtcgataatagtggggtccatttattgttcatacacaattca  
tcgaaaaacatttcaatcaattgtgtggtattattagtgaaactgtgtttcagagatagatttgatcaactaatta  
actgatcttgagcaagtgtagcagcccttatgggttttacttttgacattcctgtgtaattgacccaattgcgt  
cacaagacaaaccctttcatagaatcctgaaggccaaaggaaaagagaagatcaaagaacacattacgcccagaaa  
gggaaacagatatgagaaaaatgaacaaaaactggatagaactagaaaggaagggcgagccaggacagagtggtt  
ggagaatgctgggtcagcggcctatgctccatagggagtaacagggcgtgaagtaagtgtaattgatgagcattcaatc  
aatagaaaaatctttcgaatattttttcataagtttattatcagctctgaaatctctctgggataatcagattgtc  
**E13** gtttgccttaataactaaaattgtat**ag**GGAACTGATTGAAATATCACATGAAAACCTCAAACCTCCCGTGACAGCGAAA  
TCAGGAGCATGTCTTGTGACACCAATTTCAAGACCTGATTGGGAACTAGACAATCGTGATGTACAACACTACTACAAA  
AAATTGGTCAG**gt**atgttttccagacattgtgttttcaaaaaataataatcgattttaataatgccattgttatgct



tgttatttcgtaatgCGTgtctcctatgatggaataaactGCCAAATCGTCCACGGAGGACAACCTCACCGACTCAT  
tcaaggtaaagaccggtgttaggcaagggtgcttactctcaccctttctctttcccatggatcgactggatcat  
gaagacgtaaacatctggaggaaagcagcgggataaaagtggactggcaagatgcagcttgacgatttaggcttcgca  
gatgatctggctcctctatcacaacgcAAAAACAAATGCAGGAGAAGATTACCAGTGTAGCAGTAGCCTCAGCAG  
cagtaggtctcaacataaacaagggaagcaagactctccgatacaatacaacatgcaccaatggaattacact  
tgacggagaagccttggaggatgtggaacctttacatatctgggcagcatcatcgatgaacacagtggtctgat  
gcagatgtgaagacgCGGATCGGCAAAGCGAGAGCAGCATATCTACAATTGAAGAACATCTGGAACCTCAAAACAAT  
tgtcaaccaacccaagggttagaaatctcaatacaaatgtgaagacagttctactgtatggggcggagacgtggag  
aactacgaaagccattattcagaagatacaagtgtttatcaacagttgtctacgaatgatacttcggatctgttgg  
gcagacactatcgacaacaatctatcgtaagagagaacaaaccagactccatccggtggatgaagaaataagggaag  
aagcactggaagtggatagaaccacattgaggaaagcaccacaactgtgtcctgaaggctaaaggggaagagggaag  
GCCAAAGACCACATTACGCCGCGAAATGTAGACGGATATGAAAAGAAATGAAGAAAAATTGGATGGAAC TAGAAAGG  
aaggccgagggcagagtatgtgggttggagaatgggtggtcggcggcctatgctccattgggagtaacagggcgtaa  
taagtgtgtaagttatgtctgtatagtacagaatacttgagaaatgtaacaattgttttggaaagtattgcacatat  
ttagcagaaacttcaagccagaaacattaactctgtcctaagttattatgtaacatagccaagt cagtatcagag  
tctgaatcttcatcgattatagttattggcatatgtctatcctccgtctactgtaacgcgtaatgtctgtgtttta  
aacttcgaagcaagatttgttttttggttggagaactgaattaaacaccaggggttctttgtttttccacattacc  
tcataatggaaaaagagcattcttcaacctaactatggaatctactcatgactttcgacctgtttgagtagctac  
agtcacagatgtaatttttgggtgtacttcaacaaaggaaacgcctgattctaactcgcttctcaagtacatgcat  
gtatggttaccgcagaataattgacagctgtcagtagtttgtagtagtattcatgtagtagcgtgtattatttttta  
tattctgcatctccttctttttacagGGTAATTTGGTGATGTTTATCGAGGTGTATATAATGGATGTGAAGTTGC  
AGTGAAAACCTTGTCGTGTTGATATGACTGCATCGGATTTACGTAGGAAATTTCTACAAGGTGAAACAACCTGCATTA  
AATTTCAATCATCCAAATATTGTCAAATTAGTTGGTATTGCAGTTCAATCTTATCCAATTATGATTGTAATGGAGT  
ATGTTCCAGgttgggtgaacttttcatgtttttcatgaattctagcatatcttcatgttttattcatgttaatttggtt  
atctttactacttactagctataagtagttacatcttcaaaggataagtggttgtagcatggga

E14

>333748.c002729397.Contig1

agGGTAATTTGGTGATGTTTATCGAGGTGTATATAATGGATGTGAAGTTGCAGTGAAAACCTTGTCGTGTTGATAT  
GACTGCATCGGATTTACGTAGGAAATTTCTACAAGGTGAAACAACCTGCATTAATTTCAATCATCCAAATATTGTC  
AAATTAGTTGGTATTGCAGTTCAATCTTATCCAATTATGATTGTAATGGAGTATGTTCCAGgttgggtgaactttt  
atgtttttcatgaattctagcatatcttcatgttttattcatgttaatttgttatttttactacttactagctataa  
gtagttacatttcaaaggataagtggtttagcatgtggattgacagtaatttgtgactgcaggttaacagtggtgt  
actagtagttagaccataaatgtctttagaggattgggtctaataattccgtaaccatcggctgagccacgttgagggt  
actcttcaaaaatacagaatgggtgatgaagtagatcttcagcaacatgtattatgtatatccaacgaacgtctat  
cgtaaacgacggttactggcctgcaagtaagttggagaaagatggagacgataggtagtagcttaacgtcagtc  
atgaattaattaaaagtattctgagacataattgtagatgcaaaactaccgggttacactctcaaaacgctctca  
accacggttataaacattgtgtcatgtttgagaatcgggtcaaaagagcgtaaacatatttgggtatttgccttact  
tcggatcctcaattttgaatcattttttcacaccttatcacctaatacattctgtctacagtcacataatctgt  
ttctaacttttttcaactatcatggttctaattgtggtgttatattgggtagttttgcattttgtgtaccgaactat  
acggttgaatttgtctgggacatcattggaagtggacaaaatattgttttctcagaaaatgaagatagggtgacaggaa  
aattattgttccaataaggataaaaaaaattttttaattaagtgaatgaacttta~~gaattc~~ggttgtgaacgtcag  
actgcaatacatacaaatcccaccaactgggtccaaatgcaacaaagtgcggttcattaaattccacttgtagctac  
agttcagaactgaattaaaagtatgaaactatgggcaataaccgaggtttctgttacatctccacgtatgctatagc  
tgattgatataatcttgtatatcaataacggggaagtgtgaacgcttttcaagtgaactactgtaaaattcaagt  
aaactactattgtgattagtttgttttctgttggttcattctccgggctttcactttctctctggataatattat  
cagtacaaatatcaaatagaagtgagttatccagataacaacgccatgggtcaccattctttctgtccttatctt  
ttgtgtatatcacatcacactacttaataactactcagtagcggagtagtcatcagtcgtcacttagttcatataggat  
atacaactatacattcatttggattgaaacgatcgttttaactctcttcttcttggataagttataaaatcgtattatt  
ctaatttattccggttggttttgtattcattataaaagGTGGTTCTCTGTTAAATCATTTACGTAAATCAAAAAATG  
CATTACCTGTTATGAACTACTTCAAATGAGTTTAGATGCTGCTAATGGAATGATGATTTAGAAAGCGAGAAATTG  
TATTCATCGgtatgtttttatgtctcatttaatatcactttactttcatatgattttattttatttatacac  
ataaatattgggtacaaggaggttagtaccagatacatatacgcgcacaaatctcattcgatttgtgtgagggctgt  
gatactgctcaggtgcccAAACTGAAGCAGATGGTtttcttaaggggcccacacacgggagcatttgacctaaaggtc  
tgatccacaaggcagtggaagcagcgtgaggagatgcagtcctccttagcgggtgatcaacgattgggttcatacgc  
cattcgttccctcaggatactggagtcagccatgtgcaccattgggttggaaatgaggggttttccaactcctctag  
gtggactttccgtgtccaccaaccgggttaaagcaaccggacattccgttttccgtccctcactttgggtgaacaaca  
catccgccacgagaaggcagtgagtaggacttccctgaccgagactgtatacgcgtggccatgtgagagcatttgg

E15

agaggaagagtggactctctacactctcggcgcgtaccagggcattttgggggcactttcatataaaattcatatttg  
ataaattcaatcctgataatggttgtaaacagattgatggtttcttcttatatcaatgtagtgtcttttaatgaca  
gttggcaggaataactaatccgtaattcccatagaaggaacgttaccactaatctagtcctggcagcaataacttcc  
tccttctacctgtagtggattttccagtatagcttcttcaataactttcggagggatagagtcfaatatgtaaat  
tgaatgaactcatctttcataattttgcttaataagtaacaaagtaatttttagactttatgtaatagtgtgtagct  
tcggtcagatttcccttagtaacatttgaatggatggaatattttgttattactaaacgaaatctgtagtgttg  
gttgttatgtaagcccacatagcttattttgagcttccgaacagaccgatttggttcaatcttctggagccatactt  
taactttgaccggtgtgtgcacttcatatcctattcatcacatgtttggacttatttttatatgactaaaaat  
tgattcacgcttggcgaaatcgagttggatcacattgttctcaccgcacgctatttgcctcgtctcttccatt  
tccccaatagttgtctggatcaatgattgtatcaaatatacgtattcaaacattcatttctcgtattttgacttctt  
taatcccgttactcactaacgcgagtttaagtcaataattatataagaaaattggtgacaagtatatttcaataaca  
atcattcacatgatctaaatagagcatagggcgcgtcaccagcatttccatccaactctgttctgagccttccctt  
ccagttgttattcatcctcttcatatctgcttcaatcttagtcactatggaatatagagaagagtgactacttgta  
actcgtattgaaatattgaaactttattttatagtctatgctttaatcatcaagttaaaccatttttaaccaggatca  
gtaataataaattcgcatttggtagaataaaaacctttacatatctgagcagcatcatcgatgaacacgggtggatct  
gatgcagatgtgaagacgcggatcggcaaaccaagagcagcatatctacaaatgaagaatataggggattcaaac  
aagtgccaaccaacaccaagttcaaaaattttcaatacaaatgtcaagacaatttctactgtatggagcgaacttgg  
agaactatgaaagctatcatgcagaagatacaggtgtttatttacagttgtctatgcaagatacttcgga~~tc~~cggt  
ggctagacactatcagcaacaagttactgtgggagagaacaaaccagattccactggaggaagaaatcaggaagaa  
gctgtggaagtggatagaacacacattcagaaaatcacccaattgcatcacaaggtgagccatcacttggaaatcct  
aaagaccaaaccgggaaggggaaggctaaagaacacattgggcccagagaaatggagacagacatgagaagaatgaaca  
aaaattggatagaactaggaaggaagatccaggacagagtgagtggtgagaatgctggctcggctacctatgctcctt  
tgggagcaacagggcgtaaagtaagtgagtaagtagaataaaaacgataacaaatagtaaaatttcgcgtatttctatct  
tctcattgtatattattcattatctctgtattaaaat**ag**TGATTTAGCAGCAAGAAATTGTTTAAATTTCTGATGAT  
GGTCAATTAATAAATAGCTGATTTTGGCAGTCAAGAGAAGAACATATATATGA**ag**tatgttgaagaagtagttttat  
atatgtatagaatatcgggtgttttaggaacctcattttttatcaaatacatatgtatgtatgcatttattatttatt  
taaacacactcaaagagggcaacaaatagatctccatatagaatcaaatgatttatgtgaggggtgggatactgc  
cccagacgcccactgaagcaggtgggttcttaggaagcattccccgaaccttggacctacaagttccaatcca  
taaggcagtagatctgttatatcgtgtacgtgctgcccgtgttaatatataacgtgacgatgtcgttaacggt  
agcagtgaaattagcgtatcggaccaatacacacaaaaatccgtgcgagcagctctagagtaactaatctgtccggctatc  
tgccatagcccagccagtttaagttccagaacactaatctcagcctctgcggtataaatcattgttctcacacacactg  
ggtttatatacaaaaacaaactgaccacgtcgtaccatgaagttagagaacaaacatttgtaca~~agc~~tttagccaaatg  
tggctgtgaaataggggggacagtaattaatatactgggcataactcaggaatggtaaatcctacagtaaatagtcta  
tgggtcaaaaataaagctcataataaggggaacatgaatatgaatatttttagttacttaacaatcatacaataggaa  
atatatgtatcatattgggtctataaatagttctcaaagttaccaatcataatttctcttcaagatataacacctctt  
ttttgaaagatccgaagtt**gatatc**cggtactttaaatttcatgaatttatacccccaacgaaatagtgttggat  
tctcatataaccgacatttttctcttattcttataagaaagttcaccaggaacatgtgaaatcacaagtgatc**ga**  
**tatct**ggcaaaagcgtattgggggata~~cc~~cgataaattaggttaattagaaacttttgtcgcagcatgatttagtttcatg  
tgggtgcaattatcttatttatactgctcgtatttattttctgtttcttttggggcgggctgcagcaatttcttaa  
gcatttagcagcagatcgtctttctatgaactggataaattcactgtttctacgttctttttggggataata~~gaatt~~  
~~cg~~aaccaacaatagggcatatctctgaagtggacagtaactggactactgggctgatcgtgtgtggctgctactgct  
tcgcataaattactgtttataacggattcatggaggtttcatcggcaccatttattcattctgcaatcattttcca  
gagctgtgcgcaatttagaaaattaattcgcctaacttcccatacccccaagcatgggtgattcggcagctttatgt  
ataatttgtcgaataatcccactgactctttttgaagcaatacgttccatacatcttttgttccctcaggattctg  
gagacctgttccacagttgggttgaaccaggattttccgactcaccagatgga~~tc~~cctcctcctcctcaacatg  
gttaaagcgtcgaacattcgttgcgccttttcaactcgttaacaacagtaatgcgggtgagaaggcaatgaata  
aaattttcgtgacagtggtgcataagcgtgggttatgtgagagcattttttgaggtcgtactaattctcttccact  
tttggccgtgccattgcatctggggcaagacttgagttctgagagctatgctatattaaagtgaccgaatccact  
atcttccattatttttactgaaagattttatatacataaggaactttcgtcatataaattcaactactatgaaggagt  
ttcacgaaaatatactattaaaggtcaacacatttcttaaacattacggtttttagccgtactgtaaatattctaata  
acagtgctttaccgagtttttctttatattccttcatatggcagttgtgattcatcttctatctatctatctata  
tctttttct**ag**TAAAGTGATAAACGAGGTCAAATCCCTATCAAATGGACAGCTCCTGAAGCACTTCGTA

E16

E17

E17

atccccagctctaagacgggggtccttgctaattttataatgacacccccaaatcaccocaaacgattacctccttta  
gaaaagctcgaaaaaacctcataggcggccagtgaggactcattgggatctttaacaactctatttaagacat

>333748.c004513974.Contig1 E15 a E17 correto

E16

agTGATTAGCAGCAAGAAATTGTTTAAATTTCTGATGATGGTCAATTAATAAATAGCTGATTTTGGCATGTCAAGAGA  
AGAACATATATATGAAGtatggttgaagaagtagttttatataatgtagaatgtcgggtggttaggaacgtcatttt  
ttatcaaatacataagtagtatgtagcatttattttattttaaacacactcaaagaggcaacaaatagatatgctcc  
atatgaatcaaagatttatgtgaggggttgggatactgccccgacagcccaactgaagcaggtgggttcccttagg  
aagccattccccgaacctttgacctacaagccaatccataaggcagtagatatgttatatcgtgtacgtgagcgtgc  
cctgtttaatatacaacgtgacgatgtcgctaatoaggatagcagtgatcatcgatcggaccattgatacacaata  
tccgtgagcagcagtagtagtactaatctgtccggctatctgcctagcccagccagttaagtccagaacactaatt  
tcagcctctgcggtatgaatcattgttctcacacacactgggtttatatacaaaaacaaactgaccacgtcgtacca  
tgaagtagagaacaacatttgtacaagccttagccaaatgtggctgtgaataggggggacagtaattaataactg  
ggcataactcaggaatggtaaatcctacagtaaatgtctatgggtcaaaaataaagctcataataagggggaacatga  
atatgaatattttgtcgataaaatcccactctgactcttttgaagcaatacgttcatacatcttttgttccctcagg  
attctggagaccttgttcacagttgggttgaaccaggattttccaactcaccagatggatcctccgtagcctc  
aacatgggttaaagcgtcaaacatgagccttgacgcctttcaactcgtaaacaacagtaaatgaggtgagaagacaa  
tgaataaagtttctgtgacagtggtgcataagcgtgggttatgtgagaacatttttgggtcgtagtaattctct  
tcacacttttggcgtagcatctggtgggcaagacttgagttctgagagctatgctatattaaagtgaocga  
atccactatcttccattattcttactgaaagattttatatacacaaggaactttcgttatataattcaattactatg  
aaggagtttcacgaaaaatctatcaaagggtcaacacatttcttaaacattacggtttttagccgtagctgtaatat  
tctaataacagtgctctttaccgagttttctttatattccttcatatgacagttgtgattcatcttatctatctat  
ctatctattttttctagTTAAGTGATAAACGAGGTCAAATCCCAATCAAATGGACAGCTCCTGAAGCACTTCGTA  
CTGGTCGTTATACTATCAAATGTGATGTATGGTCTTATGGTGTACTTTTATGGGAAATATTTACTTTTCGGTGATGT  
TCCTTATCGAAATTGGTCTAATCAACAAACTAGGGATATGATTGAATCTGgtaagttatttaatttctcaaactctg  
ttatgtaagtttctttatgctaaatgaatgaatggaattaacggaaagtttggattcacttagtattgtttattt  
gaatcttcccattgatgtctaggactgcaactggctcgttcttattgacatatgtgcttactgtgagtaatgcct  
cgatataggctaaactcacaagcatggtaagtaagatggatagtggttagcaatggaatgcaggacgcgcatttt  
gtcctatttgggactcgtcagctggatgtacctgccacatctcagacttgttgatgttcaactctgggactcgaacc  
cagtagctttcacttcaaacggcatcgcgttatccactcggccactgagtcgtgatagccactttcttgtgcaatg  
gggtgaagtttaaatcttagtattgtttgtttgaaatcttccccattaatgtttag

E17

>d000017208.Contig1 d000017208.Contig1:1,4942

cttcatttccccaatagtttgtctggatcaatgattgtatcaaatatacgtattcaaacattcattctcgtattttgac  
ttctttaatcccgttactcactaacgcgagtttaagtcaataattatataagaaaaattgggtgacaagtatatttcaata  
acaatcattcacatgatctaaatagagcataggccgctcaccagcattctccatccaactctgttctgagccttccct  
tccagttgttattcatcctcttcatatctgcttcaatcttagtcactatggaaatagagaagagtgactacttgtaa  
ctcgatttgaatattgaaactttattttatagttctatgcttcaatcatcaagttaaaccatttttaaccaggatcagta  
ataataaattcagactttgttagaataaaaacctttacatatctgagcagcatcatcgatgaacacgggtggatctgatgc  
agatgtgaagacgcggatcggcaaaccaagagcagcatatctacaaatgaagaatataggggattcaaaaacagtgcc  
aaccaacaccaagttcaaaaattttcaatacaaatgtcaagacaattctactgtatggagcgaacttggagaactatg  
aaagctatcatgcagaagatacaggtgtttatttacagttgtctatgcaagatacttcggatccggttggctagacact  
atcagcaacaagttactgtgggagagaacaaaccagattccactggaggaagaatcaggaagaagcgttggagagtg  
atagaacacacattcagaaaaacoccaaattgcatcacaaggtgagccatcacttggaaatcctaaagacaaacgggaa  
ggggaaggctaaagaacacattgggcccagaaaatggagacagacatgagaagaatgaacaaaaattggatagaactag  
gaaggaagatccaggacagagtgagtggtgagaatgctggctcggctacctatgctcctttgggagcaacaggcgttaagt  
aagtgagtaagtagaataaaaacgataacaaatagtaaatctcgcgtattctatctctcattgtatattattcatta  
tatctgtattaaaatagTGATTAGCAGCAAGAAATTGTTTAAATTTCTGATGATGGTCAATTAATAAATAGCTGATTTT  
GGCATGTCAAGAGAAGAACATATATGAAGtatggttgaagaagtagttttatataatgtagaatatcgggtggttag  
gaacctcattttttatcaaatacatatgtagtatgtagcatttattttattttaaacacactcaaagaggcaacaaata  
gatatgctccatgaatgaatcaaagatttatgtgaggggttgggatactgccccgacagcccaactgaagcaggtgggt  
tccctaggaagccattccccgaacctttgacctacaagccaatccataaggcagtagatctgttatatcgtgtacgt  
gagtgccctgtttaatatacaacgtgacgatgtcgctaatoaggatagcagtgaaatagcgatcggaccaataaacacac  
aaaatccgtgagcagcagtagtagtactaatctgtccggctatctgcctagcccagccagttaagtccagaacactaa  
tctcagcctctgcggtataaatcattgttctcacacacactgggtttatatacaaaaacaaactgaccacgtcgtacca  
tgaagtagagaacaacatttgtacaagccttagccaaatgtggctgtgaataggggggacagtaattaataactg

E16

cataactcaggaatggtaaatoctacagtaatagtctatgggtcaaaaataaagctcataataaggggaacatgaatat  
gaatatttttagttacttaacaatcatacaataggaaatatatgtatcatattgggtctataaatagttctcaaagttac  
caatcataattctcttcaagatataaacacctcttttttgaagatccgaagttgatatccggactttaaaatttcatg  
aatttatacccccaacgaaatagtgttggatttctcatatacccgcacatttttcttattcttataagaaagttcac  
caggaacatgtgaatcatcaaagtgatgatgatctggcaaaagcgattgggggatcccgataaattaggttaattagaaa  
cttttgtcgcagcatgatttagtttcatgtgggtgcaattatcttatttataactgctcagtttattttctgtttctttgt  
ggcgggcgctgcagcaatttcttaagcatttagcagcagatcgtctttctatgaactggataaattcactgtttctac  
gttcttttggggataatagaattcgaaccaacaatagggcatatctctgaagtgacagactggactactgggctg  
atcgtgtgtggctgctactgcttcgcataaataactgtttataacggattcatggaggcttccatcggcacatttatt  
cattctgcaatcattttccagagctgtgcgcaatttagaaaattaattcgttaacttccatacccccaagcatgggt  
gattcggcagctttatgtataaattgtcgataaatcccactctgactcttttgaagcaatacgttcatacatctttt  
ttccctcaggattctggagacctgttccacagttggttgaaaccaggattttccgactcaccagatggatcctccg  
tatccatcaacatggttaaagcgtcgaacattcgttgcgcttttcaacttcgtaaaacaacagtaatgcggtgaga  
aggcaatgaataaaaattttcgtgacagtggtgcataaagcgtggttatgtgagagcattttttgaggtcgtactaatt  
ctcttcacacttttggcgtgccattgcatctggggcaaagacttgagttctgagagctatgctatattaaagtgacc  
gaatccactatcttccattatttttactgaaagattttatatcataaggaactttcgtcatataattcaactactatg  
aaggagtttcacgaaaatatctattaaaggtcaacacatttcttaaacattacggtttttagccgactgtaataatc  
taataacagtgctttaccgagttttctttatattccttcatatggcagttgtgattcatcttatctatctatctat  
ctatctatctatctttttctagTTAAGTGATAAACGAGGTCAAATCCCTATCAAATGGACAGCTCCTGAAGCACTTCG  
TACTGGTCTTATACTATTAAATGTGATGTATGGTCTTATGGTGTACTTTTATGGGAAATATTTACTTTCCGGTGATGT  
TCCTTATCGAAATTGGTCTAATCAACAACTAGGGATATGATTGAATCAGgttaagttctttaatttctcaaactctgtt  
atgtaagtttctttatgctaaatgaatgaatagaatgaacataaagttttggattcaattagttattgtttatttgaat  
cttcccattgatgtctaggactgcaactggctcgggtgttttttgacatatgtgcttactgtgctgattgcctcgatata  
gcctcaattcacaagcatggtaagtaagatnnnnnnnnngttatccacttggccactgagtcgtgatagccacttt  
cttgtgcaatgggggaagtttaaatttacttagtattgtttgtttgaaacttcccattgatgtttttaggactgcaa  
catggtactgggttcgagtcagagtgcaacatcaactcaagatgcaggtacatccagctgacgagtcgcaaaattg  
gacaaaacgcgctcaaacctggattccactgctagccactatccatctatgcctagaaagttttggagtttagttcta  
tgtaacttacttttatcagtaatatatcgatacaatcatttcttttttcaatatttctctgtcagGTTATAGA  
TTACCTGCACCAGATTTAATGCCAGTTTGGTTGCGTACATTAATGAATCATTGTTGGCATGATGAACCAATGAATCGG  
CCAAGTTTTTCAAAAATTTCTAATGAAATACAAATACCTAATGTCAATTCATTTACACCAAAATTTAAATAGATCTATA  
GACAATGCTCAATTGAAGAAATCTGCTATTGACAACCTATCAACAACCTCCATTACATTTATCCGTTTCTGTTAAAGAT  
AGACGAACTAGAGAAGATTAAtatggatcttttttggttactatgcccatacacatacacatatacacatacaca  
cacatacacaatagatcatagagatctatatacacatacatgtaattgtgtatattgtattgcaatctttttctgtttt  
tttttgaagcgcctttatttccctgttattcactatggcaactaacttacagttccgagtgacaaacctttctatt  
tattattatcattaatataattgattcaattcatgatatatgggtatacatagttgtactgtaaccatgcatatcatat  
aatccactacctatgatcattgtaattgtcatttatatcatttcaatttaataatttttatgattcaatttatactac  
ttatatacttatttaatocttctttcattgtttcttatctttacttatagactgtgataaaaagtagatttaatttgt  
atattctcttcagattacattacatttaccataaaacaaaataatacacataaatgaaccattctcaaagagaaagggga  
gtatgtgtgtgtgtgtgtgtaagttcaatatgacatacattcataatgtaatttatgacatcgattgagtaagtg  
ctgcttaccctgttaatacacacataatataatgacgtgatgatccccctggcaatcagtgaggagttatcattcatt  
aatctccatggtatataacatagttcat

E17

E18

>333748.c006012707.Contig1

gttcaaacacatttttttaaccattacggtttttagccggtacgataaatattttataacaaggggtctttaccgag  
gtttttctttaatattccttcatatgacagttgtgattcatcttatttatctatctatctatctttttcaggtTAA  
AGTGATAAATGAGTTCAAATCCCTATCAAATGGACAGCTCTTGAAGCACTTCGTACTGGTCTTATACTATTAAAA  
GTTGATGTTTGGTCTTATGGTGTACTTTTATGGGAAATATTCACTTTCGGTGATGTTCCCTATCGAAATTGGTCT  
AATCAACAACTAGGGATATGATTGAATCAGgttaaattatttaatttctcaaataatggttatgtaagtttctttatg  
ctaaatgaatgaatggaattaacggaaagttttggattcacttagtattgtttgtttgaaacttccccttgatggt  
tttaggactgcaacatggattaggtttgagctcagagtgcaacatcaactctgagatgcaggtatatccagctga  
cgagtcccagtgggacgaaacgcgcatcaaacctggattccactgctaaacctatccatctatgcctagaaagtt  
ttggagtttagttctatgctaacttacttttatcagtaatatatcgatacaatcatttctgtttttcaatattt  
cttttatcagGTTATAGATTACCTGCACCAGATTTAATGCCAGTTTGGTTACGTACATTAATGAATCATTGTTGGC  
ATGATGAACCAATGAATCGGCCAAGTTTTTCAAAAATTTCTAATGAAATACAAATACCTAATGTCAATTCATTTAC  
ACCAAAATTTAAATAGATCTATAGACAATGCTCAATTGAAGAAATCTGCTATTGACAATCTATCAACAACCTCCATTA  
CATTTATCCGTTTCTGTAAAGATAGACGAACTAGAGAAGATTAAtatggatctattttttgttactatgcccata  
acatacacatacacatacaatagtttatatagatctatatacacatacatgtaattgtgtatattgtattgcaatct  
tttttctgtttttttgtttttgcaagcgcctttatttactgttattcactatggcaactaacttacagttccgag  
tgacaaacctttctatttattattatcattaatataattgattcaattcatgatatatgggtatacatagttgtact

E17

E18

gtaaccatgtatatcatataatccactacctatgatcattgtaattgtcattatatcatttcattttaataatctt  
tgtgatttcaatttatactacttatatacttatttaatccttctttcattgtttcttatctttatattatagactgt  
gataaaaagtagatttaatttgtatattctcttcagattacattacatttaccataaaacaaaataatacacatcaa  
tgaatgattctcaaagagaaagggagtagtgtgtgtgtgtgtgtgtaagtgttcaatatgacatacattcataatgt  
aatttatgacatcgattgagtaagtgcctgcttaccctgttaataatacttatatatatatatgacgtgatgatccc  
cctggcaatcagtgaggagttacaattcattaatctccatggaatataacatagttcatccaatatttcattattg  
aatatagacacttaattacttatgtatacttgaactaacttacctacttataacctgttaccoccttgtaacct  
atacgacatatatacttctgtagttattctcataggatgattctttacataaccttaataggaataattgtctcattg  
tttcttttcttttcttttttatatattccattgtatacttatcaatgtaggggggtttgattttgaagggatta  
ggtaattattacaaattgtaacttataacttactttttatagaagtaataaaacaaaattgatataaaagtattagt  
agtattggattattatgtaatgcaactcgaacggatgtatgtacgtacgtacgtacgaagttctacgttgtgatt  
gactgtatgtattgatgtttgatgttattgtttatactctgttgaaattatatttagttgattgtttgtaata  
tgactgatcatctcccacc

Anexo 2: Contigs genômicos formados pelo Instituto Sanger. Em letras maiúsculas estão representados os exons. Em negrito estão representados os doadores e aceptores de cada intron. Em itálico e sublinhado estão representados os sítios para enzima *EcoR I*, apenas sublinhado os sítio para enzima *Hind III*, apenas em itálico os sítios para enzima *BamH I* e em itálico e negrito os sítios para a enzima *EcoR V*. Em negrito e sublinhado estão representados os iniciadores utilizados em *Southern-blot*.

IX.3 ANEXO III

01 MGNHTVNEVSNYFERGSDVEDDINTLASNNHTITENQSVLVSSTSDTFLVPLPSPHS LGHDKSLSSSSISTPNVDNINTKRYTDSVMLRSDHKESQVI  
02 -----ASGLHRPQPQEHSTSTSAAGTWRLTQASESRHLPHCSAAPSHQDHSAMGFGPELWC-----PKGHTEL  
03 -----MGFGSDLK-----NSQEA  
04 -----MGFSSSELCS-----PQGHGAV  
05 -----NFGFCLQ-----SGRDCI  
06 -----MGFGSDLK-----NSHEAV  
07 -----MGFGSDLK-----NSHEAV  
08 -----MGFGSDLK-----NSQEA  
09 -----MGFSSSELCS-----PQGHGAV  
10 -----MGFSSSALQS-----RAAHEAL  
11 -----  
12 -----MGFSSSELCS-----PQGHGAV  
13 -----MAEQDECNFGIDLIT-----RAGSEAL

01 VSAIDQMKSTINEMLSIFSKLVKAECTVQQAISDLLDDQKESSTVGSVDGGQSRNTAMRNLFKSNYSRGRRRRLTMNINLNKLSNSPTDNEINLKLQNL  
02 LRLQDSSELRLLELTKKMSORAKSDREYAGMHHMFSQLEK--OEGLGHLRATDHSQIGESWVW-----L  
03 LKLODELRLLELTKKFMALRIKSDREYAYTLQNLGN-----QVDKESTVQVNYVSNVSK-----  
04 QMQQEAELRLLELTKKFMALRIKSDREYAGLHHMS---L--QDSGGQSWSSGPPSPVSSQSWAE-----I  
05 NKRHDDDELKALEFSFRVMHKRKAADAEYAVTLGKIN---SQTAREMAGVGANSILVQAVNG-----A  
06 LKLODELRLLELTKKFMALRIKSDREYASTLQNLGN-----QVDKESTVQVNYVSNVSKSWLL-----M  
07 LKLODELRLLELTKKFMALRIKSDREYASTLQNLGN-----QVDKESTVQVNYVSNVSKSWLL-----M  
08 LKLODELRLLELTKKFMALRIKSDREYAYTLQNLGN-----QVDKESTVQVNYVSNVSKSWLL-----M  
09 QMQQEAELRLLELTKKFMALRIKSDREYAGLHHMS---L--QDSGGQSWRAISPDSPISQSWAE-----I  
10 TVRQDAELRLLELTKKFMALRIKSDREYATSLTAVAQ---QGLKIDRADEMOCGLISKSWRS-----Y  
11 -----  
12 QMQQEAELRLLELTKKFMALRIKSDREYAGLHHMS---L--QDGGG--RGTGPPSPISQSWAE-----I  
13 LRLNDETRILELTKKFMALRIKSDREYAYTLQNLGN-----QVDKESTVQVNYVSNVSKSWLL-----M

01 VISTNHQVRFSTRALFHKLITAGQDSDLNCEHCLRRTYLECEELKSOAR-QYMLSLKAYEKKMKFYIGQLSNIQSILTNLDCRRLIN-----  
02 ASQTEHLSQTLRRHAEDLAAQPIAKSILLRDKOOLKRVSEQWQOLSQYAWTTQOEVEKLNKQYRSVVRDSTQAKRKYOEASK-----  
03 -----VTKTELEKLNKSYRQTIKEMNSAKKRYKEALAK-----  
04 TSQTEHLSRRLRQHAEDLNSGPIKSLILIRERQHLRKYNEQWQOLOELTKTHSQDTEKLNKQYRSVVRDSTQARRKYOEASK-----  
05 LEAMDSIQKRLNCSQOQVET-IJHNIDLILKDKRGAKRAMDIGRSQDDEYR-KSCNEVQNLRRRYVOTLQATTAKKRLEQTVSKROP-----  
06 IQQTEHLSRRLRQHAEDLNSGPIKSLILIRERQHLRKYNEQWQOLOELTKTHSQDTEKLNKQYRSVVRDSTQARRKYOEASK-----  
07 IQQTEHLSRRLRQHAEDLNSGPIKSLILIRERQHLRKYNEQWQOLOELTKTHSQDTEKLNKQYRSVVRDSTQARRKYOEASK-----  
08 IQQTEHLSRRLRQHAEDLNSGPIKSLILIRERQHLRKYNEQWQOLOELTKTHSQDTEKLNKQYRSVVRDSTQARRKYOEASK-----  
09 TSQTEHLSRRLRQHAEDLNSGPIKSLILIRERQHLRKYNEQWQOLOELTKTHSQDTEKLNKQYRSVVRDSTQARRKYOEASK-----  
10 MDELHQAKQKFNABQLEV-VCDKILTHSQDTRKAKRAMQEEHAKIAARLN-HLTDEVVRKRSQKHLEGYKALRTRFENYIKAP-----  
11 -----  
12 TSQTEHLSRRLRQHAEDLNSGPIKSLILIRERQHLRKYNEQWQOLOELTKTHSQDTEKLNKQYRSVVRDSTQARRKYOEASK-----  
13 QAETDRVSSLLIAHAQALNAITVQSVDRILLNKTAKRRETTRODADAAYG-RITTEVNVKLNKSYRQTIKEMNSAKKRYKEALAK-----

01 -----NRKIDSLMNNLEYSTRKILHLHNEVCLALITAIHYHQWLYTHLRFQILKGVERTMQLASEVVWH-----  
02 DREKAKERYVRSLSKLYALHNOYVLAVQAAALHHHHYQRAIPTHESLYSLOQEMVLVLRK-----  
03 GKETEAKERYDKATMKLHMLHNQYVLALKGAQLHQYQYDITLPLILDSLOKMQBEMIKALRG-----  
04 DKORDKAKDKYVRSLSKLYALHNOYVLAVQAAALHHHHYQRAIPTHESLYSLOQEMVLVLRK-----  
05 -----VKPAEVDLNRKGFVSSAAKHLKCHNDYTLALAINANTHLEHYHKSTVPPFCLNLTQERMBLTVSQRK-----  
06 GKETEAKERYDKATMKLHMLHNQYVLALKGAQLHQYQYDITLPLILDSLOKMQBEMIKALRG-----  
07 GKETEAKERYDKATMKLHMLHNQYVLALKGAQLHQYQYDITLPLILDSLOKMQBEMIKALRG-----  
08 GKETEAKERYDKATMKLHMLHNQYVLALKGAQLHQYQYDITLPLILDSLOKMQBEMIKALRG-----  
09 DKORDKAKDKYVRSLSKLYALHNOYVLAVQAAALHHHHYQRAIPTHESLYSLOQEMVLVLRK-----  
10 -----SR--SGRLLDVRDKYQKACRKLHLTHNEVLSITFAIEVEKDFRNVLPEGLLEHQQSVQSFILLWRN-----  
11 -----  
12 -----DKORDKAKDKYVRSLSKLYALHNOYVLAVQAAALHHHHYQRAIPTHESLYSLOQEMVLVLRK-----  
13 -----SPNQDKMKRYIRAAADKLCCHNDYTLALQEVQEHQEAAYRTRILPELILNMMQVQCGHAQWQMD-----

01 LLVFGGENIHHFQSNWLKEFQSELDLTSLEQNGVKFEGPVVHTFNOALNNNTTLGGNLSPGSLIVNDLIGOSLIELTOLHKNKELLHRETVNOLPKECL  
02 ILGEVCSITSLVQEDVLAHQKVAHAVEMTD-----PATEYSSVQCHRYDSEVPVAVTFDESILLEBAE-----NLEPGELOLNETVESVQHTLT  
03 IFDDYSQITSLVTEELVNVHREIQMSVEQID-----PSTEYNNELDVHRTAAKEQELFDTSLLEENE-----NLOANEIMWNNLTADSLQVMK  
04 ILQVYLEISLQVDDVASTHRELA AAAARIQ-----PEFEYLGLRQYGSTPDVPPCVTFDESILLEDBE-----QLEPGELOLNETVESVQHTLT  
05 GMEYARTTDMT-QVYQESFGKLFNTLNLN-----PQDEYTKLQENKSDSD-EDVFTFDLSLQDYSSP-----KVEATKMSVNNLTADSLQVMK  
06 IFDEYSQITSLVTEELVNVHREIQMSVEQID-----PSTEYNNELDVHRTAAKEQELFDTSLLEENE-----NLOANEIMWNNLTADSLQVMK  
07 IFDEYSQITSLVTEELVNVHREIQMSVEQID-----PSTEYNNELDVHRTAAKEQELFDTSLLEENE-----NLOANEIMWNNLTADSLQVMK  
08 IFDDYSQITSLVTEELVNVHREIQMSVEQID-----PSTEYNNELDVHRTAAKEQELFDTSLLEENE-----NLOANEIMWNNLTADSLQVMK  
09 ILQVYLEISLQVDDVASTHRELA AAAARIQ-----PEAEYQGLRQYGSAPDVPPCVTFDESILLEEBE-----PLEPGELOLNETVESVQHTLT  
10 ILQEAAYQGDLTADKYKEIQKRIDTVIGSHN-----PTEBYGEPTEKYKTSPTPLLFQFDETLIQDIPG-----KLOSSTLTVNDLTVDWRNRLO  
11 -----  
12 ILQVYLEISLQVDDVASTHRELA AAAARIQ-----PEAEYQGLRQYGSAPDVPPCVTFDESILLEEBE-----PLEPGELOLNETVESVQHTLT  
13 ALNDSLDNIDSCRDEYREIFSRHLATLGDLD-----ESSEYNGELERNGEKLPPEAKFTFDVKLEHAP-----HASPDHLINNLTOPESSIT

01 QWEEAHSQWRAVFANPGPNPWSNLFNIHSSGVSPIVNTPQDNNSNGFHHVDNSSHLALPDTQDLWHDDRARFPVNLCEVSFRLAICEFELCFANCLADSE  
02 SIEEETLASRKAVSSK----EQRVWELQVELR-GEELALSPGERVHLLGKROGLREAAQQOIQGLVCAQAKLCAQRMMLANKLAELGSEPEPP-----  
03 TLAEELTQTQOMLLHK----EAAVLELEKRIEESFETCEKKSIVILLGQKQALBELKOSVQQLRCEAKFAQAQKALLEQKVOENDGKPEPP-----  
04 SVTDEITAVATKEVLSR----CEMVSQLORELQ-SEEQNTHPREVRVLLSKRQMLCEALQGLQIALCSQDKLCAQOELLQSKMEQLGTGEPPE-----  
05 KMEQRTEEYAKQHAER----BEKLRKAINXAAF-DAQNSEVPKAVLISNLQAEIKLLDCTLEREKVKAQKVKEDLQIGESPPLEDFLGEVGVX-----  
06 TLAEELTQTQOMLVNK----EAAVLELEKRIEESKTCCKKSDIVILLGQKQALBELKOSVQQLRCEAKFAQAQKALLEQKVOENEGKPEPP-----  
07 TLAEELTQTQOMLLNK----EAAVLELEKRIEESSETCEKKSIVILLGQKQALBELKOSVQQLRCEAKFAQAQKALLEQKVOENDGKPEPP-----  
08 TLAEELTQTQOMLLHK----EAAVLELEKRIEESFETCEKKSIVILLGQKQALBELKOSVQQLRCEAKFAQAQKALLEQKVOENDGKPEPP-----  
09 SVTDEITAVATEMVFRR----CEMVTQLOQELR-NEEQNTHPREVRVLLGKROVLOEALQGLQVALCSQAKLCAQOELLQTKLEHLGKPEPP-----  
10 ELEGAVRDCQEKQMKM----TGHVNGGSPVAN-GSIIISNGSNTSNGIQSNKDSLRCQSKDLNALRCQEKQKQKLVDMIKCALNEVGCCELPSGCDDLLT  
11  
12 SVTDEITAVATQTVLSR----CEAVACTLORELQ-NEEQNTHPREVRVLLAKKQVLOEALQALQVALCSQAKLCAQRELLQAKLEQLGKPEPP-----  
13 TMNSQLEELKKTLOTR----TDELEVVRNTEV--DENATTAEQMDQHMSTLQKYSIDISLEVEQVWLNQTKLELAINTMESSLTSLGGQPAFLFDESFG-

01 AQLVNVLTDAQRLNWLKLLTN-----  
02 -----  
03 -----  
04 -----  
05 -----  
06 -----  
07 -----  
08 -----  
09 -----  
10 LEQNFTEINGYNNEQQRNSNSTSSPGLGIMNELMRGGVLTLLRGRGRHFRRKSTPQPATPMTRSRQGRFNKLQPRSQSLGSLSVIRDGNPSPARYEPIFN  
11 -----  
12 -----  
13 -----

01 -----  
02 -----  
03 -----  
04 -----  
05 -----  
06 -----  
07 -----  
08 -----  
09 -----  
10 HRLRQAASVHYLGEEIATSSSTNPPDLTRLRRTQCSMLCLGEDEEPVVLASAPPLTQLTAAVLTNTNNNHAYADLELDKDKDTSPSPECKGEQIQPKKEQI  
11 -----  
12 -----  
13 -----

01 -----DSTTSYPN-----TTPITFSRDPDITTTTTTTTNGNSNNNNHEDIDIQNGTDDNFSQALNGHDNSIPLCNLSD  
02 -----  
03 -----  
04 -----  
05 -----  
06 -----  
07 -----  
08 -----  
09 -----  
10 RIEINQTAQNSIDAHLDRIDELNRVLDLDRKRTLQPSDDVNAIESAEENHIQTRKLAKPDSQTKRSSSSSSSECRSSKDTSHSKKRSLSFSQKSIENIF  
11 -----  
12 -----  
13 -----

01 SYSTDPPKFIEQISNSPFALSTNNTSDMSSSTEHQOQQQRPERNRLWQSVRQMSKFRSHFIRNNNIRSSSTSSIPN-----TRTYQCRLEITITITPI  
02 -----ALPQEDRCARSSTDOERSG-----  
03 -----VVNYEEDARSVTSMERKE-----R  
04 -----VPLLQDDRHSSTSSSE--REGG-----R  
05 -----TDPTSPTTPTHGCDHASPATLDTESVENGG-----KKAEEKGGNSKAAKI  
06 -----VVNYEEDARSVTSMERKE-----R  
07 -----VVNYEEDARSVTSMERKE-----R  
08 -----VVNYEEDARSVTSMERKE-----R  
09 -----VLLQDDRHSSTSSSEQEREGG-----R  
10 SNLKEFSKSPVLRMGNHILNEEQDAKRTQPSQHHSSGSDCPTNSSSSSSNNNNNKNTSSNSNHSASQSTIITSTITTTTTTTTTTTPSKENSRLKPK  
11 -----  
12 -----VLLQDDRHSSTSSSEQEREGG-----R  
13 -----NVDQOMLDHAEQVNGGQDTAAAHPTSTKLFSAMLG-----G-----VKSMKNKV

01 LNTVNDNNNSNSNSNNKPLHNPTVNLKQONHLG-----  
02 VTALKTIKNHISGIFS-PRFSLPP-----  
03 LSKFESIRHSTAGLIKSPKSVLGS-----  
04 TPTLEILKSHFSGIFR-PRFSLPP-----

05 SNTLKKSFAGKFSMFRHRMOSNTSOSIDSTEG-----  
06 LSKFESIRHSTAGIIRSPKSAVGS-----  
07 LSKFESIRHSTAGIIRSPKSAVGS-----  
08 LSKFESIRHSTAGIIRSPKSAVGS-----  
09 TPTLEILKSHISGIFR-PRFSLPP-----  
10 VPKIQKSKAIRNTRFSKLLNFQLKRSKPKQCTKRRRIHPSKSVDFFAKEFEVEQPAGSAADEQFCNCPAGQKPVKPSVQISGHKDHFPFSSSGELDE  
11 MGTVERK-----  
12 TPTLEILKSHISGIFR-PRFSLPP-----  
13 KKKALKMKTSMTATSHAQSSNLA VSSSPG-----

01 ---MDTPFYHQNRKENHQNHQSNRESQDSTSILTSFPICDMDINKEPWFHCVLPRAEVEFLLQNGDFLVROTSSKRSSGRDTIGHWNGTNGDLINGNH  
02 P--VPLIPEVOKPLCOQAWYHGAI PRSEVOELLRYSGDFLVRESQ-GK-Q  
03 STQVCDVISVGERPLAEHDWYHGAI PRIEAOELLKQOGDFLVRESH-GK-PG  
04 P--LQLVPEVOKPLAEQWYHGAI PRAEVALLTHGDFLVRESQ-GK-Q  
05 -EAGEPEGETATDAASSADPSAIVEDNESETGSTTDVPPETRVBEERWFYPNMERREAEKCKNTGDMYLRYS--SK-QN  
06 S-TFSDTIPISEKPLAEQWYHGAI PRIEAOELLKQOGDFLVRESH-GK-PG  
07 S-ALSDMISISEKPLAEQWYHGAI PRIEAOELLKQOGDFLVRESH-GK-PG  
08 STQVCDVISVGERPLAEHDWYHGAI PRIEAOELLKQOGDFLVRESH-GK-PG  
09 P--LQLIPEVOKPLAEQWYHGAI PRAEVALLVHSGDFLVRESQ-GK-Q  
10 NSDRDINDDEEEDSASDDVLSMKDHCYCVPSLAASISLSTNRPLVEEWFHCVLPRAEVEFLLNNDGDFLVRETI RNE-ES  
11 SNNDASVTDDIRSAEYHGMVPRDAEGFLKREGDFLVKRTQMPGKV-----  
12 P--LQLVPEVOKPLAEQWYHGAI PRAEVALLTHSGDFLVRESQ-GK-Q  
13 ---DMEDEFDEDEVEYDTPAEHAGVEEPEWYHGEISROEATGLLKMDFLVRS--AD-KE

01 NETILNKDHNMDGTVLRM VLSVFWHGHR--HFILYGGPE--TGEWHLDD-GHESTIRELEEMKTOPTVAKSGACVTEISRP---DWEIDNRDVOI  
02 EYVLSVLDGQPR-HFIIQAA---DNLYRLED-DGLPIPLIDHLLQSOREIETKSCIVLTRAVLKD--K--WVLNHEDVLI  
03 EYVLSVSDGQRR-HFIIQFV---DNLYRREG-TGFSNIPOLIDHFFNTRKQVITKKSQVVLINLTPKD--KKWVLNHEDVLI  
04 EYVLSVMWDCGHR-HFIIQSL---DNLYRLEG-DGFSIPLIILHLLSSOQLTKKSQVVLFRVAVPKD--K--WVLKHHEDVLI  
05 RYVLTVCWAGQOK-HEVIOEAPDEQTKTKYRFES-RSEFVSRELDHFNVSQETVTRASGAI SREILREDDK--WALKHEDIKM  
06 EYVLSVSDGQRR-HFIIQFV---DNLYRREG-TGFSNIPOLIDHFFNTRKQVITKKSQVVLINLTPKD--KKWVLNHEDVLI  
07 EYVLSVSDGQRR-HFIIQFV---DNLYRREG-TGFSNIPOLIDHFFNTRKQVITKKSQVVLINLTPKD--KKWVLNHEDVLI  
08 EYVLSVSDGQRR-HFIIQFV---DNLYRREG-TGFSNIPOLIDHFFNTRKQVITKKSQVVLINLTPKD--KKWVLNHEDVLI  
09 EYVLSVLDGCLPR-HFIIQSL---DNLYRLEG-DGFSIPLIILHLLSTQQLTKKSQVVLFRVAVPKD--K--WVLNHEDVLI  
10 QIVLSVQWNGCHKH-FIVQTTG---EGNERFEG-PFEASIQELIMHOYHSELEVTIVKSGAIDRRVCRE--R--WELSNDDVLI  
11 VLAMSVRVTDDELCHHEMLNMDP---TSNKEYFETHQESTISELINWEMTKTPEISAASGAKTRRMEERS--P--WVLNHEDVLI  
12 EYVLSVLDGQPR-HFIIQSA---DNLYRLEG-DGFSIPLIILHLLSSOQLTKKSQVVLINLTPKD--K--WVLNHEDVLI  
13 HVTLVKMEGCIKH-FIIQHNE---EG-ERFEG-DYPSIPLIILHFFKNRLEVTRKSOALRQEVKRPDLREYNISHONILI

01 LQIKGCGNFGD VYRGVY---NGCEVAVKTCRVDMTASDPRRKFLOGETTALNENHPNIVKLVGLAVOSYPTIVMVEYVGGSLNLHNR--KSKNALPVMKI  
02 GPRIGRGNFGEVFSGRIL-RADNTPVAVKSCRETLPP-ELKIKFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLSFLR--SKGPRIRKMKKI  
03 GELLGKGNFGEVYKGL-K-DKTPVAVKTCREDLPO-ELKIKFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--KRDDELKIKOL  
04 GEQIGRGNFGEVFSGRIL-RADNTPVAVKSCRETLPP-DLKAKEFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--TEGARIRVKTIL  
05 GRKLGNGAFGDVWEAAL-K-GQKVAVKTCRSTDPV-D-REKFLQEAELIKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--KKAQMTKGGKI  
06 GELLGKGNFGEVYKGL-K-DKTPVAVKTCREDLPO-ELKIKFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLSFLR--KXKDEIKIKOL  
07 GELLGKGNFGEVYKGL-K-DKTPVAVKTCREDLPO-ELKIKFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--KXKDEIKIKOL  
08 GELLGKGNFGEVYKGL-K-DKTPVAVKTCREDLPO-ELKIKFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--KXKDEIKIKOL  
09 GEQIGRGNFGEVFSGRIL-RADNTPVAVKSCRETLPP-DLKAKEFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--TEGARIRVKTIL  
10 LBRIGRGNFGEVYKGL-K-STKLDVAVKTCRMTLP-EOKRKFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--KNSNGITRQO  
11 NKKLECGAFGDVWEAAL-K-GQKVAVKTCRSTDPV-D-REKFLQEAELIKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--KXKDEIKIKOL  
12 GEQIGRGNFGEVFSGRIL-RADNTPVAVKSCRETLPP-DLKAKEFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLTFPLR--TEGARIRVKTIL  
13 GPRIGRGNFGEVFSGRIL-KTHNMEVAVKSCRESEDFP-D-KOKFLOEAKILKQYDHPNIVKLVGLAVOSYPTIVMVEYVGGDFLSFLR--GPEGRIRVKNLI

01 LQMSIDAAAGM VLEARNCIHRDLAARNCLISDDGOLKIDAFGMSREE--HIVELSDKRGQIPKWTAPEALRTGRTYTIKCDVWSYGVLLWEHFTGDPV  
02 IKMMENAAAGM VLESKCIHRDLAARNCLVTEKNTKISDFGMSRCEDDGVYASGCGKQIPVKTWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
03 VRFSLDAAAGM VLESKCIHRDLAARNCLVGENNVLKISDFGMSRCEDDGVYSSS--GLKQIPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
04 LQVGDAAAGM VLESKCIHRDLAARNCLVTEKNTKISDFGMSREEDGCVYASAGLROVVPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
05 LHMSPDACCME VLESKCIHRDLAARNCLVGENNVLKISDFGMSREEDGCVYVOSCTRQIPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
06 VKFSLDAAAGM VLESKCIHRDLAARNCLVGENNVLKISDFGMSRCEDDGVYSSS--GLKQIPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
07 VKFSLDAAAGM VLESKCIHRDLAARNCLVGENNVLKISDFGMSRCEDDGVYSSS--GLKQIPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
08 VRFSLDAAAGM VLESKCIHRDLAARNCLVGENNVLKISDFGMSRCEDDGVYSSS--GLKQIPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
09 LQVGDAAAGM VLESKCIHRDLAARNCLVTEKNTKISDFGMSREEDGCVYASAGLROVVPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
10 MCMCRDAAAGM VLESKCIHRDLAARNCLVDEHSVKISDFGMSREEDGCVYSSS--HIVSDGKQIPVKTWTAPEALNYGYSSESVDVWSFGILLWETFSRGDTP  
11 LRETTFDAAAGM VLESKCIHRDLAARNCLLSAKNBLKISDFGMSDNKDE--IKDETLEKVPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
12 LQVGDAAAGM VLESKCIHRDLAARNCLVTEKNTKISDFGMSREEDGCVYASAGLROVVPKWTAPEALNYGYSSESVDVWSFGILLWETFSLGVP  
13 TEMCVBAAAGM VLEORNCIHRDLAARNCLVGENNVLKISDFGMSREEDGCVYSSSGLKQIPKWTAPEALNYGYSSESVDVWSFGILLWETFSRGDTP

01 VRNWSNOQTRDMTES--GYRLEPAEDLMPVWLRTLMN-HCHDFEPMNRPSESKISNEIPIPNVNSFTPNLNRSIDNAQLKKSADINLSTTPLHLSVSVKDRR  
02 VANLSNOQTRAEIETQ--GYRLEPEOCPEDEVYRIMQ--RCWEYDEHRRPSEGAHVODLIAIRKRHR  
03 YPGMTNQAAREVER--GYRMSAFONCPEEVFTIMM-KCWDYKPEENRPKENDLHKELTVIKKMIT  
04 YPNLSNOQTRAEIETQ--GYRLEPEOCPEDEVYRIMQ--CCWAYDEPRKFAVLESKPKKLDGTVO  
05 YPGMSNAETRTQVDS--GYRMPPEOCPEAEITHOIMQ--LCWAYDAEDRPRKFAVLESKPKKLDGTVO  
06 YPGMTNQAAREVER--GYRISAFONCPEEDIFKIMM-KCWDYKPEENRPKENDLHKELTVIKKMIT  
07 YPGMTNQAAREVER--GYRMSAFONCPEEDIFKIMM-KCWDYKPEENRPKENDLHKELTVIKKMIT  
08 YPGMTNQAAREVER--GYRMSAFONCPEEVFTIMM-KCWDYKPEENRPKENDLHKELTVIKKMIT  
09 YPNLSNOQTRAEIETQ--GYRLEPEOCPEDEVYRIMQ--CCWAYDEPRKFAVLESKPKKLDGTVO



```

10 YSGMTNSRARERIDT-GYRMPTEKSTPEEMYRLML-CQWADAESRPHDEIYNNVVDALILLRLDNSH-----
11 YPGLTKIQTRAKIVVNDYRMKVEDGTHPTVADVVTGTCWQKNPEKRSTMDSLHKKUREFYESKK-----
12 YPNLSNQQTREIYVEK-CGRLEPOEELCFDAVFRIME-CQWAVEEGORPSESATVQELQSLRKRHR-----
13 YPGMNNRETRERIES-NYRMPREELCPHFVYKMLL-SCWEYDEDRPESAEIYDKLYSFNHQ-----

```

```

01 TRED
02 ----
03 ----
04 ----
05 ----
06 ----
07 ----
08 ----
09 ----
10 ----
11 ----
12 ----
13 ----

```

Anexo 3: Alinhamento múltiplo de SmFes e de proteínas ortólogas de outros organismos. O alinhamento foi realizado com ClustalW. As seqüências usadas e o respectivo número de acesso no banco de dados foram: 1- Fes like de *Schistosoma mansoni*, (AAP68901); 2- FBS de Fujinami sarcoma vírus (NP\_955606); 3- Fert2 proteína de *Mus musculus*, (AAH58100); 4- Fps/Fes de *Mus musculus* (AAN33122); 5- PTK de *Ephydatia fluvialis* (BAA81721); 6- Fer de *Canis familiaris* (AAF00543); 7- Fer de *Homo sapiens* (NP\_005237); 8- Fer de *Mus musculus* (AAB18988); 9- c-FES de *Homo sapiens* (P07332); 10- Fps85D(dFer) de *Drosophila melanogaster* (P18106); 11- Fer-frk-1 de *Caenorhabditis elegans* (NP\_501818); 12- Fes/Fps de *Felis catus* (TVCTFF); 13- PTK de *Sycon raphanus* (CAA76605).