

Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

Instituto Oswaldo Cruz

Curso de Pós-Graduação em Biologia Parasitária

Priscila Monnerat de Oliveira Costa

Exploração do genoma de Phytomonas serpens

Rio de Janeiro

Abril de 2006

Ficha catalográfica elaborada pela
Biblioteca de Ciências Biomédicas/ ICICT / FIOCRUZ - RJ

C837

Costa, Priscila Monnerat de Oliveira.

Exploração do genoma de *Phytomonas serpens* (Gibbs, 1957; Jankevicius et al., 1989) / Priscila Monnerat de Oliveira Costa. – Rio de Janeiro, 2006.

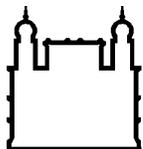
xx, 104 f. : il. ; 30 cm.

Dissertação (mestrado) – Instituto Oswaldo Cruz, Curso de Pós-graduação em Biologia Parasitária, 2006.

Bibliografia: f. 85-102

1. Genoma. 2. *Phytomonas serpens*. 3. Gene Ontology. I. Título.

CDD 575.1



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

Instituto Oswaldo Cruz

Curso de Pós-Graduação em Biologia Parasitária

Priscila Monnerat de Oliveira Costa

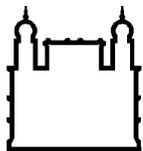
Exploração do genoma de *Phytomonas serpens*

Dissertação apresentada ao Instituto
Oswaldo Cruz como parte dos requisitos
para obtenção do título de mestre em
Biologia Parasitária

Orientador: Dr. Alberto Martin Rivera Dávila

Rio de Janeiro

Abril de 2006



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

Instituto Oswaldo Cruz

Curso de Pós-Graduação em Biologia Parasitária

Priscila Monnerat de Oliveira Costa

Exploração do genoma de *Phytomonas serpens*

Orientador: Dr. Alberto Martin Rivera Dávila

Aprovada em: 17 de Abril de 2006.

EXAMINADORES:

Dra. Ana Carolina Paulo Vicente

Dr. Mário Steindel

Dr. Rodolpho Mattos Albano

Rio de Janeiro

Abril de 2006

*Aos meus pais e minha irmã, pela ajuda para
tornarem meus sonhos realidade.*

*Para meu marido, meu pilar, por tudo o que
você significa para mim.*

Agradecimentos

A Deus,

Ao meu orientador Dr. Alberto Martin Rivera Dávila, meu “Chefinho”, pela oportunidade, pelo apoio e por me fazer uma das alunas mais invejadas pelo fato de ter um orientador assim.

À Dra Yara Traub-Cseko, pelos seus conselhos e sua amizade.

Ao CNPq, pelo apoio financeiro.

Ao Dr. Rodolpho Mattos Albano por revisar a minha tese.

A equipe do Laboratório de Genoma da UERJ, principalmente a Denise, pela ajuda.

Aos meus anjos da guarda, Silvana, Glauber, Luana e Erich, não sei o que faria sem vocês.

Ao meu amigo e irmão Cesare Bianco Junior e aos colegas do Laboratório de pesquisas em Malária pela ajuda prestada.

Aos colegas de laboratório, Amanda, Vicente, Luanda, Chris, André, Adriana, Igor, Anissa, Viviane, Anderson, Nicolas, Shejy, Marcel, Ingrid, por fazer da minha vida no laboratório muito mais divertida.

Aos colegas da bioinformática Kary, Henrique, Rafael, Ildefonso, Ricardo, Carol, Marcos, Antônio, Thomas e Paulo.

Aos meus amigos do curso de Biologia Parasitária, sempre juntos em todas as dificuldades.

Aos meus eternos amigos de faculdade por ouvirem minhas lamentações.

Aos amigos da secretaria, Elisângela, Patrícia, Marivaldo e Alberto, pela ajuda com as burocracias do dia a dia.

Aos meus amigos e parentes, desculpem pela ausência.

A todos que, mesmo que indiretamente, contribuíram para que conseguisse chegar tão longe.

"Os ideais que iluminaram meu caminho e sempre me deram coragem para enfrentar a vida com alegria foram a Verdade, a Bondade e a Beleza."

Albert Einstein

O estudo de tripanosomatídeos de plantas teve início em 1909, quando foram encontradas formas flageladas em látex, sendo depois encontradas em floema, frutas e flores. Desde a refutada tentativa de Donovan de criar o gênero *Phytomonas*, temos atualmente, várias espécies classificadas dentro do gênero *Phytomonas*, incluindo *Phytomonas serpens*, organismo alvo de nosso estudo. O ciclo de vida de *P. serpens* compreende planta e inseto. A transmissão é feita do tomate para inseto, do inseto para tomate. O vetor natural é o hemíptero *Phthia picta* (Hemiptera: Coriidae) que quando se alimenta de tomates infectados, se torna infectado após 10 a 15 dias, apresentando parasitos nas fezes, urina, tubo digestivo e glândulas salivares. Pouco se sabe sobre *P. serpens* e o conhecimento é ainda mais escasso se levarmos em conta todos os organismos dentro do gênero *Phytomonas*. O objetivo deste projeto é explorar o genoma de *P. serpens* gerando e analisando seqüências de seu genoma, aumentando o conhecimento deste organismo e comparando seu genoma com outros organismos. A cepa 9T de *P. serpens* (CT–IOC-174) foi cultivada a 27°C em meio Schneiders´insect medium suplementado com 10% de soro fetal bovino. O DNA genômico foi extraído por lise alcalina para a construção da biblioteca genômica. O DNA foi parcialmente digerido com a enzima de restrição *Sau3A*. Os fragmentos de DNA com tamanhos entre 1-3 kb foram recuperados do gel de agarose e purificados. Os fragmentos de DNA foram clonados no sítio *BamHI* do vetor de clonagem pUC18. A reação de sequenciamento foi realizada na plataforma de sequenciamento ABI3730 – 48 capilar PDTIS/FIOCRUZ e na plataforma MegaBase na UERJ. Todas as seqüências foram armazenadas em banco de dados “open source” nos servidores Linux no Laboratório de Biologia Molecular de Tripanosomatídeos e Flebotomíneos (DBBM/IOC/FIOCRUZ). Foram seqüenciados 829 clones da biblioteca de DNA genômico obtendo-se um total de 379 seqüências GSS (Genomic Sequence Survey) de alta qualidade. Foram utilizadas para a complementação deste estudo, as seqüências do projeto EST de *P. serpens* que estavam disponíveis no GenBank. Os 221 clusters GSS e os 697 clusters de EST (Expressed Sequence Tag) foram comparados com o programa de busca de similaridade Blast a diferentes bancos de dados, utilizando como parâmetro 10^{-5} , gerando 599 clusters com entradas positivas (Hits) e 303 clusters com nenhum hit, representando possíveis genes espécie específicos. Os clusters foram anotados de acordo com sua função quando comparados ao banco de dados Gene Ontology (GO) representando uma análise inédita no genoma de *P. serpens*. Inferências filogenéticas e de similaridade com o banco “Taxonomy” do NCBI foram realizados, usando inclusive seqüências do genoma ambiental, para verificar se as mesmas pertencem a Kinetoplastida, o que não se comprovou. As inferências filogenéticas realizadas com genes concatenados mostraram que *P. serpens* é mais próximo de *T. cruzi*, enquanto que inferências com genes individuais apontaram para *L. major*. Com base na literatura, inferimos que a análise da árvore de genes concatenados é correta. Foram encontradas com o GLIMMER 20 genes hipotéticos nas seqüências de GSS. Encontramos 22 cluters em GSS que não foram encontrados anteriormente em EST, como por exemplo, Piroglutamil-peptidase I, antígeno nuclear de proliferação celular (PCNA) e Peptidase_M8.

Abstract

The study of trypanosomatids of plants began in 1909, when flagellate forms were found in latex from *Euphorbiaceae*, being later found in phloem, fruits/seeds and flowers in a wide variety of plants species. Since the refuted attempt of Donovan to create the genus *Phytomonas*, a great number of species were classified in the genus *Phytomonas*, including *Phytomonas serpens*, the major organism of our study. The described life cycle of *P. serpens* involves plant (tomato) and insect. The natural vector is the Coreid bug, *Phthia picta* (Hemiptera: Coriidae), that becomes infected 10 a 15 days after feeding on infected tomatoes, presenting the parasite in excrements, urine, digestive tube and salivary glands. Very little is known about these organisms and even less is known about the genus *Phytomonas*. The goal of this project is to explore the *P. serpens* genome by generating and analyzing sequences of its genome aiming to increase the knowledge of this organism and to analyze comparatively with genomes of other organisms. *P. serpens* infects tomatoes but we still have doubts about its pathogenicity. A strain of *P. serpens* 9T (CT-IOC-174) was cultured at 27°C in Schneiders' insect medium supplemented with 10% heat-inactivated fetal bovine serum. The DNA was extracted by alkaline lyses for the construction of a genomic library. Genomic DNA was partially digested with the restriction enzyme *Sau3A*. DNA fragments with an average size of 1.0-3.0 kb were recovered and purified from agarose gel. DNA inserts were cloned into the *BamHI* restriction site of pUC18 cloning vector. The sequencing reaction was made at the Sequencing Platform ABI3730 - PDTIS/FIOCRUZ and MegaBase platform at UERJ. All the sequences were stored in an open source database in the Linux server in the Laboratory of Molecular Biology of Trypanosomatids and Phlebotomines (DBBM/IOC/FIOCRUZ). 829 clones of the *P. serpens* genomic DNA library were sequenced obtaining a total of 379 high quality sequences GSS (Genome Sequence Survey). Sequences from the EST (Expressed Sequence Tag) project of *P. serpens* available at GenBank were used for the complementation of this study. 221 GSS clusters and 697 EST clusters were compared with different databases with the similarity search program Blast, using evaluate 10^{-5} as default, obtaining 599 clusters with positive hits and 303 clusters without hit, indicating possible species-specific genes. 357 clusters that were identified when compared with Gene Ontology (GO) had their function classified, representing the first analysis of *P. serpens* genome using GO. Phylogenetics and similarity studies with Taxonomy database from NCBI were carried out confirming the position of *P. serpens* in relations to the others trypanosomatids and disclosing similarity with *T. cruzi* genome. The phylogenetic study included sequences of the enviromental genome, in order to verify if these sequences belong to the Kinetoplastid, what was not proven to be true. Phylogenetic studies using concatenated genes showed that *P. serpens* is phylogenetically closer to *T. cruzi*, but studies with individuals genes pointed to *L. major*. Based on literature, we infer that the analysis of the tree of concatenated genes is correct. 20 hypothetical genes in the GSS sequences were found with the GLIMMER. We found 22 clusters in GSS sequences that were not found in EST before, among others, Pyroglutamyl-peptidase I, proliferating cell nuclear antigen (PCNA) and Peptidase_M8.

Figura 1.1 – Flagelados derivados do látex de <i>Euphorbia</i> . “A” corresponde a grânulos de amido (Lafond, 1910).....	4
Figura 1.2 – Polimorfismo de <i>Phytomonas serpens</i> , o parasita de tomate: 1 e 2, formas encontradas no fruto; 3 e 4, flagelados encontrados no meio de cultura bifásico “blood agar”/LIT; 5 e 6, grandes flagelados encontrados em meio de cultura LIT antigo e estacionado; 7 e 8, formas não flageladas derivadas da urina do vetor; 9 e 10, flagelados do trato digestivo do vetor (Jankevicius <i>et al.</i> , 1989).....	5
Figura 1.3 – distribuição na América Latina das infecções causadas por flagelados em café (○), coco e palmeiras (▲), mandioca (■) e tomates (●) (Jankevicius <i>et al.</i> , 1989).....	6
Figura 1.4 – Ciclo de vida de <i>Phytomonas serpens</i> (Fotografia por Drees).....	10
Figura 3.1- Esquema do vetor de clonagem pUC18.....	23
Figura 3.2: Organograma do processo de anotação do programa GARS.....	28
Fig. 4.1: DNA genômico de <i>P. serpens</i> na concentração aproximada de 5 µg/µl digerido com 6U (A) e 3U (B) da enzima <i>Sau3A</i> nos tempos de 2, 4, 6 8 e 10 segundos. Legenda: M1 - marcador de peso molecular ΦX174; 2, 4, 6, 8 e 10 - DNA genômico digerido nos tempos de 2, 4, 6, 8 e 10 segundos; DI – DNA genômico não digerido; M2 e M3 - marcador de peso molecular λ <i>HindIII</i> ; M4 - marcador de peso molecular 100 pb. Gel de agarose 1% corado com brometo de etídio.....	36
Fig. 4.2: DNA genômico de <i>P. serpens</i> na concentração aproximada de 5 µg/µl (A) e de 1 µg/µl (B) digerido com 1U (A) e 0,5U (B) da enzima <i>Sau3A</i> nos tempos de 5, 15 e 45 minutos (A) e 30, 90 e 150 minutos (B). Legenda: M1 - marcador de	

peso molecular λ HindIII; 5, 15 e 45 - DNA genômico digerido nos tempos de 5, 15 e 45 minutos (A); 30, 90 e 150 - DNA genômico digerido nos tempos de 30, 90 e 150 minutos (B); DI - DNA genômico não digerido; M2 - marcador de peso molecular 100 pb. Gel de agarose 1% corado com brometo de etídio..... 37

Figura 4.3: DNA plasmidial digerido com as enzimas *EcoRI* e *PstI*. Legenda: 1 - marcador de peso molecular λ HindIII; 2 - marcador de peso molecular 100 pb. Gel de agarose 1% corado com brometo de etídio..... 37

Figura 4.4 - Árvore filogenética do gene HSP 70 (cluster PSADEST007C03.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 52

Figura 4.5 - Árvore filogenética do gene frutose-1,6-bifosfato aldolase (PSADEST006G10.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 53

Figura 4.6 - Árvore filogenética do gene Beta-tubulina (PSADEST005H07.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 53

Figura 4.7 - Árvore filogenética do gene serina-treonina kinase (PSADEST001A09.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 54

Figura 4.8 - Árvore filogenética do gene de fator de alongamento (PSADEST002G02.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 54

Figura 4.9 - Árvore filogenética do gene gliceraldeído-3-fosfato desidrogenase (PSADEST005H09.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 55

Figura 4.10- Árvore filogenética do gene adenilato ciclase (PSADSAU009F03.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 55

Figura 4.11- Árvore filogenética concatenada dos genes fator de alongamento, histona H2, HSP70, tubulina, P450 reductase e adenilato ciclase utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000..... 56

Lista de Tabelas

Tabela 1.1: Números de entradas no Entrez do NCBI nas categorias de Nucleotide, Protein e PubMed no dia 07 de fevereiro de 2006.....	11
Tabela 3.1: Meios de cultura utilizados neste estudo.....	Anexo I
Tabela 3.2: Soluções e tampões utilizados neste estudo.....	Anexo II
Tabela 3.3: Concentrações de DNA genômico e da enzima <i>Sau3A</i> e os tempos utilizados nas reações de digestão.....	21
Tabela 3.4: Concentrações de reagentes utilizados nas reações de seqüenciamento.....	26
Tabela 3.5: Condições das reações de sequenciamento.	27
Tabela 3.6: Bancos de dados utilizados neste estudo	32
Tabela 4.1: Total de clusters da biblioteca genômica classificados pelos números de reads pelos quais são formados.....	38
Tabela 4.2: Número de GSS de DNA genômico de <i>Phytomonas serpens</i> geradas pelo sequenciamento e sua distribuição após a análise de qualidade e agrupamento.....	36
Tabela 4.3: Número total de entradas (Hits) para cada banco de dados confrontados com as GSS geradas neste estudo e as seqüências de ESTs obtidas no GenBank.....	40

Tabela 4.4: Lista de genes hipotéticos encontrados nas seqüências de GSS da biblioteca <i>Sau3A</i> pelo programa GLIMMER. Os clusters sem anotação foram os que não deram hit com o Blast.....	42
Tabela 4.5: Lista de GSS de <i>P. serpens</i> que apresentaram similaridade significativa (evalue < 10 ⁻⁵) quando comparadas com os bancos de dados de domínios de proteínas CDD e Interpro.....	43
Tabela 4.6: Os valores encontrados dos sítios conservados, sítios variáveis, singletons e sítios parcimoniosos nas seqüências utilizadas nas árvores filogenéticas.	51
Tabela 4.7: Nome dos organismos e número de hits obtidos com o “Taxonomy” quando comparamos as seqüências GSS e EST de <i>P. serpens</i> com o banco de dados refseq_protein com evalue 1e ⁻⁵	57

Gráfico 4.1: Representação gráfica da tabela 4.3 onde temos o número de entradas (Hits) para cada banco de dados confrontados com as GSS geradas neste estudo e as seqüências de ESTs obtidas no GenBank.....	41
Gráfico 4.2: Relação das funções atribuídas aos clusters GSS de <i>Phytomonas serpens</i> com base nas definições do vocabulário de processo biológico do GO.....	45
Gráfico 4.3: Relação das funções atribuídas aos clusters GSS de <i>Phytomonas serpens</i> com base nas definições do vocabulário de componente celular do GO.....	46
Gráfico 4.4: Relação das funções atribuídas aos clusters GSS de <i>Phytomonas serpens</i> com base nas definições do vocabulário de função molecular do GO.....	47
Gráfico 4.5: Relação das funções atribuídas aos clusters EST de <i>Phytomonas serpens</i> com base nas definições do vocabulário de função molecular do GO.....	48
Gráfico 4.6: Relação das funções atribuídas aos clusters EST de <i>Phytomonas serpens</i> com base nas definições do vocabulário de processo biológico do GO.....	49
Gráfico 4.7: Relação das funções atribuídas aos clusters EST de <i>Phytomonas serpens</i> com base nas definições do vocabulário de componente celular do GO.....	50
Gráfico 4.8- Representação gráfica dos dados da tabela 4.6 mostrando os nomes dos organismos e número de hits obtidos com o “Taxonomy” quando comparamos as seqüências GSS e EST de <i>P. serpens</i> com o banco de dados “refseq_protein” com evalue $1e^{-5}$	58

Lista de Abreviaturas, Siglas e Símbolos

DNA -	Ácido desoxirribonucléico
RNA -	Ácido ribonucléico
mRNA -	RNA mensageiro
cDNA -	DNA complementar
EST -	Expressed Sequence Tag
GSS -	Genome Sequence Survey
kDNA -	DNA do cinetoplasto
Kb -	Kilobases
gRNA -	RNA guia
pb -	Pares de base
CSB -	Bloco de seqüência conservada
μ M -	Micrômetro
μ L -	Microlitro
μ g -	Micrograma
LB -	Luria-Bertani
X-gal -	5-bromo-4-cloro-3-indol- β -D-galactosídeo
IPTG -	Isopropil- β -D-tiogalactopiranosídeo

mM -	Milimolar
ml -	Mililitro
min. -	Minuto
rDNA -	DNA ribossomal
pmol -	Picomol
SDS -	Dodecil sulfato de sódio
SSC -	Solução salina de citrato
Rpm -	Rotação por minuto
TBE -	Tris base / Borato / EDTA
EDTA -	Etilenodiaminatetracetato
Volt -	Voltagem
Kvolt -	Kilovolt
Glimmer -	Gene Locator and Interpolated Markov Modeler
ORF -	Janela aberta de leitura (<i>open reading frame</i>)
BLAST -	Basic Local Alignment Search Tool
NCBI -	National Center for Biotechnology Information
EMBL -	European Molecular Biology Laboratory
EBI -	European Bioinformatics Institute

CDD -	Conserved Domain Database
GO -	Gene ontology
tRNA -	RNA transportador
PCR -	Reação em cadeia da polimerase (<i>polymerase chain reaction</i>)
pg -	Picograma
fg -	Fentograma
mg -	Miligrama
nr -	Não redundante
nt -	Nucleotídeo
v/v -	Volume-volume
p/v -	Peso-volume
dNTP -	Desorribonucleotídeo trifosfatado
TIGR -	The Institute for Genomic Research

1- Introdução.....	1
1.1 – Tripanosomatídeos.....	2
1.2 – Gênero <i>Phytomonas</i>	3
1.3 – <i>Phytomonas serpens</i>	8
1.4 – Descoberta de genes.....	11
1.5 – Bioinformática e Anotação de Sequências.....	14
2- Objetivos.....	17
2.1 – Objetivo.....	17
2.2 – Específicos.....	17
3- Materiais e Métodos.....	19
3.1- Organismos.....	19
3.2- Meios de cultura.....	19
3.3- Tampões e soluções.....	19
3.4- Antibiótico.....	19
3.5- Marcadores de peso molecular.....	20
3.6- Extração do DNA genômico pelo método Formol/clorofórmio.....	20
3.7- Digestão com enzimas de restrição.....	20
3.8- Eletroforese de DNA.....	22
3.9- Extração de DNA de gel de agarose.....	22
3.10- Clonagem.....	22
3.11- Transformação de bactérias.....	24
3.12- Mini-preparação plasmidial.....	24
3.13- Sequenciamento.....	26
3.14- Iniciadores utilizados para a sequenciamento dos clones (pUC 18)...	27
3.15- Precipitação da reação de sequenciamento.....	27
3.16- Análise de sequências.....	28
4- Resultados.....	35
4.1 – Digestão do DNA genômico de <i>Phytomonas serpens</i>	36
4.2 – Sequenciamento.....	38
4.3 – Análises “in silico”.....	39
4.4 – Classificação funcional das GSS e EST de <i>Phytomonas serpens</i>	44

4.5 – Filogênia de <i>Phytomonas serpens</i>	51
4.6- Similaridade utilizando o banco taxonomy do NCBI.....	57
5- Discussão.....	59
6- Conclusão.....	66
7- Bibliografia.....	69
Anexo I.....	88
Anexo II.....	89

CAPÍTULO

1

REVISÃO DA LITERATURA

1.1 – Kinetoplastida

A ordem Kinetoplastida é um grupo de protozoários que possuem um ou dois flagelos que emergem de uma bolsa flagelar. Sua característica principal é a presença do cinetoplasto, uma região próxima aos corpúsculos basais do flagelo onde encontramos o DNA mitocondrial condensado (Vickerman 1976). É desta região de onde vem a denominação da ordem.

A presente taxonomia dos kinetoplastidas é baseada na morfologia e nos ciclos biológicos. Tradicionalmente este grupo é subdividido em duas subordens: Bodonida e Trypanosomatina (Vickerman 1976, Lom 1976). O primeiro possui duas famílias: Bodonidae e Cryptobiidae, com membros que possuem um cinetoplasto extenso e dois flagelos: anterior (que sempre permanece livre) e recorrente (livre em bodonídeos e preso em cryptobiídeos). Bodonídeos possuem uma ampla variedade de formas de vida, se estendendo da vida livre, como por exemplo, *Bodo*, *Parabodo*, *Dimastigella*, *Rynchobodo* e *Cruzella*, à vida totalmente parasitária. Os Cryptobídeos são representados por ectoparasitos de pele e guelras de peixes como o *Ichtyobodo*, endoparasitos do trato reprodutivo de cobras (algumas espécies de *Cryptobia*) e trato gastro-intestinal de peixes (outra *Cryptobia*) e também por parasitas digenéticos de sangue de peixes (*Trypanoplasma*) transmitido por sanguessuga. O segundo grupo, Trypanosomatina, pertence à ordem Kinetoplastida. Os trypanosomatídeos formam uma única família, Trypanosomatidae, a qual inclui organismos parasitas obrigatórios com um único flagelo e um cinetoplasto pequeno. Tripanosomatídeos foram encontrados em todas as classes de vertebrados, bem como em alguns invertebrados, preferencialmente em insetos das ordens Díptera e Hemíptera, e também em plantas. Os invertebrados podem tanto representar hospedeiros, como para os parasitas monofiléticos *Crithidia*, *Leptomonas* e *Blastocrithidia*, ou servir como vetores, como nos gêneros digenéticos *Trypanosoma* e *Leishmania* e parasitos transmitidos de insetos para plantas, como no gênero *Phytomonas*. Um novo gênero, *Wallaceina*, foi recentemente proposto para incluir parasitos monogenéticos de insetos com uma morfologia “endomastigota”, caracterizada por uma bolsa flagelar longa e curva (Maslov *et al.*, 2001).

1.2 – Gênero *Phytomonas*

Em 1909, Alexandre Lafond, diretor do Laboratoire de Bacteriologie de Mauritius, na Costa da África, coordenou um estudo comparativo de várias *Euphorbia pilulifera* e ordenou seu técnico, o Sr. David, que examinasse o látex destas plantas. Foram encontradas formas flageladas de movimentos rápidos, a quem Lafond inicialmente denominou *Leptomonas davidi* (Lafond 1909, 1910). Logo depois, cientistas do mundo inteiro começaram a achar flagelados em outras espécies de *Euphorbia*, porém demoraram mais de 20 anos para começarem a procurar em outros tipos de plantas além das lactíferas. Até o presente momento, estes organismos foram encontrados em látex, floema, frutas e flores (Camargo, 1999). No látex, esses tripanosomatídeos foram achados nas seguintes famílias de plantas: Euphorbiaceae, Asclepiadaceae, Apocynaceae, Cecropiaceae, Compositae, Moraceae, Urticaceae e Sapotaceae. São encontrados nos tubos de seiva de Palmae e Rubiaceae e nas frutas de Anacardiaceae, Oxilidaceae, Passifloraceae, Punicaceae, Rosaceae, Rutaceae e Solanaceae (Camargo *et al.*, 1990).

Em 1909, Donovan propôs que fosse criado um novo gênero, *Phytomonas*, onde se agrupariam todas as novas espécies descobertas em plantas. Esta proposta, no entanto, gerou uma grande polêmica, sendo refutada por vários cientistas. Somente em 1927 foi dado o nome *Phytomonas* a uma espécie, *Phytomonas francai* (Aragão 1927).

A definição tradicional do gênero *Phytomonas* (família Trypanosomatidae, ordem Kinetoplastida) era baseada em dois critérios – morfologia promastigota e a ocorrência em plantas.

Ao examinarmos os organismos do gênero *Phytomonas*, podemos observar sem dificuldades, características comuns da família Trypanosomatidae: um núcleo, um cinetoplasto, um corpo celular e um flagelo. Também podemos avistar várias torções ao longo de seu eixo longitudinal, uma característica comum em flagelados de plantas, mas não exclusiva, pois também pode ser encontrada em flagelados monoxênicos de insetos (Fig. 1.1). Outra característica estrutural é a presença de grande quantidade de glicossomos espalhados pelo corpo celular. Em cultura, os parasitas exibem características comuns aos tripanosomatídeos, mas com algumas peculiaridades, como uma crista mitocondrial muito escassa enquanto que a matriz mitocondrial é muito densa (Jankevicius *et al.*, 1989).

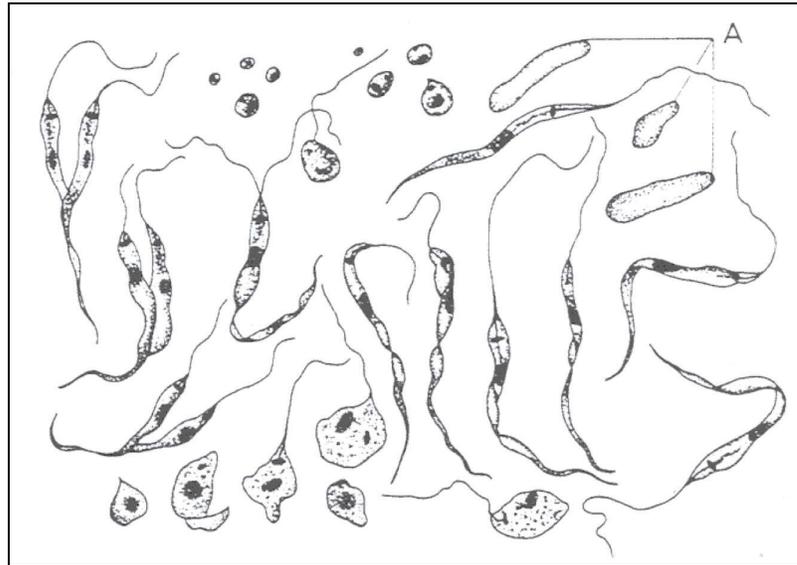


Figura 1.1 – Flagelados derivados do látex de Euphorbia. “A” corresponde a grânulos de amido (Lafond, 1910).

O ciclo de vida de *Phytomonas* inclui transmissão por insetos e propagação nos tecidos da planta (floema, látex e frutos). Todos os vetores de *Phytomonas* são insetos. A transmissão parece ser persistente e propagativa, já que quando o inseto é infectado, podemos achar os organismos deste gênero em todo o trato digestivo, hemolinfa e glândulas salivares. Todos os vetores que foram encontrados infectados com tripanosomatídeos de plantas pertencem ao grupo Pentatomomorpha: Lygaeoidea, Pentatomidae e Coreoidea (Mitchell, 2004).

Em todo seu ciclo de vida, *Phytomonas* estão sempre na forma promastigota, segundo a definição de Hoare & Wallace (1966), porém possuem alto grau de polimorfismo, isto é, seu tamanho, extensão do flagelo, forma do corpo celular e número de torções ao longo do corpo podem variar em diferentes meios de cultura e hospedeiros, sendo esse polimorfismo também encontrado na mesma cultura e no mesmo hospedeiro, mostrando uma grande plasticidade (Fig. 1.2).

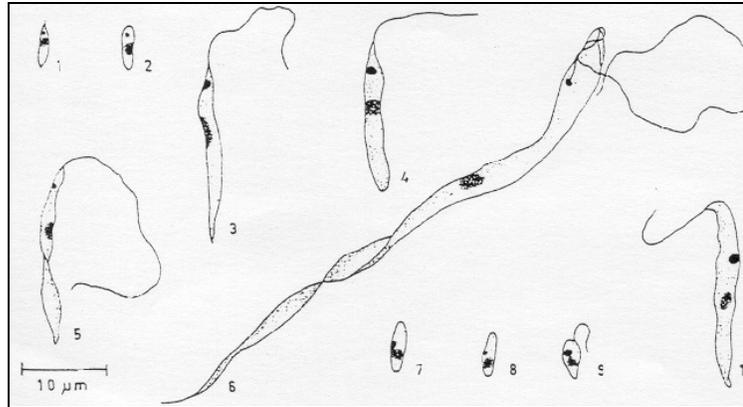


Figura 1.2 – Polimorfismo de *Phytomonas serpens*, o parasita de tomate: 1 e 2, formas encontradas no fruto; 3 e 4, flagelados encontrados no meio de cultura bifásico “blood agar”/LIT; 5 e 6, grandes flagelados encontrados em meio de cultura LIT antigo e estacionado; 7 e 8, formas não flageladas derivadas da urina do vetor; 9 e 10, flagelados do trato digestivo do vetor (Jankevicius *et al.*, 1989).

Esse polimorfismo foi comentado na revisão de Camargo (1999) como tendo duas importantes implicações:

- 1- Promastigotas morfologicamente distintos podem corresponder a estágios de desenvolvimento que ainda não foram reconhecidos no ciclo de vida dos flagelados. Em cada um dos seus hospedeiros, os flagelados de plantas passam por mudanças morfológicas e, presumivelmente, fisiológicas. Essas mudanças podem ser sequencialmente programadas e podem corresponder a passos do ciclo de vida do flagelado, como em outros dois gêneros heteroxênicos, *Trypanosoma* e *Leishmania*.
- 2- O marcante polimorfismo de flagelados de plantas inviabiliza o critério para a identificação das espécies. Devido ao polimorfismo intra-específico, promastigotas de uma espécie se assemelham em muito as de outras espécies, impedindo a distinção entre espécies de flagelados de plantas.

Sobre a patogenicidade de *Phytomonas* em plantas lactíferas, houve, inicialmente, uma controvérsia considerável. Depois foi aceito que *Phytomonas* não era

particularmente prejudicial à maioria das plantas. Mais tarde, contudo, foi mostrado que em algumas plantas economicamente importantes, como em café, coco, mandioca e palmeiras, infecções por *Phytomonas* podem causar conseqüências devastadoras (Camargo *et al.*, 1990). No café, a espécie *Phytomonas leptovosorum* causa necrose do floema e foi encontrada no Suriname, Colômbia, Guiana e nordeste do Brasil. Em palmeiras e coqueiros, *Phytomonas staheli* leva as plantas à morte pela perda de folhas e frutos sendo prevalente nas Américas do Sul e Central. Já na mandioca, uma marcante atrofia de raiz que ocorre no Brasil é causada por *Phytomonas francai*. Todavia, espécies de *Phytomonas* foram encontradas em vários frutos, como em pêssegos, tangerinas, tomates etc, mas não foi comprovado se os organismos causam algum dano nestas espécies (Camargo *et al.*, 1990) (Figura 1.3).

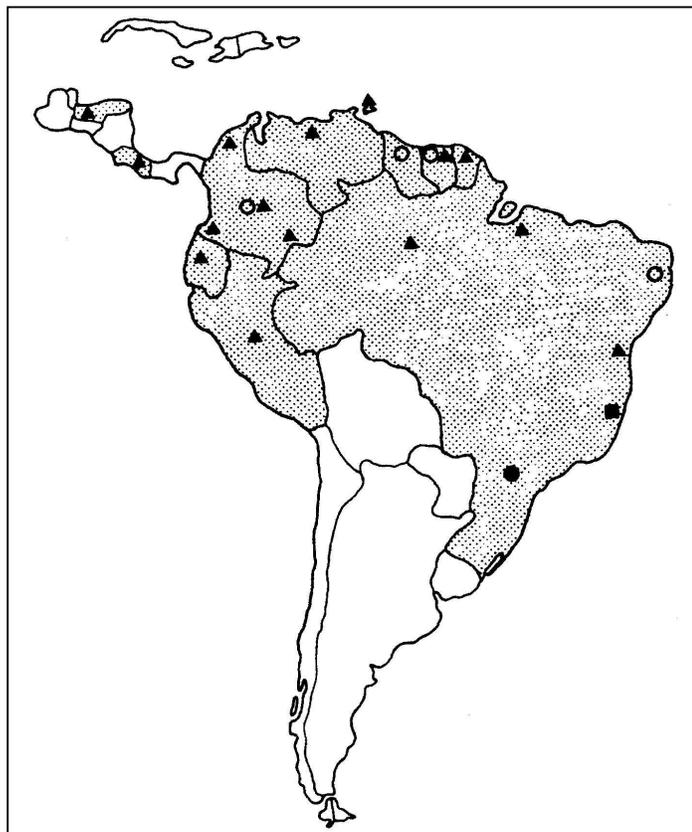


Figura 1.3 – Distribuição na América Latina das infecções causadas por flagelados em café (○), coco e palmeiras (▲), mandioca (■) e tomates (●) (Jankevicius *et al.*, 1989).

Há muitos problemas em relação à taxonomia e identificação de *Phytomonas*, pois morfologicamente não é possível distinguir tripanossomatídeos de plantas dos de insetos. Talvez seja por essa razão que os dois grupos foram inicialmente classificados como tripanossomatídeos inferiores, conceito que está em aparente desacordo com a atual classificação filogenética destes organismos (Hollar *et al.*, 1997).

Insetos não-hematófagos podem abrigar em seus tratos digestivos cinco gêneros de tripanossomatídeos: *Blastocrithidia*, *Crithidia*, *Leptomonas*, *Herpetomonas* e *Phytomonas*. Estes gêneros têm sido definidos de acordo com a morfologia das formas que ocorrem em seus ciclos de vida. Espécies dos dois primeiros gêneros podem ser relativamente fáceis de identificar devido a sua peculiar morfologia de epimastigotas em *Blastocrithidia* e coanomastigotas em *Crithidia*, mas é difícil a identificação somente por essa característica, pois em certas condições de cultura, coanomastigotas podem ser confundidas com promastigotas. Espécies dos outros três gêneros ocorrem em insetos como promastigotas, o que faz a distinção entre eles muito difícil. *Herpetomonas* spp. exibe opitomastigotas além de promastigotas em seu ciclo de vida, um fato que pode facilitar sua identificação. Contudo, opistomatigotas não estão sempre presentes e podem aparecer somente sobre certas condições de cultura especiais (Teixeira *et al.*, 1989). Os organismos pertencentes ao gênero *Phytomonas* possuem características comuns a estes, o que dificulta sua identificação. Existem métodos que são utilizados para a identificação e alguns são específicos para *Phytomonas*, como aglutinação (Sanchez-Moreno *et al.*, 1995; Schottelius *et al.*, 1987), perfil isoenzimático (Guerrini *et al.*, 1992; Fernandez-Bezerra *et al.*, 1996, Uttaro *et al.*, 1997), métodos imunológicos como anticorpos monoclonais (Teixeira *et al.*, 1989; 1995), fingerprinting de DNA genômico ou cinetoplasto (Fernandez-Bezerra *et al.*, 1996; Camargo *et al.*, 1992; Ahomadegbe *et al.*, 1990), seqüenciamento de genes ribossomais (Marché *et al.*, 1995), hibridização com oligonucleotídeos para a seqüência de “spliced leader” presente no RNAm (Teixeira *et al.*, 1996), marcadores moleculares como as enzimas do metabolismo de ornitina-arginina (Camargo *et al.*, 1987), caracterização de componentes de superfície (Abreu Filho *et al.*, 2001) e superóxido desmutase (Marin *et al.*, 2004) entre outros. A utilização destas metodologias permitiu a separação destes organismos.

1.3 – *Phytomonas serpens*

Em 1957, Gibbs ao pesquisar a seiva de tomates – *Lycopersicum esculentum* – encontrou flagelados de extremo polimorfismo, com desenvolvimento no trato digestivo de insetos, com várias torções ao longo do corpo, com movimentos serpeados, entre outras características, que também infectavam o inseto *Nezara viridula*. Estes organismos foram prontamente denominados de *Leptomonas serpens*. Não foi tentado o cultivo, portanto *L. serpens* nunca ficou disponível. Mais tarde, Podlipaev referiu-se ao flagelado descrito por Gibbs como *Phytomonas serpens* (Gibbs, 1957; Jankevicius *et al.*, 1989)

Phytomonas serpens possui as mesmas características de outros organismos pertencentes ao gênero *Phytomonas*. Possui núcleo, flagelo, cinetoplasto e glicossomos espalhados em seu corpo celular. Ao contrário das outras espécies de *Phytomonas* que são encontradas no látex e seiva de plantas, *Phytomonas serpens* é de fácil isolamento e cultivo, tornando-o um excelente modelo para estudo.

Uma característica interessante, porém não citada anteriormente, é a diferença de metabolismo que *Phytomonas* possui em relação aos outros tripanosomatídeos. Estudos realizados com *P. serpens* demonstraram que várias funções mitocondriais estavam ausentes no genoma do cinetoplasto, como respiração mediada por citocromo, produção de ATP e ciclo de Krebs, sendo que a produção energética da célula é baseada na glicólise realizada nos glicossomos localizados no corpo celular destes organismos (Maslov *et al.*, 1999).

Um dos maiores alvos de estudo sobre *P. serpens* é o cinetoplasto. Essa organela é composta de maxicírculos, que possuem função genética similar ao DNA mitocondrial de outros eucariotos e minicírculos. Representa aproximadamente 10-15% do DNA total da célula. As seqüências dos minicírculos e sua estrutura das regiões GT foram caracterizadas, bem como a edição de RNA mensageiro (RNAm) nos maxicírculos do cinetoplasto. Essas informações foram utilizadas como marcadores moleculares e para estudos taxonômicos (Fernandez-Bezerra *et al.*, 1996; Sá-Carvalho *et al.*, 1993, 1995; Maslov *et al.*, 1998).

Outro estudo interessante realizado com *P. serpens* foi a investigação da existência de uma reação antigênica cruzada entre *Phytomonas serpens* e *Trypanosoma cruzi*, que foi realizado *in vivo* com camundongos BALB/c e com soro de pacientes chagásicos e de coelhos. Este estudo comprovou que há antígenos comuns em *T. cruzi*

e *P. serpens* que são reconhecidos pelo soro humano de pacientes chagásicos e também por soro de coelhos e camundongos infectados experimentalmente. Estes achados por si não são muito surpreendentes, pois estes protozoários pertencem à mesma família de tripanosomatídeos. Mas o que foi marcante neste estudo foi o grau de proteção contra a infecção de *T. cruzi* que camundongos BALB/c altamente suscetíveis obtiveram por imunização com *P. serpens* e que sua administração oral também induz uma considerável proteção (Breganó *et al.*, 2003). Outro ponto foi que a imunidade conferida oralmente é dependente da produção aumentada de óxido nítrico produzida durante a fase aguda da infecção por *T. cruzi* (Pinge-Filho *et al.*, 2005).

O ciclo de vida de *Phytomonas serpens* compreende planta e inseto, sendo originalmente isolado de tomates, mas isto não exclui a possibilidade de outras plantas o albergarem. A transmissão é feita do tomate para inseto e do inseto para tomate (Figura 1.4). Quando insetos se alimentam de tomates contaminados, se tornam infectados após 10 a 15 dias, apresentando parasitos nas fezes, urina, tubo digestivo e glândulas salivares. O hemíptero *Phthia picta* é um candidato a ser o vetor natural no Brasil, pois se alimenta de tomates, é frequentemente encontrado na natureza infectado com promastigotas positivas para os anticorpos monoclonais específicos de *Phytomonas* e é facilmente infectado com *P. serpens* em experimentos laboratoriais, sendo posteriormente capaz de transmitir a infecção para tomates não infectados. Porém, outros insetos que também se alimentam de tomates podem ser vetores do parasito, como o inseto *Nezara viridula*, que pode ser facilmente infectado em laboratório com protozoários mantidos em cultura e foi o primeiro inseto achado infectado com *P. serpens* por Gibbs. Então, considerando a ampla variedade de insetos que se alimentam de tomates no Brasil, e que estes insetos podem se alimentar alternativamente de outros frutos, *P. serpens* pode ser descoberta em outras plantas e insetos (Jankevicius *et al.*, 1989). Com o estudo do genoma deste parasita e com a geração de novos marcadores moleculares, futuramente poderemos ampliar nossas pesquisas e explorar esta possibilidade.

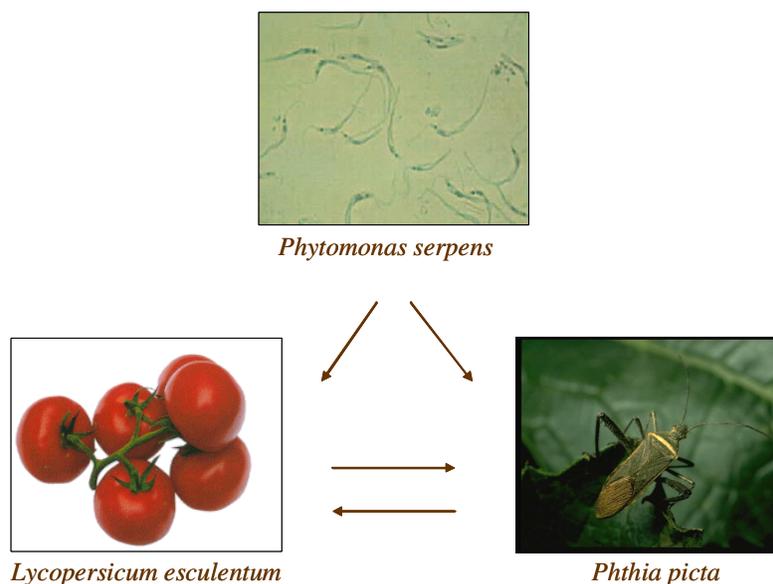


Figura 1.4 – Ciclo de vida de *Phytomonas serpens* (Fotografia por Drees).

Nada é conhecido ainda sobre a patogenicidade de *P. serpens*. Poucos estudos foram realizados abordando um possível efeito deletério dos parasitos sobre a longevidade dos tomates. Contudo, pode ser diferente a respeito das sementes e do ponto de vista comercial. Acredita-se que, em tomates, as picadas dos insetos infectados resultam em pontos circulares, amarelos e fáceis de notar na superfície das frutas. Contudo, a picada dos insetos não infectados também produz efeitos similares, mas com diferente magnitude.

Pouco se sabe sobre *Phytomonas serpens* e o conhecimento é ainda mais escasso se levarmos em conta todos os organismos dentro do gênero *Phytomonas*. Poucas pesquisas tem sido feitas sobre estes organismos. Ao fazer uma consulta no portal Entrez do NCBI (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>) no dia 07 de fevereiro de 2006, utilizando “*Phytomonas*” como palavra-chave obtivemos 118 entradas na sessão PubMed, 12 na sessão “Protein” e 829 em “nucleotide” sendo que esta categoria é dividida em três: “corenucleotide”, EST e GSS. Já em “Structure” e “genome”, nenhum item foi encontrado. Quando refinamos nossa pesquisa para a palavra-chave “*Phytomonas serpens*”, para a procura de informações sobre o organismo alvo de nosso estudo, encontramos 24 entradas no PubMed, 6 em “Protein”, 707 em “nucleotide”. Na tabela 1.1, podemos comparar os números de entradas dos três

principais tripanosomatídeos, *Trypanosoma cruzi*, *Leishmania major* e *Trypanosoma brucei*. Com isso, verificamos a pouca quantidade de informações sobre *Phytomonas serpens*, e até mesmo, sobre o gênero *Phytomonas*.

Tabela 1.1: Números de entradas no Entrez do NCBI nas categorias de Nucleotide, Protein e PubMed no dia 07 de fevereiro de 2006.

	<i>Phytomonas</i>	<i>Phytomonas serpens</i>	<i>Leishmania major</i>	<i>Trypanosoma cruzi</i>	<i>Trypanosoma brucei</i>
Corenucleotide	132	10	1820	88111	10486
EST	697	697	2191	13972	5133
GSS	0	0	18542	26684	91827
Protein	12	6	10262	41052	23339
PubMed	118	24	2196	7782	5606

Com o estudo sobre o genoma deste organismo, poderemos descobrir genes que facilitem sua identificação, filogenia e confirmem sua posição taxonômica, sem mencionar a ajuda em estudos comparativos com outros tripanosomatídeos, sendo o primeiro estudo das seqüências GSS (Genome Sequence Survey) de *Phytomonas serpens*.

1.4 – Descoberta de genes

Atualmente, com a introdução de técnicas de biologia molecular, o conhecimento sobre as espécies está aumentando a passos cada vez mais largos. Com o desenvolvimento da técnica de seqüenciamento em 1975, por Sanger e Barrel (1975) e com a automatização deste processo por Hood e Hunkapillar em 1991, obtivemos uma maior chance de conhecer melhor os organismos e aprofundar os nossos estudos, elevando o nível de compreensão das bases moleculares que formam os organismos, as diferenças e igualdades que os tornam únicos.

Com o mapeamento genético de *Trypanosoma cruzi* (Hanke *et al.*, 1996, Ferrari *et al.*, 1997, Degraeve *et al.*, 1997), *T. brucei*, *Giardia lamblia* (Smith *et al.*, 1998),

Leishmania major (Akopyants *et al.*, 2001), *Plasmodium sp.* (Carlton *et al.*, 2001) e *Leishmania braziliensis* (Laurentino *et al.*, 2004), informações sobre novos alvos para estudos em diagnóstico, quimioterapia e vacinas poderão ser identificados. É óbvia a necessidade de se estudar estes organismos, mas não podemos negar a importância de se pesquisar organismos não-patogênicos aos seres humanos e/ou organismos modelos, pois a descoberta de genes é também de grande relevância, tanto em estudos comparativos como na contribuição ao estudo da biologia dessas espécies. Este tem sido o caso de espécies como *Paramecium sp.* (Dessen *et al.*, 2001, Janssen *et al.*, 2001), *C. elegans* (Aziz & Blaxter 2003), *D. melanogaster* (Schaeffer *et al.*, 2003), entre outros.

Existem diferentes metodologias a serem aplicadas no seqüenciamento de genomas, desde o seqüenciamento de pedaços de DNA (GSS e WGSS - Whole-Genome Shotgun Sequencing) até o seqüenciamento das regiões transcritas a partir de RNAm, utilizando as técnicas de ESTs (Adams *et al.*, 1991) ou ORESTES (Nunes *et al.*, 2004; Passagia & Zaha 2003).

ESTs ou “expressed sequences tags” (etiquetas de seqüências expressas) é uma abordagem de seqüenciamento que gera seqüências pequenas derivadas do DNA complementar gerado a partir do RNAm. É uma metodologia direta de obtenção de seqüências gene-específicas de um grande número de moléculas e permite a demonstração do perfil de expressão de genes específicos em determinadas fases da célula. Segundo El-Sayed *et al.* (2000) os projetos de EST são atraentes devido ao relativo baixo investimento necessário quando comparado com os projetos de seqüenciamento de genomas inteiros, além disso, os projetos de EST são efetivos na obtenção de dados preliminares sobre regiões codificantes e uma boa ferramenta para a descoberta de genes, podendo proporcionar ampla variedade de marcadores para mapeamento STS (sequence-tagged-sites). Esta metodologia foi utilizada para a pesquisa de genes em *P. serpens* (Pappas *et al.*, 2005) e *Trypanosoma cruzi* (Porcel *et al.*, 2000).

Como se pode observar na tabela 1.1, o número de EST de *Phytomonas* é relativamente alto em relação ao número de GSS graças ao projeto de EST de *P. serpens* publicado no início de 2005 por Pappas (2005). Neste projeto foram achadas seqüências correspondentes à tradução, estrutura ribossomal e biosíntese.

Por outro lado, a metodologia GSS descrita por Peterson *et al.* (1991, 1993), tem como principal objetivo explorar o genoma do organismo, podendo identificar seqüências codificantes e não codificantes, distinguindo-se assim das ESTs, que visam

identificar os genes que estão sendo expressos em um determinado momento (Adams *et al.* 1991).

Devido à estratégia de GSS se basear na fragmentação do genoma inteiro para clonagem em plasmídeos (como por exemplo, pUC18) e no seqüenciamento dos clones obtidos, há uma tendência de identificar além de regiões que codificam para proteínas, diferentes regiões do genoma, entre elas: regiões intergênicas, íntrons, pseudogenes, etc. Os projetos GSS podem ser de pequena escala sendo chamados simplesmente de GSS, bem como de grande escala (para o seqüenciamento de genomas inteiros) neste caso sendo chamado de “Whole Genome Shotgun Sequencing – WGSS”. Da mesma maneira como na abordagem de ESTs, os GSS podem também identificar seqüências genômicas que codificam para proteínas que possam ser potencialmente usadas como alvos para quimioterapia e marcadores moleculares para diagnóstico específico. Esta metodologia foi utilizada para obter seqüências do genoma de *Leishmania braziliensis* (Laurentino *et al.*, 2004) e na exploração do genoma de *Trypanosoma vivax* (Guerreiro *et al.*, 2005).

O seqüenciamento utilizando a abordagem de GSS de *Trypanosoma vivax*, um protozoário hemoparasita que causa doença em ruminantes, demonstrou ser uma metodologia eficiente para a detecção de genes e caracterização do genoma. Neste projeto, foram geradas 455 seqüências GSS de alta qualidade correspondendo a 135,5 kb que foram clusterizadas em 331 contigs ou GSS-nr (99,6 kb), representando uma boa fonte de potenciais marcadores moleculares para esta espécie de difícil cultivo. Estes resultados aumentaram em 14 vezes o número de entradas de *T. vivax* no GenBank (consulta feita em fevereiro de 2005)(Guerreiro *et al.*,2005), o que é bastante significativo.

A técnica de “Whole Genome Shotgun Sequencing” foi utilizada no seqüenciamento dos três principais tripanosomatídeos, os “Tri-Tryps”, *T. cruzi* (El-Sayed *et al.*, 2005), *T. brucei* (Berriman *et al.*, 2005) e *L. major* (Ivens *et al.*, 2005).

Outro exemplo do uso da técnica de “WGSS” foi no projeto de seqüenciamento do genoma ambiental de amostras do mar de Sargasso onde uma amostra da superfície da água foi colhida e dessa amostra foi recolhido um “pool” de DNA que foi seqüenciado gerando várias seqüências de DNA de organismos não conhecidos. Este projeto teve como objetivo verificar se esta abordagem pode ser aplicada na descoberta de genes e espécies e caracterização de amostras ambientais (Venter *et al.*, 2004).

Dávila *et al.* (2003a) e Parkinson *et al.* (2002) têm comentado que projetos genomas de pequena escala como os ESTs e GSS, podem apresentar um melhor custo-benefício que o estudo de genes individuais ou famílias gênicas.

O uso em conjunto destas duas metodologias (seqüenciamento por GSS e por EST) no estudo de um organismo gera resultados eficazes na identificação de íntrons, éxons, regiões intergênicas e processamento dos genes. Essas técnicas foram utilizadas em conjunto em um estudo do genoma de *T. cruzi* (Agüero *et al.*, 2000) e em uma comparação entre três espécies de *Plasmodium* – *P. berghei*, *P. vivax* e *P. falciparum* (Carlton *et al.*, 2001).

1.5 – Bioinformática e Anotação de Seqüências

Na década de 90, com a automatização do processo de seqüenciamento, houve uma explosão no número de seqüências a serem armazenadas e processadas, exigindo uma ampliação na capacidade computacional, tanto da mão de obra especializada e capacitada para a análise dos dados gerados, quanto da capacidade das máquinas e programas oferecidos. Com isso, evidenciou-se a necessidade de se criar uma área multidisciplinar, que envolveria diversas áreas de conhecimento como a engenharia de softwares, matemática, estatística, ciência da computação e biologia molecular, nascendo assim, a bioinformática.

Para se ter uma idéia do crescimento do número de seqüências geradas e armazenadas no GenBank, em 1982 haviam 606 seqüências, o que não é nada comparado com 40.604.319 seqüências armazenadas em 2004 (<http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html>).

Uma vez obtidos os dados com o seqüenciamento se faz necessário saber o que estes dados representam. Para isso, utilizamos o processo de anotação. Anotar uma seqüência genômica, nada mais é do que identificar esta seqüência. O processo de anotação pode ser definido como dar um significado a uma seqüência que poderia, de outra maneira, ser quase desprovida de informação. Pela identificação de regiões de interesse e pela definição de funções putativas para estas áreas, o genoma pode ser entendido e maiores pesquisas podem ser iniciadas (Meyer *et al.*, 2003).

O processo de anotação utiliza diversos tipos de programas, partindo da análise de qualidade das seqüências, montagem ou agrupamento das seqüências para reduzir a

redundância das análises como o pacote Phred/Phrap/Consed (Ewing *et al.*, 1998a; Ewing *et al.*, 1998b) e CAP3 (Huang and Madan, 1999), identificadores de genes como GLIMMER (Salzberg *et al.*, 1998), GeneMark (Borodovsky and McIninch, 1993) e YACOP (Tech & Merkl, 2003), busca de similaridades como os pacotes BLAST (Altschul *et al.* 1997), FASTA (Pearson, 1990), InterproScan (Quevillon *et al.*, 2005) e SAM (Karplus *et al.*, 2005), alinhamentos múltiplos de seqüências como CLUSTALW (Thompson *et al.* 1994) e T-COFFEE (Notredame *et al.*, 2000), análises filogenéticas como Phylip (Felsenstein, 2005) e MEGA (Kumar *et al.*, 2004), entre outros.

Há ainda algumas plataformas de anotação que procuram integrar alguns destes programas, visando facilitar as análises e anotação das seqüências, como por exemplo, Artemis (Rutherford *et al.*, 2000) e SABIA (System for Automated Bacterial Integrated Annotation)(Almeida *et al.*, 2004) e o GARSA (Genomic Analysis Resources for Sequences Annotation)(Dávila *et al.* 2005), programa que foi utilizado em nosso estudo.

Devido ao grande número de informações geradas por seqüências de nucleotídeos e proteínas, principalmente após a criação dos projetos transcriptoma, genoma e proteoma, houve a necessidade de se criar bancos de dados para armazenar estas seqüências e deixá-las a disposição para o uso. Os bancos de dados mais utilizados são GenBank, que possui os bancos de seqüências nucleotídicas (NT, dbEST, dbGSS, RefSeq) e protéicas (nr), EBI, DDBJ, PDB, GDB, TIGR, Swiss-prot, INTERPRO, KEGG bem como consórcios de seqüenciamento FlyBase (<http://www.flybase.org>), GeneDB (<http://www.genedb.org>), entre outros.

O presente trabalho é parte integrante do consórcio Biowebdb (<http://www.biowebdb.org>) que é uma iniciativa que integra o Instituto Oswaldo Cruz (FIOCRUZ), a Universidade Federal do Rio de Janeiro e a Universidade Federal de Santa Catarina, e financiado pelo CNPq. Este consórcio procura construir uma plataforma integrada, amigável e flexível capaz de lidar com diferentes bancos de dados e projetos genomas, sendo desenhado para prover uma interação entre diferentes ferramentas de análise e algoritimos sensíveis para a detecção de homologies distantes.

CAPÍTULO

2

OBJETIVOS

2.1 - Objetivo

Gerar seqüências do genoma de *Phytomonas serpens*, através da técnica de GSS, objetivando o aumento do conhecimento sobre este organismo e a análise comparativa com os dados do genoma de outros organismos.

2.2 – Específicos

- a) Construir uma biblioteca genômica de *Phytomonas serpens* em pUC18 contendo insertos de aproximadamente 1,0-3,0Kb;
- b) Gerar seqüências do genoma;
- c) Analisar as seqüências;
- d) Anotar os genes obtidos;
- e) Comparar as seqüências obtidas com genomas de outros organismos disponíveis no Genbank.

CAPÍTULO

3

MATERIAIS E MÉTODOS

3.1- Organismos

- *Phytomonas serpens*:

A cepa de *Phytomonas serpens* 9T (CT–IOC-174) foi gentilmente cedida pela coleção de Tripanosomatídeos do Instituto Oswaldo Cruz. A manutenção das células foi feita por sucessivas passagens em meio “Schneider’s insect” enriquecido com 10% de soro fetal bovino a 27°C.

- *Escherichia coli* cepa DH5 α :

As células foram preparadas para clonagem e mantidas em meio LB com 10% de glicerol em estoque na temperatura de -70⁰ C. Depois de transformadas foram cultivadas em meio LB contendo 100 μ g/ml de ampicilina a 37⁰ C.

3.2- Meios de cultura

Anexo I – Tabela 3.1: Meios de cultura utilizados neste estudo.

3.3- Tampões e soluções

Anexo II – Tabela 3.2: Soluções e tampões utilizados neste estudo.

3.4- Antibiótico

Ampicilina

A solução estoque foi preparada na concentração de 100 mg/ml em água milliQ e esterilizada por filtração. Alíquotas de 1 ml foram armazenadas a -20°C. A concentração final foi de 100 μ g/ml.

3.5- Marcadores de peso molecular

Foram feitas diluições e alíquotas dos seguintes marcadores de peso molecular:

- **Lambda-HindIII** (New England Biolabs, EUA): DNA de fago lambda digerido com a enzima de restrição HindIII. Tamanho dos fragmentos: 23.130; 9.419; 6.557; 4.361; 2.322; 2.027; 564 e 125 pb.

- **ΦX174 RF/HaeIII** (New England Biolabs): DNA de fago ΦX174 digerido com a enzima de restrição Hae III. Tamanhos dos fragmentos: 1.353; 1.078; 872; 603; 310; 281; 271; 234; 194; 118 e 72 pb.

- **Marcador de peso molecular 100 pb** (Life Technologies, EUA): consiste de 15 fragmentos de finais não coesivos de tamanhos múltiplos de 100 pb, entre 100 pb e 1.500 pb, contendo um fragmento adicional de 2,072 pb.

3.6- Extração do DNA genômico pelo método Formol/clorofórmio

Para a obtenção do precipitado para extração de DNA genômico do parasita, culturas com aproximadamente 10^9 parasitas foram centrifugadas por 10 min, 4500 rpm à 4°C na centrífuga Labofuge 400R. O precipitado foi lavado duas vezes com PBS (pH 7,2) gelado. O precipitado foi ressuspensão com tampão de extração (4x o volume do pellet) e com 100 µg/ml de proteinase K. A mistura foi incubada a 42°C até a digestão completa do material. Após este processo, adicionou-se igual volume de fenol equilibrado, misturou-se por inversão por 10 min e centrifugou-se em temperatura ambiente por 10 min, a 14000 rpm. A fase aquosa foi transferida para um novo tubo de 1,5 ml e adicionou-se o mesmo volume de uma solução de fenol-clorofórmio (1:1). Misturou-se por inversão por 10 min e centrifugou-se nas mesmas condições anteriores. A fase líquida foi transferida para um novo tubo de 1,5 ml e adicionou-se o mesmo volume de clorofórmio. Misturou-se por 10 min por inversão e centrifugou-se nas mesmas condições anteriores. A fase aquosa foi transferida para um tubo de 1,5 ml limpo e adicionou-se o volume de 1/10 de uma solução 3M de acetato de sódio (pH 5,2) e 2,5X o volume total de etanol 100% gelado. Incubou-se a -70°C por 40 min e centrifugou-se a 14000g, em temperatura ambiente por 30 min. O sobrenadante foi descartado e lavou-se o precipitado duas vezes com etanol 70%. O precipitado foi ressuspensão com 50 µl de TE.

3.7- Digestão com enzimas de restrição

3.7a- Digestão parcial do DNA genômico de *Phytomonas serpens* com a enzima de restrição Sau3A:

Foram digeridos parcialmente o DNA genômico de *Phytomonas serpens* com a enzima de restrição Sau3A (New England Biolabs®). Variaram as concentrações tanto de DNA genômico quanto da enzima de restrição, bem como o tempo das reações.

Tabela 3.3: Concentrações de DNA genômico e da enzima Sau3A e os tempos utilizados nas reações de digestão.

<i>Concentração de DNA genômico</i>	<i>Concentração da Enzima Sau3A</i>	<i>Tempo</i>
~10 µg/µl	6 U	2, 4, 6, 8 e 10 seg
~10 µg/µl	3 U	2, 4, 6, 8 e 10 seg
~10µg/µl	1 U	5, 15 e 45 min
1 µg/µl	0,5 U	30, 90 e 150 min

3.7b- Digestão do vetor pUC18 com a enzima de restrição BamHI:

O vetor de clonagem pUC18 (Sigma®) foi digerido com a enzima BamHI (New England Biolabs®). Aproximadamente 100 ng do plasmídeo foram digeridos com 1 U de BamHI por duas horas a 37°C. A reação foi inativada por 15 min a 65°C. Após a digestão, o vetor foi defosforilado com Fosfatase alcalina (Promega®) por 30 min a 37°C e esta reação foi inativada igualmente à anterior.

Sítios de reconhecimento das enzimas de restrição	
BamHI	5'...G↑GATCC...3' 3'...CCTAG↑G...5'
Sau3A	5'...↑GATC...3'

	3'... CTAG↑...5'
--	------------------

3.8- Eletroforese de DNA

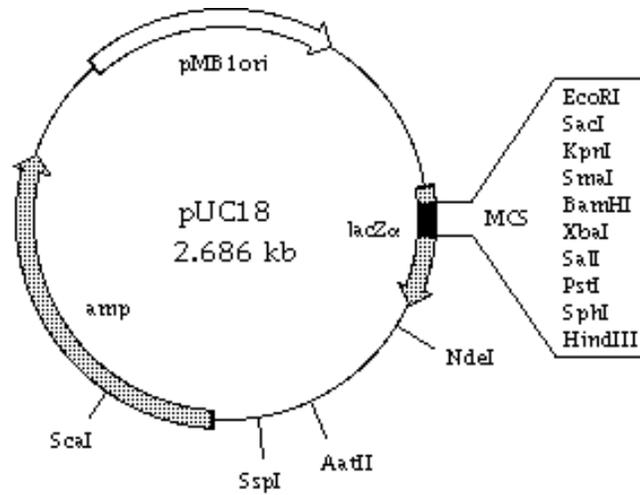
A eletroforese de DNA foi realizada em gel de agarose de 0,8% a 1% em TBE 1x contendo cerca de 0,1 µg/mL de brometo de etídeo. A diferença de potencial usualmente aplicada foi de 90-100V durante cerca de 1h. As imagens dos géis foram visualizadas e registradas em equipamento UVP *Bioimaging Systems*, com o programa *LabWorks™ Image Acquisition and Analysis*.

3.9- Extração de DNA de gel de agarose

Fragmentos de DNA de *Phytomonas serpens* nos tamanhos de 1-3 kb foram purificados do gel de agarose utilizando o kit de purificação “QIAquick™ Gel Extraction Kit” da Qiagen (EUA), de acordo com orientações do fabricante.

3.10- Clonagem

Os fragmentos de DNA obtidos na digestão por enzima de restrição foram extraídos do gel de agarose e então submetidos à reação de ligação no sítio BamHI do vetor de clonagem pUC18, como descrito por Sambrook & Russel, 2001. Foi utilizada para a ligação a enzima T4 DNA ligase (Fermentas®).



M13/pUC sequencing primer (-23), 17-mer

399 HindIII PstI SmaI BspI HindI SalI XbaI DriI Eco81 SmaI Acc65I KpnI Eco136II Eco24I SacI EcoRI XbaI 455

5' G TAA AAC GAC GGC CAG TGC CAA GCT TGC ATG CCT GCA GGT CGA CTC TAG AGG ATC CCC GGG TAC CGA GCT CGA ATT CGT AAT CAT GGT CAT AGC TGT TTC CTG 3'

3' C ATT TTG CTG CCG GTC ACG GTT CGA ACG TAC GGA CGT CCA GCT GAG ATC TCC TAG GGG CCC ATG GCT CGA GCT TAA GCA TTA GTA CCA GCA TCG ACA AAG GAC 5'

LacZ ← Val Val Ala Leu Ala Leu Ser Ala His Arg Cys Thr Ser Glu Leu Pro Asp Gly Pro Val Ser Ser Ser Ser Asn Thr Ile Met Thr Met

M13/pUC reverse sequencing primer (-28), 17-mer

Figura 3.1- Esquema do vetor de clonagem pUC18.

3.11- Transformação de bactérias

Em tubos de 1,5 ml foi colocado 100-200 µl de células competentes DH5α e 10-100 ng de DNA plasmidial. As células foram incubadas em gelo por 30 min, colocadas a 42°C para um choque térmico por 42 seg e devolvidas ao gelo por mais 2 min. Após adição de meio LB, foram colocadas por 60 min a 37°C com agitação (cerca de 200 rpm) para o crescimento. As células foram então plaqueadas em LB-agar com ampicilina (100 µg/ml), X-gal e IPTG para a seleção dos clones positivos e crescidas a 37°C por 12-16 horas. As colônias brancas, positivas para inserto foram selecionadas e congeladas em solução de glicerol (50%) e meio de cultura LB no volume de 1:1 a -70°C até o processamento.

3.12- Mini-preparação plasmidial

3.12a- Mini-preparação pelo método de fervura:

As bactérias foram colocadas para crescer em 3 ml de meio LB durante a noite à 37°C com agitação de 200 rpm. Após este período, foram centrifugadas em tubos de 1,5 ml (2x) por 2 min a 14000 rpm e ressuspensas em 400 µl da solução de STET contendo 1mg/ml de lisozima e 0,2 µl de RNase a 10 mg/ml, vortexadas, e fervidas 3 min. O lisado foi centrifugado por 15 min e o sedimento removido com o auxílio de um palito. Foram adicionados 400 µl de isopropanol e o tubo foi centrifugado por 15 min à temperatura ambiente. O sobrenadante foi descartado e o precipitado lavado com etanol 70% e ressuspensão em 50 µl de TE ou água MilliQ. Após a extração, digeriu-se o DNA plasmidial com as enzimas de restrição EcoRI (Promega®) e PstI (Promega®) para verificar a presença dos insertos de algumas amostras.

3.12b- Mini-preparação em placas de 96 poços:

Foram colocados em uma microplaca “deepwell” de 96 poços 1ml de meio LB contendo 100 µg/ml. Colônias individuais (uma colônia por poço) foram inoculadas com o auxílio do replicador. A placa foi selada com adesivo e colocada para crescer à 37°C, 200 rpm por aproximadamente 22 horas. A placa “deepwell” foi centrifugada por 6 min, 4000 rpm, a 20°C, para sedimentar as células. O sobrenadante foi descartado invertendo a placa e o excesso foi removido deixando a placa invertida sobre um papel absorvente. Foi adicionado 240 µl de GET, a placa foi selada com adesivo e agitada no

vortex por 2 min para ressuspender bem as células. Foi centrifugada por 9 min, 4000 rpm, à 20°C. Foi descartado o sobrenadante e a placa foi deixada invertida em papel absorvente por 5 min. Foi adicionado a cada poço, 80 µl de GET mais 3 µl de RNase a 10 mg/ml e a placa foi selada com um novo adesivo e agitada no vórtex por 2 min para ressuspender as células. Transferida a suspensão de células para a microplaca de 250 µl de polipropileno de fundo redondo, foi adicionado a cada poço 80 µl de NaOH 0,2N/SDS1%. A placa foi selada com adesivo, misturada 30 vezes por inversão, centrifugada rapidamente e incubada por 10 min, à temperatura ambiente. Foi adicionado a cada poço 80 µl de KOAc 3M. A placa foi selada com um novo adesivo, misturada 30 vezes por inversão, incubada por 10 min à temperatura ambiente e centrifugada rapidamente. O adesivo foi removido e a placa foi incubada em estufa à 90°C por exatos 30 min. Esfriamos a placa em gelo por 10 min e foi centrifugada por 9 min, 4000 rpm, 20°C. À parte, foi fixada com fita adesiva, uma placa Millipore (MAGV N22) no topo de uma microplaca de fundo em “V” de 250 µl de polipropileno. Verificado se os poços estavam alinhados, então, foi transferido todo o volume do sobrenadante, com extremo cuidado de não ser transferido junto os restos celulares, para a placa Millipore e as placas foram centrifugadas sem a tampa por 6 min, 4000 rpm, 20°C ou até todo o volume descer para a placa em baixo. Removida a placa Millipore, adicionaram-se 100 µl de isopropanol gelado ao filtrado que estava na microplaca de fundo em “V”. A placa foi muito bem selada com adesivo, misturada 30 vezes por inversão e centrifugada rapidamente. Foi incubada a temperatura ambiente por 15 min e centrifugada por 45 min, 4000 rpm, 20°C. Descartado o sobrenadante, adicionou-se 200 µl de etanol 70% gelado e mais uma vez centrifugado por 5 min, 4000 rpm à 20°C. Removido o sobrenadante, a placa foi invertida em papel absorvente e centrifugada rapidamente. A placa foi deixada para secar, à temperatura ambiente por 60 min coberta com papel toalha. O DNA foi ressuspensão com 40 µl de água MilliQ, coberto com adesivo e deixado à temperatura ambiente por aproximadamente 16 horas. Foi armazenado à -20°C.

3.13- Seqüenciamento

Os clones de DNA plasmidial escolhidos para serem seqüenciados que haviam sido obtidos por minipreparação por fervura, foram precipitados por PEG para maior purificação. Este protocolo foi realizado segundo Sambrook & Russel, 200, descrito acima.

As amostras de DNA plasmidial purificadas foram seqüenciadas utilizando o kit BigDye Terminator v3.1, Applied Biosystems, segundo instruções do fabricante. A reação de seqüenciamento foi realizada na plataforma PDTIS/FIOCRUZ. O seqüenciador automático capilar utilizado foi ABI 3730. Também utilizamos a plataforma de seqüenciamento Megabase situada no Laboratório de Genoma situado na Universidade do Rio de Janeiro (UERJ) sob a responsabilidade do Doutor Rodolpho Mattos Albano. As reações de seqüenciamento do Megabase foram realizadas segundo as instruções do fabricante (Tabela 3.4 e 3.5). A quantificação do DNA para a reação de seqüenciamento foi feita visualmente com base na intensidade observada em gel de agarose.

Tabela 3.4: Concentrações de reagentes utilizados nas reações de seqüenciamento.

<i>Reação de Seqüenciamento</i>	<i>PDTIS/FIOCRUZ</i>	<i>Megabase</i>
DNA	100-300 ng	200-600 ng
Iniciador M13 Universal (forward e Reverse)	3,2 pmoles	3,2 pmoles
Big Dye 3.1	1µl	-
Tampão 5X	1X	-
Mix de terminação Megabase	-	4µl
H ₂ O (q.s.p.)	10µl	10µl

Tabela 3.5: Condições das reações de sequenciamento.

<i>Plataforma de Seqüenciamento</i>	<i>Condições</i>	<i>Número de ciclos</i>
PDTIS/FIOCRUZ	96°C - 10 seg 50°C - 5 seg 60°C - 4 min	40 ciclos
Megabase Primer M13 Universal Forward	95°C - 20 seg 50°C - 15 seg 60°C - 1 min	25 ciclos
Megabase Primer M13 Universal Reverse	95°C - 20 seg 48°C - 15 seg 60°C - 1 min	27 ciclos

3.14- Iniciadores utilizados para a seqüenciamento dos clones (pUC 18)

M13 Universal Forward 5'- gta aaa cga cgg cca gt – 3'

M13 Universal Reverse 5'- cag gaa aca gct atg ac – 3'

3.15- Precipitação da reação de seqüenciamento

Foi colocado em cada poço 40 µl de isopropanol 65% a temperatura ambiente, incubado ao abrigo da luz por 15 min e centrifugado durante 45 min, 4000 rpm em temperatura ambiente. O sobrenadante foi descartado e a placa foi centrifugada rapidamente, invertida e apoiada em papel absorvente para retirar o excesso. Foi colocado em cada poço 80 µl de etanol 60% e centrifugado durante 15 min a 4000 rpm. O sobrenadante foi descartado e a placa foi centrifugada rapidamente, invertida e apoiada em papel absorvente para retirar o excesso. A placa foi deixada em temperatura ambiente para secar por aproximadamente 1 hora e o DNA foi ressuspensão com 10 µl de formamida Applied Biosystems. A reação foi desnaturada no termociclador por 3 min a 95°C e enviada para a leitura no sequenciador.

3.16- Análise de seqüências

Para análise das seqüências foi utilizado o sistema de anotação GARSA (Genomic Analysis Resources of Sequence Annotation)(Dávila *et al.*, 2005). Este programa oferece uma interface amigável, conferindo uma maior agilidade na análise das seqüências.

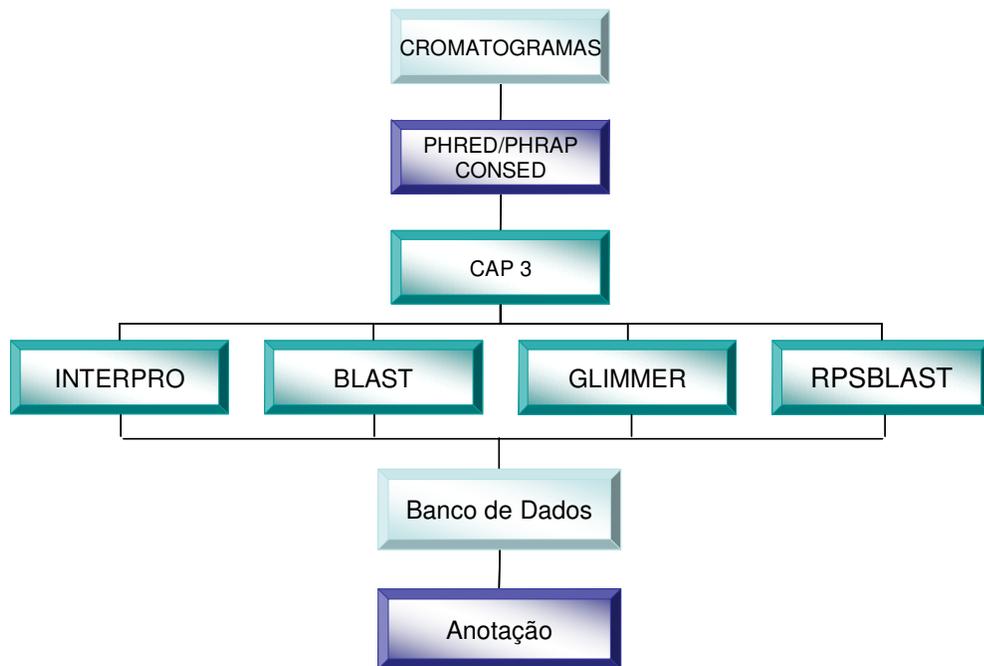


Figura 3.2: Organograma do processo de anotação do programa GARSA.

Os cromatogramas gerados pelo seqüenciador foram avaliados por sua qualidade e para a retirada de vetores pelo pacote PHRED/PHRAP/CONSED (Ewing *et al.*, 1998, Ewing and Green, 1998) (<http://www.phrap.org/>), sendo que, apenas as seqüências com qualidade PHRED acima do valor 20 foram aceitas. As seqüências que foram aprovadas foram agrupadas pelo programa Cap3 (Huang and Madan, 1999) (<http://genome.cs.mtu.edu/cap/cap3.html>).

Após os processos acima, utilizamos rotinas do programa BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1997) para busca de similaridade com organismos encontrados nos bancos de dados citados na tabela 3.6. As rotinas utilizadas foram BLASTn, que procura similaridade entre a seqüência nucleotídica contra um banco de dados de nucleotídeos; BLASTx, que busca similaridade entre a seqüência protéica contra um banco de dados de proteínas e tBLASTx, que traduz a seqüência nucleotídica e compara com um banco de dados traduzido. Também utilizamos o RPSBALST e o InterPro para a procura de domínios contra bancos de dados curados de proteínas (Tabela 3.6).

Nas análises com o programa BLAST, não foram consideradas as seqüências que apresentaram similaridade abaixo do limite estipulado de E-value $> 10^{-05}$.

Para estimar o conteúdo de G+C, utilizamos o programa geecee (<http://emboss.sourceforge.net/apps/geecee.html>) incluso no pacote emboss.

Para a predição de genes, utilizamos os programas GLIMMER2 v.2.13 (Gene Locator and Interpolated Markov Modeler)(Delcher *et al.*, 1999). O GLIMMER2 foi treinado com cds completos de tripanosomatídeos, evitando os hipotéticos e os de função desconhecida, tendo como parâmetro o comprimento mínimo de ORF de 150 nucleotídeos e tamanho mínimo de sobreposição da ORF de 20 nucleotídeos.

Para pesquisas de genes tRNA nas seqüências obtidas foi utilizado o programa tRNAscan-SE (Lowe & Eddy, 1997).

Para a validação dos cds utilizamos os resultados das rotinas do programa BLAST e o vocabulário de ontologias GO (Gene Ontology) (Ashburner *et al.*, 2000) (www.geneontology.org), fornecendo um vocabulário controlado para a descrição de genes e seus produtos, estabelecendo assim, descrições mais consistentes. O projeto GO fornece um vocabulário controlado para a descrição de genes e seus produtos em qualquer organismo, estabelecendo descrições mais consistentes em banco de dados diferentes. Este projeto começou com a colaboração entre três bancos de dados de organismos modelos: FlyBase (*Drosophila*), Saccaromyces Genome database (SGD) e o Mouse Genome Database (MGD). Desde então, o consórcio GO tem crescido para incluir vários bancos de dados, incluindo os maiores repositores de genomas de plantas, animais e micróbios.

O GO se baseia em três vocabulários estruturados, controlados (ontologias) que descrevem de maneira independente de espécies os genes, em termos associados ao

processo biológico, componente celular e função molecular. Abaixo uma breve descrição de cada termo.

A função molecular descreve atividades, como atividades catalíticas ou de ligação, em um nível molecular. Os termos do GO de função molecular representam as moléculas ou complexos que realizam as ações, mas não especifica onde ou quando e em que contexto a ação ocorre. O processo biológico é uma série de eventos realizada por um ou mais conjuntos ordenados de funções moleculares. Exemplos amplos de processos biológicos são os processos fisiológicos da célula ou transdução de sinais. A diferença entre processo biológico e função molecular é que o processo biológico precisa de mais de um passo distinto. Componente celular é como o nome descreve, mas com a condição de que seja parte de um objeto maior, que pode ser uma estrutura anatômica ou um grupo de produto de genes.

Para melhorar nossas análises, validar e aumentar o conhecimento obtido sobre *Phytomonas serpens*, incorporamos em nossos dados as seqüências que foram depositadas no banco do NCBI do sequenciamento de ESTs realizado por Pappas *et al.* (2005). Estas seqüências foram submetidas ao programa de anotação GARSA, passando pelas mesmas etapas que as seqüências GSS obtidas neste estudo.

Como as seqüências de EST obtidas no GenBank não foram analisadas previamente com o vocabulário do GO, as classificamos de acordo com as ontologias.

Foi realizada uma busca por similaridade usando o banco Taxonomy do NCBI (<http://www.ncbi.nlm.nih.gov/Taxonomy/>) para avaliar a similaridade dos clusters de GSS e EST com os organismos presentes neste banco.

As seqüências que deram similaridade significativa com os bancos de dados de aminoácidos de quando comparadas com as seqüências de GSS e EST de *P. serpens* foram utilizadas para um estudo filogenético. Os organismos utilizados foram os pertencentes aos TriTryps- *Leishmania major*, *Trypanosoma cruzi* e *Trypanosoma brucei*, alguns dos organismos pertencentes ao gênero Kinetoplastida excluindo os TriTryps, como *Phytomonas sp*, *Crithidia fasciculata*, *Leptomonas sp* entre outros. Também utilizamos seqüências do genoma ambiental do mar de Sargasso (Venter *et al.*, 2004) e seqüências de Euglenozooas.

Árvores individuais foram feitas utilizando as seqüências correspondentes aos genes de serina-treonina kinase, fator de alongamento, gliceraldeído-3-fosfato desidrogenase, fosfoglicerato kinase, frutose-1,6-bifosfato aldolase, HSP70, P450 redutase, adenilato ciclase e beta-tubulina. Também foi feita uma árvore filogenética

concatenada, utilizando genes encontrados em nosso estudo. Utilizamos os genes de fator de alongamento, histona H2, HSP70, tubulina, P450 reductase e adenilato ciclase. Os organismos escolhidos foram *Leishmania major*, *Trypanosoma cruzi*, *Trypanosoma brucei*, *Leishmania donovani* e *Euglena gracilis*.

Somente utilizamos neste estudo as seqüências que possuíssem os respectivos domínios destas proteínas. As consultas foram feitas utilizando os dados do banco Cdd do NCBI.

O alinhamento foi obtido com o programa Probcons (Probabilistic Consistency-based Multiple Alignment of Amino Acid Sequences) (Do, 2005), que roda exclusivamente em servidores Linux e mostrou ser o melhor método para nosso estudo em comparação aos outros métodos de alinhamento.

Para a inferência filogenética utilizamos o método de distância Neighbor Joining que utiliza os mesmos princípios que o método de evolução mínima. É um método rápido que tem como princípio achar pares de unidades taxonômicas operacionais (OTUs – operational taxonomic units) que minimizam o comprimento total do braço a cada estágio de clusterização das OTUs (Saitou & Nei, 1987).

Uma árvore filogenética não pode ser interpretada de maneira confiável sem um suporte estatístico para as relações evolucionárias apresentadas. Isto pode ser conseguido com o método de Bootstrap, no qual envolve uma “reamostragem” do dado para determinar a porcentagem de árvores replicadas, confirmando seus relacionamentos apresentados (Stevens *et al.*, 1999). Em nosso estudo utilizamos Bootstrap no valor de 10000.

As árvores foram obtidas com o programa MEGA 3.1 (Molecular Evolutionary Genetic Analysis) (<http://www.megasoftware.net>).

CAPÍTULO

4

RESULTADOS

4.1 – Digestão do DNA genômico de *Phytomonas serpens*

O DNA genômico de *P. serpens* foi digerido com a enzima de restrição *Sau3A* de acordo com a descrição do item 3.7, com variações no tempo de reação e nas concentrações do DNA e da enzima. Nas figuras 4.1 a 4.2 podemos observar os resultados das diferentes reações. Os fragmentos de DNA nos tamanhos entre 1 a 3 kb de todos os géis foram extraídos e ligados com o vetor de clonagem pUC18.

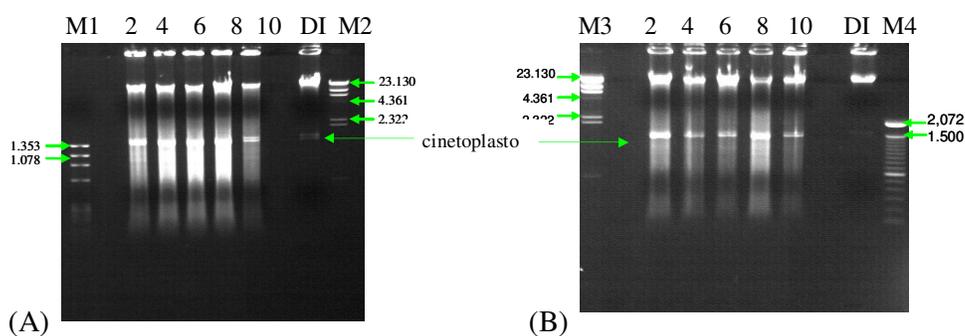


Fig. 4.1: DNA genômico de *P. serpens* na concentração aproximada de 5 µg/µl digerido com 6U (A) e 3U (B) da enzima *Sau3A* nos tempos de 2, 4, 6 8 e 10 segundos. Legenda: M1 - marcador de peso molecular ΦX174; 2, 4, 6, 8 e 10 - DNA genômico digerido nos tempos de 2, 4, 6, 8 e 10 segundos; DI – DNA genômico não digerido; M2 e M3 - marcador de peso molecular λHindIII; M4 - marcador de peso molecular 100 pb. Gel de agarose 1% corado com brometo de etídio.

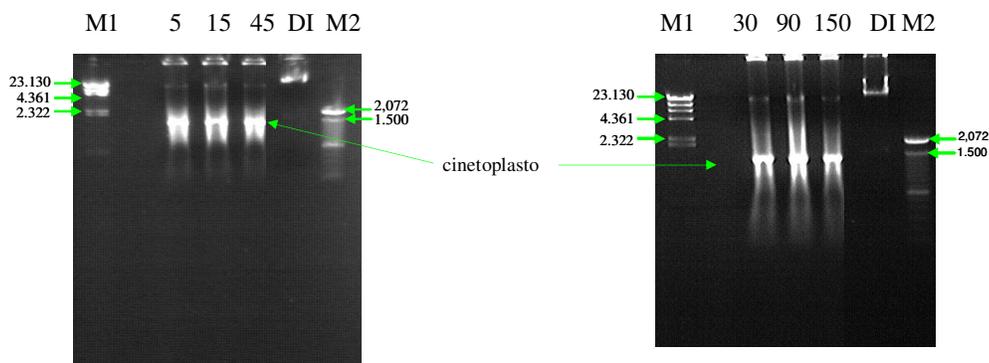


Fig. 4.2: DNA genômico de *P. serpens* na concentração aproximada de 5 $\mu\text{g}/\mu\text{l}$ (A) e de 1 $\mu\text{g}/\mu\text{l}$ (B) digerido com 1U (A) e 0,5U (B) da enzima *Sau3A* nos tempos de 5, 15 e 45 minutos (A) e 30, 90 e 150 minutos (B). Legenda: M1 - marcador de peso molecular $\lambda\text{HindIII}$; 5, 15 e 45 - DNA genômico digerido nos tempos de 5, 15 e 45 minutos (A); 30, 90 e 150 - DNA genômico digerido nos tempos de 30, 90 e 150 minutos (B); DI - DNA genômico não digerido; M2 - marcador de peso molecular 100 pb. Gel de agarose 1% corado com brometo de etídio.

A presença e o tamanho dos insertos foram confirmados pela digestão com as enzimas de restrição *EcoRI* e *PstI*, como descrito anteriormente em 3.12a (Fig. 4.3). Foram digeridas 72 amostras, sendo 53 (73,6%) positivas para o inserto.

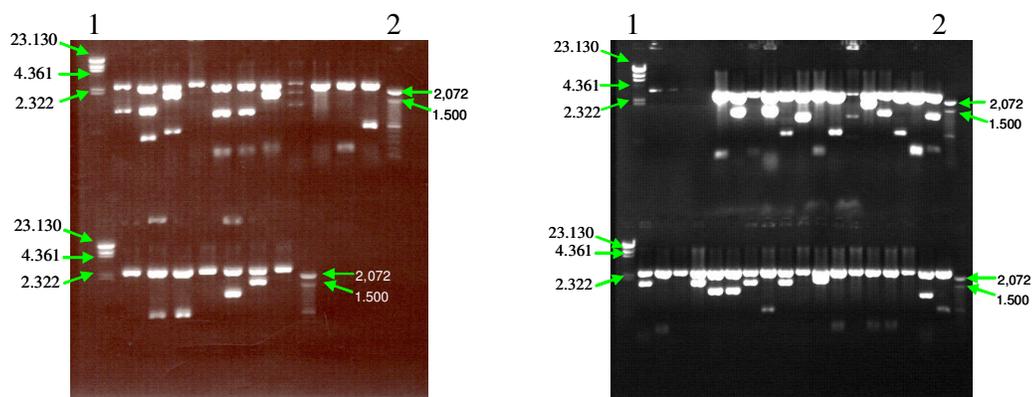


Figura 4.3: DNA plasmidial digerido com as enzimas *EcoRI* e *PstI*. Legenda: 1 - marcador de peso molecular $\lambda\text{HindIII}$; 2 - marcador de peso molecular 100 pb. Gel de agarose 1% corado com brometo de etídio.

4.2 – Seqüenciamento

Obtivemos um total de 326 clones que foram seqüenciados e 112 re-seqüenciados utilizando iniciadores “forward” e “reverse” gerando 829 seqüências. Após a utilização do pacote PRHRED/PHRAP/CONSED para avaliação de qualidade, obtivemos 379 seqüências (reads) de alta qualidade. Após a clusterização obtivemos 221 clusters, sendo 145 seqüências únicas (singlets) e 76 formados por mais de um read (Tabela 4.2).

Do total de seqüências geradas, 450 foram descartadas por baixa qualidade, tendo como motivo a presença do vetor, tamanho insuficiente dos reads e baixa qualidade dos cromatogramas. Essas seqüências não foram incluídas em nosso estudo.

O tamanho dos clusters obtidos da biblioteca genômica variou de 101pb (PSADSAU001H12.g), formado por 1 read, a 1079 pb (PSADSAU005G03.b), formado por 4 reads. O número de reads por clusters variou de 1 a 20 reads (tabela 4.1). O tamanho médio dos clusters foi de 406 pb.

Tabela 4.1: Total de clusters da biblioteca genômica classificados pelos números de reads pelos quais são formados.

Número de Reads	Total de Clusters
1	145
2	48
3	11
4	8
5	3
6	3
7	1
13	1
20	1

Foram obtidas no banco do NCBI 697 seqüências de EST de *P. serpens*. Após clusterização, foram gerados 686 clusters sendo 676 singlets e 21 formados por mais de uma seqüência. Dos clusters gerados, 676 foram formados por 1 read, 9 clusters por 2 reads e 1 cluster formado por 3 reads. O tamanho médio dos clusters foi de 508pb, sendo o menor encontrado com 143pb (PSADEST003C03.b) formado por uma

seqüência EST e o maior cluster com 1224pb (PSADEST007H05.b) também formado por uma única seqüência de EST.

Tabela 4.2: Número de GSS de DNA genômico de *Phytomonas serpens* geradas pelo sequenciamento e sua distribuição após a análise de qualidade e agrupamento.

	GSS
Clones seqüenciados	829
Seqüências de alta qualidade	379
Seqüências únicas	145
Clones em clusters	234
Clusters	76
Seqüências únicas + clusters	221

4.3 – Análises “in silico”

Todas as seqüências GSSs e ESTs foram submetidas às rotinas do programa BLAST contra os bancos de dados citados anteriormente na tabela 3.6. Em um total de 16 BLAST rodados, obtivemos 599 clusters com entradas (hits) positivas e 303 clusters com nenhum hit. Destes 303, 190 (62,7%) clusters são da biblioteca de ESTs e 113 (37,3%) da biblioteca GSS, totalizando 34% dos clusters.

O maior número de hits foi encontrado no banco de dados Lmajor_NCBI.fasta, com seqüências retiradas do NCBI derivadas do seqüenciamento completo do genoma de *L. major* liberado em julho de 2005 (Ivens *et al.*, 2005). Na tabela 4.3 podemos observar o número de clusters com hits em cada banco de dados.

Tabela 4.3: Número total de entradas (Hits) para cada banco de dados confrontados com as GSS geradas neste estudo e as seqüências de ESTs obtidas no GenBank.

Programa	Banco de dados	Hits	%	No Hits	%	Total
blastx	ecoli.aa	81	9%	821	91%	902
blastn	UniVec_Core	16	2%	886	98%	902
blastx	Lmajor_NCBI.fasta	541	60%	361	40%	902
rpsblast	Cdd	345	39%	530	61%	875
blastx	euglenozoa-not-kineto-aa.fasta	47	5%	855	95%	902
rpsblast	Pfam	311	36%	555	64%	866
rpsblast	Smart	69	10%	654	90%	723
rpsblast	Cog	269	32%	574	68%	843
rpsblast	Kog	371	42%	502	58%	873
blastx	Tbrucei_NCBI.fasta	482	53%	420	47%	902
blastx	Tcruzi_NCBI.fasta	505	56%	397	44%	902
blastx	kineto-non-tritryps.fasta	158	18%	744	82%	902
blastx	virulence-pathogenicity.fasta	2	0%	900	100%	902
blastx	uniprot_sprot.fasta	324	36%	578	64%	902
blastx	uniprot_trembl.fasta	461	51%	441	49%	902
blastx	env_nr	233	26%	669	74%	902

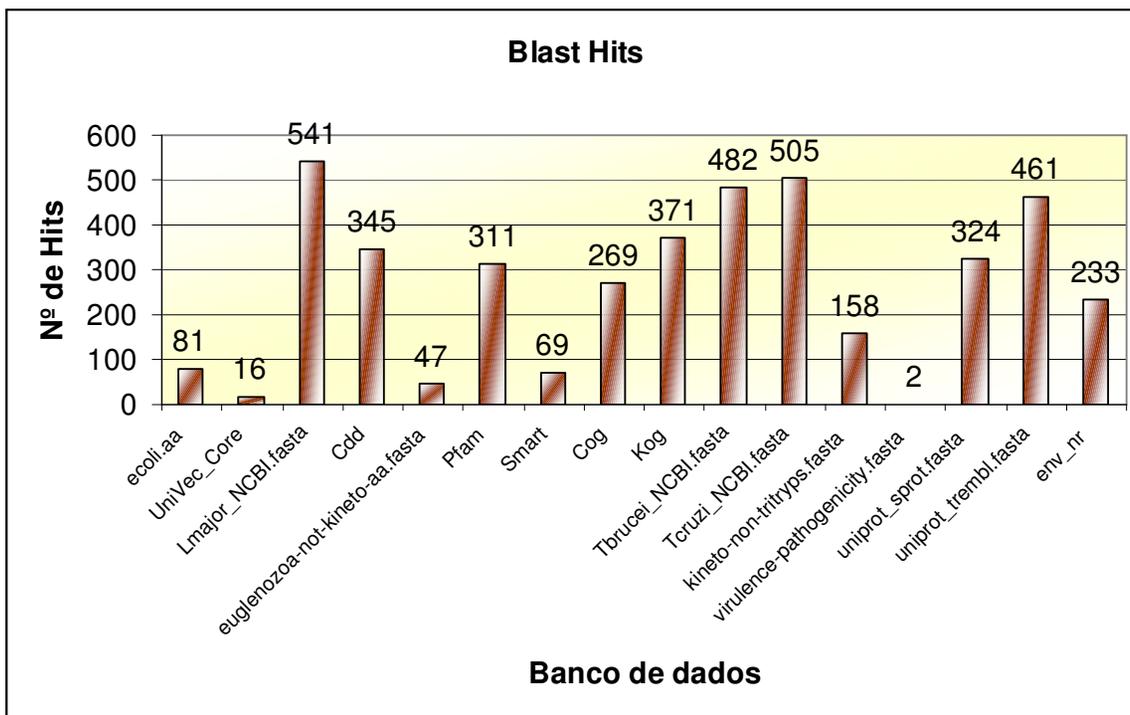


Gráfico 4.1: Representação gráfica da tabela 4.3 onde temos o número de entradas (Hits) para cada banco de dados confrontados com as GSS geradas neste estudo e as seqüências de ESTs obtidas no GenBank.

As análises feitas com o programa geecee para estimar o conteúdo G+C revelaram um conteúdo médio de 48% nas duas bibliotecas. Os valores limítrofes variaram de 77% para o cluster PSADSAU002A08.b (Proteína hipotética conservada) a 31% em PSADEST004F03.b (Proteína hipotética conservada).

As análises do GLIMMER revelaram 20 genes hipotéticos na biblioteca *Sau3A*. O tamanho dos genes variou de 156 (cluster PSADSAU005B11.b) a 516 pb (cluster PSADSAU002D10.b – proteína de membrana). O tamanho médio dos genes obtidos foi de aproximadamente 301 pb. O conteúdo de G+C dos 20 genes foi calculado, sendo a média obtida de 53%. O menor conteúdo foi de 37% no cluster PSADSAU007F10.b e o maior de 65% PSADSAU007H04.b (Tabela 4.4).

O cluster PSADSAU002D10.b é formado por uma sequência GSS e uma sequência EST, possuindo 56% de G+C, sendo anotado como proteína de membrana.

Tabela 4.4: Lista de genes hipotéticos encontrados nas seqüências de GSS da biblioteca *Sau3A* pelo programa Glimmer. Os clusters sem anotação foram os que não deram hit com o Blast.

Nome do Cluster	Conteúdo de GC	Anotação	Início do gene	Final do gene	Frame	Tamanho do gene
PSADSAU001A09.g	48%	Serina/treonina Kinase	287	51	-3	237
PSADSAU001F11.g	63%	RNAi sintetase	48	557	3	510
PSADSAU002D10.b	56%	Proteína de membrana	186	701	3	516
PSADSAU003F02.g	64%	Fator de splicing	23	523	2	501
PSADSAU003H09.g	63%	Proteína hipotética conservada	470	27	-3	444
PSADSAU005B11.b	38%	Proteína hipotética	375	220	-1	156
PSADSAU005F10.b	44%	Proteína hipotética	136	315	1	180
PSADSAU005G03.b	51%	P450 redutase	24	839	3	816
PSADSAU006B11.g	52%	Proteína hipotética	305	514	2	210
PSADSAU007A04.b	55%	Proteína hipotética conservada	271	17	-2	255
PSADSAU007C02.b	48%	Proteína hipotética	95	259	2	165
PSADSAU007C06.b	59%	Proteína hipotética conservada	516	16	-1	501
PSADSAU007F10.b	37%	Proteína hipotética	70	255	1	186
PSADSAU007H04.b	65%	Proteína hipotética	47	214	2	168
PSADSAU009A05.g	47%	Proteína hipotética conservada	83	373	2	291
PSADSAU009C04.g	52%	Proteína hipotética	364	116	-2	249
PSADSAU009E11.g	53%	Proteína hipotética conservada	236	45	-3	192
PSADSAU009H02.g	61%	Fator de alongamento	438	133	-1	306
PSADSAU010E01.g	60%	Proteína hipotética	438	181	-1	258
PSADSAU010E11.g	56%	Proteína hipotética conservada	42	233	3	192

Também foram encontrados genes hipotéticos nas seqüências ESTs, no total de 129 genes hipotéticos. O maior gene encontrado foi de 744 pb (PSADEST003B10.b- proteína hipotética conservada) enquanto que o menor foi de 165 pb (PSADEST001F11.b- proteína ribossomal L37). O conteúdo médio G+C também foi calculado, sendo aproximadamente 55%, apresentando os valores limítrofes de 67% (PSADEST006E11.b - proteína ribossomal S11) e 38% (PSADEST001F03.b - cisteína peptidase; PSADEST001F11.b - proteína ribossomal L37).

Nenhuma seqüência de RNAt foi encontrada nas seqüências GSS e EST pelas análises com o tRNAscan-SE.

Algumas das funções hipotéticas reveladas foram corroboradas “in silico” com análises realizadas usando os bancos de dados de domínios de proteínas CDD e Interpro. Foram encontradas 10 hits para os dois bancos, 9 hits somente para o CDD e 5 hits para Interpro.

Tabela 4.5: Lista de GSS de *P. serpens* que apresentaram similaridade significativa (evalue < 10⁻⁵) quando comparadas com os bancos de dados de domínios de proteínas CDD e Interpro.

Cluster	CDD		Description	Interpro	
	Acesso	E-value		Acesso	E-value
PSADSAU001F10.g	10297	9,0E-77	RNAt sintetase	IPR002314	9E-77
PSADSAU005D04.b	25767	2,0E-07	Proteína de Transporte de aa transmembrana	IPR002422	2E-7
PSADSAU005G03.b	10243	2,0E-51	P450 Redutase	IPR001433	2E-51
PSADSAU006A10.g	10851	7,0E-15	Transportador tipo-ABC multidrogas	IPR006162	7E-15
PSADSAU006C11.g	10019	2,0E-46	Proteína nucleolar	IPR001678	2E-46
PSADSAU007H02.b	14378	2,0E-49	EF-1alpha (GTPase)	IPR000795	2E-49
PSADSAU008H02.b	14378	9,0E-53	EF-1alpha (GTPase)	IPR000795	9E-53
PSADSAU009F01.g	25950	3,0E-16	Antígeno de proliferação nuclear	IPR000730	3E-16
PSADSAU009F03.g	37	2,0E-19	Adenilate ciclase	IPR001054	2E-19
PSADSAU009H02.g	14378	5,0E-29	EF-1alpha (GTPase)	IPR004160	5E-29
PSADSAU001E12.g	10681	3,0E-13	Aminoacido permease		
PSADSAU001F11.g	10028	4,0E-16	RNAt sintetase		
PSADSAU002D10.b	23872	1,0E-07	Proteína de membrana		
PSADSAU005F11.b	24480	1,0E-10	Histona deacetilase		
PSADSAU006A06.g	1998	2,0E-41	Peptidase_M8, Leishmanolysin..		
PSADSAU006F11.g	24480	9,0E-25	Histona deacetilase		
PSADSAU007A02.b	9491	1,0E-10	Possível antígeno de superfície		
PSADSAU008A02.b	9491	5,0E-10	Possível antígeno de superfície		
PSADSAU009G10.g	14302	1,0E-15	Fator de splicing associado ao U5 snRNA		
PSADSAU001A09.g			Serina/treonina Kinase	IPR000719	8E-05
PSADSAU001F08.g			Proteína kinase-like (PK-like)	IPR011009	0.0094
PSADSAU002C05.b			Proteína kinase	IPR000719	8.655
PSADSAU005G12.b			Proteína kinase	IPR000719	8E-05
PSADSAU007E11.b			Proteína kinase-like (PK-like)	IPR011009	0.0037

Alguns clusters foram anotados como proteínas hipotéticas não apresentaram similaridade com os bancos de dados utilizados, porém segundo a análise com o Glimmer estes clusters podem ser considerados como proteínas hipotéticas. Na tabela 4.6 e 4.7 podemos observar a lista das seqüências anotadas da biblioteca *Sau3A* e dos ESTs, respectivamente.

Com as análises realizadas, conseguimos 22 clusters em GSS que não foram encontradas anteriormente em *P. serpens*. Em ESTs anotamos 340 clusters em funções diversas.

Tabela 4.6: Lista das seqüências da biblioteca GSS anotadas

Cluster	Conteúdo GC	Anotação	Organismo	Evalue	Banco de dados
PSADSAU002D10.b	56%	Proteína de membrana	<i>L. major</i>	3,0E-54	Lmajor_NCBI
PSADSAU005G02.b	47%	Triptofano 2-like proteína	<i>L. major</i>	8,0e-22	Lmajor_NCBI
PSADSAU006G02.g	56%	Triptofano 2-like proteína	<i>T. brucei</i>	5,0E-53	Tbrucei_NCBI
PSADSAU007G11.b	66%	(Q9NDY5) Proteína hipotética conservada L5856.04	<i>L. major</i>	1,0E-20	Lmajor_NCBI
PSADSAU009F03.g	61%	Adenilato ciclase	<i>L. donovani</i>	5,0E-67	kineto-non-trit
PSADSAU001E12.g	51%	Amino acido permease	<i>L. donovani</i>	2,0E-52	kineto-non-trit
PSADSAU009F01.g	48%	Antígeno de proliferação celular (PCNA), putativa	<i>L. major</i>	8,0E-67	Lmajor_NCBI
PSADSAU005F03.b	58%	Dineína cadeia pesada, putativa	<i>L. major</i>	6,0E-32	Lmajor_NCBI
PSADSAU009H02.g	61%	Fator de alongamento	<i>L. major</i>	3,0E-41	Lmajor_NCBI
PSADSAU008H02.b	59%	Fator de alongamento 1-alpha	<i>L. major</i>	5,0E-61	Lmajor_NCBI
PSADSAU007H02.b	57%	Fator de alongamento 1-alpha	<i>T. brucei</i>	6,0E-59	Tbrucei_NCBI
PSADSAU003F02.g	64%	Fator splicing predito		2,0E-07	Kog
PSADSAU009F11.g	49%	Fosfatase ácida de membrana 2	<i>L. mexicana</i>	4,0E-15	kineto-non-trit
PSADSAU001A12.g	48%	Fosfatidilinositol 3-kinase, putativa	<i>L. major</i>	1,0E-110	Lmajor_NCBI
PSADSAU001F10.g	55%	Glicil RNAt sintetase, putativa	<i>L. major</i>	1,0E-134	Lmajor_NCBI
PSADSAU006F11.g	50%	Histona deacetilase, putativa	<i>T. cruzi</i>	5,0E-32	Tcruzi_NCBI
PSADSAU005F11.b	52%	Histona deacetilase, putativa	<i>L. major</i>	3,0E-29	Lmajor_NCBI
PSADSAU006A06.g	50%	Peptidase_M8: Leishmanolysin		1,0E-41	Pfam
PSADSAU006D09.g	55%	Piroglutamil-peptidase I	<i>L. major</i>	1,0e-14	Lmajor_NCBI
PSADSAU008A02.b	50%	Possível antígeno de superfície	<i>L. major</i>	8,0E-13	Lmajor_NCBI
PSADSAU007A02.b	47%	Possível antígeno de superfície		6,0E-11	Pfam
PSADSAU002C05.b	70%	Proteína kinase, putativa	<i>L. major</i>	2,0E-33	Lmajor_NCBI
PSADSAU001F08.g	69%	Proteína kinase, putativa	<i>L. major</i>	9,0E-30	Lmajor_NCBI
PSADSAU007E11.b	49%	Proteína kinase, putativa	<i>L. major</i>	6,0E-15	Lmajor_NCBI
PSADSAU007H04.b	65%	Proteína hipotética			
PSADSAU006B11.g	52%	Proteína hipotética			
PSADSAU003H09.g	63%	Proteína hipotética			
PSADSAU010E01.g	60%	Proteína hipotética			
PSADSAU009C04.g	52%	Proteína hipotética			
PSADSAU007F10.b	37%	Proteína hipotética			

Continuação da tabela 4.6.

Cluster	Conteúdo GC	Anotação	Organismo	Evalue	Banco de dados
PSADSAU007C02.b	48%	Proteína hipotética			
PSADSAU005F10.b	44%	Proteína hipotética			
PSADSAU005B11.b	38%	Proteína hipotética			
PSADSAU003A09.g	42%	Proteína hipotética conservada	<i>T. cruzi</i>	9,0E-67	Tcruzi_NCBI
PSADSAU006E10.g	54%	Proteína hipotética conservada	<i>L. major</i>	2,0E-47	Lmajor_NCBI
PSADSAU011G11.b	59%	Proteína hipotética conservada	<i>T. cruzi</i>	2,0E-06	Tcruzi_NCBI
PSADSAU011B04.b	65%	Proteína hipotética conservada	<i>L. major</i>	2,0E-18	Lmajor_NCBI
PSADSAU010G01.g	62%	Proteína hipotética conservada	<i>L. major</i>	5,0E-43	Lmajor_NCBI
PSADSAU010E11.g	56%	Proteína hipotética conservada	<i>L. major</i>	4,0E-08	Lmajor_NCBI
PSADSAU009H08.g	51%	Proteína hipotética conservada	<i>L. major</i>	4,0E-14	Lmajor_NCBI
PSADSAU009H07.g	67%	Proteína hipotética conservada	<i>L. major</i>	3,0E-11	Lmajor_NCBI
PSADSAU009F09.g	68%	Proteína hipotética conservada	<i>L. major</i>	9,0E-12	uniprot_trembl
PSADSAU009E11.g	53%	Proteína hipotética conservada	<i>L. major</i>	8,0E-21	Lmajor_NCBI
PSADSAU009D09.g	54%	Proteína hipotética conservada	<i>T. cruzi</i>	1,0E-49	Tcruzi_NCBI
PSADSAU009D06.g	52%	Proteína hipotética conservada	<i>L. major</i>	5,0E-36	Lmajor_NCBI
PSADSAU009D04.g	45%	Proteína hipotética conservada	<i>T. brucei</i>	1,0E-11	Tbrucei_NCBI
PSADSAU009C06.g	53%	Proteína hipotética conservada	<i>L. major</i>	2,0E-25	Lmajor_NCBI
PSADSAU009B05.g	53%	Proteína hipotética conservada	<i>L. major</i>	4,0E-14	Lmajor_NCBI
PSADSAU009A05.g	47%	Proteína hipotética conservada	<i>L. major</i>	2,0E-26	Lmajor_NCBI
PSADSAU008A05.b	51%	Proteína hipotética conservada	<i>L. major</i>	2,0E-19	Lmajor_NCBI
PSADSAU007H05.b	67%	Proteína hipotética conservada	<i>T. brucei</i>	6,0E-11	Tbrucei_NCBI
PSADSAU007C06.b	59%	Proteína hipotética conservada	<i>T. cruzi</i>	1,0E-14	Tcruzi_NCBI
PSADSAU007A04.b	55%	Proteína hipotética conservada	<i>L. major</i>	9,0E-23	Lmajor_NCBI
PSADSAU006F01.g	46%	Proteína hipotética conservada	<i>L. major</i>	4,0E-15	Lmajor_NCBI
PSADSAU006A11.g	44%	Proteína hipotética conservada	<i>L. major</i>	4,0E-40	Lmajor_NCBI
PSADSAU006A09.g	45%	Proteína hipotética conservada	<i>L. major</i>	7,0E-22	Lmajor_NCBI
PSADSAU005E10.b	54%	Proteína hipotética conservada	<i>L. major</i>	1,0E-45	Lmajor_NCBI
PSADSAU005A10.b	54%	Proteína hipotética conservada	<i>L. major</i>	8,0E-30	Lmajor_NCBI
PSADSAU003B12.g	69%	Proteína hipotética conservada		6,0E-08	uniprot_sprot.f
PSADSAU002G05.b	47%	Proteína hipotética conservada	<i>L. major</i>	4,0E-11	Lmajor_NCBI
PSADSAU002E09.b	53%	Proteína hipotética conservada	<i>L. major</i>	1,0E-14	Lmajor_NCBI
PSADSAU002B05.b	54%	Proteína hipotética conservada	<i>L. major</i>	1,0E-81	Lmajor_NCBI

Continuação da tabela 4.6.

Cluster	Conteúdo GC	Anotação	Organismo	Evalúe	Banco de dados
PSADSAU002A08.b	77%	Proteína hipotética conservada	<i>L. major</i>	8,0E-12	Lmajor_NCBI
PSADSAU001H12.g	61%	Proteína hipotética conservada	<i>L. major</i>	4,0E-09	Lmajor_NCBI
PSADSAU001H09.g	66%	Proteína hipotética conservada	<i>L. major</i>	4,0E-33	Lmajor_NCBI
PSADSAU001H08.g	49%	Proteína hipotética conservada	<i>L. major</i>	1,0E-19	Lmajor_NCBI
PSADSAU001G11.g	42%	Proteína hipotética conservada	<i>T. brucei</i>	8,0E-37	Tbrucei_NCBI
PSADSAU001G09.g	55%	Proteína hipotética conservada	<i>L. major</i>	9,0E-67	Lmajor_NCBI
PSADSAU001F11.g	63%	Proteína hipotética conservada	<i>T. cruzi</i>	1,0E-31	Tcruzi_NCBI
PSADSAU001C12.g	48%	Proteína hipotética conservada	<i>T. cruzi</i>	6,0E-43	Tcruzi_NCBI
PSADSAU001A09.g	48%	Proteína hipotética conservada	<i>T. cruzi</i>	5,0E-28	Tcruzi_NCBI
PSADSAU007F01.b	51%	Proteína kinase-like , putativa	<i>T. cruzi</i>	3,0E-40	Tcruzi_NCBI
PSADSAU005G03.b	51%	NADPH-citocromo p450 redutase-like	<i>L. major</i>	3,0E-88	Lmajor_NCBI
PSADSAU006C11.g	56%	Proteína nucleolar, putativa		9,0E-57	Kog
PSADSAU006H09.g	70%	Serina-treonina proteína kinase	<i>L. major</i>	3,0E-08	Lmajor_NCBI
PSADSAU006A10.g	64%	Transportador ABC , putativa	<i>L. major</i>	2,0E-35	Lmajor_NCBI
PSADSAU005D04.b	60%	Transportador amino ácido, putativa	<i>L. amazonens</i>	4,0E-37	kineto-non-trit
PSADSAU009G10.g	52%	U5 snRNA-fator associado splicing	<i>L. major</i>	9,0E-40	Lmajor_NCBI

Tabela 4.7: Lista das seqüências da biblioteca EST anotadas

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST007C03.b	57%	Proteína choque térmico 70 (hsp70)	<i>L. braziliensis</i>	1,0E-111	kineto-non-trit
PSADEST007B02.b	54%	2-hidroxiácido desidrogenase	<i>P. sp.</i>	1,0E-109	kineto-non-trit
PSADEST002C03.b	66%	ABC transportador, putativa	<i>L. major</i>	4,0E-66	Lmajor_NCBI
PSADEST001D03.b	71%	Acetoin desidrogenase e3 componente-like	<i>L. major</i>	1,0E-35	Lmajor_NCBI
PSADEST001D03.b	71%	Acetoin desidrogenase e3 componente-like	<i>L. major</i>	1,0E-35	Lmajor_NCBI
PSADEST002A02.b	53%	Acetil-CoA sintetase	<i>L. major</i>	7,0E-74	Lmajor_NCBI
PSADEST007B05.b	61%	Proteína ribossomal ácida P2 beta	<i>L. braziliensis</i>	8,0E-15	kineto-non-trit
PSADEST006E09.b	61%	Receptor proteína quinase C ácida, putativa	<i>T. cruzi</i>	4,0E-90	Tcruzi_NCBI
PSADEST001H03.b	41%	Proteína carreadora ácida, precursora mitocondrial, putativa	<i>T. brucei</i>	6,0E-43	Tbrucei_NCBI
PSADEST005G11.b	48%	Adenina fosforribosiltransferase	<i>L. donovani</i>	2,0E-81	kineto-non-trit
PSADEST002A06.b	48%	Adenosina quinase, putativa	<i>T. cruzi</i>	9,0E-58	Tcruzi_NCBI
PSADEST003C12.b	47%	Adenilate quinase, putativa	<i>T. cruzi</i>	5,0E-51	Tcruzi_NCBI
PSADEST004H01.b	45%	Adenilate quinase, putativa	<i>L. major</i>	4,0E-45	Lmajor_NCBI
PSADEST004D09.b	43%	ADP/ATP carreadora	<i>L. mexicana</i>	3,0E-32	kineto-non-trit
PSADEST001G05.b	48%	Fator ativador ADP-ribosilação GTPase, putativa	<i>L. major</i>	5,0E-35	Lmajor_NCBI
PSADEST001F12.b	49%	Fator ADP-ribosilação, putativa	<i>L. major</i>	2,0E-90	Lmajor_NCBI
PSADEST007H06.b	58%	Álcool desidrogenase, putativa	<i>T. cruzi</i>	9,0E-45	Tcruzi_NCBI
PSADEST007H03.b	61%	Alfa-tubulina	<i>T. danilewski</i>	1,0E-121	kineto-non-trit
PSADEST008B04.b	51%	oxidase alternativa	<i>T. vivax</i>	8,0E-53	kineto-non-trit
PSADEST001H01.b	62%	amino acido permease, putativa	<i>T. cruzi</i>	9,0E-21	Tcruzi_NCBI
PSADEST007A04.b	53%	amino acida permease, putativa	<i>L. donovani</i>	9,0E-52	kineto-non-trit
PSADEST006D12.b	51%	aminoacido permease, putativa; aminoacido transportadora, putativa	<i>L. major</i>	2,0E-44	Lmajor_NCBI
PSADEST007D04.b	45%	aminopeptidase, putativa	<i>L. major</i>	4,0E-82	Lmajor_NCBI
PSADEST005D02.b	51%	anion-transportador ATPase-like, putativa	<i>T. cruzi</i>	2,0E-51	Tcruzi_NCBI
PSADEST007F07.b	58%	aquagliceroporina	<i>L. mexicana</i>	2,0E-35	kineto-non-trit
PSADEST006G06.b	50%	aquagliceroporina 9	<i>L. tarentolae</i>	1,0E-35	kineto-non-trit
PSADEST004H03.b	46%	aquaporina 3, putativa	<i>L. major</i>	1,0E-19	Lmajor_NCBI
PSADEST002G04.b	52%	arginina N-metiltransferase, putativa	<i>T. cruzi</i>	2,0E-80	Tcruzi_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST004A02.b	46%	Aspartil protease		4,0E-24	Pfam
PSADEST006A04.b	53%	RNA helicase ATP-dependente, putativa	<i>L. major</i>	9,0E-69	Lmajor_NCBI
PSADEST002A05.b	51%	axoneme central apparatus, putativa	<i>T. cruzi</i>	9,0E-49	Tcruzi_NCBI
PSADEST006B02.b	58%	Fator básico de transcrição 3a, putativa	<i>L. major</i>	3,0E-43	Lmajor_NCBI
PSADEST005H07.b	60%	Beta-tubulin	<i>L. major</i>	1,0E-146	lmajor-ntta
PSADEST007A02.b	55%	Beta-tubulin	<i>T. danilewski</i>	4,0E-43	kineto-non-trit
PSADEST001G08.b	52%	branch point binding, putativa	<i>T. cruzi</i>	1,0E-63	Tcruzi_NCBI
PSADEST003G06.b	57%	C-8 esterol isomerase-like	<i>L. major</i>	2,0E-54	Lmajor_NCBI
PSADEST001E04.b	55%	Calmodulina	<i>L. major</i>	1,0E-149	lmajor-ntta
PSADEST005F02.b	39%	Calmodulina, putativa	<i>T. cruzi</i>	6.0e-22	Tcruzi_NCBI
PSADEST001E09.b	47%	Calpain-like cisteina peptidase, putativa	<i>T. brucei</i>	9,0E-45	uniprot_trembl.
PSADEST004F02.b	62%	carbamil fosfate sintase, putativa	<i>L. mexicana</i>	9,0E-86	kineto-non-trit
PSADEST005F10.b	48%	transportadora cation, putativa	<i>T. cruzi</i>	2,0E-30	Tcruzi_NCBI
PSADEST002E09.b	47%	CG10910-PB	<i>D. melanogater</i>	6,0E-07	refseq_Proteína
PSADEST003G11.b	52%	cis-trans isomerase (ciclofilin- 40), putativa		9.0e-59	Pfam
PSADEST008A11.b	57%	Co-chaperonina CPN10	<i>L. donovani</i>	2,0E-43	kineto-non-trit
PSADEST001H06.b	54%	Co-chaperonina CPN10	<i>L. donovani</i>	6,0E-19	kineto-non-trit
PSADEST004B08.b	41%	Cof: Predicted hydrolases of the HAD superfamily		2,0E-14	Cdd
PSADEST005F11.b	68%	cofactor-independente fosfoglicerato mutase	<i>L. mexicana</i>	7,0E-81	kineto-non-trit
PSADEST006E03.b	42%	ciclofilina	<i>L. donovani</i>	4,0E-74	kineto-non-trit
PSADEST001C09.b	63%	ciclofilina A		1,8E-108	Interpro
PSADEST005G10.b	54%	cistationa gamma liase, putativa	<i>T. cruzi</i>	4,0E-78	Tcruzi_NCBI
PSADEST007F02.b	59%	cistationina beta-sintetase	<i>L. tarentolae</i>	3,0E-52	kineto-non-trit
PSADEST001F03.b	38%	cisteina peptidase, putativa	<i>L. major</i>	6,0E-18	Lmajor_NCBI
PSADEST003D05.b	43%	cisteina peptidase, putativa	<i>T. cruzi</i>	2,0E-43	Tcruzi_NCBI
PSADEST004E03.b	29%	citocromo oxidase	<i>L. tarentolae</i>	9,0E-25	kineto-non-trit
PSADEST002A11.b	50%	dihidrolipoamido desidrogenase, putativa	<i>L. major</i>	4,0E-88	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST003A03.b	53%	DNA polimerase IV, putativa; DNA polimerase kappa, putativa	<i>L. major</i>	8,0E-25	Lmajor_NCBI
PSADEST006D06.b	43%	DNA-dependente RNA polimerases, putativa	<i>L. major</i>	1,0E-52	Lmajor_NCBI
PSADEST003F01.b	43%	dineína cadeia leve, putativa	<i>L. major</i>	4,0E-36	Lmajor_NCBI
PSADEST002G02.b	59%	Fator de alongamento 1-alpha	<i>L. braziliensis</i>	1,0E-128	lbraziliensis-n
PSADEST007H11.b	55%	Fator de alongamento 2	<i>L. major</i>	8,0E-75	Lmajor_NCBI
PSADEST002F02.b	42%	Fator de alongamento 2 subunidade alfa, putativa	<i>L. major</i>	2,0E-75	Lmajor_NCBI
PSADEST002E02.b	48%	Endonuclease III, putativa	<i>L. major</i>	1,0E-55	Lmajor_NCBI
PSADEST004A12.b	49%	Fator de Iniciação 5a eucariotos	<i>L. infantum</i>	2,0E-44	kineto-non-trit
PSADEST007B11.b	48%	Fator de Iniciação 5a eucariotos	<i>L. infantum</i>	4,0E-64	kineto-non-trit
PSADEST005D09.b	45%	Fator de liberação da cadeia peptídica 3 eucarioto subunidade 1	<i>T. brucei</i>	3,0E-70	Tbrucei_NCBI
PSADEST004F04.b	62%	Fator de liberação 3 eucarioto	<i>T. cruzi</i>	7,0E-66	Tcruzi_NCBI
PSADEST005G03.b	49%	fator de iniciação da tradução eucariótico 1A, putat	<i>L. major</i>	3,0E-65	Lmajor_NCBI
PSADEST002H09.b	43%	fator de iniciação da tradução eucariótico 3 (eIF-3)	<i>T. cruzi</i>	3,0E-56	Tcruzi_NCBI
PSADEST004H06.b	48%	fator de iniciação da tradução eucariótico 6, putativ	<i>T. brucei</i>	1,0E-101	Tbrucei_NCBI
PSADEST001E06.b	49%	expression-site associated gene (ESAG3), putativa	<i>L. major</i>	2,0E-12	Lmajor_NCBI
PSADEST001H12.b	49%	expression-site associated gene (ESAG3), putativa	<i>L. major</i>	8,0E-13	Lmajor_NCBI
PSADEST003E09.b	53%	farnesil sintetase, putativa	<i>L. major</i>	8,0E-45	Lmajor_NCBI
PSADEST004G02.b	49%	farnesiltransferase, putativa	<i>L. major</i>	1,0E-72	Lmajor_NCBI
PSADEST006G10.b	56%	frutose-1,6-bisfosfato aldolase	<i>L. mexicana</i>	1,0E-105	kineto-non-trit
PSADEST007H10.b	52%	glucosamina-frutose-6-fosfato aminotransferase, putativa	<i>L. major</i>	3,0E-48	Lmajor_NCBI
PSADEST004C02.b	48%	glutamina amidotransferase, putativa	<i>L. major</i>	9,0E-69	Lmajor_NCBI
PSADEST005C04.b	53%	gliceraldeído 3-fosfato desidrogenase, glicossom	<i>P. sp.</i>	2,0E-41	kineto-non-trit
PSADEST001C03.b	48%	Glicerol quinase, putativa	<i>L. major</i>	8,0E-71	Lmajor_NCBI
PSADEST005H09.b	59%	Glicossomal gliceraldeído-3-fosfato dehidrogenase	<i>P. sp.</i>	1,0E-113	kineto-non-trit
PSADEST008B02.b	63%	glicossomal fosfoenolpiruvato carboxiquinase, putativa	<i>L. major</i>	1,0E-103	Lmajor_NCBI
PSADEST001B02.b	40%	GTPase Rab4, small G superfamily		1,0E-65	Smart
PSADEST002D05.b	49%	GTP-binding, putativa	<i>L. major</i>	1,0E-101	Lmajor_NCBI
PSADEST004A11.b	51%	haloacido dehalogenase-like hidrolase-like	<i>L. major</i>	9,0E-54	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST003D09.b	58%	Heat shock 70-related	<i>L. major</i>	4,0E-75	Lmajor_NCBI
PSADEST007A09.b	61%	Heat shock 70	<i>T. cruzi</i>	1,0E-162	Tcruzi_NCBI
PSADEST005F09.b	59%	Heat shock Proteína 83, putativa	<i>T. cruzi</i>	3,0E-86	Tcruzi_NCBI
PSADEST008C01.b	53%	Hexose transportadora, putativa	<i>T. cruzi</i>	1,0E-39	Tcruzi_NCBI
PSADEST002G08.b	47%	Histona deacetilase, putativa	<i>L. major</i>	1,0E-71	Lmajor_NCBI
PSADEST006G08.b	48%	Histona H1		2,0E-18	Pfam
PSADEST006H01.b	56%	Histona H1-like	<i>C. fasciculata</i>	1.0e-10	kineto-non-trit
PSADEST006C11.b	63%	Histona H2A	<i>T. brucei</i>	5,0E-37	Tbrucei_NCBI
PSADEST006F06.b	56%	Histona H2A	<i>L. major</i>	4,0E-46	Lmajor_NCBI
PSADEST007E06.b	36%	Histona H2B, putativa		4,0E-12	Smart
PSADEST007G09.b	70%	Histona H2B, putativa	<i>T. brucei</i>	7,0E-11	Tbrucei_NCBI
PSADEST007F01.b	66%	Histona H3	<i>L. mexicana</i>	3,0E-52	kineto-non-trit
PSADEST006A02.b	61%	Histona H3	<i>L. mexicana</i>	4,0E-51	kineto-non-trit
PSADEST003B06.b	62%	Histona H4	<i>L. tarentolae</i>	4,0E-36	kineto-non-trit
PSADEST007H09.b	58%	Histona H4	<i>L. tarentolae</i>	7,0E-36	kineto-non-trit
PSADEST002D11.b	58%	Histona H4	<i>L. tarentolae</i>	6,0E-36	kineto-non-trit
PSADEST005H12.b	65%	Histona h4	<i>L. major</i>	4,0E-35	Lmajor_NCBI
PSADEST007D08.b	50%	inibidor da cisteína peptidase	<i>L. major</i>	5,0E-17	Lmajor_NCBI
PSADEST002G05.b	65%	inosina-5'-monofosfato desidrogenase	<i>L. major</i>	1,0E-71	Lmajor_NCBI
PSADEST001C08.b	46%	iron-sulfur cluster assembly, putativa	<i>L. major</i>	4,0E-44	Lmajor_NCBI
PSADEST002C02.b	46%	isocitrato desidrogenase [NADP], precursor mitocondrial	<i>L. major</i>	5,0E-53	Lmajor_NCBI
PSADEST001B05.b	58%	quinase, glicossomal, putativa	<i>L. major</i>	1,0E-07	Lmajor_NCBI
PSADEST003H10.b	47%	Proteína relacionada a quinesina	<i>L. donovani</i>	5,0E-20	kineto-non-trit
PSADEST005B08.b	40%	quinesina, putativa	<i>L. major</i>	1,0E-23	Lmajor_NCBI
PSADEST006A01.b	61%	L6 Proteína ribossomal	<i>L. braziliensis</i>	1,0E-69	kineto-non-trit
PSADEST007D03.b	59%	enzima malica, putativa	<i>L. major</i>	2,0E-80	Lmajor_NCBI
PSADEST006A10.b	44%	enzima malica, putativa	<i>T. cruzi</i>	4,0E-41	Tcruzi_NCBI
PSADEST005B02.b	45%	família maoc desidratase-like	<i>L. major</i>	6,0E-43	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST006D05.b	58%	Proteína de ligação a sequencia minicírculo (UMSBP), putativa	<i>L. major</i>	2,0E-57	Lmajor_NCBI
PSADEST001F09.b	40%	translocase mitocondrial subunidade, putativa	<i>L. major</i>	7,0E-26	Lmajor_NCBI
PSADEST005G07.b	47%	nucleotideo-binding MRP-familia	<i>L. major</i>	4,0E-78	uniprot_treml.
PSADEST008A03.b	47%	NAD(p)-dependente esteroide desidrogenase-like	<i>L. major</i>	5,0E-84	Lmajor_NCBI
PSADEST003E12.b	55%	NADH-dependente fumarato redutase, putativa	<i>L. major</i>	9,0E-37	Lmajor_NCBI
PSADEST004B07.b	51%	NADH-ubiquinona oxidoreductase, mitocondrial, put	<i>T. brucei</i>	2,0E-41	Tbrucei_NCBI
PSADEST006G07.b	50%	Proteína nucleolar, putativa	<i>L. major</i>	1,0E-74	Lmajor_NCBI
PSADEST006F08.b	48%	nucleolar RNA binding, putativa	<i>L. major</i>	3,0E-87	Lmajor_NCBI
PSADEST007B07.b	46%	Nucleosideo difosfato quinase		5.6e-98	Interpro
PSADEST001G03.b	48%	nucleosideo transportadora 2	<i>C. fasciculata</i>	2,0E-41	kineto-non-trit
PSADEST002A03.b	49%	nucleosideo transportadora-like, putativa	<i>T. cruzi</i>	8,0E-41	Tcruzi_NCBI
PSADEST003H08.b	48%	nucleosideo transportadora-like, putativa	<i>T. cruzi</i>	4,0E-40	Tcruzi_NCBI
PSADEST002H10.b	54%	nucleosome assembly protein-like	<i>L. major</i>	1,0E-10	Lmajor_NCBI
PSADEST004E12.b	52%	componente oligomérico complexo de golgi 8, putativa	<i>T. cruzi</i>	6,0E-29	Tcruzi_NCBI
PSADEST005F06.b	50%	oxidoreductase, putativa	<i>T. cruzi</i>	1,0E-63	Tcruzi_NCBI
PSADEST006C06.b	59%	p36 LACK	<i>L. amazonens</i>	1,0E-128	kineto-non-trit
PSADEST006B04.b	57%	proteína paraflagellar rod	<i>C. fasciculata</i>	1,0E-160	kineto-non-trit
PSADEST007G03.b	53%	proteína paraflagellar rod 1D	<i>L. mexicana</i>	1.0e-94	kineto-non-trit
PSADEST007H01.b	58%	proteína paraflagellar rod 2C	<i>L. major</i>	2,0E-90	Lmajor_NCBI
PSADEST004E02.b	60%	proteína de membrana peroxissomal 4, putativa	<i>L. major</i>	2,0E-62	Lmajor_NCBI
PSADEST004E02.b	60%	proteína de membrana peroxissomal 4, putativa	<i>L. major</i>	2,0E-62	Lmajor_NCBI
PSADEST007D12.b	63%	fosfoglicerato quinase, glicossomal	<i>C. fasciculata</i>	4,0E-96	kineto-non-trit
PSADEST003A10.b	58%	poli(a) binding, putativa	<i>C. fasciculata</i>	4,0E-72	kineto-non-trit
PSADEST005B01.b	54%	poliadenilato-binding 1, putativa	<i>L. major</i>	2,0E-70	Lmajor_NCBI
PSADEST001A11.b	57%	polimerase-like, putativa	<i>L. major</i>	9,0E-06	Lmajor_NCBI
PSADEST006B09.b	62%	possível polipeptideo nascente associado complexo subunidade, cópia 1	<i>L. major</i>	2,0E-54	Lmajor_NCBI
PSADEST001A01.b	40%	possível antígeno de superfície	<i>L. major</i>	1,0E-25	Lmajor_NCBI
PSADEST004B01.b	43%	prostaglandina f sintase, putativa	<i>L. major</i>	2,0E-71	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST001G12.b	60%	proteossomo alpha subunidade 3, putativa	<i>L. major</i>	3,0E-78	Lmajor_NCBI
PSADEST004B10.b	51%	proteossomo alpha subunidade 7, putativa	<i>L. major</i>	1,0E-104	Lmajor_NCBI
PSADEST007C05.b	62%	Proteína de superfície celular, putativa	<i>L. donovani</i>	2,0E-67	kineto-non-trit
PSADEST004A03.b	45%	Proteína disulfide isomerase	<i>L. donovani</i>	7,0E-31	kineto-non-trit
PSADEST007B12.b	50%	Proteína disulfide isomerase	<i>L. donovani</i>	3,0E-33	kineto-non-trit
PSADEST002B10.b	61%	Proteína hipotética			
PSADEST002D01.b	48%	Proteína hipotética			
PSADEST003C11.b	60%	Proteína hipotética			
PSADEST004F03.b	31%	Proteína hipotética			
PSADEST004G10.b	50%	Proteína hipotética			
PSADEST005E08.b	40%	Proteína hipotética			
PSADEST006B10.b	60%	Proteína hipotética			
PSADEST007F12.b	56%	Proteína hipotética			
PSADEST001A05.b	47%	Proteína hipotética conservada	<i>L. major</i>	4,0E-28	Lmajor_NCBI
PSADEST001A07.b	52%	Proteína hipotética conservada	<i>L. major</i>	4,0E-23	Lmajor_NCBI
PSADEST001B08.b	60%	Proteína hipotética conservada	<i>T. cruzi</i>	1,0E-24	Tcruzi_NCBI
PSADEST001B09.b	49%	Proteína hipotética conservada	<i>L. major</i>	9,0E-07	Lmajor_NCBI
PSADEST001C02.b	46%	Proteína hipotética conservada	<i>L. major</i>	8,0E-36	Lmajor_NCBI
PSADEST001C04.b	52%	Proteína hipotética conservada	<i>L. major</i>	2,0E-36	Lmajor_NCBI
PSADEST001C07.b	42%	Proteína hipotética conservada	<i>L. major</i>	5,0E-38	Lmajor_NCBI
PSADEST001C11.b	50%	Proteína hipotética conservada	<i>T. cruzi</i>	2,0E-14	Tcruzi_NCBI
PSADEST001D07.b	47%	Proteína hipotética conservada	<i>L. major</i>	5,0E-40	Lmajor_NCBI
PSADEST001E01.b	49%	Proteína hipotética conservada	<i>L. major</i>	2,0E-20	Lmajor_NCBI
PSADEST001E03.b	52%	Proteína hipotética conservada	<i>L. major</i>	2,0E-43	Lmajor_NCBI
PSADEST001E05.b	65%	Proteína hipotética conservada	<i>L. major</i>	6,0E-50	Lmajor_NCBI
PSADEST001E07.b	45%	Proteína hipotética conservada	<i>L. major</i>	2,0E-24	Lmajor_NCBI
PSADEST001F07.b	48%	Proteína hipotética conservada	<i>L. major</i>	8,0E-28	Lmajor_NCBI
PSADEST002A01.b	42%	Proteína hipotética conservada	<i>T. cruzi</i>	1,0E-12	Tcruzi_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST002A12.b	56%	Proteína hipotética conservada	<i>L. major</i>	1,0E-20	Lmajor_NCBI
PSADEST002B11.b	42%	Proteína hipotética conservada	<i>L. major</i>	1,0E-25	Lmajor_NCBI
PSADEST002C04.b	41%	Proteína hipotética conservada	<i>L. major</i>	4,0E-48	Lmajor_NCBI
PSADEST002C06.b	52%	Proteína hipotética conservada	<i>L. major</i>	3,0E-48	Lmajor_NCBI
PSADEST002C07.b	44%	Proteína hipotética conservada	<i>L. major</i>	3,0E-48	Lmajor_NCBI
PSADEST002C11.b	51%	Proteína hipotética conservada	<i>T. cruzi</i>	1,0E-39	Tcruzi_NCBI
PSADEST002D03.b	41%	Proteína hipotética conservada	<i>T. cruzi</i>	2,0E-25	Tcruzi_NCBI
PSADEST002D07.b	58%	Proteína hipotética conservada	<i>L. major</i>	6,0E-55	Lmajor_NCBI
PSADEST002D08.b	46%	Proteína hipotética conservada	<i>L. major</i>	4,0E-38	Lmajor_NCBI
PSADEST002E08.b	50%	Proteína hipotética conservada	<i>L. major</i>	3,0E-71	Lmajor_NCBI
PSADEST002E11.b	62%	Proteína hipotética conservada	<i>T. cruzi</i>	4,0E-09	Tcruzi_NCBI
PSADEST002F01.b	37%	Proteína hipotética conservada	<i>T. cruzi</i>	3,0E-07	Tcruzi_NCBI
PSADEST002F03.b	57%	Proteína hipotética conservada	<i>L. major</i>	4,0E-36	Lmajor_NCBI
PSADEST002F09.b	63%	Proteína hipotética conservada	<i>L. major</i>	5,0E-23	Lmajor_NCBI
PSADEST002F11.b	52%	Proteína hipotética conservada	<i>T. cruzi</i>	6.0e-40	Tcruzi_NCBI
PSADEST002F12.b	49%	Proteína hipotética conservada	<i>L. major</i>	3,0E-21	Lmajor_NCBI
PSADEST002G09.b	54%	Proteína hipotética conservada	<i>L. major</i>	4,0E-09	Lmajor_NCBI
PSADEST002G12.b	40%	Proteína hipotética conservada	<i>T. cruzi</i>	9,0E-25	Tcruzi_NCBI
PSADEST002H02.b	67%	Proteína hipotética conservada	<i>T. brucei</i>	3,0E-12	Tbrucei_NCBI
PSADEST002H04.b	49%	Proteína hipotética conservada	<i>L. major</i>	1,0E-48	Lmajor_NCBI
PSADEST002H06.b	45%	Proteína hipotética conservada	<i>L. major</i>	3.0e-12	Lmajor_NCBI
PSADEST002H07.b	47%	Proteína hipotética conservada	<i>L. major</i>	1,0E-09	Lmajor_NCBI
PSADEST003A01.b	38%	Proteína hipotética conservada	<i>T. brucei</i>	3,0E-19	Tbrucei_NCBI
PSADEST003A02.b	41%	Proteína hipotética conservada	<i>L. major</i>	1,0E-40	Lmajor_NCBI
PSADEST003B03.b	59%	Proteína hipotética conservada	<i>T. cruzi</i>	3,0E-41	Tcruzi_NCBI
PSADEST003B10.b	44%	Proteína hipotética conservada	<i>L. major</i>	9,0E-48	Lmajor_NCBI
PSADEST003C05.b	48%	Proteína hipotética conservada	<i>T. cruzi</i>	9,0E-20	refseq_Proteína
PSADEST003D08.b	49%	Proteína hipotética conservada	<i>L. major</i>	8,0E-71	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST003D11.b	49%	Proteína hipotética conservada	<i>L. major</i>	7,0E-36	Lmajor_NCBI
PSADEST003E01.b	50%	Proteína hipotética conservada	<i>L. major</i>	8,0E-41	Lmajor_NCBI
PSADEST003E04.b	51%	Proteína hipotética conservada	<i>L. major</i>	1,0E-41	Lmajor_NCBI
PSADEST003F07.b	51%	Proteína hipotética conservada	<i>L. major</i>	4,0E-30	Lmajor_NCBI
PSADEST003G01.b	56%	Proteína hipotética conservada	<i>T. cruzi</i>	4,0E-11	Tcruzi_NCBI
PSADEST004A07.b	44%	Proteína hipotética conservada	<i>L. major</i>	7,0E-23	Lmajor_NCBI
PSADEST004B02.b	51%	Proteína hipotética conservada	<i>L. major</i>	9,0E-59	Lmajor_NCBI
PSADEST004B06.b	56%	Proteína hipotética conservada	<i>L. major</i>	2,0E-56	Lmajor_NCBI
PSADEST004B12.b	56%	Proteína hipotética conservada	<i>T. cruzi</i>	2,0E-12	Tcruzi_NCBI
PSADEST004C06.b	44%	Proteína hipotética conservada	<i>L. major</i>	2,0E-10	Lmajor_NCBI
PSADEST004D12.b	50%	Proteína hipotética conservada	<i>L. major</i>	8,0E-25	Lmajor_NCBI
PSADEST004E01.b	37%	Proteína hipotética conservada	<i>L. major</i>	2,0E-31	Lmajor_NCBI
PSADEST004E04.b	58%	Proteína hipotética conservada	<i>L. major</i>	1,0E-23	Lmajor_NCBI
PSADEST004F05.b	53%	Proteína hipotética conservada	<i>T. cruzi</i>	4,0E-24	Tcruzi_NCBI
PSADEST004F07.b	58%	Proteína hipotética conservada	<i>L. major</i>	2,0E-66	Lmajor_NCBI
PSADEST004F09.b	51%	Proteína hipotética conservada	<i>T. cruzi</i>	6,0E-19	Tcruzi_NCBI
PSADEST004G03.b	44%	Proteína hipotética conservada	<i>L. major</i>	3,0E-25	Lmajor_NCBI
PSADEST004G05.b	41%	Proteína hipotética conservada	<i>L. major</i>	4,0E-58	Lmajor_NCBI
PSADEST004G07.b	52%	Proteína hipotética conservada	<i>L. major</i>	3,0E-45	Lmajor_NCBI
PSADEST004G08.b	62%	Proteína hipotética conservada	<i>T. cruzi</i>	7,0E-48	Tcruzi_NCBI
PSADEST005A02.b	43%	Proteína hipotética conservada	<i>L. major</i>	1,0E-33	Lmajor_NCBI
PSADEST005A08.b	46%	Proteína hipotética conservada	<i>T. cruzi</i>	6,0E-17	Tcruzi_NCBI
PSADEST005B10.b	47%	Proteína hipotética conservada	<i>L. major</i>	2,0E-34	Lmajor_NCBI
PSADEST005B11.b	48%	Proteína hipotética conservada	<i>L. major</i>	5,0E-33	Lmajor_NCBI
PSADEST005C03.b	45%	Proteína hipotética conservada	<i>L. amazonensis</i>	3,0E-11	kineto-non-trit
PSADEST005D08.b	45%	Proteína hipotética conservada	<i>T. cruzi</i>	6,0E-32	Tcruzi_NCBI
PSADEST005G02.b	56%	Proteína hipotética conservada	<i>L. major</i>	1,0E-35	Lmajor_NCBI
PSADEST005H06.b	69%	Proteína hipotética conservada	<i>L. major</i>	8,0E-18	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST006A07.b	62%	Proteína hipotética conservada	<i>L. major</i>	1,0E-46	Lmajor_NCBI
PSADEST006B05.b	46%	Proteína hipotética conservada	<i>L. major</i>	7,0E-25	Lmajor_NCBI
PSADEST006D01.b	58%	Proteína hipotética conservada	<i>L. major</i>	1,0E-49	Lmajor_NCBI
PSADEST006E04.b	53%	Proteína hipotética conservada	<i>T. cruzi</i>	2,0E-20	Tcruzi_NCBI
PSADEST006F09.b	70%	Proteína hipotética conservada	<i>L. major</i>	1,0E-37	Lmajor_NCBI
PSADEST006F11.b	57%	Proteína hipotética conservada	<i>L. major</i>	7,0E-53	Lmajor_NCBI
PSADEST006H05.b	50%	Proteína hipotética conservada	<i>T. cruzi</i>	3,0E-95	Tbrucei_NCBI
PSADEST007A03.b	57%	Proteína hipotética conservada	<i>L. major</i>	3,0E-72	Lmajor_NCBI
PSADEST007D11.b	53%	Proteína hipotética conservada	<i>T. cruzi</i>	3,0E-17	Tcruzi_NCBI
PSADEST007F04.b	56%	Proteína hipotética conservada	<i>T. cruzi</i>	3,0E-47	Tcruzi_NCBI
PSADEST007F10.b	49%	Proteína hipotética conservada	<i>L. major</i>	2,0E-68	Lmajor_NCBI
PSADEST007G04.b	46%	Proteína hipotética conservada	<i>L. major</i>	9,0E-14	Lmajor_NCBI
PSADEST007H12.b	57%	Proteína hipotética conservada	<i>L. major</i>	8,0E-26	Lmajor_NCBI
PSADEST008A04.b	45%	Proteína hipotética conservada	<i>T. cruzi</i>	3,0E-14	Tcruzi_NCBI
PSADEST008A08.b	48%	Proteína hipotética conservada	<i>L. major</i>	2,0E-11	Lmajor_NCBI
PSADEST001D01.b	50%	Proteína hipotética Tc00.1047053507019.45	<i>T. cruzi</i>	4,0E-23	Tcruzi_NCBI
PSADEST006C03.b	48%	Proteína quinase	<i>E. histolytica</i>	4,0E-15	refseq_Proteína
PSADEST002B01.b	54%	Proteína quinase, putativa	<i>T. cruzi</i>	5,0E-35	Tcruzi_NCBI
PSADEST005B03.b	44%	Proteína tipo fosfatase 1 regulador-like	<i>L. major</i>	3,0E-42	Lmajor_NCBI
PSADEST007G10.b	56%	Proteína ribossomal 40S S10, putativa	<i>L. major</i>	5,0E-39	Lmajor_NCBI
PSADEST006H07.b	59%	Proteína ribossomal 40S S11, putativa	<i>T. cruzi</i>	6,0E-29	Tcruzi_NCBI
PSADEST008B09.b	55%	Proteína ribossomal 40S S12, putativa	<i>L. major</i>	4,0E-49	Lmajor_NCBI
PSADEST007G06.b	58%	Proteína ribossomal 40S S13, putativa	<i>T. brucei</i>	4,0E-71	Tbrucei_NCBI
PSADEST008B06.b	57%	Proteína ribossomal 40S S13, putativa	<i>T. cruzi</i>	9,0E-66	Tcruzi_NCBI
PSADEST006F02.b	64%	Proteína ribossomal 40S S15, putativa		3,0E-58	Kog
PSADEST005H03.b	51%	Proteína ribossomal 40S S15a, putativa	<i>T. brucei</i>	5,0E-58	Tbrucei_NCBI
PSADEST006E12.b	54%	Proteína ribossomal 40S S15a, putativa	<i>L. major</i>	5,0E-63	Lmajor_NCBI
PSADEST007G12.b	57%	Proteína ribossomal 40S S16, putativa	<i>L. major</i>	2,0E-65	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST006H04.b	62%	Proteína ribossomal 40S S17, putativa	<i>T. brucei</i>	1,0E-42	Tbrucei_NCBI
PSADEST005G12.b	56%	Proteína ribossomal 40S S18, putativa	<i>L. major</i>	2,0E-43	Lmajor_NCBI
PSADEST006E10.b	57%	Proteína ribossomal 40S S2, putativa	<i>T. brucei</i>	1,0E-107	Tbrucei_NCBI
PSADEST007E08.b	66%	Proteína ribossomal 40S S21, putativa	<i>L. major</i>	4,0E-38	Lmajor_NCBI
PSADEST007B09.b	56%	Proteína ribossomal 40S S23, putativa	<i>T. cruzi</i>	7,0E-73	Tcruzi_NCBI
PSADEST005H08.b	54%	Proteína ribossomal 40S S24e, putativa	<i>L. amazonens</i>	7,0E-41	kineto-non-trit
PSADEST006E08.b	61%	Proteína ribossomal 40S S27, putativa	<i>T. brucei</i>	1,0E-40	Tbrucei_NCBI
PSADEST008B11.b	63%	Proteína ribossomal 40S S3, putativa	<i>L. major</i>	2,0E-95	Lmajor_NCBI
PSADEST006E06.b	55%	Proteína ribossomal 40S S30, putativa	<i>L. major</i>	2,0E-21	Lmajor_NCBI
PSADEST007E05.b	62%	Proteína ribossomal 40S S3a, putativa	<i>L. major</i>	1,0E-83	Lmajor_NCBI
PSADEST007E04.b	67%	Proteína ribossomal 40S S4, putativa	<i>L. major</i>	1,0E-104	Lmajor_NCBI
PSADEST007E04.b	67%	Proteína ribossomal 40S S4, putativa	<i>L. major</i>	1,0E-104	Lmajor_NCBI
PSADEST006C10.b	62%	Proteína ribossomal 40S S5, putativa	<i>L. major</i>	6,0E-86	Lmajor_NCBI
PSADEST006C08.b	61%	Proteína ribossomal 40S S8, putativa	<i>L. major</i>	9,0E-82	Lmajor_NCBI
PSADEST006H08.b	60%	Proteína ribossomal 40S S9, putativa	<i>L. major</i>	7,0E-84	Lmajor_NCBI
PSADEST008A10.b	48%	Proteína ribossomal 50S L7Ae, putativa	<i>T. cruzi</i>	2,0E-65	Tcruzi_NCBI
PSADEST007E10.b	62%	Proteína ribossomal 60S ácida P2, putativa	<i>T. brucei</i>	5,0E-21	Tbrucei_NCBI
PSADEST007E11.b	60%	Proteína ribossomal 60S ácida, putativa	<i>L. major</i>	1,0E-114	Lmajor_NCBI
PSADEST007H04.b	62%	Proteína ribossomal 60S L10, putativa	<i>L. major</i>	1,0E-111	Lmajor_NCBI
PSADEST006H11.b	53%	Proteína ribossomal 60S L10a, putativa	<i>L. major</i>	1,0E-102	Lmajor_NCBI
PSADEST005H11.b	62%	Proteína ribossomal 60S L11, putativa		3,0E-80	Kog
PSADEST008B01.b	59%	Proteína ribossomal 60S L12, putativa	<i>L. major</i>	2,0E-58	Lmajor_NCBI
PSADEST006C04.b	60%	Proteína ribossomal 60S L13, putativa	<i>L. major</i>	2,0E-78	Lmajor_NCBI
PSADEST007C01.b	57%	Proteína ribossomal 60S L13a, putativa	<i>L. major</i>	1,0E-107	Lmajor_NCBI
PSADEST003C03.b	57%	Proteína ribossomal 60S L17, putativa	<i>T. cruzi</i>	3,0E-21	Tcruzi_NCBI
PSADEST006C09.b	54%	Proteína ribossomal 60S L17, putativa	<i>L. major</i>	2,0E-33	Lmajor_NCBI
PSADEST008B03.b	59%	Proteína ribossomal 60S L18, putativa	<i>L. major</i>	2,0E-82	Lmajor_NCBI
PSADEST008B10.b	60%	Proteína ribossomal 60S L18, putativa	<i>L. major</i>	1,0E-79	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST006C05.b	63%	Proteína ribossomal 60S L18a, putativa	<i>L. major</i>	5,0E-83	Lmajor_NCBI
PSADEST007E07.b	64%	Proteína ribossomal 60S L2, putativa	<i>T. cruzi</i>	1,0E-124	Tcruzi_NCBI
PSADEST006C02.b	64%	Proteína ribossomal 60S L21, putativa	<i>L. major</i>	9,0E-73	Lmajor_NCBI
PSADEST006H03.b	53%	Proteína ribossomal 60S L22, putativa	<i>L. major</i>	1,0E-39	Lmajor_NCBI
PSADEST007B08.b	54%	Proteína ribossomal 60S L22, putativa	<i>L. major</i>	2,0E-46	Lmajor_NCBI
PSADEST006E02.b	54%	Proteína ribossomal 60S L23, putativa	<i>L. major</i>	2,0E-65	Lmajor_NCBI
PSADEST005E11.b	53%	Proteína ribossomal 60S L23a, putativa	<i>T. brucei</i>	1,0E-53	Tbrucei_NCBI
PSADEST006F04.b	62%	Proteína ribossomal 60S L26, putativa	<i>L. major</i>	3,0E-71	Lmajor_NCBI
PSADEST006F03.b	56%	Proteína ribossomal 60S L27A/L29, putativa	<i>L. major</i>	5,0E-58	Lmajor_NCBI
PSADEST008B08.b	65%	Proteína ribossomal 60S L28, putativa	<i>L. major</i>	2,0E-38	Lmajor_NCBI
PSADEST006C12.b	55%	Proteína ribossomal 60S L30, putativa	<i>T. brucei</i>	6,0E-32	Tbrucei_NCBI
PSADEST007G01.b	61%	Proteína ribossomal 60S L31, putativa	<i>L. major</i>	3,0E-84	Lmajor_NCBI
PSADEST007H08.b	57%	Proteína ribossomal 60S L32, putativa	<i>L. major</i>	8,0E-33	Lmajor_NCBI
PSADEST007C02.b	64%	Proteína ribossomal 60S L34, putativa	<i>L. major</i>	2,0E-59	Lmajor_NCBI
PSADEST006A03.b	61%	Proteína ribossomal 60S L35, putativa	<i>L. major</i>	6,0E-40	Lmajor_NCBI
PSADEST005C10.b	48%	Proteína ribossomal 60S L36, putativa	<i>L. major</i>	3,0E-34	Lmajor_NCBI
PSADEST006C01.b	59%	Proteína ribossomal 60S L36, putativa	<i>L. major</i>	2,0E-29	Lmajor_NCBI
PSADEST006B12.b	57%	Proteína ribossomal 60S L37a, putativa	<i>L. major</i>	1,0E-37	Lmajor_NCBI
PSADEST007E02.b	45%	Proteína ribossomal 60S L37a, putativa	<i>L. major</i>	3,0E-39	Lmajor_NCBI
PSADEST006E07.b	48%	Proteína ribossomal 60S L44, putativa	<i>T. brucei</i>	4,0E-43	Tbrucei_NCBI
PSADEST007A10.b	46%	Proteína ribossomal 60S L44, putativa	<i>T. brucei</i>	4,0E-43	Tbrucei_NCBI
PSADEST005H01.b	55%	Proteína ribossomal 60S L5, putativa	<i>L. major</i>	2,0E-42	Lmajor_NCBI
PSADEST005H10.b	61%	Proteína ribossomal 60S L7, putativa	<i>L. major</i>	2,0E-75	Lmajor_NCBI
PSADEST006H12.b	58%	Proteína ribossomal 60S L7a, putativa	<i>L. major</i>	1,0E-106	Lmajor_NCBI
PSADEST006H10.b	58%	Proteína ribossomal 60S L9, putativa	<i>L. major</i>	2,0E-73	Lmajor_NCBI
PSADEST007C08.b	65%	Proteína ribossomal 60S		2.3e-24	Interpro
PSADEST007C06.b	63%	Proteína ribossomal L15e		1.1e-97	Interpro
PSADEST006F05.b	55%	Proteína ribossomal L19e		4.3e-81	Interpro

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST007A01.b	67%	Proteína ribossomal L24, putativa	<i>L. major</i>	9,0E-36	Lmajor_NCBI
PSADEST006H09.b	56%	Proteína ribossomal L27, putativa	<i>T. cruzi</i>	3.0e-65	Tcruzi_NCBI
PSADEST007E09.b	64%	Proteína ribossomal I3, putativa	<i>T. cruzi</i>	1,0E-179	Tcruzi_NCBI
PSADEST006B08.b	58%	Proteína ribossomal I35a, putativa	<i>L. major</i>	1,0E-73	Lmajor_NCBI
PSADEST008B05.b	58%	Proteína ribossomal I35a, putativa	<i>L. major</i>	2,0E-63	Lmajor_NCBI
PSADEST006G12.b	53%	Proteína ribossomal L38, putativa	<i>L. major</i>	3,0E-33	Lmajor_NCBI
PSADEST007B01.b	57%	Proteína ribossomal L38, putativa	<i>T. cruzi</i>	3,0E-34	Tcruzi_NCBI
PSADEST007B06.b	57%	Proteína ribossomal L39, putativa	<i>T. cruzi</i>	3,0E-24	Tcruzi_NCBI
PSADEST007H05.b	69%	Proteína ribossomal L4/L1e		3.9e-129	Interpro
PSADEST008A12.b	52%	Proteína ribossomal L41		2,0E-06	Kog
PSADEST006E11.b	67%	Proteína ribossomal S11		3.8e-66	Interpro
PSADEST003B12.b	56%	Proteína ribossomal S11	<i>L. donovani</i>	9,0E-42	uniprot_treml.
PSADEST001B07.b	67%	Proteína ribossomal S19-like Proteína	<i>L. major</i>	3,0E-45	Lmajor_NCBI
PSADEST007C04.b	62%	Proteína ribossomal S2		7.5e-103	Interpro
PSADEST007G11.b	57%	Proteína ribossomal s20, putativa	<i>L. major</i>	9,0E-29	Lmajor_NCBI
PSADEST005G08.b	62%	Proteína ribossomal S25	<i>L. major</i>	2,0E-34	Lmajor_NCBI
PSADEST007B04.b	58%	Proteína ribossomal S25	<i>L. infantum</i>	7,0E-38	kineto-non-trit
PSADEST005H02.b	65%	Proteína ribossomal S26, putativa	<i>L. major</i>	2,0E-51	Lmajor_NCBI
PSADEST007G07.b	62%	Proteína ribossomal S26, putativa	<i>L. major</i>	7,0E-44	Lmajor_NCBI
PSADEST004E05.b	55%	Proteína ribossomal S29, putativa	<i>T. cruzi</i>	4,0E-27	Tcruzi_NCBI
PSADEST007F09.b	58%	Proteína ribossomal S33	<i>L. infantum</i>	7,0E-31	kineto-non-trit
PSADEST001C10.b	47%	Proteína ribossomal S6, putativa	<i>T. cruzi</i>	9,0E-57	Tcruzi_NCBI
PSADEST007E12.b	59%	Proteína ribossomal S7, putativa	<i>T. cruzi</i>	4,0E-80	Tcruzi_NCBI
PSADEST004A01.b	44%	Pseudouridilato sintase-like	<i>L. major</i>	2,0E-62	Lmajor_NCBI
PSADEST007C10.b	46%	Pterin-4-alpha-carbinolamina desidratase, putativa	<i>L. major</i>	2,0E-34	Lmajor_NCBI
PSADEST002C05.b	51%	Puromicina-sensível aminopeptidase e aminopeptidases relacionadas		1,0E-10	Kog
PSADEST003C07.b	49%	Piridoxal quinase, putativa	<i>L. major</i>	3,0E-39	Lmajor_NCBI
PSADEST005F01.b	50%	rab1 small GTP-binding, putativa	<i>L. major</i>	8,0E-75	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST004B04.b	56%	Redutase cadeia curta, putativa	<i>L. mexicana</i>	2,0E-65	kineto-non-trit
PSADEST001D06.b	44%	reticulon domain, 22 kDa potentially aggravating (pape22)	<i>L. major</i>	6,0E-34	Lmajor_NCBI
PSADEST007F08.b	48%	Riboflavina quinase/fmn adenililtransferase-like	<i>L. major</i>	5,0E-39	Lmajor_NCBI
PSADEST001H08.b	55%	Ribonuclease, putativa	<i>T. cruzi</i>	1,0E-14	Tcruzi_NCBI
PSADEST007D02.b	52%	Ribonuclease, putativa	<i>T. cruzi</i>	7,0E-44	Tcruzi_NCBI
PSADEST007D05.b	58%	Ribossomal DEAD box	<i>L. braziliensis</i>	7,0E-87	kineto-non-trit
PSADEST007H07.b	66%	Proteína fosfo ribossomal S6	<i>L. infantum</i>	7,0E-84	kineto-non-trit
PSADEST003H03.b	49%	Proteína regulatória biogênese ribossomal (RRS1), putativa	<i>L. major</i>	4,0E-81	Lmajor_NCBI
PSADEST001F11.b	66%	Proteína ribossomal L37	<i>L. infantum</i>	3,0E-31	kineto-non-trit
PSADEST001H10.b	54%	Proteína de ligação RNA, putativa	<i>L. major</i>	9,0E-51	Lmajor_NCBI
PSADEST007G02.b	61%	Proteína de ligação RNA, putativa	<i>L. major</i>	3,0E-47	Lmajor_NCBI
PSADEST001G02.b	69%	RNA complexo edição MP46	<i>T. brucei</i>	9,0E-17	Tbrucei_NCBI
PSADEST002E06.b	40%	RNA helicase, putativa	<i>T. cruzi</i>	5,0E-30	Tcruzi_NCBI
PSADEST003H12.b	51%	RNA-binding, putativa	<i>T. brucei</i>	1,0E-31	Tbrucei_NCBI
PSADEST003E07.b	53%	S-adenosil-L-metionina-C-24-delta-esterol-metiltransferase	<i>L. donovani</i>	9,0E-80	kineto-non-trit
PSADEST005G09.b	57%	S-adenosilmetionina sintetase	<i>L. donovani</i>	2,0E-92	kineto-non-trit
PSADEST001A09.b	44%	Serina/Treonina proteína quinase	<i>L. major</i>	6,0E-57	Lmajor_NCBI
PSADEST001B11.b	40%	Serina/Treonina proteína quinase	<i>L. major</i>	7,0E-63	Lmajor_NCBI
PSADEST002B05.b	54%	Proteína heat shock 70-relacionada 1, precursor mitocondrial	<i>L. major</i>	2,0E-30	Lmajor_NCBI
PSADEST005D01.b	37%	Peptídeo sinal peptidase, putativa	<i>T. cruzi</i>	1,0E-34	Tcruzi_NCBI
PSADEST004H10.b	50%	Partícula reconhecedora de sinal, putativa	<i>L. major</i>	5,0E-09	Lmajor_NCBI
PSADEST003H11.b	50%	Proteína estresse induzida STII	<i>L. major</i>	2,0E-72	Lmajor_NCBI
PSADEST003G02.b	48%	Succinil-CoA sintetase, subunidade alfa		1,0E-101	Kog
PSADEST006B06.b	57%	Succinil-CoA transferase 1	<i>L. major</i>	1,0E-72	Lmajor_NCBI
PSADEST005C02.b	49%	Fator traducional SUI1, putativa	<i>L. major</i>	9,0E-37	Lmajor_NCBI
PSADEST007G08.b	50%	Proteína controladora traducional de tumor (TCTP), putativa	<i>T. brucei</i>	1,0E-45	Tbrucei_NCBI
PSADEST001G06.b	49%	Tripanotona redutase	<i>C. fasciculat</i>	2,0E-28	kineto-non-trit
PSADEST002H01.b	53%	Tirosina aminotransferase, putativa	<i>L. major</i>	1,0E-43	Lmajor_NCBI

Continuação da tabela 4.7

Cluster	Conteúdo GC	Anotação	Organismo	E-value	Banco de dados
PSADEST006F01.b	60%	Ubiquitina/Proteína ribossomal S27a, putativa	<i>L. major</i>	2,0E-48	Lmajor_NCBI
PSADEST005H05.b	58%	Ubiquitina-fusão	<i>L. tarentolae</i>	2,0E-65	kineto-non-trit
PSADEST007C11.b	58%	Ubiquitina-fusão	<i>L. tarentolae</i>	2,0E-65	kineto-non-trit
PSADEST007E01.b	62%	Seqüência universal de ligação ao minicírculo	<i>C. fasciculata</i>	4,0E-42	kineto-non-trit
PSADEST006D09.b	48%	ATPase-vacuolar (H+) G subunidade		3,8E-28	Intepro
PSADEST002E12.b	54%	WD-40 repetitivo	<i>L. major</i>	3,0E-85	Lmajor_NCBI
PSADEST001B03.b	53%	Proteína transportadora-like zinco	<i>L. major</i>	3,0E-27	Lmajor_NCBI
PSADEST003D01.b	49%	Proteína ZPR1 dedos de zinco, putativa	<i>L. major</i>	3,0E-55	Lmajor_NCBI

4.4 – Classificação funcional das GSS e EST de *Phytomonas serpens*

Com base no GO, foi feita uma busca por similaridade em todas as seqüências de GSS obtendo 61 (27%) seqüências com e-value inferior à $1e<10^{-5}$. Estas seqüências foram analisadas, permitindo inferir uma classificação funcional das mesmas. Entre estas, 34 seqüências foram anotadas como proteínas hipotéticas conservadas. As 27 seqüências restantes foram anotadas de acordo com os três termos do GO.

Um total de 297 (43%) seqüências de ESTs deram hit e foram classificadas de acordo com as ontologias.

Os gráficos abaixo listam os termos em que os clusters foram anotados. Estes mostram separadamente as diferentes ontologias, as anotações feitas e quantos clusters apresentaram as mesmas anotações. Para facilitar as análises, separamos as seqüências GSS das seqüências ESTs.

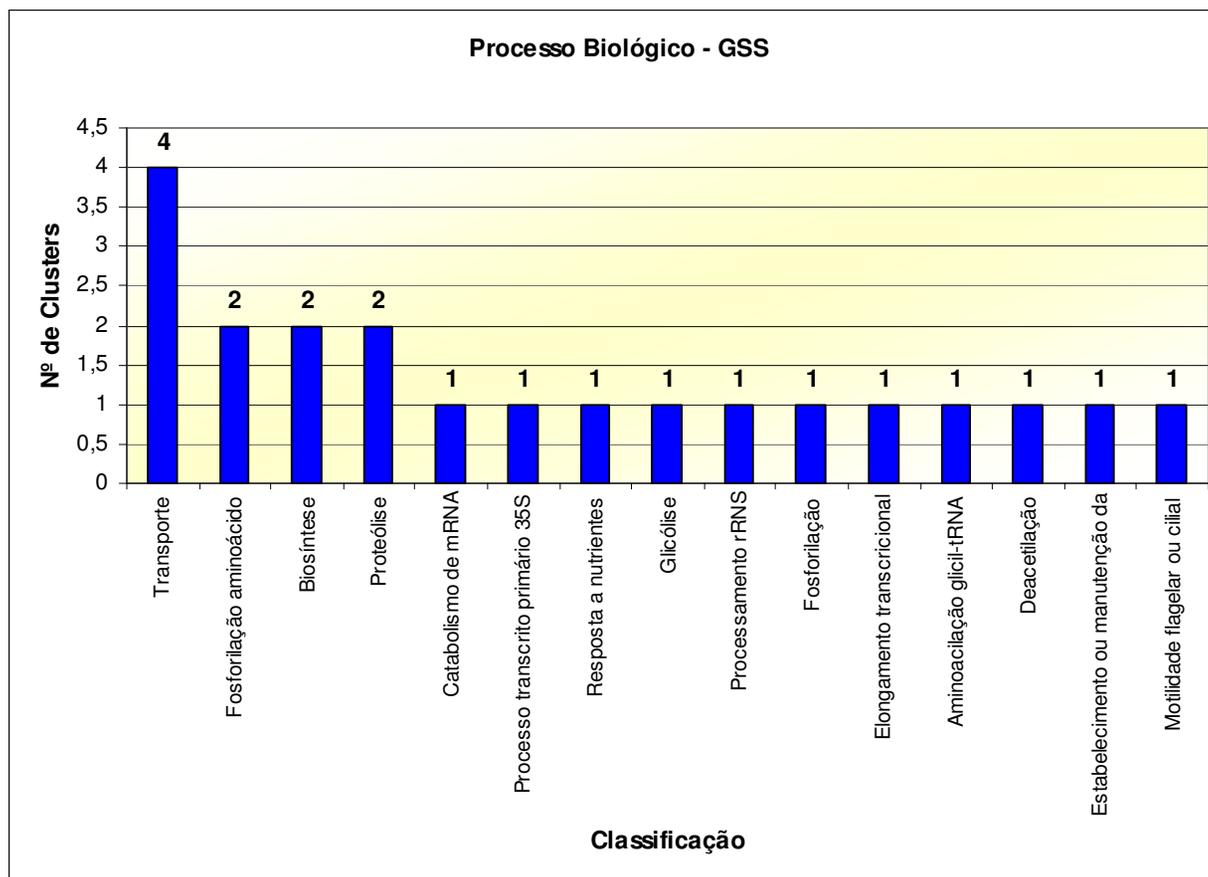


Gráfico 4.2: Relação das funções atribuídas aos clusters GSS de *Phytomonas serpens* com base nas definições do vocabulário de processo biológico do GO.

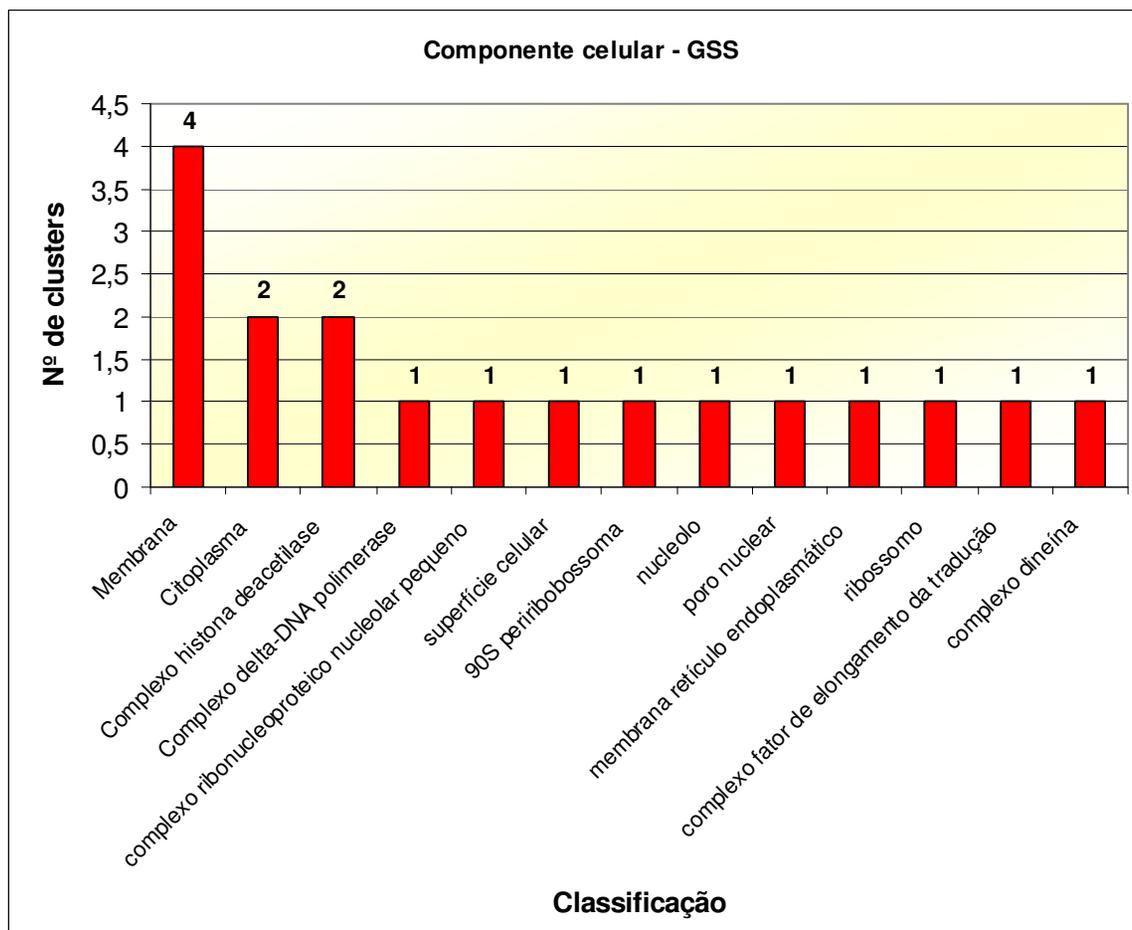


Gráfico 4.3: Relação das funções atribuídas aos clusters GSS de *Phytomonas serpens* com base nas definições do vocabulário de componente celular do GO.

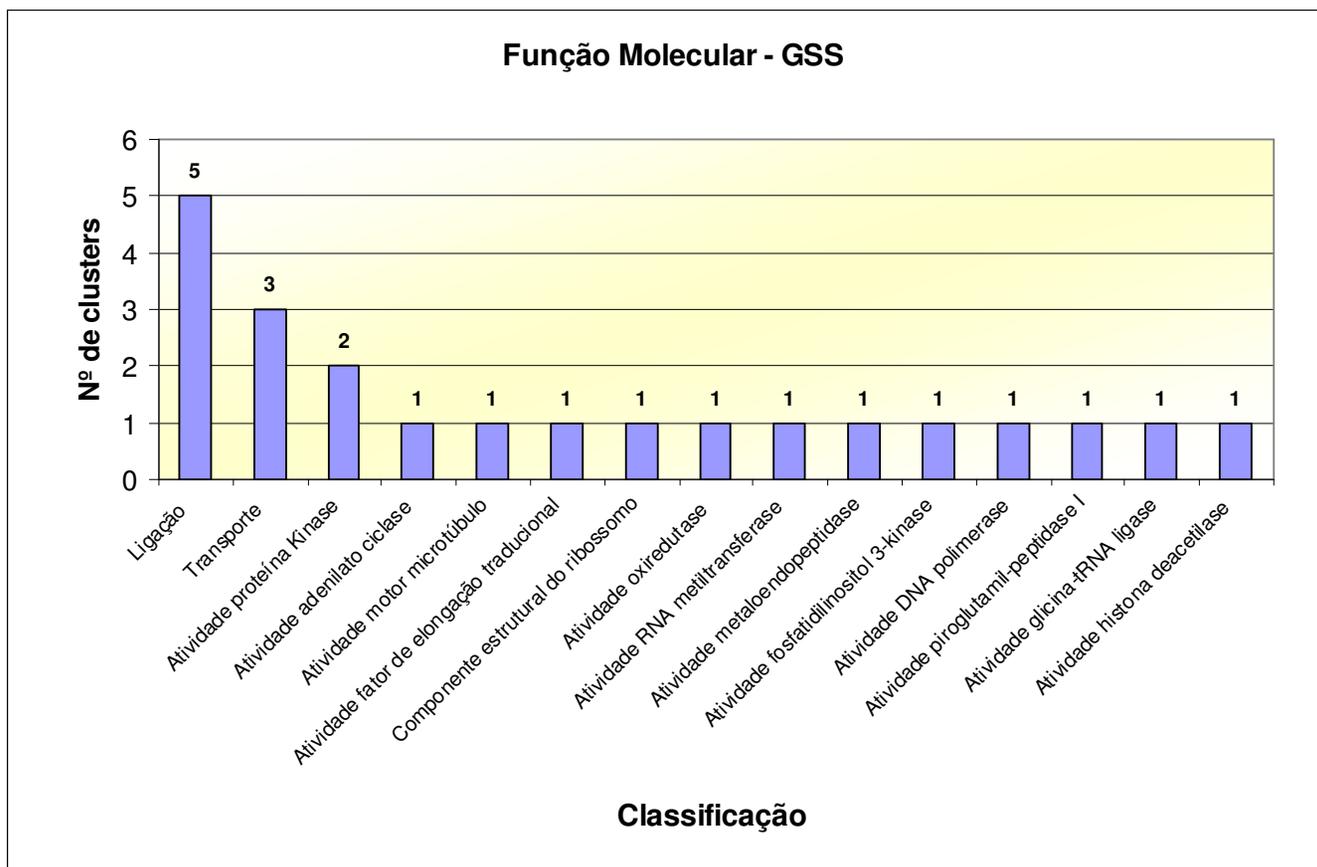


Gráfico 4.4: Relação das funções atribuídas aos clusters GSS de *Phytomonas serpens* com base nas definições do vocabulário de função molecular do GO.

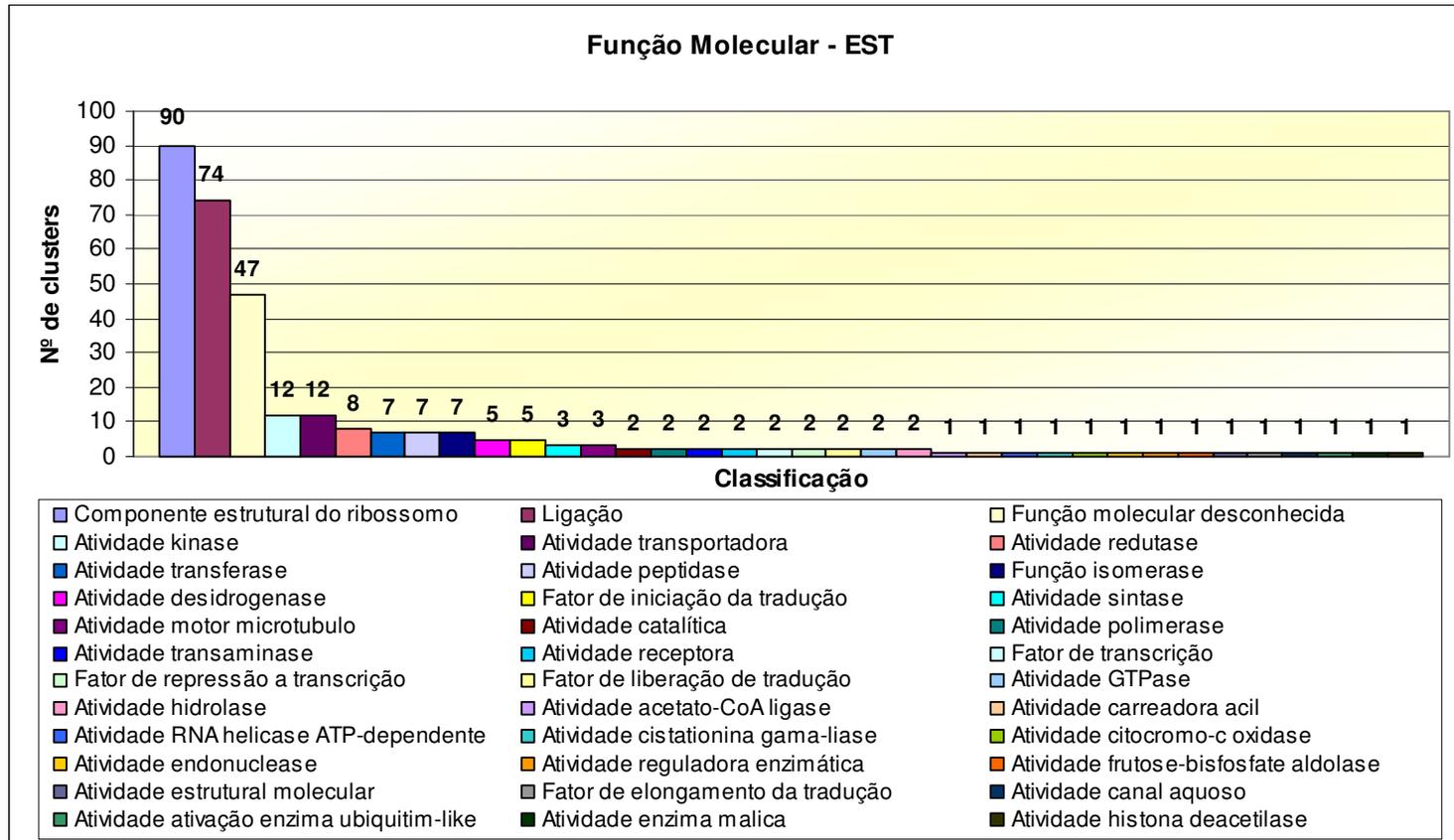


Gráfico 4.5: Relação das funções atribuídas aos clusters EST de *Phytomonas serpens* com base nas definições do vocabulário de função molecular do GO.

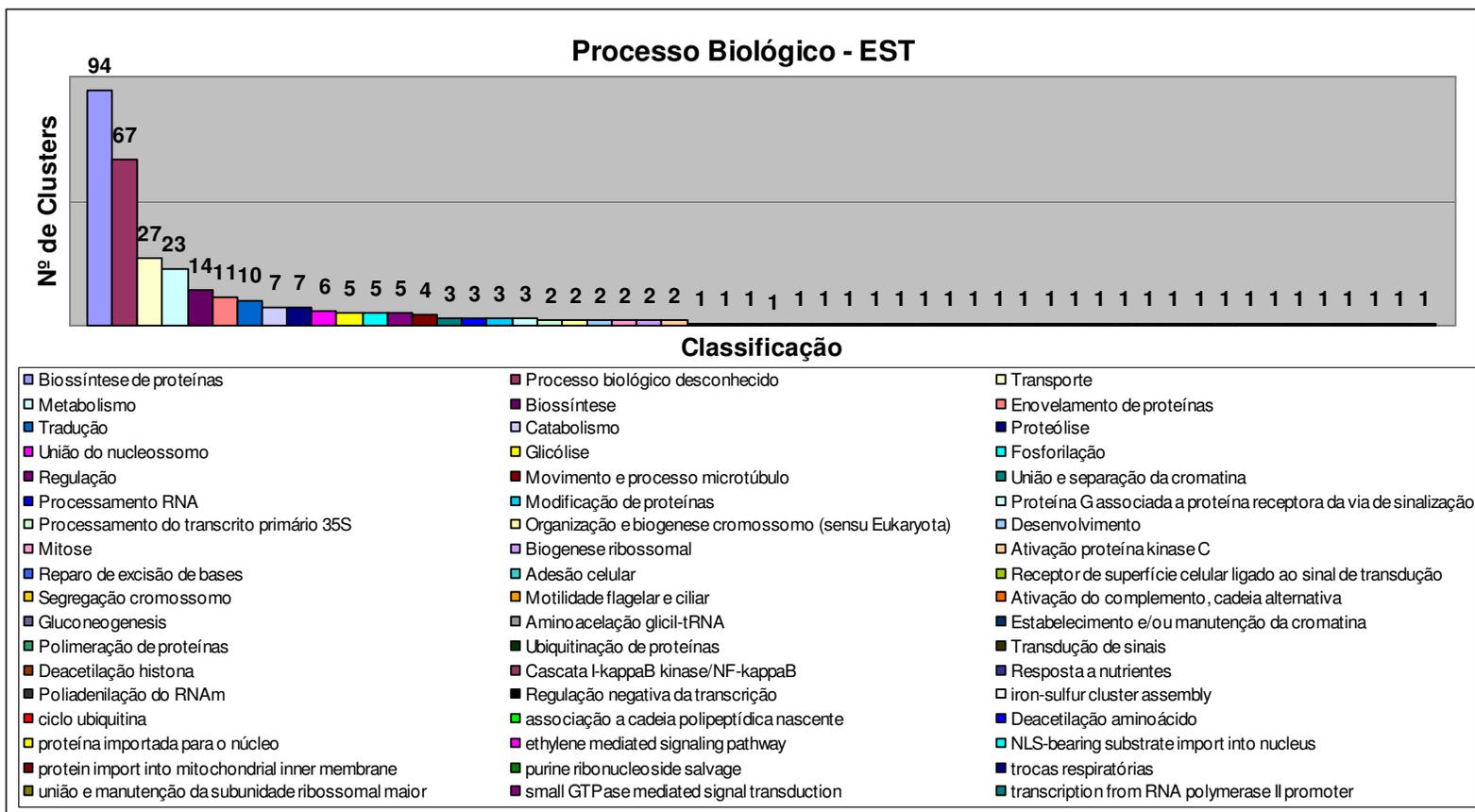


Gráfico 4.6: Relação das funções atribuídas aos clusters EST de *Phytomonas serpens* com base nas definições do vocabulário de processo biológico do GO.

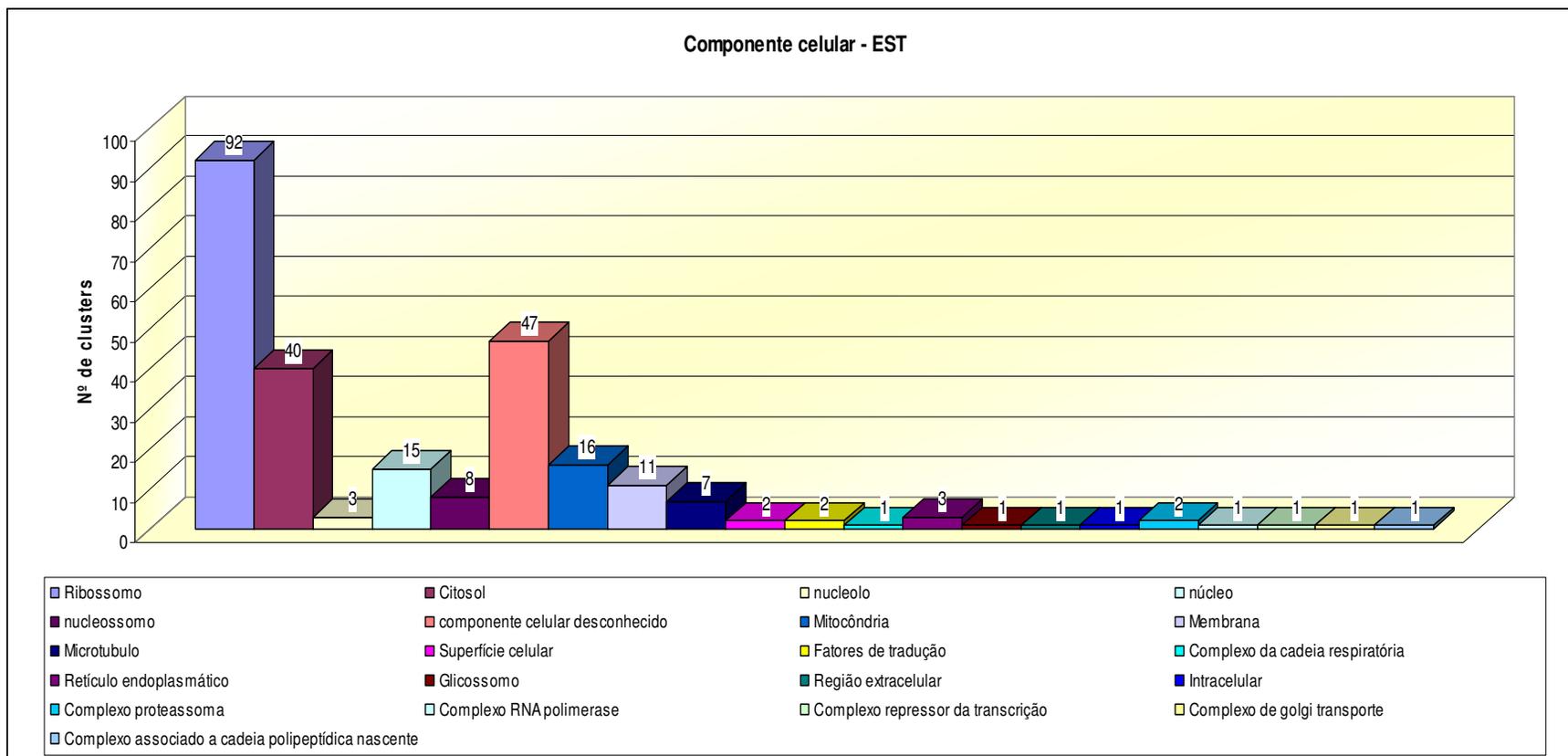


Gráfico 4.7: Relação das funções atribuídas aos clusters EST de *Phytomonas serpens* com base nas definições do vocabulário de componente celular do GO.

4.5 – Filogênia de *Phytomonas serpens*

Foram feitas inferências filogenéticas com as seqüências GSS e ESTs de *P. serpens*. Os organismos utilizados foram os pertencentes aos TriTryps- *L. major*, *T. cruzi* e *T. brucei*, alguns dos organismos pertencentes ao gênero Kinetoplastida excluindo os TriTryps como *Phytomonas sp*, *Crithidia fasciculata*, *Leptomonas sp*, Euglenozooas e seqüências do genoma ambiental do mar de Sargasso.

Árvores individuais foram feitas utilizando as seqüências correspondentes aos genes de serina-treonina kinase, fator de alongamento, gliceraldeído-3-fosfato desidrogenase, frutose-1,6-bifosfato aldolase, HSP70, beta-tubulina e adenilato ciclase (Figuras 4.4 a 4.10).

Também foi feita uma árvore filogenética concatenada, utilizando genes encontrados em nosso estudo. Utilizamos os genes de fator de alongamento, histona H2, HSP70, tubulina e adenilato ciclase. Os organismos escolhidos foram *Leishmania major*, *T. cruzi*, *T. brucei*, *L. donovani* e *E. gracilis* (Figura 4.11). Foi verificado que os organismos acima pertenciam à mesma cepa nas diferentes seqüências utilizadas.

Os valores encontrados dos sítios conservados, sítios variáveis, singletons e sítios parcimoniosos nas seqüências utilizadas nas árvores filogenéticas podem ser observados na tabela 4.6.

Para todas as árvores construímos uma duplicata sem a seqüência de *P. serpens* (dados não mostrados). Foi possível observar nas árvores de HSP70, beta-tubulina, serina-treonina kinase, adenilato ciclase e fator de alongamento uma melhora na definição da árvore quando inserimos a seqüência de *P. serpens*.

Na árvore de frutose-1,6-bifosfato aldolase (Figura 4.5) observamos que as seqüências ambientais se posicionam dentro da árvore.

Tabela 4.7- Os valores encontrados dos sítios conservados, sítios variáveis, singletons e sítios parcimoniosos nas seqüências utilizadas nas árvores filogenéticas.

Genes	Sítios conservados	Sítios variáveis	Singletons	Sítios Parcimoniosos
serina-treonina kinase (n: 473)	64	335	213	118
fator de alongamento (n: 449)	318	131	67	64
gliceraldeído-3-fosfato desidrogenase (n: 405)	174	189	53	136
frutose-1,6-bifosfato aldolase (n: 503)	105	267	78	189
HSP70 (n: 697)	350	314	184	130
beta-tubulina (n: 445)	358	87	43	43
adenilato ciclase (n: 1531)	172	1254	741	426
Concatenada (n: 6654)	2356	2823	2104	628

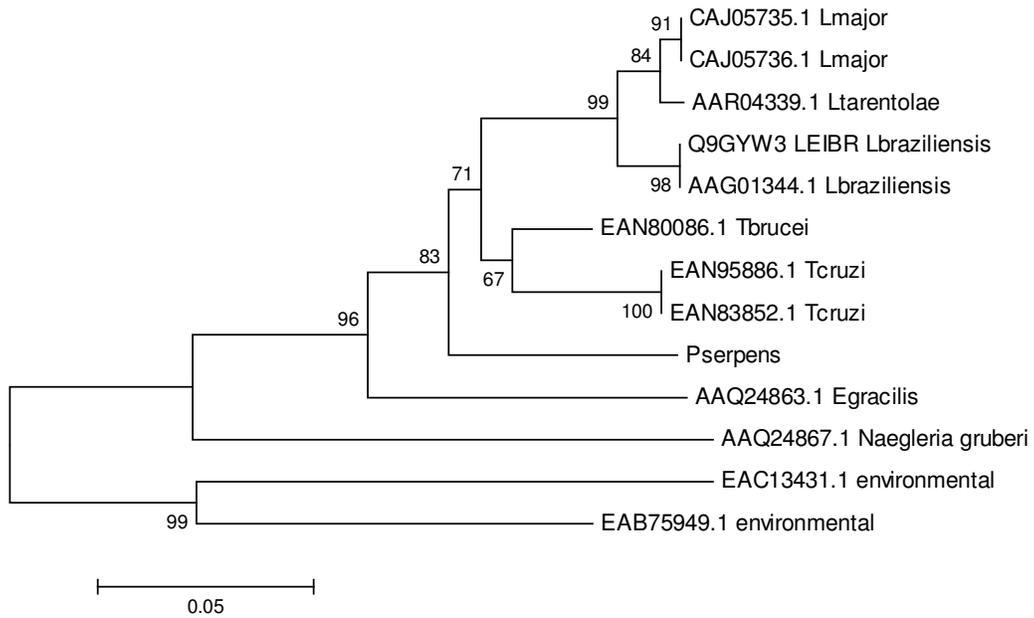


Figura 4.4 – Árvore filogenética do gene HSP 70 (cluster PSADEST007C03.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

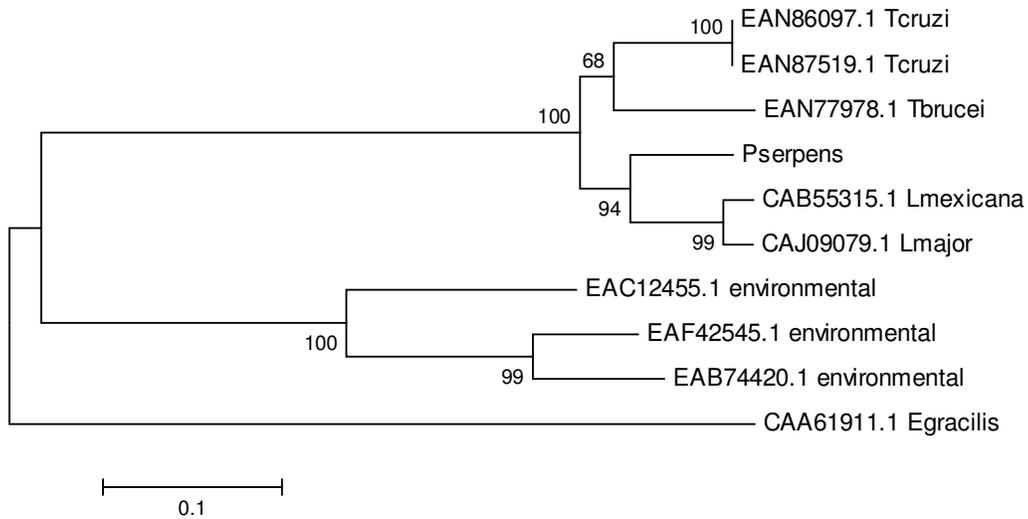


Figura 4.5 – Árvore filogenética do gene frutose-1,6-bifosfato aldolase (PSADEST006G10.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

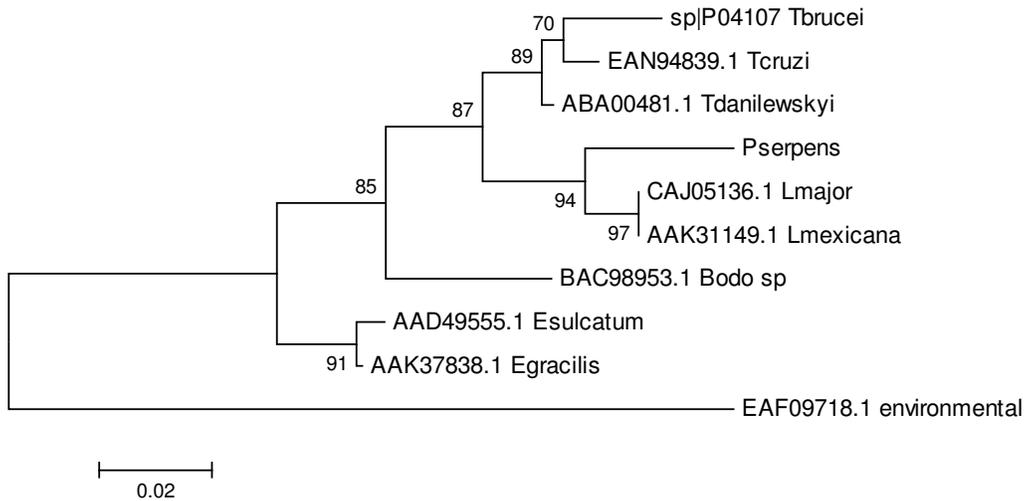


Figura 4.6 – Árvore filogenética do gene Beta-tubulina (PSADEST005H07.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

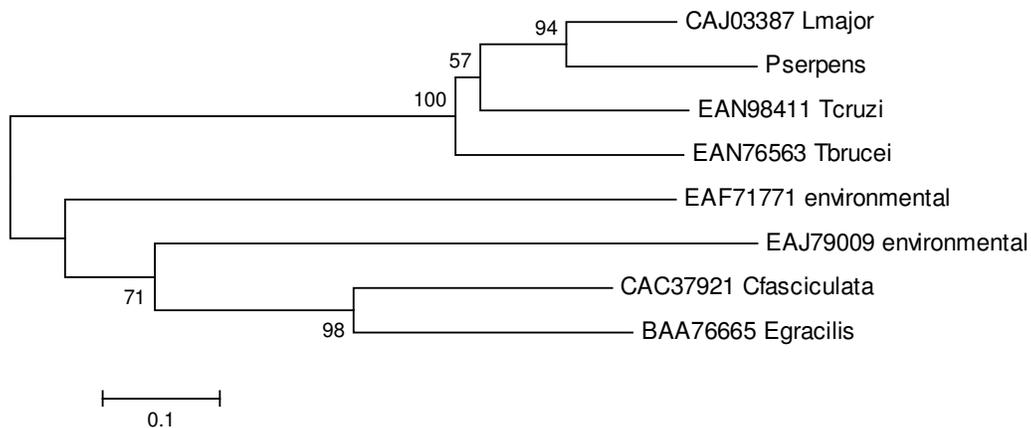


Figura 4.7 – Árvore filogenética do gene serina-treonina kinase (PSADEST001A09.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

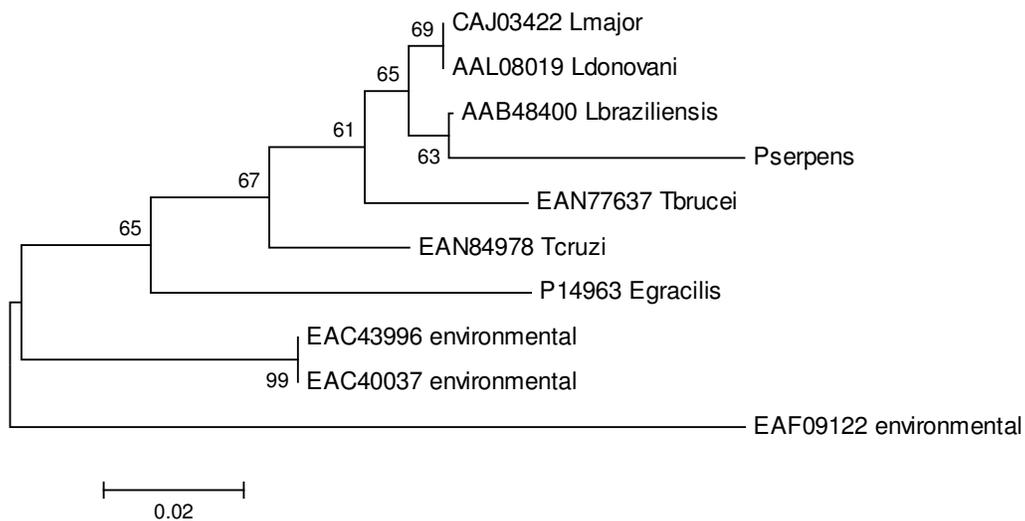


Figura 4.8 – Árvore filogenética do gene de fator de alongamento (PSADEST002G02.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

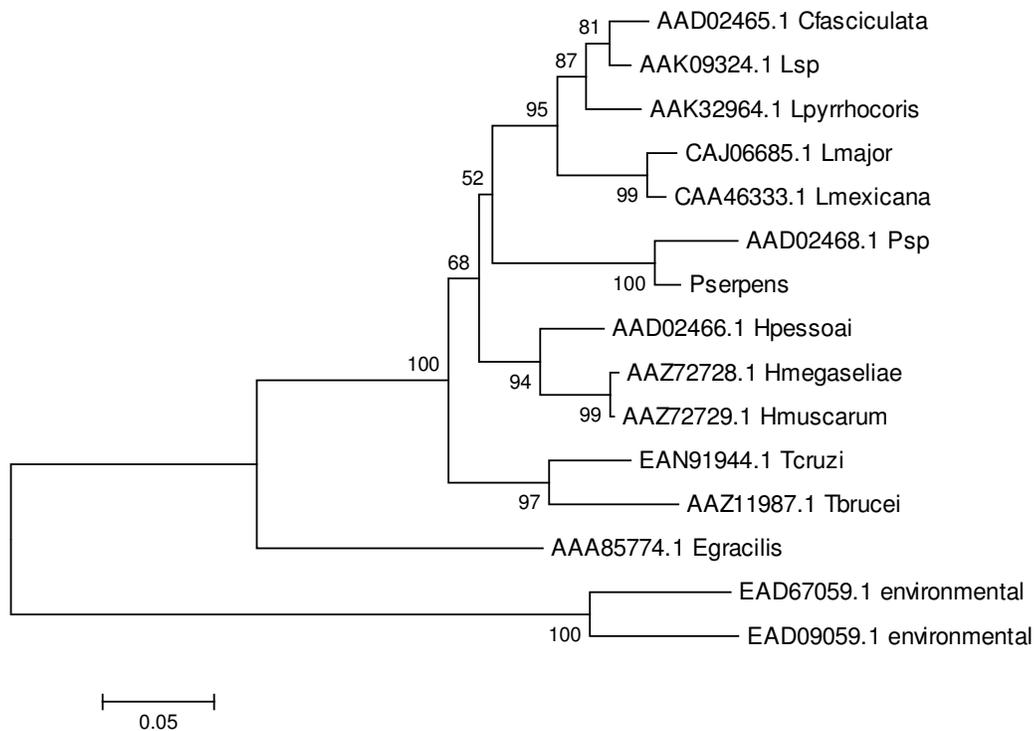


Figura 4.9 – Árvore filogenética do gene gliceraldeído-3-fosfato desidrogenase (PSADEST005H09.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

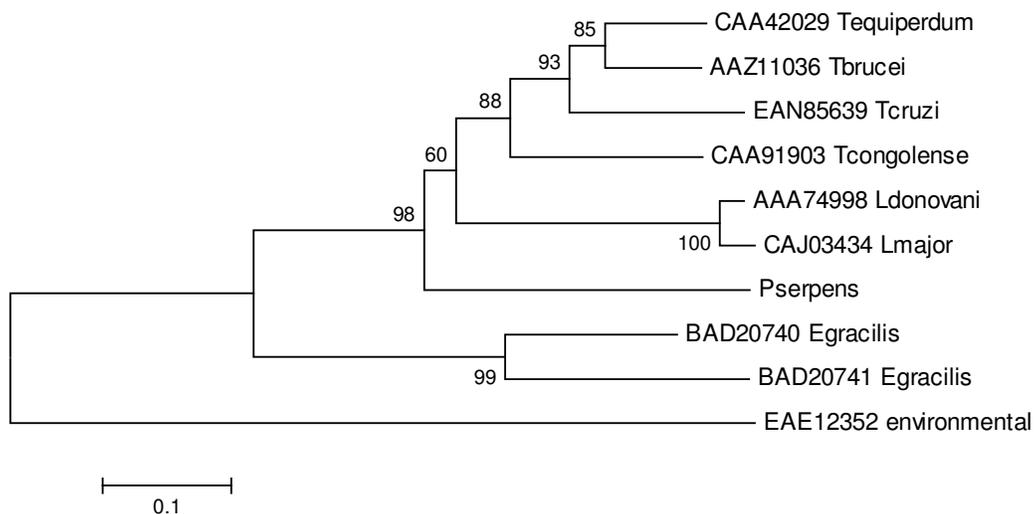


Figura 4.10 – Árvore filogenética do gene adenilato ciclase (PSADSAU009F03.b) utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

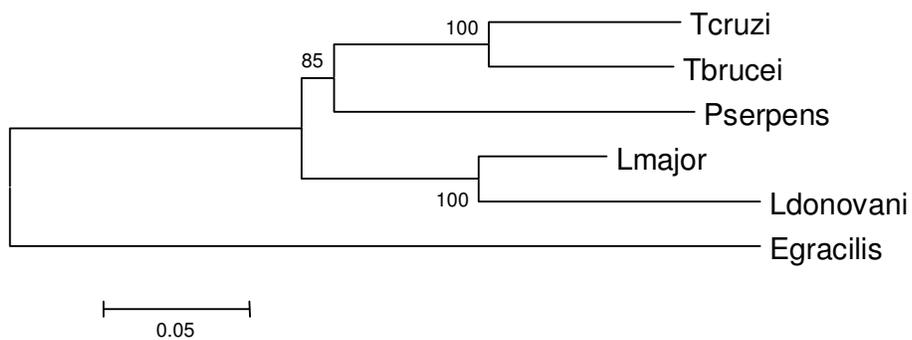


Figura 4.11- Árvore filogenética concatenada dos genes fator de alongamento, histona H2, HSP70, tubulina e adenilato ciclase utilizando o método de distância Neighbor Joining com valor de “Bootstrap” de 10000.

4.6- Similaridade utilizando o banco Taxonomy do NCBI

Na busca por similaridade as seqüências GSS e EST foram comparadas com o banco de dados Taxonomy do NCBI. O melhor hit para cada cluster com evaluate menor que $10e^{-5}$ foi encontrado com *T. cruzi* (304 hits), *T. brucei* (132 hits) e *L. major* (69 hits) (Tabela 4.7 e gráfico 4.8).

Também foram encontrados hits com mamíferos, fungos, bactérias, inseto e um apicomplexa.

Tabela 4.7: Nome dos organismos e número de hits obtidos com o “Taxonomy” quando comparamos as seqüências GSS e EST de *P. serpens* com o banco de dados refseq_protein com evaluate $1e^{-5}$.

Nome do organismo	Hit
<i>Actinobacillus succinogenes</i> 130Z	1
<i>Dictyostelium discoideum</i> AX4	1
<i>Drosophila melanogaster</i>	1
<i>Erwinia amylovora</i>	1
<i>Fucus vesiculosus</i>	1
<i>Geobacillus kaustophilus</i> HTA426	1
<i>Haemophilus somnus</i> 2336	1
<i>Homo sapiens</i>	1
<i>Pan troglodytes</i>	1
<i>Plasmodium chabaudi chabaudi</i>	1
<i>Rattus norvegicus</i>	1
<i>Rhizophyidium</i> sp. 136	1
<i>Solibacter usitatus</i> Ellin6076	1
<i>Todarodes pacificus</i>	1
<i>Yarrowia lipolytica</i> CLIB 122	1
<i>Leishmania major</i> strain Friedlin	69
<i>Trypanosoma brucei</i> TREU927	132
<i>Trypanosoma cruzi</i> strain CL Brener	304
<i>Cryptosporidium hominis</i>	3
<i>Entamoeba histolytica</i> HM-1:IMSS	5
<i>Escherichia coli</i> CFT073	3
<i>Escherichia coli</i> HS	2
<i>Schizosaccharomyces pombe</i> 972h-	2

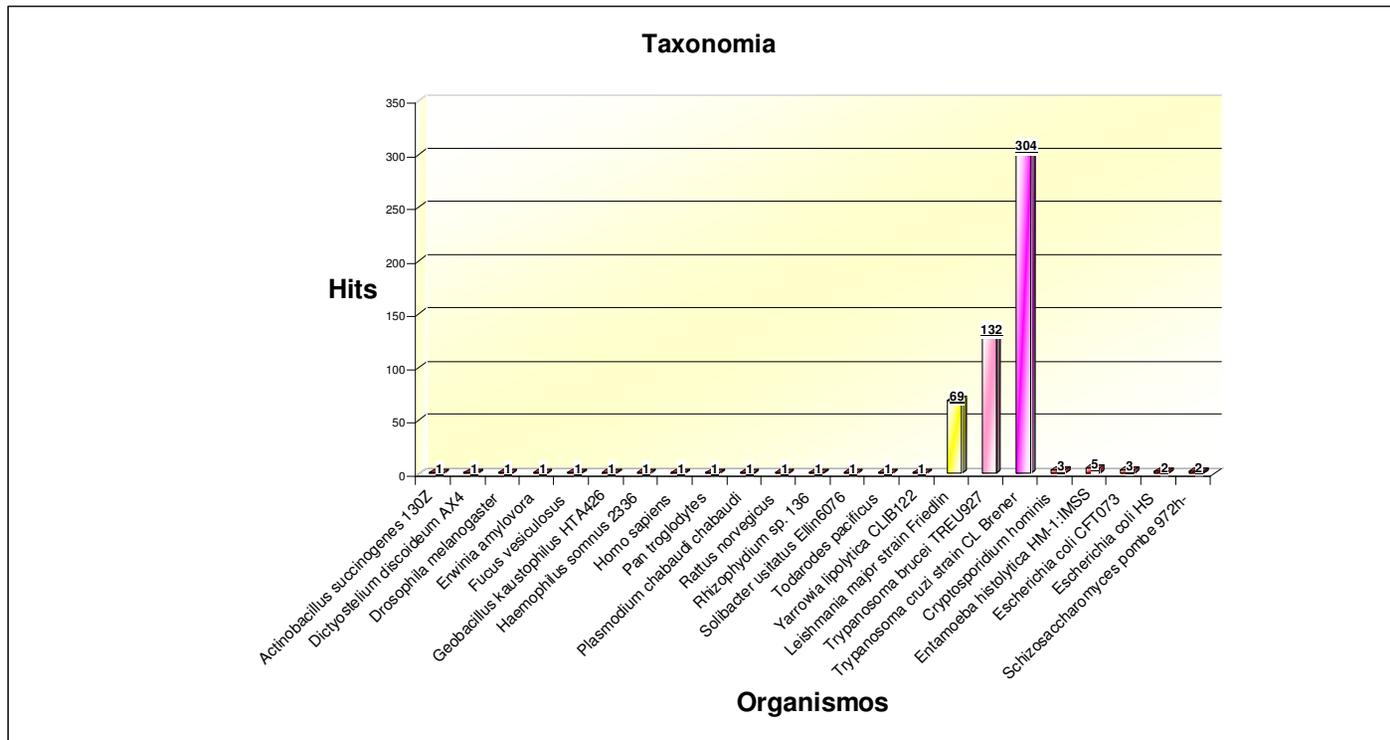


Gráfico 4.8- Representação gráfica dos dados da tabela 4.6 mostrando os nomes dos organismos e número de hits obtidos com o “Taxonomy” quando comparamos as seqüências GSS e EST de *P. serpens* com o banco de dados refseq_protein com evaluate $1e^{-5}$.

CAPÍTULO

5

DISCUSSÃO

Nas últimas décadas, técnicas moleculares foram desenvolvidas e aprimoradas gerando uma revolução na quantidade de informações sobre os organismos. Esta revolução incluiu o desenvolvimento de diferentes abordagens para o seqüenciamento de genes, entre elas os GSSs (Sánchez *et al.*, 2001) e os ESTs (Neto *et al.*, 2000), e inclusive o seqüenciamento de genomas inteiros usando WGS (Weber *et al.*, 1997). Desde então mais de 21 projetos genomas de eucariotos (<http://www.ncbi.nlm.nih.gov/genomes/leuks.cgi> - Consulta realizada dia 19/03/06), 290 de bactéria e 26 de Archeae foram completados (<http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi> - Consulta realizada dia 19/03/06). A combinação da extensa informação biológica obtida nestes projetos e a utilização de ferramentas moleculares de seqüenciamento e de análise permitiram um maior conhecimento dos genes e da biologia dos organismos vivos, e com isso, desvendar alguns dos mistérios que fazem com que estes organismos possuam características únicas.

Atualmente, as melhores e mais comuns formas de descoberta de genes são seqüenciamento de GSS e de EST. Neste estudo, utilizamos a técnica de GSS para explorar o genoma de *Phytomonas serpens*. E para ajudar nesta exploração, utilizamos as seqüências EST disponíveis no GenBank.

Neste projeto foram geradas um total de 829 seqüências GSS, onde 379 apresentaram alta qualidade e foram utilizadas em nosso estudo. Estas seqüências GSS de alta qualidade mais as seqüências EST equivalem a 166 kb. O tamanho do genoma de *P. serpens* ainda não foi estimado, portanto não podemos calcular a porcentagem do genoma estudado. Contudo, estes números são significativos já que até o presente momento não há nenhuma seqüência GSS deste organismo depositada no GenBank.

A média do conteúdo de G+C foi de 48% nas seqüências de GSS e EST, valor próximo aos dos genomas de *T. brucei*, onde o valor no genoma é de 46,4% (Berriman *et al.*, 2005), *T. cruzi* cujo valor é de 51% (El-Sayed *et al.*, 2005), mas inferior ao genoma de *L. major*, que possui 59,7% (Ivens *et al.*, 2005). O alto valor do conteúdo de GC pode refletir uma alta densidade gênica, como foi observado em *T. cruzi* (Aguero *et*

al., 2000), mas não representa uma regra, pois *Plasmodium falciparum* possui um genoma extremamente pobre em GC, de aproximadamente 18% (Musto *et al.*, 1999).

O GLIMMER encontrou 20 genes hipotéticos na biblioteca *Sau3A*. O tamanho médio foi de 331 pb e com a média do conteúdo de G+C de aproximadamente 53%. Também foram encontrados genes hipotéticos nas seqüências ESTs, sendo um total de 131 com média do conteúdo de G+C de 55%. Comparando com os genomas dos outros tripanosomatídeos, onde em *T. brucei* foi verificado que a média do comprimento de CDS é de 1.242 bp (Berriman *et al.*, 2005), em *L. major* é de 1.407 bp (Ivens *et al.*, 2005) e em *T. cruzi* os CDS tem comprimento médio 1.152 bp (El-Sayed *et al.*, 2005b), verificamos que a média obtida em nosso trabalho é mais baixa mas isto não exclui a possibilidade de *P. serpens* possuir genes maiores.

Das 221 clusters GSS de *P. serpens*, 108 (49%) apresentaram hit com e-value $<10^{-5}$ quando comparadas com bancos de dados públicos. Em um estudo de El-Sayed e Donelson (1997) com GSS de *T. brucei*, de 360 seqüências, 113 (33%) tiveram hits com bancos de dados. Em *T. cruzi*, de 3.724 GSS, 32,5% apresentaram hit (Aguero *et al.*, 2000). Já em *T. vivax*, de 455 GSS, 108 seqüências obtiveram hits (Guerreiro *et al.*, 2005).

Das 686 clusters EST, 496 (72%) obtiveram hit com evalue $<10^{-5}$ com os bancos de dados. No genoma de *T. cruzi*, do total de 797 seqüências EST, apenas 30% apresentou similaridade (Brandão *et al.*, 1997).

As 190 (28%) clusters de EST que não apresentaram hits com nenhum banco de dados serão anotados como proteínas hipotéticas e serão utilizados em estudos posteriores para procura de genes espécie específicos.

Os 113 (33%) clusters GSS que não apresentaram similaridade com nenhuma seqüência depositada nos bancos de dados públicos representam, até o momento, possíveis “genes órfãos”. Estas seqüências serão analisadas mais detalhadamente durante o processo de “re-anotação”, onde serão realizadas análises de motivos e detecção de homologias distantes. Caso seja confirmada a veracidade destas seqüências como “genes órfãos”, estas podem representar genes exclusivos da espécie, sendo utilizados como potenciais marcadores moleculares.

Usando as ontologias do GO, 358 seqüências GSS e EST puderam receber algum tipo de anotação.

Dois clusters (PSADEST007E02.b e PSADSAU005B11.b) apresentaram similaridade com a ubiquitina, uma proteína envolvida nos processos de modificação

pós-traducional de proteínas, sendo extremamente conservada de protozoários a vertebrados (Burger & Seth, 2004). Esta proteína já foi descrita anteriormente em tripanosmatídeos, sendo utilizada no diagnóstico diferencial da doença de chagas e de leishmaniose (Telles *et al.*, 2003).

O fator de alongamento é uma proteína que se associa ao ribossomo durante a fase de alongamento no processo de síntese de proteínas e cataliza a formação de ligação acil entre os resíduos de aminoácido e a cadeia peptídica. O gene que codifica para esta proteína já foi utilizado anteriormente em estudos filogenéticos, por exemplo, na construção de uma árvore universal de eucariotos (Baldauf *et al.*, 1995), estudos moleculares em artrópodes (Regier *et al.*, 1997), entre outros.

Em relação a proteínas do citoesqueleto, encontramos duas seqüências similares as Kinesinas (PSADEST003H10.b e PSADEST005B08.b) em nosso estudo. Os genomas dos tri-tryps codificam para aproximadamente 40 kinesinas, sendo proteínas pouco conservadas, estando associadas aos microtúbulos (Berriman *et al.*, 2005). Outras proteínas encontradas foram alfa-tubulina (PSADEST007H03.b) e beta-tubulina (PSADEST005H07.b e PSADEST007A02.b), sendo que no genoma dos tri-tryps são encontradas 6 formas desta proteína (Berriman *et al.*, 2005).

Tripanosomatídeos contém varias cópias de histonas (H1, H2B, H2B, H3 e H4), as quais participam do empacotamento de DNA cromossomal em nucleossomos nos eucariotos. A maioria destes genes se encontram clusterizados em “tandem” em *T. cruzi* e *T. brucei*, sendo encontrados em cópias únicas em *L. major* (Ivens *et al.*, 2005). Em nosso estudo encontramos seqüências similares aos 5 tipos de histonas – H1 (PSADEST006H01.b), H2A (PSADEST006C11.b, PSADEST006F06.b), H2B – (PSADEST007E06.b, PSADEST007G09.b), H3 – (PSADEST007F01.b, PSADEST006A02.b) e H4 (PSADEST003B06.b, PSADEST007H09.b, PSADEST002D11.b, PSADEST005H12.b)

Encontramos também Peptidase_M8: Leishmanolysin (PSADSAU006A06.g) uma enzima encontrada em eucariotos, incluindo *Leishmania* e protozoários relacionados. A endopeptidase é a proteína mais abundante na superfície celular durante o estágio promastigota do parasito (Rawlings *et al.*, 1995).

Outra proteína encontrada foi a “WD-40 repeats” (PSADEST002E12.b) também conhecida como WD ou “beta-transducin repeats”. As proteínas “WD-40 repeats” são uma grande família achadas em todos os eucariotos e estão implicadas de funções desde

transdução de sinais e regulação da transcrição até o controle do ciclo celular e apoptose (Li *et al.*, 2001; Smith *et al.*, 1999).

Encontramos em *P. serpens* um cluster similar à enzima Tripanotiona redutase (PSADEST001G06.b), que foi anteriormente purificada e caracterizada de *Leishmania* e tripanossomas da África e América do Sul (Sullivan *et al.*, 1991). Estas enzimas possuem propriedades físicas, mecânicas e cinéticas similares e são membros da família “flavoproteína dissulfide oxiredutase”. A tripanotiona é uma versão análoga dos parasitos para a glutatona, já que esta enzima é equivalente a glutatona redutase (Lantwin *et al.*, 1994).

O cluster PSADEST003E07.b é similar a S-adenosil-L-metionina-C-24-delta-esterol-metiltransferase (SCMT ou ERG6). Esta enzima está ausente na membrana de promastigotas de *L. donovani* resistentes a amphotericin B [(AmB)-resistente], sugerindo que a SCMT não é funcional ou não é expressa em parasitos AMB-resistentes (Pourshafie *et al.*, 2004).

O cluster PSADSAU009F01.g é similar ao antígeno nuclear de proliferação celular (PCNA) (Matsumoto *et al.*, 1987). Também conhecida como ciclina, é uma proteína nuclear ácida “non-histona” que tem um papel muito importante no controle da replicação do DNA em eucariotos (Hata *et al.*, 1992). Esta proteína age como co-fator para a DNA polimerase delta (Travali *et al.*, 1989). A seqüência de PCNA é muito conservada entre plantas e animais, indicando uma forte pressão seletiva para a conservação da estrutura, e sugerindo que este tipo de mecanismo de replicação do DNA é conservado em eucariotos (Suzuka *et al.*, 1991).

O cluster PSADEST005C04.b é similar ao Gliceraldeído-3-fosfato desidrogenase (GAPDH) que tem um importante papel na glicólise e gluconeogênese. Esta enzima cataliza a fosforilação oxidativa do D-gliceraldeído-3-fosfato em 1,3-difosfoglicerato na presença de NADP e fosfato inorgânico (Pavão *et al.*, 2002). Estudos estruturais das enzimas de *T. brucei* (Velieux *et al.*, 1993) e *Leishmania mexicana* (Kim *et al.*, 1995) mostraram diferenças no sítio de ligação do cofator NADP quando comparado com o homólogo humano, o que levou ao desenho e a síntese de vários inibidores específicos para estes sítios (Callens *et al.*, 1995).

O cluster PSADSAU006D09.g é similar ao Piroglutamil-peptidase I, uma enzima que participa do clã das cisteínas peptidases família C15 e foi caracterizada em *Trypanosoma brucei*, podendo ser utilizada como um possível alvo terapêutico para o tratamento contra este parasito (Morty *et al.*, 2006).

Devido à ausência de registros fósseis para a inferência da filogenia dos tripanosomatídeos, esta era até pouco tempo mantida em especulação. Atualmente, com os métodos filogenéticos associados às informações obtidas nos projetos genomas, podemos inferir a história desses organismos pelas informações descritas em seus genes.

A maior parte dos estudos filogenéticos publicados até o presente momento com os organismos do gênero *Phytomonas* foram realizados utilizando-se os genes RNAr de *P. serpens* (Maslov *et al.*, 1995), parasitos de floema da cepa HART1 (Marché *et al.*, 1995) e de látex (Marché *et al.*, 1995, Fernandes *et al.*, 1993). Estes estudos concluíram que o gênero *Phytomonas* é monofilético.

Em nosso estudo utilizamos seqüências dos TriTryps, bem como outras seqüências dos Kinetoplastida, visando identificar seqüências similares com os clusters GSS e EST gerados neste estudo. Para trazer mais robustez para nossas árvores, incluímos também os Euglenozooas, pois estes parecem ter os mesmos ancestrais dos kinetoplastidas (El-Sayed *et al.*, 2005). Incluímos também em nosso estudo as seqüências do genoma ambiental do mar de Sargasso (Venter *et al.*, 2004), com o intuito de testar se alguns destes organismos poderiam pertencer à ordem dos Kinetoplastida. Infelizmente, apesar da similaridade que estas seqüências do genoma ambiental apresentaram quando comparadas com as seqüências GSS e EST de *P. serpens*, a grande maioria não agrupou dentro da clade dos Kinetoplastida, porém não podemos descartar que estes organismos não pertençam aos Euglenozoa. Na árvore filogenética construída com a seqüência correspondente ao gene de frutose-1,6-bifosfato aldolase, as seqüências ambientais usadas se agruparam no clado dos Euglenozoa, demonstrando uma possível relação entre esses organismos. As inferências filogenéticas realizadas anteriormente com as seqüências ambientais do mar de Sargasso foram realizadas sem as seqüências dos Kinetoplastida ou Euglenozoa, utilizando apenas bactérias (Venter *et al.*, 2004; LeCleir *et al.*, 2004).

A filogenia obtida com a fusão de genes conservados gera resultados mais robustos para as análises das seqüências (Martin *et al.*, 2002; Oriane *et al.*, 2002). A árvore concatenada mostra que *P. serpens* é filogeneticamente mais próximo de *T. cruzi* do que outros tripanosomatídeos. Por outro lado, as árvores com genes individuais mostram que *P. serpens* é mais próximo filogeneticamente de *L. major*. Considerando que as árvores de genes individuais mostram a reconstrução da filogenia dos genes

usados, e não necessariamente a filogenia da espécie (Gadagkar *et al.*, 2005) acreditamos que a árvore com genes concatenados esteja mostrando a filogenia correta.

Estudos sobre as diferenças entre a árvore de genes concatenados e árvores de genes individuais foram realizados anteriormente. Olmstead e Sweere (1994) acharam que a combinação de diferentes conjuntos de dados produzem uma filogenia mais resolvida do que quando é utilizado um único gene. Rokas *et al.* (2003) observou que a análise de 20 genes concatenados derivados de um grupo de dados forneceu uma árvore única, completamente resolvida, com suporte máximo. O estudo de Gadagkar *et al.* (2005) comparou as metodologias de inferência filogenética mostrando uma maior acurácia (>95%) da árvore concatenada.

Diferenças entre os resultados das árvores de genes concatenados e as de genes individuais foram observadas anteriormente (Charles *et al.*, 2005, Miya *et al.*, 2000, Pissani *et al.*, 2004). A incongruência encontrada entre os resultados das árvores de genes individuais e as árvores de genes concatenados é devido a um número de fatores biológicos que estão relacionados à evolução dos genes e as análises filogenéticas, como a taxa de evolução, o tamanho das seqüências, a taxa de transição-transversão e conteúdo de G+C (Charles *et al.*, 2005). É importante salientar que a concatenação de genes para inferências filogenéticas funciona melhor pelo fato de não precisar levar em consideração os fatores listados acima (Gadagkar *et al.*, 2005).

Estudos prévios apontaram problemas com a inserção de seqüências de *P. serpens* nas árvores individuais de genes (Wiemer *et al.*, 1995; Marche *et al.*, 1995). O estudo realizado por Stevens & Gibson (1999) para inferir a evolução dos tripanosomatídeos mostrou que quando se inseria a seqüência 18S rRNA na análise, reduzia significativamente a definição filogenética dentro no nível de *Trypanosoma*. Em nosso estudo verificamos que o mesmo acontece na árvore de beta-tubulina e adenilato ciclase, onde quando inserimos a seqüência de *P. serpens*, houve uma perda na definição do nível de *Trypanosoma*. Entretanto, em 3 árvores (HSP70, serina-treonina kinase e fator de alongamento) observamos um aumento na robustez quando inserimos a seqüência de *P. serpens*.

CAPÍTULO

6

CONCLUSÃO

1. A metodologia GSS mostrou ser uma boa técnica para descobrimento de genes em DNA genômico.
2. Os 829 clones da biblioteca de DNA genômico de *Phytomonas serpens* cepa 9T (CT-IOC-174) obtendo-se um total de 379 seqüências GSS de *P. serpens* de alta qualidade seqüenciados significaram um bom avanço para conhecimento deste organismo, já que antes do presente estudo, não havia seqüências deste tipo registradas no GenBank.
3. Os clusters foram identificados de acordo com sua função quando comparados com o banco de dados Gene Ontology (GO), onde foram classificadas de acordo com as três ontologias. Isto representa uma análise inédita no genoma de *P. serpens*, já que em estudos anteriores nenhum cluster foi analisado em relação a este banco de dados.
4. Foram encontradas com o GLIMMER 20 genes hipotéticos nas seqüências de GSS.
5. Uma comparação de similaridade usando o “taxonomy” foi realizada comparando as seqüências de *P. serpens* com Blast e o banco de dados do taxonomy do NCBI demonstrando que o genoma de *P. serpens* é mais similar ao genoma de *T. cruzi*.
6. As inferências filogenéticas realizadas com genes concatenados mostraram que *P. serpens* é mais próximo filogeneticamente de *T. cruzi*. As inferências filogenéticas com genes individuais apontaram para *L. major*. Com base na literatura, inferimos que as análises da árvore de genes concatenados é mais correta.
7. Nos estudos filogenéticos foram inseridas seqüências do genoma ambiental, para verificar se estas seqüências pertencem à família Trypanosomatidae, o que infelizmente não se comprovou, contudo, a seqüência presente na árvore de frutose-1,6-bifosfato aldolase pode ser um euglenozoa.

8. GARS (Genomic Analysis Resources for Sequences Annotation), o sistema de anotação utilizada em nosso projeto provou ser um eficiente e rápido método para a validação de seqüências, sendo de grande ajuda para o êxito deste estudo.

CAPÍTULO

7

BIBLIOGRAFIA

Abreu Filho BA, Dias Filho BP, Vermelho AB, Jankevicius SI, Jankevicius JV, dos Santos RL. Surface component characterization as taxonomic tools for *Phytomonas* spp identification. *Parasitol Res.* 2001 Feb;87(2):138-44.

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merrill CR, Wu A, Olde B, Moreno RF. Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* 1991; 252: 1651-1656.

Agüero, F, Verdun, RE, Frasch, ACC, Sánchez, DO. A random sequencing approach for the analysis of the *Trypanosoma cruzi* genome: general structure, large gene and repetitive DNA families, and gene discovery. *Genome Res.* 2000; Dec;10(12):1996-2005.

Ahomadegbe JC, Dollet M, Coulaud D, Gargani D, Riou G. Kinetoplast DNA permits characterization of pathogenic plant trypanosomes of economic importance. *Biol. Cell.* 1990; 70:167-76.

Akopyants NS, Clifton SW, Martin J, Pape D, Wylie T, Li L, Kissinger JC, Roos DS, Beverley SM. A survey of the *Leishmania major* Friedlin strain V1 genome by shotgun sequencing: a resource for DNA microarrays and expression profiling. *Mol. Biochem. Parasitol* 2001; 113: 337–340.

Almeida LG, Paixão R, Souza RC, Costa GC, Barrientos FJ, Santos MT, Almeida DF, Vasconcelos AT. A System for Automated Bacterial (genome) Integrated Annotation--SABIA. *Bioinformatics.* 2004; Nov 1;20(16):2832-3.

Altschul SF, Madden TL, Schaffer AA, *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; Sep 25(17): 3389-3402.

Aragão, HB. Sur un flagelé du latex de *Manihot palmata*: *Phytomonas franca* sp. *C. R. Soc. Biol.*, 1927; 97:27.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 2000 May; 25 (1):25-9.

Aziz A e Blaxter M. Hox gene evolution in nematodes: novelty conserved. *Current Opinion in Gen & Develop* 2003; 13: 593-598.

Baldauf SL, Palmer JD, Doolittle WF. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. *Proc. Natl. Acad. Sci. USA.* 1996; July, Vol. 93, pp. 7749-7754.

Baxevanis AD, Oullette BFF. *Bioinformatics: A practical guide to the analysis of genes and proteins* 2nd ed. USA, Wiley-Interscience; 2001.

Benício A. Abreu Filho, Benedito P. Dias Filho, Alane B. Vermelho, Shiduca I. Jankevicius, José V. Jankevicius, Regina L. dos Santos. Surface component characterization as taxonomic tools for *Phytomonas spp.* identification. *Parasitol. Res.* 2001; 87: 138-144.

Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, *et al.* The genome of the African trypanosome *Trypanosoma brucei*. *Science*. 2005; Jul 15;309(5733):416-22.

Borodovsky, M. and McIninch J. GENMARK: Parallel Gene Recognition for both DNA Strands. *Comput. Chem.* 1993; 17, 123-133.

Brandao A, Urmenyi T, Rondinelli E, Gonzalez A, de Miranda AB, Degraive W. TcruziDB, an integrated database, and the WWW information server for the *Trypanosoma cruzi* genome project. *Mem Inst Oswaldo Cruz*. 1997 Nov-Dec; 92(6):805-9.

Bregano JW, Picao RC, Graca VK, Menolli RA, Itow Jankevicius S, Pinge Filho P, Jankevicius JV. *Phytomonas serpens*, a tomato parasite, shares antigens with *Trypanosoma cruzi* that are recognized by human sera and induce protective immunity in mice. *FEMS Immunol. Med. Microbiol.* 2003; Dec 5;39(3):257-64.

Burger AM, Seth AK. The ubiquitin-mediated protein degradation pathway in cancer: therapeutic implications. *Eur J Cancer*. 2004 Oct;40(15):2217-29.

Callens M, Hannaert V. The rational design of trypanocidal drugs: selective inhibition of the glyceraldehyde-3-phosphate dehydrogenase in Trypanosomatidae. *Ann. Trop. Med. Parasitol.* 1995; 89 (Suppl. 1), 23-30.

Camargo EP. *Phytomonas* and other trypanosomatids parasites of plants and fruit. *Adv Parasitol.* 1999; 42: 29-112.

Camargo EP, Sbravate C, Teixeira MMG, Uliana SR, Soares MBM, Affonso HT, Floter-Winter L. Ribosomal DNA restriction analysis and synthetic oligonucleotide probing in the identification of genera of lower trypanosomatids. *J. Parasitol.* 1992; 78:40-8.

Camargo, EP, Kastelain, P, Roitman, I. Trypanosomatid parasites of plants (*Phytomonas*). Parasitol. Today. 1990; Jan;6(1):22-5.

Camargo EP, Silva, S, Roitman, I, Souza, W, Jankevicius, JV, Dollet, M. Enzymes of Ornithine-Arginine metabolism in trypanosomatids of the genus *Phytomonas*. J. Protozool. 1987; 34(4):439-441.

Carlton JMR, Muller R, Yowell CA, Fluegge MR, Sturrock KA, Pritt JR, Vargas-Serrato E, Galinski MR, Barnwell JW, Mulder N, Kanapin A, Cawley SE, Hide WA, Dame JB. Profiling the malaria genome: a gene survey of three species of malaria parasite with comparison to other apicomplexan species. Mol. Biochem. Parasitol. 2001; 118: 201–210.

Dávila AM, Majiwa PA, Grisard EC, Aksoy S, Melville SE. Comparative genomics to uncover the secrets of tsetse and livestock-infective trypanosomes. Trends Parasitol. 2003; 19(10):436-439.

Dávila AM, Lorenzini DM, Mendes PN, Satake TS, Sousa GR, Campos LM, Mazzoni CJ, Wagner G, Pires PF, Grisard EC, Cavalcanti MC, Campos ML. GARSA: Genomic Analysis Resources for Sequence Annotation. Bioinformatics. 2005; Oct 6.

Degrave W, Levin MJ, da Silveira JF, Morel CM. Parasite genome projects and the *Trypanosoma cruzi* genome initiative. Mem. Inst. Oswaldo Cruz. 1997; 92: 859-862.

Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 1999; Dec 27(23): 4636-4641.

Dessen P, Zagulski M, Gromadka R, Plattner H, Kissmehl R, Meyer E, Bétermier M, Schultz JE, Linder JU, Pearlman RE, Kung C, Forney J, Satir BH, Van Houten JL, Keller AM, Froissard M, Sperling L, Cohen J. *Paramecium* genome survey: a pilot project. Trends Genet. 2001; 17: 306-308.

Neto Dias E, Correa RG, Verjovski-Almeida S, Briones MR, Nagai MA, da Silva W Jr, Zago MA, Bordin S, Costa FF, Goldman GH, Carvalho AF, Matsukuma A, Baia GS, Simpson DH, Brunstein A, de Oliveira PS, Bucher P, Jongeneel CV, O'Hare MJ, Soares F, Brentani RR, Reis LF, de Souza SJ, Simpson AJ. Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc Natl Acad Sci U S A*. 2000 Mar 28;97(7):3491-6.

Do, C.B., Mahabhashyam, M.S.P., Brudno, M., and Batzoglou, S. PROBCONS: Probabilistic Consistency-based Multiple Sequence Alignment. *Genome Res*. 2005; 15: 330-340.

El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN *et al*. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science*. 2005; Jul 15; 309(5733):409-15.

El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, Caler E, *et al*. Comparative genomics of trypanosomatid parasitic protozoa. *Science*. 2005 Jul 15;309(5733):404-9.

El-Sayed NM, Hegdea P, Quackenbusha J, Melvilleb SE, Donelsonc JE. The African trypanosome genome. *Int. J. Parasitol*. 2000; 30: 329-345.

El-Sayed NM, Donelson JE. African trypanosomes have differentially expressed genes encoding homologues of the Leishmania GP63 surface protease. *J Biol Chem*. 1997 Oct 17; 272 (42): 26742-8.

Ewing, B., Hillier, L., Wendl, M. C., Green, P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res*. 1998a; Mar;8(3):175-85.

Ewing, B., Green, P. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Res*. 1998b; Mar;8(3):186-94.

Fernandez-Bezerra C, Osuma A, Muller E, Dollet M, Sanchez-Moreno M. Characterization of *Phytomonas* isolated from fruits by electrophoretic isoenzymes and kinetoplast-DNA analysis. FEMS Microbiol Letters. 1996; 145:463-8.

Ferrari I, Lorenzi H, Santos MR, Brandariz S, Requena JM, Schijman A, Vazquez M, da Silveira JF, Ben-Dov C, Medrano C, Ghio S, Lopez Bergami P, Cano I, Zingales B, Urmenyi TP, Rondinelli E, Gonzalez A, Cortes A, Lopez MC, Thomas MC, Alonso C, Ramirez JL, Chiurrillo MA, Aldao RR, Levin MJ, *et al.* Towards the physical map of the *Trypanosoma cruzi* nuclear genome: construction of YAC and BAC libraries of the reference clone *T. cruzi* CL-Brener. Mem. Inst. Oswaldo Cruz. 1997; 92: 843-852.

Felsenstein J. Accuracy of Coalescent Likelihood Estimates: Do we need More Sites, More Sequences, or More Loci? Mol. Biol. Evol. 2005 Dec 19.

Fernandes AP, Nelson K, Berveley SM. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. Proc. Natl. Acad. Sci. U.S.A. 1993; 90:11608-12.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science. 1995; 269(5223):496-512.

Gadagkar SR, Rosenberg MS, Kumar S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J Exp Zool B Mol Dev Evol. 2005 Jan 15; 304(1):64-74.

Gibbs AJ. *Leptomonas serpens* n. sp., parasitic in the digestive tract and salivary glands of *Nezara viridula* (Pentatomidae) and in the sap of *Solanum lycopersicum* (tomato) and other plants. J. Parasitol. 1957; 47:297-303

Guerreiro, LTA., Souza, SS, Wagner, G, Souza, E, Mendes, P, Campos, L, Barros, L, Pires, P, Campos, MLM, Grisard, EC, Dávila, AMR. Exploring the genome of *Trypanosoma vivax* through GSS and *In Silico* Comparative Analysis. OMICS. 2005; 9(1):116-128.

Guerrini F, Segur S, Gargani D, Tibayrenc M, Dollet M. An isoenzyme analysis of the genus *Phytomonas*: genetic, taxonomic and epidemiologic significance. J. Protozool. 1992; 39:516-21.

Hammer J, Schneider M. Genomics Algebra: A new, integrating data model, language, and tool for processing and querying genomic information. CIDR Conference 2003.

Hanks SK, Quinn AM, Hunter T. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science. 1988 Jul 1; 241(4861):42-52.

Hanke J, Sanchez DO, Henriksson J, Aslund L, Pettersson U, Frasch AC, Hoheisel JD. Mapping the *Trypanosoma cruzi* genome: analyses of representative cosmid libraries. Biotechniques. 1996; 21:686/8-690/3.

Hata S, Kouchi H, Tanaka Y, Minami E, Matsumoto T, Suzuka I, Hashimoto J. Identification of carrot cDNA clones encoding a second putative proliferating cell-nuclear antigen, DNA polymerase delta auxiliary protein. Eur. J. Biochem. 1992; 203: 367-371.

Hoare, CA and Wallace, FG, Developmental stages of trypanosomatid flagellates: a new terminology. Nature. 1966; 212: 1385-1386.

Hollar L, Maslov DA. A phylogenetic view on the genus *Phytomonas*. Mol. Biochem. Parasitol. 1997; 89:295-299.

Huang, X., Madan A. CAP3: A DNA Sequence Assembly Program. *Genome Res.* 1999; Sep 9(9):868-77.

Hunkapiller T, Hood L. LIMS and the Human Genome Project. *Biotechnology (N Y)*. 1991 Dec;9(12):1344-5.

Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, Sisk E *et al.* The genome of the kinetoplastid parasite, *Leishmania major*. *Science*. 2005 Jul 15; 309 (5733): 436-42.

Jamie R. Stevens, Wendy C. Gibson. The evolution of pathogenic trypanosomes. *Cad. Saúde Pública, Rio de Janeiro*. 1999 out-dez; 15(4):673-684.

Jankevicius JV, Jankevicius SI, Campaner M, Conchon I, Maeda LA, Teixeira MMG, Freymuller E, Camargo EP. Life cycle and culturing of *Phytomonas serpens* (Gibbs), a Trypanosomatid parasite of tomatoes. *J. Protozool.* 1989; 36(3):265-271.

Janssen CS, Barrett MP, Lawson D, Quail MA, Harris D, Bowman S, Phillips RS, Turner CMR. Gene discovery in *Plasmodium chabaudi* by genome survey sequencing. *Mol. Biochem. Parasitol.* 2001; 113: 251–260.

Karplus K, Katzman S, Shackleford G, Koeva M, Draper J, Barnes B, Soriano M, Hughey R. SAM-T04: what is new in protein-structure prediction for CASP6. *Proteins*. 2005; 61 Suppl 7:135-42.

Kim H, Feil IK, Verlinde CLMJ, Petra PH, Hol WG. Crystal structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Leishmania mexicana*: implications for structure-based drug design and a new position for the inorganic phosphate binding site. *Biochemistry*. 1995 Nov 21;34(46):14975-86.

Kumar S, Tamura K, Nei M. MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. *Comput Appl Biosci.* 1994 Apr; 10 (2): 189-91.

Lafond A. Sur la presence d'un *Leptomonas*, parasite de la classe des Flagellés, dans le latex de trois Euphorbiacées. Ann Inst Pasteur . 1910; 24: 205-219.

Lafond A. Sur la presence d'un *Leptomonas*, parasite de la classe des Flagellés dans le latex de l'*Euphorbia pilulifera*. C. R. Seances Soc. Biol. Fil.. 1909; 66: 1011-1013.

Lantwin CB, Schlichting I, Kabsch W, Pai EF, Krauth-Siegel RL. The structure of *Trypanosoma cruzi* trypanothione reductase in the oxidized and NADPH reduced state. Proteins 1994;18: 161-173.

Laurentino EC, Ruiz JC, Fazelinia G, Myler PJ, Degrave W, Alves-Ferreira M, Ribeiro JMC, Cruz AK. A survey of *Leishmania brasiliensis* genome by shotgun sequencing. Mol. Biochem. Parasitol. 2004; 137: 81-86.

LeCleir GR, Buchan A, Hollibaugh JT. Chitinase Gene Sequences Retrieved from Diverse Aquatic Habitats Reveal Environment-Specific Distributions. Appl. Environ. Microbiol. 2004 Dec; p. 6977–6983.

Li D, Roberts R. WD-repeat proteins: structure characteristics, biological function, and their involvement in human diseases. Cell. Mol. Life Sci. 2001; 58: 2085-2097.

Lom J. Biology of the trypanosomes and trypanoplasms of fish. In WHR Lumsden, DA Evans (eds), *Biology of the kinetoplastida*, Academic Press, London/New York/San Francisco 1976; p. 35-130.

Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997; 25: 955-964.

Maike Tech and Rainer Merkl. YACOP: Enhanced gene prediction obtained by a combination of existing methods. *In Silico Biology*. 2003; 3(4):441-51.

Marché S, Roth C, Philippe H, Dollet M, Baltz TH. Characterization and detection of plant trypanosomatids by sequence analysis of the small subunit ribosomal RNA gene. *Mol. Biochem. Parasitol* 1995; 71:15-26.

Marín, C, Hitos, AB, Rodrigues-Gonzalez, I, Dollet, M, Sanchez-Moreno, M. *Phytomonas* iron superoxide dismutase: a possible molecular marker. *FEMS Microbiology Letters*. 2004; 234: 69-74.

Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D. Evolutionary analysis of Arabidopsis, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *PNAS*. 2002 September 17; vol. 99 (19):12246-12251.

Maslov, DA, Podlipaev, SA, Lukes, L. Phylogeny of the kinetoplastida: Taxonomic problems and insights into the evolution of parasitism. *Mem. Inst. Oswaldo Cruz*. 2001; April vol. 96(3): 397-402.

Maslov DA, Nawathean P, Scheel J. Partial kinetoplast-mitochondrial gene organization and expression in the respiratory deficient plant trypanosomatid *Phytomonas serpens*. *Mol Biochem Parasitol*. 1999 Apr 30; 99(2): 207-21.

Maslov, D,A, Hollar, L, Haghghat, P, Nawathean, P. Demonstration of mRNA editing and localization of guide RNA genes in Kinetoplast-mitochondria of the plant trypanosomatid *Phytomonas serpens*. *Mol. Biochem. Parasitol*. 1998; 93:225-236.

Maslov DA, Simpson L. Evolution of parasitism in kinetoplastid protozoa. *Parasitol. Today*. 1995; 11:30-2.

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J, Kalinowski J, Linke B, Rupp O, Giegerich R, Puhler A. GenDB--an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 2003 Apr 15;31(8):2187-95.

Matsumoto K, Moriuchi T, Koji T, Nakane PK. Molecular cloning of cDNA coding for rat proliferating cell nuclear antigen (PCNA)/cyclin. *EMBO J.* 1987; 6: 637-642.

Miya M, Nishida M. Use of mitogenomic information in teleostean molecular phylogenetics: a tree-based exploration under the maximum-parsimony optimality criterion. *Mol Phylogenet Evol.* 2000 Dec;17(3):437-55.

Mitchell, PL. Heteroptera as Vectors of Plant Pathogens. *Neotrop. Entomol.* 2004; 33(5):519-545.

Morty RE, Bulau P, Pelle R, Wilk S, Abe K. Pyroglutamyl peptidase type I from *Trypanosoma brucei*: a new virulence factor from African trypanosomes that de-blocks regulatory peptides in the plasma of infected hosts. *Biochem J.* 2006 Mar 15; 394 (Pt 3):635-45.

Notredame C, Higgins DG, Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* 2000; Sep 8;302(1):205-17.

Nunes FM, Valente V, Sousa JF, Cunha MA, Pinheiro DG, Maia RM, Araujo DD, Costa MC, Martins WK, Carvalho AF, Monesi N, Nascimento AM, Peixoto PM, Silva MF, Ramos RG, Reis LF, Dias-Neto E, Souza SJ, Simpson AJ, Zago MA, Soares AE, Bitondi MM, Espreafico EM, Espindola FS, Paco-Larson ML, Simoes ZL, Hartfelder K, Silva WA Jr. The use of Open Reading frame ESTs (ORESTES) for analysis of the honey bee transcriptome. *BMC Genomics.* 2004; Nov 3;5(1):84.

Olmstead RG, Sweere JA, Wolfe KH. Ninety extra nucleotide in *ndhF* gene of tobacco chloroplast DNA: a summary of revisions to the 1986 genome sequence. *Plant Mol Biol.* 1993 Sep; 22 (6):1191-3.

Oriane Matte-Tailliez O, Brochier C, Forterre P, Philippe H. Archaeal Phylogeny Based on Ribosomal Proteins. *Mol Biol Evol.* 2002 May;19(5):631-9.

Parkinson J, Guiliano DB, Blaxter M. Making sense of EST sequences by CLOBBing them. *BMC Bioinformatics.* 2002; 25(3): 1-8.

Passagia MP, Zaha A. Técnicas de biologia molecular. In: Zaha A. *Biologia Molecular Básica.* 3th ed. Porto Alegre: Mercado Aberto; 2003. p.379-413.

Pavao F, Castilho MS, Pupo MT, Dias RL, Correa AG, Fernandes JB, da Silva MF, Mafezoli J, Vieira PC, Oliva G. Structure of *Trypanosoma cruzi* glycosomal glyceraldehyde-3-phosphate dehydrogenase complexed with chalepin, a natural product inhibitor, at 1.95 Å resolution. *FEBS Lett.* 2002 Jun 5; 520 (1-3):13-7.

Pappas GJ Jr, Benabdellah K, Zingales B, Gonzalez A. Expressed sequence tags from the plant trypanosomatid *Phytomonas serpens*. *Mol Biochem Parasitol.* 2005 Aug; 142 (2):149-57.

Pearson WR. Rapid and sensitive sequence comparison with FASTP and FASTA. *Meth. Enzymol.* 1990;183:63-98.

Peterson S.N., Hu P-C, Bott K.F., Hutchison C.A. A survey of the *Mycoplasma genitalium* genome by using random sequences. *J. Bacteriol.* 1999; 175:7918-7930.

Peterson SN, Schramm N, Hu PC, Bott KF, Hutchison CA. A random sequence approach for placing markers on the physical map of *Mycoplasma genitalium*. *Nucleic Acids Res.* 1991; 19: 6027-6031.

Pinge-Filho P, Peron JP, de Moura TR, Menolli RA, Graca VK, Estevao D, Tadokoro CE, Jankevicius JV, Rizzo LV. Protective immunity against *Trypanosoma cruzi*

provided by oral immunization with *Phytomonas serpens*: role of nitric oxide. Immunol. Lett. 2005; Jan 31; 96(2):283-90.

Pissani D, Poling L L, Lyons-Weiler M, Hedges SB. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. BMC Biology 2004, 2:1.

Porcel BM, Tran AN, Tammi M, Nyarady Z, Rydaker M, Urmenyi TP, Rondinelli E, Pettersson U, Andersson B, Aslund L. Gene survey of the pathogenic protozoan *Trypanosoma cruzi*. Genome Res. 2000 Aug;10(8):1103-7.

Pourshafie M, Morand S, Virion A, Rakotomanga M, Dupuy C, Loiseau PM. Cloning of S-Adenosyl-L-Methionine: C-24-delta-Sterol-Methyltransferase (ERG6) from *Leishmania donovani* and Characterization of mRNAs in Wild-Type and Amphotericin B-Resistant Promastigotes. Antimicrob. Agents Chemother. 2004 July; p. 2409–2414.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. InterProScan: protein domains identifier. Nucleic Acids Res. 2005; Jul 1:33(Web Server issue):W116-20.

Rawlings ND, Barrett AJ. Evolutionary families of metallopeptidases. Meth. Enzymol. 1995; 248: 183-228.

Regier JC and Shultz JW. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. Mol. Biol. Evol. 1997; Volume 14, Number 9 Pp. 902-913.

Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000 Jun;16(6):276-7.

Rokas A, Williams BL, King N, Carroll SB. Genome-scale approaches to resolving incongruence in molecular phylogenies. Nature. 2003 Oct 23; 425 (6960):798-804.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. Artemis: sequence visualization and annotation. *Bioinformatics*. 2000 Oct;16(10):944-5.

Sá-Carvalho, D., Traub-Cseko, Y.M. Sequences with high propensity to form G-quartet structures in kinetoplast DNA from *Phytomonas serpens*. *Mol. Biochem. Parasitol.* 1995; 72:103-109.

Sá-Carvalho, D, Perez-Morga, D, Traub-Cseko, Y.M. Characterization of kinetoplast DNA from *Phytomonas serpens*. *J. Eukaryot. Microbiol.* 1993; 40 (3): 317-322.

Salzberg S, Delcher A, Kasif S, White O. Microbial gene identification using interpolated Markov models. *Nucleic Acids Res.* 1998; 26:2, 544-548.

Sambrook J, Russel DW. *Molecular Cloning: A Laboratory Manual Vol 2.* 3th ed. New York: Cold Spring Harbor Laboratory Press; 2001.

Sanchez DO, Zandomeni RO, Cravero S, Verdun RE, Pierrou E, Faccio P, Diaz G, Lanzavecchia S, Aguero F, Frasch AC, Andersson SG, Rossetti OL, Grau O, Ugalde RA. Gene Discovery through Genomic Sequencing of *Brucella abortus*. *Infect Immun.* 2001 Feb;69(2):865-8.

Sanchez-Moreno M, Fernandez-Bezerra C, Mascaro C, Rosales MJ, Dollet M, Osuna A. Isolation, in vitro culture, ultrastructure study and characterization by lectin-agglutination tests of *Phytomonas* isolated from tomatoes (*Lycopersicon esculentum*) and cherimoyas (*Annona cherimolia*) in Southeastern Spain. *Parasitol. Res.* 1995; 81:575-81.

Schaeffer SW, Bernhardt MJ, Anderson WW. Evolutionary rearrangement of the amylase genomic regions between *Drosophila melanogaster* and *Drosophila pseudoobscura*. *J. Hered.* 2003; 94: 464-471.

Schottelius J. Selective lectin reactions of two stocks of *Leishmania enriettii* with differing pathogenicity. *Parasitol Res.* 1987;73(1):1-8.

Smith MW, Aley SB, Sogin M, Gillin FD, Evans GA. Sequence survey of the *Giardia lamblia* genome. *Mol. Biochem. Parasitol.* 1998; 95: 267–280.

Smith TF, Gaitatzes C, Saxena K, Neer EJ. The WD repeat: a common architecture for diverse functions. *Trends Biochem. Sci.* 1999; 24: 181-185.

Stein, L. Genome annotation: from sequence to biology. *Nat. Rev. Genet.* 2001; 2, 493-503.

Stevens, JR and Gibson, WC. The Molecular Evolution of Trypanosomes. *Cad Saude Publica.* 1999 Oct-Dec;15(4):673-84.

Sullivan FX, Walsh CT. Cloning, sequencing, overproduction and purification of trypanothione reductase from *Trypanosoma cruzi*. *Mol. Biochem. Parasitol.* 1991; 44: 145-147.

Suzuka I, Hata S, Matsuoka M, Kosugi S, Hashimoto J. Highly conserved structure of proliferating cell nuclear antigen (DNA polymerase delta auxiliary protein) gene in plants. *Eur. J. Biochem.* 1991; 195: 571-575.

Tech M, Merkl R. YACOP: Enhanced gene prediction obtained by a combination of existing methods. *In Silico Biol.* 2003;3(4):441-51.

Teixeira MMG, Serrano MG, Nunes LR, Campaner M, Buck GA, Camargo EP. Trypanosomatidae: A spliced-leader-derived probe specific for the genus *Phytomonas*. *Exp. Parasitol.* 1996; 84:311-9.

Teixeira MMG, Campaner M, Camargo EP. Characterization of the target antigen of *Phytomonas*-specific monoclonal antibodies. J. Eukaryot. Microbiol. 1995; 42:232-7.

Teixeira MMG, Camargo EP. Monoclonal antibodies for the identification of trypanosomatids of the genus *Phytomonas*. J. Protozool. 1989; 36:262-4.

Telles S, Abate T, Slezzynger T, Henriquez DA. *Trypanosoma cruzi* ubiquitin as an antigen in the differential diagnosis of Chagas disease and leishmaniasis. FEMS Immunol Med Microbiol. 2003 Jun 10; 37(1):23-8.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; Nov 11;22(22):4673-80.

Travali S, Ku DH, Rizzo MG, Ottavio L, Baserga R, Calabretta B. Structure of the human gene for the proliferating cell nuclear antigen. J. Biol. Chem. 1989; 264: 7466-7472.

Uttaro, AD, Sanchez-Moreno, M, Opperdoes, FR. Genus-specific biochemical makers for *Phytomonas* spp. Mol. Biochem. Parasitol. 1997; 90: 337-342.

Vellieux FM, Hajdu J, Verlinde CL, Groendijk H, Read RJ, Greenhough TJ, Campbell JW, Kalk KH, Littlechild JA, Watson HC, et al.. Structure of glycosomal glyceraldehyde-3-phosphate dehydrogenase from *Trypanosoma brucei* determined from Laue data. Proc Natl Acad Sci U S A. 1993 Mar 15; 90 (6):2355-9.

Venter, JC, Remington, K, Heidelberg, JF, Halpern, AL, Rusch, D, Eisen, J *et al.* Environmental genome Shotgun sequencing of the Sargasso sea. Science. 2004; April (304): 66-74.

Vickerman K, Preston TM. Comparative cell biology of the kinetoplastid flagellates. In WHR Lumsden, DA Evans (eds), *Biology of the kinetoplastida*, Academic Press, London/New York/San Francisco 1976; p. 35-130.

Weber JL, Myers EW. Human whole-genome shotgun sequencing. *Genome Res.* 1997 May;7(5):401-9.

Wiemer, E.A.C. *et al.* Molecular analysis of glyceraldehyde-3-phosphate dehydrogenase in *Trypanoplasma borreli*: an evolutionary scenario of subcellular compartmentation in Kinetoplastida. *J. Mol. Evol.* 1995; 40: 443–454.

Tabela 3.1: Meios de cultura utilizados neste estudo.

Meio LB (Luria-Bertani)	Triptona 10g/l; extrato de levedura 5g/l; NaCl 10g/l. Ajuste do pH a 7,0 com NaOH 5 N. Esterilização a 121 ⁰ C por 20min.
Meio LB ágar	Triptona 10g/l; extrato de levedura 5g/l; NaCl 10g/l; ágar 15g/l. Ajuste do pH 7,0 com NaOH 5 N. Esterilização a 121 ⁰ C por 20min. Foi adicionado no meio LB é adicionado tetraciclina (12.5 µg/ml), canamicina (50 µg/ml) ou ampicilina (100 µg/ml).
“Meio Schneider’s Insect”	“Schneider’s Insect” 24,5 g/l, CaCl _{0,6} g/l, 10% Soro Fetal Bovino e ampicilina (100 µg/ml). Filtrar em membrana de 0,22µ.

Tabela 3.2: Soluções e tampões utilizados neste estudo

IPTG (Isopropil- β -D-tiogalactopiranosídeo)	100mM ou 1M
PBS (Phosphate buffered saline)	NaCl 132mM, KCl 3mM, Na ₂ HPO ₄ 8mM, KH ₂ PO ₄ 1,5mM. Ajustar para pH 7,2 com NaOH.
STET	8g de sacarose, 5mL tris-HCL pH 8, 5g de triton X100, água q.s.p. 100mL.
Tampão de amostra de DNA	TBE 2X, azul de bromofenol, 0,25%, xileno cianol 0,25%, Ficoll 400 20%.
TBE (Tris/ Borato /EDTA)	Tris base 90 mM, ácido bórico 90 mM, EDTA 2mM, pH 8,0.
TE (Tris EDTA)	10 mM tris-HCl pH 7,5, 1mM EDTA
Xgal (5-bromo-4-cloro-3-indoil- β -D-galactopiranosídeo)	2% em dimetilformamida.
Tampão de lise	10 mM Tris pH 7,4, 10 mM NaCl, 25 mM EDTA, 1% SDS 10 %
GET (Glicose, EDTA e Tris)	23 ml de glicose anidra 20% filtrada, 10 ml 0,5 M EDTA pH 8,0 – autoclavado, 13 ml Tris-HCl pH 7,4 – autoclavado, H ₂ O MilliQ q.s.p. 500 ml- autoclavada
Acetato de Potássio 3M (KOAc)	60 ml de Kac 5 M, 11,5 ml de ácido acético glacial, 28,5 ml H ₂ O MilliQ autoclavada