

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE MICROBIOLOGIA**

**RETROTRANSPOSONS NOS GENOMAS DE
FUNGOS DO COMPLEXO
PARACOCCIDIOIDES: IDENTIFICAÇÃO,
CARACTERIZAÇÃO, DISTRIBUIÇÃO E
EXPRESSÃO.**

Marco Aurélio Soares

Belo Horizonte, Minas Gerais

2013

Marco Aurélio Soares

**RETROTRANSPOSONS NOS GENOMAS DE
FUNGOS DO COMPLEXO
PARACOCCIDIOIDES: IDENTIFICAÇÃO,
CARACTERIZAÇÃO, DISTRIBUIÇÃO E
EXPRESSÃO.**

Tese apresentada ao programa de Pós-graduação em Microbiologia do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais como requisito para obtenção do título de doutor em Microbiologia.

Orientadora: Prof^a. Dr^a. Patrícia Silva Cisalpino

Co-orientador: Dr. Jerônimo da Conceição Ruiz

(Centro de Pesquisas René Rachou, FioCruz / MG)

Belo Horizonte
Marco de 2013

Este estudo foi realizado no Laboratório de Biologia de Microrganismos do Departamento de Microbiologia, do Instituto de Ciências Biológicas (UFMG) e no Laboratório de Parasitologia Celular e Molecular da Fundação Osvaldo Cruz – Fiocruz-MG.

DEDICATÓRIA

A DEUS, PAI, FILHO E ESPÍRITO SANTO

Pelo dom da vida, pela experiência pessoal e por todas as bênçãos.

Aos meus Pais

José Eustáquio Soares e Maria de Fátima Soares, pelo amor, pelos ensinamentos, pela contribuição alélica e por tudo que vocês representam para mim. Amo vocês.

A minha família:

Suede Naves Narciso Soares, obrigado pelo amor, paciência compreensão e por ser meu porto seguro durante os momentos difíceis do meu curso. Obrigado por ser o anjo que me conduz nesta vida.

Davi Narciso Soares: louvo a DEUS pela sua vida, meu filho. Você ainda em gestação conseguiu me mostrar os valores da confiança em DEUS e da família. Obrigado por ser a fonte de paz. Eu te amo.

Aos meus irmãos

Paulo Henrique Soares e Shirley de Fátima Soares, pelo carinho, amizade e incentivo. Amo vocês.

Aos meus sobrinhos,

Nestor, Isabel, Samira, Guilherme, Leonardo, Elloah e Lara, por me ensinarem o valor da inocência.

Aos meus tios

Dulce, José Rodrigues, Maria Aparecida e Jorge. Obrigado pelo carinho e por ajudar na realização deste sonho.

AGRADECIMENTO

À professora orientadora Patrícia. Obrigado por me conduzir no caminho da ciência, apontar a perseverança e paciência como pré-requisitos na conquista dos objetivos. Sempre me lembrarei dos exemplos de humildade e de uma frase que levarei pela vida..

Ao meu co-orientador Dr. Jeronimo pelos ensinamentos em bioinformática, pela amizade e pelas conversas aconselhadoras sobre o valor próprio. Lembrarei também da sua frase.

À professora Viviane pela atenção, acolhida e ensinamentos. Obrigado por mostrar que não há situação nesta vida em que devemos desanimar. Você é um exemplo de superação. Quando crescer quero ser igual você.

À professora Fátima Noronha, pela acolhida, amizade, e pelas conversas agradáveis que para mim serviram como lições de vida.

À professora Daniele pela contribuição científica, pela convivência, alegria e disponibilidade.

À professora Sueli pelas conversas sobre os netos. Aqueles momentos me fizeram pensar no valor da família.

A todos os professores do departamento de Microbiologia pelos ensinamentos durante todo o curso.

Às professoras Adlane e Mônica do departamento de genética pelo carinho, incentivo e confiança. A acolhida de vocês foi fundamental na minha caminhada.

À Gilvânia, técnica do laboratório, pela ajuda no preparo do material e principalmente pela alegria contagiante que suavizou os dias difíceis cansativos.

Aos pesquisadores da Fiocruz: Raul, Patrícia Ruy, Nesley e Édson, obrigado pela amizade. Luciana, e Antônio, obrigado pela amizade e ajuda incondicional na condução das análises de bioinformática.

Aos alunos do laboratório de Biologia de Microrganismos: Diogo, Fernanda, Ludmila, Nívea, Patrícia Campi e Rosana. Obrigado pela amizade, companheirismo e ajudas, incondicionais.

Aos alunos do laboratório BIOGEN: Fábio, Marliete, Vanessa, Ludmila, César. Obrigado pela acolhida, convivência e por tornar meus dias mais alegres.

Aos alunos do laboratório LIMHO: Ana Carol, Celso, Lívia Resende, Lívia Tavares, Luana, Lucas, Mila, Nívea, Patrícia Campi, Rebeca, Raquel, Renata, Sílvia, Talles, Vivian, Zélia. Obrigado pela amizade e convivência. Daniel Cisalpino e Fernando, não sei como agradecer a ajuda dos experimentos de qRT-PCR. Muito obrigado.

A todos os alunos do Departamento de Microbiologia pela amizade.

Às agências CNPq e FAPEMIG pelo apoio financeiro.

043

Soares, Marco Aurélio.

Retrotransposons nos genomas de fungos do complexo *Paracoccidioides*: identificação, caracterização, distribuição e expressão [manuscrito] / Marco Aurélio Soares. - 2013.

208 f. : il. ; 29,5 cm.

Orientadora: Prof^a. Dr^a. Patrícia Silva Cisalpino. Co-orientador: Dr. Jerônimo da Conceição Ruiz.

Tese (doutorado) - Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas.

1. Microbiologia - Teses. 2. Paracoccidioides - Teses. 3. Retroelementos. 4. Elementos de DNA transponíveis. I. Cisalpino, Patrícia Silva. II. Ruiz, Jerônimo da Conceição. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 579

Sumário

Lista de abreviaturas	ix
Lista de figuras	xiii
Lista de tabelas	xv
Abstract	xvi
Resumo	xvii
1 – INTRODUÇÃO E JUSTIFICATIVA	18
2 – REVISÃO DE LITERATURA	20
2-1 <i>Paracoccidioides brasiliensis</i> e Paracoccidioidomicose	20
2-2 Estudos genéticos e genômicos de <i>P. brasiliensis</i>	24
2-3 Elementos genéticos móveis ou transponíveis: transposons e retrotransposons.	30
2.3.1 Descoberta, ubiquidade e abundância	30
2.3.2 Características dos elementos transponíveis das Classes I e II e sua classificação.	31
2.3.3 Principais características de retrotransposons (elementos da Classe I) segundo a classificação (Wicker, Sabot <i>et al.</i> , 2007).	35
2.3.4 Potencial dos elementos transponíveis em causar variações genéticas	43
2-4 Elementos transponíveis em fungos	45
3- OBJETIVOS	53
3.1 Objetivo Geral	53
3.2 Objetivos Específicos	53
4 - MATERIAL E MÉTODOS	55
4.1 Análise de bioinformática	55
4.1.1 Workflow computacional	56
4.1.2 Sequências de EST (<i>Expressed Sequence Tags</i>)	57
4.1.3 – Mineração de dados (data mining)	57
4.1.4 Processamento dos dados	58
4.1.5 Buscas por Similaridade de Sequências	59
4.1.6 Construção da tabela <i>hyperlinkada</i>	61
4.2 - Anotações de Sequências	62
4.2.1 Anotação estrutural e funcional dos retrotransposons	62

4.2.2	Identificação de domínios conservados	62
4.2.4	– Artemis	62
4.2.5	- Análise filogenética	63
4.3	– Isolamento e caracterização de putativos retrotransposons	63
4.3.1	– Microrganismos.....	64
4.3.2	– Manutenção de amostras.....	66
4.3.3	– Extração de DNA genômico de isolados do complexo Paracoccidioides..	66
4.3.4	- PCR convencional	67
4.3.5	– Extração de RNA	69
4.3.6	Quantificação do RNA.....	69
4.3.7	- Transcrição Reversa para reações qRT-PCR	69
4.3.8	– Transcrição reversa para reações RT-PCR iniciador específico	70
4.3.9	– PCR quantitativo em tempo real (qRT-PCR)	70
4.3.10	– Clonagem	73
4.3.11	– Preparo de bactérias Escherichia coli competentes	73
4.3.12	– Transformação bacteriana	73
4.3.13	– Extração de DNA plasmidial (preparação em pequena escala).....	74
4.3.14	– Sequenciamento de DNA.....	74
4.3.15	– Restrição de DNA genômico	75
4.3.16	– <i>Southern blot</i>	77
4.3.17	– Fluxograma de validação experimental	77
5	– RESULTADOS	80
5.1	– Análise das sequências de EST e geração de grupos de sequências.	80
5.2	– Busca por similaridade dos grupos de <i>EST</i> e anotação funcional.....	83
5.3	– Caracterização dos elementos móveis	93
5.3.1	– Caracterização estrutural do elemento RtPb1	93
5.3.2	– Caracterização estrutural do elemento RtPb2	94
5.3.3	– Caracterização estrutural do elemento RtPb3.....	94
5.3.4	– Caracterização estrutural do elemento RtPb4	97
5.3.5	– Caracterização estrutural do elemento RtPb5	97

5.4 – Classificação dos elementos caracterizados	99
5.5 – Relação filogenética dos retrotransposons	100
5.6 – Distribuição e mapeamento, <i>in silico</i> , dos retrotransposons	102
5.6.1 Número total de cópias	102
5.6.2 Número de cópias completas e íntegras	104
5.7 – Validação experimental dos elementos preditos.....	109
5.7.1 – Amplificação por PCR dos elementos completos nos genomas dos isolados Pb01, Pb03 e Pb18.....	109
5.7.2 – Amplificação por PCR de fragmentos específicos das transcriptases reversas dos elementos RtPb3, RtPb4 e RtPb5 nos genomas de diferentes isolados.	111
5.7.3 – Comprovação, por sequenciamento, da identidade das sequências.	113
5.7.4 – Mapeamento dos elementos retrotransponíveis por <i>Southern blot</i>	115
5.8 – Análise dos padrões de transcrição dos cinco elementos retrotransponíveis por RT-PCR.	121
5.9 – Quantificação da transcrição dos elementos retrotransponíveis por qRT-PCR.	123
6 – DISCUSSÃO	125
7 - CONCLUSÕES	137
8 – PERSPECTIVAS.....	138
9 – REFERÊNCIAS BIBLIOGRÁFICAS	139

Lista de abreviaturas

aa – aminoácidos

AIDS (*Acquired immunodeficiency syndrome*) – síndrome da imunodeficiência adquirida.

BLAST (*Basic local alignment search tool*) – ferramenta de busca computacional, algoritmo básico para alinhamento local de sequências.

cDNA – DNA complementar

Chromo (*chromodomain*) – cromodomínio

CIE – contra-imunoelektroforese

DIRS-like elements – elementos semelhantes a DIRS1, um retrotransposon descrito em *Dictyostellium discoideum*. Apresentam terminações longas repetidas, invertidas e não idênticas, e codificam uma tirosina recombinase (YR)

CT (Cycle Threshold) – ciclo da reação em que fluorescência ultrapassa a linha básica

DNA – ácido desoxirribonucleico

EDTA (*Ethylenediamine tetraacetic acid*) – (ácido etilendiamino tetraacético)

ELISA (*enzyme-linked immunosorbent assay*) – ensaio enzimático de imunoadsorção.

ENV – proteína do envelope viral codificada pelo gene *env*

EST (*expressed sequence tags*) – etiquetas de sequências expressas.

Gag – proteínas do nucleocapsídeo codificadas pelo gene *gag*

GAPDH – gliceraldeído fosfato desidrogenase.

GSS – *genome survey sequences*

ID – imunodifusão.

IFI – imunofluorescência indireta.

IN (*integrase*) – integrase, domínio enzimático ou enzima responsável por catalisar a integração do DNA originado de um elemento transponível no genoma hospedeiro

IS (*insertion sequences*) – sequências de inserção, transposons de DNA de estrutura muito simples que existem em genomas bacterianos.

Kb – kilobases

LINE (*Long interspersed elements*) – elementos longos intercalados

LTR (*Long terminal repeats*) – terminações longas repetidas

Mb – Megabases

NCBI (*National Center for Biotechnology Information*) – Centro Nacional para Informação Biotecnológica (EUA)

NR (*Non-redundant protein sequences database*) – banco de dados de sequencias não redundantes de proteínas, abrigado no sítio do NCBI

nt – nucleotídeo

ORF (*open reading frame*) – fases de leitura aberta

Pb – pares de base

PBS (*primer binding site*) – sítio de anelamento do primer de tRNA

PCM – paracoccidioidomicose

PCR (*Polymerase chain reaction*) – reação em cadeia da polimerase

PERL (*Practical extraction and report language*) – linguagem de programação computacional

PFGE (*Pulsed field gel electrophoresis*) – eletroforese de campos pulsados ou de campos alternados

PLE (*Penelope-like elements*) – elementos semelhantes ao retrotransposon Penelope, descrito em *Drosophila virilis*, que apresenta um domínio de transcriptase reversa divergente do de outras sequências previamente descritas

Pol protein (*poliprotein*) – proteína precursora de enzimas e de proteínas estruturais de vírus e de partículas semelhantes a vírus. Normalmente processada por clivagem por proteases originando outras moléculas funcionalmente autônomas.

PPT – trato de polipurina

qRT-PCR (*quantitative reverse transcription real time PCR*) – reação de PCR quantitativa em tempo real a partir de molde cDNA

RAPD (*Random amplification of polymorphic DNA*) – amplificação randômica de DNA polimórfico

Replibase – banco de dados de análise de sequências repetitivas desenvolvido e abrigado pelo GIRI (*Genetic Information Research Institute*)

RFLP – polimorfismos de tamanho de fragmentos de DNA gerados por endonucleases de restrição

RNA – ácido ribonucleico

RNase H – ribonuclease H

RT (*reverse transcriptase*) – transcriptase reversa

SINE (*Small interspersed elements*) – elementos curtos intercalados

SSC (*Saline Sodium Citrate*) – salina citrato de sódio

TAE – Tris acetato EDTA

TBE – Tris borato EDTA

TE (*transposable element*) - elemento transponível.

TEClass – banco de dados local de elementos transponíveis desenvolvido pelo Dr Jose Marcos Ribeiro.

TEfam – banco de dados relacional para a submissão, recuperação e análise de sequencias de elementos transponíveis abrigado no sítio do Virginia Tech (Instituto politécnico da Universidade estadual da Virgínia, EUA).

TIR – terminações repetidas e invertidas

TIR (*terminal inverted repeats*) – terminações invertidas e repetidas.

TRIS – Tris-hidroximetil-aminometano

tRNA – ácido ribonucléico transportador

TSD (*target site duplication*) – duplicação do sitio alvo de inserção

VLP (*virus-like particle*) – partícula semelhante ao capsídeo viral

Lista de figuras

Figura 1 – Sistema de classificação	34
Figura 2 – Ciclo teórico de replicação de retrotransposon	39
Figura 3 – Mecanismo de transcrição reversa de retrotransposon com LTR. ..	40
Figura 4 – Representação esquemática das análises computacionais para identificação de potenciais elementos retrotransponíveis.	56
Figura 5 – Representação esquemática dos testes de validação experimental dos dados obtidos por análises de bioinformática.....	79
Figura 6 – Agrupamento e distribuição de ESTs.....	82
Figura 7 – Resultados das buscas por similaridade	84
Figura 8 - Número de <i>contigs</i> no banco de dados NR correspondentes a retrotransposons.	87
Figura 9 – Grupos de sequências representativos de elementos genéticos móveis.....	88
Figura 10 – Figura representativa da estrutura do retrotransposon RtPb1.	95
Figura 11 - Figura representativa das estruturas dos retrotransposons RtPb2 e RtPb3	96
Figura 12 – Figura representativa das estruturas dos retrotransposons RtPb4 e RtPb5.	98
Figura 13 - Análise filogenética dos elementos RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5 utilizando o método Neighbor-joining e o modelo <i>p-distance</i>	100
Figura 14 – Número total de cópias* dos elementos Gypsy-RtPb1, LINE-RtPb2, Copia-RtPb3, Copia-RtPb4 (painel A) e Gypsy-RtPb5 (painel B) nos isolados Pb01, Pb03 e Pb18.	103
Figura 15 – Número de cópias* completas e íntegras dos elementos RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5.....	107
Figura 16 – Resolução eletroforética dos amplicons obtidos para Copia-RtPb3(A), Copia-RtPb4(B) e Gypsy-RtPb5(C) em diferentes isolados do complexo <i>Paracoccidioides</i>	110
Figura 17- Identificação dos elementos RtPb3, RtPb4 e RtPb5 em diferentes isolados do complexo <i>Paracoccidioides</i>	112

Figura 18 - Alinhamento dos amplicons referentes aos elementos Copia-RtPb3(A), Copia-RtPb4(B) e Gypsy-RtPb5(C) sequenciados com suas respectivas predições <i>in silico</i>	114
Figura 19 - Determinação do número de cópias genômicas do elemento Copia-RtPB3 nos isolados Pb01, Pb03 e Pb18 do complexo <i>Paracoccidioides</i>	116
Figura 20 - Determinação do número de cópias genômicas do elemento Copia-RtPB4 nos isolados Pb01, Pb03 e Pb18 do complexo <i>Paracoccidioides</i>	118
Figura 21 - Determinação do número de cópias genômicas do elemento Gypsy-RtPB5 nos isolados Pb01, Pb03 e Pb18 do complexo <i>Paracoccidioides</i>	120
Figura 22 – Análise da transcrição dos elementos retrotransponíveis Gypsy-RtPb1, LINE-RtPb2, Copia-RtPb3, Copia-RtPb4, Gypsy-RtPb5 e dos genes controles β -tubulina e <i>gp43</i>	122
Figura 23 - Quantificação relativa do número de transcritos dos elementos RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5.	124

Lista de tabelas

Tabela 1 - Bancos de dados utilizados para as análises.....	58
<i>Tabela 2 - Superfamílias de elementos transponíveis e retrotransponíveis representadas no banco de dados TEClass.....</i>	<i>59</i>
Tabela 3 - Algoritmos utilizados para buscas de similaridade local entre pares de sequências.	60
Tabela 4 - Parâmetros utilizados na execução das análises pelos algoritmos do <i>Blast</i>	60
Tabela 5 - Isolados fúngicos do complexo <i>Paracoccidioides</i>	65
Tabela 6 - Relação de oligonucleotídeos (PCR convencional).....	68
Tabela 7 – Relação de oligonucleotídeos (qRT-PCR).....	72
Tabela 8 – Relação das endonucleases de restrição.....	76
Tabela 9 – Grupos representativos de putativos elementos retrotransponíveis e resultados de similaridade (<i>hits</i> , <i>e-values</i> , identidades, escores) encontrados nos bancos NR, TEfam, TE-Class e Repbase.	91
Tabela 10 – Prováveis retrotranspos e seus grupos constituintes.	92
Tabela 11 – Classificação dos elementos identificados	99
Tabela 12 – Numero de cópias completas e de copias truncadas dos cinco elementos (Gypsy-RtPb1, LINE-RtPb2, Copia-RtPb3, Copia- RtPb4 e Gypsy-RtPb5) encontradas nos genomas dos isolados Pb01 (<i>P. lutzi</i>), Pb03 e Pb18 (<i>P. brasiliensis</i>).....	108

Abstract

Evidences from the literature and the previous experience of our research group indicate the occurrence of mobile genetic elements in the genomes of fungi of the *Paracoccidioides* species complex. Aiming to identify and characterize potentially active retrotransposons in these genomes we employed bioinformatics analysis of the EST deposited in NCBI public databases (www.ncbi.nlm.nih.gov). The 41,558 EST were assembled into a local bank and grouped by sequence alignments, generating 12,922 groups distributed in 8110 singlets and 4812 contigs. These groups were compared with the non-redundant database (NR, NCBI) and databases specialized in transposable elements (TE, RepBase and TEClass), through the BLAST tool. 10,378 groups presented similarity against NR, 142 with retrotransposons characteristic annotator terms as reverse transcriptase (the most common term, with 84 records). In specific banks, 809 groups were found, 342 of which showing regions of similarity to retrotransposons. The remaining 437 were similar to DNA transposons. Only 52 groups were common to both the TE specific and NR banks, and those that showed the highest similarity allowed the identification of five distinct elements in the genomes: - RtPb1, RtPb2, RtPb3, RtPb4 and RtPb5. The elements were characterized and classified in the LTR and LINE orders, superfamilies *Copy*, *Gypsy* and *I*, sustained by phylogenetic analyzes. The coding regions of these elements revealed the presence of characteristic regions corresponding to endonucleases, integrases, proteases, RNases H, reverse transcriptases and a chromo-domain. We identified LTR regions in elements of the Copy and Gypsy superfamilies (RtPb1, RtPb3, RtPb4 and RtPb5) and a poly-A tail LINE-like element (RtPb2). The elements RtPb1 and RtPb2 (Araújo, 2008) had their complete sequences mapped and characterized *in silico*. The genomic mapping *in silico* identified 538 copies of all five elements (RtPb1 = 52; RtPb2 = 101; RtPb3 = 18; RtPb4 = 65 and RtPb5 = 302) distributed in the three genomes studied (Pb01= 366; Pb03= 57; Pb18= 115) (www.broadinstitute.org). PCR allowed the amplification of the entire elements by primer walking method. The identity of the elements was confirmed by sequencing and alignment of the reverse transcriptase regions. The sequences were used to verify the presence of each element in the DNA of 31 different fungal isolates. Southern blots analysis of genomic DNA in isolates representative of each of the four known phylogenetic species indicated their presence in multiple copies. The expression of distinct retrotransposons was studied by RT-PCR and by quantitative qRT-PCR. The most expressed element is RtPb1 in isolate Pb01. The analysis showed different expression patterns between distinct fungal isolates. It was observed a peculiar expression of each different retrotransposon in fungal isolates each representative of a known and distinct phylogenetic species.

Resumo

Evidências da literatura e a experiência prévia do nosso grupo indicaram a ocorrência de elementos genéticos móveis nos genomas de fungos do complexo *Paracoccidioides*. Objetivando a identificação e a caracterização de retrotransposons potencialmente ativos nestes genomas empregaram-se abordagens de bioinformática na análise de ESTs depositadas em banco de dados público do NCBI (www.ncbi.nlm.nih.gov). As 41558 ESTs foram reunidas num banco local e agrupadas por alinhamento, gerando 12922 grupos de sequências distribuídos em *singlets* (8110) e *contigs* (4812). Estes grupos foram comparados ao banco de dados não redundantes (NR, NCBI) e bancos especializados em elementos transponíveis (TE, RepBase e TEClass), por meio da ferramenta BLAST. Os resultados contra o NR mostraram que 10378 grupos apresentaram similaridade com sequências depositadas sendo que 142 apresentaram termo anotador característico de retrotransposons (transcriptase reversa foi o termo mais comum, com 84 registros). Contra os bancos específicos, 809 grupos revelaram similaridade, dos quais 342 com regiões de similaridade com retrotransposons; os outros 437 foram similares a transposons de DNA. Apenas 52 grupos foram comuns tanto ao NR quanto aos bancos específicos de elementos móveis. Destes, os que apresentaram os maiores valores de similaridade permitiram a identificação de cinco retrotransposons distintos nos genomas: - RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5. Os elementos foram caracterizados e classificados nas ordens LTR e LINE, superfamílias *Copia*, *Gypsy* e *I*, sustentáveis por análises filogenéticas. As ORFs desses elementos revelaram a presença de regiões características correspondentes a endonucleases, integrases, proteases, RNase H, transcriptases reversas e chromo-domínio. Foram identificadas as regiões de LTR dos elementos das superfamílias *Copia* e *Gypsy* (RtPb1, RtPb3, RtPb4 e RtPb5) e uma região de cauda poliA (RtPb2). Os elementos RtPb1 e RtPb2 (Araújo, 2008) cujas sequências completas não eram conhecidas tiveram nas completamente caracterizadas e mapeadas *in silico*. O mapeamento genômico, *in silico*, permitiu, ainda, identificar 538 cópias distribuídas entre os cinco elementos (RtPb1=52; RtPb2=101; RtPb3=18; RtPb4=65 e RtPb5=302), nos três genomas estudados (Pb01=366; Pb03=57; Pb18=115) (www.broadinstitute.org). Análises moleculares, usando a metodologia *primer walking*, por meio da técnica de PCR permitiram a amplificação dos elementos inteiros. A identidade das sequências foi comprovada por sequenciamento e alinhamento da região das transcriptases reversas. A presença de cada elemento em 31 diferentes isolados do fungo foi comprovada. Análises por hibridização de *Southern blots* do DNA genômico em isolados representativos das quatro espécies filogenéticas conhecidas indicaram a presença desses elementos em múltiplas cópias. A expressão dos retrotransposons foi observada empregando-se a técnica de PCR quantitativo qRT-PCR que revelou padrões de expressão diferentes entre os isolados estudados. O elemento mais expresso foi o RtPb1 no isolado Pb01. As análises indicaram a expressão peculiar de sequências características de diferentes retrotransposons em padrões distintos nos isolados representativos das quatro espécies filogenéticas conhecida.

1 – INTRODUÇÃO E JUSTIFICATIVA

O desenvolvimento da pesquisa genômica tem revelado a importância de se analisar e de se coordenar a análise da vasta informação acumulada com vistas a facilitar a manipulação de dados e avançar no conhecimento da biologia dos organismos estudados. Com relação aos fungos, estão hoje disponíveis 63 genomas completos e de cerca de 600 outros estão em diferentes etapas de sequenciamento (<http://www.genomesonline.org/cgi-bin/GOLD/index.cgi>). Para a aplicabilidade dos dados por eles gerados serão exigidos instrumentos eficazes de análise, coordenação e integração de informações.

Entre estes projetos citados acima destacamos o recém-concluído sequenciamento genômico do fungo dimórfico *Paracoccidioides brasiliensis* (isolados Pb01, Pb03 e Pb18), o patógeno responsável pela paracoccidioidomicose (PCM), doença de importância epidemiológica indiscutível na América Latina e no Brasil. A informação gerada será, certamente, um instrumento fundamental para a busca de métodos mais eficazes para a compreensão da patogênese da PCM, seu diagnóstico, prevenção e tratamento da doença, além de ser fundamental para a compreensão da biologia de seu agente.

Apesar da existência de trabalhos sobre genômica comparativa de fungos, existem poucos relatos da presença ou caracterização de elementos de DNA repetitivos no genoma de *P. brasiliensis*. Representando desde 0,5 a 10% do genoma dos fungos, as repetições mais frequentes são os micros e minisatélites, as repetições teloméricas, os retrotransposons e os transposons de DNA. Estas duas últimas classes de repetições são vistas como agentes centrais na evolução e reestruturação do genoma, atuando por meio de diferentes mecanismos como transposição, recombinação ectópica, transmissão horizontal e inativação epigenética. Os retrotransposons representam o objeto de estudo da presente proposta.

Tendo esta perspectiva, o projeto envolveu uma etapa inicial que utilizou abordagens de bioinformática associadas à utilização de matrizes de domínios, mapeamento de ESTs (*Expressed Sequence Tags*, etiquetas de sequências

expressas) e buscas por similaridade de sequências contra bancos de dados de domínio público visando à identificação, caracterização, mapeamento e anotação estrutural e funcional de retrotransposons no genoma de *P. brasiliensis*.

Uma vez definida a diversidade desses elementos no genoma, tendo como critérios iniciais de seleção a integridade de sequências codificadoras de enzimas envolvidas na mobilidade desses elementos e o contexto genômico, partiu-se para a caracterização experimental e funcional de alguns dos putativos retrotransposons.

O presente trabalho buscou integrar o estudo genômico *in silico* com a comprovação experimental, e tem potencial para contribuir para a compreensão de aspectos da biologia do fungo relacionados à estrutura, organização e funcionalidade do genoma, dos processos de recombinação e variação genética.

2 – REVISÃO DE LITERATURA

2-1 *Paracoccidioides brasiliensis* e Paracoccidioidomicose

Paracoccidioides brasiliensis é um fungo termodimórfico, capaz de alterar sua morfologia entre filamentosos e leveduriforme em função da temperatura. Este fungo é agente causador da paracoccidioidomicose (PCM), micose sistêmica prevalente na América Latina (Lacaz, 1994). No Brasil, a PCM constitui a primeira causa de morte entre as micoses sistêmicas quando a taxa de mortalidade é comparada com criptococose, candidíase, pneumocistose e histoplasmose, e a oitava causa de morte entre as doenças infecciosas e parasitárias recorrentes, sendo precedida pela doença de Chagas, tuberculose, AIDS, malária, esquistossomose, sífilis e hanseníase (Coutinho, Silva *et al.*, 2002; Prado, 2009). A doença é restrita à América Latina, havendo relatos na Colômbia, Venezuela e Argentina. O Brasil, país com alta taxa endêmica, responde por 80% dos casos relatados na literatura (Restrepo, Salazar *et al.*, 1984; Brummer, Castaneda *et al.*, 1993).

A doença e seu agente etiológico foram descritos pela primeira vez por Adolfo Lutz, em 1908, quando foi isolado de uma lesão na cavidade oral, apresentando aspecto de levedura com múltiplos brotamentos. A entidade foi denominada de *Granuloma pseudococcidium*. Posteriormente, com o aumento do número de casos envolvendo esta micose, houve avanço na descrição da doença e caracterização do agente etiológico que passou a ser chamado de *Zymonema brasiliensis*. No final da década de 1920, Floriano de Almeida reviu a classificação, adotando a denominação de *Paracoccidioides brasiliensis* (Bethlem, Capone *et al.*, 1999).

A classificação de *P. brasiliensis*, considerado até recentemente como pertencente ao filo Deuteromycota por não apresentar ou não se conhecer a fase sexuada ou teleomórfica, foi revista a partir de análises moleculares baseadas no alinhamento e comparação de sequências de genes RNA ribossômicos, sendo o *P. brasiliensis* classificado atualmente como pertencente ao reino Fungi, filo Ascomycota; subfilo Pezizomycotina, classe Eurotiomycetes, subclasse Eurotiomycetidae, ordem Onygenales e família Ajjelomycetaceae (Leclerc, Philippe *et al.*, 1994; Peterson e Sigler, 1998; Untereiner, Scott *et al.*, 2004).

À temperatura ambiente (23°C a 28°C), o fungo apresenta-se na forma micelial, crescendo como colônias brancas, aderindo firmemente ao meio com hifas delgadas, septadas que apresentam clamidósporos terminais ou intercalares. Quando cultivado entre 35°C e 37°C, o fungo apresenta-se leveduriforme formando colônias de coloração creme, aspecto cerebriforme, formadas por células arredondadas e multinucleadas (contendo de 2 a 6 núcleos), com parede celular espessa, entre 4 e 30 µm. Durante o ciclo celular são encontradas subpopulações celulares com diferentes conteúdos de DNA, característicos de células com múltiplos brotamentos ou polinucleadas. A primeira população é formada por células pequenas com apenas um núcleo, provavelmente as células filhas. A segunda população engloba células um pouco maiores, uninucleadas e binucleadas, que podem corresponder a células em diferentes estágios do ciclo celular e o terceiro grupo com células com morfologia heterogênea e um elevado número de núcleos por célula (Mcewen, Restrepo *et al.*, 1987; Feitosa, Cisalpino *et al.*, 2003; Almeida, Martins *et al.*, 2006).

O habitat ocupado por *P. brasiliensis* ainda é pouco conhecido, sabendo-se apenas que é exógeno ao ser humano. Neste caso, o fungo estaria presente no ambiente, viveria saprofiticamente sob a forma de micélio, em regiões úmidas como solos de florestas tropicais e subtropicais, com pequenas variações de temperatura, ou associados a vegetais, e produziria conídios capazes de dispersão aérea (Brummer, Castaneda *et al.*, 1993). Diversos trabalhos relataram o isolamento do fungo a partir do solo (Negroni, 1966; De Albornoz, 1971; Restrepo, Salazar *et al.*, 1984; Montenegro, Miyaji *et al.*, 1996; Lacaz, Vidal *et al.*, 1999). Há relatos do seu isolamento a partir de duas espécies de tatus, *Dasypus novemcinctus* (Bagagli, Sano *et al.*, 1998; Franco, Bagagli *et al.*, 2000; Bagagli, Franco *et al.*, 2003) e *Cabassous centralis* (Corredor, Peralta *et al.*, 2005), além do relato da sua presença em ração de cachorro, possivelmente contaminada com o solo, numa fazenda em área endêmica (Ferreira, Freitas *et al.*, 1990). Em 2002, relatou-se o primeiro caso confirmado de paracoccidioomicose em cães, diagnosticado por biópsia numa fêmea da raça Doberman, de vida estritamente urbana, que apresentava como manifestação clínica o enfartamento de linfonodos da cadeia ganglionar cervical. O diagnóstico etiológico foi confirmado sucessivamente por imuno-histoquímica com

anticorpos para antígeno específico do fungo e por técnicas moleculares (Fagundes, 2002). Recentemente, a doença foi diagnosticada novamente em cães da raça Doberman na região endêmica de Botucatu, SP, que apresentavam manifestações clínicas da doença (Ricci, Mota *et al.*, 2004; Farias, Werner *et al.*, 2005).

Um novo aspecto da ecologia do fungo foi o relato recente de sua ocorrência em animais silvestres. Por meio de técnicas moleculares detectou-se a presença de DNA do *P. brasiliensis* em carcaças de animais selvagens mortos em estradas na região endêmica de Botucatu, mostrando que a infecção desses animais silvestres deve ser comum em áreas endêmicas. Neste estudo foram detectados amplicons de genes específicos do *P. brasiliensis* por *Nested-PCR* em tecidos de diferentes animais, como a preá (*Cavia aperea*), porco espinho (*Sphiggurus spinosus*), quati (*Procyon cancrivorus*) e gambá (*Gallictis vittata*) (Richini-Pereira, Bosco *et al.*, 2008).

A paracoccidiodomicose (PCM) é uma doença sistêmica que apresenta duas formas clínicas progressivas, a forma aguda (juvenil ou linfática) e a forma crônica (do adulto ou pulmonar). Na forma aguda, a doença progride rapidamente por disseminação do fungo pelo sistema reticuloendotelial (baço, fígado, linfonodos, medula óssea), via linfática e hematogênica. A forma aguda acomete principalmente adultos jovens e crianças de ambos os sexos, sendo responsável por cerca de 3 a 5% dos casos de descritos, enquanto a forma crônica é a mais comum (90% dos casos), afetando principalmente trabalhadores rurais, do sexo masculino, na faixa etária entre 30 e 60 anos. Esta forma se desenvolve lentamente nos pulmões, podendo apresentar longo período de latência em focos quiescentes, manifestando lesões pulmonares ou evoluindo para outras lesões, disseminadas, na dependência da preservação da imunidade, pois ocorre comprometimento progressivo da resposta imune celular (Franco, Montenegro *et al.*, 1987).

A transmissão da doença não ocorre de indivíduo para indivíduo, sendo a maioria dos pacientes residentes em zonas rurais, podendo também acometer eventualmente pessoas que possuem pouco contato com o solo. A frequência do desenvolvimento da doença é maior em pessoas do sexo masculino, embora

a infecção ocorra em ambos os sexos. As evidências experimentais e dados histopatológicos indicam que a infecção ocorria por inalação de esporos ou conídios (formas infectantes) que no pulmão se converteriam em leveduras (formas parasitárias teciduais). É interessante destacar que antes da adolescência, não há diferença estatística na incidência de casos em crianças do sexo masculino e feminino. Assim, a ocorrência predominante no sexo masculino é justificada por evidências experimentais que indicam que fatores hormonais participariam da patobiologia da doença, especialmente o hormônio 17- β -estradiol que em níveis semelhantes aos níveis sanguíneos encontrados nas mulheres foi capaz de inibir a conversão de conídios para leveduras, *in vitro* (Restrepo, Salazar *et al.*, 1984; Brummer, Castaneda *et al.*, 1993; Coutinho, Silva *et al.*, 2002).

O diagnóstico da PCM é clínico e laboratorial, sendo realizado, neste último caso, por meio da visualização direta das leveduras de *P. brasiliensis* em amostras biológicas por meio do exame micológico direto e por meio de exames histopatológicos (biópsias, pus de linfonodos, escarro, aspirado brônquico ou lavado broncoalveolar, material granulomatoso da base de úlceras, líquido céfalo-raquidiano, medula óssea). A identificação é possibilitada pela observação de células com morfologia leveduriforme apresentando multibrotamentos característicos, muitas vezes comparados a roda-de-leme. Além do exame direto realizam-se cultivos a partir da semeadura de espécimes clínicos na tentativa de isolar o fungo, o que aumenta a confiabilidade do diagnóstico. Nos cultivos, além da morfologia leveduriforme característica pode-se induzir a conversão para a fase micelial, a partir da alteração da temperatura de incubação.

A combinação de exames sorológicos e micológicos é de grande importância para o diagnóstico e prognóstico da PCM, podendo a detecção de anticorpos específicos alcançar 100 % de positividade em pacientes não tratados. Entre as técnicas empregadas nesses exames podem ser citadas a imunodifusão dupla (ID), contraímunoeletroforese (CIE), imunofluorescência indireta (IFI), ELISA (*enzyme-linked immunosorbent assay*) e o *immunoblot*. A reação de ID é o principal método de diagnóstico sorológico da PCM devido ao baixo custo, simplicidade e especificidade do teste (Shikanai-Yasuda, Telles

Filho *et al.*, 2006). As preparações utilizadas são provenientes do sobrenadante de culturas, concentrado ou liofilizado, constituído de múltiplos componentes antigênicos, em geral glicoconjugados. Os resultados obtidos com o método ELISA, pela sua alta sensibilidade, também são de grande utilidade (Camargo, Gesztesi *et al.*, 1994). A especificidade dessas reações é atribuída ao principal componente antigênico do fungo, a glicoproteína de 43000 Dáltons (Gp43). Esta molécula foi muito bem caracterizada e sua reatividade comprovada estudando-se número elevado de soros de pacientes (Puccia, Schenkman *et al.*, 1986; Cisalpino, Puccia *et al.*, 1996; Diniz, Carvalho *et al.*, 2002).

O tratamento da PCM é prolongado, começando com dosagem agressiva de antifúngicos seguida por manutenção da terapia por um ou dois anos, às vezes mais. O tratamento da PCM emprega derivados sulfamídicos (sulfadiazina, associação sulfametoxazol-trimetoprim), anfotericina B, e derivados azólicos (cetoconazol, fluconazol e principalmente, o itraconazol). Os azólicos têm mostrado eficácia contra *P. brasiliensis*, causando menos efeitos colaterais quando comparados com a anfotericina B (Travassos, Silva *et al.*, 2004; Visbal, San-Blas *et al.*, 2005; Shikanai-Yasuda, Telles Filho *et al.*, 2006). Sugere-se o itraconazol como a melhor opção terapêutica para controle de formas leves e moderadas da doença, mas como não é usualmente fornecido pela rede pública, a combinação sulfametoxazol-trimetoprim tem sido a alternativa mais utilizada. Pacientes com formas mais graves devem receber anfotericina B ou associação de sulfametoxazol-trimetoprim por via intravenosa (Shikanai-Yasuda, Telles Filho *et al.*, 2006).

Embora os azólicos e outras drogas possam impedir a progressão da PCM, focos quiescentes podem persistir, constituindo-se em uma fonte de reincidência da doença após o término do tratamento. De fato, uma característica peculiar do *P. brasiliensis* é sua capacidade de reativar a doença (recidiva) após um período de dormência (Brummer, Castaneda *et al.*, 1993).

2-2 Estudos genéticos e genômicos de *P. brasiliensis*

Os primeiros relatos da utilização da tecnologia do DNA recombinante para estudos de taxonomia em *P. brasiliensis* surgiram a partir dos anos 90 (Leclerc, Philippe *et al.*, 1994), década que marcou também o início da utilização

das primeiras técnicas de biologia molecular para a clonagem e caracterização de genes de *P. brasiliensis*. Entre as primeiras técnicas de genotipagem empregadas estão a amplificação randômica de DNA polimórfico (RAPD) (Soares, Madlun *et al.*, 1995) eletroforese de campo pulsado (PFGE) (Cano, Cisalpino *et al.*, 1998) e análises do polimorfismo de tamanho de fragmentos gerados por endonucleases de restrição (RFLP) (Calcagno, Nino-Vega *et al.*, 1998).

Soares e colaboradores (1995) empregando a técnica de RAPD, (Soares, Madlun *et al.*, 1995) utilizaram cinco oligonucleotídeos iniciadores aleatórios para distinção de sete isolados de *P. brasiliensis*. O perfil de amplificação do DNA permitiu a separação dos isolados em dois grupos com apenas 35 % de identidade genética. Neste estudo foi possível observar que, com todos os iniciadores utilizados, os isolados Pb01 (Brasil, Goiás) e 7455 (Equador) sempre se apresentaram separados dos outros. A eletroforese de pulsos alternados (PFGE) permite a caracterização genômica, mapeamento gênico e biotipagem epidemiológico-molecular de amostras de microrganismos refratários à análise genética clássica. Cano e colaboradores (1998) utilizaram esta técnica para mapear genes específicos às bandas cromossômicas determinando a identidade de algumas bandas em dois isolados que apresentam perfis cariotípicos distintos (Cano, Cisalpino *et al.*, 1998). Feitosa e colaboradores descreveram a resolução e comparação do cariótipo eletroforético de 12 isolados do *P. brasiliensis* de diferentes regiões geográficas, identificaram seis perfis cariotípicos distintos e extenso polimorfismo cromossômico entre os isolados. O mapeamento de nove marcadores genéticos, correspondentes a genes previamente caracterizados, permitiu observar que alguns genes que hibridizavam em uma mesma banda cromossômica em alguns isolados, hibridizavam em bandas diferentes em outros isolados, sugerindo a ocorrência de translocações e translocações recíprocas, observando-se também duplicações e deleções, alterações que no seu conjunto devem contribuir para o polimorfismo cromossômico em *P. brasiliensis* (Feitosa, Cisalpino *et al.*, 2003).

A técnica de RFLP foi utilizada para caracterizar 32 isolados de *P. brasiliensis* de diferentes regiões da América Latina (Calcagno, Nino-Vega *et al.*, 1998). Dezesesseis diferentes padrões de restrição foram identificados com a

enzima de restrição *Hinf* I, e apenas cinco padrões com a enzima *Hinc* II. Os dados obtidos na análise por RFLP, usando ambas as endonucleases de restrição, permitiram, com pequenas exceções, agrupar os isolados de acordo com sua origem geográfica.

Entre os primeiros genes de *P. brasiliensis* a serem clonados e sequenciados estão vários genes que codificam antígenos, pois a intensa busca por um diagnóstico imunológico específico da doença nas décadas de 70 e 80. O primeiro gene do fungo a ser clonado correspondeu ao gene codificador do antígeno Gp43 (glicoproteína de 43000 Da), antígeno imunodominante na resposta humoral (Cisalpino, Puccia *et al.*, 1996).

Além das técnicas de biologia molecular utilizadas nos estudos com *P. brasiliensis*, surgiu também, no final da década de 90, a aplicação de análises proteômicas, importante ferramenta para identificação de proteínas expressas nas fases micelial e leveduriforme do fungo (Salem-Izacc, Jesuino *et al.*, 1997; Cunha, Sousa *et al.*, 1999; Da Fonseca, Jesuino *et al.*, 2001). Logo depois, a técnica de eletroforese bidimensional associada à detecção imunológica das frações proteicas do fungo, permitiu o sequenciamento peptídico, clonagem e caracterização de antígenos que apresentam homologia com HSP60 (Izacc, Gomez *et al.*, 2001; Cunha, Zancopé-Oliveira *et al.*, 2002), manosiltransferase (Costa, Gómez *et al.*, 2002), gliceraldeído fosfato desidrogenase (GAPDH) (Barbosa, Cunha Passos *et al.*, 2004), catalase P (Moreira, Bailão *et al.*, 2004), de um antígeno homólogo a triose fosfato isomerase (Moreira, Bailão *et al.*, 2004) e da formamidase (Goldman, Dos Reis Marques *et al.*, 2003; Borges, Pereira *et al.*, 2005).

Os primeiros trabalhos relatando a aplicação de técnicas genômicas em larga escala em *P. brasiliensis* surgiram no ano de 2003 quando dois grupos de pesquisas brasileiros desenvolveram, de forma independente, procedimentos para a utilização de etiquetas de sequências expressas (*EST*) como ferramenta para estudos genômicos neste fungo (Felipe, Andrade *et al.*, 2003; Goldman, Dos Reis Marques *et al.*, 2003). No primeiro trabalho, os autores descreveram o transcriptoma do isolado Pb01, identificando genes preferencialmente expressos na fase micelial ou leveduriforme. Foram anotadas 3938 *EST* da fase micelial e

leveduriforme, agrupados em 597 *contigs* e 1563 *singlets*, resultando em 2160 genes putativos, que poderiam corresponder a um quarto dos genes do fungo. Desses, 894 foram classificados em 18 categorias funcionais, sendo 44% do metabolismo celular, 25% do armazenamento e processamento de informações, 19% associados à divisão celular e processamento pós-transcricional, e 12% não tiveram sua função determinada. As análises computacionais indicaram vários genes que devem estar envolvidos na transição dimórfica, genes expressos somente em uma das fases (micelial ou leveduriforme), ou seja, genes em estágios específicos e genes envolvidos na resistência a drogas (Felipe, Andrade *et al.*, 2003)

No segundo trabalho os autores, analisaram 13490 *EST* perfazendo 4692 genes expressos do isolado Pb18. A comparação com sequências depositadas em banco de dados de outros microrganismos permitiu a identificação de vários genes homólogos a fatores de virulência e patogenicidade em *Candida albicans*, ligados ao metabolismo de aminoácidos e nucleotídeos (7 genes), a biossíntese da parede celular e adesão (13 genes), transdução de sinais (6 genes) e alguns outros (12 genes) com potencial para envolvimento tanto na transição dimórfica quanto no processo infeccioso, como aqueles para proteínas de choque térmico e calmodulina. A análise da expressão de 13 desses genes durante a transição de micélio para levedura, e vice-versa foi confirmada por RT-PCR em tempo real. Os genes foram classificados em três grupos, dois grupos que apresentaram alta expressão na fase micelial, enquanto o terceiro inclui genes com significativa expressão durante a fase leveduriforme (Goldman, Dos Reis Marques *et al.*, 2003).

Felipe e colaboradores (2005) identificaram 19718 *EST* (9777 de levedura e 9941 de micélio) distribuídos em 6022 grupos, representando genes expressos. Este estudo revelou que 68.5% dos genes de *P. brasiliensis* são relacionados com genes de outros fungos, sendo 52,9% ortólogos aos genes de *Aspergillus nidulans*. No entanto, 30,2% dos genes encontrados aparentemente representam novos genes, exclusivos de *P. brasiliensis*. Os grupos foram distribuídos em 18 categorias funcionais: metabolismo celular (29%), transcrição (12%), síntese de proteína (15%), produção bioenergética (9%), controle da

organização celular (4%), transdução de sinais, relacionados com o dimorfismo (4%) e outros (Felipe, Andrade *et al.*, 2005).

Matute (2006), baseando em análise filogenética de 65 isolados de *P. brasiliensis*, propuseram a existência de três diferentes espécies: S1, (espécie filogenética 1 com 38 dos isolados estudados, distribuídos pelo Brasil, Argentina, Paraguai, Peru e Venezuela), PS2 (espécie filogenética 2 com 6 isolados, cinco do Brasil e um da Venezuela) e PS3 (espécie filogenética 3 com 21 isolados). As genealogias de quatro das regiões estudadas apoiam fortemente o grupo PS2, composto por cinco isolados brasileiros e um venezuelano. O segundo grupo, PS3, composto somente por 21 isolados colombianos, encontra apoio na genealogia do gene da α -tubulina. A distribuição de duas das três linhagens, S1 e PS2, territorialmente ampla, sugere que haja outras barreiras para o fluxo genético, além do isolamento geográfico. Esse estudo sugeriu a possível ocorrência de reprodução sexuada em *P. brasiliensis*, na espécie filogenética S1, mas não nas outras duas linhagens propostas (Matute, Sepulveda *et al.*, 2006).

Em outro estudo filogenético, utilizando sequências codificadoras (*CHS4*, *Actin*, *ODC*, *URA3*, *CHS2*, *HSP70*, *FKS1*, *Hydrophobin*, *Kex1*, *Catalase A*, *Catalase P*, *Formamidase*, *Glyoxalase*) e sequências ITS (sequências ribossômicas), englobando 21 isolados de *P. brasiliensis*, foi possível agrupar todos os isolados nas mesmas três linhagens filogenéticas, exceto para o isolado Pb01 que se mostrou filogeneticamente muito distante de todos os outros. Os autores sugeriram que o isolado Pb01 poderia corresponder a uma quarta linhagem filogenética ainda desconhecida no gênero *Paracoccidioides*; contudo seria necessário comprovar a existência e estudar outros isolados que se assemelham ao Pb01 (denominados *Pb01-like*) (Carrero, Niño-Vega *et al.*, 2008). De fato, em 2009, Teixeira e colaboradores ao estudarem 88 isolados de diferentes origens geográficas verificaram que 17 deles, oriundos da região centro-oeste do Brasil (estados de Mato Grosso, Mato Grosso do Sul e Goiás) eram geneticamente semelhantes ao Pb01. Os isolados *Pb01-like* constituíram um *cluster* ou linhagem de tal modo divergente do *cluster* de *P. brasiliensis* que continha as três linhagens previamente descritas (S1, PS2 e PS3) (divergindo entre 8,3 e 9,8 no conjunto das análises dos 13 *loci* sequenciados com

percentual de divergência semelhante àquele observado entre as espécies *Magnaporthe grisea* e *M. oryzae*, por exemplo) que os autores propuseram ser esta uma espécie nova no gênero que receberia o nome de *Paracoccidioides lutzii* em homenagem a Adolf Lutz, descobridor da doença e de seu agente etiológico (Teixeira, Theodoro *et al.*, 2009a).

A iniciativa de realizar o sequenciamento genômico deste patógeno veio associada ao projeto denominado *Fungal Genome Initiative* (FGI), executado pelo *Broad Institute of MIT and Harvard* que previa o sequenciamento de fungos patogênicos dimórficos, entre os quais *Coccidioides immitis*, *Blastomyces dermatitidis* e *Histoplasma capsulatum* (*Ajjelomyces deramtitidis*, *A. capsulatum*). O projeto teve como objetivo examinar a diversidade genética, determinar as características genômicas comuns e peculiares dessas linhagens, bem como comparar os genomas completos de todos os fungos patogênicos dimórficos, entre si e com genomas de outros fungos patogênicos.

Atualmente, está disponível no sítio do *Broad Institute* (<http://www.broad.mit.edu/science/projects/projects>) o resultado do sequenciamento gnômico de fungos patogênicos que inclui várias linhagens desses fungos e três isolados de *P. brasiliensis* (Pb01, Pb03 e Pb18). Para tanto, estabeleceu-se um banco de dados no qual se encontram disponíveis as sequências genômicas das linhagens de todos os patógenos, inclusive as três de *Paracoccidioides*. É interessante destacar que o isolado Pb01 representa a nova espécie proposta, ou seja, *Paracoccidioides lutzii*, e os isolados Pb03 e Pb18 representam respectivamente as linhagens filogenéticas PS2 e S1 de *P. brasiliensis*. No banco de dados do *Broad Institute* as sequências genômicas dos isolados foram reunidas e apresentadas em supercontigs (*scaffolds*), assim distribuídos: 111 *supercontigs* do isolado Pb01, 65 *supercontigs* do isolado Pb03 e 57 *supercontigs* para o isolado Pb18.

De posse dos resultados de sequenciamento, Desjardins *et al* (2011) apresentaram resultados de uma análise genômica comparativa dos agentes da paracoccidioidomicose entre si e com outros fungos, buscando entender melhor as bases genéticas do seu crescimento e patogenicidade. Utilizando abordagens computacionais os autores caracterizaram as famílias gênicas e rotas

metabólicas relacionadas com o metabolismo de carboidratos e proteínas de *Paracoccidioides* e do grupo de fungos dimórficos. Pela comparação dos genomas foi possível analisar outros parâmetros, tais como: características dos genomas, evolução de famílias gênicas, presença de reprodução sexual e identificação de potenciais alvos de drogas. Considerando as características dos genomas, os autores mostraram diferenças entre os isolados. Os genomas de *Paracoccidioides brasiliensis* Pb03 e Pb18 se mostraram similares entre si em tamanho (30.0 Mb e 29.1 Mb, respectivamente, com cerca de 8000 genes preditos) e um pouco menores que o de *P. lutzii* (32.9 Mb, com 9.132 genes preditos). Conquanto, os três genomas apresentassem alta sintonia, a similaridade de sequências entre Pb03 e Pb18 (*P. brasiliensis*) é de 96% enquanto destes com *P. lutzii* (Pb01) é de 90%. Indicou-se que os elementos repetitivos compunham aproximadamente 8% do genoma de *P. brasiliensis* (Pb03 e Pb18, 7,7 e 9,2%, respectivamente) e quase o dobro em *P. lutzii* (Pb01, 16%). Os autores relataram a expansão de uma família de quinases específica de fungos (*FUNK1*) em *Paracoccidioides* e por outro lado, relataram que os fungos dimórficos, como um grupo, perderam muitos genes no metabolismo de carboidratos ao passo que retiveram a maioria dos genes para protease (Desjardins, Champion *et al.*, 2011).

2-3 Elementos genéticos móveis ou transponíveis: transposons e retrotransposons.

2.3.1 Descoberta, ubiquidade e abundância

Os elementos genéticos móveis foram descobertos na década de 40 do século 20, por Bárbara McClintock, ao estudar o padrão variegado de cores do milho. Neste trabalho, foram identificados dois loci denominados *Activator* (Ac) e *Dissociator* (Ds) que possuíam capacidade de interagir entre si. McClintock percebeu, pelas mudanças de coloração do pericarpo dos grãos, que os elementos eram capazes de se mover de um lugar para outro no cromossomo (McClintock, 1950). Graças ao seu trabalho, no ano de 1983, 40 anos após a descrição de seus elementos, McClintock recebeu o prêmio Nobel. Seus experimentos, 60 anos atrás, levaram-na a propor que células sob estresse ambiental ativariam elementos transponíveis para reestruturar o genoma celular (Ebina e Levin, 2007).

Em trabalhos que foram realizados em diferentes espécies a partir do final da década de 70, foram identificados elementos transponíveis em bactérias, levedura e *Drosophila* (Capy, Gasperi *et al.*, 2000; Biémont e Vieira, 2006). Atualmente, sabe-se que esses elementos estão presentes no genoma da maioria das espécies de Eubactéria, Archeobactérias e Eucariotos, exceto em *Plasmodium falciparum* e provavelmente em espécies relacionadas (Wicker, Sabot *et al.*, 2007).

Os elementos genéticos móveis são definidos como sequências de DNA capazes de se mover pelo genoma, contendo toda a informação necessária para sua mobilização (Daboussi e Capy, 2003; Bigot, Brillet *et al.*, 2005; Biémont e Vieira, 2006). Atualmente, são denominados como elementos transponíveis (TE – *Transposable elements*). Apresentam grande diversidade de sequências e de distribuição nas diferentes espécies constituindo, em plantas até 80% ou mais do DNA genômico total (Morgante, Brunner *et al.*, 2005; Wicker, Sabot *et al.*, 2007). Estima-se que a abundância desses elementos corresponda a, respectivamente, 22% e 50% dos genomas de *Drosophila melanogaster* e do milho. De acordo com os dados do *International Human Genome Sequence Consortium*, 45% do genoma humano é constituído por TE, entre os quais os retrotransposons LINE1 que chegam a representar 17% do genoma (Lander, Linton *et al.*, 2001). Em fungos, esses elementos também têm sido identificados como parte substancial do genoma (3 – 20%) (Daboussi e Capy, 2003; Wicker, Sabot *et al.*, 2007).

2.3.2 Características dos elementos transponíveis das Classes I e II e sua classificação.

Os elementos transponíveis encontrados nos genomas eucariotos estão divididos em duas classes (I e II) que diferem segundo o seu mecanismo de transposição e sua organização estrutural. Os elementos agrupados na classe I são denominados retrotransposons e são todos aqueles que possuem ou apresentam um intermediário de RNA no evento de transposição. Nenhum dos membros dessa classe cliva ou transfere fitas de DNA do sítio doador. Neste caso, o intermediário de RNA é transcrito a partir de uma cópia do elemento inserida no genoma (cópia genômica, DNA, semelhante a um retrovírus endógeno ou ERV). Esse RNA transcrito é, sucessivamente, utilizado como

molde no processo de transcrição reversa, ou seja, sofre a ação de uma enzima denominada transcriptase reversa (RT) resultando na formação de DNA de dupla fita, uma nova cópia à semelhança de um provírus, que será inserido no genoma. Os elementos de classe II, conhecidos com transposons de DNA, são encontrados em quase todos os eucariotos e, em procariotos, recebem o nome de sequências de inserção (IS). Os elementos de classe II se movimentam diretamente como sequências de DNA e nenhum deles se move por meio de um intermediário de RNA (Daboussi e Capy, 2003; Wicker, Sabot *et al.*, 2007).

A primeira classificação dos elementos transponíveis foi feita, portanto, com base na presença desses diferentes intermediários da transposição e os dividiu considerando esses mecanismos: Intermediário de RNA (Classe I ou retrotransposon) ou de DNA (Classe II ou DNA transposon). O mecanismo de transposição da classe I é comumente chamado de copiar-e-colar e o da classe II chamado de cortar-e-colar (Finnegan, 1989). Com a descoberta de elementos em bactérias e eucariotos capazes de copiar e colar sem um intermediário de RNA e de elementos não-autônomos altamente reduzidos chamados de MITES (*Miniature inverted repeats transposable elements*) tornou-se necessária a elaboração de um novo sistema de classificação. O grande volume de dados referentes à identificação, caracterização e à diversidade de elementos transponíveis nas diferentes espécies reforçou os argumentos a favor da necessidade de se estabelecerem critérios e modelos de comparação que organizassem as semelhanças e diferenças estruturais e funcionais observadas no conjunto de elementos já descritos. Tal procedimento facilitaria a identificação de novos elementos, seus reconhecimentos e classificações.

Wicker e colaboradores (2007) foram os autores que propuseram o primeiro sistema unificado e hierarquizado de classificação que, mantendo as classes utilizadas anteriormente e aplicando critérios mecânicos e baseados na análise da estrutura e organização das enzimas responsáveis pela mobilização dos elementos transponíveis, permite o seu agrupamento em níveis menores. Este sistema inclui, em ordem hierárquica, os níveis de classe, subclasse, ordem, superfamília, família e subfamília (Figura 1) (Wicker, Sabot *et al.*, 2007). Classe: divide os TEs considerando a presença ou ausência de um intermediário de RNA. Subclasse: usado para distinguir elementos que utilizam

o mecanismo cortar-e-colar daqueles que deixam o sítio doador e se integram em outro lugar no genoma. Ordem: considera a organização enzimática. Superfamília: agrupam elementos que compartilham a mesma estratégia de replicação e diferem na estrutura proteica e em domínios codificadores, na presença e tamanho dos TSD (*target site duplication*) que são duplicações diretas geradas nas regiões flangeadoras da região de inserção. Família: definida pela conservação das sequências de DNA (A similaridade no nível de aminoácidos é geralmente elevada entre as proteínas codificadas por famílias diferentes que pertencem à mesma superfamília). Subfamília: é definida com base em dados filogenéticos e podem, em casos específicos, servir para distinguir internamente populações de elementos autônomos daqueles não autônomos (Wicker, Sabot *et al.*, 2007).

Classificação	Estrutura	Duplicação	Código	Ocorrência
Ordem	Superfamília			
Classe I (retrotransposons)				
LTR	<i>Copia</i>	→ GAG AP INT RT RH →	4-6	RLC P, M, F, O
	<i>Gypsy</i>	→ GAG AP RT RH INT →	4-6	RLG P, M, F, O
	<i>Bel-Pao</i>	→ GAG AP RT RH INT →	4-6	RLB M
	<i>Retrovirus</i>	→ GAG AP RT RH INT ENV →	4-6	RLR M
	<i>ERV</i>	→ GAG AP RT RH INT ENV →	4-6	RLE M
DIRS	<i>DIRS</i>	→ GAG AP RT RH YR ←	0	RYD P, M, F, O
	<i>Ngaro</i>	→ GAG AP RT RH YR →	0	RYN M, F
	<i>VIPER</i>	→ GAG AP RT RH YR →	0	RYV O
PLE	<i>Penelope</i>	← RT EN →	Variable	RPP P, M, F, O
LINE	<i>R2</i>	— RT EN —	Variable	RIR M
	<i>RTE</i>	— APE RT —	Variable	RIT M
	<i>Jockey</i>	— ORF1 — APE RT —	Variable	RIJ M
	<i>L1</i>	— ORF1 — APE RT —	Variable	RIL P, M, F, O
	<i>I</i>	— ORF1 — APE RT RH —	Variable	RII P, M, F
SINE	<i>tRNA</i>	— — —	Variable	RST P, M, F
	<i>7SL</i>	— — —	Variable	RSL P, M, F
	<i>5S</i>	— — —	Variable	RSS M, O
Classe II (transposons de DNA) - Subclasse 1				
TIR	<i>Tc1-Mariner</i>	→ Tase* ←	TA	DTT P, M, F, O
	<i>hAT</i>	→ Tase* ←	8	DTA P, M, F, O
	<i>Mutator</i>	→ Tase* ←	9-11	DTM P, M, F, O
	<i>Merlin</i>	→ Tase* ←	8-9	DTE M, O
	<i>Transib</i>	→ Tase* ←	5	DTR M, F
	<i>P</i>	→ Tase ←	8	DTP P, M
	<i>PiggyBac</i>	→ Tase ←	TTAA	DTB M, O
	<i>PIF-Harbinger</i>	→ Tase* — ORF2 ←	3	DTH P, M, F, O
	<i>CACTA</i>	→ Tase — ORF2 ←	2-3	DTC P, M, F
Crypton	<i>Crypton</i>	→ YR ←	0	DYC F
Classe II (transposons de DNA) - Subclasse 2				
Helitron	<i>Helitron</i>	→ RPA — Y2 HEL ←	0	DHH P, M, F
Maverick	<i>Maverick</i>	→ C-INT — ATP — CYP — POL B ←	6	DMM M, F, O

Aspectos estruturais			
→ Terminações longas repetidas	← Terminações invertidas repetidas	Região codificadora	Região não codificadora
— Motivo diagnóstico em região não codificadora		— / — Região que pode conter uma ou mais ORF adicionais	
Domínios das proteínas codificadas			
AP, Proteinase aspártica	APE, Endonuclease apurínica	ATP, ATPase	C-INT, C-integrase
ENV, Proteína do envelope	GAG, Proteína do capsídeo	HEL, Helicase	INT, Integrase
POL B, DNA Polimerase B	RH, RNase H	RPA, Proteína de replicação	CYP, Proteinase de cisteína
Tase, Transposase (* com domínio DDE)		YR, Recombinase de tirosina	EN, Endonuclease
			ORF, Fase de leitura aberta
			RT, Transcriptase reversa
			Y2, YR com domínio YY
Species groups			
P, Planta	M, Metazoários	F, Fungos	O, Outros

Figura 1 – Sistema de classificação

Figura adaptada de Wicker, 2007. O sistema hierárquico de classificação divide os elementos transponíveis em duas classes baseado na presença e ausência de RNA, como intermediário de transposição. As subdivisões em subclasse, ordem e superfamília são baseadas no mecanismo cortar-colar (classe II), organização enzimática e nos domínios codificadores, respectivamente. O tamanho da duplicação do sítio de inserção (TSD) é utilizado na classificação. DIRS, *Dictyostelium intermediate repeat sequence*; LINE, *long interspersed nuclear element*; LTR, *long terminal repeat*; PLE, *Penelope-like elements*; SINE, *short interspersed nuclear element*; TIR, *terminal inverted repeat*.

2.3.3 Principais características de retrotransposons (elementos da Classe I) segundo a classificação (Wicker, Sabot *et al.*, 2007).

Os retrotransposons (Classe I) foram divididos em cinco ordens: retrotransposons LTR (long terminal repeat), elementos DIRS-like (*Dictyostelium intermediate repeat sequence*), elementos Penelope-like (PLEs), LINE (*long interspersed nuclear element*) e SINE (*short interspersed nuclear element*) (Wicker, Sabot *et al.*, 2007).

É interessante observar que, por sua semelhança estrutural com retrovírus (família Retroviridae), os retrotransposons da ordem LTR (Wicker, Sabot *et al.*, 2007), podem ser também estudados de acordo com a classificação do Comitê Internacional de Taxonomia Viral (ICTV), nomenclatura utilizada por muitos autores (www.ncbi.nlm.nih.gov/ICTVdb; <http://ictvonline.org>). Nessa abordagem, os retrotransposons são subdivididos em três grupos distintos: Pseudoviridae (elementos *Copia*/Ty1), Metaviridae (*Gypsy*/Ty3), sendo estes dois grupos existentes em plantas, fungos e metazoários; Semotivirus (Bel/Pao), um grupo de ocorrência exclusiva em metazoários, que nessa classificação é apresentado como um gênero de Metaviridae. Há comentários na literatura de que a classificação do ICTV precisa ser revisada para atender a diversidade atual dos retrotransposons (Benachenhou, Sperber *et al.*, 2013).

Por outro lado, é interessante destacar que os retrovírus diferem dos retrotransposons, a partir do quais podem ter evoluído por meio da aquisição de sequências podendo citar-se os genes *env* (envelope viral) que conferem infectividade e transmissibilidade, enfim, mobilidade extracelular. Entre os aspectos distintos que podemos mencionar, e até decorrente da não transmissibilidade extracelular dos retrotransposons, está o fato de a evolução viral ser mais influenciada pela epidemiologia do que pelas filogenias dos organismos. Outro aspecto a diferenciá-los se refere ao fato de apesar das semelhanças da ORF codificadora da Gag entre TE da superfamília Gypsy e retrovírus, a dos primeiros codifica apenas a função da matriz enquanto a de retrovírus codifica capsídeo e matriz. Importante também lembrar a existência de retrovírus que, por perda de infectividade associada à deleção de domínios que

habilitam a mobilidade extracelular, são descritos como retrovírus endógenos (ERV) (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b).

Os elementos da ordem LTR podem variar em tamanho atingindo até 25 Kb. Possuem duas terminações longas repetidas (LTRs), repetições diretas, semelhantes às de retrovírus, que flanqueiam a sequência codificadora. As LTR apresentam as regiões U3-R-U5, que abrigam motivos regulatórios para transcrição e regulação da integração, caracterizando-se também pela presença das bases 5'-TG-3' no início da sequência e as bases 5'-CA-3' no final da sequência. A região central R é encontrada repetida nas duas extremidades dos transcritos. Para a maioria dos retrotransposons a síntese da primeira fita é iniciada pelo anelamento da região 3' de um tRNA com um sítio de ligação próximo à LTR esquerda; já a segunda fita é iniciada a partir de uma trilha de poli-purinas próximo à LTR da direita. As LTR podem variar, tanto em tamanho, desde centenas de pares de base até 5 – 6 Kb, como também apresentam grande variabilidade nas sequências de bases o que dificulta estudos comparativos. Esses elementos possuem ORF (“Open Reading frame”, fase aberta de leitura) para GAG, proteína estrutural que compõe partículas semelhantes a vírus, e para POL, uma poliproteína. Esta poliproteína pode possuir domínios catalíticos de proteinase aspártica (PR AP), integrase (IN), transcriptase reversa (RT) e ribonuclease H (RNase H ou RH) (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008a; b; Finnegan, 2012)

Segundo Wicker e colaboradores, há 5 superfamílias na ordem LTR: Copia, Gypsy, Bel-Pao, Retrovirus e Retrovirus endógenos (ERV). As superfamílias Gypsy e Copia, da ordem LTR, receberam essa denominação a partir da caracterização no organismo modelo *Drosophyla melanogaster* de elementos representativos, que são frequentemente mencionadas em associação com elementos dessas mesmas superfamílias, respectivamente, Ty3 e Ty1, descritos noutro organismo modelo, a levedura *Saccharomyces cerevisiae*. As estruturas consenso desses elementos são semelhantes às de retrovírus, exceto pela ausência do gene *env*. Na maioria dos elementos, observam-se os genes *gag* e *pol*, havendo também numerosos exemplos de fusão desses dois genes em uma única ORF. O gene *pol* codifica os domínios enzimáticos já mencionados (PR, RT RH e IN). Admite-se que haja equivalência

funcional desses domínios com os de retrovírus, pois ensaios funcionais foram realizados apenas em levedura, e na maioria dos casos admite-se que os mecanismos de transposição dos retrotransposons LTR seja semelhante ao de retrovírus (Eickbush & Jamburuthugoda, 2008).

As superfamílias *Copia* e *Gypsy* diferem na posição ou ordem sequencial das ORF que codificam a RT e IN na POL (Wicker, Sabot *et al.*, 2007). Há duas maneiras pelas quais o arranjo dos domínios proteicos codificados pelos genes *pol* estão alterados comparativamente aos de retrovírus: 1) na superfamília *Copia* (Ty1) o domínio IN localiza-se na região amino-terminal relativamente aos domínios RT e RH enquanto em *Gypsy* e *Bel-Pao* o domínio IN se localiza na região C-terminal (Figura 1); 2) nos retrovírus a RT e RH estão separadas por um domínio de conexão (domínio *tether*). As sequências das RH de retrovírus são muito divergentes daquelas dos retrotransposons LTR; esse domínio RH atual seria um novo domínio adquirido pelos retrovírus e a conexão (*tether*) seria o antigo domínio RH ancestral, hoje inativo como tal, mas provavelmente envolvido na regulação da RH ativa. No retrovírus, os genes *env* seguem-se ao domínio IN (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b).

A figura 2 mostra um ciclo teórico de vida de um retrotransposon LTR. Inicialmente, ocorre a transcrição da informação inserida no genoma hospedeiro (DNA) em uma molécula de mRNA (A) pela ação da enzima RNA *pol* II celular. Em seguida a informação é traduzida gerando as proteínas (B) *Gag* e *Pol*, características de retrotransposons e retrovírus. A poliproteína *Pol* é, posteriormente, clivada pela protease em três subunidades, AP (protease aspártica), RT-RNaseH e IN (Integrase). Um processo de dimerização das moléculas de mRNA ocorre (C) antes do empacotamento desse dímero pelas VLP (*virus-like particles*) que, por sua vez, são formadas pela polimerização das *Gags*. As proteínas RT-RNaseH e IN são envolvidas juntamente com o dímero de RNA iniciando-se o processo de transcrição reversa pela ação da enzima RT-RNaseH utilizando o mRNA como molde (D). Após a síntese da primeira fita de cDNA ocorre a degradação da fita de RNA utilizada como molde e síntese da segunda fita (E). Em (F) o processo de síntese é finalizado e ocorre a ligação das enzimas IN nas extremidades das LTRs. Para finalizar o ciclo ocorre a integração da nova cópia do elemento (G) pela quebra das fitas de DNA no sítio

de inserção, catalisada pela ação das integrases ligadas às LTR do novo elemento.

O processo de transcrição reversa é complexo e composto de várias etapas que ocorrem nas VLP, no citoplasma celular (Figura 3). Inicialmente ocorre a síntese da molécula de DNA a partir de RNA molde na sua extremidade 5'. A síntese começa a partir de um iniciador tRNA oriundo da célula hospedeira (a,b), seguido-se a degradação da molécula de RNA (c). Em seguida, ocorre um salto do fragmento de DNA recém-sintetizado para a extremidade 3' do RNA (d) que permite a síntese completa da primeira fita de cDNA (e). A próxima etapa é catalisada pela enzima RT-RNaseH que degrada grande parte da fita de RNA utilizada como molde (f). O restante da molécula de RNA que não sofreu degradação é então utilizado como iniciador para a síntese da fita complementar de DNA (g). Um segundo salto ocorre (h) e permite a síntese completa desta segunda fita. Durante todo o processo as LTRs do elemento são formadas.

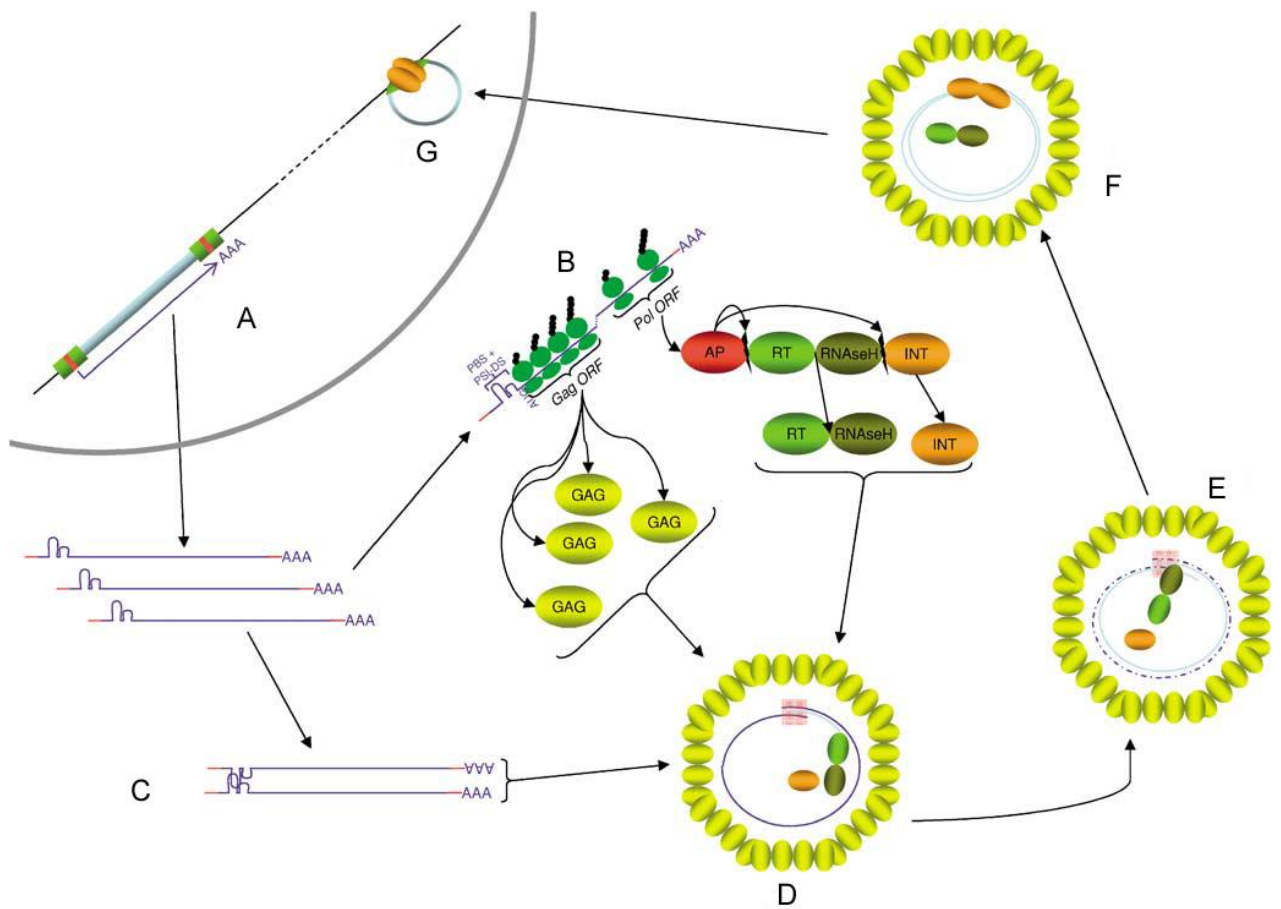


Figura 2 – Ciclo teórico de replicação de retrotransposon

Figura adaptada (Sabot e Schulman, 2006). Em (A) Transcrição do mRNA. (B) Tradução e síntese das proteínas codificadas pelos genes *gag* e *pol* dos elementos ativos. A poliproteína codificada pelo gene *pol* é posteriormente clivada pela protease (AP) em AP, Transcriptase reverse-RNaseH (RT-RNaseH) e Integrase (IN). (C) Dimerização de RNA. (D) Empacotamento do dímero de RNA juntamente com as proteínas RT-RNaseH e IN na vesícula (VLP- *Virus-like particle*) formada pela polimerização dos produtos do gene *gag*. Dentro dessa vesícula ocorre a transcrição reversa e síntese do cDNA. (E) Degradação do RNA e início da síntese da segunda fita de DNA. (F) Término da síntese de DNA e ligação das proteínas integrase (IN) nas terminações longas repetidas (LTR). (G) integração da nova cópia do elemento no genoma hospedeiro.

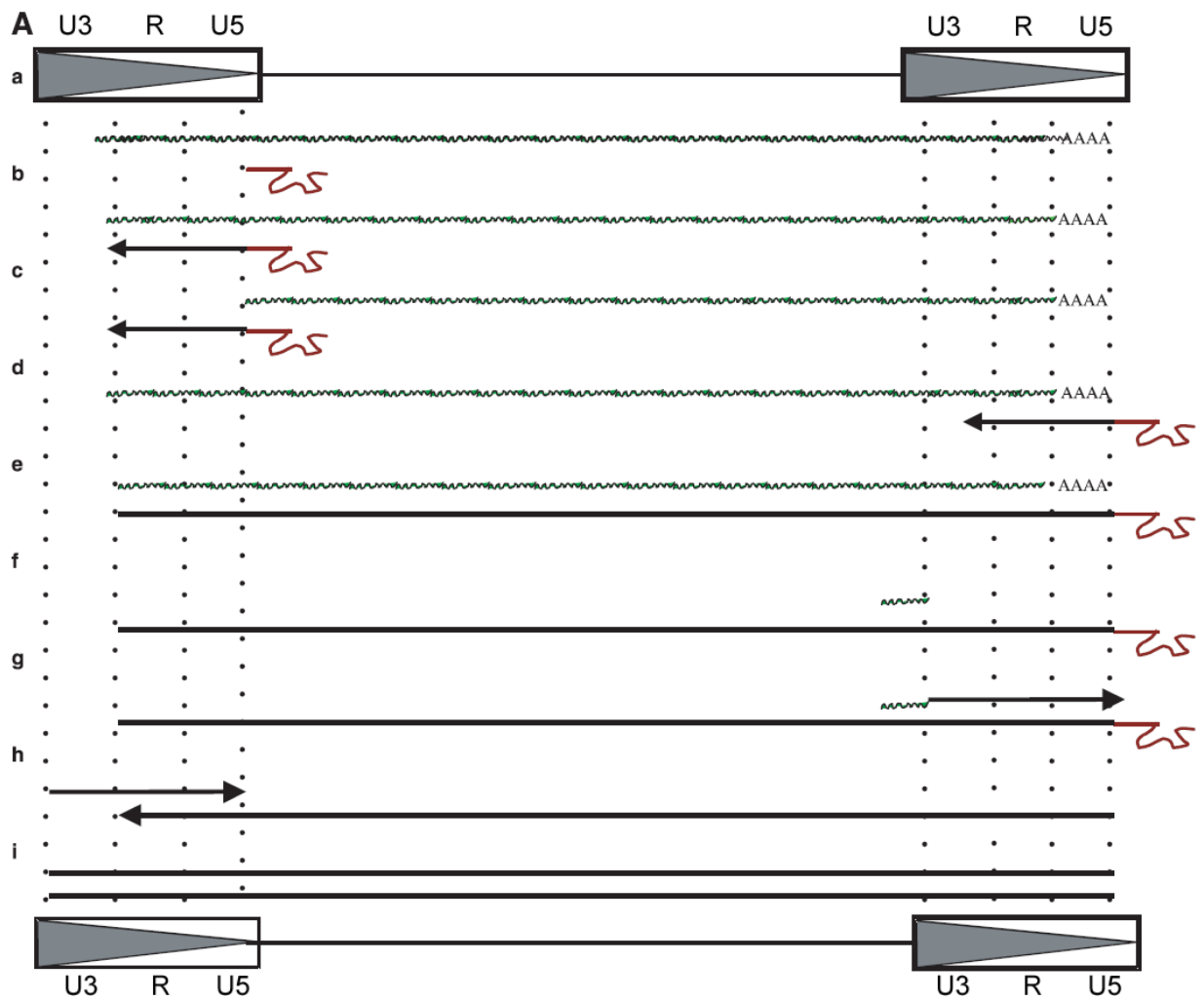


Figura 3 – Mecanismo de transcrição reversa de retrotransposon com LTR.

Caixas: LTRs; linhas verticais pontilhada: regiões das mesmas U3-R-U5; linha seguida de A: mRNA; linha irregular: tRNA; seta: direção da síntese. A transcrição reversa começa com a síntese de uma molécula de cDNA na região 5' do RNA usando um iniciador de tRNA (a, b), seguida da degradação da molécula de RNA nessa região (c), salto da molécula de cDNA para a região 3' do RNA (d) e síntese da fita completa de cDNA (e). Em seguida, o RNA tem sua fita parcialmente degradada pela RNaseH, codificada pelo próprio elemento (f). A pequena porção restante de RNA serve de iniciador para a síntese da porção final da segunda fita de DNA, tendo a primeira fita como molde (g). Ocorre outro salto para a porção inicial do DNA (h), seguido da síntese completa da segunda fita de DNA (i). Durante o processo, as terminações longas repetidas são formadas. Retângulos coloridos indicam LTRs. Linhas verticais pontilhadas indicam as regiões das mesmas (U3-R-U5); linha seguida de A: mRNA; linha irregular: tRNA; seta preta: direção da síntese. Figura adaptada de (Kazazian, 2004).

Os membros da ordem DIRS, também chamados de retrotransposons YR, diferem tanto de retrotransposons LTR quanto de retrovírus, pois apresentam repetições LTR de orientação invertida, não codificam um domínio IN (do tipo DDE) ou AP sendo caracterizados pela presença de um gene que codifica uma tirosina recombinase (YR). Um segmento das sequências LTR é repetido no interior do elemento dando origem ao que se denominam repetições internas complementares (ICR). O domínio ICR teria uma importância crítica na transcrição reversa para um intermediário de DNA, propondo-se que um intermediário circular se recombine com o sítio cromossômico, gerando a inserção por recombinação. Não formam duplicações no sítio de inserção. As tirosinas recombinases observadas nessa ordem são similares às recombinases dos bacteriófagos λ . Esse aspecto é especialmente interessante nessa ordem e levou autores que a estudaram a indagar, ainda sem resposta, como na sequência de um retrotransposon com gene para RT se associou um gene que codifica uma proteína relacionada às recombinases sítio-específicas dos fagos λ ? (Goodwin e Poulter, 2001; Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b)

Os elementos da ordem PLE (elementos Penelope-like) possuem sequências semelhantes a LTR que podem ser encontradas em orientação direta ou invertida e são caracterizados pela presença de um tipo de transcriptase reversa mais relacionada com as telomerasas do que com as transcriptases reversas de retrotransposons de outras ordens, como LTR ou LINE, ou de retrovírus. É interessante destacar que esses elementos são encontrados em inserções próximas aos telômeros nos cromossomos hospedeiros numa orientação consistente com a possibilidade de encontrar uma extremidade cromossômica livre que possa ser utilizada para iniciar a transcrição reversa. A autêntica atividade de RT desse novo tipo de enzima foi comprovada experimentalmente. Os PLE apresentam outra peculiaridade que são sequências codificadoras semelhantes a endonucleases Uri (GIY YIG) dos introns bacterianos móveis do grupo I e com as UvrC que são endonucleases bacterianas envolvidas em reparo do DNA (excisão de nucleotídeos) (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b).

Membros da ordem LINE (*long interspersed nuclear elements*) são elementos constituídos por milhares de pares de bases, encontrados em todos os eucariotos e, provavelmente, perderam as LTRs apresentando na extremidade 3' uma cauda poli-A (Figura 1). Apresentam estrutura e mecanismo de inserção diferentes dos elementos da ordem LTR e de retrovirus tendo sido por muito tempo denominados de retrotransposons não-LTR ou de retroelementos de famílias não-virais. Os elementos da ordem LINE foram classificados e subdivididos em 5 superfamílias: R2, L1, I, RTE e Jockey. De um modo geral parece haver variações estruturais nessa ordem: - alguns representantes (como as superfamílias RTE e R2) apresentam uma ORF única que codifica RT; na superfamília R2 há na extremidade C-terminal da RT um domínio EN o qual tem similaridade com enzimas de restrição do tipo II. É interessante destacar que esses elementos se inserem de modo sítio-específico em sequências de genes altamente conservados dos genomas hospedeiros como os genes rRNA nas diferentes espécies onde foram estudados. Já nas superfamílias L1, I e Jockey se observam duas ORF e há um domínio de endonuclease denominada APE (dada a semelhança com as endonucleases apurínicas-apirimidínicas envolvidas no reparo de DNA) na região N-terminal da RT. Apenas os membros da superfamília I apresentam ORF para a RNase H. Demonstrou-se experimentalmente que esta região está diretamente envolvida no reconhecimento e clivagem do sítio alvo no genoma hospedeiro (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b).

Na ordem LINE, o mecanismo de retrotransposição foi bem determinado experimentalmente para elementos das superfamílias R2 e L1 e se denomina transcrição reversa iniciada pelo sítio de inserção (*target-site primed reverse-transcription, TPRT*). O RNA mensageiro completo correspondente ao elemento é transcrito a partir de um promotor interno para RNA polimerase II, localizado na região 5'-UTR. O mRNA é exportado para o citoplasma e traduzido. Após tradução as proteínas exibem afinidade *cis* (ou preferência *cis*) e se ligam assim, preferencialmente, ao seu próprio mRNA codificador, formando uma partícula ribonucleoprotéica (RNP) a qual é transportada para o núcleo. O processo de transcrição reversa se inicia, então, a partir da clivagem da primeira fita de DNA pela endonuclease e utiliza o 3'-OH liberado pela clivagem para iniciar a reação

de transcrição reversa. Ocorre também a clivagem da segunda-fita que por sua vez permite a utilização da extremidade liberada pra iniciar a síntese da segunda fita do elemento (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b).

A ordem SINE (*short interspersed nuclear elements*) caracteriza-se por elementos que não codificam proteínas, elementos não-autônomos, e apesar disso, não são simples derivados degenerados de elementos de classe I, sendo gerados por eventos acidentais de retrotransposição de vários transcritos de polimerase III. Eles apresentam promotores internos para Pol III, o que permite que sejam expressos, e cujas sequências definem as superfamílias da ordem SINE e sua origem: - tRNA, 7SL RNA e 5SL RNA. São elementos pequenos (80-500 pb) e geram duplicações do sítio-alvo de 5-15 pb. A maioria dos SINE consiste em dois ou mais módulos: “cabeça” 5'-terminal (região onde se encontram os promotores), “corpo” (região de origem pouco definida provavelmente derivada de LINE) e “cauda” 3'-terminal. Para a função de transposição eles dependem de RT de LINE autônomos sendo que alguns SINE exigem parceiros definidos enquanto outros são mais generalistas. Os transcritos derivados dos SINE se ligam a RT de um LINE seguindo-se a sua transcrição reversa e integração de uma cópia no genoma de modo similar ao descrito anteriormente para os LINE. Os elementos SINE estão ausentes da maioria dos genomas de eucariotos unicelulares. É interessante destacar que a sequência *Alu*, amplamente presente e estudada no genoma humano é um exemplo representativo de retrotransposon da ordem SINE (Wicker, Sabot *et al.*, 2007; Kramerov e Vassetzky, 2011).

2.3.4 Potencial dos elementos transponíveis em causar variações genéticas

Os TE possuem potencial para causar variações nos genomas hospedeiros. Ao mesmo tempo em que sua presença e atividades podem ser lesivas e causar doenças eles representam uma importante fonte de variação genética. A transposição é, em geral, altamente regulada e capaz de influenciar a expressão de genes quando ocorre inserção dentro desses genes ou nas suas proximidades, contribuindo para ativação ou inativação, total ou parcial de um ou mais genes (Hayes, 2003). Há alguns exemplos interessantes, estudados em

Saccharomyces cerevisiae, de como as integrases selecionam seus sítios de inserção. O retrotransposon Ty5 de *S. cerevisiae* tem uma região com 9 aminoácidos no C-terminal (seletora de domínios alvo) da IN que precisa ser fosforilada para interagir diretamente com uma proteína, Sir4p (*silent information regulator 4*), interação essa crítica para dirigir a integração do elemento para regiões de heterocromatina, ou seja, aquelas regiões onde sua inserção seria menos deletéria. Contudo, mutantes para IN não fosforilados se tornam mutagênicos, pois passam a se integrar em regiões onde genes são ativamente expressos (Cowley e Oakey, 2013).

Há atualmente um crescente conjunto de evidências sugerindo que TE podem ser funcionalmente importantes e não apenas “lixo genômico” ou “DNA parasita” ou “DNA egoísta”, algo bem além de elementos que apenas fazem cópias de si mesmos. Quão bons podem ser os TE? Como os TE influenciam a expressão gênica? TE contribuem no remodelamento dos genomas, constituindo uma das principais fontes de mutações, devido à sua capacidade de transposição para um novo sítio, por possibilitarem a recombinação homóloga e rearranjos cromossômicos (Zou, Kim *et al.*, 1996). Há considerável evidência de que significativo número de cópias domesticadas de TE agem como elementos regulatórios transcricionais ou que contribuem para regiões codificadoras de genes celulares. Temos um exemplo em *Leishmania* spp., um protista unicelular responsável por doenças humanas, que reciclou e expandiu uma família completa de pequenos TEs extintos (retroposons) preenchendo um importante caminho biológico, a regulação da expressão gênica. Outra contribuição importante é a geração de diversidade de transcritos. TE podem fornecer novos promotores, novos sítios de “*splicing*” (splicing alternativo) e novos sinais de poliadenilação. Há trabalhos mostrando como o mobiloma (coletivo dos TE) pode influenciar o transcriptoma (conjunto dos transcritos), tanto na sua diversidade como na sua regulação. Um transcrito alternativamente processado pode evoluir para uma nova função; TE em várias cópias podem dispersar elementos regulatórios e tecer novas redes de interações (Cowley e Oakey, 2013).

Do ponto de vista de aplicabilidade prática, os conhecimentos relativos à estrutura e organização dos TE de determinada população, considerando os padrões polimórficos gerados a partir de suas inserções, podem ser ferramentas

úteis para a tipagem de isolados clínicos e ambientais de microrganismos, entre os quais fungos patogênicos. Keller e colaboradores utilizaram uma sonda derivada de um retrotransposon (TCN1) para discriminar linhagens de *Cryptococcus neoformans*. Os resultados indicam que TCN1 pode ser utilizado devido à sua sensibilidade e especificidade, para discriminar isolados dos sorotipos A e D, podendo ser uma ferramenta útil em estudos epidemiológicos (Keller, Hettler *et al.*, 2006). Um método de tipagem denominado RISC (caracterização do contexto de sítios de inserção de retrotransposons) foi proposto para discriminar isolados de *A. fumigatus*. Este método é baseado na amplificação de sequências que flanqueiam elementos retrotransponíveis (100 a 500 pb), digestão com endonucleases de restrição e observação de padrão de “*fingerprinting*” resultante de variações do número de cópias e diferenças nos sítios de inserção (De Ruiter, De Valk *et al.*, 2007).

2-4 Elementos transponíveis em fungos.

A presença de elementos genéticos móveis em fungos foi descrita na década de 70, em leveduras *Saccharomyces cerevisiae*, e apenas 10 anos depois sua ocorrência foi descrita em fungos filamentosos (Daboussi e Capy, 2003; Muszewska, Hoffman-Sommer *et al.*, 2011). Destacando também o seu papel como agentes centrais na evolução e reestruturação dos genomas por meio de mecanismos de transposição, recombinação ectópica, transmissão horizontal, amplificação e inativação epigenética. Daboussi e Capy, em 2003, dão conta da existência, naquela época, de aproximadamente 60 elementos descritos em fungos, de diferentes tipos e pertencentes aos maiores grupos previamente descritos em outros eucariotos. Destes, aproximadamente 30 elementos correspondiam a retrotransposons LTR (*Ty1/copia* e *Ty3/gypsy*) e a retrotransposons não-LTR (LINE e SINE) (Daboussi e Capy, 2003). A grande maioria correspondia ao grupo *Ty3/gypsy*, que está distribuída em três filos de fungos: Ascomycota, Basidiomycota e Zygomycota. Este número hoje pode ser considerado diminuto hoje, uma vez que muitos outros elementos transponíveis foram identificados nos últimos anos, em diferentes espécies de fungos (Goldman, Dos Reis Marques *et al.*, 2003; Fujino, Sekiguchi *et al.*, 2005; Monroy e Sheppard, 2005; Bouvet, Jacobi *et al.*, 2007; Araújo, 2008; Nunes, Gowda *et al.*, 2011; Labbe, Murat *et al.*, 2012).

De modo geral essas descrições se referem à ocorrência, abundância e diversidade de retrotransposons no genoma de uma determinada espécie de fungo, em particular, não havendo descrições ou análises recentes relativas a essas características dos retrotransposons no reino Fungos. Por outro lado, numa revisão recente, relativa apenas aos retrotransposons LTR em fungos, Muszewska e colaboradores relataram que, realizando buscas em cerca de 60 genomas de fungos disponíveis hoje em bancos de dados públicos, foram capazes de relacionar 66.000 elementos (somando-se completos e cópias remanescentes identificáveis) pertencentes às superfamílias Gypsy/Ty3 e Copia/Ty. Quanto à abundância da ordem de retrotransposons LTR nos genomas fúngicos, Muszewska e colaboradores relatam serem estes escassos em alguns genomas (~50 elementos) ao passo que em outros genomas se observa grande expansão dessa ordem (~8000 elementos). Os dados mostram que as expansões de elementos móveis em fungos se referem tanto ao aumento dos números de cópias de elementos individuais quanto à diversidade ou número de tipos de retroelementos encontrados nesses genomas (Muszewska, Hoffman-Sommer *et al.*, 2011). Os genomas possuem mecanismos contra as expansões desses elementos e os fungos tornaram-se especialistas nesse campo: - 3 dos principais mecanismos de silenciamento de transposons (*repeat induced point mutation* - RPI; *methylation and quelling*; *sex-induced silencing* - SIS) foram descritos em fungos (Muszewska, Hoffman-Sommer *et al.*, 2011).

Pouco ainda se sabe sobre as sequências repetitivas e os elementos móveis em *Paracoccidioides brasiliensis*. A primeira menção à presença de elementos genéticos móveis no genoma de *P. brasiliensis* pode ser encontrada na publicação de Felipe e colaboradores (2003): na descrição do transcriptoma parcial do fungo, retroelementos foram listados entre as 18 sequências de cDNAs majoritariamente expressas, mencionando-se, sem caracterizá-los, a ocorrência de oito diferentes grupos de retroelementos *Ty-like* (Felipe, Andrade *et al.*, 2003). Sucessivamente, a ocorrência de transcritos com escores de similaridade significativos com retrotransposons do gênero *Chromovirus* (família Metaviridae) foi relatada por Gorinsek (Gorinsek, Gubensek *et al.*, 2004) entre as ESTs de fungos (Ascomycetos) depositadas em bancos de dados públicos, citando-se *Magnaporthe grisea*, *Colletotrichum trifolli*, *Schizosaccharomyces*

pombe, *S. cerevisiae*, *Blumeria graminis* e também *P. brasiliensis*, num estudo de genômica comparativa desse gênero de retroelementos em eucariotos. Ainda em 2004, Nascimento e colaboradores com base na caracterização de microsátélites em uma coleção de sequências genômicas randômicas (GSS) (1117 padrões distintos em 3,8 Mb, a maioria em regiões não codificadoras), estimaram que 0,3% deles estariam localizados em transposons. De fato, em seguida, Matute e colaboradores (2006) empregando iniciadores baseados naqueles padrões de microsátélites, conseguiram discriminar entre duas (S1 e PS2) das três prováveis espécies filogenéticas previamente propostas.

Reinoso e colaboradores (2005) descreveram um fragmento gênico em *P. brasiliensis* com similaridade a transcriptases reversas. Foi construída uma biblioteca genômica de *P. brasiliensis* em vetor pBlueScript KS+ cujo tamanho médio dos insertos era 8,0 Kb. Selecionaram-se sete clones ao acaso para sequenciamento. O clone 2 (4439 pb) (*accession number* AF443/89) apresentou duas ORFs, sendo que uma delas (474 pb) (*PbrRT*, AF443189) apresentou identidade com transcriptases reversas de *M. grisea* (24,8%), indicando a presença de retrotransposons (Reinoso, Niño-Vega *et al.*, 2005).

Nos últimos cinco anos, o Laboratório de Biologia de Microrganismos do Departamento de Microbiologia, ICB / UFMG, vem desenvolvendo uma linha de pesquisa cujo objetivo geral é a identificação e caracterização de elementos genéticos móveis, transposons e retrotransposons, no genoma de *P. brasiliensis*, tendo sido concluídas duas teses de doutorado e uma dissertação de mestrado, neste período. Na primeira delas, Araújo e colaboradores (2008) realizaram num primeiro momento buscas por estes elementos entre as sequências de cDNA (ESTs) anotadas de *P. brasiliensis*, presentes no Banco de Dados do Projeto Genoma Diferencial e Funcional do Centro-oeste (*PbAESTs*) (Felipe, Torres *et al.*, 2005). Nesta busca foram encontradas 369 ESTs com similaridades a elementos genéticos móveis (TEs) e virais distribuídas em 30 *singlets* e 19 contigs. A partir desse achado inicial, caracterizaram-se dois retrotransposons majoritariamente transcritos do fungo, representados pelos contigs 400 e 1609, que apresentavam escores de similaridade elevados com retrotransposons bem estudados em outros fungos, como MGLR3 e MAGGY, elementos ativos no genoma do fitopatógeno *Magnaporthe grisea* (Araújo, 2008). Os putativos

retrotransposons de *P. brasiliensis*, denominados RTPb1 e RTPb2, foram preliminarmente caracterizados por meio do sequenciamento de seis clones de cDNA oriundos das bibliotecas que originaram as informações do transcriptoma no banco de dados *PbAESTs*. A análise das fases de leitura aberta (ORF) dos elementos RtPb1 e RtPb2 revelaram, respectivamente, três ORF similares a transcriptase reversa (RT), integrase (Int) e cromodomínio (Chromo), típicos de retrotransposons do tipo LTR (*Long Terminal Repeats*, família viral) e uma ORF similar a transcriptase reversa (domínio RNase H), típica de retrotransposons do tipo não-LTR, não virais (Araújo, 2008).

Na segunda tese de doutorado executada nessa linha de pesquisa, em nosso laboratório, Marini e colaboradores (2010) partiram do levantamento realizado por Araújo e colaboradores (2008) no banco de dados do Projeto Genoma Funcional e Diferencial de *P. brasiliensis* (elaborado e anotado por pesquisadores da Rede Genoma Centro-Oeste/ CNPq - Ministério da Ciência e Tecnologia, UnB e UFG) (Felipe et al., 2003, Felipe et al., 2005). Nesta busca, foram então encontrados dois contigs apresentando similaridade com transposons de DNA. Estas sequências (contig 530, 2 “reads”; e contig 1938, 2 “reads”) apresentaram similaridade com transposases de *Aspergillus fumigatus*. Foram desenhados oligonucleotídeos iniciadores para amplificar os contigs 530 e 1938 que foram clonados e tiveram sua identidade confirmada por sequenciamento (Marini, Zanforlin et al., 2010)

As sequências dos contigs 530 e 1938 foram então utilizadas como objeto de busca em análises de BLASTN na primeira montagem do genoma da linhagem Pb03 de *P. brasiliensis*, o primeiro a ser liberado pelo “*Fungal Genome Initiative*” (FGI) do *Broad Institute of MIT and Harvard* (www.broadinstitute.org/annotation/genome/paracoccidioides_brasiliensis/MultiHome.html). Como um grande número de sequências foi obtido buscou-se uma estratégia mais eficiente para sua identificação e caracterização. As sequências dos contigs 530 e 1938 foram utilizadas como objeto de busca por BLASTN modificado contra o banco de dados compilado localmente. Os resultados obtidos com o BLASTN modificado foram combinados com uma estratégia de BLAST reverso que permitiu localizar as terminações repetidas invertidas (TIR)

dos transposons de DNA, adotando-se o programa Artemis para a sua anotação (Marini, Zanforlin *et al.*, 2010).

Um total, 1332 *hits* ou sequencias alvos não ambíguas para transposons de DNA foram identificados, sendo 384 no genoma da linhagem Pb01 (*P. lutzii*), 475 em Pb03 e 473 em Pb18 (ambas *P. brasiliensis*), com características compatíveis com transposons da superfamília *Tc1/mariner*. Uma parte significativa destes *hits* (868, 65,2%) correspondeu a elementos altamente defectivos, com baixa conservação dos domínios típicos de transposases (HTH, CENPB e DDE) e, portanto, não foram incluídos nas análises então realizadas (Marini, Zanforlin *et al.*, 2010).

Marini e colaboradores conseguiram caracterizar oito novas famílias de transposons de DNA da superfamília *Tc1/Mariner* no genoma das três linhagens sequenciadas. Na linhagem Pb01 foram caracterizadas apenas três famílias de transposons de DNA, correspondendo a 102 (26,6%) de 384 *hits* encontrados. Na linhagem Pb03 foram identificados dois elementos aparentemente completos, potencialmente funcionais e autônomos que codificam para putativas transposases e apresentam TIR bem definidas, além de outros quatro elementos truncados. Esses elementos totalizaram 177 *hits*, o que equivale a 37,3% dos 475 *hits* similares a transposons de DNA naquela linhagem. Dos 473 *hits* observados na linhagem Pb18, 185 (39,1%) foram identificados e relacionados como elementos *Tc1/mariner*.

Os oito elementos identificados compartilham 65 a 68% de similaridade em nucleotídeos com diferentes transposons de DNA *mariner* depositados no banco de dados Repbase (Jurka et al. 2005) do *Genetic Information Research Institute* (GIRI). Seis destes elementos são flanqueados pelo dinucleotídeos TA, assim como os transposons *Tc1/mariner*. Devido a sua similaridade com essa superfamília e a inserção específica da maior parte deles (61,6%) no sítio alvo TA, eles foram nomeados Trem (abreviação de Transposable element mariner) e identificados por letras de A – H. A integridade dos transposons foi confirmada pela presença de TIR em ambas as extremidades e pelo alinhamento múltiplo com sequências de elementos bem estudados. Em sete dos oito elementos identificados foi possível encontrar as repetições terminais invertidas,

características dos transposons de DNA da subclasse 1, que transpõem pelo mecanismo de “cortar e colar”. Os elementos TremA, TremB e TremD apresentaram cópias completas, com ORFs intactas (Marini, Zanforlin *et al.*, 2010).

Os elementos Trem estão distribuídos de forma desigual no genoma das três linhagens sequenciadas (Pb01, Pb18 e Pb03) e utilizados no estudo *in silico*. Os elementos TremC e H foram encontrados em Pb01, Pb18 e Pb03; os elementos TremA, B, F e G figuraram em Pb03 e Pb18 (*P. brasiliensis*); o elemento TremD foi observado apenas em Pb18; e o elemento TremE foi exclusivo de Pb01 (*P. lutzii*). Ampliou-se experimentalmente a pesquisa da presença de elementos Trem em 17 isolados clínicos e um ambiental do fungo, constatando-se a tendência de manutenção destes padrões no DNA genômico em isolados filogeneticamente relacionados: o elemento TremE foi encontrado apenas nos isolados representativos de *P. lutzii*; os elementos TremA, TremB e TremF nos isolados de *P. brasiliensis* S1 e PS2, e o elemento TremD apenas em *P. brasiliensis* S1 (Marini, Zanforlin *et al.*, 2010).

O mapeamento físico dos elementos TremA e TremB pelas técnicas de *Chromoblot* e *Southern Blot* apontou a presença de múltiplas cópias com distribuição polimórfica nos isolados denominados Pb03 e Pb18 e corroborou os dados obtidos *in silico* indicando a ausência destes elementos no genoma de Pb01 (Marini, Zanforlin *et al.*, 2010). Assim, os elementos Trem pareceram ser bons candidatos para o reconhecimento de linhagens e espécies do gênero *Paracoccidioides*, o que foi realizado recentemente (Alves, 2012). Alves e colaboradores (2012) relataram o emprego de reações de amplificação em cadeia da polimerase (PCR) para avaliar a utilidade de três marcadores moleculares na discriminação de espécies filogenéticas de *Paracoccidioides* spp: 1) par de iniciadores destinados à região de *indel* do gene *hsp70*; 2) marcadores microssatélites de DNA); 3) Trem A-H. Foram estudados 48 isolados de *Paracoccidioides* spp, sendo trinta previamente classificados segundo a espécie filogenética (10 de *P. lutzii*, 15 de *P. brasiliensis* S1, 3 de PS2 e 2 de PS3). Puderam ser classificados como *P. lutzii* 14 isolados, por meio do uso dos iniciadores dirigidos ao *indel* do gene *hsp70*, sendo os demais 34 isolados classificados como *P. brasiliensis*. Os marcadores microssatélites utilizados não

conseguiram discriminar os isolados entre as três espécies filogenéticas de *P. brasiliensis*. Os resultados obtidos pelo uso da PCR para detecção de elementos Trem A-H mostram que Trem A, B, F e G são encontrados apenas em *P. brasiliensis* (S1, PS2 e PS3); Trem E está presente apenas em *P. lutzii*; Trem C e H estão presentes em *P. lutzii* e *P. brasiliensis* (S1, PS2 e PS3). É interessante ressaltar que em *P. brasiliensis*, a amplificação de Trem C e Trem H mostrou um padrão de duas bandas enquanto que em *P. lutzii*, observou-se apenas uma banda de aproximadamente 2 kb. Em isolados PS3, Trem D mostra um padrão de três bandas. Estes resultados confirmaram que os padrões de amplificação observados para os oito elementos Trem A-H e seu padrão esperado de bandas, podem ser úteis como marcadores moleculares espécie-específicos no gênero *Paracoccidioides*.

Dois outros trabalhos recentes apresentaram informações referentes aos elementos transponíveis no genoma de fungos do complexo *Paracoccidioides* (Desjardins, Champion *et al.*, 2011; Shankar, Wu *et al.*, 2011). No primeiro trabalho, Desjardins e colaboradores usaram abordagem de genômica comparativa com o objetivo de entender as bases genéticas relacionadas ao crescimento e patogenicidade em fungos do complexo *Paracoccidioides*. Para isso, os autores utilizaram os genomas de dois isolados de *Paracoccidioides brasiliensis* (Pb03 e Pb18) e um isolado de *Paracoccidioides lutzii* (Pb01). Pelo estudo comparativo os autores afirmaram que aproximadamente 8-9% do genoma de *P. brasiliensis* seja constituído por elementos pertencentes às classes I e II (retrotransposons e DNA transposon) e que essa percentagem pode chegar a 16% em *P. lutzii*. Segundo este estudo, observa-se que o número de elementos transponíveis pode variar aproximadamente entre 700-1300, com variações intra e inter-específicas.

No segundo trabalho, Shankar e colaboradores estudaram o perfil de expressão de um isolado da espécie *P. lutzii* (Pb01) com o objetivo de delinear os efeitos de 17- β -estradiol (E2) enquanto na transição da forma micelial para leveduriforme. Inicialmente, os pesquisadores examinaram a transição morfológica para confirmar que E2 era capaz de inibir a transição e então determinaram, a partir da massa fúngica proveniente da mesma cultura, o perfil transcricional durante a transição. Os resultados, obtidos por análises de PCR

em tempo real, mostraram que E2 afeta a expressão de genes relacionados à formação de membrana e parede celulares e genes relacionados ao metabolismo energético. Além disso, os autores relataram o efeito regulatório sobre um elemento retrotransponível. Os autores mostraram que se trata de um elemento retrotransponível, do grupo Ty/*gypsy* de 5360pb, cujas LTR possuem 198 pb.

Diante do grande número de elementos genéticos móveis presentes no genoma de *P. lutzi* e *P. brasiliensis*, do número limitado de retrotransposons identificados e caracterizados, da crescente importância que vem sendo atribuída na literatura devido à presença e atividade desses elementos como agentes de variabilidade genética e na diversidade do transcriptoma, este trabalho se propôs realizar uma busca detalhada visando a localização e caracterização de elementos potencialmente ativos no genoma de fungos do complexo *Paracoccidioides*.

3- OBJETIVOS

3.1 Objetivo Geral

Identificar e caracterizar, elementos genéticos móveis, retrotransposons, no genoma de fungos dimórficos do complexo *Paracoccidioides*; realizar anotação estrutural e funcional e avaliar a expressão, buscando contribuir para o conhecimento da estrutura, organização e funcionalidade do genoma e da biologia do fungo.

3.2 Objetivos Específicos

1. Obtenção das sequências de ESTs (etiquetas de sequências expressas) de *Paracoccidioides* disponíveis no banco de dados dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>), “clusterização” de dados, anotação funcional;
2. Utilização dos alinhamentos significativos com elementos móveis, especialmente retrotransposons, para o mapeamento, aprimoramento da anotação genômica disponível e identificação de novos elementos.
3. Utilização de diferentes matrizes de domínios funcionais (transcriptase reversa, RNase H, integrase e cromodomínios – domínios modificadores de cromatina, poliproteínas gag/env/pol) disponíveis nos bancos de dados PFAM – *Protein Families* (<http://pfam.sanger.ac.uk/>) e CDD – *Conserved Domain Database* (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>) para mapeamento genômico dos elementos móveis;
4. Elaboração de *scripts* na linguagem de programação PERL para extração de relatório das análises e, posterior, integração de resultados na plataforma de anotação genômica Artemis (<http://www.sanger.ac.uk/Software/Artemis/>) visando viabilizar a identificação sistemática de alvos para estudo experimental e funcional.
5. Avaliar a relação filogenética entre os elementos identificados e o contexto genômico nos quais eles estão inseridos.

6. Amplificar, clonar e caracterizar as sequências correspondentes aos retrotransposons selecionados, potencialmente ativos no genoma de espécies do complexo *Paracoccidioides*.
7. Comprovar a presença e determinar a organização dos retrotransposons classificados e selecionados, em diferentes isolados do complexo *Paracoccidioides*.
8. Determinar experimentalmente a distribuição e organização genômica dos elementos potencialmente ativos por *Southern blot*.
9. Avaliar o padrão de transcrição dos retrotransposons classificados e selecionados por RT-PCR.
10. Avaliar quantitativamente por qRT-PCR o padrão de expressão dos retrotransposons classificados e selecionados.

4 - MATERIAL E MÉTODOS

4.1 Análise de bioinformática

Para as abordagens de bioinformática, contou-se com a orientação do Professor Dr. Jeronimo Ruiz do Centro de Pesquisas René Rachou – FIOCRUZ. Desenvolveram-se *scripts* na linguagem de programação PERL (*Practical Extraction and Report Language*) para formatar e permitir as análises do conjunto de dados. Além disso, utilizaram-se ferramentas computacionais desenvolvidas para o sistema operacional *Windows* (programas não publicados) cedidas pelo Dr. José M. C. Ribeiro do *Laboratory of Malaria and Vector Research - NIH (National Institutes of Health)* colaborador do grupo de bioinformática. Esses programas, quando citados no decorrer do trabalho serão referidos junto ao nome JM-Ribeiro.

4.1.1 Workflow computacional

A mineração (*data mining*) e o processamento dos dados de ESTs de *P. brasiliensis* para identificação de elementos retrotransponíveis envolveram diferentes abordagens como: agrupamento por identidade das sequências e buscas por similaridade aplicadas em várias etapas da análise. A figura 4 representa o fluxograma consolidado desses passos.

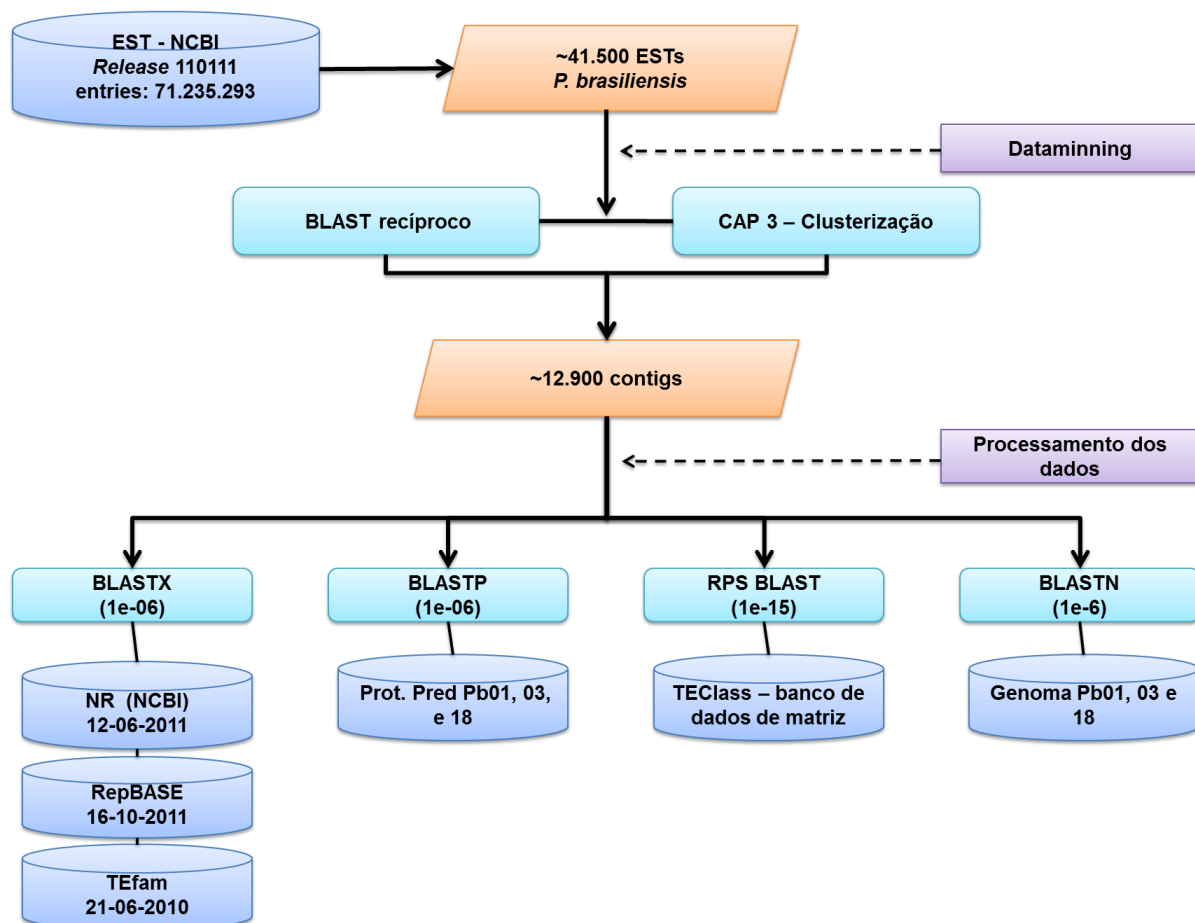


Figura 4 – Representação esquemática das análises computacionais para identificação de potenciais elementos retrotransponíveis.

Os cilindros representam os bancos de dados utilizados: **EST** – banco de dados de EST depositado no NCBI; **NR** – banco de dados de proteínas não redundantes (NCBI); **RepBase** – banco de sequências repetitivas do GIRI (Genetic Information Research Institute); **TEfam** – banco de elementos transponíveis da Universidade da Virgínia, EUA; **Prot. Pred** – proteoma predito dos isolados Pb01, 03 e 18; **TEClass** – banco de matrizes de domínios conservados representativos de elementos transponíveis; **Genoma** – banco de sequências genômicas dos isolados Pb01, 03 e 18. Os paralelogramos representam os bancos de EST e de grupos analisados. Os retângulos representam os diferentes tipos de análises realizadas pelo algoritmo Blast (vide tabela 3).

4.1.2 Sequências de EST (*Expressed Sequence Tags*)

Para a identificação e avaliação de retrotransposons presentes nos genomas de fungos do complexo *Paracoccidioides*, utilizou-se, para a obtenção das sequências, a base de dados dbEST (*Expressed Sequence Tags data base, release 110111*) do NCBI (*National Center for Biotechnology Information*) <ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA>. Para a recuperação das sequências de ESTs utilizou-se o gênero do microrganismo como palavra chave, “*Paracoccidioides*”, objeto de busca na base de dados, recuperando-se aproximadamente 41 mil sequências associadas ao organismo.

4.1.3 – Mineração de dados (data mining)

Na abordagem computacional utilizou-se, para mineração de dados, a estratégia do *best hit* recíproco (*Reciprocal Best Hit - RBH*) seguido do agrupamento dos dados (Huang e Madan, 1999). Esta estratégia é largamente utilizada para a detecção de genes ortólogos. Nesse trabalho, extrapolou-se a técnica para eliminar a redundância de informações dos dados.

Utilizou-se o algoritmo BLASTN (*nucleotide Blast*) para fazer o BLAST recíproco entre as sequências de *ESTs* e gerar o RBH. Nesta comparação, desligou-se o filtro de baixa complexidade para evitar a quebra de longos HSPs (*High Segment Score Pairs* – regiões de alto escore de alinhamento) em segmentos menores devido à baixa complexidade e repetições das sequências (Altschul, Gish *et al.*, 1990).

As sequências oriundas do RBH foram utilizadas como entrada para o programa CAP3 (Huang e Madan, 1999) que utiliza o método de alinhamento múltiplo de sequências para a geração dos *contigs*. No contexto de *clusterização*, um *contig* pode ser definido como sequência consenso obtida a partir do alinhamento de múltiplas sequências de EST. No entanto, no processo de *clusterização*, sequências que não apresentam similaridade com outras ficam isoladas formando os *singlets*. Neste trabalho, tanto os *singlets* quanto os *contigs* foram tratados como grupos. O programa foi parametrizado para agrupar sequências com sobreposição de, no mínimo, 100 nucleotídeos e identidade igual ou superior a 80%.

4.1.4 Processamento dos dados

O próximo passo foi identificar regiões de similaridades correspondentes a retrotransposons nos *contigs* gerados pelo processo conhecido como *data mining*. Para tal, utilizou-se o método de busca por similaridade de sequências por meio da ferramenta *BLAST* (Altschul, Gish *et al.*, 1990). Seis bancos de dados de domínio público foram compilados localmente além de um banco de dados de matriz para efetuar as análises (Tabela 1). As sequências utilizadas correspondem a regiões codificadoras de elementos transponíveis e retrotransponíveis pertencentes a diversas superfamílias (Tabela 2). Estas sequências foram obtidas a partir do banco de dados NR (*Non-redundant*) do NCBI.

Tabela 1 - Bancos de dados utilizados para as análises

Nome do banco de dados	Endereço eletrônico	Data de atualização
NR	ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA	12/06/2011
RepBASE	http://www.girinst.org/replib/index.html	16/10/2011
TEfam	http://tefam.biochem.vt.edu/tefam/search.php	21/06/2010
TEClass*	Banco de dados de matrizes de alinhamento	10/2011
Proteoma predito (Pb01, 03 e 18)	http://www.broadinstitute.org/annotation/genome/paracoccidioides_brasiliensis/MultiHome.html	12/2010

* **NR** – banco de dados de proteínas não redundantes (NCBI); **TEfam** – banco de elementos transponíveis da Universidade da Virgínia, EUA; **TEClass**, banco de matrizes de domínios conservados representativos de elementos transponíveis; **RepBase** – banco de sequências repetitivas do GIRI (Genetic Information Research Institute); O banco de dados TEClass foi construído localmente utilizando a ferramenta *PSIblast* (*Position-Specific Iterative Blast*) para construção de matrizes. As sequências utilizadas correspondem a regiões codificantes de diversas superfamílias que representam elementos transponíveis e retrotransponíveis (Tabela 2), obtidas no banco de dados NR.

Tabela 2 - Superfamílias de elementos transponíveis e retrotransponíveis representadas no banco de dados TEClass.

Classes	Total de famílias	Total de sequências
Classe I		
L1	1773	1980
Pao Bel	81	97
Penelope	290	328
Ty3_gypsy	1644	1867
Rnase_H	24	26
Classe II		
hAT	48	51
Mariner	74	91
Merlin	3	3
Mutator	19	56
P	54	58
PIF	40	43
piggyBac	80	90
Fragmentos não classificados	1338	2693
Total	5468	7383

4.1.5 Buscas por Similaridade de Sequências

Para buscas por similaridade local entre pares de sequências foi utilizado o algoritmo *BLAST* (Altschul, Gish *et al.*, 1990). A tabela 3 mostra os diferentes tipos de algoritmos e os bancos de dados que foram usados na execução do trabalho.

Tabela 3 - Algoritmos utilizados para buscas de similaridade local entre pares de seqüências.

Algoritmo	Descrição	Bases de dados usadas
<i>blastn</i>	Compara uma seqüência de nucleotídeos com um banco de dados de seqüências de nucleotídeos	Genoma dos isolados Pb01, Pb03 e Pb18.
<i>rpsblast</i>	Compara uma seqüência nucleotídeos com um banco de dados de matrizes	TEClass
<i>blastx</i>	Compara uma seqüência de nucleotídeos traduzidos com um banco de dados de seqüências de aminoácidos	NR, Proteoma e TEfam
<i>tblastx</i>	Compara uma seqüência de nucleotídeos traduzidos com um banco de dados de seqüências de nucleotídeos traduzidos	RepBase

Todas as buscas, realizadas por alinhamentos entre seqüências, com os diferentes bancos de dados, foram parametrizadas com o objetivo de aperfeiçoar as comparações entre os *contigs* e as seqüências contidas nesses bancos. A tabela 4 apresenta a relação destes parâmetros. Todas as análises de busca de similaridade foram realizadas na plataforma LINUX, empregando-se linhas de comando contendo todos os parâmetros descritos anteriormente.

Como parâmetro de corte utilizou-se um *e-value* de 10^{-6} . Ele é um indicativo estatístico que representa a chance de que um alinhamento em questão, considerando seu valor de *score* e o número de seqüências presentes na base de dados, tenha sido encontrado ao acaso. Para as nossas análises com o algoritmo *rpsblast* (Reverse PSI-BLAST), utilizou-se um *e-value* mais restrigente (10^{-15}) em função das seqüências utilizadas para gerar a matriz.

Tabela 4 - Parâmetros utilizados na execução das análises pelos algoritmos do *Blast*.

Parâmetro	Descrição	Valores aceitos
-p	Indica o algoritmo a ser usado na análise	blastn
		blastp
		blastx
		tblastn
		tblastx
-i	Identifica a sequência que será comparada como o banco de dados	Nome da sequência a ser confrontada
-d	Identifica o banco de dados	Nome do banco de dados
-o	Nome do arquivo de saída (output file)	Nome do arquivo resultante da análise
-F	Desliga o filtro de baixa complexidade	T para filtro ligado e F para filtro desligado
-e	Fornece o valor de <i>e-value</i>	Valores $\leq 10^{-6}$
-m	Gera resultado no formato tabular	8 ou 9

4.1.6 Construção da tabela *hyperlinkada*

Para organizar os resultados e relacionar cada um dos *contigs* aos diferentes resultados gerados nas análises, utilizou-se o programa *Excel* que permitiu uma melhor visualização dos dados. Esta planilha foi construída de forma automatizada utilizando o programa *TableMaker* (versão 2011 para *Windows*, programa não publicado). Além disso, ferramentas desenvolvidas para o sistema operacional *Windows*, tal como o *Excel*, são comumente mais amigáveis ao usuário comum.

4.2 - Anotações de Sequências

4.2.1 Anotação estrutural e funcional dos retrotransposons

A etapa de anotação estrutural e funcional foi realizada utilizando-se os bancos não redundantes (NR) de proteínas e nucleotídeos do NCBI disponíveis em <ftp://ftp.ncbi.nih.gov/blast/db/>, a fração SeqDBLite do banco de dados Gene Ontology <http://www.geneontology.org/>, os bancos de dados de domínios conservados PFAM (Protein Families - <http://pfam.sanger.ac.uk/>) e CDD (Conserved Domain Database <http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml>)

e o Interproscan <http://www.ebi.ac.uk/Tools/InterProScan/>. Cópias dos bancos de dados acima descritos foram mantidas em servidor local para buscas por similaridade local.

4.2.2 Identificação de domínios conservados

Para buscas sensíveis nos bancos de domínio contendo descrições estatísticas da sequência consenso de famílias de proteínas (Modelos Ocultos de Markov) utilizou-se o algoritmo HMMER versão 2.3.2. O programa possui código fonte livre para uso acadêmico e está disponível em <http://hmm.janelia.org/>.

4.2.3 Identificação de elementos retrotransponíveis

A identificação estrutural de elementos retrotransponíveis nos genomas foi realizada por meio do programa LTR_Finder (Xu e Wang, 2007). Este programa identifica características específicas dos retrotransposon, tais como terminações longas repetidas e duplicações do sítio de inserção.

4.2.4 – Artemis

O programa Artemis foi utilizado como ferramenta de anotação das sequências dos elementos retrotransponíveis encontrados no genoma de *Paracoccidioides*. Este programa, disponível em <http://www.sanger.ac.uk/Software/Artemis>, é uma ferramenta para anotação e visualização de características e resultados de análises de sequências, dados de “*next generation sequencing*” e permite visualiza-las nas suas seis possíveis fases de leitura aberta (ORF). O programa é livre, escrito em linguagem Java

podendo ser utilizado em vários sistemas operacionais como Windows, UNIX e MAC (Rutherford, Parkhill *et al.*, 2000).

4.2.5 - Análise filogenética

Realizou-se o alinhamento múltiplo de sequências de aminoácidos correspondentes à região de transcriptase reversa dos elementos identificados, RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5 por meio do algoritmo ClustalW (Thompson, Higgins *et al.*, 1994). A reconstrução filogenética das sequências alinhadas pelo programa foi realizada utilizando o método de distância *Neighbor-joining* implementado no programa MEGA5 (Tamura, Peterson *et al.*, 2011). A árvore foi construída usando o modelo *p-distance*, sendo a sustentação dos grupos avaliada por 5000 réplicas de *bootstrap*. Este modelo calcula as distâncias entre duas sequências considerando o número de aminoácidos diferentes entre elas e divide este número pelo total de aminoácidos avaliados. A confiabilidade dos dados foi verificada pelo cálculo da média Jukes-Cantor (JC) (Hall, 2008). As sequências utilizadas nessa análise estão relacionadas no anexo 9.

4.3 – Isolamento e caracterização de putativos retrotransposons

As sequências dos potenciais retrotransposons selecionadas para validação experimental, a partir das análises de dados gerados *in silico*, foram amplificadas por PCR convencional, clonadas e caracterizadas por sequenciamento. Os DNAs insertos correspondentes a cada elemento em estudo foram purificados de acordo com as instruções do Kit Wizard® SV Gel and PCR Clean-UP System e utilizados como sondas para a hibridação em *Southern blots* do DNA genômico de diferentes isolados do fungo, digeridos com diferentes endonucleases de restrição.

4.3.1 – Microrganismos

Foram utilizados dois isolados de *Paracoccidioides brasiliensis* (Pb18 e Pb03) e um isolado de *P. lutzii* (Pb01). Estes isolados foram objeto de estudo nos projetos de EST (Pb18 e Pb01) (Felipe, Andrade *et al.*, 2003; Goldman, Dos Reis Marques *et al.*, 2003), e junto com o isolado Pb03, foram sequenciadas e tornadas disponíveis no sitio do banco de dados do FGI Fungal Genome Initiative (www.broadinstitute.org/annotation/genome/paracoccidioides_brasiliensis/). Estes isolados estão disponíveis na coleção do Laboratório de Biologia de Microrganismos, ICB/UFMG, cedidos por diferentes pesquisadores. Além dos isolados acima, foram utilizados outros 31 isolados oriundos de diferentes regiões do Brasil (Tabela 5).

Tabela 5 - Isolados fúngicos do complexo *Paracoccidioides*.

Nome do Isolado	Classificação Filogenética
Pb01	Pb01- <i>like</i>
ED01	Pb01- <i>like</i>
1578	Pb01- <i>like</i>
Pb03	PS2
1017	PS2
1925	PS2
EPM83	PS3
EPM83	PS3
Pb18	S1
AP1	S1
B339	S1
MG4	S1
MG5	S1
Pb09	S1
Peru	S1
Pinguim	S1
Útero	S1
SS1	S1
470	S1
924	S1
6326	S1
9673	S1
EPM59	Não classificado
EPM62	Não classificado
EPM69	Não classificado
EPM81	Não classificado
EPM82	Não classificado
EPM87	Não classificado
EPM92	Não classificado
EPM97	S1
EPM101	Não classificado
EPM102	Não classificado
EPM117	Não classificado
EPM141	Não classificado

4.3.2 – Manutenção de amostras

Os isolados foram mantidos na fase micelial com repiques mensais, e a fase leveduriforme (L) foi induzida, quando necessário, por meio do crescimento a 37°C, e mantida com repiques semanais, ambas em tubos inclinados de ágar YPD (Extrato de levedura, 5 g; Peptona de caseína, 5 g; Dextrose, 15 g; Água bidestilada 1l; ágar a 1,5 % para meio sólido; pH 7,0).

4.3.3 – Extração de DNA genômico de isolados do complexo *Paracoccidioides*

A extração do DNA total foi realizada de acordo com Girardin e Latgé (Girardin e Latgé, 1994), com modificações. Para a obtenção de massas celulares, a amostra de *Paracoccidioides* previamente convertida à fase L foi submetida a três cultivos sucessivos em ágar YPD, com intervalos de 5 em 5 dias, para indução de crescimento exponencial (San-Blas, Suzuki *et al.*, 1994), e sucessivamente cultivada em caldo YPD, a temperatura de 37°C, sob constante agitação, por 7 dias. Após esse período o cultivo foi interrompido, as massas no estado leveduriforme foram centrifugadas a 8000 g por 10 min a 4°C e lavadas 3 vezes em EDTA (50 mM) [pH 8,0], por 10 min. As células foram então ressuspensas em volume conhecido de EDTA (50 mM).

Aproximadamente 300 mg de massa celular foi macerada na presença de nitrogênio líquido, num gral de porcelana esterilizado, até obtenção de um pó fino, não permitindo seu descongelamento. A massa celular macerada foi transferida para um tubo de 2 mL contendo 800 µL de tampão de lise e 8 µL de proteinase K (10 mg/µL) e incubada a 65°C por 10 min. Após, foram adicionados 600 µL de fenol e 200 µL de clorofórmio, sob agitação, até a obtenção de uma suspensão homogênea. A suspensão foi centrifugada a 13.000 rpm por 1 h a 8°C e a fase aquosa coletada e transferida para um novo tubo de 2 mL. Foram adicionados 800 µL de clorofórmio, centrifugando-se a 13.000 rpm por 20 min a 8°C. A fase aquosa foi transferida para um novo tubo contendo 12 µL de solução de RNase A (10 mg/µL) e incubada por 2 h à 37°C. Após a incubação foi adicionado igual volume, aproximadamente 800 µL, de clorofórmio e a suspensão foi agitada e centrifugada a 13.000 rpm por 10 min. A fase aquosa foi coletada e transferida para um novo tubo de 2 mL. O DNA foi precipitado com 840 µL de isopropanol e 40 µL de NaCl 4 M (0,2 mM NaCl final), solubilizado em

100 µL de água e mantido a -20°C. A avaliação qualitativa do DNA foi realizada após eletroforese convencional em agarose a 0,8%, em TBE, a 70 V, por 1 h e a concentração do DNA, determinada através leitura de OD a 260 nm em nanodrop.

4.3.4 - PCR convencional

Foram desenhados oligonucleotídeos iniciadores de acordo com a sequência alvo a ser amplificada (Tabela 6). O DNA genômico dos isolados foi submetido, à amplificação por PCR segundo o protocolo do kit Taq DNA Polimerase Recombinante – LGC Biotecnologia. A cada tubo de reação foram adicionados os seguintes reagentes: 0,2mM de cada dNTP, 1,5 mM MgCl₂ (50mM), 1 pmol dos iniciadores senso e anti-senso, 50 ng de DNA genômico, 2,5 U Taq DNA polimerase (5 U/µL) e água para um volume final de 25 µL.

O material foi homogeneizado e incubado em termociclador, observando-se os ciclos de desnaturação, anelamento e amplificação usuais (essencialmente 94°C por 5 minutos, 25 ciclos de 94°C por 30 segundos; “a temperatura de anelamento de cada *primer* por 30 segundos; e 72°C por 1 minuto para cada Kb a ser amplificado, finalizando em 72°C por 5 minutos), variando-se o programa essencialmente em função da temperatura de anelamento dos iniciadores a serem empregados (Tabela 6).

Para as reações de PCR para amplificação, a partir do cDNA, das sequências correspondentes à transcriptase reversa dos elementos estudados e dos genes controles gp43 e β-tubulina, utilizou-se os seguintes reagentes: 0,2mM de cada dNTP, 1,5 mM MgCl₂ (50mM), 100 nmol dos iniciadores senso e anti-senso, 2,5 U Taq DNA polimerase (5 U/µL) e água para um volume final de 25 µL. Foram feitas reações com diferentes volumes de cDNA (1 µL, 2 µL, 3 µL e 4 µL).

Tabela 6 - Relação de oligonucleotídeos (PCR convencional)

Primer		Sequência	Temperatura de melting
RtPb1	1F	5'-TGTTAGGCTCAGCGTCC3'	54,5 °C
	1R	5'-GTCCCATCAGTCCGGT-3'	57,0 °C
	2F	5'-ATTTGAACACTGGCTCTCCG3'	55,0 °C
	2R	5'-TCCATTTGCTCCAGTATCA3'	51,0 °C
	3F	5'-GTGAGAGGTTGAGATGAA-3'	49,0 °C
	3R	5'-CTTTGCTAGGTTGGCTTC-3'	51,0 °C
	4F	5'-CCAGCTGATGTTGACCTG3'	55,6 °C
	4R	5'-TCCGTCCAGAGCGTCAGT-3'	58,5 °C
	5F	5'-TGCCTTACTATGACCCTGAGG3'	56,7 °C
	5R	5'-CGGGTGGTTGGAATGGAGAT-3'	57,3 °C
	6F	5'-TGTTAGGCTCAGCGTCC3'	57,5 °C
	6R	5'-GTCCCATCAGTCCGGT-3'	57,3 °C
RtPb3	1F	5'-TGTGGGATGGTTGAATCT-3'	51,0 °C
	1R	5'-GTGCAAGATTTCCACCTG3'	51,3 °C
	2F	5'-CTAGGAGGGATGCCGAC3'	58,0 °C
	2R	5'-TACTTGGTTGAGGATGGCGAA-3'	56,3 °C
	3F	5'-TAGGGTCCAGGTTTTCCG3'	54,0 °C
	3R	5'-CAAATCAGGACTATTCTGCG3'	51,0 °C
	4F	5'-ATCTCGTCGAGCGTCTT-3'	53,0 °C
	4R	5'-GAGTTAGCCACACAAGATGA3'	52,0 °C
	5F	5'-TGTTATCGCCATTCTGGT-3'	54,3 °C
	5R	5'-TGTTGGGCGTGTGTTGC-3'	57,0 °C
RtPb4	1F	5'-TGTCAGGCATCCGCCTTCT-3'	60,4 °C
	1R	5'-AGAGCGTTCGGCTTGGTGGC-3'	63,2 °C
	2F	5'-GATCCGAACCTGAGGTCCG3'	55,0 °C
	2R	5'-TCAGCAGAGTACGCAGGCAG3'	59,7 °C
	3F	5'-GCTATCGAATGGGCGGACG-3'	58,5 °C
	3R	5'-AGACCTTTGCGGGTGGCAGC-3'	63,2 °C
	4F	5'-TTCGGAATCGCCATCATCGG3'	57,6 °C
	4R	5'-GGGTGAGGTTCTTCGAACTG-3'	57,7 °C
	5F	5'-AAGAACCTGCCTCCGCCGG3'	63,3 °C
	5R	5'-TGCAGAGAGCCAACCCCGA3'	63,8 °C
	6F	5'-TATGTGGACGATGTTCCGCTA-3'	55,7 °C
	6R	5'-TGTGAAGGATAGATTGCTTG-3'	50,0 °C
RtPb5	1F	5'-CCGAACGAGATAAAGCGGG3'	56,3 °C
	1R	5'-CCTGTGTCTGATTTACCAA3'	49,0 °C
	2F	5'-CTGATGGATCTCTCTCTGG-3'	53,8 °C
	2R	5'-GGCAGCTTTCTCTAGGGT-3'	54,5 °C
	3F	5'-GCAAAAGCTGATGTGATTCCG3'	52,3 °C
	3R	5'-GCTTGCAGAGAAGAAGCA3'	57,7 °C
	4F	5'-TCAATGTTTATCTCCGCTCG3'	52,2 °C
	4R	5'-GCCACATGCGTTCAGGAAT3'	56,1 °C
	5F	5'-TAAGACGTGAGTTCTACTGG3'	51,0 °C
	5R	5'-TTATACTACCCCTTCT3'	48,0 °C
M13	F	5'-GTAAAACGACGGCCAGT3'	52,0 °C
	R	5'-CAGGAAACAGCTATGAG3'	50,0 °C
T7		5'-TAATACGACTCACTATAGGG3'	56,0 °C
SP6		5'-ATTTAGGTGACACTATAG3'	48,0 °C
Sonda RtPb3	F	5'-TCTTAACCTGCTCGCTAACG3'	54,0 °C
	R	5'-TGTGATGGCGTATCTTCATGG-3'	55,0 °C
Sonda RtPb4	F	5'-TTAACCGCCTCAAGCAGATTG3'	55,5 °C
	R	5'-GCATCAATTCGCATACCAGAT3'	53,0 °C
Sonda RtPb5	F	5'-TCAGCGAGACTCTACAGCG3'	58,0 °C
	R	5'-CCCCTAGACGCATAAGCA3'	55,5 °C

4.3.5 – Extração de RNA

As amostras fúngicas foram homogeneizadas em 1 mL de TRIZOL (GIBCO/BLR Laboratories, Grand Island, N.Y., EUA). As amostras homogeneizadas foram incubadas por 10 min à temperatura ambiente (TA). Em seguida, foram adicionados 200 µL de Clorofórmio para cada 1 mL de TRIZOL seguidos de agitação em *vortex* e incubação por 3 min TA. O homogenato foi centrifugado a 14.000 rpm por 15 min a 8° C. O RNA foi precipitado com isopropanol (MERCK) (500µL de isopropanol para cada 1ml de TRIZOL). As amostras foram incubadas por 10 minutos em temperatura ambiente e então centrifugadas a 14.000 rpm por 10 min a 8°C. Após descarte do sobrenadante, foi adicionado 1 mL de etanol 75% (diluído em água tratada com DEPC) para 1mL de TRIZOL. Foi realizada nova centrifugação a 10.000 rpm por 5 min a 8° C. O RNA foi estocado a -70° C. A avaliação qualitativa do RNA foi verificada em gel desnaturante de agarose/formaldeído, observado sobre transluminador UV (540nm) e fotografado para verificação da integridade.

4.3.6 Quantificação do RNA

A quantificação das amostras de RNA foi feita em espectrofotômetro NanoDrop ND1000 (Nano Drop Technologies, Wilmington, DE, EUA). O RNA foi quantificado pela absorção a 260nm utilizando-se 2µL da amostra pura. O aparelho forneceu a concentração em ng/µL e a correlação 260/280 (RNA/proteínas).

4.3.7 - Transcrição Reversa para reações qRT-PCR

A reação de transcrição reversa foi realizada utilizando-se 2,0 µg de RNA e 50µM do primer Oligo dT (15) (Promega Cor., Madison WI, EUA). O tubo foi incubado por 5 minutos a 70°C e em seguida por 5 minutos no gelo para permitir a separação das fitas de RNA e o anelamento do *primer* e para impedir a formação de artefatos inespecíficos. Em seguida, foi adicionado 14,5µL de uma solução contendo: 1,5µL de dNTPs a 100 mM (Promega Cor., Madison, WI, EUA) , 4 µL tampão M-MLV 5X (Promega Cor., Madison WI, EUA) e 200 U da enzima Transcriptase reversa-M-MLV (Promega Cor., Madison WI, EUA) e 8,5µL de água Milli-Q autoclavada. A solução final foi incubada por 2 horas a 42°C para a

transcrição reversa do mRNA. As amostras de cDNA foram conservadas a -20°C.

4.3.8 – Transcrição reversa para reações RT-PCR iniciador específico

A reação de transcrição reversa iniciador específico foi realizada utilizando-se 2,0 µg de RNA total, tratado com DNase e 100nM do iniciador reverso correspondente à transcriptase reversa de cada elemento e de regiões dos controles gp43 e β-tubulina. O tubo foi incubado por 5 minutos a 70°C e em seguida por 5 minutos no gelo para permitir a separação das fitas de RNA e o ancoramento do iniciador e para impedir a formação de artefatos inespecíficos. Em seguida, foi adicionado 14,5µL de uma solução contendo: 1,5µL de dNTPs a 100 mM (Promega) , 4 µL tampão 5X (Promega) e 1 µL da enzima Transcriptase reversa ImpromII(Promega) e 8,5µL de água Milli-Q autoclavada, de acordo com as especificações do fabricante. A solução final foi incubada por 2 horas a 42°C para a transcrição reversa do mRNA. As amostras de cDNA foram conservadas a -20°C.

4.3.9 – PCR quantitativo em tempo real (qRT-PCR)

Inicialmente foi necessário verificar qual a concentração ideal dos pares de iniciadores seriam utilizados nas reações de qRT-PCR e avaliar a eficiência das mesmas por meio de análise de curva padrão. Foram testadas as concentrações de 100, 250 e 500 nM, temperaturas de 60, 62 e 64°C e diluições do cDNA molde de 10^0 a 10^{-5} . Constatou-se que a concentração ideal para todos os iniciadores foi de 100 nM, temperatura de 62°C e diluição de 10^{-1} para o cDNA molde. Foram determinadas também as porcentagens de eficiência para cada reação.

Para as reações subsequentes de amplificação foram utilizados 5µL da amostra de cDNA, diluído 1/10. A PCR foi realizada em um volume final de 20µL contendo 10µL de Power SYBR Green PCR Master Mix 2X (Applied biosystems, Foster City, CA, EUA), 1,0µL de cada iniciador (Tabela 7) a 1µM e 2µL de água ultrapura. A reação foi realizada utilizando-se o protocolo para PCR com SYBR Green do termociclador 7500 Fast System (Applied Biosystems, Foster City, CA, EUA). A reação foi realizada nas seguintes etapas: incubação inicial de 1 minuto

a 95°C e, 15 segundos a 95°C para desnaturação, 1 minuto para anelamento e extensão a 62°C. Esses dois últimos passos foram repetidos 40 vezes. Os resultados foram obtidos com auxílio do *software* 7500 V2.05 (Applied Biosystems, Foster City, CA, EUA). Os dados foram analisados utilizando-se o método de ΔCt em que $\Delta Ct = Ct$ do gene alvo – Ct do gene constitutivo. Dois genes constitutivos foram utilizados nestas análises: β -tubulina e L35 (codifica a proteína da subunidade ribossomal 60S). O valor atribuído ao C_T do gene constitutivo foi calculado pela média geométrica dos dois genes. Os valores de expressão foram obtidos pela fórmula: $2^{-\Delta Ct}$. Para confirmar a viabilidade dos genes β -tubulina e L35 como genes de referência foram utilizados os algoritmos de análise geNORM e NORMfinder (<http://www.gene-quantification.de/hkg.html#genorm>).

Tabela 7 – Relação de oligonucleotídeos (qRT-PCR)

Oligonucleotídeo		Sequência	Temperatura de <i>melting</i>
RtPb1	1F	5'-CAGGTCCTCAGACTCAGGGTC-3'	61,0 °C
	1R	5'-CCTACGCCTCCCAATCCTG-3'	65,0 °C
RtPb2	1F	5'-TCAGACGAAGGATGCTGCGG-3'	64,0 °C
	1R	5'-CCTGGAGAGTAACAACATTC-3'	61,0 °C
RtPb3	1F	5'-TACGGGTCGTATGCCGCATA-3'	65,0 °C
	1R	5'-CTCCACGGACATTGGTTTGTGC-3'	65,3 °C
RtPb4	1F	5'-CCCAGATATGGCGCTTCTACG-3'	62,0 °C
	1R	5'-CCGATACAACGTGGTCCGAATG-3'	64,0 °C
RtPb5	1F	5'-GCGAATTGGACAGGCTAAATGG-3'	65,3 °C
	1R	5'-CCGAGCAGAAGTCGTCCAGATAT-3'	64,0 °C
gp43	F	5'-CAGCTGTGTTCTCGCATGGG-3'	65,0 °C
	R	5'-TTAACCCCGCCGACGAAT-3'	65,0 °C
B-tubulina	F	5'-TCGTCCGTCGTGAGGCTG-3'	65,0 °C
	R	5'-CAGAAGGTCTCATCAGCGTGCTC-3'	61,0 °C
L35	F	5'-AATCGCTGGAGGTGCCG-3'	63,0 °C
	R	5'-CACAGCATATTTCCGCTGAGG-3'	63,0 °C

Estes oligonucleotídeos ancoram em regiões internas correspondentes aos genes de transcriptase reversa.

4.3.10 – Clonagem

Os fragmentos de DNA de *Paracoccidiodies sp.* amplificados por PCR foram clonados utilizando-se o *kit p-GEM-T Easy Vector System I* –Promega. Os fragmentos amplificados por PCR foram separados por eletroforese em gel de agarose e purificados com o *Kit Wizard® SV Gel and PCR Clean-Up System* – Promega. A seguir, os fragmentos foram ligados com o vetor *pGEM-T Easy* em um volume final de 10 µL. A reação de ligação: 5 µL do tampão de reação *2X Rapid ligation Buffer*, 1 µL *pGEM-T Easy Vector* (50 ng), 1-2 µL do produto de PCR, 1 µL T4 DNA ligase (3 U/µL) e água deionizada para o volume final de 10 µL. A ligação foi processada por 12-16 h a 4°C.

4.3.11 – Preparo de bactérias *Escherichia coli* competentes

Uma colônia da bactéria *E. coli* DH5α foi inoculada em 3 mL de meio LB (triptona 10 g/l, extrato de levedura 5 g/l, NaCl 10 g/l) e incubada com agitação a 37°C por 16h. Posteriormente, 3 mL da cultura foram inoculados em 250 mL de meio LB, e o frasco incubado com agitação a 37°C. O crescimento das bactérias foi determinado pela leitura da absorbância a 550 nm em espectrofotômetro.

Em seguida, o material foi incubado 60 minutos no gelo. Após esse período, a suspensão bacteriana foi centrifugada a 3000 rpm por 15 minutos a 4°C, descartando-se o sobrenadante. O precipitado foi lavado com solução FB (KCl 100mM, CaCl₂ – 2H₂O 50mM, glicerol 10%, acetato de potássio 10mM, pH 6,2) frio. Após nova centrifugação e descarte do sobrenadante, as bactérias foram ressuspensas em FB e incubadas por mais 60 min no gelo. Em seguida, o material foi centrifugado, ressuspensão em 20 mL de FB, distribuído em alíquotas de 500 µL em tubos de microcentrífuga e mantidos a –70°C até o momento do uso.

4.3.12 – Transformação bacteriana

Os plasmídios recombinantes foram utilizados para transformar bactérias DH5α em um volume de 100 µL. A suspensão bacteriana foi incubada por 1 h em gelo, e posteriormente, submetida a choque térmico à 42°C por 2 min, seguido de incubação por 10 min no gelo. A seguir, foram acrescentados 400 µL

de meio LB e incubação a 37°C por 1 hora sob agitação. Após esse período, 200 µL do cultivo de bactérias transformadas foram semeados em placas de Petri contendo LB sólido (triptona 10 g/l, extrato de levedura 5 g/l, bacto-ágar 15 g/l, NaCl 10 g/l, ampicilina 100µg/mL) e incubadas em estufa a 37°C durante 12-16 h.

4.3.13 – Extração de DNA plasmidial (preparação em pequena escala)

Alíquotas de bactérias recombinantes foram inoculadas em 3 mL de meio LB contendo ampicilina (100µg/mL) e incubadas por 18 h a 37°C. As bactérias foram precipitadas por centrifugação a 3.500 rpm por 5 min, o precipitado ressuspenso em 100 µL de solução A (glicose 50 mM, EDTA 10 mM, Tris HCl 25 mM, pH 8,0) contendo Rnase (1,6 µg/mL) e transferido para tubos de microcentrífuga de 1,5 mL. Após adição de 200 µL de solução B (SDS 1%, NaOH 0,2 M) e incubação a temperatura ambiente por 5 min, foram acrescentados 200 µL de solução C (acetato de potássio 3 M, ácido acético 5 M) e a mistura incubada por 30 min no gelo.

Após centrifugação a 13.000 rpm por 15 min a 8°C, 400 µL do sobrenadante foram transferidos para outro tubo de microcentrífuga contendo 300 µL de isopropanol. O DNA foi precipitado por centrifugação a 13.000 rpm por 8 min a 13°C. O sobrenadante foi desprezado e o precipitado lavado com 400 µL de etanol 70%. O precipitado contendo DNA foi seco em estufa a 42°C por 30 min e ressuspenso em 30 µL de solução TE.

4.3.14 – Sequenciamento de DNA

Foi utilizado o kit Big Dye 3.1, Applied Biosystem. Três clones de cada amostra foram sequenciados duas vezes com cada iniciador (direto e reverso). Cada reação de sequenciamento foi feita com os seguintes componentes: 2 µl de Big Dye, 1 µl de tampão 5X, acrescentando-se 1,6 pmoles de *primer*, 400-600 ng de DNA plasmidial e água para um volume final de 10 µl. Os componentes da reação foram homogeneizados por rápida centrifugação e submetidos à reação e PCR nas seguintes condições: 25 ciclos a 96°C por 20 segundos, 50°C por 10 segundos e 60°C por 4 min. Para a precipitação do DNA foram adicionados 100 µL de etanol 80% gelado, seguindo-se incubação por 4°C durante 30 min no

escuro e centrifugação a 4.000 rpm durante 40 min. O DNA foi lavado com etanol 70% gelado (200 µL) por centrifugação nas mesmas condições. O material foi seco e ressuspensão em 2,5 µL de tampão “Loading Buffer” (Perkin Elmer).

As amostras foram submetidas à eletroforese em gel de poliacrilamida no aparelho ABI Prisma 377, Applied Biosystem. Uma solução contendo acrilamida (9 g de uréia, 2,5 mL de acrilamida/bisacrilamida, 2,5 mL TBE 10X, 13 mL H₂O Milli-Q) foi homogeneizada com agitador magnético por 20-30 min e polimerizada pela adição de 125 µL de APS (persulfato de amônia 10%) e 17,5 µL TEMED. As amostras (2,5 µL) foram incubadas a 90°C durante 2 min e aplicadas no gel.

As sequências obtidas foram analisadas empregando-se o programa Phred Phrap Consed, disponível em (<http://www.phrap.org/phredphrapconsed.html>). As sequências consideradas de boa qualidade com os respectivos cromatogramas foram alinhadas. Após este processo foram obtidas sequências consenso, que foram utilizadas para busca por homologia no banco de dados do NCBI e analisadas.

4.3.15 – Restrição de DNA genômico

O DNA genômico de diferentes isolados do complexo *Paracoccidioides* foi digerido com 10U (unidades) das seguintes enzimas de restrição (Tabela 7). Em um tubo de microcentrífuga foram adicionados 10 µg do DNA genômico, 2 µL de tampão de reação da enzima de restrição 10X, 1 µL de enzima (10U) e H₂O suficiente para um volume final de 20 µL. A reação foi incubada a 37°C por 18 horas e os fragmentos de restrição gerados foram submetidos a uma eletroforese em géis de agarose a 0,8% a 50v por 15 horas. Os géis foram fotografados sob luz UV e transferidos para membranas de náilon.

Tabela 8 – Relação das endonucleases de restrição.

Retroelemento	Enzima	Sítio de corte	Observação
RtPb3	<i>EcoRI</i>	G▼AATTC	Não corta o elemento
	<i>Hind III</i>	A▼AGCTT	Corte interno
RtPb4	<i>Nco I</i>	C▼CATGG	Não corta o elemento
	<i>Hind III</i>	A AGCTT	Corte interno
RtPb5	<i>Sal I</i>	G▼TCGAC	Não corta o elemento
	<i>Eco RI</i>	G▼AATTC	Corte interno

4.3.16 – Southern blot

Os fragmentos de DNA resolvidos eletroforéticamente foram transferidos para membranas de náilon (Southern, 1975) (*Hybond-N – Amersham Pharmacia Biotech*), utilizando o sistema de transferência a vácuo “VacuGene™ XL” (Amersham Pharmacia Biotech).

O gel contendo o DNA genômico íntegro ou digerido com enzimas de restrição foi posicionado sobre a membrana, previamente umedecida com água bidestilada estéril, no sistema de transferência; em seguida foi tratado com as soluções de depurinação (HCl 250 mM), desnaturação (NaCl 1,5 M e NaOH 500 mM, pH 7,5) e neutralização (NaCl 0,6M, Tris-HCl 1M, pH 8,0), durante 30 min cada, com pressão entre 55 e 60 psi. O DNA foi transferido com solução de transferência (SSC 20 X – NaCl 150 mM e citrato de sódio 15 mM) por 3 h. Após a transferência, o DNA foi fixado às membranas irradiando-se as membranas úmidas, com luz UV, durante 3 min. As membranas obtidas foram estocadas à -20 °C.

Fragmentos correspondentes às sequências das transcriptases reversas dos elementos estudados foram marcados radioativamente e utilizados para hibridação de *Southern blots* utilizando-se marcação isotópica com [³²P] dCTP conforme protocolo descrito por (Feinberg e Vogelstein, 1983). A pré-hibridação das membranas foi feita em 15 mL da solução G (1% BSA, 500 mM NaH₂PO₄, 1 mM EDTA e 7% SDS) (Church e Gilbert, 1984) por 2h à temperatura de 48°C. As sondas foram adicionadas à solução G após desnaturação (95°C durante 5 min e logo após 5 min no gelo) e incubadas durante 16h à 48°C. Após a hibridação, as membranas foram lavadas 4 vezes com SSC 2 X (SSC 1X 150 mM NaCl, 15 mM citrato de sódio) e 0,1% SDS a temperatura ambiente. Após a lavagem, as membranas foram expostas ao filme de raio X e incubadas a -70°C. Após três dias, os filmes de raio X (Kodak) foram revelados e fixados.

4.3.17 – Fluxograma de validação experimental

Como tentativa inicial de validar as previsões computacionais realizadas, utilizou-se a estratégia de *primer walking* para amplificação do elemento inteiro nos genomas de fungos do complexo *Paracoccidioides*. Essa estratégia viabiliza a amplificação de grandes fragmentos de DNA a partir do desenho de iniciadores

específicos que permitem a amplificação de trechos de sequências menores que quando sobrepostos pelas regiões terminais permitem a reconstrução do fragmento original. A figura 5 representa de forma esquemática a abordagem utilizada na validação experimental dos elementos RtPb3, RtPb4 e RtPb5. Esta validação constou das seguintes etapas: a) amplificação, por PCR (*primer walking*) dos elementos inteiros nos genomas dos isolados Pb01, Pb03 e Pb18; b) amplificação por PCR das regiões de transcriptase reversa de cada elemento nos genomas de 31 isolados; c) clonagem e sequenciamento de aproximadamente 300pb correspondente à transcriptase reversa de cada elemento; d) mapeamento por *Southern blot* dos elementos estudados e e) análises dos transcritos por PCR quantitativo em tempo real

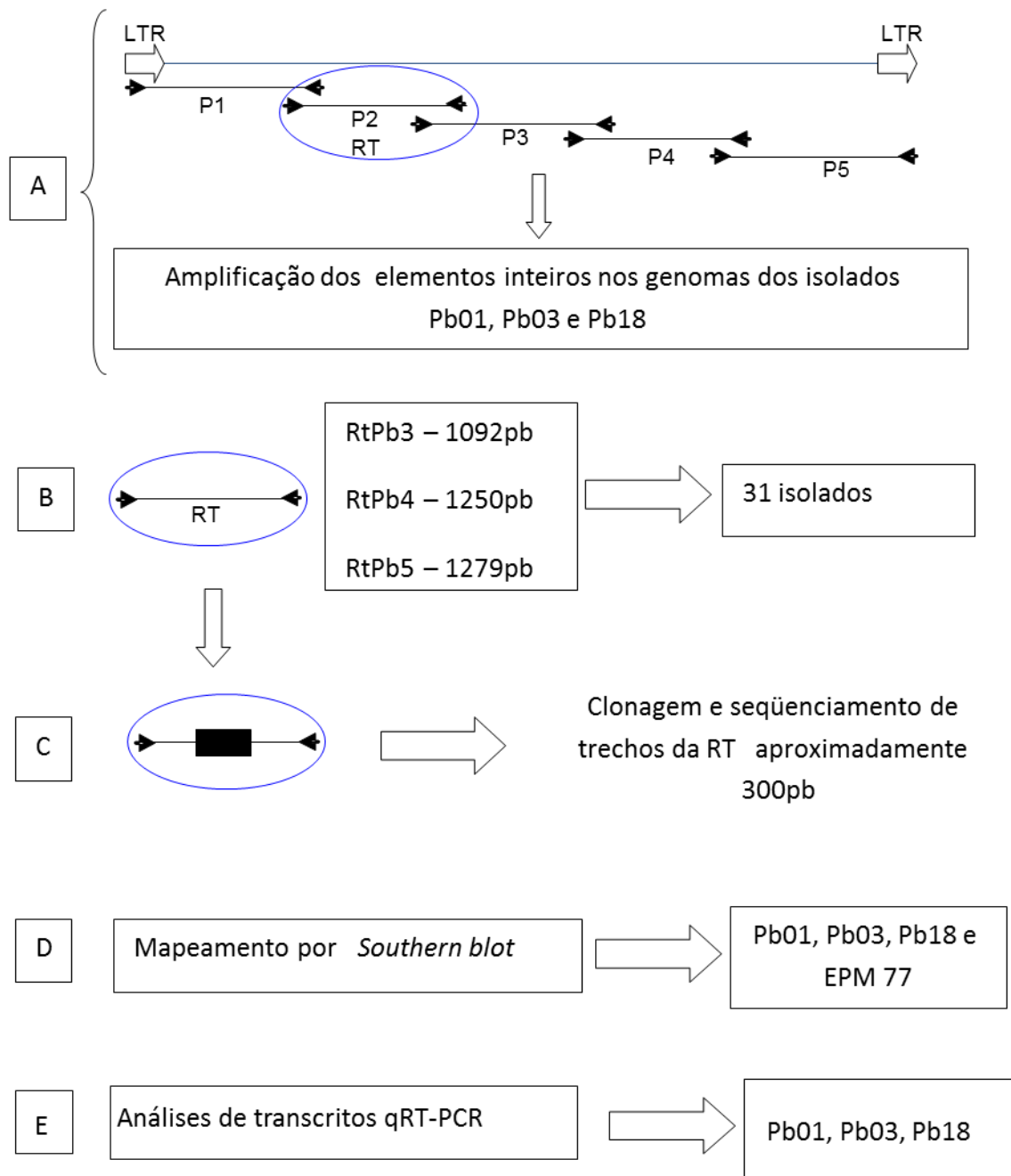


Figura 5 – Representação esquemática dos testes de validação experimental dos dados obtidos por análises de bioinformática

LTR – terminações longas repetidas; P1, P2, P3, P4 e P5 – pares de iniciadores; Pb01, Pb03 Pb18 e EPM87– Isolados de *P. brasiliensis*; RT – fragmento correspondente à transcriptase reversa; pb – pares de bases; qRT-PCR – Reação em cadeia da polimerase quantitativa em tempo real.

5 – RESULTADOS

5.1 – Análise das sequências de EST e geração de grupos de sequências.

Tendo como objetivo a identificação de elementos transponíveis nos genomas de fungos do complexo *Paracoccidioides*, estabeleceu-se como ponto de partida a análise computacional de dados presentes em bancos de domínio público. Apesar de existir um trabalho recente de genômica comparativa apresentando um *pool* de elementos transponíveis nos genomas do complexo de espécies *Paracoccidioides*, realizado por Desjardins e colaboradores (Desjardins, Champion *et al.*, 2011), pouca ênfase foi dada à caracterização de sequências repetitivas sendo realizada apenas uma classificação geral dos putativos elementos transponíveis.

Dentro desse contexto, frente à escassez de informações detalhadas sobre a estrutura e organização dos retrotransposons nos genomas de fungos do complexo *Paracoccidioides*, e do conhecimento relativo à presença desses elementos entre as sequências mais transcritas, estabeleceu-se uma estratégia que utilizou as informações contidas no banco de dados de EST (*expressed sequence tags*) do NCBI (<http://www.ncbi.nlm.nih.gov/>) para a identificação de segmentos dos elementos móveis possivelmente expressos, visando sua posterior caracterização. A metodologia estabelecida envolveu, inicialmente, a criação de um banco de dados local de EST e o agrupamento dessas sequências por meio do programa Cluster 5 (JM-Ribeiro), como descrito no item 4.1.

Foram analisadas 41558 EST de *Paracoccidioides* sendo gerados 12922 grupos de sequências, perfazendo um percentual de aproximadamente 31% (12922/41558) do número inicial (Figura 6 – painel A). Como explicitado nesta figura, ~37% dos grupos (4812/12922) foram compostos por mais de duas sequências de EST, denominados *contigs*, sendo que os outros 63% (8110/12922) apresentaram sequências individuais contabilizadas como *singlets*. Os *singlets* representam aproximadamente 19% do total das EST (8110/41558). No conjunto de grupos gerados, o maior agrupamento (contig 4177) apresentou 2315pb de comprimento enquanto que o menor (contig 75) tem 217pb.

O perfil de distribuição das sequências de EST é mostrado no painel B da figura 6. *Contigs* formados por apenas duas sequências contabilizam 13,8% (1791/12922) e com o progresso do agrupamento observou-se que houve uma diminuição gradativa do número de *contigs* no intervalo entre uma e dez sequências de EST agrupadas. Apenas sete *contigs* apresentaram um agrupamento contendo mais de 200 EST e apenas um deles apresentou mais de 1000 EST agrupadas (Figura 6). É interessante destacar que três destes sete *contigs* apresentaram similaridades com elementos genéticos móveis nos bancos de dados pesquisados, a saber

- db_est-contig 54 (411 EST, formamidase, transposase, DNA transposon);
- db_est-contig 194 (361 EST, transcriptase reversa, retrotransposon LTR);
- db_est-contig 959 (286 EST, transposase, DNA transposon);

Contudo, não foi possível verificar a presença de ORF íntegras associadas a esses *contigs* e o seus alinhamentos mostraram evidências sugestivas de regiões de baixa complexidade.

A

Clusterização – CAP 3							
<i>P. brasiliensis</i>	Número inicial de ESTs	Número de grupos	%	<i>Singlets</i>	Contigs com mais de 2 sequências	> contig (nt*)	< contig (nt*)
Atualização 100111	41558	12922	31,09%	8110	4812	4177 (2315)	75 (217)

*(nt) nucleotídeo

B

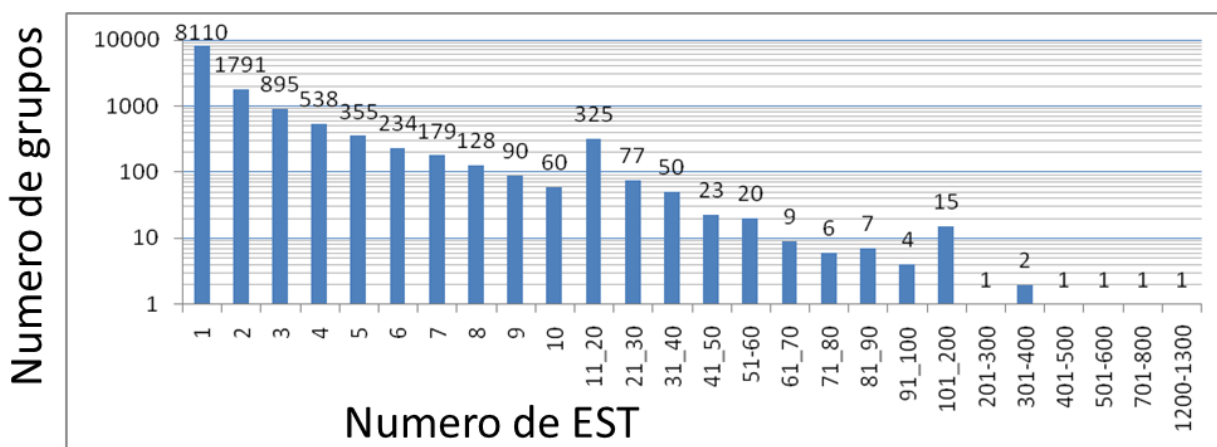


Figura 6 – Agrupamento e distribuição de ESTs.

Painel A - Perfil dos grupos de sequências, *contigs* e *singlets*, gerados após agrupamento por meio do programa CAP3. Número de grupos – número de *contigs* e *singlets* analisados; % - percentual de agrupamento; *singlets* – sequência única que não apresentou alinhamento com outras sequências; *contigs* – sequência consenso resultante do alinhamento duas ou mais sequências; > contig – os números fora e dentro do parêntese representam respectivamente a identidade e o tamanho, em nucleotídeos, do maior *contig*; < contig - os números fora e dentro do parêntese representam, respectivamente, a identidade e o tamanho, em nucleotídeos, do menor *contig*; **Painel B** - Perfil de distribuição das EST nos grupos formados. Os números sobre as barras representam os números de grupos formados desde por 1 EST (*singlets*, 8110) até por 1200-1300 EST (1).

5.2 – Busca por similaridade dos grupos de *EST* e anotação funcional.

Tendo-se gerado os grupos a serem analisados era necessário identificá-los. Esse processo foi realizado pela busca de similaridades com sequências conhecidas e depositadas em bancos de dados públicos utilizando a ferramenta de alinhamento local BLAST (*Basic Local Alignment Search Tool*) (item 4.1.5). Os 12922 grupos foram comparados com sequências depositadas em diferentes bancos de dados (Figura 7). A comparação foi feita contra cada banco separadamente, em servidor local, objetivando a identificação de regiões dos grupos que pudessem apresentar similaridade significativa com trechos característicos de retrotransposons. Todos os resultados foram reunidos e armazenados em uma única tabela do Excel em formato *hiperlincado* (Tabela *hiperlincada* 1). Esta tabela, pelo seu grande tamanho e impossibilidade de ser apresentado de modo impresso, não consta deste documento, podendo ser visualizada na forma digital. Esse formato viabiliza consultas e acessos rápidos às informações importantes tais como: a sequência original dos grupos, resultado do alinhamento entre as sequências analisadas frente aos diferentes bancos de dados e os valores de confiança estatística das análises.

Como pode ser observado na figura 7, no contexto das análises por similaridade de sequências, foram utilizados inicialmente quatro bancos de dados diferentes (item 4.1.5): NR, banco de dados de proteínas não redundantes (NCBI); TEfam, banco de elementos transponíveis da Universidade da Virgínia, EUA; TEClass, banco de matrizes de domínios conservados representativos de elementos transponíveis (JM-Ribeiro); RepBase, banco de dados de elementos repetitivos pertencente ao GIRI (*Genetic Information Research Institute*).

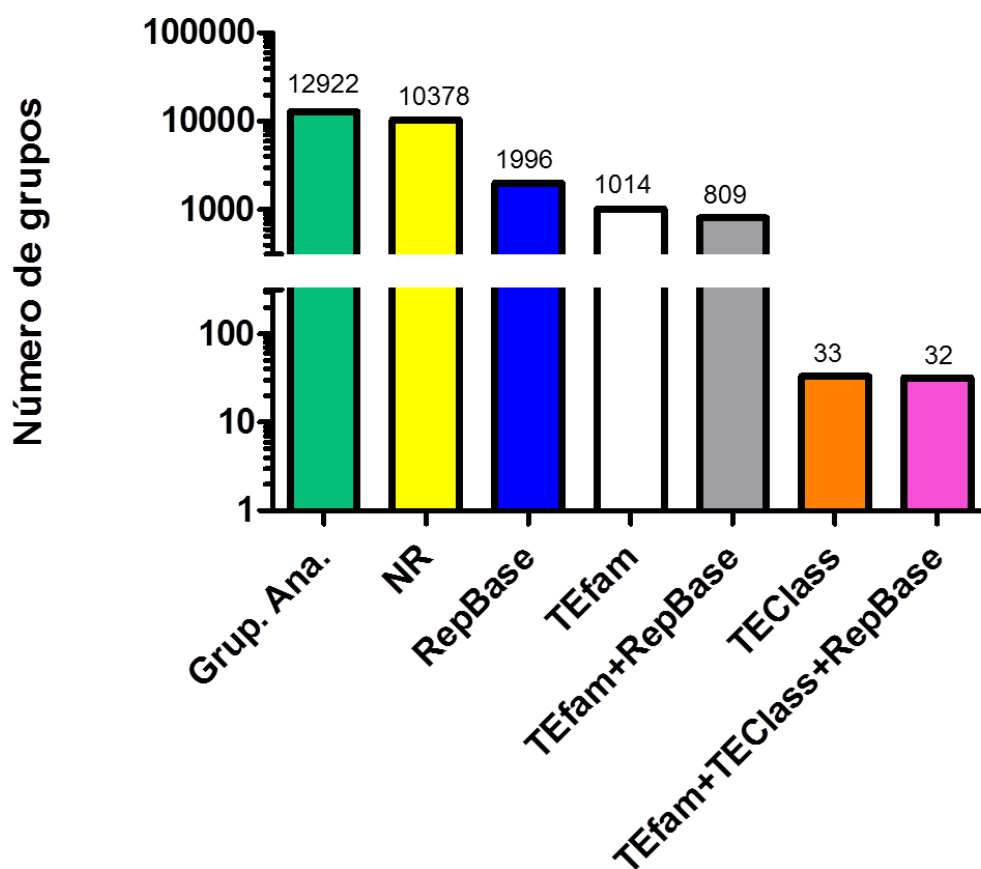


Figura 7 – Resultados das buscas por similaridade

Grup. Ana.– Grupos analisados; **NR** – banco de dados de proteínas não redundantes (NCBI); **RepBase** – banco de sequências repetitivas do GIRI (*Genetic Information Research Institute*); **TEfam** – banco de elementos transponíveis da Universidade da Virgínia, EUA; **TEClass** - banco de matrizes de domínios conservados representativos de elementos transponíveis; O banco de dados TEClass foi construído localmente utilizando a ferramenta *PSblast* (*Position-Specific Iterative Blast*) para construção de matrizes. As sequências utilizadas correspondem a regiões codificadoras de diversas superfamílias que representam elementos transponíveis e retrotransponíveis (Tabela 2), obtidas no banco de dados NR. **TEfam+RepBase** - Grupos comuns nestes dois bancos; **TEfam+TEClass+RepBase** - Grupos comuns nestes três bancos. Os valores (acima das barras) representam o número de grupos que apresentaram similaridade em cada um dos bancos pesquisados.

As análises contra esses bancos foram divididas em duas etapas:

1^a) Análise contra o banco de dados (NR) do NCBI (<ftp://ftp.ncbi.nlm.nih.gov/blast/db/>); 2^a) Análise contra bancos de dados específicos de elementos transponíveis (TEfam, TEClass e RepBase). Contra o NR, dos 12922 grupos analisados aproximadamente 80% (10378/12922) apresentaram similaridade com sequências já depositadas, dentro dos parâmetros estabelecidos (item 4.3.1). Os grupos que não apresentaram similaridade com nenhuma sequência depositada no NR, aproximadamente 20% (2544/12922), foram submetidos a uma nova análise e os resultados validados, ou seja, não foi evidenciada similaridade. Na segunda etapa, nos resultados dos bancos de dados específicos (TEfam, TEClass e RepBase), observou-se, quando da comparação com o banco de dados TEfam (<http://tefam.biochem.vt.edu/tefam/search.php>), que 7.8% dos grupos analisados (1014/12992) apresentaram regiões de similaridade com sequências de elementos transponíveis; 15.4% dos grupos (1996/12992) apresentaram similaridades com sequências do banco RepBase (<http://www.girinst.org/replib/index.html>) e apenas 0,25% dos grupos (33/12922) revelou similaridades com sequências depositadas no banco TEClass (JM-Ribeiro). Os grupos comuns aos dois principais bancos de elementos transponíveis (TEfam e RepBase) foram 809, o que equivale a 6,2% do total de grupos analisados. Este valor caiu para 32, ou seja, 0,24% do total, quando foi considerada a sobreposição de resultados dos três bancos (TEfam, RepBase e TEClass).

Tendo em vista que os valores encontrados no banco NR (80%, 10378/12922 grupos) não representam exclusivamente *hits* com elementos móveis, uma vez que o banco alberga todas as proteínas depositadas no NCBI, fez-se necessário a filtragem, tendo como objetivo a extração acurada e direcionada, relacionados a elementos retrotransponíveis. Essa análise inicial envolveu o desenvolvimento de um *script* (Anexo1) objetivando a identificação de oito palavras-chave (Figura 8) provavelmente presentes nas anotações. Essa abordagem envolveu a busca léxica de oito termos anotadores e auxiliou na identificação, dentre os 40 melhores *hits* para cada grupo, daqueles que pudessem representar regiões características de retrotransposons. Como resultados dessa abordagem de filtragem, foi possível concentrar os resultados,

obtendo-se um número menor de 238 grupos (Figura 8) que ficou reduzido a 142 grupos (anexo 2), quando se eliminou a redundância de termos anotadores, a serem considerados para análises futuras.

A frequência de aparecimento dos termos anotadores encontrados (em ordem decrescente) (Figura 8) foi: ~35% (84/238) transcriptase reversa; ~20% (49/238) gag; ~10% (25/238 e 24/238) retrotransposon e endonuclease; ~8% (19/238) Retrovírus e Rnase H; ~5% (12/238) integrase; e ~2.5% *Copia-like*. A descrição completa dos resultados obtidos incluindo a identidade do *hit* seguido de seu termo anotador, os valores de score e de *e-value* correspondentes encontram-se no anexo 3. É importante ressaltar que foram considerados, para construção dessas tabelas, apenas os *hits* que apresentaram *e-value* menores ou iguais a 10^{-6} .

Uma visão geral dos resultados das análises realizadas nas duas etapas anteriormente descritas (análises contra NR e bancos específicos) está representada nos painéis A e B da figura 9. No painel A é mostrado o número de grupos obtidos em cada análise. Contra o banco NR (vermelho) foram obtidos 142 grupos. Todos os 142 grupos do NR foram representativos apenas de retrotransposons. Pela análise em bancos específicos (azul) 809 grupos apresentaram similaridade com elementos móveis. A partir desses dados foi possível verificar quais deles apresentavam similaridade especificamente com transposons de DNA e retrotransposons. A discriminação preferencial para transposon de DNA se baseou na presença do termo anotador transposase. O painel B da figura 9 mostra essa subdivisão: dos 809 grupos encontrados nos bancos específicos, 437 foram representativos de transposons de DNA e 372 de retrotransposons.

Um passo natural, então, e que permitiria refinar os resultados, foi verificar quais eram os grupos de EST apontados paralelamente pelas duas abordagens como correspondentes a putativos retrotransposons. O painel C da figura 9 mostra esse resultado indicando que 52 grupos de EST eram comuns (Anexo 4) aos resultados das duas abordagens (NR e bancos específicos). ou seja, estavam presentes nos resultados finais de cada uma das duas abordagens.

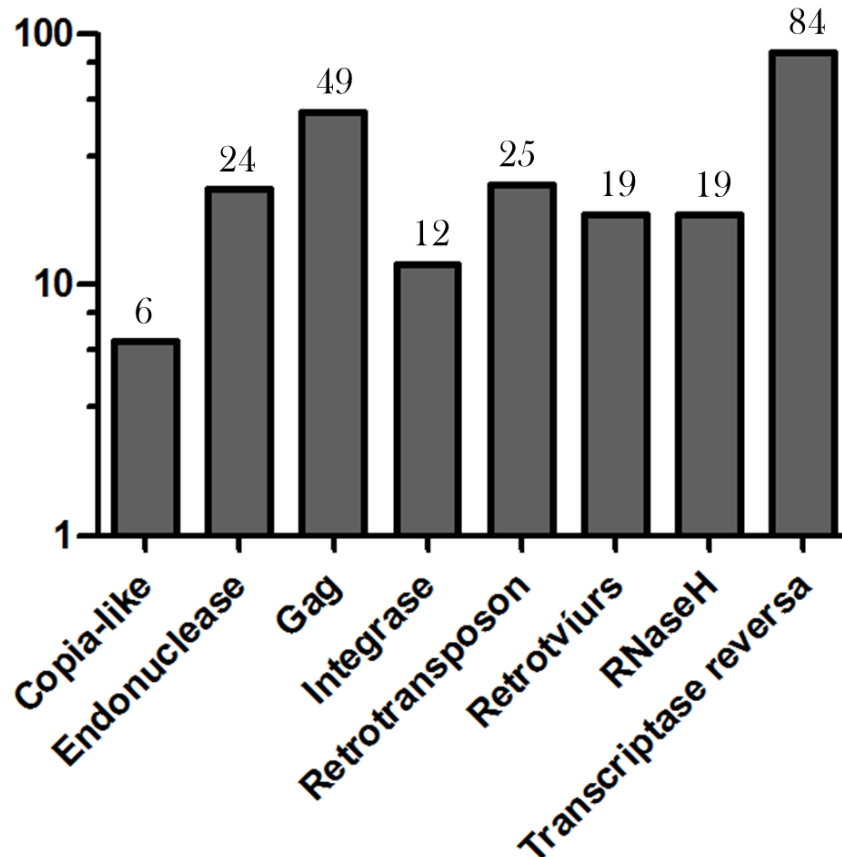


Figura 8 - Número de *contigs* no banco de dados NR correspondentes a retrotransposons.

O eixo X representa os termos descritores utilizados para identificação dos elementos nos bancos de dados. Empregaram-se termos presentes em anotações de elementos retrotransponíveis (termos anotadores). *Copia-like*, termo utilizado para designar elementos da ordem *Copia*; Endonuclease, enzima responsável por clivar as fitas de DNA para inserção do elemento; Gag, proteínas formadoras da VLP (*Virus-like particle*); Integrase, proteína responsável pela integração do elemento no genoma hospedeiro; Retrotransposon, elementos de transposição que possuem intermediário de RNA; Retrovírus, termo utilizado na denominação de elementos móveis; RNaseH, denominação do sítio catalítico da proteína transcriptase reversa cuja função é degradar a fita molde de RNA; Transcriptase reversa, proteína com função e DNA polimerase que catalisa a replicação de DNA a partir de um molde de RNA. Os números sobre as barras representam os grupos encontrados com cada termo anotador.

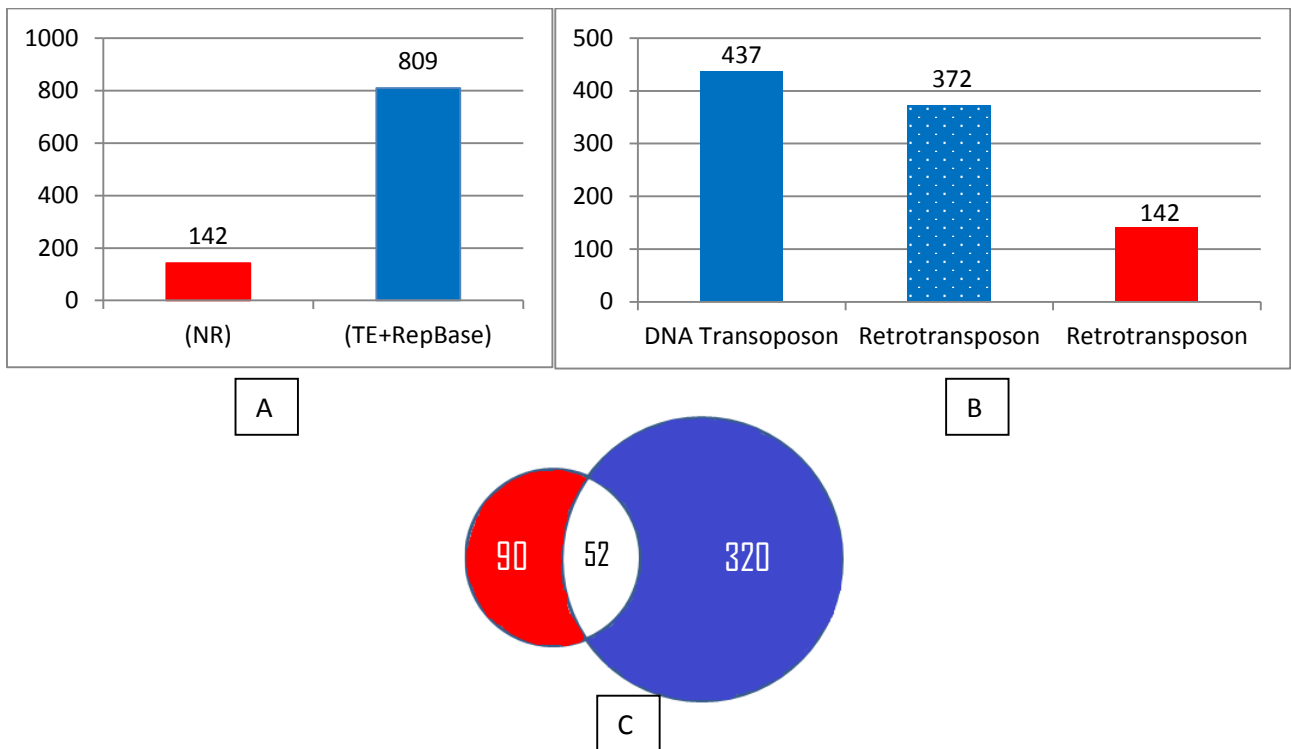


Figura 9 – Grupos de seqüências representativos de elementos genéticos móveis

Grupos que apresentaram *hits* nos bancos NR (NCBI) e bancos específicos (TEfam+Repbase), distinguindo-se os grupos similares aos elementos de classe 1 (retrotransposons) dos de classe 2 (transposons de DNA). Painel A – O gráfico representa o número de grupos encontrados no banco de proteínas não redundantes NR (vermelho, 142) e em bancos específicos (azul, 809). Painel B – Os gráficos representam subconjuntos do painel A - Número de grupos que apresentaram *hits* com elementos das classes 1 (Retrotransposon) (azul pontilhado em branco, 437) e 2 (transposons de DNA) (azul sólido, 437) nos bancos específicos. Em vermelho, número de grupos do NR correspondentes a putativos retrotransposons (142). Painel C – O gráfico destaca os grupos comuns ao banco NR e aos bancos específicos de elementos móveis (52 grupos, setor branco) e aqueles exclusivos de cada abordagem, outros 90 grupos com *hits* indicativos de similaridade com retrotransposons no NR e 320 grupos com *hits* indicativos de similaridades com seqüências de bancos específicos (TEfam, TEClass e Repbase).

A partir desse conjunto de 52 grupos, buscou-se por aqueles que, além de estarem presentes em todos os bancos analisados, apresentassem uma maior densidade de termos anotadores mapeados e pudesse servir de ponto de partida para a localização de elementos completos ou íntegros. Isso permitiu que o processo de análise ficasse concentrado inicialmente em cinco grupos de EST mais informativos (1145, 1945, 8104, 9844 e 12801) (Tabela 9). Estes grupos apresentaram similaridade com sequências pertencentes a retrotransposons de diferentes grupos taxonomicos. O grupo 1145 foi identificado como relacionado a elementos da ordem LTR e superfamília *Gypsy* (RepBase e TE-Class), cujos resultados de *Blast* (bancos NR e TEfam) apontam similaridades com regiões ou domínios da poliproteína POL. Quanto ao grupo 1945, verificou-se a existência de similaridades com regiões de transcriptase reversa (NR e TEfam) e com elementos pertencentes à ordem LINE (RepBase e TEfam).

Já o grupo 8104 apresentou similaridade com elementos pertencentes à ordem LTR, porém, foi classificado na superfamília *Copia* (RepBase e TE-Class). De maneira semelhante, o grupo 9844 apresenta similaridade com elementos da superfamília *Copia* e também com poliproteína POL. Por fim, o grupo 12801 foi classificado como pertencente à superfamília *Gypsy* (TE-Class e RepBase) enquanto que as análises dos outros dois bancos relataram similaridade com sequências de poliproteína de retrovírus.

Empregando-se esses cinco grupos como sequências de busca (Tabela 9) contra as sequências genômicas disponíveis de *Paracoccidoides* foi possível identificar regiões que apresentaram identidade significativa com os putativos retrotransposons, encontrando-se cópias completas de cada elemento, em pelo menos um dos 3 genomas. Observou-se, também, que nessas regiões ocorria o ancoramento de outros grupos: - dos 52 grupos que estavam presentes simultaneamente tanto nos resultados das buscas no NR quanto nos bancos específicos, 27 foram ancorados a cinco regiões genômicas correspondentes aos cinco prováveis retrotransposons que apresentaram pelo menos uma cópia completa em algum dos genomas do complexo *Paracoccidoides*. A Tabela 10 mostra os grupos constituintes de cada elemento, o número e a origem das EST constituintes de cada grupo. O elemento que apresentou o maior número de etiquetas expressas foi o RtPb1 (98), sendo observado sequências descritas em quatro trabalhos (Felipe, Andrade *et al.*, 2003; Goldman, Dos Reis Marques *et*

al., 2003; Felipe, Andrade *et al.*, 2005; Felipe, Torres *et al.*, 2005; Bastos, Bailão *et al.*, 2007). O elemento RtPb2 também apresentou um número considerável de sequências (23) quando comparado com os outros: RtPb3(3), RtPb4(2) e RtPb5(1). Os outros 25 grupos restantes (anexo 4) foram analisados quanto ao contexto genômico e foram mapeados em regiões com grande número de códons de terminação, não sendo possível localizar e associar a eles cópias íntegras ou completas de outros retrotransposons.

Tabela 9 – Grupos representativos de putativos elementos retrotransponíveis e resultados de similaridade (*hits*, *e-values*, identidades, escores) encontrados nos bancos NR, TEfam, TE-Class e Repbase.

Grupo	Tamanho(pb)	Hit NR	E value	Identidade	Score	id %
1145	2295	retrotransposon polyprotein - <i>Talaromyces stipitatus</i> ATCC10500	e-169	gi 242760779	535	51
1945	2375	reverse transcriptase - <i>Talaromyces stipitatus</i> ATCC10500	0.0	gi 242826474	325	38
8104	494	retrotransposon protein, putative - <i>Oryza sativa</i>	5,00E-21	gi 77552925	104	35
9844	979	copia-like retrotransposon - <i>Coccidioides posadasii</i>	4,00E-12	gi 303321121	77	43
12801	341	retrovirus polyprotein <i>Penicillium marneffe</i> ATCC18224	4,00E-35	gi 212530907	151	60
Grupo	Tamanho(pb)	Hit TEfam	E value	Identidade	Score	id %
1145	2295	hypothetical protein (T18348 probable pol polyprotein, truncated - rice blast	1E-087	gi 149209059	322	38
1945	2375	reverse transcriptase	0.0	gi 162568710	1108	99
8104	494	Putative retroelement - <i>Oryza sativa</i>	5,00E-21	gi 23477804	98	35
9844	979	putative retroelement pol polyprotein	3,00E-13	gi 4914361	73.6	35
12801	341	probable pol polyprotein - rice blast fungus gypsy retroelement (fragment)	7,00E-28	gi 7493955	119	48
Grupo	Tamanho(pb)	Hit TE-CLASS	E value	Identidade	Score	id %
1145	2295	Gypsy_noGag	1,00E-116	Gypsy_noGag	407	30
1945	2375	LINE	7E-44	Outcast_Ele5_ORF2	167	37
8104	494	Copia_all	1,00E-29	Copia_all	117	31
9844	979	Copia_all	2,00E-28	Copia_all	115	31
12801	341	Gypsy_noGag	1,00E-25	Gypsy_noGag	104	34
Grupo	Tamanho(pb)	Hit REPBASE	E value	Identidade	Score	id %
1145	2295	Gypsy-9_BG-I LTR retrotransposon from barley powdery mildew: internal portion	1,00E-179	Gypsy-9_BG-I	567	47
1945	2375	Tad1-4_BG Tad1 Non-LTR retrotransposon from barley powdery mildew, consensus	1,00E-104	Tad1-4_BG	105	45
8104	494	copia-2-I_AN Internal portion of copia-2_AN LTR retrotransposon - a consensus	6,00E-31	copia-2-I_AN	91.3	53
9844	979	MTANGA_I Anopheles gambiae mtanga retrotransposon (internal portion)	1,00E-17	MTANGA_I	68.0	42
12801	341	Gypsy-1_ARO-I LTR retrotransposon from the <i>Arthroderma otae</i>	3,00E-50	Gypsy-1_ARO-I	195	64

Tabela 10 – Prováveis retrotranspos e seus grupos constituintes.

Elemento	Grupos	Nº EST	Origem das ESTs				
			CA	CN	BQ	DR	EH
RtPb1 (98 EST)	1144	2	1	1			
	1145	76	12	64			
	1146	1		1			
	1147	1		1			
	1148	1	1				
	1149	1		1			
	1150	1		1			
	1151	1		1			
	1152	1		1			
	1153	1	1				
	1154	1		1			
	1155	1			1		
	1156	1		1			
	1157	1		1			
	1158	1	1				
	1160	1	1				
	1161	1		1			
	3336	5	1	3			1
RtPb2 (23 EST)	6641	2	1	1			
	9783	1		1			
	1945	20	2	16	2		
RtPb3 (3 EST)	8104	1				1	
	11330	1			1		
	12427	1			1		
RtPb4 (2 EST)	8707	1		1			
	9844	1		1			
RtPb5 (1 EST)	12801	1			1		
Total	27	127					

Grupos – representa a identidade dos grupos de EST que foram alinhados com cada elementos.
 Nº EST – quantidade de EST que compõem cada grupo CA (Felipe, Andrade *et al.*, 2003), CN (Felipe, Torres *et al.*, 2005), BQ e DR (Goldman, Dos Reis Marques *et al.*, 2003), EH (Bastos, Bailão *et al.*, 2007).

5.3 – Caracterização dos elementos móveis

A partir das análises anteriores, foram realizadas buscas (BlastN) visando localizar e realizar as anotações estruturais e funcionais de elementos íntegros (cópias inteiras apresentando regiões codificadoras) nos *supercontigs* que constituem os genomas dos fungos do complexo *Paracoccidioides*: *P. brasiliensis* (Pb03 e Pb18) e *P. lutzii* (Pb01) (<http://www.broadinstitute.org/>). No contexto de montagem genômica o termo *supercontig* se refere à sequência consenso oriunda da junção de vários *contigs* menores. Os elementos RtPb1 e RtPb2 apresentaram correspondência com aqueles estudados e validados experimentalmente por Araújo (2008) quando ainda não havia disponibilidade de sequências genômicas. Os resultados apresentados neste trabalho, referentes a esses elementos constituem análises *in silico* para a caracterização estrutural completa, determinação do número de cópias nos genomas e de quantificação da expressão, não realizado anteriormente.

5.3.1 – Caracterização estrutural do elemento RtPb1

O elemento RtPb1 (Figura 10 – anexo 5-a,b), cópia completa, foi identificado, inicialmente, no *Supercontig_1.10* do isolado Pb01 por apresentar similaridade com 18 grupos (setas alinhadas abaixo do elemento). Dentro da sequência do *supercontig*, a cópia foi localizada com as coordenadas (52418..531157). Este elemento possui 5740 pb (pares de base) e pela comparação com o banco de domínios conservados do NCBI (CDD) apresentou duas *ORFs* correspondentes ao gene *gag* (codifica proteínas que constituem a VLP – *virus-like particle*) e ao gene *pol* (codifica poliproteína que depois de processada dará origem às enzimas transcriptase reversa-Rnase H, integrase, e protease). Flanqueando as *ORFs* foram encontradas terminações longas repetidas (LTRs). A região codificadora deste elemento possui 5103pb (1700aa), está compreendida entre as coordenadas (363..5464). As LTRs possuem 246pb, estão situadas nas posições 5'(1..246) e 3'(5495..5740) e possuem os nucleotídeos TG iniciais e CA finais. A primeira *ORF* relativa ao gene *gag* (363..1550) possui 1188pb (396aa), apresenta um domínio conservado correspondente ao dedo de zinco (1404..1454) com 51pb (17aa). A segunda *ORF*, relativa à poliproteína (1551..5464), apresenta um desvio de leitura, possui

3915pb (1304aa) e apresenta os seguintes domínios conservados: Protease (1631..1927), 297pb (99aa); Transcriptase reversa (2588..3115), 528pb (176aa); Rnase H (3395..3772), 378pb (126aa); Integrase (4256..4597), 342pb (114aa); e Cromo-domínio (5162..5317), 156pb (52aa). Além disso, os sítios PBS (266.276), 11pb; e PPT (5427..5444), 18pb foram identificados.

5.3.2 – Caracterização estrutural do elemento RtPb2

A primeira cópia completa do elemento RtPb2 (5905pb) (Figura 11 – anexo 5-a,b) foi identificada no *Supercontig_1.1* do isolado Pb01, na posição (74991..80895), por apresentar similaridade com três grupos. A região codificadora deste elemento possui 5655 pb (1884aa), está compreendida entre as coordenadas (2..5681), e possui duas *ORFs*. A primeira delas (2..1717) com 1716pb (572aa) não apresentou similaridade com nenhuma proteína. Já a segunda (1743..5681) possui 3939pb, (1312aa) e compreende os domínios conservados característicos: endonuclease (1897..2492), 666pb (222aa); transcriptase reversa (3276..4076), 801pb (267aa); e Rnase H (4719..5153), 435pb (145aa). Este elemento não possui LTRs, mas um sítio correspondente à cauda poli A foi localizado (5880..5905), possui 26pb e é constituído pela sequência AAATAAAAACACACACACACACACAC.

5.3.3 – Caracterização estrutural do elemento RtPb3.

A primeira cópia completa do elemento RtPb3 (5181pb) (Figura 11 – anexo 5-a,c) que permitiu sua caracterização foi identificada no isolado Pb03. A localização desta cópia foi feita por meio do alinhamento do *Supercontig_1.2* com três grupos na região delimitada pelas coordenadas (1307933..1302753). Pela comparação com o banco de domínios conservados, o elemento apresentou *ORF* única (379..4965) de 4587pb (1528aa) com três domínios conservados: integrase (2431..2808), 378pb (126aa); transcriptase reversa (3451..4209), 759pb (253aa); Rnase H (4495..4923), 429pb (143aa). Flanqueando essa região codificadora foram encontradas as LTRs 5'(1..103) e 3'(5079..5181). Os sítios PBS(164..182) e PPT(5064..5078) também foram mapeados nesta cópia com, respectivamente, 19 e 15 pares de bases.

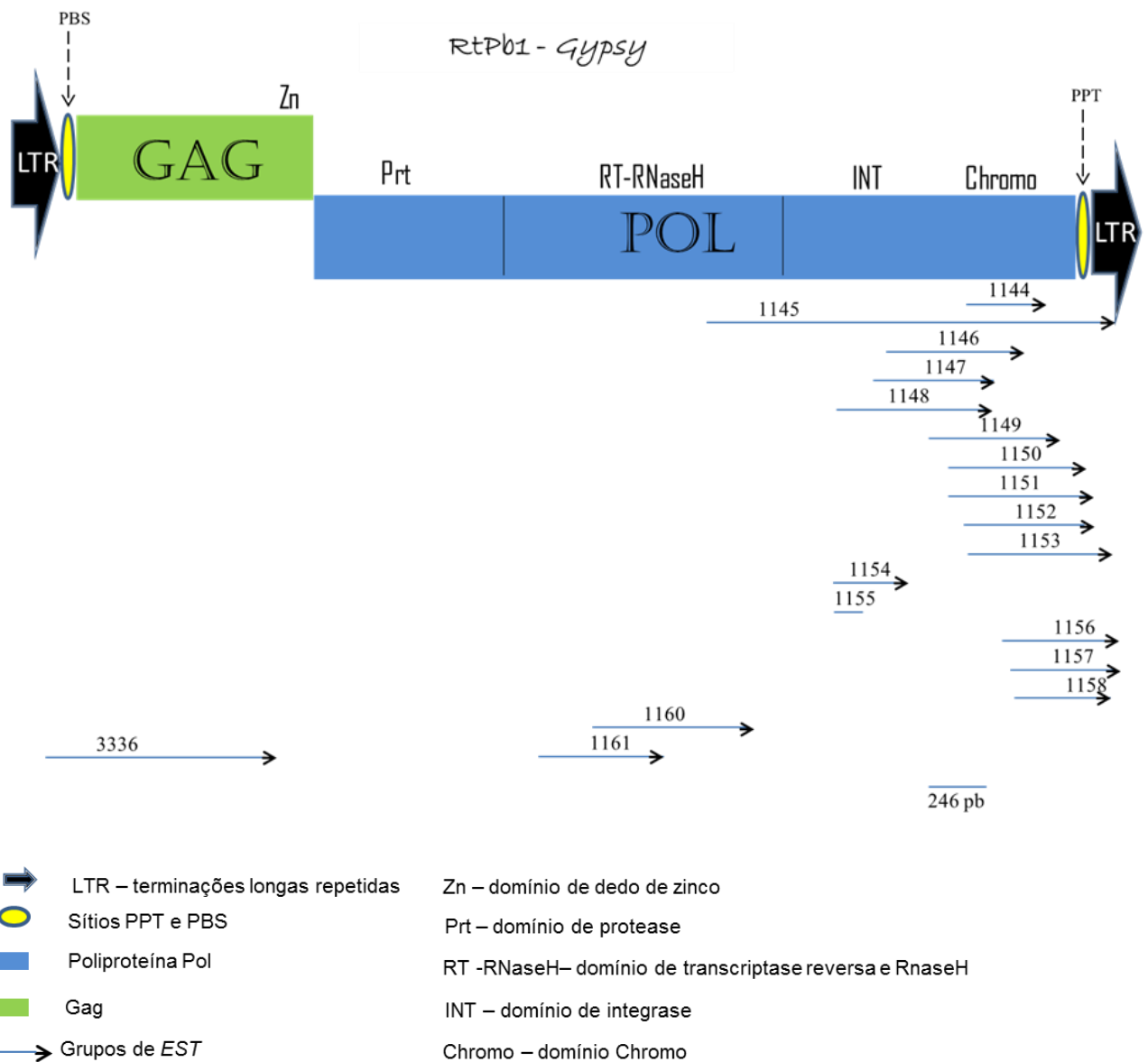


Figura 10 – Figura representativa da estrutura do retrotransposon Rtpb1.

Zinc – motivo de ligação de íons zinco; Gag – ORF relativa ao gene *gag* que codifica proteínas do VLP (*Virus-like particle*); Pol – ORF relativa ao gene *pol* que codifica poliproteína; Protease – enzima que cliva a poliproteína; Transcriptase reversa – enzima que realiza a transcrição reversa; RNaseH – domínio catalítico da transcriptase reversa que degrada o molde de RNA ; Integrase – enzima que realiza a integração do retrotransposon no genoma hospedeiro ; Chromo – domínio da integrase associado ao reconhecimento de padrões de metilação de histonas e inserção do elemento; LTR – regiões regulatórias da transcrição; PBS e PPT – Sítios de ligação dos iniciadores tRNA e trato de polipurina. O termo *Gypsy* associado ao nome do elemento (Rtpb1) refere-se à superfamília.

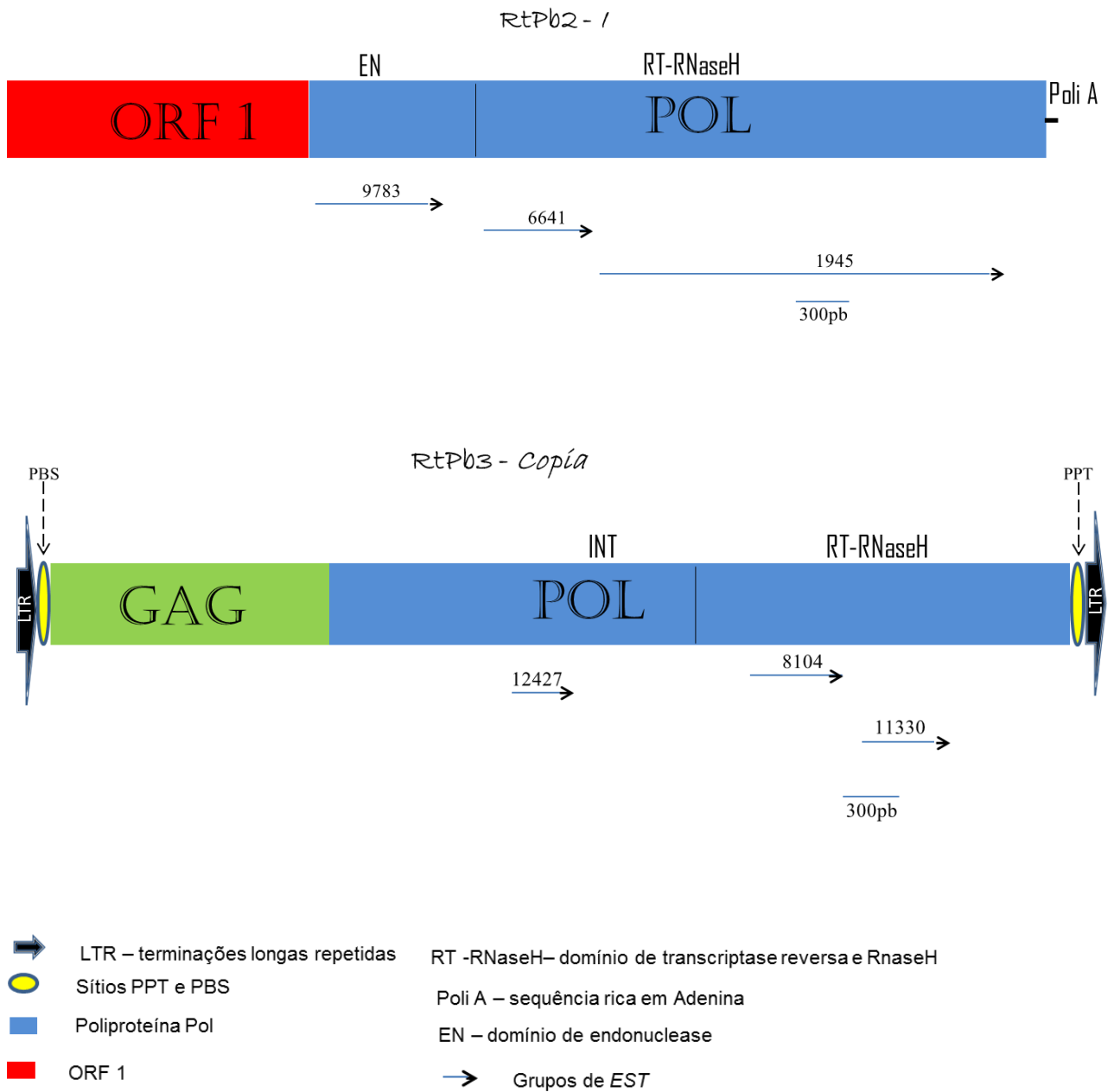


Figura 11 - Figura representativa das estruturas dos retrotransposons RtPb2 e RtPb3

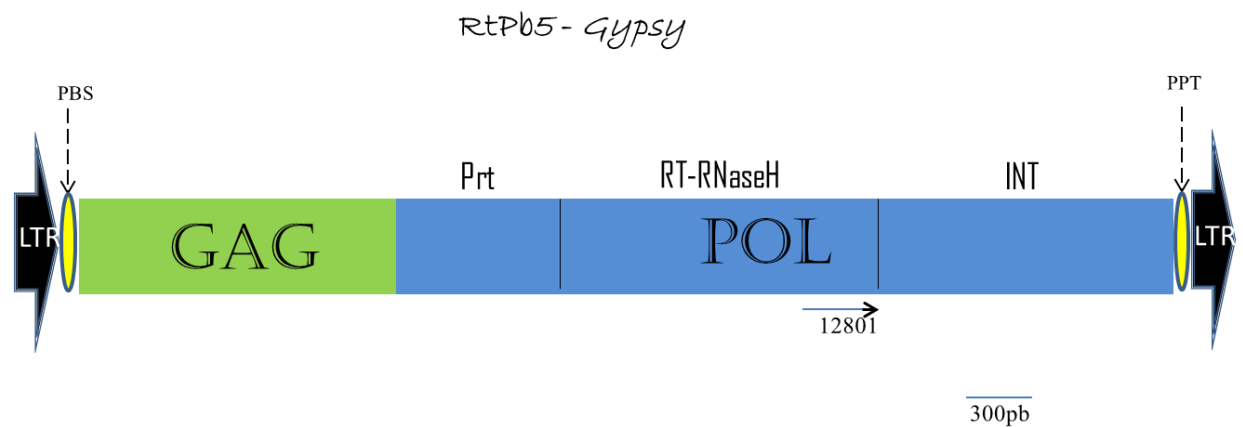
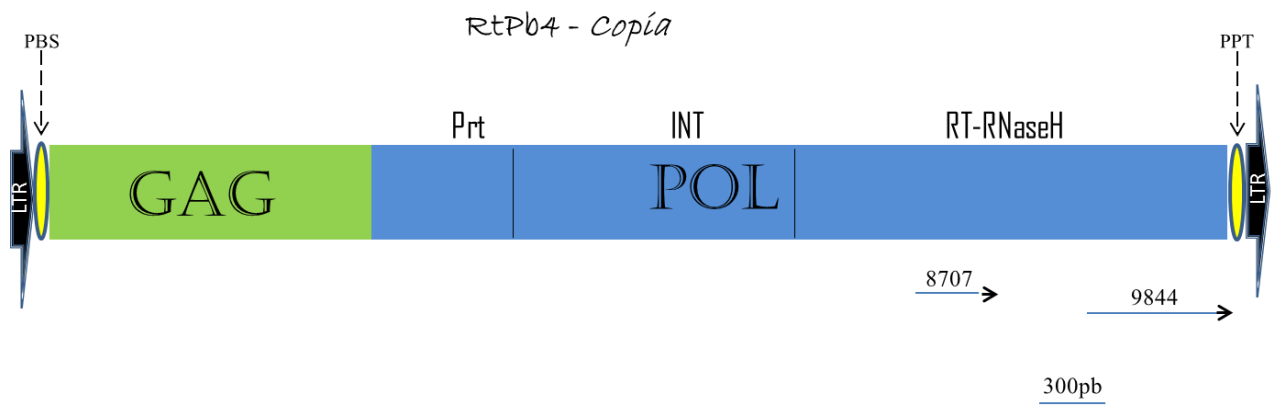
Gag – ORF relativa ao gene *gag* que codifica proteínas do VLP (*Virus-like particle*); Pol – ORF relativa ao gene *pol* que codifica poliproteína; Transcriptase reversa – enzima que realiza a transcrição reversa; RNaseH – domínio catalítico da transcriptase reversa que degrada o molde de RNA; Integrase – enzima que realiza a integração do retrotransposon no genoma hospedeiro; Endonuclease – enzima responsável pela clivagem das fitas de DNA hospedeiro no sítio de inserção; LTR – regiões regulatórias da transcrição; PBS e PPT – Sítios de ligação dos iniciadores tRNA e trato de polipurina. Os termos I e Cópia associados aos nomes dos elementos referem-se às superfamílias.

5.3.4 – Caracterização estrutural do elemento RtPb4

O elemento RtPb4 (5685pb) (Figura 12 – anexo 5-a,c), teve sua primeira cópia identificada no isolado Pb18 no *Supercontig_1.13* (81482..87166). Sua localização foi possível por meio do alinhamento com os grupos 8707 e 9844. A única ORF (403..5463) com 5061pb e (1686aa) é caracterizada por quatro domínios conservados: Gag (2431..2691), 261pb (87aa); integrase (2764..3105), 342pb (114aa); transcriptase reversa (3934..4668), 735pb (245aa) ; RnaseH (4930..5409), 480pb (160aa). As sequências das LTRs (98pb) que flanqueiam essa região codificadora foram localizadas: 5'(1..98) e 3'(5588..5685). Além disso, foi possível mapear os sítios PBS(178..192) com 15pb e PPT(5523..5542) com 20pb.

5.3.5 – Caracterização estrutural do elemento RtPb5

A caracterização da primeira cópia do elemento RtPb5 (Figura 12 – anexo 5-a,c) foi realizada, após sua localização, pelo alinhamento do grupo 12801 com o *supercontig_1.5* do isolado Pb18 (70558..75969). A cópia completa possui 5412pb de tamanho, ORF única correspondente a POL constituída por domínios característicos de protease, transcriptase reversa, Rnase H e integrase) flanqueadas por LTRs. A região codificadora deste elemento (194..4978) possui 4785pb (1594aa). Apresentou vários domínios conservados: Protease (3482..3769), 288pb (96aa); Transcriptase reversa (2504..3037), 534pb e 178aa; Rnase H (1853..2221), 369pb (123aa); Integrase (983..1321), 339pb (113aa). Suas LTRs (188pb) estão situadas nas posições 5'(1..188) e 3'(5225..5412). Possui os sítios PBS (207..221) com 19pb e PPT (5186..5204) com 15pb.



- ➡ LTR – terminações longas repetidas
- Sítios PPT e PBS
- Poliproteína Pol
- Gag
- Prt – domínio de protease
- RT -RNaseH – domínio de transcriptase reversa e RNaseH
- INT – domínio de integrase
- ➡ Grupos de EST

Figura 12 – Figura representativa das estruturas dos retrotransposons RtPb4 e RtPb5.

Gag – ORF relativa ao gene *gag* que codifica proteínas do VLP (*Virua-like particle*); Pol – ORF relativa ao gene *pol* que codifica poliproteína; Protease – Proteína que cliva a poliproteína; Transcriptase reversa – Proteína que realiza a transcrição reversa; RNaseH – domínio catalítico da transcriptase reversa que degrada o molde de RNA; Integrase – Proteína que realiza a integração da molécula do retrotransposon no genoma hospedeiro; LTR – regiões regulatórias da transcrição; PBS e PPT – Sítios de ligação dos iniciadores tRNA e trato de polipurina. Os termos *Copia* e *Gypsy* associados ao nome do elemento referem-se às superfamílias.

5.4 – Classificação dos elementos caracterizados

A classificação dos elementos (tabela 11) foi realizada com base nos resultados de alinhamentos com sequências de outros retrotransposons constantes do banco de dados do Giri (*Genetic Information Research Institute*) <http://www.girinst.org/censor/index.php> (anexo 6) e nos critérios estabelecidos por Wicker e colaboradores (2007).

Tabela 11 – Classificação dos elementos identificados

Elemento	Classificação			Identidade do Hit	
	Classe	Ordem	Superfamília	Organismo	% Similaridade
RtPb1	I (Retrotransposons)	LTR	Gypsy	<i>Aspergillus fumigatus</i>	64,9
RtPb2	I (Retrotransposons)	LINE	I	<i>Blumeria graminis</i>	63,9
RtPb3	I (Retrotransposons)	LTR	Copia	<i>Tuber melanosporum</i>	72
RtPb4	I (Retrotransposons)	LTR	Copia	<i>Blumeria graminis</i>	69,8
RtPb5	I (Retrotransposons)	LTR	Gypsy	<i>Arthroderma otae</i>	68,3

Todos os elementos estudados apresentaram similaridade com retrotransposons (Classe I). O RtPb1 foi classificado na ordem LTR, superfamília *Gypsy* e apresentou 64,9% de similaridade com um elemento identificado em *Aspergillus fumigatus* em uma região com 2842pb (2561..5403). Na ordem LINE e superfamília I foi classificado o elemento RtPb2. Este elemento apresentou, num segmento de 599pb (3135..3734), similaridade de 63,9% com um retrotransposon identificado em *Blumeria graminis*. Os elementos RtPb3 e RtPb4 foram classificados nas mesmas ordem e superfamília: LTR e *Copia*, respectivamente. O RtPb3 apresentou, em uma região de 101pb (4497..4598), 72% de similaridade com um elemento identificado em *Tuber melanosporum*. No caso do elemento RtPb4, um trecho de 82pb (5017..5099) mostrou 69,8% de similaridade com um retrotransposon de *Blumeria graminis*. O *hit* mais significativo do elemento RtPb5 foi uma região de 1482pb (1845..3327) que apresentou similaridade de 68,3% com um retroelemento identificado em *Arthroderma otae*. Este elemento foi classificado na ordem LTR e superfamília *Gypsy*.

5.5 – Relação filogenética dos retrotransposons

As sequências de aminoácidos referentes ao domínio da transcriptase reversa dos elementos RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5 foram utilizadas para alinhamento múltiplo global, implementado no programa MEGA4 como descrito na seção 4.2.5, juntamente com sequências de elementos retrotransponíveis descritos em outros organismos (Anexo 9). A análise permitiu separar as sequências em três grupos: LTRGypsy, LTR Copia e Não-LTR ou LINE (Figura 13).

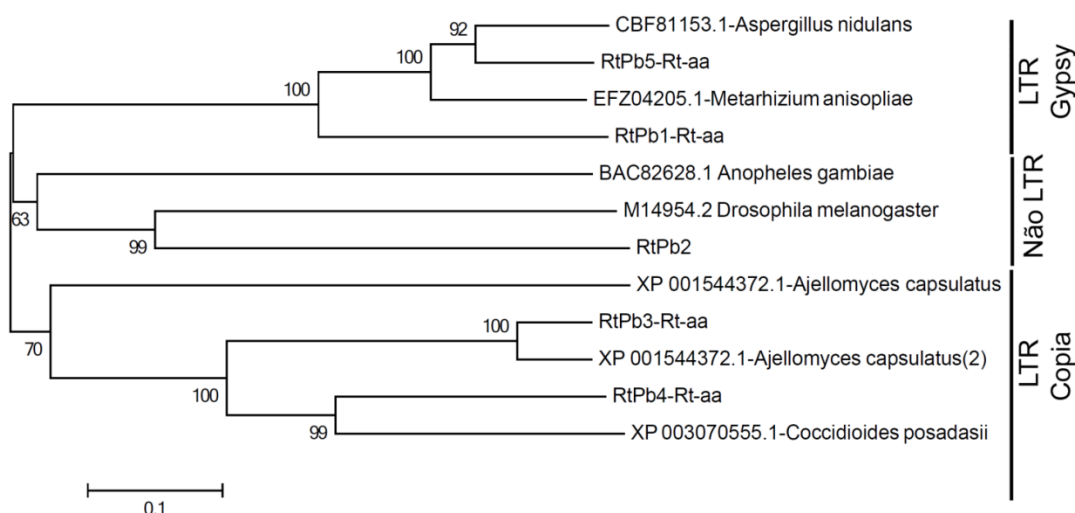


Figura 13 - Análise filogenética dos elementos RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5 utilizando o método Neighbor-joining e o modelo *p-distance*.

A sustentação dos grupos foi avaliada com 5000 réplicas de bootstrap. A barra corresponde a 0.1 substituições por sítio. Os números próximos aos nós representam os valores de bootstrap. A lista das sequências utilizadas na análise é mostrada no anexo 9. **LTR e n-LTR** representam ordens dentro de Classe I (retrotransposons). **Gypsy e Copia** representam superfamílias dentro da ordem LTR. O sistema de classificação utilizado foi proposto por (Wicker, Sabot *et al.*, 2007).

Inicialmente observou-se a formação de três grupos, sendo dois compreendendo elementos da ordem LTR (superfamílias Gypsy e Copia, respectivamente) e o outro da ordem não-LTR. O primeiro, no qual se encontram os elementos RtPb1 e RtPb5 abriga as sequências de elementos retrotransponíveis classificados na superfamília *Gypsy*. Este grupo foi confirmado com bootstrap de 100%. Observou-se que os elementos RtPb1 e RtPb5 distinguem-se em dois grupos separados. O RtPb5 apresentou-se, em 92% das análises, próximo a elementos presentes em fungos do gênero *Aspergillus*. O segundo grupo, ainda dentro da ordem LTR, foi constituído por elementos da superfamília *Copia*, confirmado em 100% das análises. Observou-se que os elementos RtPb3 e RtPb4 separaram-se em grupos distintos, embora sejam próximos estruturalmente entre si. Eles agruparam com retrotransposons descritos no gênero *Ajellomyces*, que abriga outros fungos dimórficos filogeneticamente próximos de *P. brasiliensis*. Como já era esperado, o elemento RtPb2 ficou agrupado juntamente com outros elementos da ordem não-LTR e apresentou maior similaridade com um elemento descrito *Drosophila melanogaster* em 99% das análises.

5.6 – Distribuição e mapeamento, *in silico*, dos retrotransposons

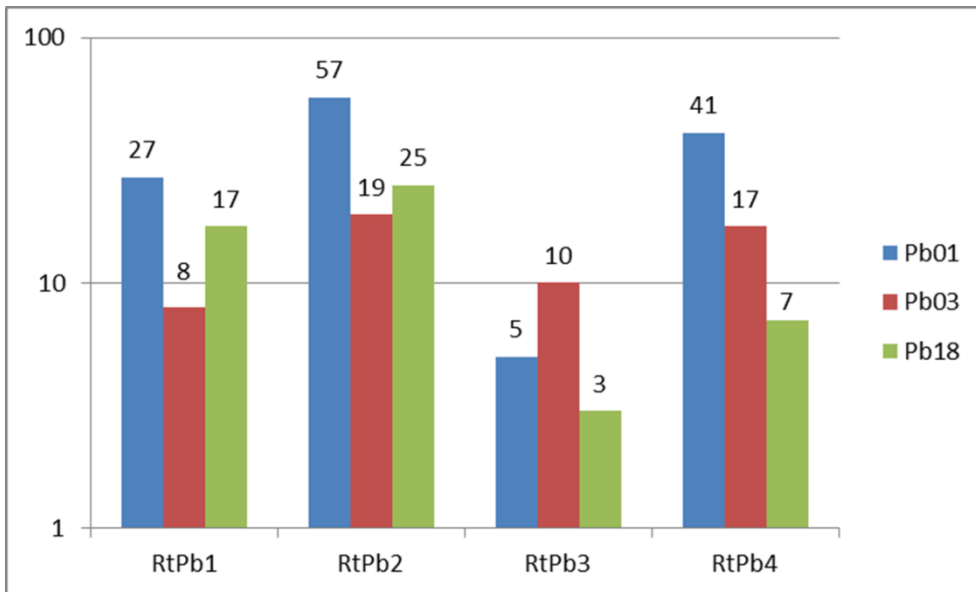
5.6.1 Número total de cópias

Após a localização e caracterização dos cinco elementos, suas sequências (Anexo 7), foram utilizadas com a finalidade de identificar e mapear outras cópias nos genomas (Anexo 8). A figura 14 (painéis A e B) e Tabela 12 mostram o número de cópias de cada elemento nos três genomas estudados. As análises foram feitas com base nos critérios (80, 80, 80,) estabelecidos na literatura. Foram consideradas as sequências que apresentaram 80% de similaridade em 80% da região codificadora e, no mínimo, 80 pares de base (Wicker, Sabot *et al.*, 2007)

No total, considerando os cinco elementos caracterizados, foram encontradas 538 cópias de retrotransposons potencialmente ativos (Gypsy-RtPb1=52; LINE-RtPb2=101; Copia-RtPb3=18; Copia-RtPb4=65, figura 14, painel A; Gypsy-RtPb5=302, figura 14, painel B) nos genomas dos três isolados. Mais da metade desse número, aproximadamente 68% (366/538) foram localizados no genoma do isolado Pb01 (*P. lutzii*). No isolado Pb18 foram encontradas 115 cópias (21,37%), enquanto no isolado Pb03 foram encontradas 57 cópias (10,59%). Determinou-se a distribuição e as coordenadas das cópias de cada um dos cinco elementos nos diferentes *supercontigs* que compõem o genoma de cada isolado (Anexo 8).

Do elemento Gypsy-RtPb1 foram encontradas 52 cópias, sendo 27 no isolado Pb01, 17 no isolado Pb18 e apenas 8 no isolado Pb03 (figura 14, painel A). O mesmo padrão de distribuição, Pb01>Pb18>Pb03, foi encontrado para o elemento LINE-RtPb2, porém, com número maior de cópias (101 cópias no total): Pb01 (57), Pb18 (25) e Pb03 (19) (figura 14, painel A). Já para o elemento Copia-RtPb3, com o menor número total de cópias (18), este perfil se alterou observando-se no isolado Pb03 um número maior de cópias (10), seguido pelos isolados Pb01 (5) e Pb18 (3) (figura 14, painel A). Em relação ao elemento Copia-RtPb4, observou-se que a maioria de suas 65 cópias foi encontrada no isolado Pb01 (41), seguido do Pb03 (17) e Pb18 (7) (figura 14, painel A).

A



B

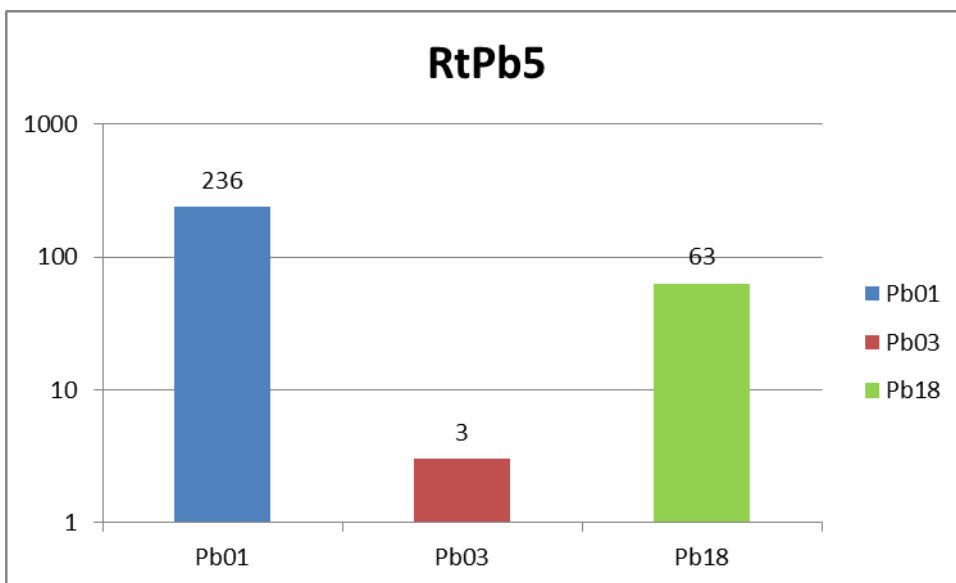


Figura 14 – Número total de cópias* dos elementos Gypsy-RtPb1, LINE-RtPb2, Copia-RtPb3, Copia-RtPb4 (painel A) e Gypsy-RtPb5 (painel B) nos isolados Pb01, Pb03 e Pb18.

*Os números sobre as barras representam as cópias encontradas por meio dos critérios estabelecidos por Wicker, Sabot et al. (2007).

Quanto ao elemento Gypsy-RtPb5, é interessante destacar que foi este o de maior representatividade no total de cópias genômicas dos retrotransposons, com 302 cópias, compreendendo cerca de 56% das cópias encontradas de todos os elementos (302/538), e apresentou ainda uma peculiaridade, 78,15% delas (236/302) estavam presentes no isolado Pb01 (*P. lutzii*). Nos isolados Pb03 e Pb18 foram encontrados, respectivamente, 3 e 63 cópias do elemento Gypsy-RtPb5.

5.6.2 Número de cópias completas e íntegras

A partir das informações acima, foram feitas análises para determinar os números de cópias completas (aquelas que possuam o mesmo tamanho da cópia caracterizada) e de cópias íntegras (aquelas que não possuam códons de terminação dentro da região codificadora) e, por contraposição, caracterizar também o número de cópias truncadas (aquelas que apresentam códon de terminação dentro da região codificadora).

A Figura 15 apresenta gráficos que indicam o número de cópias completas e íntegras dos cinco retrotransposons em estudo, identificadas nos genomas de cada um dos isolados Pb01, Pb03 e Pb18, respectivamente. O elemento Gypsy-RtPb1 apresentou, no genoma do isolado Pb01 (*P. lutzii*), 14 cópias completas das quais 13 estavam íntegras. Portanto, nesse genoma, o número de cópias truncadas foi de 14. Já no genoma do isolado Pb03 (*P. brasiliensis*), foram encontradas apenas 2 cópias completas, ambas íntegras. Já o número de cópias truncadas, nesse genoma, foi igual a 6. No genoma da Pb18 (*P. brasiliensis*) foram observadas 14 cópias completas, porém, apenas 2 estavam íntegras. Considerando o total de 17 cópias do RtPb1 presentes no genoma, 15 dessas cópias eram truncadas.

O retrotransposon RtPb5, também um elemento Gypsy, por sua vez apresentou 90 cópias completas no genoma do isolado Pb01, nenhuma das quais era íntegra. Ou seja, as 236 cópias do RtPb5 encontradas nesse genoma eram truncadas. Já no genoma do isolado Pb03, onde foram observadas apenas 3 cópias desse elemento, 2 delas sendo cópias completas, e nenhuma era íntegra, o que resultou na presença de 3 cópias truncadas desse elemento nesse genoma. No genoma do isolado Pb18, por outro lado, o elemento RtPb5,

presente em 63 cópias, tinha 29 cópias completas, e apresentou as 2 únicas cópias íntegras desse elemento (61 das 63 cópias eram truncadas).

O elemento LINE-RtPb2, apresentou 19 cópias completas no genoma do isolado Pb01, 10 das quais eram íntegras. Considerando a presença no genoma de 57 cópias desse elemento, verifica-se que 47 delas eram truncadas. Já no genoma do isolado Pb03 foram encontradas 11 cópias completas, sendo 9 íntegras. Neste isolado o número de cópias truncadas foi de 10. No genoma do isolado Pb18 o elemento RtPb2 apresentou 6 cópias completas, sendo apenas 1 íntegra. O total de cópias truncadas desse elemento nesse genoma foi de 24.

Dos elementos *Copia*, aqui identificados, denominados RtPb3 e RtPb4, o primeiro apresentou 5 cópias no genoma da Pb01, sendo 4 delas completas e íntegras, ou seja, apenas uma cópia era truncada; no genoma do isolado Pb03 foram encontradas 10 cópias, das quais 8 eram completas e também íntegras; no genoma da Pb18, foram encontradas apenas 3 cópias do RtPb3, sendo as 3 completas, porém, apenas 1 era íntegra. Portanto, nesse caso eram as cópias truncadas. Já o segundo elemento *Copia*, o RtPb4, encontrou-se no genoma da Pb01 41 cópias, sendo 32 completas, porém, nenhuma estava íntegra. No genoma do isolado Pb03, das 17 cópias encontradas, 13 eram completas, porém, apenas 1 era íntegra. Por fim, no genoma do isolado Pb18, onde foram encontradas 7 cópias do RtPb4, 3 delas eram completas mas apenas 2 eram íntegras (figura 15).

Os dados acima são sintetizados na Tabela 12 onde destacamos os percentuais relativos às cópias truncadas relativamente ao das cópias de cada elemento no genoma dos três isolados, em particular, e nos 3 genomas, como um todo. Assim, no isolado Pb01, percebe-se que apenas 43,4% das cópias dos elementos são completas e ainda menos, apenas 7,4% são cópias íntegras, registrando-se, por outro lado, elevado percentual de cópias truncadas (92,6%). Neste isolado, todas as cópias (100%) dos elementos *Copia*-RtPb4 e *Gypsy*-RtPb5 estão truncadas, bem como 82,4% das cópias do elemento LINE-RtPb2 e 51,8% das cópias do *Gypsy*-RtPb1. Já o elemento *Copia*-RtPb3, por outro lado, apresenta 80% das suas cópias completas e íntegras e apenas 20% são truncadas. No genoma do isolado Pb03 também estão truncadas todas as cópias

do elemento Gypsy-RtPb5 e 94,1% das do Copia-RtPb4. Contudo, neste isolado, o percentual de cópias truncadas do Gypsy-RtPb1 (75%) supera o das cópias truncadas do elemento LINE-RtPb2 (52,6%). Já o elemento Copia-RtPb3, registrou, no isolado Pb03, o mesmo percentual de cópias truncadas (20%) observado no isolado Pb01. No isolado Pb18, o elemento com maior percentual de cópias truncadas foi novamente o Gypsy-RtPb5 (96,8%), tendo sido o único isolado a apresentar cópias completas e integras deste elemento (2 cópias, 0,56% do total). Nesse isolado, os elementos LINE-RtPb2 e Gypsy-RtPb3 apresentaram os percentuais mais elevados de cópias truncadas entre os 3 genomas, respectivamente 96% e 88,2%. Por outro lado, foi neste isolado que Copia-RtPb4 apresentou o menor percentual de cópias truncadas entre os 3 genomas (71,4%).

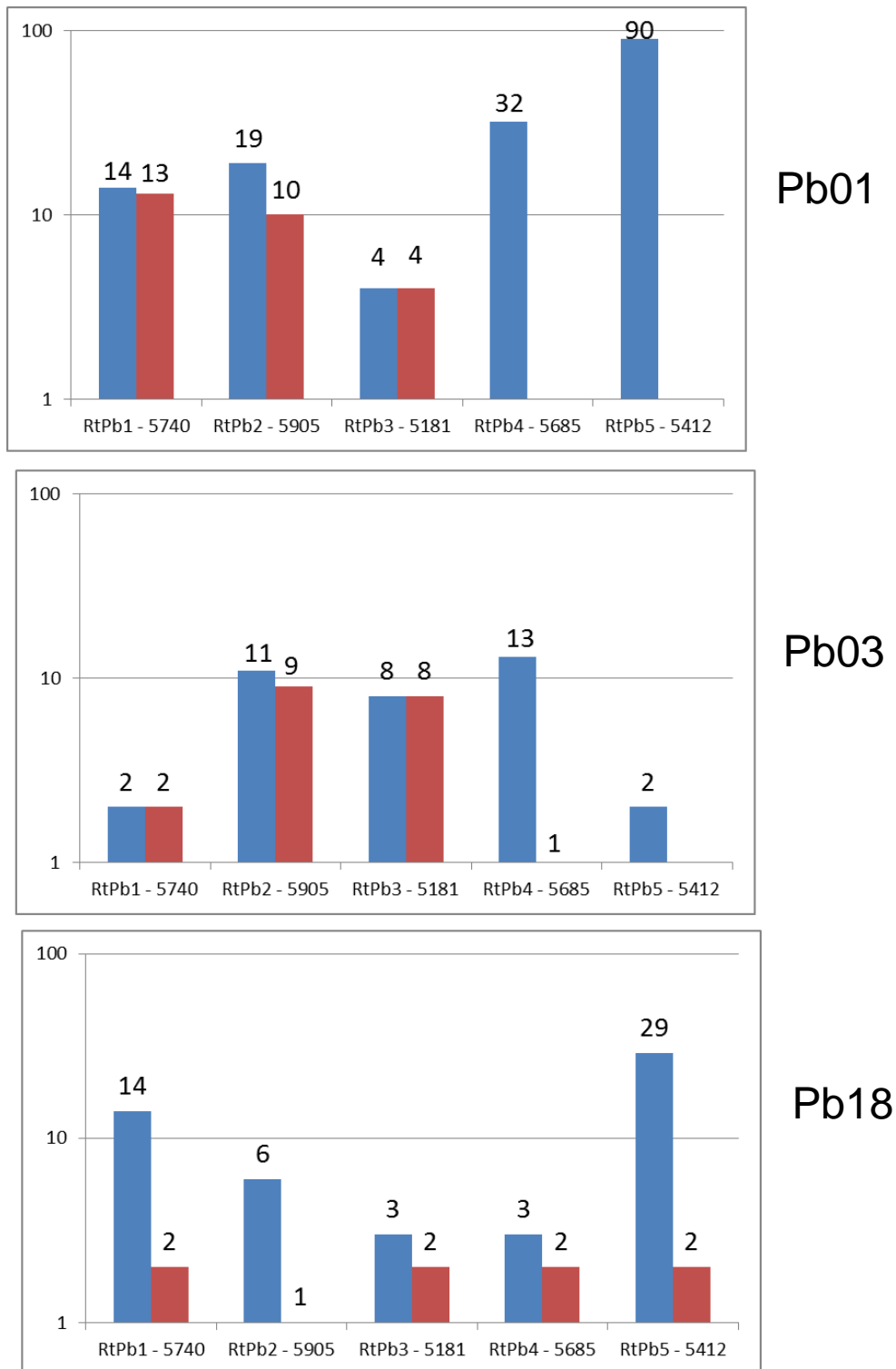


Figura 15 – Número de cópias* completas e íntegras dos elementos RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5.

Os painéis apresentam os números de cópias nos isolados Pb01, Pb03 e Pb18. Os números sobre as barras representam respectivamente os números de cópias completas (azul) e íntegras (vermelho) referentes a cada elemento. Os valores após os nomes dos elementos representam seus tamanhos em nucleotídeos.

Tabela 12 – Numero de cópias completas e de cópias truncadas dos cinco elementos (Gypsy-RtPb1, LINE-RtPb2, Copia-RtPb3, Copia- RtPb4 e Gypsy-RtPb5) encontradas nos genomas dos isolados Pb01 (*P. lutzi*), Pb03 e Pb18 (*P. brasiliensis*).

Elementos	Isolado Pb01 (<i>P. lutzi</i>)			Isolado Pb03 (<i>P. brasiliensis</i>)			Isolado Pb18 (<i>P. brasiliensis</i>)			Totais nos 3 genomas		
	Nº copias	completas/ integras	truncadas	Nº copias	completas/ integras	truncadas	Nº copias	completas/ integras	truncadas	Nº copias	completas/ integras	truncadas
Gypsy-RtPb1	27	14/13	14 (51,8%)	8	2/2	6 (75%)	17	14/2	15 (88,2%)	52	30/17(32,7%)	35 (67,3%)
LINE-RtPb2	57	19/10	47 (82,4%)	19	11/9	10 (52,6%)	25	6/1	24 (96%)	101	36/20 (19,8%)	81 (80,2%)
Copia-RTPb3	5	4/4	1 (20%)	10	8/8	2 (20%)	3	3/1	2 (66,6%)	18	15/13 (72,2%)	5 (27,7%)
Copia-RtPb4	41	32/0	41 (100%)	17	13/1	16 (94,1%)	7	3/2	5 (71,4%)	65	48/3 (4,6%)	62 (95,4%)
Gypsy-RtPb5	236	90/0	236 (100%)	03	2/0	03 (100%)	63	29/2	61 (96,8%)	302	121/2 (0,66%)	300 (99,4%)
Sub-total	366	159(43,4%) / 27 (7,4%)	339 (92,6%)	57	36(63,15%) / 20(35,08%)	37 (64,9%)	115	55(47,8%) / 8(6,95%)	107 (93,04%)	538 (100%)	250 (46,4%) / 55 (10,2%)	483 (89,7%)

O termo “completas” se refere a cópias dos retrotransposons estudados que possuem o tamanho completo (Gypsy-RtPb1= 5740 pb; LINE-RtPb2= 5905pb; Copia-RtPb3= 5181pb; Copia-RtPb4= 5685pb; Gypsy-RtPb5= 5412pb), “integras” se refere a aquelas que não apresentam códons de terminação; “truncadas” se refere a aquelas copias que tendo ou não o tamanho completo apresentam códons de terminação nas regiões codificadoras.

5.7 – Validação experimental dos elementos preditos.

5.7.1 – Amplificação por PCR dos elementos completos nos genomas dos isolados Pb01, Pb03 e Pb18.

A partir das sequências completas dos elementos, foram desenhados iniciadores (*primers*) (seção 4.3.4) para serem empregados na estratégia de *primer walking* para amplificação das sequências inteiras dos elementos (seção 4.3.16, figura 5, painel A). Os iniciadores foram desenhados de forma que a sequência completa, dividida em fragmentos, fosse amplificada e que houvesse sobreposições entre esses fragmentos. As amplificações foram realizadas inicialmente nos isolados Pb01-*like*, Pb03 e Pb18 (Figura 16).

O painel **A** da figura 16 representa o gel contendo as amplificações dos fragmentos P1(1029), P2(1090), P3(1123), P4(1265) e P5(1462) correspondentes ao elemento Copia-RtPb3. Observou-se a presença de amplificações de todos os fragmentos em todos os isolados. No painel **B** está representado o gel da amplificação dos seis fragmentos P1(986), P2(1082), P3(1060), P4(1237), P5(1250) e P6(1248) correspondentes ao elemento Copia-RtPb4. Os fragmentos P2 e P6 que apresentaram falhas na primeira reação foram amplificados em uma segunda reação. Todos os fragmentos foram amplificados em todos os isolados. Observou-se a presença de amplificações inespecíficas principalmente nos fragmentos P2, P4 e P6. Considerando a amplificação dos fragmentos P1(979pb), P2(866pb), P3(1279pb), P4(961pb) e P5(1265pb) correspondentes ao elemento RtPb5 (Painel C) foram observadas no gel apenas as bandas correspondentes ao elemento no isolado Pb18. Observou-se também neste resultado que o fragmento correspondente à transcriptase reversa de 1279pb apresentou amplificação nos três isolados estudados.

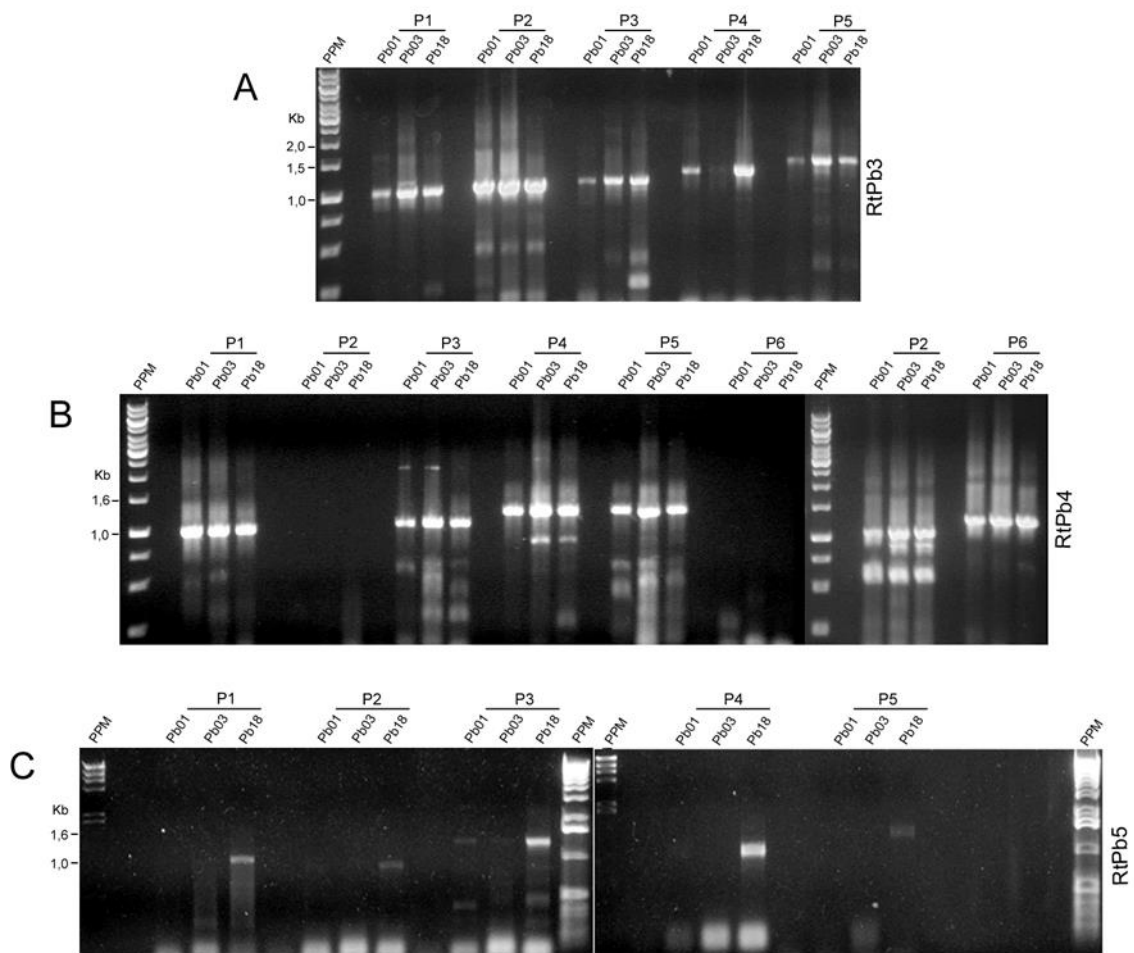


Figura 16 – Resolução eletroforética dos amplicons obtidos para Copia-RtPb3(A), Copia-RtPb4(B) e Gypsy-RtPb5(C) em diferentes isolados do complexo *Paracoccidioides*.

PPM, padrão de peso molecular (1Kb DNA ladder - Invitrogen); P1 a P6 – pares de primers utilizados (tabela 6); Pb01, Pb03 e Pb18 – isolados fúngicos.

5.7.2 – Amplificação por PCR de fragmentos específicos das transcriptases reversas dos elementos RtPb3, RtPb4 e RtPb5 nos genomas de diferentes isolados.

Tendo sido observadas experimentalmente a amplificação dos elementos RtPb3, RtPb4 e RtPb5 nos genomas dos isolados Pb01, Pb03 e Pb18, utilizou-se um trecho correspondente à transcriptase reversa (sequência mais conservada pela análise *in silico*) para amplificação em outros 31 isolados (vide figura 5 e tabela 5).

Os amplicons gerados de aproximadamente 1092pb para o elemento Copia-RTPb3, 1250pb para o Copia-RtPb4 e 1279pb para o Gypsy-RtPb5 são mostrados na figura 17. Os resultados referentes ao elemento Copia-RtPb3, evidenciaram amplificação em todos os isolados, exceto os isolados Pb01-*like* e EPM69. Com relação ao elemento Copia-RtPb4, foi observado amplificação em todos os isolados, com exceção do isolado EPM 82. No resultado referente ao elemento Gypsy-RtPb5, observou-se que em alguns isolados não houve amplificação do fragmento esperado (Pb01, EPM81, EPM82 e EPM102).

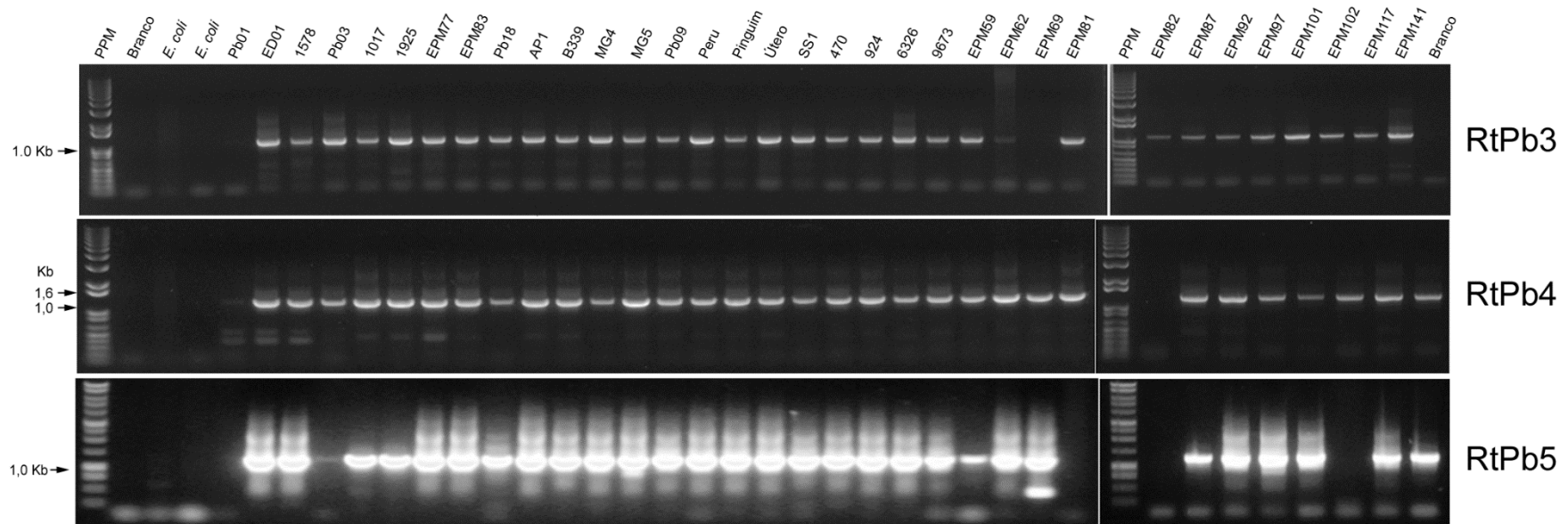


Figura 17- Identificação dos elementos RtPb3, RtPb4 e RtPb5 em diferentes isolados do complexo *Paracoccidioides*

Resolução eletroforética de amplicons de 1092pb, 1250pb e 1279pb correspondentes à regiões de transcriptase reversa dos elementos RtPb3, RtPb4 e RtPb4, respectivamente. As identidades dos isolados são mostradas em cada canaleta. PPM, padrão de peso molecular (1Kb DNA ladder - Invitrogen); controles negativos: Branco, PCR sem DNA molde; DNAd *E. coli*.

5.7.3 – Comprovação, por sequenciamento, da identidade das sequências.

Visando a confirmação da identidade dos amplicons, realizou-se o sequenciamento de uma região de aproximadamente 300pb correspondente à transcriptase reversa (figura 5, painel C) dos elementos Copia-RtPb3, Copia-RtPb4 e Gypsy-RtPb5.

As sequências consenso resultantes do sequenciamento foram então submetidas à análise de identificação através do programa ORF-FINDER (<http://www.ncbi.nlm.nih.gov/projects/gorf/>), do NCBI. A figura 18 mostra os resultados dessas análises juntamente com os alinhamentos das sequências consensos obtidas e suas respectivas predições computacionais. Os três painéis A, B e C mostram claramente o correto alinhamento das sequências dos elementos Copia-RtPb3, Copia-RtPb4 e Gypsy-RtPb5 e suas identidades, a partir de análises do ORF-FINDER. Como esperado, as sequências referentes aos elementos Copia-RtPb3 e Copia-RtPb4 apresentaram *hit* com transcriptases reversas identificadas como RVT-2, características de elementos retrotransponíveis da superfamília *Copia*, enquanto que o elemento RtPb5 apresentou *hit* com RVT-1, características de elementos da superfamília *Gypsy*. Um fato interessante foi que a sequência consenso do elemento Gypsy-RtPb5 apresentou alguns eventos de substituição. Um desses eventos, representado pelo círculo no painel C da figura 18, indica a geração de um códon de parada, levando a uma interrupção da sequência de leitura, fato que na descrição dos resultados deste trabalho foi descrito como sequências truncadas. Observou-se que a substituição de uma citosina por uma timina levou o aparecimento de um códon TGA. Este fato foi evidenciado pela barra superior ao alinhamento.



Figura 18 - Alinhamento dos amplicons referentes aos elementos Copia-RtPb3(A), Copia-RtPb4(B) e Gypsy-RtPb5(C) sequenciados com suas respectivas predições *in silico*.

A barra acima do painel A representa a ORF da sequência truncada da transcriptase reversa do elemento RtPb4. O círculo no mesmo painel mostra a mudança que resultou no surgimento de um códon de parada dessa ORF. As figuras abaixo dos alinhamentos mostram a identidade das sequências alinhadas por meio da comparação com o banco de dados de domínios conservados do NCBI.

5.7.4 – Mapeamento dos elementos retrotransponíveis por *Southern blot*.

Como abordagem complementar para validação experimental dos elementos preditos *in silico*, utilizou-se a metodologia de *Southern blot* (vide figura 5, painel D). Nesses experimentos foram utilizadas sondas correspondentes às sequências parciais dos elementos que tiveram suas identidades comprovadas por sequenciamento. As membranas de *nylon* contendo os fragmentos de DNA imobilizados foram, então, hibridizadas com essas sondas. As figuras 19, 20 e 21 mostram os resultados para cada elemento separadamente, utilizando DNA genômico de quatro isolados (Pb01, Pb03, Pb18 e EPM83). É interessante destacar aqui que cada uma dessas amostras do fungo pertence a uma das 4 linhagens filogenéticas conhecidas, descritas na literatura, respectivamente, Pb01-*like* ou *P. lutzii*, PS2, S1 e PS3 (estas 3 em *P. brasiliensis*). As moléculas de DNA foram digeridas com diferentes enzimas de restrição (Seção 4.3.14) de forma que as enzimas utilizadas para cada elemento pudessem: a) clivar a sequência em apenas um sítio e b) clivar o DNA em regiões fora da sequência de interesse.

No caso do elemento Copia-RtPb3 as enzimas utilizadas foram *EcoRI* e *HindIII*. A primeira não apresentou sítio de restrição na sequência e a segunda apresentou dois sítios nas posições 2068 e 4374, gerando fragmentos em torno de 2 Kb. A sonda correspondente ao elemento Copia-RtPb3 foi sintetizada utilizando a sequência interna ao fragmento P2 (1092pb) possuindo tamanho de aproximadamente 329pb. Os sinais gerados (Figura 19) foram compatíveis com o tamanho dos fragmentos esperados (em torno de 2 Kb). Apenas o isolado Pb01 não apresentou sinais de hibridação. O isolado Pb03 apresentou maior número de fragmentos com tamanhos superiores a 2Kb. O perfil de restrição e hibridação gerado pela enzima *EcoRI* sugere a presença de múltiplas cópias do elemento neste genoma indicando também a provável presença de cópias inteiras nestes fragmentos. O perfil obtido com a enzima *Hind III* apresenta fragmentos com tamanhos esperados em torno de 2-3Kb indicando a clivagem da sequência em fragmentos esperados.

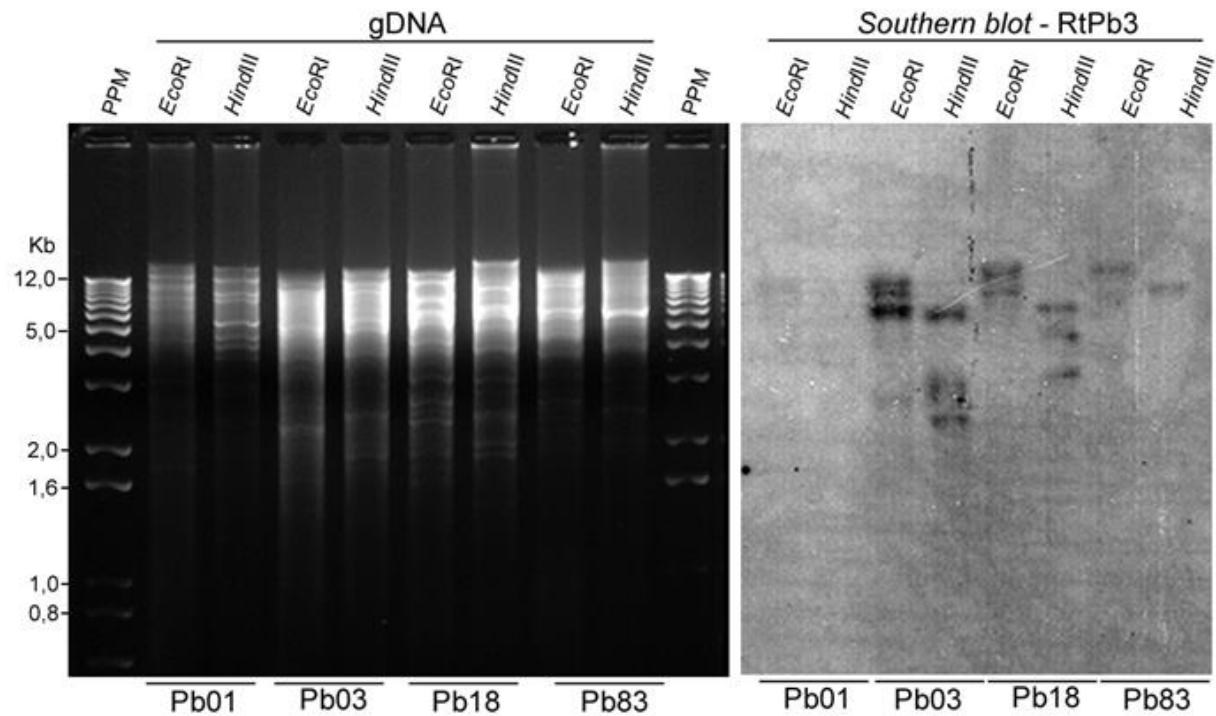


Figura 19 - Determinação do número de cópias genômicas do elemento Copia-RtPB3 nos isolados Pb01, Pb03 e Pb18 do complexo *Paracoccidioides*.

Painel da esquerda, resolução eletroforética do DNA genômico após restrição com as endonuclease *EcoRI* ou *HindIII*. Painel da direita, *Southern blot* do DNA genômico após restrição utilizando sonda radioativa para o elemento RtPb3. PPM, padrão de peso molecular (1Kb Ladder – Invitrogen).

Os perfis obtidos com o isolado Pb18 apresentaram sinais em fragmentos de tamanhos diferentes quando comparados com o isolado anterior, Pb03.

Nota-se a presença de sinais em fragmentos com tamanhos superiores a 12 Kb contendo prováveis cópias inteiras do elemento. Três sinais com tamanhos superiores a 2Kb foram observados na canaleta 6, indicando possivelmente a presença de sítios de restrições alternativos nas sequências deste isolado. Quando se observa os perfis de hibridação nos isolado EPM83 pode-se perceber que os sinais diferiram dos outros anteriores pela presença de apenas um sinal em cada canaleta. Ambos os sinais apontam fragmentos com tamanhos superiores a 10Kb, indicando provavelmente a presença de cópias inteiras e também prováveis alterações nos sítios de restrição correspondentes à enzima *HindIII*, responsável por clivar a sequência do elemento em fragmentos menores.

As análises envolvendo o elemento Copia-RtPb4 foram feitas a partir de digestão do DNA genômico com as enzimas *NcoI* e *HindIII* usadas pelo fato de não clivarem e clivarem em mais de um ponto dentro da sequência respectivamente. A enzima *HindIII* apresentou três sítios de restrição dentro da sequência nas posições 347, 3260 e 4794. Os fragmentos gerados pela clivagem desta enzima produzem fragmentos entre 1-3Kb. A sonda para esse elemento (375pb) foi sintetizada a partir da sequência interna do fragmento P5 (1250pb). Os resultados gerados com essa sonda (Figura 20) mostram perfis semelhantes entre os isolados, exceto Pb01. Neste isolado, foram observados sinais fracos de hibridação com dois fragmentos, quando comparados com os outros isolados. Nos isolados Pb03, Pb18 e EPM83 foram observados sinais semelhantes com fragmentos de aproximadamente 2Kb sempre que sequências foram digeridas com a enzima *HindIII*. Fragmentos com tamanhos a partir de 12Kb também foram observados nos três isolados quando a digestão foi realizada com a enzima *NcoI*. Outro fato observado foi a presença de um sinal, presente apenas no isolado Pb18. Este mesmo sinal não foi observado nos outros isolados.

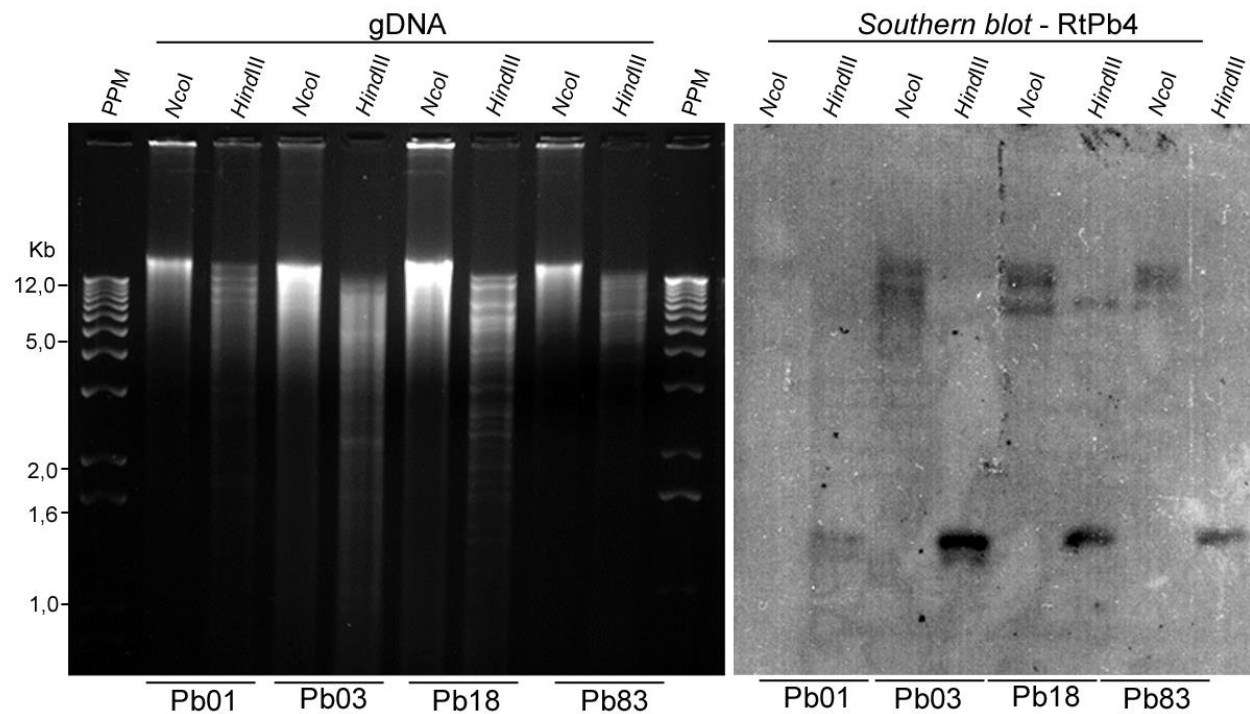


Figura 20 - Determinação do número de cópias genômicas do elemento *Copia-RtPB4* nos isolados Pb01, Pb03 e Pb18 do complexo *Paracoccidioides*.

Painel da esquerda, resolução eletroforética do DNA genômico após restrição com as endonuclease *NcoI* ou *HindIII*. Painel da direita, *Southern blot* do DNA genômico após restrição utilizando sonda radioativa para o elemento RtPb4. PPM, padrão de peso molecular (1Kb Ladder – Invitrogen).

Para as análises referentes ao elemento Gypsy-RtPb5, os DNAs genômicos foram digeridos com as enzimas *SalI* e *EcoRI*. A primeira foi escolhida por não possuir sítio de corte dentro da sequência do elemento. Já a *EcoRI* foi escolhida por clivar a sequência na posição 2998. A sonda para esse elemento (321pb) foi sintetizada a partir da sequência interna do fragmento P5 (1279pb).

Quando as análises foram feitas na tentativa de localizar os elementos RtPb5 observou-se a presença de poucas cópias nos genomas estudados (Figura 21). Os resultados apontaram a presença de sinais apenas nos isolados Pb01, Pb18 e EPM83. O isolado Pb03 não apresentou sinal de hibridação com nenhuma das enzimas utilizadas. Os outros isolados apresentaram sinais de hibridação semelhantes, destacando-se a hibridação em um fragmento com tamanho acima de 12Kb, sugerindo a presença do elemento inteiro. No isolado Pb18 observou-se a presença de sinais em fragmentos com tamanho superiores a 5Kb, indicando provavelmente a presença de outras cópias no genoma. Este fato foi observado também no isolado EPM83 que além de apresentar sinal no fragmento de aproximadamente 12Kb, apresentou em fragmentos menores, indicando também presença de múltiplas cópias neste isolado.

Um fato interessante que foi observado com essas análises foi que em todas as buscas foi possível verificar a presença de cópias dos elementos estudados nos genomas dos isolados, apresentando sinais mais fortes naqueles isolados que previamente tinham apresentado cópias inteiras em maior número.

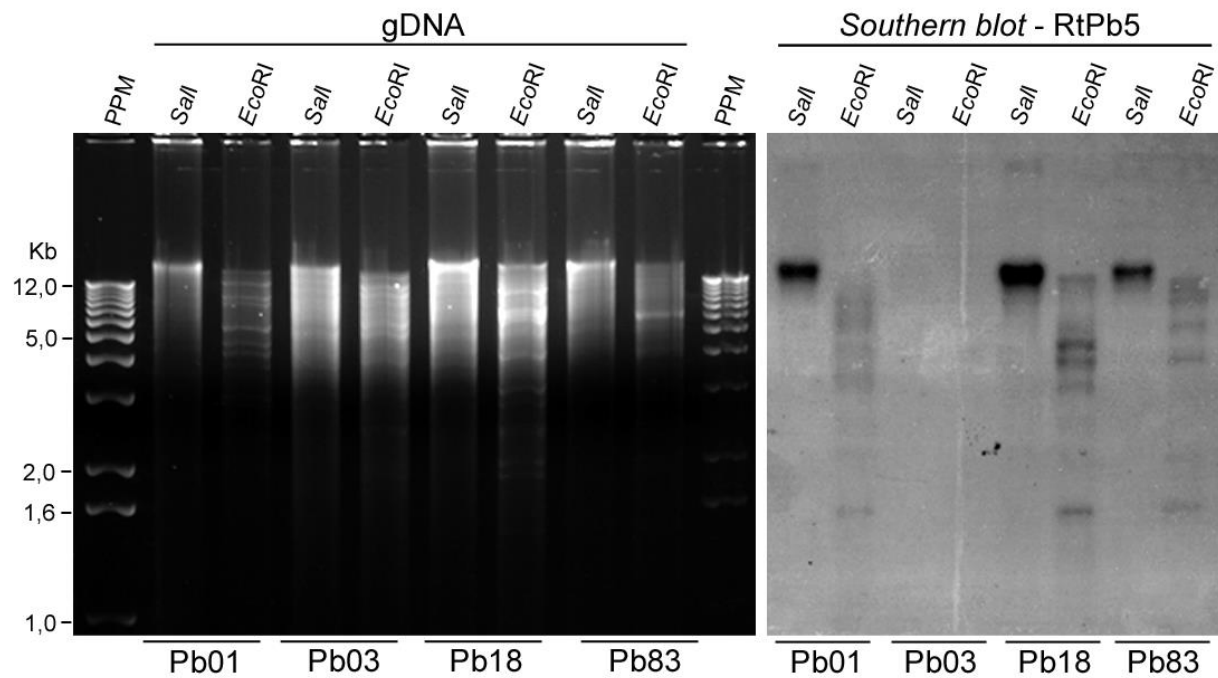


Figura 21 - **Determinação do número de cópias genômicas do elemento Gypsy-RtPB5 nos isolados Pb01, Pb03 e Pb18 do complexo *Paracoccidioides*.**

Painel da esquerda, resolução eletroforética do DNA genômico após restrição com as endonuclease *SalI* ou *EcoRI*. Painel da direita, *Southern blot* do DNA genômico após restrição utilizando sonda radioativa para o elemento RtPb5. PPM, padrão de peso molecular (1Kb Ladder – Invitrogen).

5.8 – Análise dos padrões de transcrição dos cinco elementos retrotransponíveis por RT-PCR.

Os elementos caracterizados foram estudados quanto ao seu padrão de transcrição nos genomas dos isolados Pb01 (*P. lutzi*), Pb03(PS2), Pb18(S1) por meio da técnica de RT-PCR (Ítem 4.3.8). As análises demonstram que os cinco elementos retrotransponíveis, objetos de estudo deste trabalho, e os genes controles (β -tubulina e gp43) são transcritos na fase de levedura desses três isolados (Figura 22).

Considerando o elemento Gypsy-RtPb1, observou-se que o amplicom de maior intensidade de sinal entre os isolados apareceu no isolado Pb01 e que os isolados Pb03 e Pb18 apresentaram intensidades semelhantes de sinal. O mesmo fato aconteceu para o elemento LINE-RtPb2. Os sinais de Copia-RtPb3 apresentam-se de forma homogênea nos três isolados. O elemento Copia-RtPb4 apresentou sinais mais fortes nos isolados Pb03 e Pb18 quanto comparados com Pb01. É interessante desatacar que os sinais de amplificação do elemento Gypsy-RtPb5 foram notados apenas no isolado Pb18. Isso aconteceu em todas as reações com variação do volume de cDNA entre 1 - 4 μ L (Ítem 4.3.4). Os amplicons referentes ao gene constitutivo β -tubulina apresentaram sinais de mesma intensidade nos três isolados. Para o outro gene controle (gp43) observou-se que sinais de amplificação apresentaram intensidades diferentes, sendo mais intenso nos isolados Pb03 e Pb18, quando comparados com o isolado Pb01.

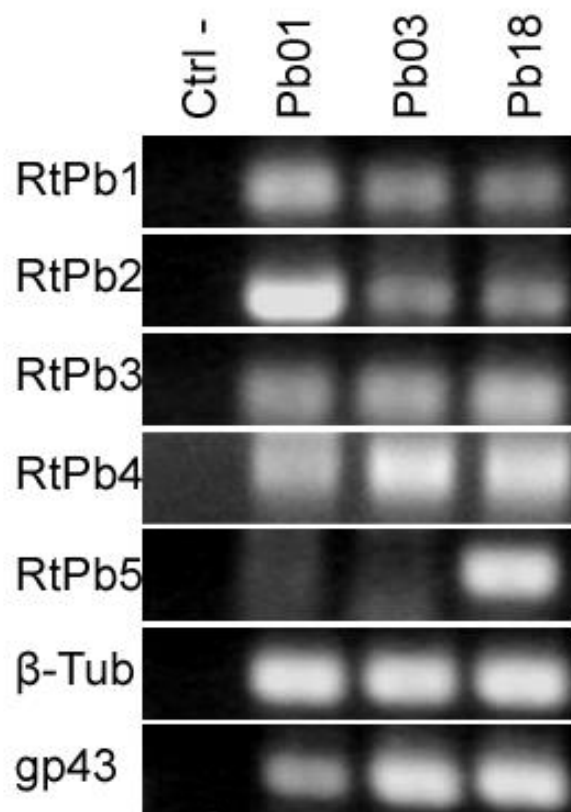


Figura 22 – Análise da transcrição dos elementos retrotransponíveis Gypsy-RtPb1, LINE-RtPb2, Copia-RtPb3, Copia-RtPb4, Gypsy-RtPb5 e dos genes controles β -tubulina e gp43.

Resolução eletroforética, em gel de agarose, de amplicons obtidos a partir de cDNA correspondentes à regiões de transcriptase reversa dos elementos retrotransponíveis RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5 e dos genes controles β -tubulina e gp43 (glicoproteína de 43 KDa). As identidades dos isolados são mostradas em cada canaleta. Ctrl(-) – controle negativo da reação (RNA total).

5.9 – Quantificação da transcrição dos elementos retrotransponíveis por qRT-PCR.

A quantificação da transcrição dos retrotransposons caracterizados foi realizada por meio da técnica PCR quantitativo em tempo real (qRT-PCR) (Ítem 4.3.9). Nestas análises foram utilizados os isolados: Pb01 (*P. lutzi*), Pb03 (PS2) e Pb18 (S1) (*P. brasiliensis*). Na padronização de cada par de iniciadores foram obtidas curvas de dissociação indicando a amplificação específica do respectivo gene alvo (Anexo 10). Constatou-se que a concentração ideal dos *primers* e a diluição ideal de cDNA, mediante a observação das curvas-padrão, foi de 100 nM e 10^{-1} , respectivamente (Anexo 11). A figura 23 apresenta os dados de quantificação referentes aos retrotransposons Gypsy-RtPb1, LINE-RtPb2, Copia-RtPb3, Copia-RtPb4 e Gypsy-RtPb5.

Os resultados da expressão (transcrição) dos cinco elementos são apresentados na figura 23. Dentre os elementos estudados, observou-se que o Gypsy-RtPb1 foi o mais expresso, aproximadamente (~11 UR) no isolado Pb01. A expressão deste elemento foi observada também nos isolados Pb03 (~5 UR) e Pb18 (~3 UR). O elemento LINE-RtPb2 também apresentou sinais de expressão nos três isolados Pb01 (~4 UR) e Pb03 (~3 UR) e Pb18 (<1).

O elemento Copia-RtPb3 apresentou sinais de expressão em todos os isolados e teve sua maior expressão em Pb03 (~8 UR) e Pb18 (~2 UR), enquanto que no isolado Pb01 a sua taxa de expressão apresentou valores menores que 0,5 UR. O Copia-RtPb4 apresentou em todos os isolados uma baixa quantidade de transcrito, com valores menores que 0,8 UR: Pb01 (0,000033 UR); Pb03 (0,76 UR) e Pb18 (0,28 UR). Por fim, as análises do elemento Gypsy-RtPb5 permitiram observar que este foi o elemento com a menor taxa de expressão, com valores em torno de 0,2 ou menores que 0,4 UR, sendo que o maior valor foi encontrado no isolado Pb18 (0,44 UR), seguido por Pb01 (0,0025 UR) e Pb03 (0,000009 UR).

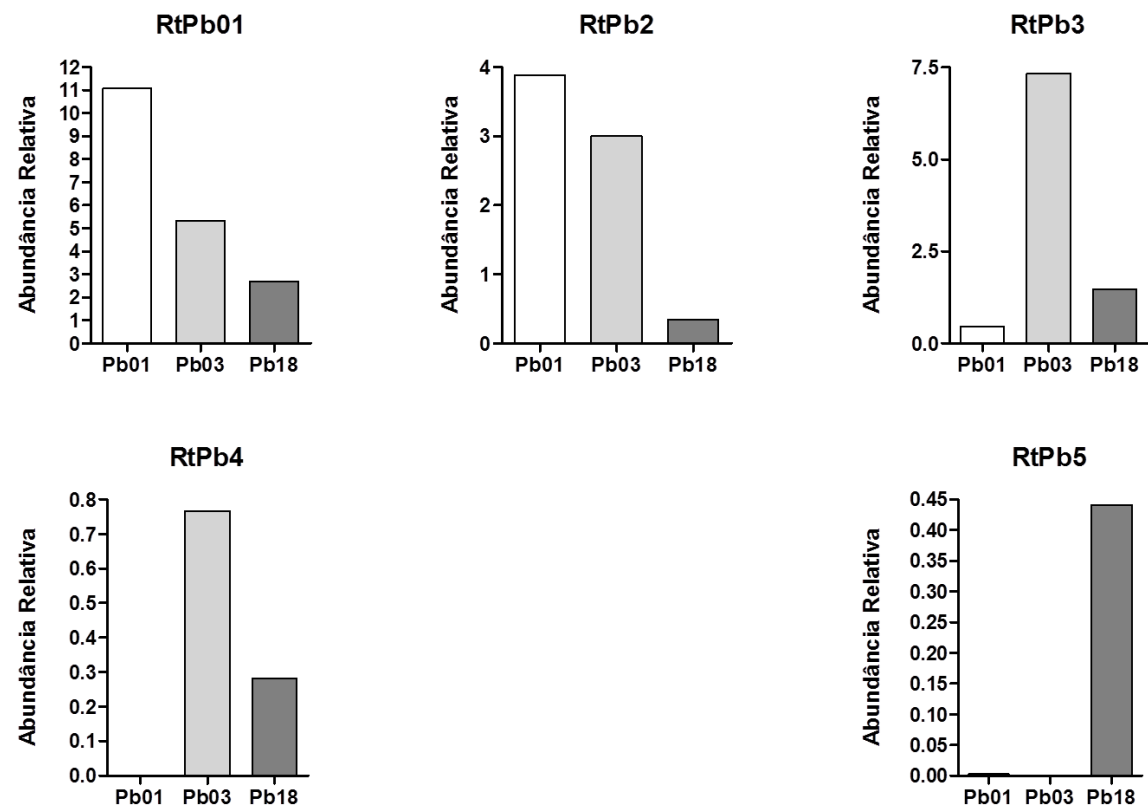


Figura 23 - Quantificação relativa do número de transcritos dos elementos RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5

Os valores de expressão de cada elemento foram obtidos por meio da normalização (valores de C_T) com dois genes de referência: β -tubulina e L35. A escala no eixo Y representa a abundância relativa dos transcritos, determinada pelo método C_T comparativo - Valores C_T plotados contra o logaritmo do número de cópias de cDNA ($2^{-\Delta C_T}$). Pb01, Pb03 e Pb18 - isolados de *Paracoccidioides*.

6 – DISCUSSÃO

Procurou-se no presente trabalho, identificar e caracterizar retrotransposons potencialmente ativos no genoma de fungos dimórficos do complexo *Paracoccidioides*, realizar a anotação estrutural e funcional de suas sequências e avaliar a sua transcrição (expressão).

O interesse particular em estudar e conhecer melhor os elementos genéticos móveis, componentes dos genomas cuja importância biológica crescente foi indicada em capítulo anterior, decorreu de resultados de trabalhos que, visando correlacionar perfis cariotípicos distintos de *P. brasiliensis*, obtidos por PFGE, empregou-se o mapeamento cromossômico a partir de um painel de marcadores moleculares, nove genes previamente caracterizados, como sondas em “Southern blot”. Revelaram-se perfis cariotípicos com 4 ou 5 bandas cromossômicas cujos tamanhos variaram entre 2,5 e 9,5 Mb. Após análise dos padrões de hibridização foram perceptíveis sete perfis distintos entre os 12 isolados em estudo. Os resultados, além de demonstrarem a ocorrência de polimorfismo de tamanho e número de bandas cromossômicas em *Paracoccidioides* indicaram a existência de variações genéticas intraespecíficas em isolados naturais e evidenciaram expressivos rearranjos cromossômicos, sugestivos da ocorrência de translocações, deleções, inversões. É também interessante mencionar que a utilização do DNA correspondente a uma das bandas cromossômicas foi removido do gel e marcado para utilização como sonda, obtendo-se um padrão de sinais que indicou a existência de sequências repetitivas no genoma do fungo (Feitosa, Cisalpino *et al.*, 2003).

A detecção de polimorfismos é um evento comum em fungos, revela a fluidez na organização cromossômica de eucariotos e indica a ocorrência de rearranjos de sequências envolvendo recombinação, translocações, inversões e deleções de segmentos. Transposons e retrotransposons, descritos entre as sequências de média repetitividade nos genomas eucariotos, foram caracterizados em diversos fungos. Entre os fungos patogênicos como *Candida albicans*, *Cryptococcus neoformans* e *Aspergillus fumigatus*, numerosas famílias de TE vem sendo reportadas. Numerosos exemplos de rearranjos cromossômicos mediados por TE têm sido relatados em drosófilas, leveduras e humanos. Em S.

cerevisiae retrotransposons Ty foram encontrados associados às regiões limítrofes de deleções, duplicações, inversões ou translocações cromossômicas. Elementos móveis podem produzir uma série de rearranjos por meio de recombinação ectópica entre elementos dispersos, localizados no mesmo ou em diferentes cromossomos. A instabilidade do cariótipo tem sido investigada em algumas espécies de fungos, e sua análise extensiva tem evidenciado a presença de translocações, deleções, e duplicações de largos fragmentos cromossômicos associados à presença de numerosas famílias de TE (Zou, Kim *et al.*, 1996; Hayes, 2003).

Portanto, há indicações de que a instabilidade do cariótipo observada em *Paracoccidioides* possa estar associada, a exemplo de outros fungos, com a presença no genoma de elementos genéticos móveis. Esta percepção coincidiu com a publicação das primeiras análises do transcriptoma do fungo. Em 2003, Felipe e colaboradores apresentaram a descrição do transcriptoma parcial do fungo, relatando a presença de elementos Ty-like, retrotransposons, entre as 18 sequências majoritariamente expressas (Felipe, Andrade *et al.*, 2003; Goldman, Dos Reis Marques *et al.*, 2003).

No presente trabalho, para atender ao objetivo de identificar e caracterizar retrotransposons potencialmente ativos no genoma de *Paracoccidioides* spp., todas as sequências de EST disponíveis em banco de dados público, dbEST (<http://www.ncbi.nlm.nih.gov/dbEST/>), foram obtidas por acesso computacional procedendo-se à “clusterização” de dados e anotação funcional. As EST de *Paracoccidioides* foram armazenadas num banco de dados local, tratadas com o objetivo de retirar a redundância e aumentar a qualidade (programa CAP3 que compara as sequências entre si, agrupando-as de acordo com os parâmetros estabelecidos). Felipe e colaboradores (2003) relataram a rapidez e eficiência do programa na análise de geração de grupos de sequências.

Os depósitos de EST de *Paracoccidioides* realizados no dbEST por diversos grupos de pesquisadores a partir de 2003, e sucessivamente, acumularam um número elevado, aproximadamente 41.500 sequências (Felipe, Andrade *et al.*, 2003; Goldman, Dos Reis Marques *et al.*, 2003; Felipe, Andrade *et al.*, 2005; Bastos, Bailao *et al.*, 2007; Costa, Borges *et al.*, 2007). No presente

trabalho, essas ~41.500 EST formaram 12922 grupos (31% de agrupamento), distribuídas de forma heterogênea entre singlets e contigs. Isso se deve, provavelmente ao fato de este conjunto de EST, armazenadas no banco local, conter representantes oriundos de trabalhos realizados por diferentes autores, obtidas em diferentes situações de estimulação do fungo e de diferentes isolados, inclusive expressas nas formas de micélio e/ou levedura (Felipe, Andrade et al., 2003; Goldman, Dos Reis Marques et al., 2003; Costa, Borges et al., 2007).

O mecanismo de retrotransposição baseia-se na existência de uma molécula de RNA como intermediário de um processo que tem seu primeiro passo, no seu ciclo replicativo (e de mobilidade), dependente do processo de transcrição (Wicker, Sabot *et al.*, 2007; Finnegan, 2012), sendo evidenciados entre as sequências expressas. De fato, a primeira abordagem de estudo dos elementos móveis de *Paracoccidioides* foi realizada por Araújo e colaboradores (2008), num período em que não havia disponibilidade de dados genômicos do fungo. Baseou-se no levantamento dos transcritos existentes no Banco de Dados do Projeto Genoma Funcional e Diferencial de *P. brasiliensis* (PbDBEST) (www.biomol.unb.br/Pb), elaborado e anotado pelos pesquisadores dos grupos ligados à Rede Genoma Centro-Oeste/ CNPq - Ministério da Ciência e Tecnologia, cujo acesso foi possível sob condições de sigilo e senha fornecida pelo grupo responsável, cuja anotação trazia registros de similaridades como sequências de genes característicos de TE (busca léxica). Araújo (2008) realizou a caracterização preliminar de dois dos retrotransposons estudados no presente trabalho (Gypsy-RtPb1 e LINE-RtPb2). É interessante mencionar que no banco PbDBEST estavam disponíveis, naquele momento, anotações relativas a 6022 grupos de EST, menos da metade do que se encontrou aqui após a otimização do banco local.

A abordagem seguinte no estudo dos TE em *Paracoccidioides* foi realizada por Marini e colaboradores, e a estratégia inicial foi também a de pesquisar transcritos compatíveis com transposons de DNA, partindo do mesmo banco de dados PbDBEST. Encontrou-se apenas duas sequências: contigs 530 (673 pb) e 1938 (637 pb) (números de acesso GenBank CN240498 e CN247880, respectivamente), anotados como putativas transposases. De posse de acesso

autorizado aos primeiros *drafts* do genoma das três linhagens de *Paracoccidioides*, então em andamento, utilizaram-se essas duas sequências em buscas por similaridade contra o banco de dados do projeto FGI (*Fungal genome initiative*), *Broad Institute of MIT and Harvard*, baixadas do “*Paracoccidioides brasiliensis Database*” (http://www.broad.mit.edu/annotation/genome/paracoccidioides_brasiliensis/Downloads.html) e utilizadas nas análises *in silico*. Cabe comentar que a partir desse diminuto número (2) de grupos de busca um grande número de *hits* com similaridade a transposons de DNA foi encontrado nos três isolados. Uma análise destes *hits* permitiu a identificação inequívoca de oito novas famílias de transposons de DNA, denominados Trem (*transposable elements mariner*) A-H, todos pertencentes à superfamília *Tc1/mariner*. Nos genomas das 3 linhagens registrou-se a ocorrência de 484 cópias desses elementos nos genomas (Marini, Zanforlin *et al.*, 2010).

No presente trabalho, após se realizar a otimização do banco local de EST, buscou-se encontrar, entre os 12922 seus grupos componentes quais eram os que apresentavam similaridade com retrotransposons. As buscas por similaridade empregando o algoritmo BLAST baseou-se em comparações com o banco de proteínas não redundantes do NCBI (NR) e bancos de dados específicos de elementos repetitivos e/ou móveis (RepBase, TEfam). Por se tratar de bancos de dados especializados, cujas sequências passaram por processo de curadoria manual, a utilização dos dados dos bancos RepBase e TEfam tem sido relatada em diferentes trabalhos envolvendo elementos transponíveis (Desjardins, Champion *et al.*, 2011; Fernandez-Medina, Struchiner *et al.*, 2011; Muszewska, Hoffman-Sommer *et al.*, 2011). Nestas buscas diferentes *e-value* têm sido utilizados no processo de anotação de elementos móveis: mais estrigente (10^{-15}) (Fernandez-Medina, Struchiner *et al.*, 2011) ou menos estrigente (10^{-4}) (Shankar, Wu *et al.*, 2011). Neste trabalho, empregou-se um *e-value* (10^{-6}) que pode ser considerado rigoroso nestas buscas, pois valores menores tendem a ser menos significantes (Goldman, Dos Reis Marques *et al.*, 2003).

As abordagens adotadas selecionaram, respectivamente, 142 grupos de EST usando a busca lexica nas anotações de *hits* com o NR, e 809 grupos que

apresentaram similaridade com sequências depositadas em bancos específicos. No primeiro caso o ajuste de parâmetros de busca do algoritmo BLAST para o número de hits foi feito no sentido de obter 40 *hits* para cada comparação e sucessivamente se fez necessária a utilização de *script* para a localização e seleção de sequências apresentando termos anotadores característicos de retrotransposons. No segundo caso, 437 grupos que apresentaram termos anotadores sugestivos de transposons de DNA, como transposase, foram reservados para análises futuras. Já aqueles com termos anotadores estritamente relacionados a retrotransposons, 372 grupos, foram utilizados preferencialmente para a continuidade dos trabalhos. É interessante, porém, destacar que embora apresentando termo anotador sugestivo de elementos da classe II, os 437 grupos “excluídos” podem albergar sequências que façam parte de retrotransposons. Isso se deve ao fato de que não há um padrão universal para o processo de anotação e que as sequências de integrase, por exemplo, características de retrotransposons, apresentam similaridade estrutural com as transposases de elementos da Classe II (Rice e Baker, 2001). Na literatura podemos encontrar tanto trabalhos que utilizaram a primeira abordagem, com comparação inicial contra o banco NR (Zhu e Oudemans, 2000), quanto estudos que utilizaram preferencialmente os bancos RepBase e TEfam (Desjardins, Champion *et al.*, 2011; Fernandez-Medina, Struchiner *et al.*, 2011), sendo esta última mais frequente, provavelmente por sua acurácia.

A necessidade de definir um conjunto preferencial de EST para o processo inicial de identificação e caracterização dos putativos retrotransposons ativos no genoma de *Paracoccidioides* suscitou a indagação de se havia grupos redundantes entre os 142 oriundos do NR e filtrados por meio da busca léxica e os 809 grupos derivados das comparações com sequências de bancos específicos. A resposta a esta indagação resultou num processo de triagem e permitiu restringir a 52 o conjunto preferencial de grupos, os primeiros a serem analisados na etapa subsequente.

Cinco destes 52 grupos se destacaram por seu elevado escore de similaridade com retrotransposons, em todos os banco de dados e pela maior densidade de termos anotadores. Saliente-se que esses cinco grupos permitiram identificar regiões genômicas com fases de leitura aberta e elevado potencial de correspondência com retrotransposons, pois nelas foi possível ancorar um total

de 27 dos 52 grupos do conjunto preferencial. Trabalhos de transcriptoma relatam a construção de bibliotecas de cDNA a partir de moléculas de mRNA e consideram para análises futuras sequências que possuam tamanho maior que 200pb (Felipe, Andrade *et al.*, 2003; Goldman, Dos Reis Marques *et al.*, 2003; Felipe, Andrade *et al.*, 2005). Desta forma, é possível que grupos de sequências diferentes permitam a identificação de regiões genômicas que correspondam a um mesmo retrotransposon (Tabela 10). A identificação dessas regiões genômicas por meio do ancoramento de diferentes grupos de ESTs foi de fundamental importância para a caracterização de cópias inteiras dos elementos estudados facilitou a caracterização dos cinco elementos encontrados na execução deste estudo. Tendo sido mapeadas as regiões genômicas com potencial de serem sequências completas de retrotransposon e com o auxílio dos programas LTR-FINDER (Xu e Wang, 2007) e ORF-FINDER (<http://www.ncbi.nlm.nih.gov/projects/gorf/>) foi possível determinar a estrutura e a localização das sequências repetidas que flanqueiam as sequências codificantes e domínios protéicos característicos (Figuras 10-12).

Foram identificados e caracterizados 4 retrotransposons da ordem LTR e um retrotransposon da ordem LINE (Wicker, Sabot *et al.*, 2007). O elemento da ordem LINE (RtPb2), não-LTR, apresentou características da superfamília I. As cópias autônomas desses elementos codificam pelo menos uma proteína transcriptase reversa e uma endonuclease (endonuclease apurinica ou apirimidinica) e apresentam uma ORF semelhante a *gag*, além de domínio de transcriptase reversa; apresentarem cauda poli A (Wicker, Sabot *et al.*, 2007) Os outros quatro elementos caracterizados, todos da ordem LTR, tipicamente apresentam terminações longas repetidas (LTR), comuns também a retrovirus, as quais abrigam regiões promotoras do processo de transposição (Finnegan, 2012). Dos quatro elementos caracterizados dois apresentaram estrutura característica da superfamília Ty3-Gypsy (RtPb1 e RtPb5) e dois Ty1-Copia (RtPb3 e RtPb4). É interessante comentar que, Muzewska e colaboradores (2011) explicitam que apenas duas famílias de elementos LTR são encontrados em fungos, até o momento, justamente Ty3-Gypsy e Ty1-Copia e que a maioria dos representantes de Ty3-Gypsy encontrados são majoritariamente Chromovirus (Metaviridae) (Muszewska, Hoffman-Sommer *et al.*, 2011).

As LTR são sequências que podem variar em tamanho entre 80 pb e 25kb, e possuem as bases TG e CA nas suas extremidades 5' e 3', respectivamente. As LTR não codificam proteína mas desempenham um importante papel no controle da atividade dos elementos com LTR (Gao, Jimenez-Lopez *et al.*, 2012). Além disso, as LTR de determinados elementos possuem sinais regulatórios que conferem a capacidade de perceber estímulos externos e ativar o processo de transposição. Takeda e colaboradores (1999) relataram a presença de um elemento retrotransponível em tabaco cuja transposição pode ser ativada por vários estresses, incluindo infecção viral. Os autores demonstraram que uma repetição de 13 pb presente na região promotora da LTR funciona como um elemento de regulação e que altos níveis de expressão foram vistos em resposta ao estresse induzido quimicamente (Takeda, Sugimoto *et al.*, 1999).

Entre os elementos da ordem LTR observa-se a presença de duas fases abertas de leitura (ORF). A primeira para GAG, proteína estrutural que compõe partículas semelhantes a vírus, e a segunda para a poliproteína POL. Esta poliproteína pode possuir domínios catalíticos de proteinase aspártica (PR AP), integrase (IN), transcriptase reversa (RT) e ribonuclease H (RNase H ou RH) (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b; Finnegan, 2012). Baseado na divergência das sequências e na ordem dos produtos gênicos, os elementos retrotransponíveis com LTR podem ser subdividido em 5 superfamílias, entre as quais Ty1-Copia e Ty3-Gypsy (Wicker, Sabot *et al.*, 2007; Eickbush e Jamburuthugoda, 2008b; Finnegan, 2012).

É interessante destacar dois aspectos estruturais peculiares observados na organização de dois dos elementos LTR aqui caracterizados: - o primeiro foi observado no elemento Gypsy-RtPb1: - além de regiões típicas de transcriptase reversa (RT), RNase H, e integrase (Int) foi observada a presença de um cromodomínio (*Chromo-domain – chromatin organization modifier*), uma integrase que apresenta um cromodomínio na região C-terminal. Um cromodomínio é descrito como constituído por 40-50 aminoácidos encontrado em proteínas envolvidas na reorganização (remodelamento) da cromatina. Muitas vezes essas regiões apresentam especificidade para interação com proteínas histonas H3 atuando como metiltransferases, dirigindo os elementos durante o processo de integração e implicadas na regulação da expressão

genica ou no silenciamento transcricional mediado por RNA (“heterocromatinização” de certos *loci*) (Horita, Ivanova *et al.*, 2001; Hughes e Hawley, 2009).

Quanto a sua ocorrência em *Paracoccidioides*, de fato, desde 2004, Gorinsek e colaboradores relataram a existência de retrotransposons do gênero Chromovirus (família Metaviridae, Gypsy/Ty3) entre as ESTs de fungos (Ascomycetos) depositadas em bancos de dados públicos, nominando, entre outros fungos, *P. brasiliensis*, num estudo de genômica comparativa desse gênero de retroelementos em eucariotos (Gorinsek, Gubensek *et al.*, 2004). Como mencionado anteriormente, a maioria dos representantes da superfamília Ty3-Gypsy encontrados nos genomas de fungos são Chromovirus, considerados os elementos mais bem sucedidos nos fungos, duas vezes mais bem sucedidos na dominação desses genomas do que os elementos Ty1-Copia (Muszewska, Hoffman-Sommer *et al.*, 2011).

O outro aspecto organizacional a se destacar foi o não mapeamento no elemento Copia-RtPb3 do domínio catalítico de proteinase aspártica. Todos os domínios catalíticos dos elementos deste trabalho foram mapeados com por meio de comparação com o banco de dados de domínios conservados do NCBI (<http://www.ncbi.nlm.nih.gov/cdd/>). O fato de algum domínio não ter sido mapeado em algum elemento não implica na sua inexistência, e sim no fato de não ter sido possível identifica-lo por essa metodologia (Wicker, Sabot *et al.*, 2007).

A comparação das sequências desses elementos com aquelas depositados no banco de dados GIRI (*Genetic Information Research Institute*) mostrou que os elementos deste estudo fazem parte de famílias distintas daquelas já depositadas (identidade < 80%), ou seja, provavelmente está-se diante de novas famílias de retrotransposons das ordens LTR e LINE, a exemplo do que ocorreu com os elementos Trem A-H, descritos por Marini e colaboradores (2010) como 8 novas famílias de transposons de DNA (Marini, Zanforlin *et al.*, 2010).

Análise filogenética utilizando sequências de aminoácidos de RT tem sido relatada. Isso porque os agrupamentos realizados com base em características

estruturais comuns e identidade de genes em nível de nucleotídeos podem não contemplar todos os elementos. Desta forma, a presença do domínio da transcriptase reversa em todos os elementos retrotransponíveis torna-o de grande importância em estudos filogenéticos (Wicker, Sabot *et al.*, 2007; Eickbush, Ye *et al.*, 2008; Eickbush e Jamburuthugoda, 2008b). Vale ressaltar aqui que a metodologia utilizada no breve estudo filogenético realizado foi adequada, pois a questão a ser respondida por essa análise estava relacionada apenas à classificação em nível de Ordem e Superfamília, não havendo a intenção de mostrar um caminho evolutivo, mas confirmar a classificação por meio da análise de distância entre pares de sequências.

Depois de caracterizadas e classificadas, uma cópia (completa e íntegra) de cada elemento foi utilizada como sequência de busca para localização de outras cópias nos três genomas estudados. Utilizou-se para isso os critérios estabelecidos 80-80-80, pelos quais as sequências representam o mesmo elemento se houver 80% de identidade, em 80% das sequências alinhadas, em pelo menos 80 pares de bases (Wicker, Sabot *et al.*, 2007). De acordo com esses critérios, 538 cópias correspondendo aos cinco elementos foram encontradas. Desjardins e colaboradores (2011), no artigo que realizou a primeira comparação genômica de fungos do complexo *Paracoccidioides* apresentaram um *pool* de elementos móveis nos genomas dos mesmos isolados utilizados neste trabalho. Os autores afirmam que todos os tipos básicos de elementos transponíveis estão presentes e que compõem entre 8-9% do genoma dos isolados de Pb03 e Pb18, ao passo que Pb01 apresenta duas vezes mais, 16% de TE. Neste trabalho, não foi observado um padrão de distribuição entre as cópias, mas os valores refletem o mesmo padrão descrito por Desjardins e colaboradores (2011), tendo-se encontrado mais da metade delas (366 cópias) no isolado Pb01, seguidas por Pb18 (115) e Pb03 (57). Foi interessante observar que esse número tão maior pode ser atribuído à expansão elemento Gypsy-RtPb5 registrada no isolado Pb01, com 236 cópias, todas truncadas. Essa expansão não foi observada nos outros isolados, Pb03 e Pb18, com respectivamente 56 e 63 cópias. Um padrão distinto de distribuição dos elementos da classe II (transposons de DNA) foi encontrado por Marini e colaboradores (2010) quando mostraram que a distribuição dos elementos TREM (**T**ransposable **e**lements **m**ariner)

caracterizados também é semelhante entre os isolados Pb03 (475) e Pb18 (473), diferindo do isolado Pb01 (384). Neste caso, porém, o número de elementos Trem foi menor no isolado Pb01 (Marini, Zanforlin *et al.*, 2010).

Considerando apenas as cópias completas foi possível notar que novamente o isolado Pb01 apresentava o maior número (159 cópias), seguidos por Pb03 (33) e Pb18 (55). Quando considerado no número de cópias íntegras (aquelas que possuem o potencial de codificar as enzimas necessárias à mobilidade do elemento) foi observado o maior número em Pb01. É interessante notar que o mesmo padrão de distribuição é observado. Além das diferenças genômicas apresentadas por Desjardins (2011), trabalhos de filogenia utilizando diferentes loci gênicos e regiões ITS (Carrero, Niño-Vega *et al.*, 2008; Teixeira, Theodoro *et al.*, 2009b), considerando os três isolados, têm mostrado que o isolado Pb01 apresenta diferença significativa com relação aos demais isolados. Estes estudos levaram a considerar o isolado Pb01 uma nova espécie denominada *P. lutzii*.

Embora tenha sido observado um grande número de cópias truncadas (cópias não autônomas) em todos os elementos, há a possibilidade dessas cópias serem ativadas por enzimas codificadas por cópias íntegras (autônomas), pois a ativação de elementos não-autônomos é possível desde que as regiões promotoras presentes nas terminações longas repetidas (LTR) estejam inalteradas, ou seja, não sofreram mutação (Wicker, Sabot *et al.*, 2007). Dos elementos descritos neste trabalho, observou-se que no isolado Pb01 não foi encontrada cópia íntegra dos elementos Copia-RtPb4 e Gypsy-RtPb5. No isolado Pb03, não foi observada cópias íntegras do elemento Gypsy-RtPb5 (Figura 14). Embora tenham sido inativados, esses elementos ainda representam potencial de mobilidade, pois alguns deles poderiam ser ativados pela ação de outros pertencentes a famílias diferentes (Wicker, Sabot *et al.*, 2007). Sendo assim, todas as cópias devem ser consideradas como potenciais geradoras de mudanças genômicas, contribuindo para a diferenciação entre isolados. Desta forma talvez pudesse ser interessante investigar se a presença e o elevado número de cópias de elementos móveis em Pb01 *P. lutzii* (isolados Pb01-*like*) em relação aos demais isolados, que constituem linhagens distintas

da espécie *P. brasiliensis*, poderia ter contribuído de alguma forma na especiação em *Paracoccidioides*.

Finalmente, pareceu-nos importante verificar experimentalmente se todos os elementos encontrados eram transcricionalmente ativos e se era possível quantificar esta atividade. Observou-se amplificação a partir de cDNA indicativa da presença de transcritos para 4 (Gypsy-RtPb1, LINE-RtPb2, CopiaRtPb3 e Copia-RtPb4) dos 5 elementos caracterizados. Por outro lado, não foram obtidos produtos de amplificação a partir de cDNA, nos isolados Pb01 e Pb03, que correspondessem ao elemento Gypsy-RTPb5, sugerindo seu silenciamento pelos genomas em questão. Já no isolado Pb18 observou-se a presença do amplicon correspondente ao elemento Gypsy-RTPb5. Na quantificação por qRT-PCR o elemento Gypsy-RtPb1 e o elemento LINE-RtPb2 foram mais expressos no isolado Pb01 sendo transcritos também nos dois outros isolados. A expressão de um elemento da superfamília Gypsy foi relatada, no isolado Pb01, por Shankar e colaboradores (2011) após desafiarem este isolado com o hormônio β -17 estradiol (Shankar, Wu *et al.*, 2011).

O elemento Copia-RtPb3 também apresentou sinais de transcrição em todos os isolados, mas teve maior atividade em Pb03. O Copia-RtPb4 apresentou baixa quantidade de transcritos em todos os isolados, o que contraria os resultados do RT-PCR quando foi observada amplificação a partir do cDNA. Vale ressaltar neste ponto que a amplificação a pela técnica Rt-PCR desse elemento só foi possível com um volume maior de cDNA, o que pode corroborar com os baixos valores obtido por qRT-PCR. Por fim, as análises do elemento Gypsy-RtPb5 permitiram observar que este foi o elemento com a menor taxa de expressão, com alguma expressão apenas no isolado Pb18 (0,44 UR), o que corrobora a ausência de amplicons nos isolados Pb01 e Pb03.

É inevitável indagar se esses resultados apresentam correlação com os dados *in silico* relativos ao numero de copias integras e cópias truncadas. E podemos observar que há coerência entre esses dados no isolado Pb01, que tem o maior numero de copias integras e a maior abundancia relativa de transcritos para os elementos Gypsy-RtPb1 e LINE-RtPb2. É também coerente o resultado do Copia-RtPb3 onde o maior número de cópias integras foi

observado no isolado Pb03, aquele que apresentou maior abundância desse transcrito. O elemento Copia-RtPb4 que não apresentou cópias integras no isolado Pb01 teve nesse isolado a menor abundância relativa desse transcrito. Para esse elemento, a abundância de transcritos foi maior no isolado Pb03 (com apenas uma cópia íntegra) do que no Pb18 (com duas cópias íntegras). Gypsy-RtPb5, um elemento que não apresentou cópias íntegras nos isolados Pb01 e Pb03, assinalando-se apenas duas cópias íntegras no isolado Pb18, mostrou naqueles primeiros isolados a menor abundância relativa de transcritos.

Quanto ao padrão de transcrição dos elementos nos diferentes isolados, observa-se que, no Pb01, os elementos que apresentaram maior abundância relativa de transcritos foram Gypsy-RtPb1 (11UR), LINE-RtPb2 (4UR) e Copia-RtPb3 (2UR). No isolado Pb03 esse padrão foi Copia-RtPb3 (7UR), Gypsy-RtPb1 (5UR), LINE-RtPb2 (3UR). No isolado Pb18, aquele que, em geral, apresentou a menor abundância relativa de transcritos para todos os elementos, em ordem decrescente, Gypsy-RtPb1 (3UR), Copia-RtPb3 (~2UR).

As expansões de elementos móveis em fungos se referem tanto ao aumento dos números de cópias de elementos individuais quanto à diversidade ou número de tipos de retroelementos encontrados nesses genomas. Os genomas tendem a impedir as expansões desses elementos empregando para isso 3 mecanismos de silenciamento: RPI (*repeat induced point mutation*), cuja ocorrência em *Paracoccidioides* observamos preliminarmente por meio de alinhamentos realizados durante as análises (Figura 18) e SIS (*sex-induced silencing*) (Muszewska, Hoffman-Sommer *et al.*, 2011). Um dos desdobramentos que podem decorrer desse trabalho indicam a necessidade de se estudar mais detalhadamente por meio de que estratégias *Paracoccidioides* silencia os elementos.

Foi possível, portanto, no presente trabalho, identificar 5 retrotransposons, no genoma de fungos dimórficos do complexo *Paracoccidioides*, caracterizados como das ordens LTR (4) e LINE (1), das superfamílias Gypsy (2), Copia (2) e I (1), transcricionalmente ativos, contudo com padrões distintos nos genomas estudados.

7 - CONCLUSÕES

- As EST depositadas nos banco de dados do NCBI albergam sequências representativas de retrotransposons, indicando a expressão desses elementos;
- Cinco putativos retrotransposons expressos no transcriptoma de fungos do complexo *Paracoccidioides*, denominados RtPb1, RtPb2, RtPb3, RtPb4 e RtPb5, foram identificados e caracterizados.
- As sequências completas dos elementos foram localizadas e caracterizadas. As regiões correspondentes à Protease, Transcriptase reversa-RNaseH, Integrase e Endonuclease foram mapeadas estabelecendo-se suas respectivas coordenadas.
- As regiões correspondente às terminações longas repetidas (LTRs) e cauda PoliA foram descritas juntamente com suas coordenadas. A classificação foi feita por meio de alinhamento com sequências de elementos descritos e permitiu agrupa-los nas superfamílias LINES, *Copia* e *Gypsy*.
- O mapeamento físico indicou a presença de múltiplas cópias dos elementos (538), distribuídos de forma heterogênea em vários *supercontigs* nas sequencias dos genomas conhecidos (isolados Pb01, Pb03 e Pb18).
- A presença dos cinco retrotransposons foi demonstrada no DNA genômico de 34 isolados do fungo por amplificação específica (PCR) de segmentos correspondentes as regiões correspondentes à transcriptase reversa.
- Análises de mapeamento por *Southern* evidenciaram a presença de múltiplas cópias dos elementos nos três genomas estudados.
- Análise dos por meio de RT-PCR e qRT-PCR permitiu observar níveis diferentes de transcritos dos diferentes elementos nos três genomas.

8 – PERSPECTIVAS

- Verificar a possibilidade de identificação e caracterização de outros retrotransposons nos grupos ainda em estudo (NR=90 grupos e Bancos específicos = 320).
- Analisar a presença dos elementos estudados nos genomas de outros fungos dimórficos.
- Verificar a relação dos retrotransposons com genes de virulência dos fungos do complexo *Paracoccidioides*.

9 – REFERÊNCIAS BIBLIOGRÁFICAS

ALMEIDA, A. J. et al. New insights into the cell cycle profile of *Paracoccidioides brasiliensis*. **Fungal Genet Biol**, v. 43, n. 6, p. 401-9, Jun 2006. ISSN 1087-1845. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16631397> >.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **J Mol Biol**, v. 215, n. 3, p. 403-10, Oct 1990. ISSN 0022-2836. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2231712> >.

ALVES, F. L. **Papel dos transposons de DNA da Superfamília *TC1/mariner* (TREM A-H) no reconhecimento de linhagens de fungos patogênicos do gênero *Paracoccidioides***. 2012. 95 UFMG

ARAÚJO, R. A. D. C. **Elementos genéticos móveis no transcriptoma do fungo patogênico *Paracoccidioides brasiliensis*: caracterização de dois retrotransposons majoritários**. 2008

BAGAGLI, E. et al. High frequency of *Paracoccidioides brasiliensis* infection in armadillos (*Dasyus novemcinctus*): an ecological study. **Med Mycol**, v. 41, n. 3, p. 217-23, Jun 2003. ISSN 1369-3786. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12964713> >.

BAGAGLI, E. et al. Isolation of *Paracoccidioides brasiliensis* from armadillos (*Dasyus novemcinctus*) captured in an endemic area of paracoccidioidomycosis. **Am J Trop Med Hyg**, v. 58, n. 4, p. 505-12, Apr 1998. ISSN 0002-9637. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9574800> >.

BARBOSA, M. S. et al. The glyceraldehyde-3-phosphate dehydrogenase homologue is differentially regulated in phases of *Paracoccidioides brasiliensis*: molecular and phylogenetic analysis. **Fungal Genet Biol**, v. 41, n. 7, p. 667-75, Jul 2004. ISSN 1087-1845. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15275662> >.

BASTOS, K. P. et al. The transcriptome analysis of early morphogenesis in *Paracoccidioides brasiliensis* mycelium reveals novel and induced genes potentially associated to the dimorphic process. **BMC Microbiol**, v. 7, p. 29, 2007. ISSN 1471-2180 (Electronic)

1471-2180 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17425801> >.

BASTOS, K. P. et al. The transcriptome analysis of early morphogenesis in *Paracoccidioides brasiliensis* mycelium reveals novel and induced genes

potentially associated to the dimorphic process. **BMC Microbiol**, v. 7, p. 29, 2007. ISSN 1471-2180. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17425801> >.

BENACHENHOU, F. et al. Conserved structure and inferred evolutionary history of long terminal repeats (LTRs). **Mob DNA**, v. 4, n. 1, p. 5, Feb 1 2013. ISSN 1759-8753 (Print). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23369192> >.

BETHLEM, E. P. et al. Paracoccidioidomycosis. **Curr Opin Pulm Med**, v. 5, n. 5, p. 319-25, Sep 1999. ISSN 1070-5287. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10461538> >.

BIÉMONT, C.; VIEIRA, C. Genetics: junk DNA as an evolutionary force. **Nature**, v. 443, n. 7111, p. 521-4, Oct 2006. ISSN 1476-4687. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17024082> >.

BIGOT, Y.; BRILLET, B.; AUGÉ-GOUILLOU, C. Conservation of Palindromic and Mirror Motifs within Inverted Terminal Repeats of mariner-like Elements. **J Mol Biol**, v. 351, n. 1, p. 108-16, Aug 2005. ISSN 0022-2836. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15946679> >.

BORGES, C. L. et al. The antigenic and catalytically active formamidase of *Paracoccidioides brasiliensis*: protein characterization, cDNA and gene cloning, heterologous expression and functional analysis of the recombinant protein. **Microbes Infect**, v. 7, n. 1, p. 66-77, Jan 2005. ISSN 1286-4579. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15716068> >.

BOUVET, G. F.; JACOBI, V.; BERNIER, L. Characterization of three DNA transposons in the Dutch elm disease fungi and evidence of repeat-induced point (RIP) mutations. **Fungal Genet Biol**, v. 44, n. 5, p. 430-43, May 2007. ISSN 1087-1845. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17291794> >.

BRUMMER, E.; CASTANEDA, E.; RESTREPO, A. Paracoccidioidomycosis: an update. **Clin Microbiol Rev**, v. 6, n. 2, p. 89-117, Apr 1993. ISSN 0893-8512. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/8472249> >.

CALCAGNO, A. M. et al. Geographic discrimination of *Paracoccidioides brasiliensis* strains by randomly amplified polymorphic DNA analysis. **J Clin Microbiol**, v. 36, n. 6, p. 1733-6, Jun 1998. ISSN 0095-1137 (Print)

0095-1137 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9620409> >.

CAMARGO, Z. P. et al. Monoclonal antibody capture enzyme immunoassay for detection of *Paracoccidioides brasiliensis* antibodies in paracoccidioidomycosis. **J Clin Microbiol**, v. 32, n. 10, p. 2377-81, Oct 1994. ISSN 0095-1137 (Print)

0095-1137 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/7814469> >.

CANO, M. I. et al. Electrophoretic karyotypes and genome sizing of the pathogenic fungus *Paracoccidioides brasiliensis*. **J Clin Microbiol**, v. 36, n. 3, p. 742-7, Mar 1998. ISSN 0095-1137. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9508305> >.

CAPY, P. et al. Stress and transposable elements: co-evolution or useful parasites? **Heredity**, v. 85 (Pt 2), p. 101-6, Aug 2000. ISSN 0018-067X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11012710> >.

CARRERO, L. L. et al. New *Paracoccidioides brasiliensis* isolate reveals unexpected genomic variability in this human pathogen. **Fungal Genet Biol**, v. 45, n. 5, p. 605-12, May 2008. ISSN 1096-0937. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18364259> >.

CHURCH, G. M.; GILBERT, W. Genomic sequencing. **Proc Natl Acad Sci U S A**, v. 81, n. 7, p. 1991-5, Apr 1984. ISSN 0027-8424. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/6326095> >.

CISALPINO, P. S. et al. Cloning, characterization, and epitope expression of the major diagnostic antigen of *Paracoccidioides brasiliensis*. **J Biol Chem**, v. 271, n. 8, p. 4553-60, Feb 1996. ISSN 0021-9258. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/8626811> >.

CORREDOR, G. G. et al. The naked-tailed armadillo *Cabassous centralis* (Miller 1899): a new host to *Paracoccidioides brasiliensis*. Molecular identification of the isolate. **Med Mycol**, v. 43, n. 3, p. 275-80, May 2005. ISSN 1369-3786. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16010854> >.

COSTA, A. A. et al. Characterization of a gene which encodes a mannosyltransferase homolog of *Paracoccidioides brasiliensis*. **Microbes Infect**, v. 4, n. 10, p. 1027-34, Aug 2002. ISSN 1286-4579. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12191652> >.

COSTA, M. et al. Transcriptome profiling of *Paracoccidioides brasiliensis* yeast-phase cells recovered from infected mice brings new insights into fungal response upon host interaction. **Microbiology**, v. 153, n. Pt 12, p. 4194-207, Dec 2007. ISSN 1350-0872 (Print)

1350-0872 (Linking). Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/18048933>>.

COUTINHO, Z. F. et al. Paracoccidioidomycosis mortality in Brazil (1980-1995). **Cad Saude Publica**, v. 18, n. 5, p. 1441-54, 2002 Sep-Oct 2002. ISSN 0102-311X. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/12244377>>.

COWLEY, M.; OAKLEY, R. J. Transposable elements re-wire and fine-tune the transcriptome. **PLoS Genet**, v. 9, n. 1, p. e1003234, Jan 2013. ISSN 1553-7404 (Electronic)

1553-7390 (Linking). Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/23358118>>.

CUNHA, A. F. et al. Identification, N-terminal region sequencing and similarity analysis of differentially expressed proteins in *Paracoccidioides brasiliensis*. **Med Mycol**, v. 37, n. 2, p. 115-21, Apr 1999. ISSN 1369-3786. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/10361267>>.

CUNHA, D. A. et al. Heterologous expression, purification, and immunological reactivity of a recombinant HSP60 from *Paracoccidioides brasiliensis*. **Clin Diagn Lab Immunol**, v. 9, n. 2, p. 374-7, Mar 2002. ISSN 1071-412X. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/11874881>>.

DA FONSECA, C. A. et al. Two-dimensional electrophoresis and characterization of antigens from *Paracoccidioides brasiliensis*. **Microbes Infect**, v. 3, n. 7, p. 535-42, Jun 2001. ISSN 1286-4579. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/11418327>>.

DABOUSSI, M. J.; CAPY, P. Transposable elements in filamentous fungi. **Annu Rev Microbiol**, v. 57, p. 275-99, 2003. ISSN 0066-4227. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/14527280>>.

DE ALBORNOZ, M. B. Isolation of *Paracoccidioides brasiliensis* from rural soil in Venezuela. **Sabouraudia**, v. 9, n. 3, p. 248-53, Nov 1971. ISSN 0036-2174. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/4944202>>.

DE RUITER, M. T. et al. Retrotransposon insertion-site context (RISC) typing: a novel typing method for *Aspergillus fumigatus* and a convenient PCR alternative to restriction fragment length polymorphism analysis. **J Microbiol Methods**, v. 70, n. 3, p. 528-34, Sep 2007. ISSN 0167-7012. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/17692974>>.

DESJARDINS, C. A. et al. Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. **PLoS Genet**, v. 7, n. 10, p. e1002345, Oct 2011. ISSN 1553-7404. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22046142> >.

DINIZ, S. N. et al. Expression in bacteria of the gene encoding the gp43 antigen of paracoccidioides brasiliensis: immunological reactivity of the recombinant fusion proteins. **Clin Diagn Lab Immunol**, v. 9, n. 6, p. 1200-4, Nov 2002. ISSN 1071-412X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12414750> >.

EBINA, H.; LEVIN, H. L. Stress management: how cells take control of their transposons. **Mol Cell**, v. 27, n. 2, p. 180-1, Jul 20 2007. ISSN 1097-2765 (Print) 1097-2765 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17643368> >.

EICKBUSH, D. G. et al. Epigenetic regulation of retrotransposons within the nucleolus of Drosophila. **Mol Cell Biol**, v. 28, n. 20, p. 6452-61, Oct 2008. ISSN 1098-5549. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18678644> >.

EICKBUSH, T. H.; JAMBURUTHUGODA, V. K. The diversity of retrotransposons and the properties of their reverse transcriptases. **Virus Res**, v. 134, n. 1-2, p. 221-34, Jun 2008a. ISSN 0168-1702 (Print) 0168-1702 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18261821> >.

EICKBUSH, T. H.; JAMBURUTHUGODA, V. K. The diversity of retrotransposons and the properties of their reverse transcriptases. **Virus Res**, v. 134, n. 1-2, p. 221-34, Jun 2008b. ISSN 0168-1702. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/18261821> >.

FAGUNDES, R. Q. **Pesquisa da paracoccidioidomicose em cães (canis familiaris) na região endêmica de Botucatu, São Paulo**. 2002. Faculdade de Medicina Veterinária e Zootecnia de Botucatu da Universidade Estadual Paulista, Unesp

FARIAS, M. R. et al. Canine paracoccidioidomycosis: case report of generalized lymphadenitis. **Rev. Inst. Med. Trop. S. P.**, v. 47, 2005.

FEINBERG, A. P.; VOGELSTEIN, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. **Anal Biochem**, v. 132, n. 1, p. 6-13, Jul 1983. ISSN 0003-2697. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/6312838> >.

FEITOSA, L. O. S. et al. Chromosomal polymorphism, syntenic relationships, and ploidy in the pathogenic fungus *Paracoccidioides brasiliensis*. **Fungal Genet Biol**, v. 39, n. 1, p. 60-9, Jun 2003. ISSN 1087-1845. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12742064> >.

FELIPE, M. S. et al. Transcriptional profiles of the human pathogenic fungus *Paracoccidioides brasiliensis* in mycelium and yeast cells. **J Biol Chem**, v. 280, n. 26, p. 24706-14, Jul 2005. ISSN 0021-9258. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15849188> >.

FELIPE, M. S. et al. Transcriptome characterization of the dimorphic and pathogenic fungus *Paracoccidioides brasiliensis* by EST analysis. **Yeast**, v. 20, n. 3, p. 263-71, Feb 2003. ISSN 0749-503X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12557278> >.

FELIPE, M. S. et al. Functional genome of the human pathogenic fungus *Paracoccidioides brasiliensis*. **FEMS Immunol Med Microbiol**, v. 45, n. 3, p. 369-81, Sep 2005. ISSN 0928-8244. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16061364> >.

FERNANDEZ-MEDINA, R. D.; STRUCHINER, C. J.; RIBEIRO, J. M. Novel transposable elements from *Anopheles gambiae*. **BMC Genomics**, v. 12, p. 260, 2011. ISSN 1471-2164 (Electronic)

1471-2164 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21605407> >.

FERREIRA, M. S. et al. Isolation and characterization of a *Paracoccidioides brasiliensis* strain from a dogfood probably contaminated with soil in Uberlândia, Brazil. **J Med Vet Mycol**, v. 28, n. 3, p. 253-6, 1990. ISSN 0268-1218. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2213439> >.

FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends Genet**, v. 5, n. 4, p. 103-7, Apr 1989. ISSN 0168-9525. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2543105> >.

FINNEGAN, D. J. Retrotransposons. **Curr Biol**, v. 22, n. 11, p. R432-7, Jun 5 2012. ISSN 1879-0445 (Electronic)

0960-9822 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22677280> >.

FRANCO, M. et al. A critical analysis of isolation of *Paracoccidioides brasiliensis* from soil. **Med Mycol**, v. 38, n. 3, p. 185-91, Jun 2000. ISSN 1369-3786. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10892985> >.

FRANCO, M. et al. Paracoccidioidomycosis: a recently proposed classification of its clinical forms. **Rev Soc Bras Med Trop**, v. 20, n. 2, p. 129-32, 1987 Apr-Jun 1987. ISSN 0037-8682. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/3507739> >.

FUJINO, K.; SEKIGUCHI, H.; KIGUCHI, T. Identification of an active transposon in intact rice plants. **Mol Genet Genomics**, v. 273, n. 2, p. 150-7, Apr 2005. ISSN 1617-4615. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15803319> >.

GAO, D. et al. Functional and structural divergence of an unusual LTR retrotransposon family in plants. **PLoS One**, v. 7, n. 10, p. e48595, 2012. ISSN 1932-6203 (Electronic)

1932-6203 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/23119066> >.

GIRARDIN, H.; LATGÉ, J. P. **DNA extraction and quantification. In: Molecular Biology of Pathogenic Fungi. A laboratory Manual.** 1994.

GOLDMAN, G. H. et al. Expressed sequence tag analysis of the human pathogen *Paracoccidioides brasiliensis* yeast phase: identification of putative homologues of *Candida albicans* virulence and pathogenicity genes. **Eukaryot Cell**, v. 2, n. 1, p. 34-48, Feb 2003. ISSN 1535-9778. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12582121> >.

GOODWIN, T. J.; POULTER, R. T. The diversity of retrotransposons in the yeast *Cryptococcus neoformans*. **Yeast**, v. 18, n. 9, p. 865-80, Jun 30 2001. ISSN 0749-503X (Print)

0749-503X (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11427969> >.

GORINSEK, B.; GUBENSEK, F.; KORDIS, D. Evolutionary genomics of chromoviruses in eukaryotes. **Mol Biol Evol**, v. 21, n. 5, p. 781-98, May 2004. ISSN 0737-4038. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/14739248> >.

HALL, B. G. **Phylogenetic Trees Made Easy.** Terceira. Sunderland: Sinaeur Associates, 2008.

HAYES, F. Transposon-based strategies for microbial functional genomics and proteomics. **Annu Rev Genet**, v. 37, p. 3-29, 2003. ISSN 0066-4197. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/14616054> >.

HORITA, D. A. et al. Solution structure, domain features, and structural implications of mutants of the chromo domain from the fission yeast histone methyltransferase Clr4. **J Mol Biol**, v. 307, n. 3, p. 861-70, Mar 30 2001. ISSN 0022-2836 (Print)

0022-2836 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11273706> >.

HUANG, X.; MADAN, A. CAP3: A DNA sequence assembly program. **Genome Res**, v. 9, n. 9, p. 868-77, Sep 1999. ISSN 1088-9051. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10508846> >.

HUGHES, S. E.; HAWLEY, R. S. Heterochromatin: a rapidly evolving species barrier. **PLoS Biol**, v. 7, n. 10, p. e1000233, Oct 2009. ISSN 1545-7885 (Electronic)

1544-9173 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19830257> >.

IZACC, S. M. et al. Molecular cloning, characterization and expression of the heat shock protein 60 gene from the human pathogenic fungus *Paracoccidioides brasiliensis*. **Med Mycol**, v. 39, n. 5, p. 445-55, Oct 2001. ISSN 1369-3786. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/12054056> >.

KAZAZIAN, H. H. Mobile elements: drivers of genome evolution. **Science**, v. 303, n. 5664, p. 1626-32, Mar 2004. ISSN 1095-9203. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15016989> >.

KELLER, S. M.; HETTLER, E. A.; WICKES, B. L. A retrotransposon-derived probe for discriminating strains of *Cryptococcus neoformans*. **Mycopathologia**, v. 162, n. 6, p. 377-87, Dec 2006. ISSN 0301-486X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17146581> >.

KRAMEROV, D. A.; VASSETZKY, N. S. Origin and evolution of SINEs in eukaryotic genomes. **Heredity (Edinb)**, v. 107, n. 6, p. 487-95, Dec 2011. ISSN 1365-2540 (Electronic)

0018-067X (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21673742> >.

LABBE, J. et al. Characterization of transposable elements in the ectomycorrhizal fungus *Laccaria bicolor*. **PLoS One**, v. 7, n. 8, p. e40197, 2012. ISSN 1932-6203 (Electronic)

1932-6203 (Linking). Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/22870194>>.

LACAZ, C. A. S. et al. *Paracoccidioides brasiliensis*. A mycologic and immunochemical study of two strains. **Rev Inst Med Trop Sao Paulo**, v. 41, n. 2, p. 79-86, 1999 Mar-Apr 1999. ISSN 0036-4665. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/10413954>>.

LACAZ, C. S. F., M., RESTREPO-MORENO, A. DEL NEGRO, G. **Paracoccidioidomycosis**. Boca raton, 1994.

LANDER, E. S. et al. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, Feb 2001. ISSN 0028-0836. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/11237011>>.

LECLERC, M. C.; PHILIPPE, H.; GUÉHO, E. Phylogeny of dermatophytes and dimorphic fungi based on large subunit ribosomal RNA sequence comparisons. **J Med Vet Mycol**, v. 32, n. 5, p. 331-41, 1994. ISSN 0268-1218. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/7844699>>.

MARINI, M. M. et al. Identification and characterization of Tc1/mariner-like DNA transposons in genomes of the pathogenic fungi of the *Paracoccidioides* species complex. **BMC Genomics**, v. 11, p. 130, 2010. ISSN 1471-2164. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/20178623>>.

MATUTE, D. R. et al. Microsatellite analysis of three phylogenetic species of *Paracoccidioides brasiliensis*. **J Clin Microbiol**, v. 44, n. 6, p. 2153-7, Jun 2006. ISSN 0095-1137. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/16757613>>.

MCCLINTOCK, B. The origin and behavior of mutable loci in maize. **Proc Natl Acad Sci U S A**, v. 36, n. 6, p. 344-55, Jun 1950. ISSN 0027-8424. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/15430309>>.

MCEWEN, J. G. et al. Nuclear staining of *Paracoccidioides brasiliensis* conidia. **J Med Vet Mycol**, v. 25, n. 5, p. 343-5, Oct 1987. ISSN 0268-1218. Disponível em: <
<http://www.ncbi.nlm.nih.gov/pubmed/2448446>>.

MONROY, F.; SHEPPARD, D. C. Taf1: a class II transposon of *Aspergillus fumigatus*. **Fungal Genet Biol**, v. 42, n. 7, p. 638-45, Jul 2005. ISSN 1087-1845. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15896988> >.

MONTENEGRO, M. R. et al. Isolation of fungi from nature in the region of Botucatu, state of São Paulo, Brazil, an endemic area of paracoccidioidomycosis. **Mem Inst Oswaldo Cruz**, v. 91, n. 6, p. 665-70, 1996 Nov-Dec 1996. ISSN 0074-0276. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9283642> >.

MOREIRA, S. F. et al. Monofunctional catalase P of *Paracoccidioides brasiliensis*: identification, characterization, molecular cloning and expression analysis. **Yeast**, v. 21, n. 2, p. 173-82, Jan 2004. ISSN 0749-503X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/14755642> >.

MORGANTE, M. et al. Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. **Nat Genet**, v. 37, n. 9, p. 997-1002, Sep 2005. ISSN 1061-4036. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16056225> >.

MUSZEWSKA, A.; HOFFMAN-SOMMER, M.; GRYNBERG, M. LTR retrotransposons in fungi. **PLoS One**, v. 6, n. 12, p. e29425, 2011. ISSN 1932-6203 (Electronic)

1932-6203 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22242120> >.

NEGRONI, P. [The *Paracoccidioides brasiliensis* lives saprophytically in the soil of Argentina]. **Prensa Med Argent**, v. 53, n. 39, p. 2381-2, Sep 1966. ISSN 0032-745X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/5998765> >.

NUNES, C. C. et al. Diverse and tissue-enriched small RNAs in the plant pathogenic fungus, *Magnaporthe oryzae*. **BMC Genomics**, v. 12, p. 288, 2011. ISSN 1471-2164 (Electronic)

1471-2164 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21635781> >.

PETERSON, S. W.; SIGLER, L. Molecular genetic variation in *Emmonsia crescens* and *Emmonsia parva*, etiologic agents of adiaspiromycosis, and their phylogenetic relationship to *Blastomyces dermatitidis* (*Ajellomyces dermatitidis*) and other systemic fungal pathogens. **J Clin Microbiol**, v. 36, n. 10, p. 2918-25, Oct 1998. ISSN 0095-1137 (Print)

0095-1137 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9738044> >.

PRADO, M. S., M.B; LAURENTI, R; TRAVASSOS, L.R; TABORDA, C. P. Mortality due to systemic mycoses as a primary cause of death or in association with AIDS in Brazil: a review from 1996 to 2006. **Mem Inst Oswaldo Cruz**, v. 104, n. 3, p. 513-521, May 2009 2009.

PUCCIA, R. et al. Exocellular components of *Paracoccidioides brasiliensis*: identification of a specific antigen. **Infect Immun**, v. 53, n. 1, p. 199-206, Jul 1986. ISSN 0019-9567. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/2424841> >.

REINOSO, C. et al. Random sequencing of *Paracoccidioides brasiliensis* genes. **Med Mycol**, v. 43, n. 8, p. 681-9, Dec 2005. ISSN 1369-3786. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16422297> >.

RESTREPO, A. et al. **Estrogens inhibit mycelium-to-yeast transformation in the fungus *Paracoccidioides brasiliensis*: implications for resistance of females to paracoccidioidomycosis.** *Infect Immun*. 46: 346-53 p. 1984.

RICCI, G. et al. Canine paracoccidioidomycosis. **Med Mycol**, v. 42, n. 4, p. 379-83, Aug 2004. ISSN 1369-3786. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/15473365> >.

RICE, P. A.; BAKER, T. A. Comparative architecture of transposase and integrase complexes. **Nat Struct Biol**, v. 8, n. 5, p. 302-7, May 2001. ISSN 1072-8368 (Print)

1072-8368 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11774877> >.

RICHINI-PEREIRA, V. B. et al. Molecular detection of *Paracoccidioides brasiliensis* in road-killed wild animals. **Med Mycol**, v. 46, n. 1, p. 35-40, Feb 2008. ISSN 1369-3786. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/17885959> >.

RUTHERFORD, K. et al. Artemis: sequence visualization and annotation. **Bioinformatics**, v. 16, n. 10, p. 944-5, Oct 2000. ISSN 1367-4803. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/11120685> >.

SABOT, F.; SCHULMAN, A. H. Parasitism and the retrotransposon life cycle in plants: a hitchhiker's guide to the genome. **Heredity (Edinb)**, v. 97, n. 6, p. 381-8, Dec 2006. ISSN 0018-067X. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16985508> >.

SALEM-IZACC, S. M. et al. Protein synthesis patterns of *Paracoccidioides brasiliensis* isolates in stage-specific forms and during cellular differentiation. **J Med Vet Mycol**, v. 35, n. 3, p. 205-11, 1997 May-Jun 1997. ISSN 0268-1218. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/9229337> >.

SAN-BLAS, G. et al. Fungal polysaccharides. **J Med Vet Mycol**, v. 32 Suppl 1, p. 321-8, 1994. ISSN 0268-1218. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/7536840> >.

SHANKAR, J. et al. Influence of 17 β -estradiol on gene expression of *Paracoccidioides* during mycelia-to-yeast transition. **PLoS One**, v. 6, n. 12, p. e28402, 2011. ISSN 1932-6203. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/22194832> >.

SHIKANAI-YASUDA, M. A. et al. [Guidelines in paracoccidioidomycosis]. **Rev Soc Bras Med Trop**, v. 39, n. 3, p. 297-310, 2006 May-Jun 2006. ISSN 0037-8682. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/16906260> >.

SOARES, C. M. et al. Characterization of *Paracoccidioides brasiliensis* isolates by random amplified polymorphic DNA analysis. **J Clin Microbiol**, v. 33, n. 2, p. 505-7, Feb 1995. ISSN 0095-1137. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/7714219> >.

TAKEDA, S. et al. A 13-bp cis-regulatory element in the LTR promoter of the tobacco retrotransposon Tto1 is involved in responsiveness to tissue culture, wounding, methyl jasmonate and fungal elicitors. **Plant J**, v. 18, n. 4, p. 383-93, May 1999. ISSN 0960-7412 (Print)

0960-7412 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/10406122> >.

TAMURA, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Mol Biol Evol**, v. 28, n. 10, p. 2731-9, Oct 2011. ISSN 1537-1719 (Electronic)

0737-4038 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/21546353> >.

TEIXEIRA, M. M. et al. Phylogenetic analysis reveals a high level of speciation in the *Paracoccidioides* genus. **Mol Phylogenet Evol**, v. 52, n. 2, p. 273-83, Aug 2009a. ISSN 1095-9513 (Electronic)

1055-7903 (Linking). Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/19376249> >.

_____. Phylogenetic analysis reveals a high level of speciation in the Paracoccidioides genus. **Mol Phylogenet Evol**, v. 52, n. 2, p. 273-83, Aug 2009b. ISSN 1095-9513. Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/19376249> >.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Res**, v. 22, n. 22, p. 4673-80, Nov 1994. ISSN 0305-1048. Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/7984417> >.

TRAVASSOS, L. R. et al. Therapeutic activity of a killer peptide against experimental paracoccidioidomycosis. **J Antimicrob Chemother**, v. 54, n. 5, p. 956-8, Nov 2004. ISSN 0305-7453. Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/15448128> >.

UNTEREINER, W. A. et al. The Ajellomycetaceae, a new family of vertebrate-associated Onygenales. **Mycologia**, v. 96, n. 4, p. 812-21, Jul-Aug 2004. ISSN 0027-5514 (Print)

0027-5514 (Linking). Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/21148901> >.

VISBAL, G. et al. Paracoccidioides brasiliensis, paracoccidioidomycosis, and antifungal antibiotics. **Curr Drug Targets Infect Disord**, v. 5, n. 3, p. 211-26, Sep 2005. ISSN 1568-0053. Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/16181141> >.

WICKER, T. et al. A unified classification system for eukaryotic transposable elements. **Nat Rev Genet**, v. 8, n. 12, p. 973-82, Dec 2007. ISSN 1471-0064. Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/17984973> >.

XU, Z.; WANG, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. **Nucleic Acids Res**, v. 35, n. Web Server issue, p. W265-8, Jul 2007. ISSN 1362-4962 (Electronic)

0305-1048 (Linking). Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/17485477> >.

ZHU, P.; OUDEMANS, P. V. A long terminal repeat retrotransposon Cgret from the phytopathogenic fungus Colletotrichum gloeosporioides on cranberry. **Curr Genet**, v. 38, n. 5, p. 241-7, Dec 2000. ISSN 0172-8083 (Print)

0172-8083 (Linking). Disponible em: < <http://www.ncbi.nlm.nih.gov/pubmed/11191207> >.

ZOU, S.; KIM, J. M.; VOYTAS, D. F. The Saccharomyces retrotransposon Ty5 influences the organization of chromosome ends. **Nucleic Acids Res**, v. 24, n. 23, p. 4825-31, Dec 1996. ISSN 0305-1048. Disponível em: < <http://www.ncbi.nlm.nih.gov/pubmed/8972872> >.

ANEXOS

```

#!/usr/bin/perl
#Programa para leitura de arquivos gerados pelo
blast #Escrito por Marco Aurelio;
use Getopt::Long; #modulo que processa sua linha de comando
my %prots;
my ($in_file,$out_file); #variaveis que recebem os parametros da linha
de comando
my $usage = "\n$0 -i fasta_file -o output_file\n
    GetOptions ('i=s' => \$in_file, 'o=s' => \$out_file); # Carrega as opções
dadas na linha de comando
if(!$in_file or !$out_file) { die $usage;}          # Se não existir o arquivo de
entrada, mata o programa
if (! -e $in_file) {
    die "\nCannot find file: $in_file\n\n";        # Mata o programa se não
existir o arquivo de entrada
}
open (INPUT,"$in_file");                          # Abre o arquivo de entrada
open (OUTPUT,">$out_file");                       # Abre o arquivo de saída
while($linha=<INPUT>) {
    chomp $linha; #utilizado para retirar caracteres ocultos.
    if ($linha=~m/Query= /) {                      # Apresenta match com query
        #$count=0;
        $nome_contig='$';
        $nome_contig=~s/\>//g;
    }
    if ($linha=~m/reverse transcriptase/ && $linha=~m/^gi/){
        print OUTPUT "$nome_contig\t$linha\n";
    }
}
}

```

	RNaseH	RT		Gag		Endo	Copia-like	Retrotransposon	Retrovirus	Integrase
1	*1144	103	3251	3	1160	114	5815	*1144	*1144	*1144
2	*1145	331	3384	267	1161	115	8104	1145	*1145	*1147
3	*1146	332	3546	268	1200	1945	8707	1146	*1146	1148
4	*1147	333	3608	269	1295	2603	9844	1147	*1147	1155
5	1148	334	3695	270	2094	3881	11330	1148	1148	7932
6	1149	335	3705	271	2095	4106	12427	1149	1149	8104
7	1151	336	3959	272	2096	5404		1150	1150	8707
8	1152	337	4248	273	2579	5590		1151	1151	9844
9	1153	338	4852	274	2657	8104		1152	1152	10445
10	1155	339	5181	275	2658	8192		1153	1153	11330
11	1157	340	5182	770	3101	8251		1154	1154	12427
12	1160	341	5680	938	3336	8697		1155	1155	12801
13	8115	342	5906	1144	3451	8794		1156	1156	
14	8289	343	6641	1145	4219	9035		1157	1157	
15	9661	347	8065	1146	4852	9348		1158	1158	
16	10066	393	8104	1147	5537	9606		1160	1161	
17	10174	395	8192	1148	8104	9629		1161	8289	
18	12795	396	8289	1149	8289	9783		5815	12795	
19	12801	397	8707	1151	9053	10023		8104	12801	
20		399	8766	1152	9844	10215		8289		
21		402	9025	1153	11330	10533		8707		
22		624	9102	1154	12427	10739		9844		
23		778	9346	1155	12715	10778		11330		
24		795	9418	1157	12795	11330		12427		
25		1036	9617	1158				12795		
26		1144	9625							
27		1145	9748							
28		1146	9783							
29		1147	9844							
30		1148	9967							
31		1149	9981							
32		1151	10197							
33		1152	10827							
34		1153	10919							
35		1155	11125							
36		1157	11330							
37		1160	11486							
38		1161	11701							
39		1945	12330							
40		2546	12795							
41		2843	12801							
42		2873	12878							

*Os grupos que apresentaram redundância (foram identificados com o uso de dois ou mais termos anotadores)

Anexo 2 – Grupos encontrados com a utilização de script perl (Banco de dados NR)

Contig	Hit	Score	E-value
8104	gi 303320891 ref XP_003070440.1 copia-like retrotransposon	102	2,00E-20
11330	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	96	3,00E-18
12427	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	81	7,00E-14
9844	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	77	4,00E-12
5815	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	60	2,00E-07
8707	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	52	2,00E-12

Contig	Hit	core	E-value
9629	gi 168752009 ref ZP_02777031.1 endonuclease/exonuclease/phos	374	e-103
1945	gi 242826378 ref XP_002488629.1 endonuclease/reverse transcript	343	e-101
8794	gi 121704746 ref XP_001270636.1 ap endonuclease	285	7,00E-75
10215	gi 93004917 ref YP_579354.1 endonuclease/exonuclease/phos	192	4,00E-47
8251	gi 226279960 gb EEH35526.1 endonuclease/exonuclease/phos	189	2,00E-87
9783	gi 242826378 ref XP_002488629.1 endonuclease/reverse transcriptase	133	2,00E-29
8697	gi 226279960 gb EEH35526.1 endonuclease/exonuclease/phos	127	3,00E-28
5404	gi 76818411 ref YP_335567.1 putative endonuclease/exonuclease	97	8,00E-18
10533	gi 170739969 ref YP_001768624.1 endonuclease/exonuclease	90	2,00E-16
9348	gi 121710988 ref XP_001273110.1 AP endonuclease	88	1,00E-28
8104	gi 288597 emb CAA49283.1 gag,protease,endonuclease, reverse transcriptase	87	6,00E-16
8192	gi 242826378 ref XP_002488629.1 endonuclease/reverse transcriptase	87	7,00E-16
9035	gi 121704746 ref XP_001270636.1 ap endonuclease	87	4,00E-16
4106	gi 76818411 ref YP_335567.1 putative endonuclease/exonuclease	86	2,00E-14
115	gi 76818411 ref YP_335567.1 putative endonuclease/exonuclease	82	1,00E-13
9606	gi 76818411 ref YP_335567.1 putative endonuclease/exonuclease	77	1,00E-12
114	gi 237654677 ref YP_002890991.1 DNA mismatch endonuclease	76	8,00E-12
10023	gi 240107054 gb EER25241.1 AP endonuclease	73	4,00E-14
5590	gi 209881833 ref XP_002142354.1 endonuclease/exonuclease	70	3,00E-10
11330	gi 288597 emb CAA49283.1 gag,protease,endonuclease, reverse transcriptase	67	5,00E-10
10778	gi 86604847 ref YP_473610.1 endonuclease/exonuclease/pho	66	1,00E-09
3881	gi 76818411 ref YP_335567.1 putative endonuclease/exonuclease	63	9,00E-08
10739	gi 237834449 ref XP_002366522.1 endonuclease/exonuclease/phos	57	8,00E-07
2603	gi 237834449 ref XP_002366522.1 endonuclease/exonuclease/phos	55	9,00E-07

Anexo 3 - Grupos obtidos com os termos anotadores copia-like e endonuclease.

Contig	Hit	Score	E-value
3	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	175	3,00E-42
267	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	173	4,00E-41
268	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	175	9,00E-42
269	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	174	1,00E-41
270	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	175	1,00E-41
271	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	176	7,00E-42
272	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	175	7,00E-42
273	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	175	9,00E-42
274	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	162	4,00E-38
275	gi 126335377 ref XP_001372617.1 PREDICTED: similar to gag polyp	52	6,00E-15
770	gi 126342077 ref XP_001376670.1 PREDICTED: similar to gag prote	67	5,00E-10
938	gi 28569857 dbj BAC57901.1 gag-like protein [Anopheles gambiae	99	1,00E-18
1144	gi 242760222 ref XP_002339946.1 gag	130	4,00E-48
1145	gi 242760222 ref XP_002339946.1 gag	385	1,00E-104
1146	gi 242760222 ref XP_002339946.1 gag	201	9,00E-50
1147	gi 242760222 ref XP_002339946.1 gag	186	3,00E-45
1148	gi 242760222 ref XP_002339946.1 gag	177	3,00E-46
1149	gi 242760222 ref XP_002339946.1 gag	127	3,00E-27
1151	gi 242760222 ref XP_002339946.1 gag	89	1,00E-15
1152	gi 242760222 ref XP_002339946.1 gag	122	8,00E-26
1153	gi 242760222 ref XP_002339946.1 gag	86	8,00E-15
1154	gi 242760222 ref XP_002339946.1 gag	62	2,00E-08
1155	gi 242760222 ref XP_002339946.1 gag	151	2,00E-35

Anexo 3 - Grupos obtidos com o termo anotador Gag.

Contig	Hit	Score	E-value
1157	gi 242760222 ref XP_002339946.1 gag	59	6,00E-07
1158	gi 242760222 ref XP_002339946.1 gag	57	3,00E-06
1160	gi 242760222 ref XP_002339946.1 gag	132	5,00E-40
1161	gi 242760222 ref XP_002339946.1 gag	276	1,00E-72
1200	gi 34015312 gb AAQ56503.1 putative gag protein [Oryza sativa (j	62	3,00E-07
1295	gi 25229103 gb AAN71720.1 putative gag protein [Danio rerio	59	2,00E-06
2094	gi 213138273 gb ACJ44801.1 gag protein [Human immunodeficiency	61	3,00E-07
2095	gi 31415910 gb AAP50931.1 putative retrotransposon gag protein	52	9,00E-05
2096	gi 37728011 gb AAR02376.1 gag protein [Simian immunodeficiency	51	2,00E-04
2579	gi 28569857 dbj BAC57901.1 gag-like protein [Anopheles gambiae	59	7,00E-07
2657	gi 38503582 gb AAR22580.1 gag protein [Small ruminant lentivirus	63	8,00E-08
2658	gi 37935900 gb AAO47146.1 gag protein [Human immunodeficiency v	63	4,00E-08
3101	gi 538540 pir A40831 gag-akt polyprotein - AKT8 murine leukemia	79	8,00E-13
3336	gi 165974304 dbj BAF99127.1 gag protein [Magnaporthe oryzae	72	1,00E-10
3451	gi 28569857 dbj BAC57901.1 gag-like protein [Anopheles gambiae	48	3,00E-09
4219	gi 28569857 dbj BAC57901.1 gag-like protein [Anopheles gambiae	59	1,00E-06
4852	gi 242816502 ref XP_002486791.1 gag	99	9,00E-36
5537	gi 13435243 gb AAK26118.1 AC084406_1 putative gag-pol polyprotei	68	3,00E-10
8104	gi 7444442 pir T02087 gag	97	7,00E-19
8289	gi 242760222 ref XP_002339946.1 gag	94	2,00E-17
9053	gi 28569857 dbj BAC57901.1 gag-like protein [Anopheles gambiae	57	8,00E-07
9844	gi 241993361 gb ACS74199.1 putative gag-pol polyprotein [Fragar	69	1,00E-09
11330	gi 109391001 emb CAJ09951.2 putative gag-pol polyprotein [Citru	76	1,00E-12
12427	gi 47824985 gb AAT38758.1 Putative gag-pol polyprotein, identic	67	8,00E-10
12715	gi 538540 pir A40831 gag-akt polyprotein - AKT8 murine leukemia	109	1,00E-22
12795	gi 83753113 gb AAR29046.2 gag-pol polyprotein [Aspergillus flavus	53	6,00E-10

Anexo 3 - Grupos obtidos com o termo anotador Gag (Continuação).

Contig	Hit	Score	E-value
1144	gi 7493982 pir S71898 probable integrase Mars4 - fungus (Ascobo	82	6,00E-14
1147	gi 34452675 gb AAQ72730.1 putative integrase [Petunia x hybrida	100	4,00E-19
1148	gi 34452675 gb AAQ72730.1 putative integrase [Petunia x hybrida	111	1,00E-22
1155	gi 7493982 pir S71898 probable integrase Mars4 - fungus (Ascobo	74	6,00E-12
7932	gi 47176964 ref YP_015575.1 integrase [Oligotropha carboxidovor	59	6,00E-07
8104	gi 41469360 gb AAS07202.1 putative integrase [Oryza sativa (ja	87	6,00E-16
8707	gi 41469360 gb AAS07202.1 putative integrase [Oryza sativa (ja	40	7,00E-06
9844	gi 158578541 gb ABW74566.1 integrase [Boechera divaricarpa	65	8,00E-09
10445	gi 47176964 ref YP_015575.1 integrase [Oligotropha carboxidovor	55	4,00E-06
11330	gi 41469360 gb AAS07202.1 putative integrase [Oryza sativa (ja	72	2,00E-11
12427	gi 158578541 gb ABW74566.1 integrase [Boechera divaricarpa	66	1,00E-09
12801	gi 4140712 gb AAD04177.1 putative integrase [Oryza sativa	84	5,00E-15

Contig	Hit	Score	E-value
1145	gi 242775227 ref XP_002478601.1 retrovirus polyprotein	516	e-162
1161	gi 242775227 ref XP_002478601.1 retrovirus polyprotein	285	4,00E-75
1146	gi 212536843 ref XP_002148577.1 retrovirus polyprotein	240	2,00E-61
1147	gi 212536843 ref XP_002148577.1 retrovirus polyprotein	224	2,00E-56
1148	gi 212536843 ref XP_002148577.1 retrovirus polyprotein	203	3,00E-57
1149	gi 212536843 ref XP_002148577.1 retrovirus polyprotein	200	4,00E-49
1152	gi 212530973 ref XP_002145643.1 retrovirus polyprotein	181	2,00E-43
1153	gi 212530973 ref XP_002145643.1 retrovirus polyprotein	151	1,00E-34
12801	gi 212530907 ref XP_002145610.1 retrovirus polyprotein	151	4,00E-35
1155	gi 212545550 ref XP_002152929.1 retrovirus polyprotein	150	7,00E-35
1144	gi 212536843 ref XP_002148577.1 retrovirus polyprotein	149	9,00E-71
8289	gi 242775227 ref XP_002478601.1 retrovirus polyprotein	127	7,00E-38
1156	gi 212530973 ref XP_002145643.1 retrovirus polyprotein	117	3,00E-24
1150	gi 212536843 ref XP_002148577.1 retrovirus polyprotein	110	2,00E-35
1157	gi 212530973 ref XP_002145643.1 retrovirus polyprotein	110	5,00E-22
1151	gi 212545550 ref XP_002152929.1 retrovirus polyprotein	95	4,00E-21
1158	gi 212530973 ref XP_002145643.1 retrovirus polyprotein	91	3,00E-16
1154	gi 212530969 ref XP_002145641.1 retrovirus polyprotein	77	5,00E-12
12795	gi 212525970 ref XP_002143142.1 retrovirus polyprotein	77	2,00E-19

Anexo 3 - Grupos obtidos com os termos anotadores integrase e retrovírus.

Contig	Hit	Score	E-value
1144	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	149	5,00E-70
1145	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	535	e-169
1146	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	222	6,00E-56
1147	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	212	1,00E-52
1148	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	211	1,00E-53
1149	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	200	4,00E-49
1150	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	107	5,00E-21
1151	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	94	4,00E-17
1152	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	200	5,00E-49
1153	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	165	9,00E-39
1154	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	83	2,00E-14
1155	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	151	4,00E-35
1156	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	131	2,00E-28
1157	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	124	2,00E-26
1158	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	97	3,00E-18
1160	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	149	1,00E-54
1161	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	291	7,00E-77
5815	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	60	2,00E-07
8104	gi 77552925 gb ABA95721.1 retrotransposon proteintative, Ty	104	5,00E-21
8289	gi 242760779 ref XP_002340058.1 retrotransposon polyprotein	135	5,00E-36
8707	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	52	2,00E-12
9844	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	77	4,00E-12
11330	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	96	3,00E-18
12427	gi 303321121 ref XP_003070555.1 copia-like retrotransposon	81	7,00E-14
12795	gi 322712221 gb EFZ03794.1 retrotransposon nucleocapsid protein	49	3,00E-19

Anexo 3 - Grupos obtidos com o termo anotador retrotransposon

Contig	Hit	Score	E-value
1144	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	98	5,00E-38
1145	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	294	3,00E-77
1146	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	138	7,00E-31
1147	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	129	4,00E-28
1148	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	118	9,00E-25
1149	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	107	2,00E-21
1151	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	59	1,00E-06
1152	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	103	3,00E-20
1153	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	74	3,00E-11
1155	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	79	2,00E-13
1157	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	59	6,00E-07
1160	gi 7229089 dbj BAA92424.1 reverse transcriptase-RNase H [Tricho...	106	1,00E-28
8115	gi 240107561 gb EER25740.1 RNase H domain containing protein [C...	128	2,00E-28
8289	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	91	1,00E-16
9661	gi 71998627 ref NP_001022508.1 RNase H family member (rnh-1.1) ...	48	8,00E-04
10066	gi 240107561 gb EER25740.1 RNase H domain containing protein [C...	125	8,00E-27
10174	gi 17536263 ref NP_495796.1 RNase H family member (rnh-2) [Caen...	85	1,00E-14
12795	gi 7229089 dbj BAA92424.1 reverse transcriptase-RNase H [Tricho...	49	3,00E-10
12801	gi 5002510 dbj BAA78625.1 reverse transcriptase-RNase H-integra...	87	4,00E-16

Anexo 3 - Grupos obtidos com o termo anotador RNaseH.

Contig	Hit	Score	E-value
103	gi 7159927 emb CAB76395.1 -reverse-transcriptase-homologue-ND5-i	88	1,00E-15
331	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	75	1,00E-11
332	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	75	2,00E-11
333	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	75	1,00E-11
334	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	59	6,00E-07
335	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	75	2,00E-11
336	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	62	8,00E-08
337	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	74	2,00E-11
338	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	73	5,00E-11
339	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	73	1,00E-10
340	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	64	2,00E-10
341	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	75	1,00E-11
342	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	71	2,00E-10
343	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	60	3,00E-07
347	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	49	1,00E-07
393	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	72	9,00E-11
395	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	74	2,00E-11
396	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	70	6,00E-10
397	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	65	1,00E-08
399	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	59	7,00E-07
402	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	64	1,00E-08
624	gi 72255555 ref NP_001026800.1 -telomerase-reverse-transcriptase	44	0E-018-
778	gi 32490473 dbj BAC79156.1 -reverse-transcriptase-like-protein	45	0E-011-
795	gi 188593505 emb CAQ53852.1 -putative-reverse-transcriptase-matu	44	0E-022-
1036	gi 166197456 dbj BAG06167.1 -reverse-transcriptase-homolog-[Pyla	63	4,00E-08
1144	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	98	5,00E-38
1145	gi 50402537 gb AAT76520.1 -putative-reverse-transcriptase-RNaseH	296	7,00E-78
1146	gi 50402537 gb AAT76520.1 -putative-reverse-transcriptase-RNaseH	138	7,00E-31
1147	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	129	4,00E-28
1148	gi 50402537 gb AAT76520.1 -putative-reverse-transcriptase-RNaseH	134	1,00E-29
1149	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	107	2,00E-21

Anexo 3 - Grupos obtidos com o termo anotador transcriptase reversa.

Contig	Hit	Score	E-value
1151	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	59	1,00E-06
1152	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	103	3,00E-20
1153	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	74	3,00E-11
1155	gi 50402537 gb AAT76520.1 -putative-reverse-transcriptase-RNaseH	88	3,00E-16
1157	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	59	6,00E-07
1160	gi 20502510 dbj BAB91415.1 -marY1-like-reverse-transcriptase-[Tr	105	5,00E-33
1161	gi 50897341 gb AAT85792.1 -reverse-transcriptase-(RNA-dependent	218	5,00E-55
1945	gi 162568710 gb ABY19378.1 -reverse-transcriptase-[Paracoccidiosi	1108	0-0
2546	gi 241567434 ref XP_002402294.1 -reverse-transcriptase,-putative	37	3E-9-
2843	gi 225556128 gb EEH04418.1 -telomerase-reverse-transcriptase-[Aj	210	6,00E-53
2873	gi 14388177 gb AAK60396.1 -telomerase-reverse-transcriptase-[Cry	36	8E-8-
3251	gi 242827044 ref XP_002488757.1 -reverse-transcriptase,-putative	43	0E-056-
3384	gi 32490473 dbj BAC79156.1 -reverse-transcriptase-like-protein	43	0E-038-
3546	gi 242827044 ref XP_002488757.1 -reverse-transcriptase,-putative	38	1E-9-
3608	gi 241567434 ref XP_002402294.1 -reverse-transcriptase,-putative	37	6E-8-
3695	gi 166368946 ref YP_001661219.1 -reverse-transcriptase-[Microcys	38	0E-11-
3705	gi 32490473 dbj BAC79156.1 -reverse-transcriptase-like-protein	35	7E-8-
3959	gi 53370713 gb AAU89208.1 -reverse-transcriptase-(RNA-dependent	38	1E-6-
4248	gi 34392547 dbj BAC82601.1 -reverse-transcriptase-[Ciona-intesti	40	0E-76-
4852	gi 7248355 dbj BAA92704.1 -reverse-transcriptase-[Tricholoma-bak	61	3,00E-15
5181	gi 220683915 gb ACL80770.1 -telomerase-reverse-transcriptase-[St	39	0E-17-
5182	gi 220683915 gb ACL80770.1 -telomerase-reverse-transcriptase-[St	37	1E-8-
5680	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	39	0E-61-
5906	gi 77918975 ref YP_356790.1 -reverse-transcriptase/maturase-[Pel	33	9E-6-
6641	gi 242826474 ref XP_002488649.1 -reverse transcriptase, putat	238	8,00E-61
8065	gi 34015265 gb AAQ56457.1 -putative-reverse-transcriptase-[Oryza	34	5E-6-
8104	gi 68164467 gb AA87200.1 -reverse-transcriptase-[Sciadopitys-ve	94	5,00E-18
8192	gi 242826474 ref XP_002488649.1 -reverse-transcriptase,-putative	90	9,00E-17
8289	gi 5002510 dbj BAA78625.1 -reverse-transcriptase-RNase-H-integra	91	1,00E-16
8707	gi 68164451 gb AA87192.1 -reverse-transcriptase-[Cycas-revoluta]	44	7,00E-08
8766	gi 32490473 dbj BAC79156.1 -reverse-transcriptase-like-protein	37	0E-88-
9025	gi 6063425 dbj BAA85382.1 -DNA-polymerase/reverse-transcriptase-	35	9E-6-
9102	gi 241567434 ref XP_002402294.1 -reverse-transcriptase,-putative	37	0E-52-
9346	gi 1513119 gb AAB06732.1 -reverse-transcriptase-[Drosophila-pseu	34	9E-9-
9418	gi 156046084 ref XP_001589597.1 -reverse-transcriptase-[Scleroti	41	0E-12-

Anexo 3 - Grupos obtidos com o termo anotador transcriptase reversa (Continuação).

Contig	Hit	Score	E-value
9617	gi 34015152 gb AAQ56348.1 -putative-reverse-transcriptase-[Oryza	35	6E-6-
9625	gi 32490473 dbj BAC79156.1 -reverse-transcriptase-like-protein	37	1E-9-
9748	gi 241567434 ref XP_002402294.1 -reverse-transcriptase,-putative	38	0E-89-
9783	gi 242826474 ref XP_002488649.1 -reverse-transcriptase,-putative	128	6,00E-28
9844	gi 159464070 ref XP_001690265.1 -reverse-transcriptase,-gag-pol-	67	3,00E-09
9967	gi 224588076 ref YP_002640605.1 -reverse-transcriptase-[Saccharo	35	2E-6-
9981	gi 225556128 gb EEH04418.1 -telomerase-reverse-transcriptase-[Aj	145	2,00E-33
10197	gi 32490473 dbj BAC79156.1 -reverse-transcriptase-like-protein	39	0E-67-
10827	gi 32490473 dbj BAC79156.1 -reverse-transcriptase-like-protein	41	0E-043-
10919	gi 30017567 gb AAP12989.1 -putative-reverse-transcriptase-[Oryza	35	2E-6-
11125	gi 242814149 ref XP_002486313.1 -reverse-transcriptase,-putative	62	3,00E-11
11330	gi 2415368 gb AAB70784.1 -protease/reverse-transcriptase-[Volvox	74	5,00E-12
11486	gi 242810815 ref XP_002485660.1 -reverse-transcriptase-family-pr	38	0E-41-
11701	gi 71411762 ref XP_808116.1 -telomerase-reverse-transcriptase-[T	32	0E-007-
12330	gi 220683899 gb ACL80762.1 -telomerase-reverse-transcriptase-[St	37	0E-67-
12795	gi 7248353 dbj BAA92703.1 -reverse-transcriptase-[Tricholoma-mag	45	1,00E-10
12801	gi 50402537 gb AAT76520.1 -putative-reverse-transcriptase-RNaseH	93	1,00E-17
12878	gi 1166376 gb AAA85635.1 -reverse-transcriptase-[Colletotrichum-	74	5,00E-12

Anexo 3 - Grupos obtidos com o termo anotador transcriptase reversa (Continuação).

Grupos identificados em bancos específicos									Grupos identificados no NR							
									Grupo	Situação	Grupo	Situação	Grupo	Situação	Grupo	Situação
104	346	1154	1682	2650	3962	5359	8317	10468	3	N	1149	S*(1)	5590	N	11125	S●
4	352	1155	1704	2657	4022	5411	8343	10487	103	N	1150	S*(1)	5680	N	11330	S*(3)
15	396	1156	1728	2658	4035	5503	8525	10526	114	N	1151	S*(1)	5815	N	11486	N
21	414	1157	1793	2715	4103	5620	8554	10798	115	N	1152	S*(1)	5906	N	11701	N
43	417	1158	1806	2753	4106	5621	8583	11027	267	S●	1153	S*(1)	6641	S*(2)	12330	N
55	423	1160	1809	2766	4108	6135	8601	11125	268	N	1154	S*(1)	7932	N	12427	S*(3)
59	428	1161	1818	2788	4124	6230	8670	11308	269	S●	1155	S*(1)	8065	N	12715	N
63	438	1169	1824	2796	4131	6289	8707	11330	270	S●	1156	S*(1)	8104	S*(3)	12795	S●
83	532	1195	1832	2807	4162	6319	8727	11438	271	S●	1157	S*(1)	8115	N	12801	S*(1)
101	533	1196	1840	2821	4201	6337	8758	11703	272	N	1158	S*(1)	8192	S●	12878	S●
133	535	1197	1871	2872	4223	6478	8773	11832	273	N	1160	S*(1)	8251	N		
134	547	1200	1889	2922	4224	6479	8783	12095	274	N	1161	S*(1)	8289	S●		
149	676	1201	1893	2957	4231	6480	8881	12328	275	N	1200	S●	8697	N		
156	684	1204	1945	2960	4232	6489	8884	12427	331	S●	1295	N	8707	S*(4)		
163	686	1205	1988	2966	4237	6495	8885	12438	332	N	1945	S*(2)	8766	N		
177	715	1208	1992	2967	4319	6504	8886	12443	333	S●	2094	S●	8794	N		
187	739	1209	1995	3039	4413	6531	8956	12795	334	N	2095	S●	9025	N		
193	770	1217	1997	3075	4494	6553	8969	12801	335	N	2096	S●	9035	N		
194	779	1301	2005	3143	4547	6554	9036	12878	336	N	2546	N	9053	N		
225	783	1328	2027	3150	4549	6557	9045	12884	337	S●	2579	N	9102	N		
234	844	1345	2030	3177	4573	6588	9128		338	N	2603	N	9346	N		
235	918	1350	2031	3179	4650	6632	9136		339	S●	2657	S●	9348	N		
236	939	1366	2059	3187	4670	6641	9185		340	N	2658	S●	9418	N		
241	958	1367	2094	3232	4674	6659	9196		341	S●	2843	N	9606	N		
245	970	1369	2095	3300	4676	6678	9395		342	N	2873	N	9617	N		
248	972	1375	2096	3336	4697	6709	9409		343	N	3101	N	9625	N		
250	987	1414	2099	3376	4727	6736	9482		347	N	3251	N	9629	N		
260	990	1424	2136	3379	4740	6785	9570		393	N	3336	S*(1)	9661	N		
261	1036	1447	2150	3392	4752	6824	9618		395	N	3384	N	9748	N		
264	1101	1469	2158	3420	4756	7008	9689		396	S●	3451	N	9783	S*(2)		
266	1108	1484	2181	3440	4802	7012	9715		397	N	3546	N	9844	S*(4)		
267	1117	1512	2203	3457	4803	7032	9782		399	N	3608	N	9967	N		
269	1121	1543	2228	3478	4852	7084	9783		402	N	3695	N	9981	N		
270	1125	1552	2248	3517	5011	7163	9844		624	N	3705	N	10023	N		
271	1144	1554	2264	3572	5070	7269	9869		770	S●	3881	N	10066	N		
278	1145	1555	2283	3584	5156	7435	9986		778	N	3959	N	10174	N		
292	1146	1556	2289	3636	5197	7931	9998		795	N	4106	S●	10197	N		
299	1147	1570	2293	3672	5265	8104	10250		938	N	4219	N	10215	N		
305	1148	1606	2314	3711	5277	8186	10298		1036	S●	4248	N	10445	N		
331	1149	1607	2425	3722	5279	8192	10314		1144	S*(1)	4852	S●	10533	N		
333	1150	1625	2429	3725	5307	8246	10330		1145	S*(1)	5181	N	10739	N		
337	1151	1650	2522	3751	5312	8274	10334		1146	S*(1)	5182	N	10778	N		
339	1152	1651	2528	3761	5316	8275	10337		1147	S*(1)	5404	N	10827	N		
341	1153	1654	2598	3770	5347	8289	10462		1148	S*(1)	5537	N	10919	N		

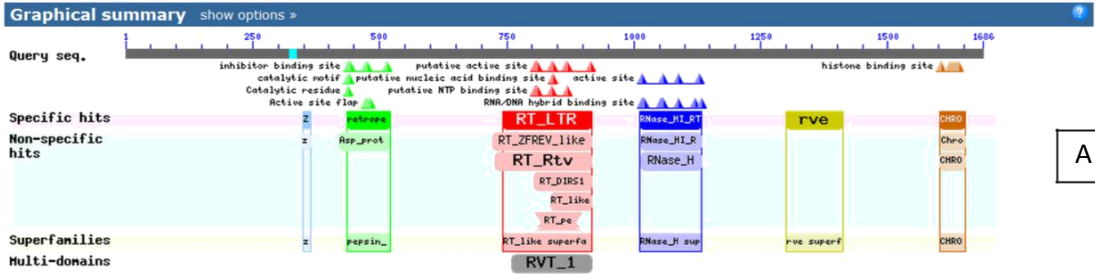
A situação dos grupos (S-SIM/N-Não) indica que os mesmos foram mapeado ou não simultaneamente nos bancos NR e especializados (TE e RepBase). S* (Grupos caracterizados); S●(Grupos não caracterizados); DBs(Banco de dados TE e RepBase); NR(Banco de dados de proteínas não redundantes do NCBI).

Anexo 4 – Relação dos grupos de ESTs que apresentaram similaridade com retrotransposons nos bancos NR e bancos especializados (TE e RepBase).

Conserved domains on [cl|25446]

[View concise result](#)

Local query sequence

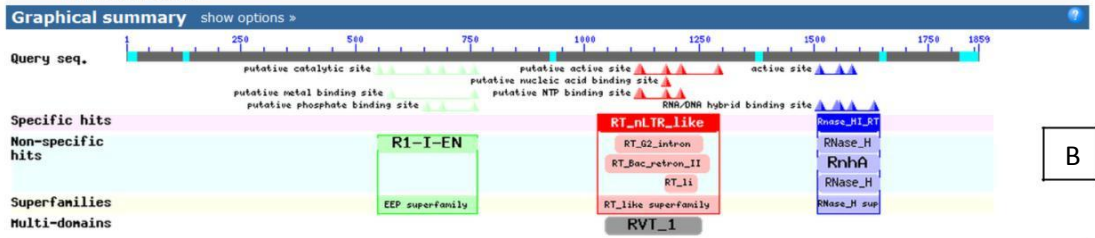


A

Conserved domains on [cl|25187]

[View concise result](#)

Local query sequence

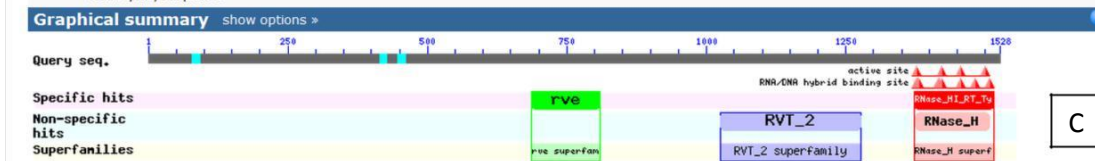


B

Conserved domains on [cl|18086]

[View concise result](#)

Local query sequence



C

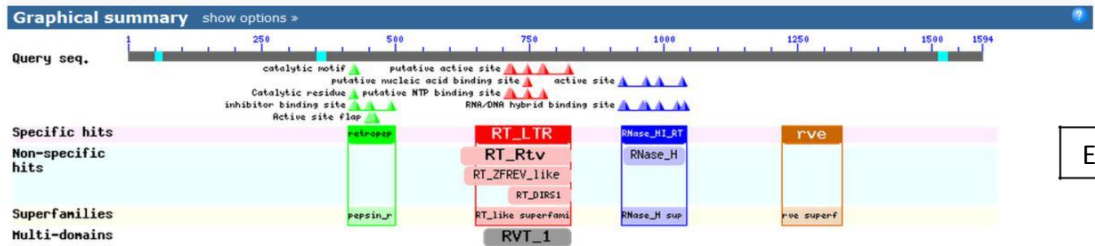


D

Conserved domains on [cl|19492]

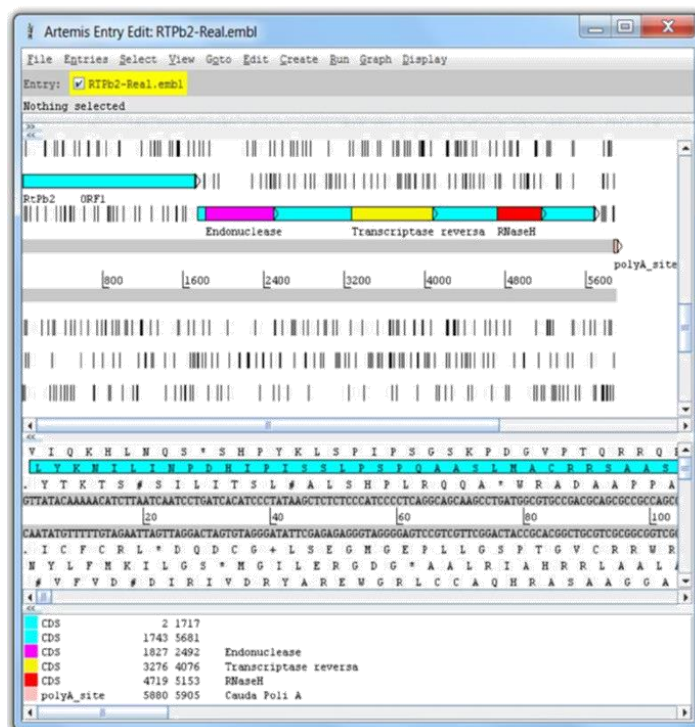
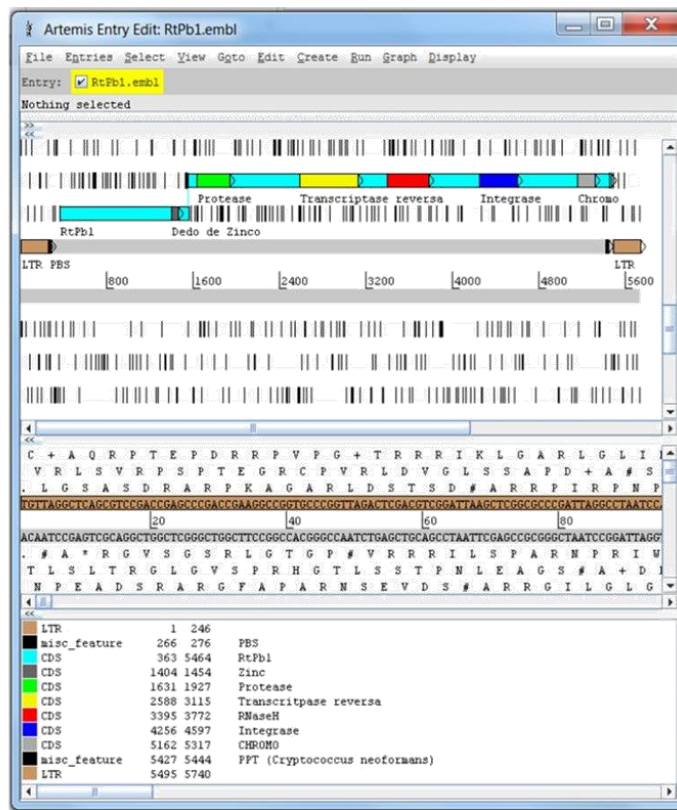
[View concise result](#)

Local query sequence

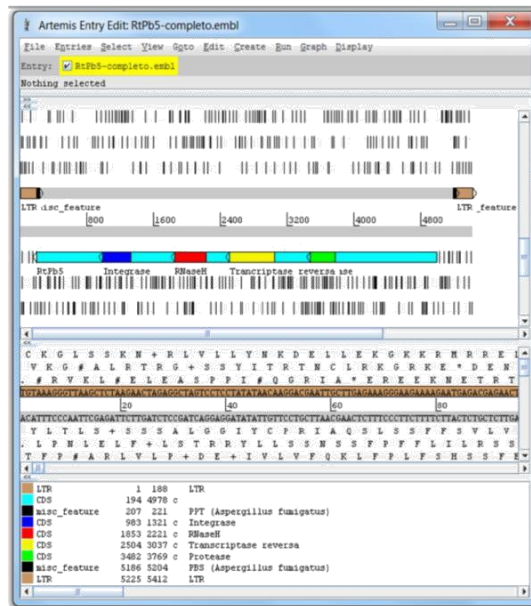
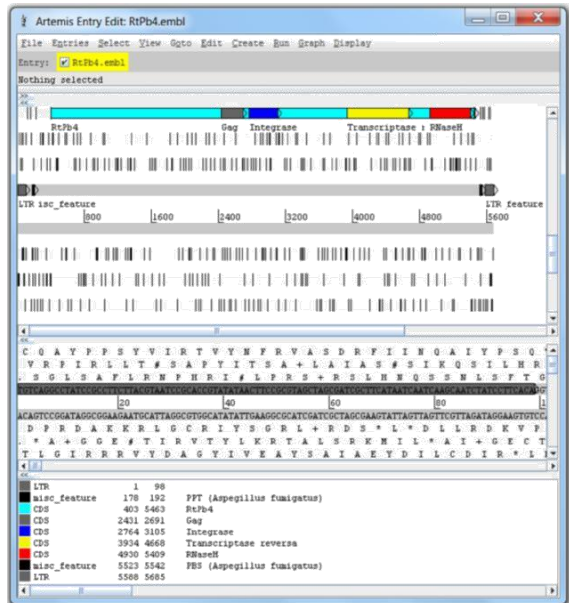
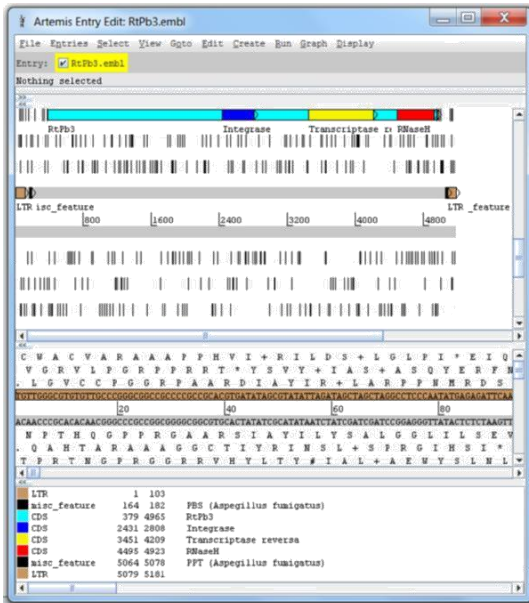


E

Anexo 5 (a) – Identificação dos domínios conservados (NCBI-Conserved domains) dos elementos RtPb1(A), RtPb2(B), RtPb3(C), RtPb4(D) e RtPb5(E).



Anexo 5 (b) – Anotação dos elementos RtPb1 e RtPb2(B) – Programa Artemis.



Anexo 5 (c) – Anotação dos elementos RtPb3, RtPb4 e RtPb5 – Programa Artemis.

giri

Home About GIRI Conference Repbase Repeat Masking Downloads Research

Map of Hits

[SVG viewer](#) is required to view graphical representation of the map as Scalable Vector Graphics (SVG plot).

1660.tmp/data.ori ([SVG Plot](#); [Alignments](#); [Masked](#))

1660.tmp/data.ori

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
1660.tmp/data.ori	1447	1496	Gypsy-23_ADe-I	3850	3899	LTR/Gypsy	d	0.8000	2.5000	242
1660.tmp/data.ori	2561	5403	Gypsy-3-I_AF	2797	5630	LTR/Gypsy	d	0.6490	1.7817	3560

```

ID Gypsy-3-I_AF rebase; DNA; PLN; 5726 BP.
XX
AC .
XX
DT 28-FEB-2006 (Rel. 11.02, Created)
DT 07-MAR-2006 (Rel. 11.02, Last updated, Version 1)
XX
DE Internal portion of the Gypsy-3_AF LTR retrotransposon - a
DE consensus sequence.
XX
KW Gypsy; LTR Retrotransposon; Transposable Element;
KW Interspersed repeat; Gypsy superfamily; Gypsy-3_AF;
KW Gypsy-3-LTR_AF; Gypsy-3-I_AF.
XX
OS Aspergillus fumigatus
OC Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
OC Eurotiomycetes; Eurotiomycetidae; Eurotiales; Trichocomaceae;
OC mitosporic Trichocomaceae; Aspergillus.
XX
RN [1]
RP 1-5726
RA Galagan J.E., Calvo S.E., Cuomo C., Ma L.J., Wortman J.R.,
RA Batzoglou S., Lee S.I., Spevak C.C., Basturkmen M. et al.;
RT "Sequencing of Aspergillus nidulans and comparative analysis with
RT A. fumigatus and A. oryzae.";
RL Nature 438(7071), 1105-1115 (2005).
XX
RN [2]
RP 1-5726
RA Kapitonov V.V. and Jurka J.;
RT "Gypsy-3_AF, a family of gypsy LTR retrotransposons in the
RT Aspergillus fumigatus genome.";
RL Rebase Reports 6(2), 64-64 (2006).
XX
DR [2] (Consensus)
XX
CC This is an internal portion of the Gypsy-3_AF LTR
CC retrotransposon. Two proviral copies are 94% identical to each
CC other. All 374 mismatches are transitions. Therefore, most of the
CC mutations was induced by RIP. Its PBS (pos. 2-13) is
CC complementary to Gypsy-3-LTR_AF (pos. 207-218 (self-primed
CC reverse transcription).

```

Anexo 6 – Alinhamento do elemento RtPb1 com retrotransposon descrito em *Aspergillus fumigatus*

Map of Hits

[SVG viewer](#) is required to view graphical representation of the map as Scalable Vector Graphics (SVG plot).

2160.tmp/data.ori ([SVG Plot](#); [Alignments](#); [Masked](#))



Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
2160.tmp/data.ori	1815	1909	I-1_AF	1775	1869	NonLTR/Tad1	d	0.6842	1.6667	327
2160.tmp/data.ori	2736	2782	Tad1-11_BG	2751	2797	NonLTR/Tad1	d	0.7660	1.3750	237
2160.tmp/data.ori	3135	3734	Tad1-6_BG	3027	3620	NonLTR/Tad1	d	0.6393	1.9798	1018
2160.tmp/data.ori	3963	4017	Tad1-15_BG	4355	4409	NonLTR/Tad1	d	0.7455	1.5556	245
2160.tmp/data.ori	4748	4864	I-1_AF	4696	4812	NonLTR/Tad1	d	0.6638	1.5217	239
2160.tmp/data.ori	5267	5332	Tad1-22_BG	5138	5203	NonLTR/Tad1	d	0.7424	1.5455	319

```
Tad1-6_BG rebase; DNA; PLN; 5779 BP.
XX
AC .
XX
DT 22-MAR-2011 (Rel. 16.09, Created)
DT 22-MAR-2011 (Rel. 16.09, Last updated, Version -1)
XX
DE Tad1 Non-LTR retrotransposon from barley powdery mildew,
DE consensus.
XX
KW Tad1; Non-LTR Retrotransposon; Transposable Element; Tad1-6_BG.
XX
OS Blumeria graminis
OC Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
OC Leotiomycetes; Erysiphales; Erysiphaceae; Blumeria.
XX
RN [1]
RP 1-5779
RA Spanu P.D., Abbott J.C., Amselem J., Burgis T.A., Soanes D.M.,
RA St??ber K., Ver Loren van Themaat E., Brown J.K. et al.;
RT "Genome expansion and gene loss in powdery mildew fungi reveal
RT tradeoffs in extreme parasitism.";
RL Science 330(6010), 1543-1546 (2010).
XX
RN [2]
RP 1-5779
RA Bao W. and Jurka J.;
RT "Non-LTR retrotransposons from barley powdery mildew.";
RL Rebase Reports 11(9), 2540-2540 (2011).
```

Anexo 6 – Alinhamento do elemento RtPb2 com retrotransposon descrito em *Blumeria graminis*


giri

Home About GIRI Conference Rebase Repeat Masking Downloads Research

Map of Hits

[SVG viewer](#) is required to view graphical representation of the map as Scalable Vector Graphics (SVG plot).

2705.tmp/data.ori ([SVG Plot](#); [Alignments](#); [Masked](#))

2705.tmp/data.ori 

Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
2705.tmp/data.ori	4497	4598	Copia-7_TMe-I	3977	4075	LTR/Copia	d	0.7200	1.9231	298
2705.tmp/data.ori	4633	4691	Copia-48_BG-I	4351	4409	LTR/Copia	d	0.7458	1.5000	275
2705.tmp/data.ori	4763	4902	PCretroS_I	4428	4573	LTR/Copia	d	0.6690	1.5769	287

```

Copia-7_TMe-I rebase;   DNA;   PLN; 4718 BP.
XX
AC   .
XX
DT   22-MAR-2011 (Rel. 16.11, Created)
DT   22-MAR-2011 (Rel. 16.11, Last updated, Version -1)
XX
DE   LTR retrotransposon from the Perigord black truffle genome:
DE   internal portion.
XX
KW   Copia; LTR Retrotransposon; Transposable Element; Copia-
KW   7_TMe_;
KW   Copia-7 TMe-LTR; Copia-7 TMe-I.
XX
OS   Tuber melanosporum
OC   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
OC   Pezizomycetes; Pezizales; Tuberaceae; Tuber.
XX
RN   [1]
RA   Martin F., Kohler A., Murat C., Balestrini R., Coutinho
RA   P.M.,
RA   Jaillon O., Montanini B., Morin E., Noel B. et al.;
RT   "P??rigord black truffle genome uncovers evolutionary
RT   origins and
RT   mechanisms of symbiosis.";
RL   Nature 464(7291), 1033-1038 (2010).
XX
RN   [2]
RP   1-4718
RA   Jurka J.;
RT   "LTR retrotransposons from the Perigord black truffle
RT   genome.";
RL   Rebase Reports 11(11), 2840-2840 (2011).

```

Anexo 6 – Alinhamento do elemento RtPb3 com retrotransposon descrito em *Tuber melanosporum*.

Map of Hits

[SVG viewer](#) is required to view graphical representation of the map as Scalable Vector Graphics (SVG plot).

4386.tmp/data.ori ([SVG Plot](#); [Alignments](#); [Masked](#))



Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
4386.tmp/data.ori	4031	4142	Copia-47_MLP-I	3279	3390	LTR/Copia	d	0.6726	1.6500	270
4386.tmp/data.ori	4298	4343	Copia-41_BG-I	3525	3570	LTR/Copia	d	0.8043	3.0000	225
4386.tmp/data.ori	4780	4832	Copia-6_MLP-I	3963	4015	LTR/Copia	d	0.7358	1.5556	234
4386.tmp/data.ori	5017	5099	Copia-51_BG-I	4229	4311	LTR/Copia	d	0.6988	1.7857	296

```

Copia-51_BG-I rebase;   DNA;   PLN; 4651 BP.
XX
AC   .
XX
DT   22-MAR-2011 (Rel. 16.09, Created)
DT   22-MAR-2011 (Rel. 16.09, Last updated, Version -1)
XX
DE   LTR retrotransposon from barley powdery mildew: internal
      portion.
XX
KW   Copia; LTR Retrotransposon; Transposable Element; Copia-51_BG-
      I.
XX
OS   Blumeria graminis
OC   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
OC   Leotiomycetes; Erysiphales; Erysiphaceae; Blumeria.
XX
RN   [1]
RP   1-4651
RA   Spanu P.D., Abbott J.C., Amselem J., Burgis T.A., Soanes D.M.,
RA   St??ber K., Ver Loren van Themaat E., Brown J.K. et al.;
RT   "Genome expansion and gene loss in powdery mildew fungi reveal
RT   tradeoffs in extreme parasitism.";
RL   Science 330(6010), 1543-1546 (2010).
XX
RN   [2]
RP   1-4651
RA   Bao W. and Jurka J.;
RT   "LTR retrotransposons from barley powdery mildew.";
RL   Rebase Reports 11(9), 2411-2411 (2011).
    
```

Anexo 6 – Alinhamento do elemento RtPb4 com retrotransposon de *Blumeria graminis*.

Map of Hits

[SVG viewer](#) is required to view graphical representation of the map as Scalable Vector Graphics (SVG plot).

4094.tmp/data.ori ([SVG Plot](#); [Alignments](#); [Masked](#))



Name	From	To	Name	From	To	Class	Dir	Sim	Pos/Mm:Ts	Score
4094.tmp/data.ori	476	604	Gypsy-1_ARO-I	5070	5198	LTR/Gypsy	c	0.6484	1.9545	285
4094.tmp/data.ori	605	1578	Gypsy-1-I_AN	3470	4452	LTR/Gypsy	c	0.6224	2.3241	1490
4094.tmp/data.ori	1845	3327	Gypsy-1_ARO-I	2218	3700	LTR/Gypsy	c	0.6831	1.8477	4596
4094.tmp/data.ori	3699	3771	Gypsy-1_ARO-I	1453	1525	LTR/Gypsy	c	0.6986	1.8333	267
4094.tmp/data.ori	4283	4633	Gypsy-1_ARO-I	610	963	LTR/Gypsy	c	0.6351	1.8667	500

```
Gypsy-1_ARO-I rebase;   DNA;   PLN; 5423 BP.
XX
AC   ABVF01000027;
XX
DT   02-MAR-2011 (Rel. 16.03, Created)
DT   02-MAR-2011 (Rel. 16.03, Last updated, Version -1)
XX
DE   LTR retrotransposon from the Arthroderma otae genome: internal
DE   portion.
XX
KW   Gypsy; LTR Retrotransposon; Transposable Element; Gypsy-1_ARO ;
KW   Gypsy-1_ARO-LTR; Gypsy-1_ARO-I.
XX
OS   Arthroderma otae
OC   Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina;
OC   Eurotiomycetes; Eurotiomycetidae; Onygenales; Arthrodermataceae;
OC   Arthroderma.
XX
RN   [1]
RP   1-5423
RA   Jurka J. and Kohany O.;
RT   "LTR retrotransposons from the Arthroderma otae genome.";
RL   Direct Submission to RU (02-MAR-2011).
XX
```

Anexo 6 – Alinhamento do elemento RtPb5 com retrotransposon de *Arthroderma otae*

tgtaggctcagcgtccgaccgagcccgacggaagccggtgcccgggttagactgacgctggatgaagctcggcggcggataggcctaatccatcacgctgaccacagaccatgca
ggtaggtataatccacgagcccccagcacccttccatccatcaagcttagtatttcgttatatagcagttagcaacgagagagctttcttctatgagcctgccttcgctcc
ctaacaacaatacagcttcttcttctcactcaggtttcagcagaaaacttgagctcagttgaattagaacctgctagttcagagcctgacaacgctgagatcaactgct
aagacaacagcagcttgaagacaacgctcaatcaacatcagatggccaccacagcctcccaacccctggaagggggcgaagaggatggcgatacctcaacaagcaactgacc
aacacagcgtcctcaatggaggaattcagagaggaatcactcggacagcagcgtgttataatgagttggcggcgtcttgcagatccttgaagcagaaataaaggata
caatcacagctgagggatggaacggaacaacgctcactcagcagcagcgtgtcaagcaagcagagttgagtgagcctaaatgctgaacgctgagccttagccacccctggctccgctga
cagtcagcaaacagcagcaagtcgaggtgtcaccctcctgagactcaccggaggtctgtcaagattgagtgaccacaaacacctcactgacggcaagggcgaataatgaaactgccc
tctccggatgaaacaagctcggggagaagcgtgaccactaccctcagcagcagatgaggtatgcatacctgaaacccggagctgaggggacggcagcgtcacatgcccctgta
tggaaagggaccaccggagcagatcagactgcagaagagatattgagaccctcaagctcaatataatgaggtgctaaacagctgcaaaatgccaagagtgattatcgcaagctgatta
tgcgcaatggggcaaatatcatgaattgtcacaataatcttacactggcaggtgagggcaagattgcccggggggacacaagacagactcaatgacaagctgtccttgacctcc
agaagatgtagcagtagccaatgagcactgatacatatgctgagttcagaaaaatctgtgcagggcagctcacacctccaggcaatcaatgccaccagaataaaggagctcgc
ggggaatgggctaatcctcacaagacaacacctgcccggcggcgaacagctcaaacccccagcctcctcaacaataatccccactgtgagggagggcctgagctctctgcccgg
atattcagtgctactactgtaaggaataaaggccacattgcccaggaactgcccctgcaagcagaagctccagcaggtgagttgctgagctcacaggggaatgagccccgagccttcagcag
atcgggaaagagtgagctgagagaaagactccactcaggtagttcaatagctgtgagaggtttgagatgagggactgattggggggaaagccttaatgtgccaatcaaatatgac
ctacaatggaaatgctacttcaacaatctcttctgctgatactggagcaaatggagcgaatttlatgatacccgatagccatagaatggcgaagattcttlatgaggaagtccaaga
actcccaatcaaatgcccagatgaaaggaatacaatggcggcccccgggggtgtagagaccgcaatgaccctaaacctctgggtggatggcggagattcagaatgtccctctgctggt
aacagacctaggccagcaccctgtcactctgggacggaagttggtggcagcagagatataatggctgagttaaaaaaccaacgcttggctggagtgaaagggctatccccgaa
ggtagggagccgagttgaaagatgtagcttggctagctgaaagcctcagaccctcaagcctgagcagctcagggcagatggaaacgagagagatcaagctggatgagagatcagaca
ggagcctcagctcagctccgagatgaaagggacagctcagctgagctgagcagggcagggcagctgggagggagcagcaccacagatgaggtatgaaactcaattgg
ggctgcttctccatggttatagaaacaagaagctcaacagcttcaataaccagctgtagtgaattgataaaacagcttggagagaagcggggttagtggagctgaggtgagct
ttagtggagctgagcagcaatcaccagaatggttgaagggcgggagagctgcttcaaaagaagcagatcagatgaactccccctacggaagaatgtagtttaaaatagagct
ccagcctgagttgacctcaccagggcattggacatgcccctctataatagatgactagaggaactagagggcagtgaaagcctcactgaaagcaactagcaaggggtttatcac
ccccagttcagccccctacgctcccccaatcctgattgcccgttagtgggaagaagctcagctctgctgtagcttgcgagctgaatgctataacaaagagatcagatccccctgccc
ctgtagatgagctgtaggacgactcaatggggcgaatacttcaaaagtggacattcggcaggggttccatcagctcgaatggacctgagctgagggactgaccacttttag
aacccgctcagggactataaatacgggtggtccctttggccttcaaacgcccctgctgcttcaacgggtttgttaacagttctctttgactaccggacaagttatgacagc
atcctaggtgagcttctaaatctatagcaagaatctggagggacacaagcagctgaaactcagctgtagaccgactaaaagacagcgggttgcgaagcagcaccagaatgtaggt
tgcagctcagagggacaaaatactagggttctatcagtagcagagggatgagggcagaccagagaagcagctgcccattgtagagtgaggccccaccacagctcaaaagagctca
gtacttctaggatttggcaatttctaccgagcattcataaggaactacagccgaatcgaacagcactgtaccaatgacaagaagactaccctatttggtagacatcagaatgcaa
atcagccttgaaaaaatgaaagggcagctggtggaggttccacactgcttactatgaccctgagcgggagagctgattgagcagcagcctcagcggagtagtgctggggact
aagtcaagatttgaataatgaccttggcaccagtagccttctattcaagacaatgaaatggtgagagcttaattatgaaattcagataaggaatgctggcagtagtccgggctt
acaagaatggcggcagagcttggaggactacaaggggaggaagcattcaagatcctcactgactcagcagctggaatctctcagcactcaagaagctgaaatgacagctcaagccc
ctggcagaggttctctccggttctactttgagatcaagtagcagcagggcaagggcaactgctgagcagcagcactgagcagggctgagcaggttgaagggcaacagggcggctat
gagatcctcctaggaccagagaccctgacccccagatcactccctctacagacattgctagcactcaactcctcagagaagagagcattgctgaatgcatagttgcttaaacagac
agccactcaatgagccctcggccagcagggcagatgagagaatgctccatacaatggaggaactgctacggtacaatggaagcagtagtggtgctgctcaggaaggaagaaac
ctgtagcagcagcactcagaggaatcagtagcaagcttcaacagcagcaccctggagggcgaagagaccctggcctggtgcaatcagcgtattggtgggcccctggcggacagagct
ctcagtagcagctgcccgaagcaaacaggttatgagcattttggtcagtagtggcggctaaagcaagaagcaatctccattccaacaccgggggagttacagccagggagctggc
ccgacttctcagcttctactattcaacaatggggcctgactcaattgctctgagcagggcctccagttcattcaagctttgggagtagttacacagatcctgggcaaaa
actgaaacttcaacggcagcactgcaacaactgtaggacagcagaagttgtagtcaatcaatctgcccctctctgcccctatgtagttactcaacagcagctgggtaggtgagtg
gctgcccattatagaccatgctcggcagcagcagcctcccaaacgcttccacagggatctccccctcctcgcagcaaatggatgaaacccgaaatgcttagctggcaaacctggcggg
gaaatggagcagatgagagaagaaatcagagaagcagcaagggagcttcaatcaccgaatgagagagatctggaaccagggcagtagaaagcacaagcagcagcaggaacagatgaggg
gcagactgacaagcaccatcgagagcctgacttgcagtaggagatttggttatcacaatgaaatcacaagatggatcgaccagcgtaaactggcggagcagaaatcggcggcc
ctcagaatcctggaataaggtgggcaatgctcaaacctggactccccatacaaatgaaatctcaatctctcccagacaagctacgcgaagactccagagaccctcacttgg
ccagactcagaccaccagaccctaaatgaaatagatggtgagaatgaaatgggaaatgagctgatactggcagtagaataagccggagcaagctgagatcgggtgctggaaggg
gtttgtaggagatttctcaatggtaccctcctgagcttcaaaaggtcaccacagcctacagcagcttccatgagggctaacctcaaaaagctggaaccccaagggcagctagatgagtg
gctgaagcctgggagacagacagctaccctgaagcagaggtttagtagcagctaccggcagtaggagcagctgcccctagaccggggggtaggttagctcagcgtccagcagggc
cgaccgagggcgggtgcccgggttagactcagcgtcaggttaagctcggcggcggataggcctaatccatcagctgaccacagcactgcaaggtggttataatccacagcagggcccc
agcacccttctacccatcaagcttagtttctctatagcagttagcaacagcagagagcttctctctatgagcctgactctgctccctaa

Anexo 7 - Sequência de nucleotídeos do elemento RtPb1 (5740 nt).

gttatacaaaaacatcttaataactctgatcacatccctataagctctctccatccctcaggcagcaagcctgatggcgtgccgagcagccgcccagccgcccccaac
accacgaacgctcccgagcttgacgagcctctctcccatacggcgtccaccgaagagaccctctctacccaatcgcccggccacatgtgaactgaatgacatggagtc
tccaagctccctacaccagcagcagctcaagagcagaatgtcgacaaattcaccctaacctacgtggccggtgagcagcagctctagcggccagcagggctgagcagaga
atcccatgcagaagcactagcagctgcaaatatgcttgatcggtgaagaagacctagatgtcatgaaggcaaaaggcctatcaaggcactggccttgaggccgtggaaggcgtt
agtcaggagctggtacaatggagctgagaacagctaaagcagctccacacaacaaagctgctggcctcactggtgcaaatgatgcaagagctcctaccacgcaaaagcagctgcga
cagcacaaatgtaaaaaacaccagcaacatactgcagaaccctgcttaggaatacccagagaaggaaaaaccacgaatcctcgccgctgcttcaagatcatctagctcg
caaagcaagccaattgctactcttagaagctgaaaggctgaactccagggccttggcagcagcctacacagcagttcaagcctccacagggctgcatctctagaagctt
gcaagacatgaaggcctgtagcccaaaagaatgttttagcagccaccctaccaatctcagatagagaaggaagaacattggatcacctgctgattcctgggctgcccattggagat
tagtgactatgggtggaactctgctctcagcagaatccatgcagaagaattcactcacaaccagaatcaagcctgctcagagctgctgggcaaaatacaccctggaacc
aacagcctcttaatactacatcttgcagagacagcaataaacctggcctcaaaagtggaactatggatcaaaagcagctgacagaaagccctaaagctcaaaatacaaca
atgcagcaaatgctgggatttcatcaagagagatactgctcctgctcctctcctgctcaaaagagcagcttcaagaatcatgtaaggaagaaatccatctgctcaagagagaaatgaaac
atgggatgcccgaatgctgactaaactctgtagccacatgcccggatgatccctctgcccaccagaccagggataaagaatgggactgttcaaaagcctcaagaaccagct
gagggatctcgcgaatgcaagagatgagcactactggcgaagcagtaaacgcaaatgctggctgctcaagggcctcagcagacagggctggaatgatctcctgagcgtagatcaa
cctcacagacatacaacacagcaccagaacagcattgagcagcagcattcagcagcagcattaggcctggcccaaccacaactctaatgcaacccaaccacagctctctcagcaac
aacacctctctcagctactcaatgaagaagaactctagatctcgcaatctcctgagatgtaagaagcaactgtagaagcaagcaagggggccagcacagagagccctcaatctagcgt
ttgaagaaaagctgacatccctctcatacaggaaccctggattcacaggaacttaagcaacgactctcaaaaagcctcagcactcaactgcttggcccattgaaaagtgggaca
ggagccagggctcctcactatgtgataaacccccacaatcctgctctgagctccagggccctgggcaaaaacgggacatccacatgctgcacatacagggctggggtctac
agcttcagacagtaacataatacaatgcacccccagggcagagagctctggacaaggtggtggctgctctctcctggctgctccagcactcctggctgctctgctggcagctca
atcaagcaccctggtggcaccagagagctcacaagcaggtggggaacacactcctgcaatggaccatgaattcaacttctcctcaccatggaccagacactccacacagctg
ggcataacatgatagactggcctgggcccacccgggaatfagtatacaagaggtggcagacaggtggagagactccttcaactctctgaccacagactcttcacaaaataca
gcctccacagggcgaagggctgacagggcacaaccctttctcctgctgacacatgagtagaaggtcttaaccagacactggagggctataagccaaaatagacaactatgctg
atctccagagtagcccaatcacagatcagctggatgcccctggcctaaaggtatcacagatgcaactaatgctcctggaaagttccacaagaagagcagggcgaagcaggtcagc
catggtggaatggaggctgcccagaagcagtgacaacacatcatctgctcagcagcactaggctctagtgttttagagaagcaaaaagcaatgtagcaggaccaggaagcaaga
gagaaactgctggcaggaagatagcagagggcctcagtggggaagatctttagagatgcaaatggggccttcaactggcctcttcaatcccccctctgtgatgtagaagactg
ggggaacgcaacaacccttagagaagaagagacctctgcaacgggctgctttagagaagctgcagagagagagatacccaactcagctggtccaaccacacatctgcaaattac
ctttctgctgctcagcactgagatacaaaagtctctgctgaacagggtaaacactccaggaactgtagaggtgcaaacagaggtgctcagaaagggatggccacataatgaag
tatcagtggaagggctgttcagagctgcttagggctggcggcaccacactcctttagagatgctcactggctgcttaggtcaacagagatccaacacctcccgcagctct
acaggtctatagcctctctcaactgctggggaagggactggaacgctggttagccctgagctggcctggagcagctatacaagaagaagatgctcaccaccagctttagagccctcc
ctgggggggctgctcagacatggtgagcagatggttcatgatataagagaagcctggggcctggaagagtagctctctcttctcctgacatcaagggctctttgatgctgctac
tcagaggaagggatgctgggagattgcaagaacaagggctggccacaaaatttaattagatgggtggaagcttcatgacagggcagcagggaggtcagattggatggactcctgggag
agctctttgatctccagcagggcttccacaagggctcgccagctctcactatcttctctagctgttttattcagccctgcttcttcttaggaactatacaagaagagcgtgcacgcttag
gctacgcagatgatctaggactgcttagtctggaaccactctggaaggaatgtgttactcctcagggagatttcaagttgctaaatgattgggcccagcaagaagggctgacatttg
attttggcaaaaatgagatgacacactcaccagggcagaaacccttcaaacctcactgactggaacaacactggcactcactctatcacagaaagcagtgaaagctcca
taagatcttaggaatcctggctagatgaaaaactgagctcagaacaacgtggagacctggggcctagggcagcaggggtcttaatggcagacagctaaagcaacacagcaagag
gcgccccagtaaacctgcttagacaatcattcagggctgctgctctctgctgctgcttagggagcagagggcctggggcctggagccacagaaatggagaagggaaatccca
atcaaatggatgcttagtcagaagctagatctagcccttagagatgctttggagggggcactggcagctacaagaccacaccatctcagctcaccagtgagctagaatctagc
caatgaaagagacactagacaacgggctgctgattgtagtaagaatcaaaagcttagacaagagccagcggctagtgagagccttaataaagtagctaccacaacacatcaaca
cagctgcttagagcgtccagctgctggtggatagaacatagacccttactcactgcccctggatgagatctccaggaagaaagctgtagagattgggtatgagggcagggc
gcacaaaggaacagggcagcacaagacttaagagctggctacaacaacgcaactcccctggatctggtggctttagcagatggttcacagatagagaccctcagcgtgagctggtgag
gatgggctatagctggggtagcaaacactcactaactccacaggaacactggcaatggataaagcagagggctcttgcagcagaggaatggcagccctgagggcctaaaggaagca
ccacttgcctcagggcaagttgcccagcaacatcactattgcttagacaatttggaggtagcagagccctggatctgctgcaaacacctctagcagcagggctctccgagctt
tgaagccataaagggctggcagagcagggataggtcccatggccacccagagcaggggtggtcattagatgggtcccagggcagcagggctgggtggttaagaaaagcagactttg
aagccaaaagcagccgagcagggcagcagagtggaagctggatcttcaactgacagatagggcacttagcttataaagcagacacatggaacaacacacagggccttggaga
actactggggcaaaaatcgcccacaagctataatggagctggagatcccaactaaagagagcctcaacagaggtgaaacttccagcttcaactaggaagctctatgctcggcctgag
gtcagaaagcggccggcctatcaaatctccgagagcactggcaactcgggtggggccctagaatcaataactggtcaagagaacagccttctatagaatattgctcaaatgc
gcgtgcgcaactccaagcagctccacagcacttctctctctgcttggcccactctctcaccacct
ctttctacacctcctctcctcctcccaaacctaaaacctcaaaacctctgctctatggcaacagaccatgggctctgctgggtgagcaatcagcaaccattttgctc
atgtgaagatagctactagataggtgctgctacgggctagctctaccggcctaagtttgagcgggactctactagatcagctgcccctaaaggcagctggactaaccgggctta
aaaaaacacacacacacacac

Anexo 7 - Sequência de nucleotídeos do elemento RtPb2 (5905 nt).

tgtgggctgtgtgtggccggggccggcccccggccacgtgataatagcgtatattagatagctagctaggcctcccaatatgagagattcaaccatcccacaggttatggccggatag
cagtttttttaagcaaaacagacgttttaattagaatagcgtgcaaccaacgcgataatggatgagttcgactagatttccttataaaacgctcttttacactaaaaaacaagaat
cccaaggggatactctctgttggagaaagcgatcgcgtccggaaggaatcaaaacaaaacatcattttctctttttcctgatttctttctttaaattttgatctgaggata
ccaagtcttggcttagatgagtagatcaccctactcctcaatcagaacgaagagatctctacgttcccatcaaaagcccaagagacgttccctacactcagagaagggtga
gtggatcaggtgcccacctcaggagacgaaggagaaccgaatcgcgttagtggaaagcaatggaaagcaatcacaacccgtgatacagggaaggccgtacgcgacgaatca
gaggtgacgggacccctagcccaaatcctggccctacccaatcctaccccaatgccaacgcatgatacagatgaccgggccaatctctacgtttttgcaacagatgtagacc
gcccggggaacgcctcagtagtactacccctgctgataccctaaacaggagatcagtgaaatacaaaagggtgcaagcttcgatgacgaagaccgtgacgacttgat
ggaaacactaccaagcagtagggatggcatcctgggggacggggatattgggcccactgatatccttataaaaaaacacatgaaccctgaaagatacagaccaccaggttg
gataaagcgttggacagtagtacaaggccttataccgtcgaattgtgcagctccctactgggacctgtacgtaacaagatggagatagaaagcggataacgcggctgcccattgg
cagaagatcagccagaaatcagctgtctcattggcagaggacgtatgaattgcatgaaggagctctcaaaccttgcgtcaaggataacgactatctgctaccaaaaggagattgccc
tacctggtgacagcgaataagggttagccacacaagatgacgattttatcacgactttctcttattggcctacgtgatacagaagcgtttgtaagaccctgctgtagacttc
tacgcaactggccaagatccaatcgtgaaatggatattgacgatttggatgaaagcgttaaccaatcgtacaagcaagcaaaagagtcgatacgtccgaagaccttaagggcaacgct
gccaccagaaatggcgataaacaagcagccaatcagtcgactactcctcgaatgtagctgtttgtgacccgctagggacaatacaaaagaccctgctggccgacaccggagttagcc
accgactcagtagggagagcgaataagataaagatcgtcagagaaatccgtagagaaatggggagcaagcaaaagcaataaagggaagagtagataaaggatgacaaaaacaac
gagaagacgcagatctccttgcctcacaactagtgccaacggcgcgatacgcactatccgctcgtccgagatacacaacaaatggatattggataccctggccagctccaatag
agcggcgagagcattcagatgacgtacgtgcttttgaaagtaattcaaaagtttaaacggcggcgggagcagggagaggttaagaaagttggggatctcctcagaaatcggat
gggggggatctatggtaagagacgtccattggatcccaaatcaacccgtcaacccgtgtgctgttcgcgaagctagaagaccaggtttccgattgaaatataccccatcaccctct
tattcattatcagcggccaacaggtgctcattaccgctcagcaacggccaacagctccaatacaagatcgtcgaagcagaaatcgggaattgagctccaccatcgtccaaagc
gtcaacttcggcccacaatcctcagatcacggagacgcgagtggtgacggctgagaaatcgtggcgataaaacgcaataagatattgaaaacccctcgcggaatggcaggttgacgtgggtoat
atctgtgtggggcgaataatcgtttggcaagaaatccactatcagggctacgaatacaaaagcccaaacatcaggactatctgacgacccctgctcaaaagcggatgacacaaaaatc
tcgacagaccgagtcggcagggcgctacgagcgatggaacgcgtgcatatcgtatcgcagggcggcgaagcctcgcagagatgaacgatacccccattgaaatccgcggaggggacg
aaaattacatgctaataccagatgagccacccttataaggtagtctacttttgatgcaagggcaacagctgctgattttaccgcccactgggtccaattcagaaaaacccctggt
ttcaccggctcgtcagtggaacggatattgacggcgtattccgttccaaatagatgctacgattatgaccagtcaggggagcattgggagcctacgacccccatcccccaac
caagcggcgtatcggagcgccgcatctgtagcgcgtgcccggcagtccttattgggaagataatccctaaagagttctggggcgatattttgaaacggtagtgtagatc
acgaaacaacggcctacttccactgtcttgataatcgaaatttggcctatggaatacagatcagatgaaatcgcagctgtacgctgtcctcactacgataccgttctctgctcctaaac
gtcccggagaggtttgcaactcggcgagctcggatcggagcgttttagctcactcaccggatcgcgaacacggcggaatagagctgctggcagggcggaattgtaagcttatggg
tttcgctgaaatcaatttttctgttggatccccggacagcctcctgataccaggtgtagtatttccctcaaaagagcgtacgaaatgaaacaaaatgtagaaggactagtg
tccaatcctacgacctacttggtaggtagggcgaattccccgaaatcgaagacgcgaaatggaataagacctcgaataatcgttgcctacgtacgttccctctcctatacaggttcattg
ccaagacgtacaaggatgcccgtgaaatcagaccaggttagggagaatattggaagaaatcaatgcaacgtgagaatcgcagatacctgaaacagggaaaacctggaccctaaatccccgagatca
gacgtcccggcagacgaaccagtggcaccagggcgttggatcttcgtagaaagatcgcggcgaacggcgaaccccaacgcgataccggagggatattacgaaaaatcgtgggtagcgtgc
ggtaattggttgaaaagggatcgcggagagacgtacgcccctgttggtaaaatgacacatcgaatacgtatgagttcgcgactcagcgcaaaaaggaatggcactacgtcaagtcgat
attgtagtgccgatctcctatggcggactgggacgaacgaatcattatgagacagcccccggatgtgaaagcaggagctcccacggacattgggtttgtgctgtggaaagatctctataggaa
ttggatccggcagcgaagatttggtagacttactgacggcattgtgcgtagaaacggatgtgcccctcctccctacgccccccctatggcggcactacgaccctgagccactctat
atgacgatcagctgacgatcttcacataatggcggaaagcgctgcccagccgattggcgtatgaaacgttagcgagcaagttagagatacaagatcacaagagcctcgaatatt
ttgggtatggaggtgaaacgtcgccaggatggcggaaatcaaaatgacgctgaaacggataatcgatcgttttagtcaagatattccactgggagcagctgtcaccgttctgtagcccaatg
gaagaagacttcgaaatcagcagactcccctgtctcccgaaaaagaagcgtgtttacaacacagcgtatacaaacggcgttggctgcttacaataatcgcgggtcaaatcagctccgagat
atcggccgtgcccgatcccctcctagcgcagaaaaacggcgttccccccccaaagcgtgaaagcttactcaccgctcgtacgctatttgaaggaagcagcagatcagggatctcctaac
gagggggcccaacgatcgtgagcggcttcaatccccaaatgctacgtttacaccatggatatacggatcggatgggagggcaccctgaaatggcagatcgggaatcagctcaagtttc
gtctcttctgtagcgggtgccccgatttcaatggcgatcttcaagcaaccttgcgtatctttagactcaacgaagctgaaatcagtagcgttagtgaagcgtcagcgggctacgtgg
attcgaatattgattgggataatgggatttattacgaagggaccctcagatattggggcgataaaatagggttcaatcgaatcgcacccgtgcccagcagataccgcgcgacgaacgat
atcgacactagatccactatcccgtacgcaatgaaatgagggggtgatttcgatgaccagcttccacagggagatggcgccagacgggctacaacacccctaggcacaagtgcca
tttaaacgtttatcagtttgcctaggtcttccacaaggaataatgctcactcagcaaaatctgtagtagcgtgttccgttgcactcaggaatagatagctagatttgaatcagc
gaggacataatccccgacgttgggtgctgggggggttggggcgtgtgtggccggggcggcccccggccacgtgataatagatagctagctaggcctcccaatat
gagagattcaaccatcccaca

Anexo 7 - Sequência de nucleotídeos do elemento RtPb3 (5181 nt).

tgtagcggcctaccgctcttacgtaatccgaccgtataaactccgcgtagctagcgtcctcaataacaagcaatctacctcacaggttatgagaccggtttgcgc
aaataactgcaatccaggggaattctgcaaatatagcgtataataactcgtactcgtctcaatcaataaataatgctctattttatgtaatgatcgatttgggtgctcttata
ttattttttccgctgttcaatccgtctcgttagcgggtgaaaaactcgtcaactcgtcaatttttcagcagctgctgcacagaagctgccaatccgctcaatcaagcttatagcga
aatctcgcggaataaagattactcgttataaagaagataaacgacatggcggaacgtatgtagctcgtcgcggtcgcgattcgagcccaatccagccagctgaccgatcgaaat
gaccaggggggctgagctaccgacccccggggatccaagattccaatattggacccccgaaaccccaatcggcaagccgggctgattacagccccaccagctgagggccg
gtccgctacgtcgtgctgaataggcgaaccgactccttcagaagatgggggtggagacccctgcgggtgaccccaactgcccggcaactgactccgaagaattgcaactgaggaga
ttggatctcgaagctcaaaagattgattagaactcaaaaagagggcgcaaaagattgcagctcaactcaattggaggggctccgctcagaatgacagggccgctgcccga
aatcccgctcgggtcgcagctcagcatgggacccgaactggaggtccggaatgcaataaacctctacgaccagaagccgctcggcaagggcagccagctcaaacagggact
aagtatgccaccaagccagcctcgtgctcgtatctcgaatctcctcctggcagcaaatgaaagcccaagctaatgacgctcaatgctgggttatattggaggcaagcaagtc
aaaagccctgatggagacccccgaattgggctcttttggggggcaaaaactcgtcggctttacgcaatcaattcgaacgctggcagccagcaattgagccccggttataaagatgat
gagggcaggaatccttacccttggcagcgggtgagggaggttctccatccggagaacgagctctgacgggaagcagctcctggcattacatcaatgggtgggtgccaagctact
gacgtacacacgcttttctgataaagtcacatgacgttatatgaaattggaaatctcgaattttactgcaaaatgacgtggtatccacgtgttctattctgcaactcccaatataatg
cgaactacgtcgaagaaaatcagaggggtacaagagtcgaactcgaattctgctgttggagctcattcgatcattgagggaaatcagctctcgaattgaaaccccccaagaggca
aagaagggcccccagggcgttaattctcggccaactggctcaggctgaaatcgggctcgggttcccaataacagcagcaactcgaatcaaaaataatccagccccggggggccgaccg
ggacgggaggtatcgaaatggcgagcgttcaaatctcgacctcccaatctgcccggaaatggagcaatcgctcaatggcgggctcgtggggcagcaatccggatagcggccccgaagaa
tggagaccgccaacggcgagaaatgggaaataataataagaagagtagtgggggcaaaaagggcccccaactgcgaatggcagctccgagagcttcaaaaaggcgaatctagtg
caggtcaattgggggtgctgctactcgtgcaaccaatcgcaacccgctcaccatattgggtagcggggccggatgcaataatggcgtgtagcagctcggattttcccaaatcaagag
gttagtggccctcagatcaagtggcgaataagggctacttgaaggttcaagggaatgggggtgcctccatccagctcggggaaagattaaatcgaatccgaccagctctatttgggtgcca
ggcttaccgcaacccatataatccgctggcctattggagccagggctataattcggctgactcgcaggtggagcgggaaatcacttgaagataatagcaccacaggaacgacatc
ttctacgcaattgaaatcccaataataataatgctctacgacacccccctatgttcgggggtattatgaaatgatctggagctggaccacatggcaacttggcggtcccgaagccc
agggcatttcaacggatataatccgaacgctgcacactgggtctttcaaccggatgacgtggagcagctaatcaaccgaaggaatccggataccgataccttcaatttggctccca
ccaaccgactaccgccaatattcgggaagcagctaccgattgacaggtggcaccggcgtatgggcaataatggcagggagaatcgaataaaatgctcggggagccccgttctggaatc
gcaatcactcggcccgaagaagctcctgctcgcgccagctcgtttaaagccaagcaaaaactgcaagttataaaaaacccccgctacggctgcccaccgcaaaaggtcttaagatctgg
gtggacctattgggggtggagaaactcgtgtaataagaagaatacaccatcttattcaggttaagaagctcgcgattgtagctgtagtaccacagaaatcgggtggaccttctata
ttggagcccgcaatgaggtctcaaaatataacggacctcaaccaatttatgaagaatctgctcaatcgcgggtgtgcaatatacctctgacggcgaggtgaaatcaaacgcc
gcagccgctataatgagtgaggtatataatgggaaccaacccccccccacactcgcagcaagaatgggtgctcgaacgcttctatggcctgattatgcaaaaagccgcaactta
ttaaagattcctcctccccccccagatggtggggcgagggcgttaattggcgccagccgaaatgcttaaccttatgcaacgctcagcaagctctataacgggtgtagaggatccggatccg
aagccggttactcctataatgagctggctcggggccctcctcgcagcggggatcgcgcatgggggtgcaagggctattataaaatagtcaccaagccgaaatagaagatggcctc
cgatccgagagcttataattacgggctatgaggttactaataatagaatggctcgaactctcaacggctcggctttacagctcgaagaagcgtgataattcaagaagatctcgaattg
gcggtactcctggccccgctgagcaggtgcaaaagcctgctcgcggcccgctggcgcggttgcgactccctcccccacagcgatatacaacagagcttcaatggtggac
ggggagtttctctcgtcccctcgggtatttctaacacctcaaaaataatgggtgtagacttagatctgctcggatattccgcaattttgggtgggaacactgcccctccaggg
tccgcttagagacggaggaatcctcaatgccactgagaatataactcagttggggaggcctcagttggggatctaccgacaactcgaactgacgtcccccgagccccgagatcgc
ggtttgcaagtccgaagaacctcaccgggctcgggctcgtcagattgaaataatcttcaatgctcgttaaggagataaacccctcaagcagatgggtgcaatggcggttagtt
gaaagaccacagataatggcgtatctcagtgcccaatggggttgcgaaaaaagataataatggggtagaaatcgaccgataaagggccgctgggttagctgcgcaatatacaagaa
aaggtgtcaactattcgaagacatctgcaccagttgatacggcctccactcccgagcactatggcgaatcgggtatcccgaggaatggcagcttctcgaatcgaatggttacggca
tctttaaagcccgctcccccaatgaactataatatacgaaccaaccccggggtcccgagggagcggcgatctggtatgcaaatgtagcagggactgattggcctgaagcaa
gcagcactatattggtatgaggagctgcgaaaaatgctattgaaaacaggcttccgctcagctcctcagcggcgtttttcaataatgacgtccgcaaggtctataataacatctat
gtggagcaggttcgctatagcggccgaatataatcggtatgaaatggcaaaaagccagatcgccgctaatatgagatcaaggatggtggcgatcgaaccgggtatttgggtatgaag
attgtagggatgcccgaatcattatcgtctcccagccatgttcaataaagaactcgtgaggaaatttcaatggatgactgccaccccaaaaatcctcccctatggaggagaatggtgat
ttcatgagcgaatacgaagaaccaaattgcttatacagcggaatttaccattacgttccatcaaacgggggtggctctctgcaatactcgtcgcacacacagccccgataaagctc
gccgctgcttctcagggagattcagctcccggcctaccctgaagaatggggcgcatccaacgaattcttagatactacaagggcactcccagacatagggtgtcttaccgaaagac
agcagacgggctttcgaatcactaccgagccgaataatcgcactcctcggggcgccctatgagctgtctcgtcccggaaatgctgaggatgcccggctcgtgctcctcccgc
ctcaataatctcagatgataatgacggggcagatcgaaatcccgcccaataagctagcgtatagcattcggggcggaataaaaaacggcgatgaaactggcccaagagtgaggtctccgc
cgatcaagcataagaggtccgcttccatctctcgggattctcgtccgcaagaacagagatctcgtattacgtcccgtcccgcaagacggcgagagggctgaccaacggcct
ggggcagctagatagccctatggggcgccgacttagggcgcggttaagggcggaatgattatagaggaaatgactggttttcaacgataatgggttttccagcaaatatgt
accgatacctagctacagggctattttttagtataatctcgcgtacagttacatcggtagtgggggtcagggctatccccttctacgtaatccgaccgataataactccgct
hgctagcagctcctcaataatcaatcaagcaatctactcctca

Anexo 7 - Sequência de nucleotídeos do elemento RtPb4 (5685 nt).

tgtaaagggttaagctctaaagaactagaggctagctcctctataatacaaggacgaattgcttgagaaagggaagaaagaaatgagacgagaactcgagcttaagcaagctcgcgtagca
tgaatcagctgagcagctgcaaaacgatgcttaggctgagctcagccagccagagcaatctaggggtcaaacccctataactaccoccttctccctctttgctgagggcctttacgactat
ttaaagccctctttacacctaatttgcctgctgagaggctcccaagcgtcaaggccttcgctctctgtaggagactcgcgctatcccacgtaggctcagcccaacctaccatttaa
cgaggatgctctttctccggtaggtctgctccgctgaatttctcaatcctcaaacctcaggtctctccggggtctctatagcgggtggctgcccagctcagttgctctgactagga
ggggatccgctcgccagcagatgagaagctttacatataaagacattgtagaagcgggtggagcctccagctcgtaagcaatgggatcctatcttctagcacagatattgtagtatt
tcgctccagctcttgacagggcatttagttctcattgctctcaagctcaaacagccttatctccaggttgaagctgaaagccgggtctctctgctatctcgctatctctctgtag
attgctgagcactccgcaattttacttgagcccaactccaacacttcgagggcctctctcattgataatccgcaaacctcaatgggatcacggttcaaacgggggtgcatcggctat
tatcaatgggagagcgtggaaggggtgctgtaaaagaaacggggtcaccacccctgagggcagcactctattatgtagggcagctcgtcagaggggacattcaccgacgatctct
gttgaagctaaagcttctgcttaggttaggtttccagggctgggtcattcgcctctgtagctccgctcagctctctgggtgtaggctgtaggagctgctcttaatgcctaatctctac
acaccgagtcacaaactgcttgtagaagctgacctcagctcagcactaccgcttagggatgcaatgctcagggatgtagttgcaagaaacacttagccactgctccgcatcag
tttcttaactgcttttaataataaccccttggctcagctcgtccgttaaacctatgagctcgtgtagtctcttagagctcggtaagccacgataaagctcagggggtcctgcccaca
tgcgttcaggaaatggggagggcttaagcagctccctgacgctcagctctcctcagcgggtttgtagctgcaaaagtgagctcctggtcattctctcacatcggggcagtagctggcc
agtagaacctcagctttagaactcctgcccgttgcctcgcggggatggccggttaagaggtgagctatgagcctcgtgcaactaacctggttcgaagaggttcgaagctggcaccact
taccagctcgaagaaagctccccctctcatttaagagcactcagagacggggccttaagcaggttcagaagaaatctggaatctccctcagcaagagtttgcctcagacgct
gcaggtcttgctattttgagggcctcattccactgagcttgaagagacatctcaggttccacgctggcagcggccactattggttccatcccaggtgcccagatataaacttaagta
gttagcttgctctgctgtagctctcactcagcgccttggcagcttctctcctcgcgtaggggtccggctccagcgttcaactccctggccggttaacactctctgaag
tgaatcgtatagagaaatccatccaacgtagctgcttctcatttagctcgcgtgtagctcagaagattctagattcttagctgtaggacttgaatctcttaactctgtagct
cagctcccatctcttaaacactcaattatcgtcaataactccttatcagtaactctgtagtgcatcccgaggtgaaatgctgtagagaaagagggcacaatgggcttagaaatgctc
gctcgtcataattgtagcagcactcctcctgagggcgaactgagggagctccgttcccaatagctcagcgtcagggctgaattgaagcaggtcggagcggagataaacattgattta
gagctttaaagcttctcagccttgcctgctcccaaaaaaggaactcttgcgtgtagaaatcgtcagtgagggtcgcgactttgagaaatcccctgataaaatcgtctatagaaatgg
caagcctaaaaatcctcgtacaccttaagagtagtaggtgttgccattccaaaacagccttaacctctccggatccatcgtatacccttgcccgtctcaatgatacctcaagat
atltgtgctgttcaactcaactcactcttctgtagtctgctcagggcctgctcccgtagacgcaatagcaccctttgtacatgctcgcgactcctccgtaagctcttgaggaaa
atacaagatgctatctaaatagggcagcagaagctgcccagataccacgttaagcagctgtagatatttgaagctgctcggagcattggccaaacaaaaggggttaaccagcc
actcaaatagcccgtatctggttctaagggcgtcattccactcactcctcctgtagctcgaatctgtgaaagctgcaataatgcccagcttagtgaacatttagcctgcaaatc
gctgtagagctcgcgtgattaaaggtagaggaatacgtactctcgggtgattttgttaagagctcgtatgctccacacagaactctgcccacccccaggctctctcggaaatgtagc
ggggggcaggggaggtgctcaccctaaataaaccttatctagataaactgtaagagctcttcgcaagccagtagtctctcagagacataataatagcggggcccaaggggtt
gtccggcctaccatcggccttttaactcagttcaatcccagatctcactcggggcagaaagtgagggtgagagctcggcggcttcttatcaaatcgtccaaaactcagataat
gttcagggcagctttctcagggcaactcttggttcactcttagagcctttgaatgctggcccaagctcagcggcaagatctcagattctgctctcctcagagctcgcagacgtcag
agaagcttacagcagagatgggaatcacaacagctttgcaattgaaactagggctcactctatgcccactctctcaaatcacaactctcagtagagggatcagggctcctttgct
gggcccagggggtcacttagatacacaatcagcattgaaacctcaaccactcagcagtagaaagggcggctcagttacgtgaccgctcctatcaatcgaatgcccagctcactatgctc
tgaagagctcccattggagcttcaagccggctcatttaaacgggggaaattcaatgacacagggagcttatgcttgaatcagtaggattgagaaataagacctagacatacactcactg
ctatcattgcttagttgaaagcattatctactagagcactcagcagtagaaggggttggatatactcttagattaaatcctcctcattgcttctctaaagctcctattctgctct
tccagttctttgattggagctactcttttctgaaatcctcgtcgtatccacatcgtcgtcgtccagtagccgctcgggatacacttagctattgtgggggtggcagggctctgg
gaggttgaagcggcagctcttaactcagacctcggctcggcactcagggcagggcctgctcctctctcagataaacctctgctgattaccaaattggcgtggggggcgtctg
gggtcagcaccagggagagatccatcagatcatcagcggagcttgggctgggttcgctcagtaggagacagcagacgaagtagagagtaggattgggattgggctttagatctgaag
aagatcgtccctcgcagcttttagaccatgcttagagcttcaagctcagagaggttagatgatacaactaactaagctctcagtagaactcctcagtagagagcgttccctcagccca
ccaagcttccctcagctccttagggggatttcaaatcccttttaacggcagatcaccatccgtgaccgagcttccagcagtagctgataaatcggggcagaaagtcga
agagagctctctaccctgctgtagattgtagcttttgtagcttttctctgctgtagggatcacaagagcagtagcgtgctgctccagtaagctctcaccgtaagcttagct
tgccttattatctcaaccagggttaactcagagctgctgcaatctctccagctcagtagacataaccaagcacaatcctcattatgctcaaaagcttcccaacagcttaagct
ttagattgaagcttagctcccactgagggtagtctccagatagaaactcactctcagcagctccggtagcttggctttggctctctgactgcaatggggtgctgagcttagat
gagagaggtttagttaggggttgaacagcttgcgtagatagctattcagctgctgagctgctataaaacccgttctcagcgtggacctgctcctgcttgggagaaagccca
tccattgttctgaagctgctggggcgtcctgctcattgctcgcgtgtagcatttggtagatttggtagacaagctgcccgttatactcgtcggtagagctttcaatgtgct
ttgttctgttggtagcagattgctaacccagctatcattacagctgctatcagtagaactcgtgtagcaggaattaaagctcccaatagcctcactatgtagcagcagcttgcgta
gctctgtagacagattaaagctcccagtggtcattcaatgaaagggtagaagcaatcgttaaggggttaagctcaagaaactagaggcttagctcctctataatacaaggacga
attgcttgagaaagggaagaaagaaatgagacgagaactcgagcttaagcaagctcgcgttagcattgtagcagctgcccagctgcaaacgatggctaggctgagctcagccagcagcgaa
tctaggggtca

Anexo 7 - Sequência de nucleotídeos do elemento RtPb5 (5412 nt).

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb1	supercontig_1.6	99.74	5741	14	1	1	5740	12970	18710	0.0	1,13E+07
RtPb1	supercontig_1.2	99.65	5741	18	2	1	5740	2151332	2145593	0.0	1,12E+07
RtPb1	supercontig_1.22	99.91	5740	5	0	1	5740	459765	465504	0.0	1,13E+07
RtPb1	supercontig_1.19	99.84	5740	9	0	1	5740	184753	190492	0.0	1,13E+07
RtPb1	supercontig_1.19	99.74	5740	15	0	1	5740	593665	587926	0.0	1,13E+07
RtPb1	supercontig_1.29	99.86	5740	6	2	1	5740	27325	21588	0.0	1,13E+07
RtPb1	supercontig_1.10	99.83	5740	10	0	1	5740	525418	531157	0.0	1,13E+07
RtPb1	supercontig_1.46	99.81	5740	11	0	1	5740	58430	52691	0.0	1,13E+07
RtPb1	supercontig_1.16	99.79	5740	12	0	1	5740	257175	251436	0.0	1,13E+07
RtPb1	supercontig_1.11	99.79	5740	12	0	1	5740	96919	102658	0.0	1,13E+07
RtPb1	supercontig_1.3	99.56	5740	23	2	1	5740	1515018	1509281	0.0	1,12E+07
RtPb1	supercontig_1.12	99.49	5740	29	0	1	5740	132309	126570	0.0	1,12E+07
RtPb1	supercontig_1.12	99.32	5740	39	0	1	5740	116772	122511	0.0	1,11E+07
RtPb1	supercontig_1.28	99.22	5740	45	0	1	5740	182688	188427	0.0	1,10E+07
RtPb1	supercontig_1.8	90.78	5664	510	3	79	5740	186820	181167	0.0	7083
RtPb1	supercontig_1.14	89.72	5662	582	0	79	5740	652794	647133	0.0	6609
RtPb1	supercontig_1.12	91.41	5588	470	1	26	5613	217979	223556	0.0	7281
RtPb1	supercontig_1.52	88.80	5538	620	0	80	5617	44189	38652	0.0	6062
RtPb1	supercontig_1.22	89.51	5537	564	4	79	5613	586874	581353	0.0	6363
RtPb1	supercontig_1.4	90.17	5535	544	0	79	5613	1140015	1134481	0.0	6659
RtPb1	supercontig_1.9	89.11	5535	602	1	79	5613	862792	857259	0.0	6183
RtPb1	supercontig_1.66	88.94	5535	609	1	79	5613	33518	39049	0.0	6116
RtPb1	supercontig_1.68	88.98	5534	610	0	79	5612	16262	10729	0.0	6133
RtPb1	supercontig_1.53	90.71	5525	502	2	79	5602	77869	72355	0.0	6887
RtPb1	supercontig_1.3	90.01	5524	542	1	79	5602	1621079	1615566	0.0	6583
RtPb1	supercontig_1.1	90.11	5187	512	1	79	5264	543762	538576	0.0	6207
RtPb1	supercontig_1.25	99.71	4856	11	2	1	4854	358504	363358	0.0	9501

Anexo 8 – Coordenadas das cópias do elemento RtPb1 no isolado Pb01.

Pb03

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb1	supercontig_1.22	98.64	5740	78	0	1	5740	42117	47856	0.0	1,08E+07
RtPb1	supercontig_1.1	88.22	5740	664	2	1	5740	255709	261436	0.0	6022
RtPb1	supercontig_1.10	83.04	5630	914	8	1	5613	1024455	1018850	0.0	3590
RtPb1	supercontig_1.34	94.02	5536	310	4	79	5613	8797	3282	0.0	8352
RtPb1	supercontig_1.1	98.64	5535	74	1	79	5613	2179038	2184571	0.0	1,04E+07
RtPb1	supercontig_1.39	93.66	5535	349	2	79	5613	15837	21369	0.0	8173
RtPb1	supercontig_1.17	91.62	5522	462	1	81	5602	48228	53748	0.0	7267
RtPb1	supercontig_1.17	90.58	4669	438	2	79	4747	67326	62660	0.0	5751

Pb18

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb1	supercontig_1.22	98.04	5760	93	1	1	5740	27574	21815	0.0	1,06E+07
RtPb1	supercontig_1.14	98.80	5743	61	8	1	5740	377603	371866	0.0	1,08E+07
RtPb1	supercontig_1.3	99.90	5740	6	0	1	5740	2211801	2206062	0.0	1,13E+07
RtPb1	supercontig_1.1	99.90	5740	6	0	1	5740	3741086	3746825	0.0	1,13E+07
RtPb1	supercontig_1.27	99.72	5740	14	2	1	5740	91042	96779	0.0	1,12E+07
RtPb1	supercontig_1.27	92.77	5740	413	2	1	5740	28950	23213	0.0	8072
RtPb1	supercontig_1.11	98.94	5740	57	4	1	5740	801036	795301	0.0	1,09E+07
RtPb1	supercontig_1.10	98.94	5740	57	4	1	5740	31881	26146	0.0	1,09E+07
RtPb1	supercontig_1.6	98.94	5740	57	4	1	5740	1123406	1129141	0.0	1,09E+07
RtPb1	supercontig_1.7	98.92	5740	58	4	1	5740	16024	10289	0.0	1,09E+07
RtPb1	supercontig_1.4	98.92	5740	58	4	1	5740	2446710	2440975	0.0	1,09E+07
RtPb1	supercontig_1.14	92.54	5740	426	2	1	5740	245507	251244	0.0	7969
RtPb1	supercontig_1.31	98.34	5740	91	4	1	5740	22914	17179	0.0	1,06E+07
RtPb1	supercontig_1.21	94.76	5740	299	2	1	5740	57993	63730	0.0	8976
RtPb1	supercontig_1.34	84.01	5646	879	4	100	5729	17578	23215	0.0	4040
RtPb1	supercontig_1.6	98.88	5514	60	2	227	5740	451155	445644	0.0	1,04E+07
RtPb1	supercontig_1.18	92.18	5279	412	1	462	5740	44559	49836	0.0	7182

Anexo 8 – Coordenadas das cópias do elemento RtPb1 nos isolados Pb03 e Pb18.

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb2	supercontig_1.37	99.10	5920	35	16	1	5904	47417	41500	0.0	1,12E+07
RtPb2	supercontig_1.83	99.10	5916	40	13	1	5904	14472	20386	0.0	1,12E+07
RtPb2	supercontig_1.39	99.32	5910	34	6	1	5905	132379	138287	0.0	1,14E+07
RtPb2	supercontig_1.34	87.81	5907	716	3	1	5905	38899	32995	0.0	5979
RtPb2	supercontig_1.2	99.39	5906	34	2	1	5905	1455026	1460930	0.0	1,14E+07
RtPb2	supercontig_1.28	99.34	5906	38	1	1	5905	55395	61300	0.0	1,14E+07
RtPb2	supercontig_1.35	99.32	5906	39	1	1	5905	200618	194713	0.0	1,14E+07
RtPb2	supercontig_1.40	99.31	5906	39	2	1	5905	100719	106623	0.0	1,14E+07
RtPb2	supercontig_1.1	100.00	5905	0	0	1	5905	74991	80895	0.0	1,17E+07
RtPb2	supercontig_1.17	99.53	5905	28	0	1	5905	425230	431134	0.0	1,15E+07
RtPb2	supercontig_1.13	99.53	5905	28	0	1	5905	968343	974247	0.0	1,15E+07
RtPb2	supercontig_1.11	99.44	5905	33	0	1	5905	817998	823902	0.0	1,14E+07
RtPb2	supercontig_1.8	99.46	5905	31	1	1	5905	714986	709083	0.0	1,14E+07
RtPb2	supercontig_1.39	99.12	5905	52	0	1	5905	110118	104214	0.0	1,13E+07
RtPb2	supercontig_1.35	99.00	5905	58	1	1	5905	140793	146696	0.0	1,12E+07
RtPb2	supercontig_1.43	99.29	5905	42	0	1	5905	58699	52795	0.0	1,14E+07
RtPb2	supercontig_1.43	98.88	5905	66	0	1	5905	64631	58727	0.0	1,12E+07
RtPb2	supercontig_1.51	98.88	5905	66	0	1	5905	44226	50130	0.0	1,12E+07
RtPb2	supercontig_1.33	87.01	5905	762	2	1	5905	65901	60002	0.0	5614
RtPb2	supercontig_1.6	99.54	5904	27	0	1	5904	863513	869416	0.0	1,15E+07
RtPb2	supercontig_1.41	99.47	5904	31	0	1	5904	100785	106688	0.0	1,15E+07
RtPb2	supercontig_1.31	99.42	5904	34	0	1	5904	144694	150597	0.0	1,14E+07
RtPb2	supercontig_1.39	99.41	5904	34	1	1	5904	31751	25849	0.0	1,14E+07
RtPb2	supercontig_1.67	99.36	5904	38	0	1	5904	23088	17185	0.0	1,14E+07
RtPb2	supercontig_1.35	86.37	5904	803	1	2	5905	76321	82222	0.0	5315
RtPb2	supercontig_1.1	99.36	5903	38	0	1	5903	481878	487780	0.0	1,14E+07
RtPb2	supercontig_1.63	99.39	5903	36	0	1	5903	10272	16174	0.0	1,14E+07
RtPb2	supercontig_1.35	99.03	5903	57	0	1	5903	97592	103494	0.0	1,13E+07
RtPb2	supercontig_1.19	99.29	5903	42	0	1	5903	654150	648248	0.0	1,14E+07
RtPb2	supercontig_1.16	98.93	5903	63	0	1	5903	738343	744245	0.0	1,12E+07
RtPb2	supercontig_1.3	99.54	5902	26	1	1	5901	1534503	1528602	0.0	1,15E+07
RtPb2	supercontig_1.29	98.61	5902	81	1	4	5905	75553	81453	0.0	1,10E+07

Anexo 8 – Coordenadas das cópias do elemento RtPb2 nos isolados Pb01.

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb2	supercontig_1.35	85.53	5901	851	2	2	5902	194673	188776	0.0	4912
RtPb2	supercontig_1.54	99.32	5898	40	0	8	5905	9265	3368	0.0	1,14E+07
RtPb2	supercontig_1.40	83.79	5898	954	1	8	5905	111375	117270	0.0	4105
RtPb2	supercontig_1.30	99.10	5898	53	0	1	5898	71341	65444	0.0	1,13E+07
RtPb2	supercontig_1.51	84.21	5898	929	1	8	5905	38316	44211	0.0	4304
RtPb2	supercontig_1.44	83.33	5898	981	1	8	5905	8980	14875	0.0	3891
RtPb2	supercontig_1.38	86.35	5898	801	2	8	5905	126512	120619	0.0	5297
RtPb2	supercontig_1.39	85.89	5895	831	1	8	5902	37683	31790	0.0	5081
RtPb2	supercontig_1.40	99.19	5894	48	0	1	5894	49197	43304	0.0	1,13E+07
RtPb2	supercontig_1.31	87.27	5893	749	1	13	5905	271106	265215	0.0	5727
RtPb2	supercontig_1.14	97.40	5893	153	0	13	5905	29729	35621	0.0	1,05E+07
RtPb2	supercontig_1.64	99.00	5891	58	1	1	5891	1551	7440	0.0	1,12E+07
RtPb2	supercontig_1.86	87.78	5891	715	5	13	5903	7322	13207	0.0	5929
RtPb2	supercontig_1.23	99.47	5884	31	0	22	5905	196498	190615	0.0	1,14E+07
RtPb2	supercontig_1.34	85.75	5881	832	5	28	5905	32942	27065	0.0	4976
RtPb2	supercontig_1.44	99.11	5852	51	1	1	5851	40129	45980	0.0	1,12E+07
RtPb2	supercontig_1.33	84.54	5839	898	4	8	5846	47139	52972	0.0	4385
RtPb2	supercontig_1.6	88.76	5703	641	0	8	5710	379226	384928	0.0	6223
RtPb2	supercontig_1.63	84.75	5567	849	0	8	5574	4301	9867	0.0	4304
RtPb2	supercontig_1.16	99.32	5415	37	0	1	5415	724467	719053	0.0	1,04E+07
RtPb2	supercontig_1.66	85.10	5261	770	1	8	5268	28144	33390	0.0	4230
RtPb2	supercontig_1.28	88.17	5064	595	1	15	5078	105691	100632	0.0	5287
RtPb2	supercontig_1.5	85.89	5017	708	0	8	5024	121990	127006	0.0	4331
RtPb2	supercontig_1.44	99.11	4959	44	0	1	4959	14899	19857	0.0	9482
RtPb2	supercontig_1.34	99.04	4921	43	3	1	4921	44907	39991	0.0	9361

Anexo 8 – Coordenadas das cópias do elemento RtPb2 nos isolados Pb01 (Continuação).

Pb03

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb2	supercontig_1.14	98.88	5912	57	4	1	5905	182751	176842	0.0	1,12E+07
RtPb2	supercontig_1.2	98.97	5909	56	3	1	5905	674585	680492	0.0	1,12E+07
RtPb2	supercontig_1.1	99.00	5908	55	2	1	5905	1704744	1698838	0.0	1,12E+07
RtPb2	supercontig_1.2	97.88	5908	121	2	1	5905	1178238	1172332	0.0	1,07E+07
RtPb2	supercontig_1.33	97.90	5908	120	2	1	5905	22144	16238	0.0	1,07E+07
RtPb2	supercontig_1.3	98.98	5907	56	2	1	5904	794237	800142	0.0	1,12E+07
RtPb2	supercontig_1.27	89.91	5907	588	2	6	5905	43803	49708	0.0	6980
RtPb2	supercontig_1.1	98.95	5906	57	3	1	5905	964340	958439	0.0	1,12E+07
RtPb2	supercontig_1.10	98.36	5906	93	2	1	5903	1071688	1077592	0.0	1,09E+07
RtPb2	supercontig_1.23	98.92	5905	60	2	1	5902	130832	124929	0.0	1,12E+07
RtPb2	supercontig_1.33	90.96	5905	531	2	1	5905	16207	10306	0.0	7458
RtPb2	supercontig_1.40	81.43	5902	1089	4	9	5905	13013	18912	0.0	2983
RtPb2	supercontig_1.17	89.18	5898	636	2	9	5905	137338	131442	0.0	6617
RtPb2	supercontig_1.10	86.93	5897	770	1	9	5905	1077625	1083520	0.0	5568
RtPb2	supercontig_1.30	83.92	5897	947	1	9	5905	23885	17990	0.0	4165
RtPb2	supercontig_1.6	82.33	5897	1040	2	9	5905	1826187	1820293	0.0	3412
RtPb2	supercontig_1.8	82.31	5891	1041	1	15	5905	1480437	1474548	0.0	3408
RtPb2	supercontig_1.3	98.93	5885	59	2	21	5902	1630922	1625039	0.0	1,12E+07
RtPb2	supercontig_1.11	83.16	5765	970	1	15	5779	35416	41179	0.0	3721

Pb18

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb2	supercontig_1.19	87.44	5914	723	3	9	5905	150170	156080	0.0	5842
RtPb2	supercontig_1.19	85.87	5910	820	3	9	5905	126781	120874	0.0	5095
RtPb2	supercontig_1.19	83.83	5907	949	5	1	5905	52652	46750	0.0	4100
RtPb2	supercontig_1.9	87.49	5906	736	2	1	5905	983714	989617	0.0	5834
RtPb2	supercontig_1.8	98.53	5905	85	1	1	5905	598597	604499	0.0	1,10E+07
RtPb2	supercontig_1.4	92.67	5905	431	1	1	5905	447220	453122	0.0	8266
RtPb2	supercontig_1.21	85.91	5904	829	2	3	5905	122122	116221	0.0	5093
RtPb2	supercontig_1.3	93.00	5903	394	5	1	5901	3185182	3191067	0.0	8415
RtPb2	supercontig_1.12	85.75	5902	839	1	1	5902	779567	773668	0.0	5025
RtPb2	supercontig_1.9	86.53	5900	779	3	10	5905	1069225	1063338	0.0	5394
RtPb2	supercontig_1.26	89.30	5899	628	2	7	5905	52943	58838	0.0	6677
RtPb2	supercontig_1.36	83.44	5898	971	3	1	5898	10916	16807	0.0	3927
RtPb2	supercontig_1.20	94.95	5897	295	2	10	5905	47579	41685	0.0	9313
RtPb2	supercontig_1.20	91.74	5897	483	3	10	5905	53492	47599	0.0	7807
RtPb2	supercontig_1.12	98.25	5896	101	1	10	5905	203243	209136	0.0	1,09E+07
RtPb2	supercontig_1.6	98.22	5896	103	1	10	5905	577100	571207	0.0	1,09E+07
RtPb2	supercontig_1.4	98.20	5896	104	1	10	5905	1373202	1379095	0.0	1,08E+07
RtPb2	supercontig_1.21	97.83	5896	126	1	10	5905	77324	71431	0.0	1,07E+07
RtPb2	supercontig_1.5	86.08	5896	819	1	7	5902	1955450	1949557	0.0	5172
RtPb2	supercontig_1.11	98.17	5894	106	1	10	5903	201300	207191	0.0	1,08E+07
RtPb2	supercontig_1.14	80.63	5876	1134	3	15	5890	189543	195414	0.0	2603
RtPb2	supercontig_1.21	84.92	5698	851	4	213	5905	69803	64109	0.0	4460
RtPb2	supercontig_1.12	85.04	5427	811	1	6	5432	918838	924263	0.0	4312
RtPb2	supercontig_1.27	97.38	5162	132	3	10	5168	1877	7038	0.0	9139
RtPb2	supercontig_1.5	87.47	4925	613	2	3	4927	80903	75983	0.0	4859

Anexo 8 – Coordenadas das cópias do elemento RtPb2 nos isolados Pb03 e Pb18.

Pb01

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb3	supercontig_1.48	96.26	5186	187	2	1	5181	108599	103416	0.0	8736
RtPb3	supercontig_1.14	89.87	5184	522	2	1	5181	24337	29520	0.0	6100
RtPb3	supercontig_1.14	92.75	5183	372	2	1	5181	742915	748095	0.0	7281
RtPb3	supercontig_1.31	89.36	5181	549	1	1	5181	276379	271201	0.0	5896
RtPb3	supercontig_1.1	93.26	4542	305	1	640	5181	2211146	2206606	0.0	6570

Pb03

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb3	supercontig_1.2	99.81	5183	8	2	1	5181	1852917	1847735	0.0	1,02E+07
RtPb3	supercontig_1.7	99.92	5182	3	1	1	5181	1350256	1355437	0.0	1,02E+07
RtPb3	supercontig_1.29	98.53	5182	75	1	1	5181	50199	45018	0.0	9662
RtPb3	supercontig_1.6	98.38	5182	82	2	1	5181	1580299	1575119	0.0	9591
RtPb3	supercontig_1.8	95.60	5182	227	1	1	5181	1404094	1398913	0.0	8457
RtPb3	supercontig_1.2	100.00	5181	0	0	1	5181	1307933	1302753	0.0	1,03E+07
RtPb3	supercontig_1.17	99.90	5181	5	0	1	5181	34461	29281	0.0	1,02E+07
RtPb3	supercontig_1.1	99.85	5181	8	0	1	5181	467718	472898	0.0	1,02E+07
RtPb3	supercontig_1.17	98.51	5108	73	3	77	5181	42945	48052	0.0	9499
RtPb3	supercontig_1.2	99.79	4231	6	3	1	4231	1956012	1960239	0.0	8292

Pb18

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb3	supercontig_1.5	99.21	5185	33	7	1	5181	1885058	1890238	0.0	9900
RtPb3	supercontig_1.5	99.44	5182	26	2	1	5181	1241203	1246382	0.0	1,00E+07
RtPb3	supercontig_1.2	99.31	5182	32	3	1	5181	1687159	1681981	0.0	9965

Anexo 8 – Coordenadas das cópias do elemento RtPb3 nos isolados Pb01, Pb03 e Pb18.

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb4	supercontig_1.7	95.23	5703	250	6	1	5685	435105	440803	0.0	9133
RtPb4	supercontig_1.11	81.51	5701	1027	7	1	5683	15190	9499	0.0	2928
RtPb4	supercontig_1.85	82.13	5696	1002	4	1	5681	10067	15761	0.0	3211
RtPb4	supercontig_1.3	85.84	5693	796	5	1	5685	1016923	1011233	0.0	4865
RtPb4	supercontig_1.1	84.47	5693	874	5	1	5685	973912	968222	0.0	4246
RtPb4	supercontig_1.28	81.73	5693	1030	4	1	5685	370506	376196	0.0	3019
RtPb4	supercontig_1.8	85.45	5692	820	4	1	5685	873730	879420	0.0	4694
RtPb4	supercontig_1.43	82.73	5692	975	3	1	5685	74578	80268	0.0	3475
RtPb4	supercontig_1.28	82.19	5692	1001	4	1	5681	111549	105860	0.0	3229
RtPb4	supercontig_1.1	84.66	5691	862	3	1	5681	23477	29166	0.0	4351
RtPb4	supercontig_1.40	82.87	5691	964	3	1	5681	128793	123104	0.0	3542
RtPb4	supercontig_1.33	90.04	5690	561	2	1	5685	171455	177143	0.0	6776
RtPb4	supercontig_1.17	84.25	5690	884	5	1	5681	363263	357577	0.0	4149
RtPb4	supercontig_1.47	83.50	5690	923	7	1	5685	49560	55238	0.0	3796
RtPb4	supercontig_1.50	97.28	5689	150	3	1	5685	86612	92299	0.0	1,00E+07
RtPb4	supercontig_1.15	88.31	5689	660	2	1	5685	26929	21242	0.0	5995
RtPb4	supercontig_1.15	83.55	5689	927	4	1	5681	96086	90399	0.0	3834
RtPb4	supercontig_1.10	84.92	5689	848	5	1	5681	403923	398237	0.0	4444
RtPb4	supercontig_1.39	81.35	5689	1048	3	9	5685	44051	38364	0.0	2861
RtPb4	supercontig_1.51	88.34	5688	654	4	1	5681	78772	84457	0.0	5997
RtPb4	supercontig_1.5	86.16	5688	778	4	1	5681	38999	44684	0.0	5013
RtPb4	supercontig_1.17	85.07	5688	840	4	1	5681	671132	665447	0.0	4522
RtPb4	supercontig_1.1	84.93	5688	849	3	1	5681	1164767	1159081	0.0	4466
RtPb4	supercontig_1.4	84.86	5688	853	3	1	5681	1020142	1014456	0.0	4435
RtPb4	supercontig_1.13	83.83	5688	912	3	1	5681	983494	977808	0.0	3967
RtPb4	supercontig_1.43	83.33	5688	940	3	1	5681	105226	110912	0.0	3745
RtPb4	supercontig_1.31	80.91	5688	1075	6	1	5681	290984	285301	0.0	2627
RtPb4	supercontig_1.31	80.71	5688	1087	4	1	5681	259441	265125	0.0	2557
RtPb4	supercontig_1.28	83.52	5687	926	4	1	5677	94938	100623	0.0	3826
RtPb4	supercontig_1.43	83.69	5685	922	3	1	5681	51714	46031	0.0	3899
RtPb4	supercontig_1.84	83.08	5685	952	6	3	5681	4400	10080	0.0	3602
RtPb4	supercontig_1.39	81.32	5685	1045	8	17	5685	140179	145862	0.0	2803
RtPb4	supercontig_1.38	85.52	5684	815	3	1	5677	116518	110836	0.0	4728
RtPb4	supercontig_1.33	80.37	5613	1089	5	83	5685	157302	162911	0.0	2365
RtPb4	supercontig_1.55	82.68	5590	960	3	1	5583	30462	24874	0.0	3392
RtPb4	supercontig_1.22	85.11	5580	820	5	1	5571	185588	191165	0.0	4444
RtPb4	supercontig_1.15	78.06	5333	1167	2	347	5677	6982	1651	0.0	1281
RtPb4	supercontig_1.39	81.10	5274	989	3	9	5275	66972	72244	0.0	2535
RtPb4	supercontig_1.64	81.56	5032	911	7	1	5021	29710	24685	0.0	2581
RtPb4	supercontig_1.5	79.04	4942	1028	6	746	5685	90529	85594	0.0	1538
RtPb4	supercontig_1.1	82.93	4820	823	0	862	5681	1014440	1009621	0.0	3029

Anexo 8 – Coordenadas das cópias do elemento RtPb4 nos isolados Pb01.

Pb03

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb4	supercontig_1.1	98.67	5714	46	18	1	5685	923249	928961	0.0	1,06E+07
RtPb4	supercontig_1.1	94.25	5701	311	3	1	5685	982458	988157	0.0	8705
RtPb4	supercontig_1.1	93.88	5701	332	3	1	5685	988680	994379	0.0	8538
RtPb4	supercontig_1.5	98.82	5699	51	4	1	5685	165664	171360	0.0	1,08E+07
RtPb4	supercontig_1.3	99.21	5696	32	5	1	5685	141828	136135	0.0	1,09E+07
RtPb4	supercontig_1.10	87.82	5691	686	2	1	5685	1062623	1056934	0.0	5781
RtPb4	supercontig_1.11	85.36	5691	825	4	1	5685	58110	63798	0.0	4653
RtPb4	supercontig_1.8	88.43	5688	654	3	1	5685	1459327	1465013	0.0	6036
RtPb4	supercontig_1.7	89.64	5687	586	2	1	5685	1183889	1189574	0.0	6589
RtPb4	supercontig_1.2	93.14	5687	387	3	1	5685	515815	521500	0.0	8157
RtPb4	supercontig_1.10	89.73	5687	581	2	1	5685	882445	888130	0.0	6629
RtPb4	supercontig_1.10	87.99	5687	680	3	1	5685	963121	957436	0.0	5834
RtPb4	supercontig_1.9	88.41	5685	656	2	1	5685	857422	851741	0.0	6030
RtPb4	supercontig_1.24	78.27	5647	1223	4	40	5683	67409	73054	0.0	1433
RtPb4	supercontig_1.22	78.71	5595	1181	5	40	5628	55311	60901	0.0	1618
RtPb4	supercontig_1.7	97.36	5079	126	8	613	5685	947784	942708	0.0	8942

Pb18

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb4	supercontig_1.10	89.90	5693	563	5	1	5685	967639	973327	0.0	6700
RtPb4	supercontig_1.2	99.95	5686	2	1	1	5685	702997	697312	0.0	1,12E+07
RtPb4	supercontig_1.13	100.00	5685	0	0	1	5685	81482	87166	0.0	1,13E+07
RtPb4	supercontig_1.25	85.66	5678	812	2	9	5685	34840	40516	0.0	4785
RtPb4	supercontig_1.1	94.94	5652	278	4	40	5685	3165262	3170911	0.0	8913
RtPb4	supercontig_1.4	97.68	5636	120	6	1	5628	2219761	2225393	0.0	1,01E+07
RtPb4	supercontig_1.10	78.36	5059	1088	3	37	5089	138778	133721	0.0	1330

Anexo 8 – Coordenadas das cópias do elemento RtPb4 nos isolados Pb03 e Pb18.

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.6	85.60	5425	765	4	1	5412	7401	12822	0.0	4553
RtPb5	supercontig_1.14	87.73	5421	646	4	1	5412	8085	13495	0.0	5471
RtPb5	supercontig_1.13	86.21	5417	740	3	1	5412	903053	908467	0.0	4799
RtPb5	supercontig_1.2	88.81	5415	594	3	1	5412	2166362	2160957	0.0	5923
RtPb5	supercontig_1.1	87.66	5415	655	4	1	5412	87996	82592	0.0	5424
RtPb5	supercontig_1.73	84.93	5415	804	3	1	5412	14759	20164	0.0	4258
RtPb5	supercontig_1.47	86.61	5414	717	4	1	5412	108254	102847	0.0	4960
RtPb5	supercontig_1.76	87.09	5414	695	3	1	5412	8218	2807	0.0	5168
RtPb5	supercontig_1.46	86.02	5414	754	2	1	5412	92317	97729	0.0	4716
RtPb5	supercontig_1.27	86.64	5413	721	2	1	5412	159832	154421	0.0	4982
RtPb5	supercontig_1.28	86.22	5413	740	5	1	5412	174820	169413	0.0	4777
RtPb5	supercontig_1.21	88.86	5413	592	4	1	5412	6787	12189	0.0	5931
RtPb5	supercontig_1.18	87.75	5413	660	3	1	5412	112037	117447	0.0	5449
RtPb5	supercontig_1.15	86.26	5413	742	2	1	5412	165059	159648	0.0	4815
RtPb5	supercontig_1.37	85.52	5413	781	3	1	5412	105521	110931	0.0	4490
RtPb5	supercontig_1.3	88.34	5413	627	3	1	5412	386077	380668	0.0	5705
RtPb5	supercontig_1.1	88.29	5413	624	3	1	5412	1113365	1118768	0.0	5693
RtPb5	supercontig_1.1	88.29	5413	632	2	1	5412	220613	215202	0.0	5687
RtPb5	supercontig_1.38	88.18	5413	636	3	1	5412	138536	133127	0.0	5634
RtPb5	supercontig_1.22	88.01	5413	646	3	1	5412	442199	436789	0.0	5560
RtPb5	supercontig_1.8	87.47	5413	676	2	1	5412	749732	755143	0.0	5338
RtPb5	supercontig_1.8	86.79	5413	713	2	1	5412	1127305	1121894	0.0	5045
RtPb5	supercontig_1.47	87.59	5413	670	2	1	5412	68239	62828	0.0	5386
RtPb5	supercontig_1.12	87.57	5413	670	3	1	5412	215457	210047	0.0	5370
RtPb5	supercontig_1.58	86.31	5413	728	4	3	5412	19041	13639	0.0	4841
RtPb5	supercontig_1.64	85.85	5413	755	4	1	5412	34154	39556	0.0	4639
RtPb5	supercontig_1.27	88.97	5412	596	1	1	5412	370353	375763	0.0	5987
RtPb5	supercontig_1.27	86.92	5412	699	2	1	5412	63962	58560	0.0	5112
RtPb5	supercontig_1.28	88.91	5412	597	2	1	5412	162765	168173	0.0	5957
RtPb5	supercontig_1.28	87.31	5412	686	1	1	5412	363100	368510	0.0	5273
RtPb5	supercontig_1.34	88.84	5412	604	0	1	5412	157573	152162	0.0	5939
RtPb5	supercontig_1.34	88.23	5412	636	1	1	5412	18791	24201	0.0	5670
RtPb5	supercontig_1.34	87.27	5412	680	2	1	5412	217551	222953	0.0	5263
RtPb5	supercontig_1.21	87.34	5412	683	2	1	5412	586198	591607	0.0	5281
RtPb5	supercontig_1.2	87.12	5412	696	1	1	5412	1384018	1378608	0.0	5194
RtPb5	supercontig_1.2	85.88	5412	763	1	1	5412	2118023	2112613	0.0	4662
RtPb5	supercontig_1.2	85.79	5412	768	1	1	5412	1796372	1801782	0.0	4623
RtPb5	supercontig_1.2	85.20	5412	800	1	1	5412	615891	621301	0.0	4369
RtPb5	supercontig_1.2	84.65	5412	830	1	1	5412	1411770	1406360	0.0	4131
RtPb5	supercontig_1.25	88.49	5412	622	1	1	5412	9099	3689	0.0	5781
RtPb5	supercontig_1.18	88.43	5412	617	2	1	5412	738462	733060	0.0	5763
RtPb5	supercontig_1.18	87.32	5412	676	3	1	5412	104257	98856	0.0	5279
RtPb5	supercontig_1.18	85.96	5412	759	1	1	5412	8727	14137	0.0	4694
RtPb5	supercontig_1.15	86.88	5412	709	1	1	5412	48330	53740	0.0	5091
RtPb5	supercontig_1.37	88.36	5412	630	0	1	5412	10120	4709	0.0	5733
RtPb5	supercontig_1.37	86.86	5412	710	1	1	5412	87519	82109	0.0	5083
RtPb5	supercontig_1.3	88.16	5412	640	1	1	5412	890123	895533	0.0	5638
RtPb5	supercontig_1.35	88.30	5412	623	3	1	5412	46231	40830	0.0	5699
RtPb5	supercontig_1.1	87.86	5412	657	0	1	5412	1021716	1016305	0.0	5519

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb01.

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.1	87.84	5412	657	1	1	5412	1028013	1033423	0.0	5503
RtPb5	supercontig_1.1	87.01	5412	693	3	1	5412	1188040	1182639	0.0	5144
RtPb5	supercontig_1.1	86.77	5412	715	1	1	5412	2098604	2093194	0.0	5043
RtPb5	supercontig_1.1	86.35	5412	737	2	1	5412	49046	54455	0.0	4853
RtPb5	supercontig_1.1	85.40	5412	787	3	1	5412	481382	475974	0.0	4440
RtPb5	supercontig_1.33	88.14	5412	633	2	1	5412	9709	15111	0.0	5636
RtPb5	supercontig_1.29	88.05	5412	637	3	1	5412	275578	280979	0.0	5588
RtPb5	supercontig_1.13	87.71	5412	664	1	1	5412	637544	642954	0.0	5448
RtPb5	supercontig_1.13	87.64	5412	660	2	1	5412	823802	829204	0.0	5422
RtPb5	supercontig_1.13	86.86	5412	710	1	1	5412	957595	963005	0.0	5083
RtPb5	supercontig_1.9	87.82	5412	650	2	1	5412	996267	990865	0.0	5501
RtPb5	supercontig_1.9	87.53	5412	673	2	1	5412	4779	10188	0.0	5360
RtPb5	supercontig_1.14	87.34	5412	684	1	1	5412	647052	641642	0.0	5289
RtPb5	supercontig_1.14	85.37	5412	791	1	1	5412	786713	781303	0.0	4440
RtPb5	supercontig_1.31	87.73	5412	663	1	1	5412	301660	307070	0.0	5455
RtPb5	supercontig_1.8	87.73	5412	663	1	1	5412	1022621	1028031	0.0	5455
RtPb5	supercontig_1.5	87.68	5412	658	2	1	5412	127007	132409	0.0	5438
RtPb5	supercontig_1.5	87.51	5412	667	2	1	5412	95942	90540	0.0	5366
RtPb5	supercontig_1.5	86.77	5412	715	1	1	5412	26991	21581	0.0	5043
RtPb5	supercontig_1.47	86.92	5412	696	3	1	5412	80448	85847	0.0	5109
RtPb5	supercontig_1.11	87.53	5412	665	3	1	5412	337954	343355	0.0	5366
RtPb5	supercontig_1.11	87.01	5412	679	4	1	5412	495994	490607	0.0	5162
RtPb5	supercontig_1.11	86.55	5412	727	1	1	5412	163065	157655	0.0	4948
RtPb5	supercontig_1.11	86.51	5412	721	2	1	5412	1018052	1012650	0.0	4938
RtPb5	supercontig_1.26	87.32	5412	677	2	1	5412	118966	113564	0.0	5287
RtPb5	supercontig_1.49	87.21	5412	690	2	1	5412	4105	9514	0.0	5225
RtPb5	supercontig_1.10	86.55	5412	727	1	1	5412	398015	392605	0.0	4948
RtPb5	supercontig_1.39	87.12	5412	696	1	1	5412	72373	77783	0.0	5194
RtPb5	supercontig_1.23	86.94	5412	698	2	1	5412	360295	365697	0.0	5120
RtPb5	supercontig_1.23	86.22	5412	745	1	1	5412	450653	456063	0.0	4805
RtPb5	supercontig_1.98	86.94	5412	695	3	1	5412	2879	8278	0.0	5116
RtPb5	supercontig_1.19	86.92	5412	697	3	1	5412	615914	621314	0.0	5107
RtPb5	supercontig_1.43	86.88	5412	709	1	1	5412	13299	7889	0.0	5091
RtPb5	supercontig_1.44	86.77	5412	715	1	1	5412	109995	115405	0.0	5043
RtPb5	supercontig_1.44	85.00	5412	799	3	1	5412	74136	68738	0.0	4286
RtPb5	supercontig_1.57	86.59	5412	724	2	1	5412	4935	10344	0.0	4956
RtPb5	supercontig_1.75	86.49	5412	719	3	1	5412	23613	18214	0.0	4926
RtPb5	supercontig_1.48	86.40	5412	727	2	1	5412	60167	54765	0.0	4890
RtPb5	supercontig_1.58	85.99	5412	758	0	1	5412	49625	55036	0.0	4718
RtPb5	supercontig_1.52	86.16	5412	740	2	1	5412	74031	79433	0.0	4787
RtPb5	supercontig_1.61	86.05	5412	752	3	1	5412	36889	31481	0.0	4718
RtPb5	supercontig_1.46	85.66	5412	767	2	1	5412	68435	63033	0.0	4573
RtPb5	supercontig_1.15	88.39	5411	623	5	3	5412	7020	12426	0.0	5707
RtPb5	supercontig_1.13	87.90	5411	646	2	2	5412	757855	763256	0.0	5531
RtPb5	supercontig_1.30	85.75	5411	765	2	7	5412	40383	34974	0.0	4605
RtPb5	supercontig_1.44	85.31	5411	793	2	1	5410	130671	136080	0.0	4407
RtPb5	supercontig_1.34	87.60	5410	671	0	3	5412	59288	53879	0.0	5404
RtPb5	supercontig_1.34	84.84	5410	804	3	10	5405	209932	204525	0.0	4224

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb01 (Continuação).

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.35	87.06	5410	699	1	3	5412	85322	90730	0.0	5166
RtPb5	supercontig_1.1	87.26	5410	688	1	3	5412	730065	735473	0.0	5253
RtPb5	supercontig_1.4	87.47	5410	669	2	3	5412	8593	13993	0.0	5346
RtPb5	supercontig_1.4	85.82	5410	762	3	6	5412	119589	124996	0.0	4623
RtPb5	supercontig_1.5	86.97	5410	696	2	3	5412	1312566	1317966	0.0	5132
RtPb5	supercontig_1.68	87.45	5410	670	2	3	5412	23131	28531	0.0	5338
RtPb5	supercontig_1.30	86.95	5410	701	3	3	5412	49136	54540	0.0	5107
RtPb5	supercontig_1.51	86.93	5410	705	2	1	5410	7154	12561	0.0	5103
RtPb5	supercontig_1.59	86.27	5410	742	1	1	5410	30697	25289	0.0	4825
RtPb5	supercontig_1.66	84.51	5410	837	1	1	5410	14901	9493	0.0	4072
RtPb5	supercontig_1.1	85.28	5409	792	1	1	5405	405261	399853	0.0	4409
RtPb5	supercontig_1.55	86.39	5409	733	2	4	5412	4219	9624	0.0	4873
RtPb5	supercontig_1.1	87.32	5408	685	1	3	5410	1150998	1145592	0.0	5273
RtPb5	supercontig_1.15	85.11	5407	803	2	1	5407	649479	644075	0.0	4320
RtPb5	supercontig_1.22	85.69	5407	760	4	10	5412	3952	9348	0.0	4569
RtPb5	supercontig_1.16	87.28	5407	679	2	6	5412	699230	693833	0.0	5261
RtPb5	supercontig_1.5	86.79	5406	703	4	8	5412	1326186	1320791	0.0	5037
RtPb5	supercontig_1.20	85.13	5406	795	4	8	5412	34640	29243	0.0	4320
RtPb5	supercontig_1.70	86.59	5406	715	3	1	5406	20031	25426	0.0	4958
RtPb5	supercontig_1.1	87.01	5405	689	2	1	5404	406329	411721	0.0	5154
RtPb5	supercontig_1.29	87.33	5405	684	1	1	5405	317080	311677	0.0	5275
RtPb5	supercontig_1.8	86.60	5405	721	3	1	5405	195158	200559	0.0	4950
RtPb5	supercontig_1.8	85.86	5405	761	3	1	5404	820624	826026	0.0	4633
RtPb5	supercontig_1.20	87.53	5405	664	3	1	5405	69388	74782	0.0	5360
RtPb5	supercontig_1.20	86.05	5405	737	3	9	5412	591756	586368	0.0	4740
RtPb5	supercontig_1.54	86.53	5405	724	4	1	5404	64332	58931	0.0	4910
RtPb5	supercontig_1.90	83.77	5405	875	2	1	5405	7482	12884	0.0	3745
RtPb5	supercontig_1.3	87.08	5404	689	2	1	5404	1361133	1366527	0.0	5176
RtPb5	supercontig_1.3	85.03	5404	808	1	9	5412	25653	31055	0.0	4290
RtPb5	supercontig_1.3	84.88	5404	817	0	9	5412	625260	619857	0.0	4234
RtPb5	supercontig_1.35	87.64	5404	656	5	3	5405	212026	206634	0.0	5390
RtPb5	supercontig_1.4	88.05	5404	645	1	1	5404	1145718	1140316	0.0	5582
RtPb5	supercontig_1.12	87.55	5404	673	0	9	5412	444189	449592	0.0	5376
RtPb5	supercontig_1.70	87.40	5404	669	3	1	5404	2610	8001	0.0	5307
RtPb5	supercontig_1.17	86.45	5404	714	2	1	5404	420503	415118	0.0	4924
RtPb5	supercontig_1.88	86.68	5404	715	3	1	5404	11471	6073	0.0	4984
RtPb5	supercontig_1.54	86.53	5404	727	1	1	5404	25463	30865	0.0	4932
RtPb5	supercontig_1.60	86.51	5404	728	1	1	5404	8012	2610	0.0	4924
RtPb5	supercontig_1.15	85.67	5403	764	6	10	5412	122832	128224	0.0	4534
RtPb5	supercontig_1.8	86.91	5403	705	2	3	5405	1139547	1144947	0.0	5089
RtPb5	supercontig_1.91	86.71	5402	708	3	3	5404	10907	5516	0.0	5005
RtPb5	supercontig_1.28	85.04	5401	795	2	12	5412	9441	14828	0.0	4306
RtPb5	supercontig_1.54	84.58	5401	823	4	10	5410	36698	31308	0.0	4082
RtPb5	supercontig_1.52	84.45	5401	837	3	1	5401	6720	1323	0.0	4022
RtPb5	supercontig_1.58	85.79	5398	765	2	16	5412	37017	31621	0.0	4603
RtPb5	supercontig_1.37	86.92	5397	697	2	1	5397	25513	20126	0.0	5099
RtPb5	supercontig_1.3	84.03	5397	860	2	17	5412	497750	492355	0.0	3848
RtPb5	supercontig_1.40	88.27	5397	632	1	1	5397	60135	65530	0.0	5672

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb01 (Continuação).

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.3	86.01	5396	752	3	10	5404	557259	562652	0.0	4686
RtPb5	supercontig_1.8	86.77	5389	713	0	1	5389	192410	187022	0.0	5029
RtPb5	supercontig_1.26	86.45	5389	719	4	10	5397	45553	50931	0.0	4877
RtPb5	supercontig_1.19	85.17	5382	787	4	31	5412	423037	417667	0.0	4324
RtPb5	supercontig_1.4	85.08	5377	783	4	6	5364	17843	23218	0.0	4298
RtPb5	supercontig_1.59	86.26	5373	737	1	40	5412	30698	36069	0.0	4791
RtPb5	supercontig_1.29	84.90	5298	798	2	115	5412	62616	57321	0.0	4143
RtPb5	supercontig_1.1	84.80	5297	804	1	117	5412	1069717	1075013	0.0	4109
RtPb5	supercontig_1.12	86.45	5297	716	2	117	5412	21701	26996	0.0	4791
RtPb5	supercontig_1.29	84.50	5296	820	1	117	5412	35319	30025	0.0	3981
RtPb5	supercontig_1.13	86.56	5296	698	7	117	5410	730215	724932	0.0	4811
RtPb5	supercontig_1.14	86.01	5296	732	2	117	5412	36750	42036	0.0	4621
RtPb5	supercontig_1.54	84.27	5296	832	1	117	5412	44962	39668	0.0	3885
RtPb5	supercontig_1.13	85.19	5293	782	2	120	5412	847417	852707	0.0	4260
RtPb5	supercontig_1.35	84.90	5290	794	2	117	5406	176083	170799	0.0	4141
RtPb5	supercontig_1.1	87.15	5285	660	3	138	5412	1085192	1079917	0.0	5101
RtPb5	supercontig_1.22	87.72	5278	639	4	138	5412	426570	431841	0.0	5303
RtPb5	supercontig_1.4	84.65	5277	809	1	136	5412	1372212	1366937	0.0	4030
RtPb5	supercontig_1.9	87.09	5276	679	2	138	5412	35087	40361	0.0	5043
RtPb5	supercontig_1.56	86.81	5276	681	8	138	5412	63283	68544	0.0	4890
RtPb5	supercontig_1.61	85.75	5276	740	5	138	5412	43544	48808	0.0	4470
RtPb5	supercontig_1.1	86.29	5275	714	2	138	5412	2078863	2084128	0.0	4722
RtPb5	supercontig_1.4	86.81	5275	687	2	138	5412	1372385	1377650	0.0	4936
RtPb5	supercontig_1.11	87.01	5275	673	3	138	5412	1007233	1012495	0.0	5019
RtPb5	supercontig_1.10	87.49	5275	651	2	138	5412	5977	11242	0.0	5222
RtPb5	supercontig_1.30	87.34	5275	659	2	138	5412	10367	15632	0.0	5158
RtPb5	supercontig_1.17	87.01	5275	676	2	138	5412	30477	35742	0.0	5023
RtPb5	supercontig_1.57	86.81	5275	695	1	138	5412	64385	59112	0.0	4930
RtPb5	supercontig_1.78	86.10	5275	726	3	138	5412	5375	10642	0.0	4629
RtPb5	supercontig_1.8	87.88	5274	630	2	139	5412	278593	283857	0.0	5386
RtPb5	supercontig_1.16	86.93	5273	680	2	138	5410	106745	112008	0.0	4988
RtPb5	supercontig_1.71	86.90	5273	680	3	138	5410	18559	23820	0.0	4966
RtPb5	supercontig_1.33	87.29	5271	669	1	138	5408	45892	40623	0.0	5128
RtPb5	supercontig_1.45	86.60	5269	697	2	144	5412	47796	53055	0.0	4845
RtPb5	supercontig_1.94	86.24	5269	714	4	138	5405	3883	9141	0.0	4678
RtPb5	supercontig_1.32	85.63	5269	755	2	137	5405	156843	162109	0.0	4427
RtPb5	supercontig_1.8	87.36	5267	663	2	138	5404	592757	587494	0.0	5146
RtPb5	supercontig_1.42	85.98	5265	728	3	149	5412	109498	114753	0.0	4575
RtPb5	supercontig_1.5	87.32	5261	655	5	138	5397	77713	82962	0.0	5114
RtPb5	supercontig_1.5	86.49	5261	701	3	152	5412	53052	58302	0.0	4781
RtPb5	supercontig_1.39	86.77	5259	687	2	138	5396	156860	162109	0.0	4904
RtPb5	supercontig_1.53	86.52	5206	700	2	207	5412	72164	66961	0.0	4738
RtPb5	supercontig_1.95	86.45	5202	695	3	211	5412	3839	9030	0.0	4712
RtPb5	supercontig_1.9	86.98	5185	672	3	3	5186	687263	692445	0.0	4902
RtPb5	supercontig_1.8	86.42	5175	702	1	7	5181	1130341	1135514	0.0	4676
RtPb5	supercontig_1.3	83.53	5161	848	2	1	5161	567829	562671	0.0	3475
RtPb5	supercontig_1.15	89.10	5149	547	4	1	5138	181563	186708	0.0	5747
RtPb5	supercontig_1.16	86.67	5131	683	1	1	5131	262304	257175	0.0	4740
RtPb5	supercontig_1.14	87.99	5129	607	2	1	5129	911711	916830	0.0	5281
RtPb5	supercontig_1.8	87.13	5129	650	3	1	5129	693428	698546	0.0	4924

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb01 (Continuação).

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.11	84.80	5119	773	3	1	5115	17967	23084	0.0	3959
RtPb5	supercontig_1.23	84.15	5116	804	6	1	5115	4622	9731	0.0	3665
RtPb5	supercontig_1.37	84.82	5113	772	3	1	5110	133915	139026	0.0	3961
RtPb5	supercontig_1.46	84.62	5111	784	2	1	5111	109420	114528	0.0	3883
RtPb5	supercontig_1.64	83.86	5110	824	1	1	5110	24683	19575	0.0	3580
RtPb5	supercontig_1.45	85.10	5106	759	2	10	5115	18569	13466	0.0	4072
RtPb5	supercontig_1.12	85.72	5041	716	3	1	5039	745810	740772	0.0	4262
RtPb5	supercontig_1.29	84.74	5038	765	3	1	5035	67183	72219	0.0	3868
RtPb5	supercontig_1.37	85.50	5036	728	2	10	5044	30500	35534	0.0	4179
RtPb5	supercontig_1.28	84.74	5008	762	2	7	5014	22233	17228	0.0	3854
RtPb5	supercontig_1.39	86.23	5004	687	2	1	5004	116715	111714	0.0	4440
RtPb5	supercontig_1.30	85.91	5004	703	2	1	5004	326343	321342	0.0	4314
RtPb5	supercontig_1.55	86.11	5004	694	1	1	5004	39702	34700	0.0	4401
RtPb5	supercontig_1.81	85.93	5003	695	2	138	5140	5303	10296	0.0	4333
RtPb5	supercontig_1.65	83.71	4997	811	3	9	5004	30085	25091	0.0	3427
RtPb5	supercontig_1.2	85.37	4995	722	2	116	5110	1399486	1404471	0.0	4103
RtPb5	supercontig_1.27	86.40	4994	678	1	9	5002	99144	104136	0.0	4508
RtPb5	supercontig_1.44	84.12	4988	789	2	117	5104	55556	60540	0.0	3594
RtPb5	supercontig_1.22	86.19	4966	677	2	401	5366	591955	586999	0.0	4403
RtPb5	supercontig_1.52	85.96	4964	687	2	1	4955	66811	61849	0.0	4314
RtPb5	supercontig_1.12	86.89	4943	638	3	470	5412	116771	111839	0.0	4651
RtPb5	supercontig_1.3	86.20	4921	670	2	220	5140	205181	200270	0.0	4369
RtPb5	supercontig_1.15	85.27	4903	722	0	502	5404	90394	85492	0.0	3994
RtPb5	supercontig_1.43	85.65	4892	700	2	117	5008	92180	87291	0.0	4115
RtPb5	supercontig_1.62	85.14	4891	725	2	117	5007	47056	42168	0.0	3915
RtPb5	supercontig_1.26	84.72	4889	744	3	115	5002	246964	251850	0.0	3745
RtPb5	supercontig_1.19	84.64	4889	749	2	117	5004	576528	571641	0.0	3721
RtPb5	supercontig_1.38	86.47	4877	657	3	139	5014	67606	72480	0.0	4411
RtPb5	supercontig_1.38	84.35	4868	757	3	136	5002	26147	31010	0.0	3588
RtPb5	supercontig_1.22	86.97	4737	615	2	1	4737	14092	9358	0.0	4482
RtPb5	supercontig_1.11	84.41	4714	733	2	401	5113	33560	38272	0.0	3501
RtPb5	supercontig_1.22	84.83	4595	695	2	10	4604	19376	14784	0.0	3566
RtPb5	supercontig_1.40	84.42	4556	707	3	145	4699	132781	137334	0.0	3378
RtPb5	supercontig_1.67	84.37	4550	708	3	496	5044	4646	9193	0.0	3358
RtPb5	supercontig_1.47	86.93	4513	576	6	901	5412	7620	3121	0.0	4236
RtPb5	supercontig_1.1	85.45	4468	646	3	900	5364	1069577	1065111	0.0	3681
RtPb5	supercontig_1.38	85.69	4444	634	2	1	4443	56044	60486	0.0	3751
RtPb5	supercontig_1.19	88.35	4436	508	2	977	5412	423344	427770	0.0	4692
RtPb5	supercontig_1.4	84.30	4433	695	1	1	4433	1411302	1406871	0.0	3261
RtPb5	supercontig_1.33	87.01	4371	560	2	1	4364	207832	212201	0.0	4157
RtPb5	supercontig_1.48	86.88	4338	565	4	144	4479	103411	99076	0.0	4056

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb01 (Continuação).

Pb03

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.48	85.39	5415	787	2	1	5412	12643	7230	0.0	4450
RtPb5	supercontig_1.26	84.87	5412	818	1	1	5412	16234	21644	0.0	4226
RtPb5	supercontig_1.1	86.71	5409	717	2	1	5409	1536268	1530862	0.0	5005

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb03.

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.19	86.81	5430	697	3	1	5412	46221	40793	0.0	5095
RtPb5	supercontig_1.23	85.69	5417	770	2	1	5412	44503	49919	0.0	4583
RtPb5	supercontig_1.12	89.83	5417	545	2	1	5412	810055	815470	0.0	6361
RtPb5	supercontig_1.19	89.33	5416	574	1	1	5412	73634	68219	0.0	6151
RtPb5	supercontig_1.31	85.41	5416	783	4	1	5412	8403	2991	0.0	4446
RtPb5	supercontig_1.3	85.74	5415	767	4	1	5412	1508149	1513561	0.0	4583
RtPb5	supercontig_1.6	88.38	5415	626	2	1	5412	735470	740884	0.0	5733
RtPb5	supercontig_1.15	99.41	5413	31	1	1	5412	114815	109403	0.0	1,05E+07
RtPb5	supercontig_1.3	98.61	5413	74	1	1	5412	1514838	1520250	0.0	1,01E+07
RtPb5	supercontig_1.28	86.92	5413	707	1	1	5412	12562	17974	0.0	5109
RtPb5	supercontig_1.19	85.00	5413	808	4	1	5412	108157	102748	0.0	4260
RtPb5	supercontig_1.21	85.24	5413	787	3	3	5412	85958	80555	0.0	4389
RtPb5	supercontig_1.40	85.26	5413	797	1	1	5412	16089	10677	0.0	4395
RtPb5	supercontig_1.5	100.00	5412	0	0	1	5412	70558	75969	0.0	1,07E+07
RtPb5	supercontig_1.5	86.47	5412	725	1	1	5412	1942513	1947917	0.0	4928
RtPb5	supercontig_1.5	85.81	5412	767	1	1	5412	1942098	1936688	0.0	4631
RtPb5	supercontig_1.9	99.39	5412	32	1	1	5412	998717	993307	0.0	1,05E+07
RtPb5	supercontig_1.22	87.12	5412	677	1	1	5412	98567	93176	0.0	5231
RtPb5	supercontig_1.3	89.93	5412	530	1	1	5412	2343629	2349025	0.0	6427
RtPb5	supercontig_1.26	91.59	5412	449	2	1	5412	15324	9919	0.0	7113
RtPb5	supercontig_1.19	91.50	5412	459	1	1	5412	141300	135890	0.0	7073
RtPb5	supercontig_1.19	83.61	5412	884	2	1	5412	6870	12278	0.0	3681
RtPb5	supercontig_1.6	91.08	5412	481	1	1	5412	1445329	1439920	0.0	6893
RtPb5	supercontig_1.10	90.13	5412	532	2	1	5412	995262	1000671	0.0	6478
RtPb5	supercontig_1.10	86.29	5412	740	2	1	5412	960797	966206	0.0	4829
RtPb5	supercontig_1.27	88.62	5412	614	2	1	5412	20744	15335	0.0	5828
RtPb5	supercontig_1.4	87.92	5412	653	1	1	5412	3031016	3036426	0.0	5535
RtPb5	supercontig_1.31	87.25	5412	690	0	1	5412	55155	49744	0.0	5257
RtPb5	supercontig_1.14	87.16	5412	695	0	1	5412	62615	57204	0.0	5218
RtPb5	supercontig_1.14	85.57	5412	781	0	1	5412	106019	100608	0.0	4536
RtPb5	supercontig_1.22	99.28	5411	37	2	1	5410	91483	86074	0.0	1,04E+07
RtPb5	supercontig_1.23	91.28	5411	471	1	1	5410	11469	6059	0.0	6976
RtPb5	supercontig_1.28	91.70	5410	447	2	1	5410	46120	51527	0.0	7148
RtPb5	supercontig_1.28	89.83	5410	550	0	3	5412	91211	85802	0.0	6363
RtPb5	supercontig_1.20	91.39	5410	465	1	1	5410	54063	59471	0.0	7021
RtPb5	supercontig_1.20	90.48	5410	515	0	3	5412	68710	74119	0.0	6641
RtPb5	supercontig_1.6	91.33	5410	466	3	1	5410	1439589	1434183	0.0	6982
RtPb5	supercontig_1.21	83.95	5409	867	1	5	5412	26335	20927	0.0	3832
RtPb5	supercontig_1.23	88.29	5405	633	0	1	5405	106387	100983	0.0	5695
RtPb5	supercontig_1.21	89.27	5405	580	0	1	5405	50272	55676	0.0	6116
RtPb5	supercontig_1.30	88.84	5405	602	1	1	5404	43194	37790	0.0	5925
RtPb5	supercontig_1.19	89.88	5404	547	0	1	5404	24762	19359	0.0	6375
RtPb5	supercontig_1.18	86.80	5403	712	1	3	5405	12158	6757	0.0	5049
RtPb5	supercontig_1.22	86.34	5397	736	1	9	5404	63790	58394	0.0	4847
RtPb5	supercontig_1.22	88.19	5350	631	1	1	5350	31743	37091	0.0	5586
RtPb5	supercontig_1.24	84.93	5347	797	3	73	5412	58131	52787	0.0	4197
RtPb5	supercontig_1.12	87.04	5269	681	1	1	5269	832358	827092	0.0	5023

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb18 .

Query id	Subject id	% id	alig_len	mis	gap	q. start	q. end	s. start	s. end	e-value	bit score
RtPb5	supercontig_1.23	87.97	5160	612	1	262	5412	35884	30725	0.0	5313
RtPb5	supercontig_1.21	87.00	5114	665	0	1	5114	104141	109254	0.0	4865
RtPb5	supercontig_1.24	88.28	4913	575	1	500	5412	89505	84594	0.0	5164
RtPb5	supercontig_1.33	85.36	4911	719	0	495	5405	22852	27762	0.0	4034
RtPb5	supercontig_1.16	86.69	4900	648	2	1	4896	54228	49329	0.0	4532
RtPb5	supercontig_1.22	83.59	4852	783	1	554	5405	63845	68683	0.0	3322
RtPb5	supercontig_1.31	84.40	4596	710	4	1	4592	12578	17170	0.0	3400
RtPb5	supercontig_1.12	86.33	4090	557	2	1324	5412	902337	898249	0.0	3659
RtPb5	supercontig_1.18	86.04	4076	566	2	1338	5412	41304	37231	0.0	3554
RtPb5	supercontig_1.1	88.84	4067	451	1	1	4064	3246487	3242421	0.0	4458
RtPb5	supercontig_1.14	86.32	4036	547	2	1	4035	5759	9790	0.0	3614
RtPb5	supercontig_1.24	90.00	4028	403	0	1	4028	89506	93533	0.0	4789
RtPb5	supercontig_1.28	83.66	3820	621	2	1593	5412	77985	81801	0.0	2611
RtPb5	supercontig_1.1	87.64	3778	467	0	1	3778	3894842	3898619	0.0	3786
RtPb5	supercontig_1.23	86.66	3711	491	2	1	3710	79473	83180	0.0	3420
RtPb5	supercontig_1.20	90.63	3659	342	1	1755	5412	98046	94388	0.0	4526

Anexo 8 – Coordenadas das cópias do elemento RtPb5 no isolado Pb18 (Continuação).

>RtPb1-Rt-aa

GFITPSSAPYASPIILIARSGKKLRFCVDFRKLNAITKRDQYPLPLIDELMDRLNGAKYF
TKLDIRQGFHRIRMDPESEDLTTFRTRYGTYKYRVVPPFGLTNAPACFQRFVNSIFFDYL
DKFMTAFMDDLIIYSKNLEEHKQQVNSVLDRLKTAGLQASIQKCEFAVQRTKYLGFII

>RtPb2

GLERLVARRLAWTAIQEKIVHPQHFGALPGRAASDLVAAMVHDIEEAWAHGKVVSLTL
DIKGAFDAVLRGRMLRRLQEQGWPPNLIWVESFMTGRTGRIRLDGLLGELFDI PAGLP
QGSPVSPILFMLFIQPLFFLGTIQRRRARLGYADDLGLLSAGTTLEGNVVTLQEDFKLL
NDWASKEGLTFDFAKTEIAHFTRRRRLSNPSIDLETTSGTHSITAKAVKSSIRYLGIIW

>RtPb3-Rt-aa

RKTWTLI PRSDVPADEPVAPGRWIFVEKDRPNADPDGILRKSRRWVACGNWLKKGSR
EETYAPVVNDTTNRMMFAISAKKGWTLRQVDIVMAYLHGRLGRTIYMRQPTGFEAGAPR
TLVCAVEGSLYGLDPAAKIWIYDYL TALLRKNGFAPSPYDPALWRHTTRRHLYMTIYVDD
LHIMAEPPADADWL VETLASKLEIKDLKEPSKYLGMEEVERRQDGGIKLTLKRYIDRLVQ
DFHLEDAVTVRTPM

>RtPb4-Rt-aa

WRLVERTPDMAILRGQVWFDKKINNAGRIDRYKARWVVCNGYQEKGVNYSKTFAPVVS
STSRALMAISVSRGWHVLQIDMVTAFLNARLPANERIYMHQPAGFREGRGDLVCELMQG
LYGLKQAALLWYEELRKMLLETGFRVSPHDAGFFINDVRKVYIT IYVDDVRLYGPNI SD
IEWTKAQIAANYEIKDVGDSNRYLGMKIDRDAESLSLSQPLFIKNLLEEFMSMDDCHPKS
SPMEEN

>RtPb5-Rt-aa

GFIRVSNSPAAPVLF AKKPGGGLRFCVDYRALNKITRKDRYPLPLISETLQORIGQAKW
FTKLDIIAAFHKIRITEGDEWMTAFRTRYGLFEWLVTPFGLANAPSTFQKYINWALRGY
LDDFCSAYLDDILIFSSRSLREHREHVQKVLMLRLREAGLQADIRKSEFEVNSTKYLGFII
I

>XP_001544372.1-Ajellomyces_capsulatus

RNTWTMI PRSDVPSGEPVAPGRWVYVEKDRPQDPNADPDGILRKSRRWVVCNGNLKKSAR
EETYAPVVNDATNRMMFAIAAKRGWTLRQVDIVMAYLHGRLGKTIYMRQPTGFEVGT
PHAMVCAVGGSLYGLDPAAKIWIYDFLTALLRENGFKPSYDPALWRHTSRNHLYMTIYVDD
LHIIAEDPADADWLVDLANKLEIKDLKEPSKYLGMEEVERRPDGAIKLTLRRYIDRLVQ
DFRLDDAVPVPTPI

Anexo 9 – Sequências de aminoácidos relativo às regiões de transcriptase reversa dos elementos utilizados nas análises filogenéticas.

>EFZ04205.1-Metarhizium_anisopliae
GFIRASSSAASAPVLFVRKPGGGLRFCVDYRALNAITRKDRYPLPLINETLRAISKAKW
LTKLDVIAAFHKIRIQEGEEWKTAFRTRYGLFEWLVTPFGMTGAPATFQRYINRTLQDY
LDEFCSAYIDDILIYSDGSLKDHRQKVRQVLKRLRDAGLQVDIDKCEFETKSTKYLGF I
V

>XP_001538733.1-Ajellomyces capsulatus
WRLIERTPNMAVLRGQWVFDKKNMVKVRYKARWVVCNGYQQKGVNYSKTFAPVVSA
STSRALMEISASRGWHVLQIDMVTAFLNAQLPADERIYMSQPTGFREGRGDLVCELMQG
LYGLKQAALLWYEELRKMLLQSGFRVSPYDAGLFIHDARKVYMTVYVDDVRLYGPKIED
LNWAKAAIAANYEIKDVGDSNRYLGMKIERDSTSLTSLQPLFITNLLEEFAMDDCHPTS
TPIEEN

>CBF81153.1-Aspergillus_nidulans
VRKPGGGLRFCVDYRALNAITKKDRYPLPLIHETLNQIGQARWFTKLDVSAAFHKIRIA
KGQEWMTAFRTRYGLFEWLVTPFGLANAPSTFQKYINWTLREYLDEFCSAYIDDVLVYT
NGDLRQHRKHVRMVLKLEEAGLYLDIKKCEFECKETKYLGF I I

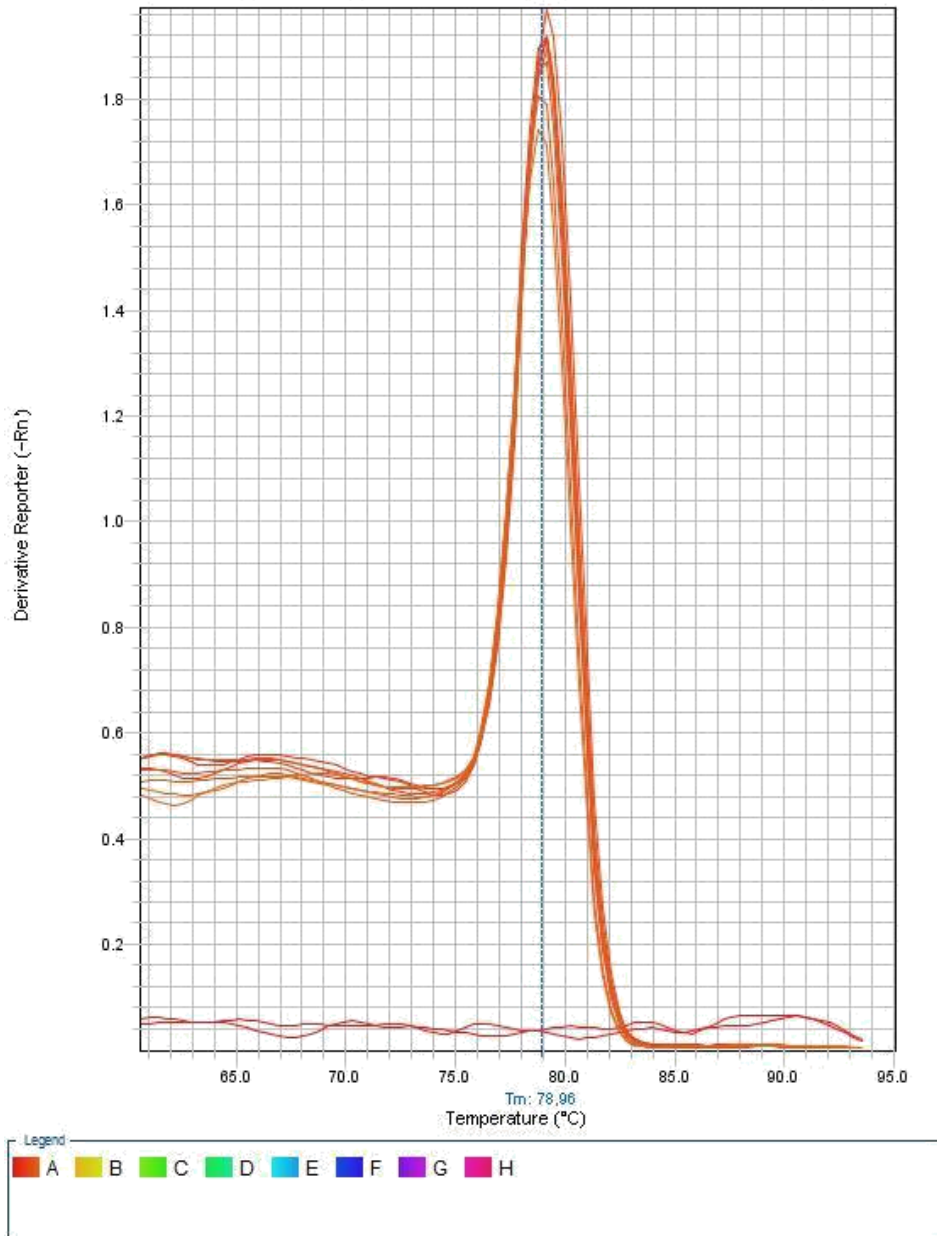
>XP_003070555.1-Coccidioides posadasii
GTWILVEKEPQMQLSGKWVDRKLNAGGTVRFKARWVVRGFEQQEGVHYKKTAAVV
SGPTTRAFFAISTAKNWKVRVDFVTAFLNGTLPEDERVYVQLPTGFKHGRGNLVGLLR
QGLYGLKQAARLWYQAATDYLKSLGFQISPYDAGLLYHKRKQIYTLHVDDCRIMGPNE
AEIEWLIEKLQERFEIKEVTEQGRYLGMRVQKQPNGDLFVNQEQYIEDLLTEYGMEDCK
PSKTPM

>BAC82628.1_Anopheles gambiae
MKRKCGKLVSFDSLQAFDRVDRGFLFNNMVSMGFNPGLVGLLRRFGDQSSSRILVNGSL
SPPIISIRRSVRQGDPLSMHLFILYLHPLITRLEGICCDQDDLVAAYADDISVVTSSQK
IELV

>M14954.2_Drosophila_melanogaster
LIIPILKPNTDKTKTSSYRPISLNCCIAKILDKIIAKRLWWLVTYNNLINDKQFGFKKG
KSTSDCLLYVDYLITKSKMHTSLVTLDFSRAFDVRGVHSIIQQQLQEWKTGPKIIKYIKN
FMSNRKITVRVGPHTSSPLPLFNGIPQGSPISVILFLIAFNKLSNIIISLHKEIKFNAYA
DDFFLIINFNKNTNTNFNLDNLFDDIENWCSYSGASLSLSKCQHLHICRKRHCTCKISC
NNFQIPSVTSLKILGIT

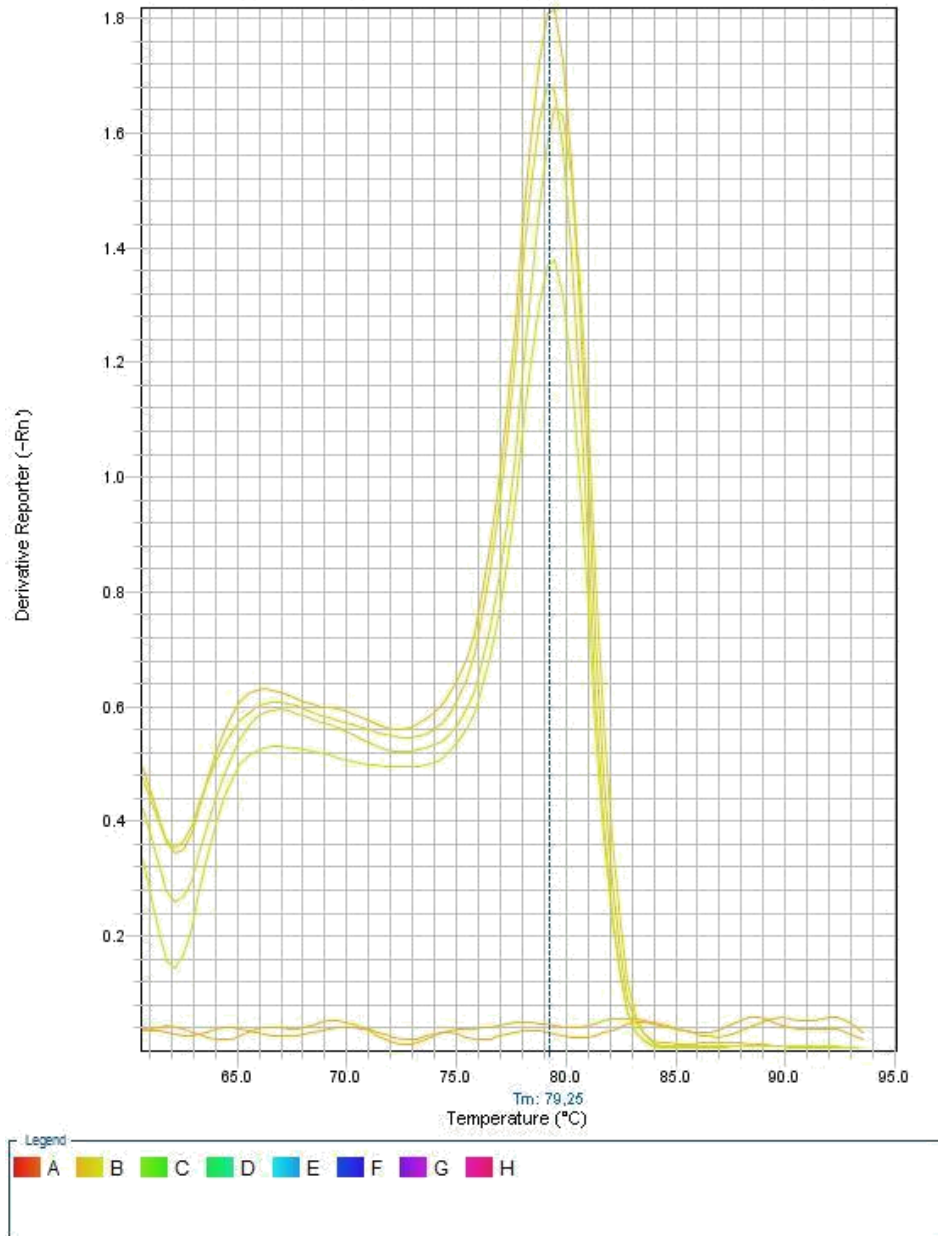
Anexo 9 – Sequências de aminoácidos relativo às regiões de transcriptase reversa dos elementos utilizados nas análises filogenéticas (Continuação).

Melt Curve



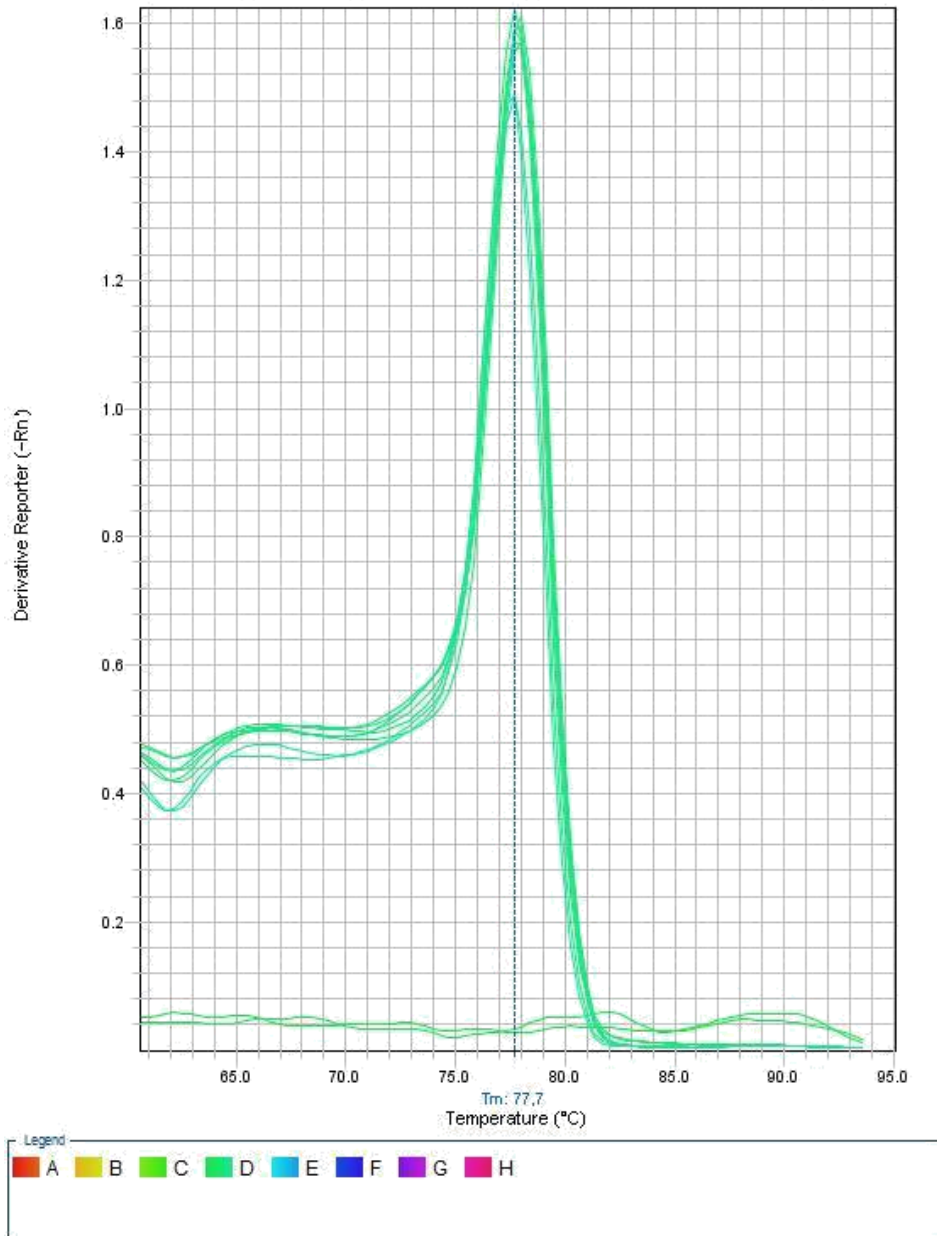
Anexo 10 – Curva de dissociação do retrotransposon RtPb1.

Melt Curve



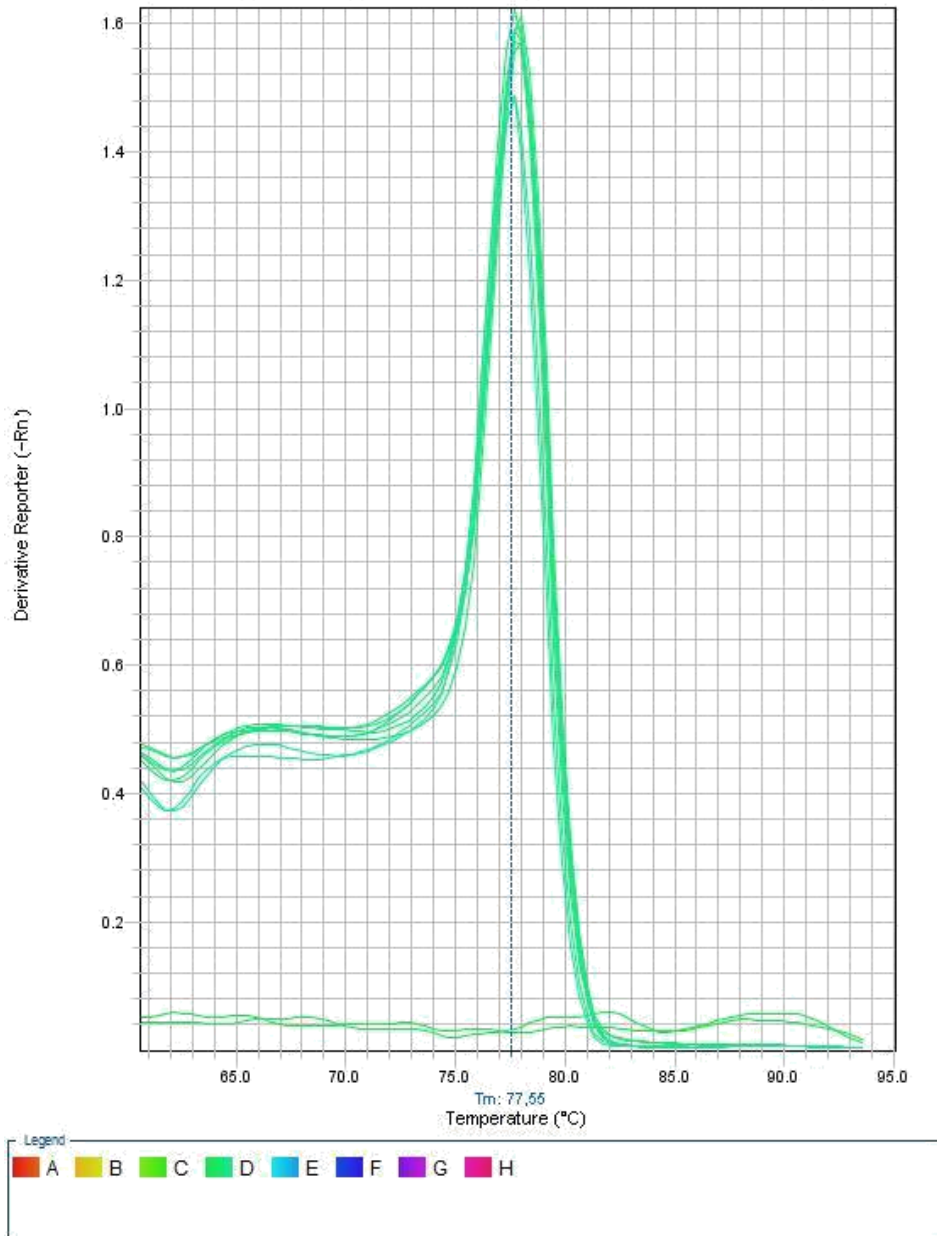
Anexo 10 – Curva de dissociação do retrotransposon RtPb2.

Melt Curve



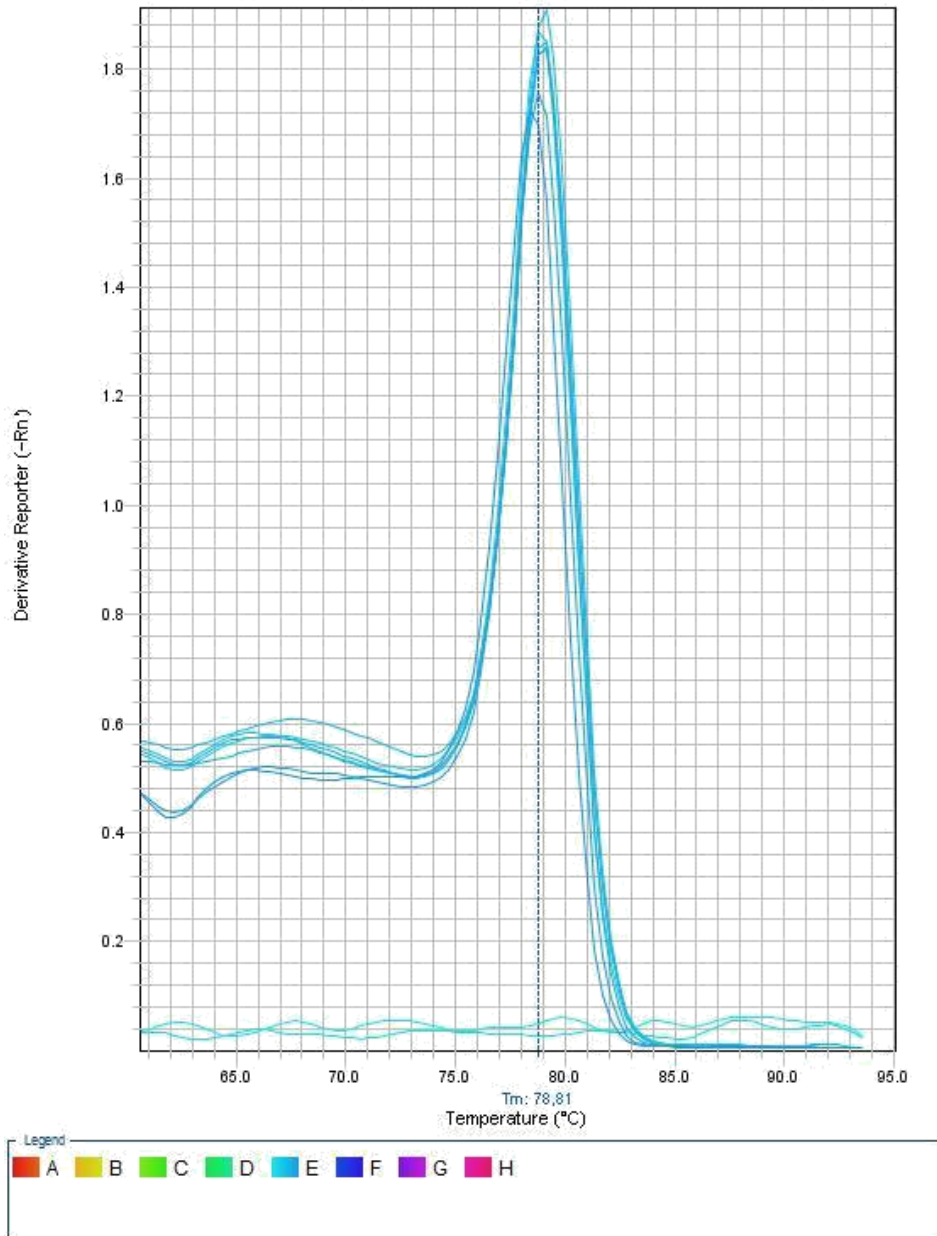
Anexo 10 – Curva de dissociação do retrotransposon RtPb3

Melt Curve

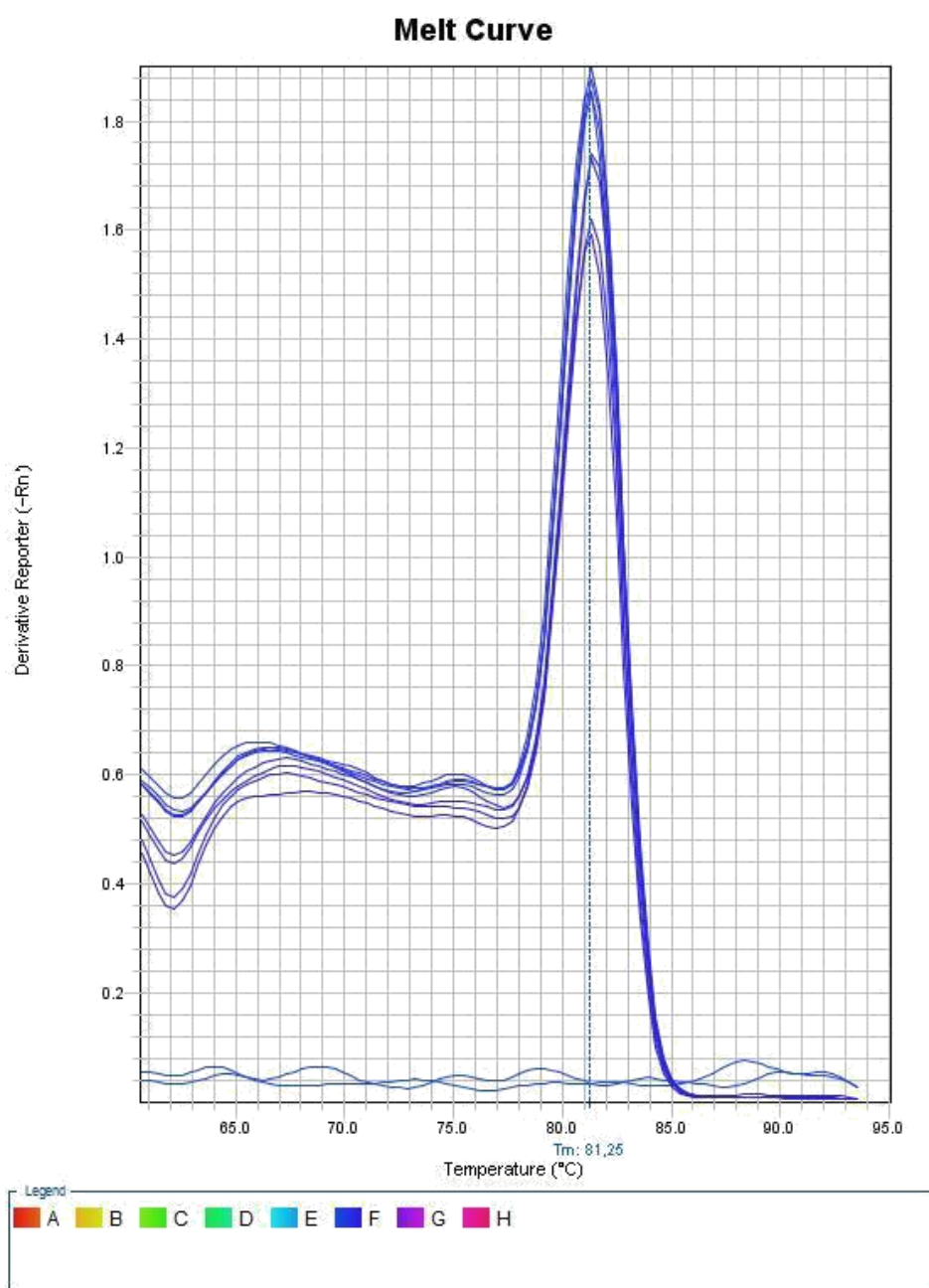


Anexo 10 – Curva de dissociação do retrotransposon RtPb4.

Melt Curve

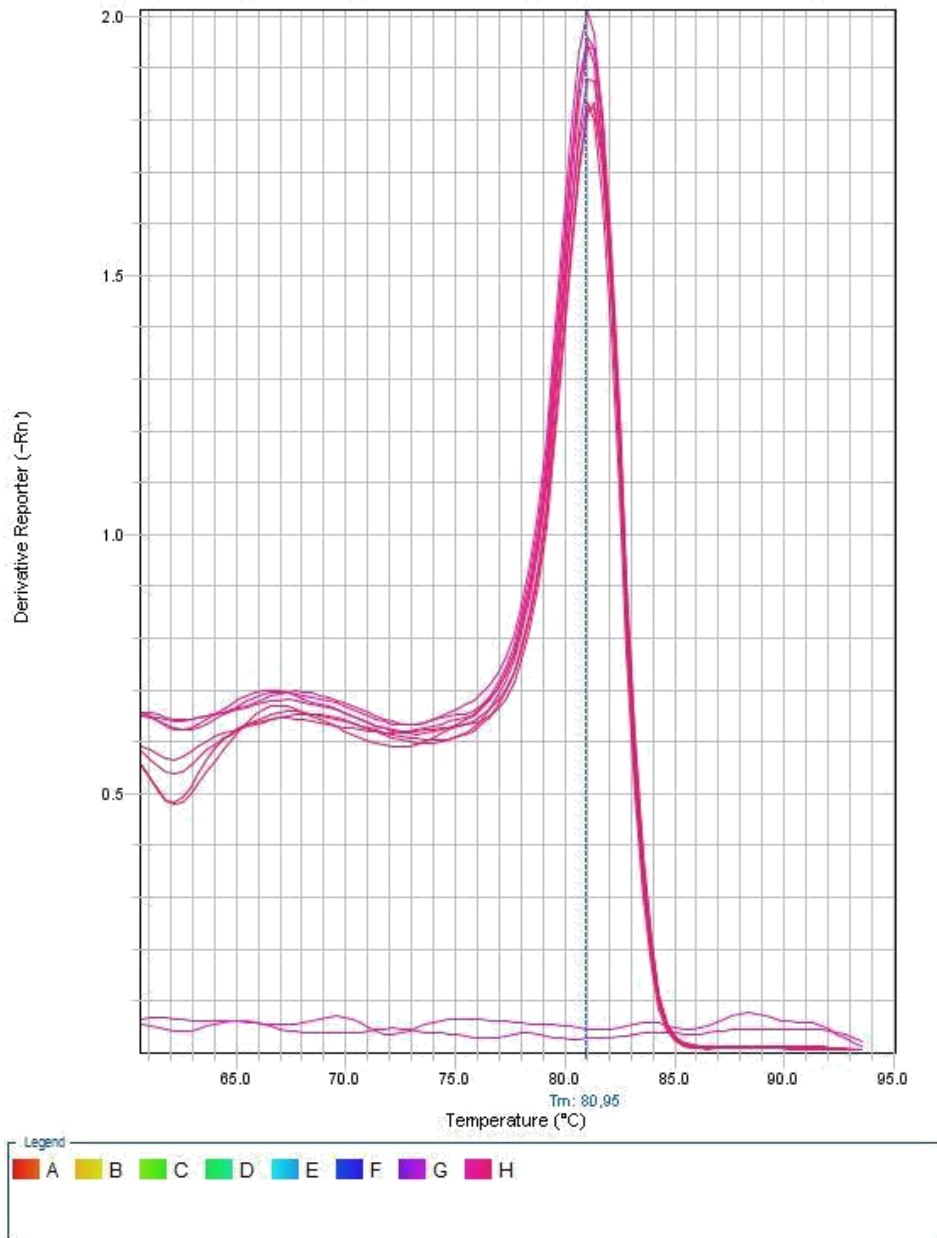


Anexo 10 – Curva de dissociação do retrotransposon RtPb5.

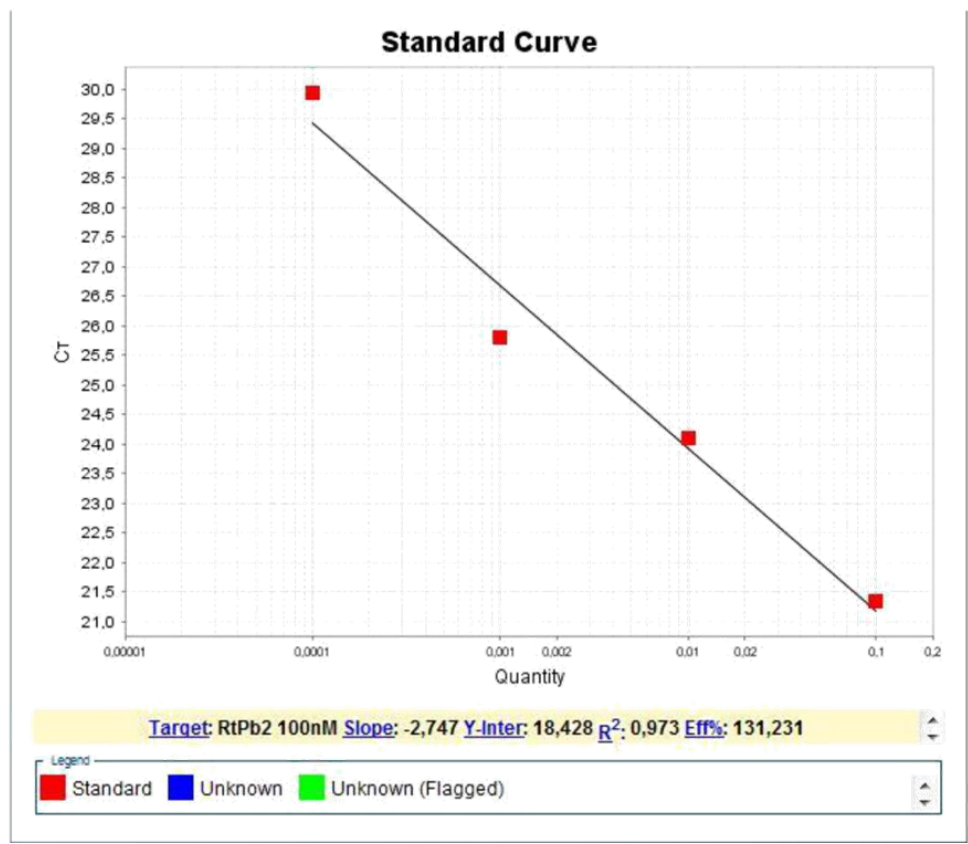
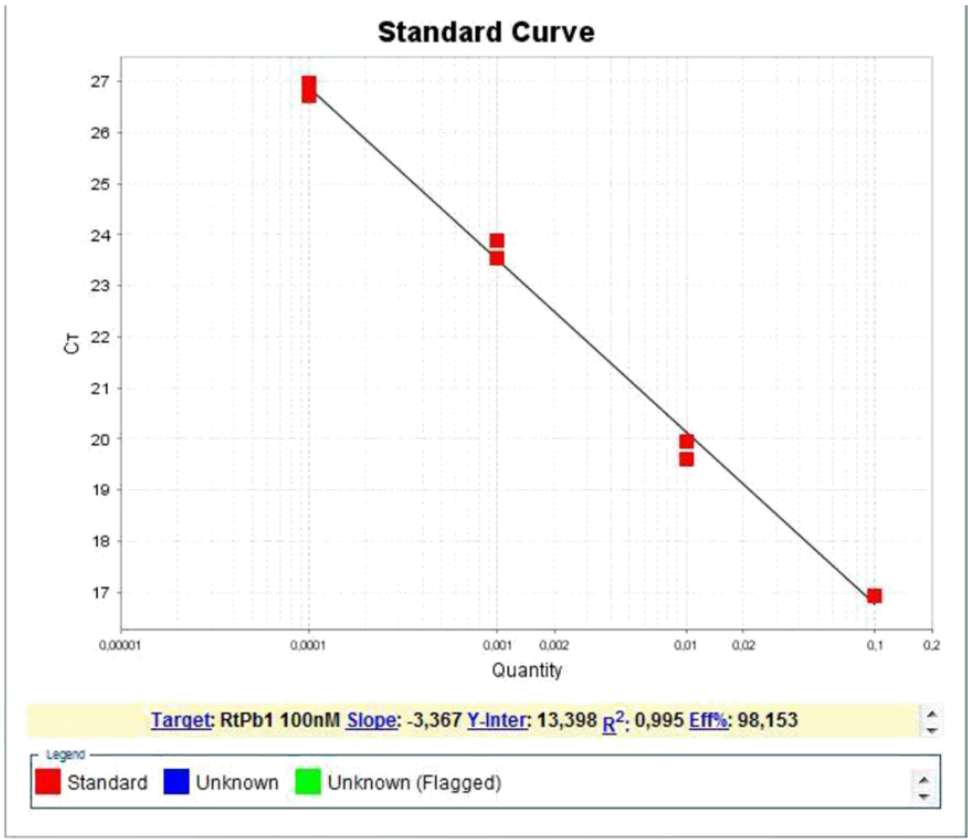


Anexo 10 – Curva de dissociação do gene constitutivo β -tubulina.

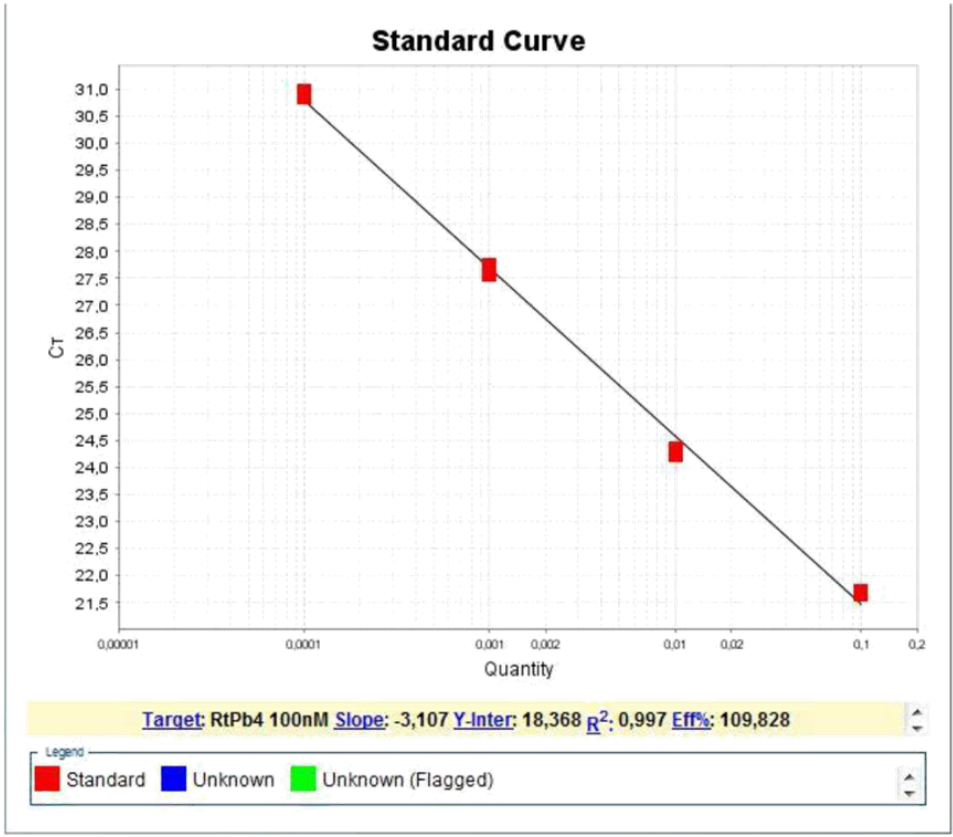
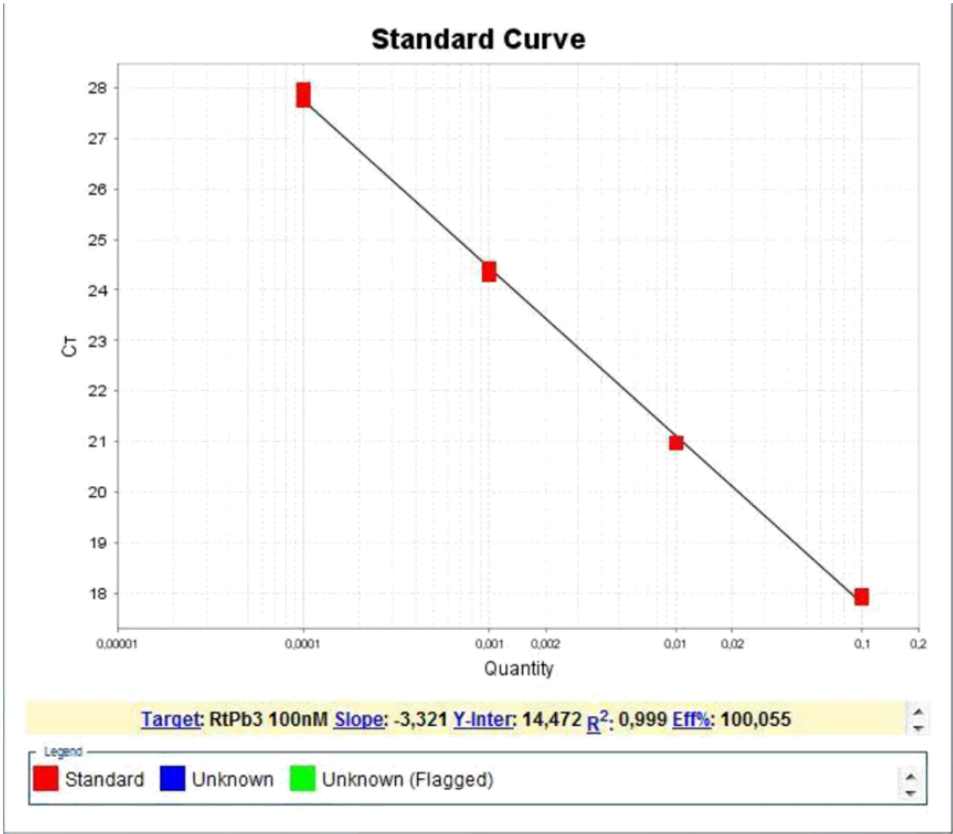
Melt Curve



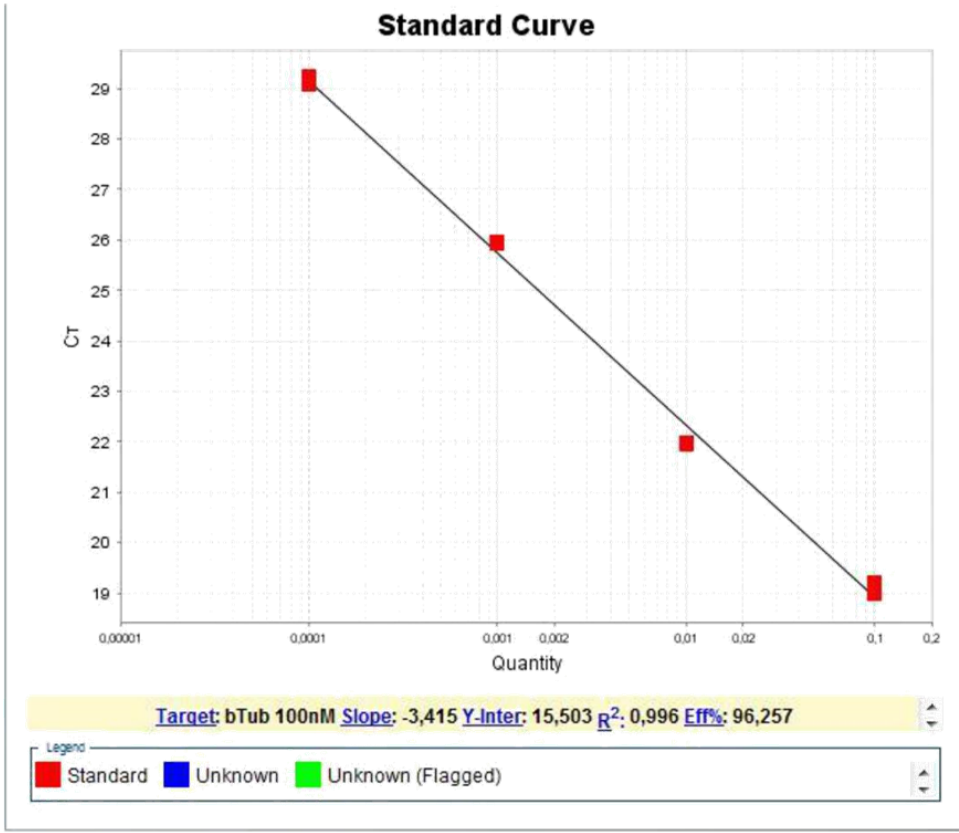
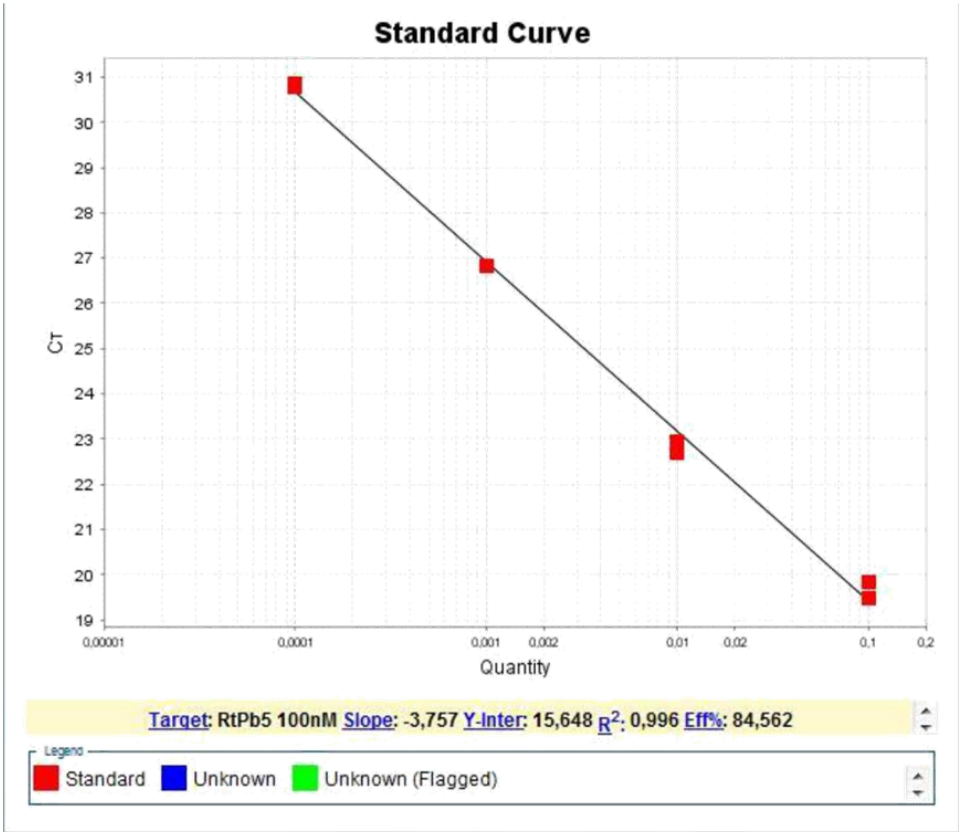
Anexo 10 – Curva de dissociação do gene constitutivo L35.



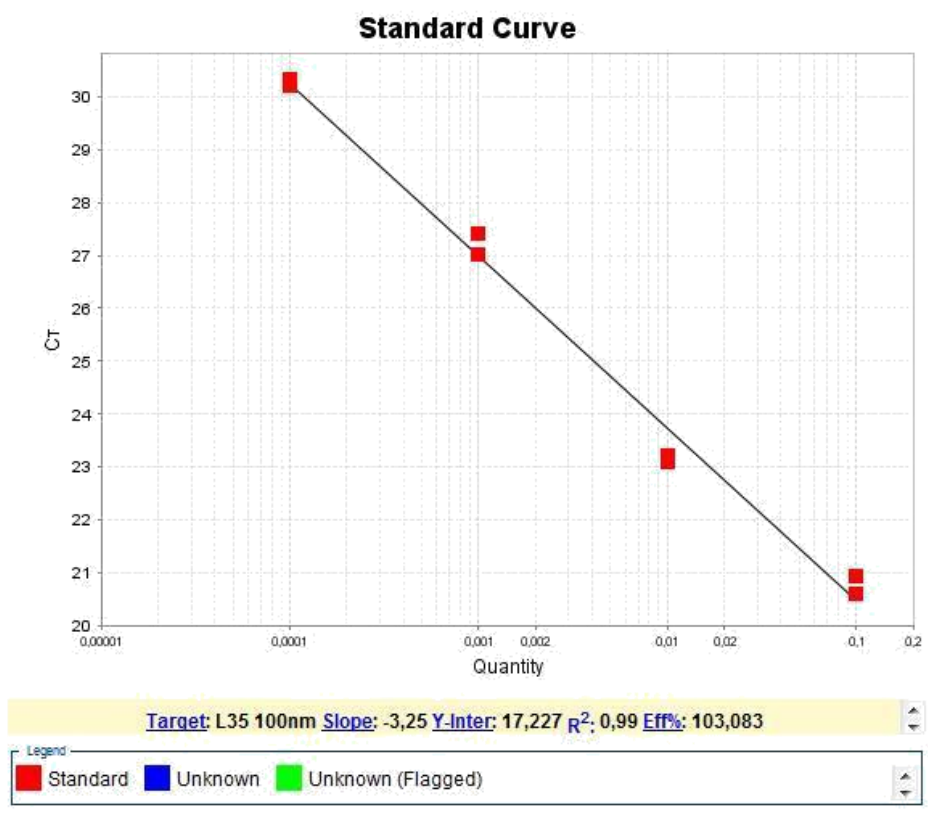
Anexo 10 – Curva de eficiência dos elementos RtPb1 e RtPb2.



Anexo 10 – Curva de eficiência dos elementos RtPb3 e RtPb4.



Anexo 10 – Curva de eficiência do elementos RtPb5 e do controle β-tubulina.



Anexo 10 – Curva de eficiência do controle L35.