

UNIVERSIDADE FEDERAL DO PAMPA

FILIPPE ZIMMER DEZORDI

**CARACTERIZAÇÃO E ANÁLISE EVOLUTIVA DE ELEMENTOS
TRANSPONÍVEIS PRESENTES NO GENOMA DE *Leptopilina boulardi***

São Gabriel - RS

2016

FILIPPE ZIMMER DEZORDI

**CARACTERIZAÇÃO E ANÁLISE EVOLUTIVA DE ELEMENTOS
TRANSPONÍVEIS PRESENTES NO GENOMA DE *Leptopilina boulardi***

Trabalho de Conclusão de Curso apresentado ao curso de Biotecnologia pela Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel em Biotecnologia.

Orientador: Prof. Dr. Paulo Marcos Pinto

Co-orientador: Dr. Gabriel da Luz Wallau

São Gabriel

2016

FILIPE ZIMMER DEZORDI

**CARACTERIZAÇÃO E ANÁLISE EVOLUTIVA DE ELEMENTOS
TRANSPONÍVEIS PRESENTES NO GENOMA DE *Leptopilina boulandi***

Trabalho de Conclusão de Curso apresentado ao curso de Biotecnologia pela Universidade Federal do Pampa, como requisito parcial para obtenção do Título de Bacharel em Biotecnologia.

Área de concentração: Bioinformática.

Trabalho de Conclusão de Curso defendido e aprovado em 21 de Novembro de 2016.

Banca Examinadora:

Prof. Dr. Paulo Marcos Pinto
Orientador
UNIPAMPA

Dr. Adriana Koslovski Sassi
UNIPAMPA

Bel. Pablo Echeverria Macedo
UNIPAMPA

AGRADECIMENTOS

Aos meus pais, Gilmar e Angelita, por colocarem minha felicidade e de meu irmão sempre em primeiro lugar, nos permitindo ter uma vida feliz e seguir nossos sonhos, às vezes, nos momentos de correria da vida acadêmica deixei de expressar a importância que vocês tiveram em cada objetivo que alcancei, mas saibam que o que fiz e que tenho feito é uma forma de agradecer aos senhores. Ao meu irmão, Vitchenzo, pelo companheirismo e pela rivalidade saudável ao longo dos anos.

Ao professor Paulo Pinto, pela oportunidade da iniciação científica, pelas horas de dedicação em tirar minhas dúvidas, pela compreensão nos momentos de “rebeldia”, pelas palavras bonitas e expressões irônicas que acabei adicionando ao meu vocabulário e por me mostrar como é bom passar o tempo no laboratório.

Ao professor Gabriel Wallau por me apresentar ao estudo dos elementos transponíveis, e que mesmo não podendo ter seu nome na orientação do trabalho, sempre esteve presente para tirar dúvidas, discutir resultados e conversar sobre meu futuro acadêmico.

Aos colegas do grupo de pesquisa LPA, pelos momentos de companheirismo, conversas filosóficas e trabalho, em especial, ao Pablo Echeverria, pela coorientação em algumas linhas de pesquisa e amizade.

A Alexandre Freitas, pela amizade, apoio, e horas de estudo e pesquisa com elementos transponíveis.

À Tainah Oliveira, pela amizade e apoio durante a graduação.

Aos colegas e amigos da turma 17, pelo companheirismo nesta jornada, pelo compartilhamento de tristezas e alegrias, em especial aos amigos Dennis, Gleison, Helber, Hudson e Lucas.

Aos professores que permitiram o ganho de conhecimento e meu crescimento pessoal nos últimos 4 anos, em especial ao Juliano Boldo e ao Luis Fabiano Costa, pela dedicação para com os alunos do curso.

“Não sejam macacos treinados.”

Juliano Tomazzoni Boldo

RESUMO

Elementos transponíveis (TEs, do inglês *Transposable Elements*), ou segmentos móveis do material genético, foram descobertos em milho (*Zea mays*) por Barbara McClintock na década de 1940 e posteriormente caracterizados em dois grandes grupos baseado no seu sistema de transposição. Desde então, estes elementos têm sido descritos em uma grande variedade de eucariotos. Com o desenvolvimento das técnicas de sequenciamento, recentemente foi identificado que estes elementos podem constituir grande parte do genoma de diversos organismos, chegando a 50% do genoma dos primatas e 85% do genoma do milho, tendo importância crucial para a evolução de organismos, atuando, por exemplo, no desenvolvimento de sistema imunológico e na constituição de cromossomos. Estas modificações provocadas por elementos transponíveis geralmente estão associadas ao processo de mobilização. Fenômenos de transferência horizontal de genes são essenciais para a manutenção de TEs (HTT, do Inglês *Horizontal Transposon Transfer*), culminando em sua invasão à novos genomas. Um dos grandes problemas em estudar mecanismos de transposição é mimetizar um ecossistema em laboratório e quando contornado este problema vem o alto custo com o sequenciamento de genomas de diversos organismos e à dificuldade de análises mobilômicas com organismos não-modelo. Dessa forma este trabalho tem como objetivo a padronização de um *pipeline* de bioinformática que permite a utilização de sequências provenientes de sequenciamento de alto desempenho (HTS, do inglês *High Throughput Sequencing*) com baixa cobertura juntamente com a caracterização dos elementos presentes no genoma de uma vespa parasitoide de drosofilídeos (*Leptopilina boulardi*) e o parentesco destes elementos com elementos de outros organismos. Foi realizado sequenciamento de próxima geração *Solexa-Illumina HiSeq 2000* de 20 indivíduos da espécie *L. boulardi*, gerando 30912162 *reads* com tamanho aproximado de 100 pares de bases, os quais foram submetidos a várias ferramentas de bioinformática resultando em *clusters* (agrupamentos de vários *reads*) os quais foram montados em *contigs* (sequências definidas pela sobreposição de vários *reads*). Destes uma parcela apresentou homologia com elementos de um banco de dados de TEs representando então 5,19% do genoma da vespa pertencendo a 15 superfamílias de elementos diferentes. Foi realizada uma análise de elementos descritos na literatura, e a partir das sequências da vespa e das provenientes da literatura foram utilizados *softwares* para alinhamento, melhor modelo de substituição de aminoácidos e montagem da árvore filogenética. Foram construídas 11 árvores de diferentes superfamílias de elementos

que apresentaram domínios de proteínas já descritas na literatura como pertencentes à estrutura destes elementos. 26 elementos de *L. bouleari* formaram agrupamento com elementos já descritos na literatura e 52 formaram agrupamento com elementos ainda não classificados, podendo indicar novos táxons de TEs. Além disso, foram identificados possíveis eventos de HTT entre vespas e formigas. O *pipeline* construído neste trabalho apresentou eficiência quando analisados a qualidade das sequências e o alto valor de suporte de ramo presente nas árvores filogenéticas, indicando a eficiência da metodologia utilizada, a qual possibilita uma economia em linhas de pesquisa que utilizem o sequenciamento simultâneo de diversos organismos.

Palavras-Chave: Elementos Transponíveis, Bioinformática, Transferência Horizontal, Filogenia, Vespa parasitoide.

ABSTRACT

Transposable elements, or mobile segments of genetic material, were discovered in maize (*Zea mays*) by Barbara McClintock in the 1940s and later featured in two major groups based on their transposition systems. Since then elements have been described in a wide variety of eukaryotes. With the development of sequencing techniques, it was recently identified that these elements may constitute a large part of the genome of various organisms, reaching 50% of the genome in primates and 85% of maize genome, and crucial for the growth of organisms acting for example, in the development of the immune system and the recognition of chromosomes. These changes caused by transposable elements are generally associated with mobilization processes. Horizontal Transfer phenomena of genes are essential for maintenance of TEs culminating in its invasion of new genomes. A major problem when studying transposition mechanisms is to mimic an ecosystem in the laboratory and when this is solved, comes the high cost of sequencing the genomes of various organisms. Thus this study aims to standardize a bioinformatic pipeline that allows the use of sequences from next-generation sequencing (NGS) with low coverage along with the characterization of the elements present in the genome of a wasp parasitoid drosophilid (*Leptopilina boulardi*) and the relationship of these elements with other organisms elements. Next generation sequencing Illumina Solexa-Hiseq 2000 was performed using 20 individuals of the specie *L. boulardi*, generating multiple reads of low coverage with an approximate size of 100 base pairs, which were subjected to various bioinformatics tools resulting in clusters (groups multiple reads) which were assembled in contigs (overlapping sequences defined by multiple reads), a portion of these showed homology with elements of TE's database then representing 5.19% of wasp genome belonging to 15 different elements superfamilies. An analysis of the elements described in the literature was performed, with the wasp sequences and literature sequences were used some softwares for alignment, best amino acid substitution model and mounting the phylogenetic tree. 11 trees of different elements superfamilies that presented proteins domains described in the literature as belonging to the structure of these elements were constructed. 26 taxa of elements described in the literature and 52 not yet described were characterized, which may indicate a new taxa of TEs, besides possible HTT processes between wasps and ants were identified. The pipeline built in this study was efficient when

analyzing the quality of the threads and the high value of branch support present in phylogenetic trees, indicating the effectiveness of the methodology used, which allows an economy in lines of research using the simultaneous sequencing of various organisms.

Keywords: Transposable elements, Bioinformatics, Horizontal Transfer, Phylogeny, parasitoid wasp.

SUMÁRIO

SUMÁRIO.....	6
1 Introdução.....	12
1.1 Classificação Geral dos Elementos Transponíveis.....	12
1.2 Superfamílias de Elementos Transponíveis.....	13
1.3 Elementos Transponíveis em Insetos	14
1.4 Vespas parasitoides e Transferência Horizontal de Elementos Transponíveis	14
1.5 <i>Leptopilina boulardi</i>	15
2 Justificativa.....	16
3 Objetivos	17
3.1 Objetivo geral	17
3.2 Objetivos específicos.....	17
4 Metodologia	18
4.1 Obtenção dos reads brutos.....	18
4.2 Identificação Genômica.....	19
4.3 Montagem dos Contigs.....	19
4.4 Criação de bancos de dados.....	19
4.4.1 Busca por proteínas homólogas	20
4.4.2 Busca por elementos homólogos	20
4.4.3 Análise da literatura.....	20
4.5 Alinhamento	20
4.6 Modelo de substituição de aminoácidos.....	21
4.7 Construção de árvores filogenéticas	21
5 Resultados	22
5.1 Polinton	23
5.2 Helitron.....	27
5.3 Chapaev	30
5.4 R1	30
5.5 Penelope	32
5.6 Loa.....	35
5.7 L2.....	35
5.8 I.....	36
5.9 Gypsy.....	39
5.10 BEL	39

5.11 Cópia	45
6 Discussão.....	48
7 Considerações Finais.....	49
REFERÊNCIAS	50
APÊNDICE	55

1 Introdução

1.1 Classificação Geral dos Elementos Transponíveis

Elementos transponíveis (TEs, do inglês *Transposable Elements*), ou segmentos móveis do material genético, foram descobertos em milho (*Zea mays*) por Barbara McClintock na década de 40, os quais foram descritos nos primeiros artigos como segmentos duplicados, modificações cromossômicas, aberrações cromossômicas, eventos de transposição e por fim em 1956 de elementos de transposição¹⁻⁵. Descrições mais atualizadas definem Elementos Transponíveis como genes móveis que possuem a capacidade de replicarem-se e moverem-se de uma região cromossômica para outra dentro de um genoma hospedeiro ou entre genomas⁶.

Desde então, estes elementos têm sido descritos em uma grande variedade de eucariotos⁷. Com o desenvolvimento das técnicas de sequenciamento e o subsequente sequenciamento dos primeiros genomas, identificou-se que estes elementos podem constituir grande parte do genoma de diversos organismos, chegando a 50% do genoma dos primatas⁸⁻¹⁰ e 85% do genoma do milho¹¹. Sua grande prevalência e distribuição sugerem que esses elementos influenciam diretamente a evolução de organismos que estes parasitam, atuando, por exemplo, no desenvolvimento de sistema imunológico¹² e na dinâmica dos cromossomos¹³. Estas modificações provocadas por elementos geralmente estão associadas primeiramente à processos deletérios e, eventualmente, podendo resultar a benefícios ao organismo, levando a eventos de domesticação, onde parte desses elementos passam a exercer funções importantes para o genoma do organismo.

Vários fatores são observados para caracterizar estes elementos, o tipo de ácido nucleico utilizado no mecanismo de transposição, regiões terminais de sequência invertida (TIRs do inglês *Terminal Inverted Repeats*), sítios alvo de duplicação (TSD do inglês *Target Site Duplication*), regiões longas de terminação (LTRs do inglês *Long Terminal Repeats*) e também as proteínas que compõem a estrutura do elemento e que estão relacionadas à mecanismos de transposição¹⁴.

A primeira classificação desses elementos foi proposta por Finnegan em 1989, este modelo era baseado unicamente no mecanismo de transposição dos elementos, separando os elementos em Classe I, os que utilizavam RNA como intermediário, e Classe II, os que utilizavam DNA como intermediário¹⁵.

Até o presente estudo não existe uma regra universal de classificação para os TEs,

existem dois principais sistemas de classificação, o de *Wicker et al*¹⁶ e o de *Kapitonov&Jurka*¹⁷. Ambos os sistemas classificam elementos transponíveis em dois grupos distintos, Transposons de DNA e Retrotransposons, baseado na maneira de mobilização destas sequências, mas não se limitando apenas ao mecanismo de transposição, utilizando também as características específicas estruturais presentes nas sequências de TEs, resultando numa classificação mais específica, com diferentes classes de TEs e sistemas de nomenclatura. Neste trabalho foi utilizada a classificação de *Kapitonov&Jurka*, a mesma utilizada na ferramenta RepeatExplorer¹⁸ e no banco de dados RepBase¹⁹, que classifica os elementos em Tipo 1 e Tipo 2.

TEs do Tipo 1, que utilizam o DNA como intermediário para sua transposição, realizam o processo por um mecanismo conhecido como recorta-e-cola (do inglês *cut-and-paste*)²⁰. Estes elementos também são encontrados em procariotos, em formas simples, chamadas de sequência de inserção (IS, do inglês *insertion sequence*)²¹. O mecanismo recorta-e-cola dos elementos de tipo 1 pode ser realizado de duas formas. O que divide este grupo em dois subgrupos. O primeiro grupo apresenta regiões TIRs que servem como sinal para a enzima Transposase cortar a dupla fita de DNA e iniciar a transposição do elemento. O segundo grupo realiza a transposição utilizando apenas uma fita de DNA como intermediário, sem necessidade de cortar a dupla fita, estes elementos são conhecidos como elementos de círculo rolante¹⁶.

De maneira diferente, os Retrotransposons, TEs do Tipo 2, que utilizam o RNA como intermediário para sua transposição, realizam o processo por um mecanismo conhecido como copia-e-cola (do Inglês, *copy-and-paste*), dessa forma, resultando na duplicação do elemento em uma nova localização no genoma²². TEs de tipo 2 subdividem-se em dois grandes grupos: Retrotransposons LTR e não-LTR. Retrotransposons LTR variam de centenas a milhares de pares de bases, possuem dois quadros de leitura aberta (ORFs, do inglês *open reading frames*), GAG e POL [17-19]. GAG codifica proteína para replicação, e POL codifica protease (PR), Integrase (INT), Transcriptase Reversa (RT) e RNase H (RH). Retrotransposons não-LTR possuem geralmente um tamanho entre 3-8kb, e um domínio de Transcriptase Reversa (RT) essencial para sua transposição²³.

1.2 Superfamílias de Elementos Transponíveis

De acordo com a classificação de *Kapitonov&Jurka* no Repbase¹⁷ foram definidas, até o momento, 61 superfamílias de elementos, sendo 24 superfamílias representadas por

elementos de Tipo 1 (Polinton, Helitron, EnSpm/CACTA, Chapaev, hAT, Harbinger, ISL2EU, Kolobok, Mariner/TC1, Merlin, MuDR, Novosib, P, piggyBac, Transib, Crypton, Sola, Zator, Ginger1, Ginger2/TDD, Academ, Zisupton, IS3EU, Dada), 33 superfamílias representadas por elementos de tipo 2 Não-LTR (Sine, CRE, NeSL, R4, R2, L1, RTE, I, Jockey, CR1, Rex1, RandI, Penelope, Tx1, RTEC, Crack, Nimb, Proto1, Proto2, RTETP, Hero, L2, Tad1, Loa, Ingi, Outcast, R1, Daphne, L2A, L2B, Ambal, Vingi, Kiri) e 4 superfamílias representadas por elementos de tipo 2 LTR (Copia, Gypsy, Bel, DIRS).

1.3 Elementos Transponíveis em Insetos

O Genoma nuclear de insetos pode variar em até duas ordens de grandeza, partindo do genoma de 90 Mb do díptero *Mayetiola destructor* até 16 Gb no gafanhoto da montanha *Podisma pedestris*^{24,25}. Apesar da existência destas grandes variações em organismos de uma mesma classe, o paradoxo de valor C^{26} postula que o tamanho do genoma não possui correlação com a aparente complexidade de um organismo. Este paradoxo foi parcialmente resolvido através da descoberta de grandes quantidades de DNA sem função aparente, chamados de DNA lixo, fazendo parte deste grupo todos os elementos repetitivos (TEs, elementos virais endógenos, repetições em série, duplicações segmentais, ribozimas), sendo predominantemente responsável pela variação de tamanho de genomas haplóides²⁷.

Em análises de genomas sequenciados de insetos foi identificado que TEs contribuem diretamente no tamanho do genoma, representando menos de 1% no genoma de 99 Mb de um mosquito antártico (*Belgica antarctica*)²⁸ e mais de 60% no genoma de *Locusta migratoria*, um gafanhoto com o maior genoma de inseto sequenciado até o momento (6.5 Gb)²⁹. Esta correlação de quantidade de TEs e tamanho de genoma ocorre também em organismos muito próximos evolutivamente, em 12 espécies do gênero *Drosophila* o conteúdo de TEs varia de 2,7% a 25%³⁰, e em mosquitos se encontra uma diferença de até 5 vezes no que diz respeito ao conteúdo de TEs, variando de 11% em *Anophele gambiae* (genoma de 278 Mb) a 47% em *Aedes aegypti* (genoma de 1.3 Gb)²⁷.

A abundância de TEs em genomas de insetos é altamente variável, e num panorama geral, elementos de Tipo 1 são mais abundantes²⁷.

1.4 Vespas parasitoides e Transferência Horizontal de Elementos Transponíveis

Eventos de transferência gênica podem ocorrer de duas formas, de organismos ancestrais para seus descendentes, também conhecido como transferência vertical, ou entre organismos de linhagens diferentes, conhecido como transferência horizontal ou lateral. Caso o evento de transferência horizontal envolva TEs, é denominado Transferência Horizontal de Elementos Transponíveis (HTT, do inglês *Horizontal Transposon Transfer*), processo importante no ciclo de vida de TEs, o qual equivale à colonização de um novo nicho ecológico por um determinado TE³¹.

Vespas parasitoides são himenópteros que parasitam vários estágios de desenvolvimento de outros artrópodes, como por exemplo, insetos, carrapatos e ácaros. As fêmeas deste grupo injetam veneno e põem ovos sobre ou dentro do hospedeiro, culminando no consumo e eventualmente na morte do hospedeiro durante o desenvolvimento de sua prole³².

Para que este processo tenha êxito duas etapas são essenciais, a queda do sistema imunológico sem que ocorra a morte do hospedeiro, e a proteção dos ovos da vespa. Deste modo, o veneno destas vespas parasitoides são compostos por diversas moléculas, sendo as VLPs (do inglês *Viral Like Particles*) fatores chave neste processo, responsáveis por destruir lamelócitos, um tipo de célula da hemolinfa, produzidas em respostas à infecções parasitárias, que agem reconhecendo e destruindo grandes objetos estranhos³³.

Estudos sugerem que VLPs podem carregar DNA e TEs, dessa forma, podendo atuar como vetor potencial para transferência horizontal de elementos transponíveis³⁴. Fragmentos de TEs de diferentes superfamílias (Mariner, Gypsy, Maverick e DIRs) foram identificados como parte do genoma de Polyadenovirus, estes são os vírus injetados como VLPs por algumas vespas parasitoides de borboletas³⁵.

1.5 *Leptopilina bouvardi*

Leptopilina bouvardi (Figitidae, superfamília Cynipoidea) é uma vespa parasitóide especializada em parasitar organismos do subgrupo *melanogaster* do gênero *Drosophila*. Análises da glândula de veneno de indivíduos fêmea da espécie *L. bouvardi* demonstraram a existência de VLPs como parte da composição do veneno³⁶. Dessa forma, a vespa *L. bouvardi* surge como um possível organismo modelo no estudo de HTT.

2 Justificativa

Com o início da era genômica a caracterização de TEs aumentou exponencialmente devido à disponibilidade de genomas de espécies modelos. Entretanto, em espécies não modelo, como *Leptopilina boulardi*, por exemplo, o conhecimento é escasso sobre essas sequências repetitivas. Associado a ausência de conhecimento está o alto custo de um projeto genoma para caracterizar as sequências repetitivas e a ausência de uma metodologia que demande baixo poder computacional.

3 Objetivos

3.1 Objetivo geral

Caracterizar o conteúdo de Elementos Transponíveis e avaliar a história evolutiva destes elementos presentes no genoma de *Leptopilina boulardi* utilizando sequenciamento de baixa cobertura.

3.2 Objetivos específicos

- Padronização de um *pipeline* que permita o uso de sequenciamento de baixa cobertura para caracterização e análise evolutiva de Elementos Transponíveis;
- Caracterização e análise evolutiva dos Elementos Transponíveis em *Leptopilina boulardi*;
- Detecção de possíveis eventos de Transferência Horizontal de Elementos Transponíveis.

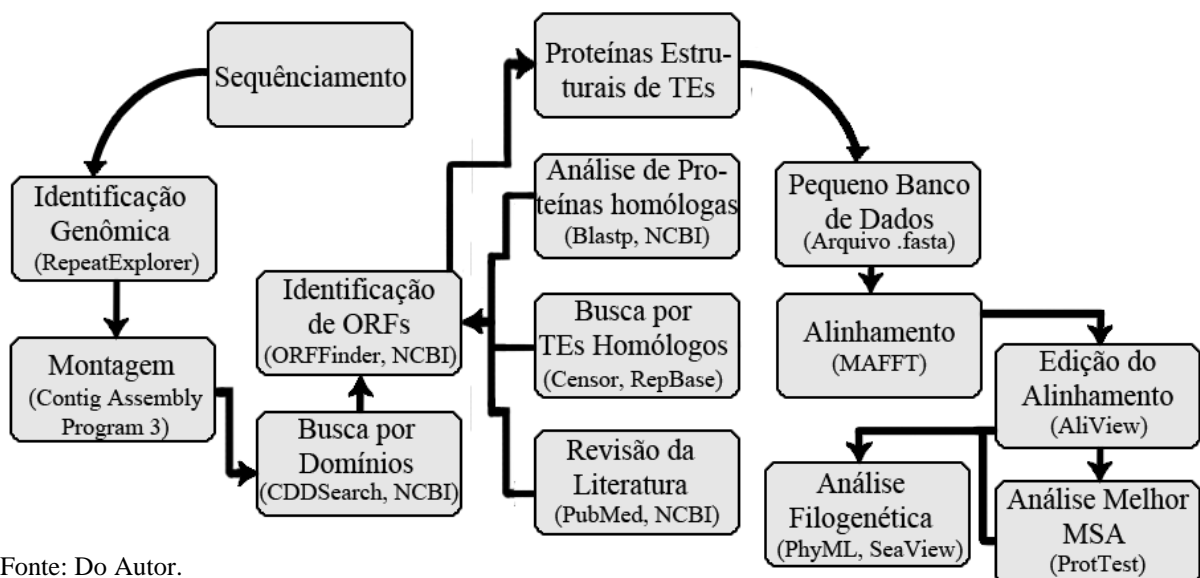
4 Metodologia

Foi desenvolvido um *pipeline* (Figura 1), permitindo identificar diferentes superfamílias de TEs em determinado genoma e sua proporção.

4.1 Obtenção dos reads brutos

O genoma de *L. boulandi* foi cedido pelo LabDros (Laboratório de Biologia Molecular e Sequenciamento, de Santa Maria-RS), o qual seguiu o seguinte procedimento: DNA total foi isolado a partir de 20 indivíduos da espécie *L. boulandi*, usando NucleoSpin Tissue XS (MACH-EREY-NAGEL), seguindo o protocolo do fabricante. DNA genômico foi sequenciado através do serviço Fasteris DNA Sequencing Service, em dispositivo Solexa Illumina-HiSeq 2000 New Generation Sequencing (NGS) de acordo com as instruções do fabricante, gerando um *dataset* com *reads* de 100 pb, apresentando cobertura média de 12.36x valor obtido multiplicando o número total de *reads* obtidos no sequenciamento pelo tamanho médio dos *reads* e dividindo pelo tamanho do genoma de *Nasonia vitripennis* (250mb), vespa próxima evolutivamente de *L. Boulandi* com o genoma sequenciado e montado.

Figura 1 - Pipeline para caracterização genômica e análise filogenética de TEs utilizando *reads* brutos provenientes de *High Throughput Sequencing* ou *Next Generation Sequencing*.



Fonte: Do Autor.

4.2 Identificação Genômica

Primeiramente foi realizado o *upload* do arquivo de sequenciamento no *webserver* RepeatExplorer da plataforma Galaxy utilizando o *software* FilleZila. Seguido de um pré-processamento para conversão das sequências para formato FastQ sanger com a ferramenta FASTQGroomer. Terminada a conversão foi utilizada a ferramenta Filter by quality, onde os *reads* foram filtrados com um valor mínimo de *Cut-off* 20, os arquivos foram então convertidos para o formato Fasta. Terminadas as etapas de pré-processamento foi utilizada a ferramenta Clustering, resultando em *clusters* representados por *reads* altamente similares, juntamente com a busca de similaridade destes clusters com TEs já descritos no banco de dados RepBase. Foi realizada uma reclusterização utilizando a ferramenta Cluster merger, selecionando *clusters* que apresentavam mais de 100 *hits* e pertenciam à uma mesma superfamília de TEs presente no RepBase, para melhorar a qualidade dos *clusters*. Finalizadas estas etapas foi realizado o *download* dos arquivos via servidor FTP, resultando em diretórios contendo *reads* em arquivos formato Fasta separados de acordo suas respectivas superfamílias de TEs.

4.3 Montagem dos Contigs

Foi utilizada a ferramenta CAP3³⁷ (Contig Assembly Program v.3) nos *top clusters* (*clusters* mais representativos e com homologia a TEs) provenientes do RepeatExplorer para montagem dos *contigs*. Esta ferramenta foi executada por linha de comando em terminal Linux utilizando os seguintes parâmetros: -a 20 -b 20 -c 12 -d 200 -e 30 -f 20 -g 6 -m 2 -n 5 -p 80 -r 1 -s 900 -t 300 -u 3 -v 2 -o 40. Gerando sequências nomeadas *contigs*, ou seja, sequências formadas a partir da sobreposição de vários *reads*.

4.4 Criação de bancos de dados

Através da revisão da literatura foram identificadas quais proteínas estão presentes na estrutura de cada superfamília de TEs a ser analisada, permitindo definir qual proteína utilizar para montar cada banco de dados. Após esta primeira análise foram seguidas três metodologias:

4.4.1 Busca por proteínas homólogas

Foi utilizada a ferramenta CDSearch³⁸ para verificar a presença de domínios proteicos nos *contigs* correspondentes à TEs, seguido do uso da ferramenta ORF Finder³⁹, contra o banco de dados de proteínas não redundantes do NCBI, para buscar sequências de aminoácidos correspondentes aos domínios encontrados com a ferramenta CD-Search. Após a obtenção destas sequências, foi realizado um Blastp “jogando” a proteína do *contig* mais representativo de determinada superfamília contra o banco de dados do NCBI, e dos resultados foram recuperadas 50 sequências proteicas de diferentes espécies que obtiveram melhor *score* durante a análise de Blastp.

4.4.2 Busca por elementos homólogos

Os *contigs* que apresentaram ORFs específicas em sua estrutura, na análise anterior, foram jogados contra um banco de dados de TEs, RepBase, retornando vários elementos homólogos com diferentes similaridades, destes, o elemento com maior valor de *score* para cada elemento identificado em *L. bouvardi* foi utilizado na ferramenta ORF Finder para *download* da sequência proteica alvo.

4.4.3 Análise da literatura

Para maior robustez da análise foi realizada uma revisão da literatura para cada superfamília de TEs analisada, identificando elementos que ainda não foram depositados no banco de dados RepBase, as sequências destes elementos foram utilizadas na ferramenta ORF Finder e foi realizado o *download* da sequência proteica alvo.

4.5 Alinhamento

Foi utilizado o *software* MAFFT com método progressivo G-INS-i, método baseado no algoritmo de *Needleman-Wunsch* que utiliza uma matriz de *score* para verificar similaridade entre sequências de diferentes tamanhos⁴⁰. Este *software* disponibiliza um

arquivo com extensão fasta, este arquivo foi então utilizado no *software* AliView para edição manual do alinhamento, excluindo regiões pouco sobrepostas e mantendo blocos conservados de aminoácidos. Após a edição manual do alinhamento foram analisadas o tamanho das ORFs visando manter um padrão de tamanho, sequências que sofreram muitas deleções durante a edição foram retiradas do arquivo.

4.6 Modelo de substituição de aminoácidos

Nesta etapa foi utilizado o *software* ProtTest3.4⁴¹ baseado em um algoritmo que busca identificar qual o melhor modelo de substituição entre 112 possíveis modelos de substituição de aminoácidos.

4.7 Construção de árvores filogenéticas

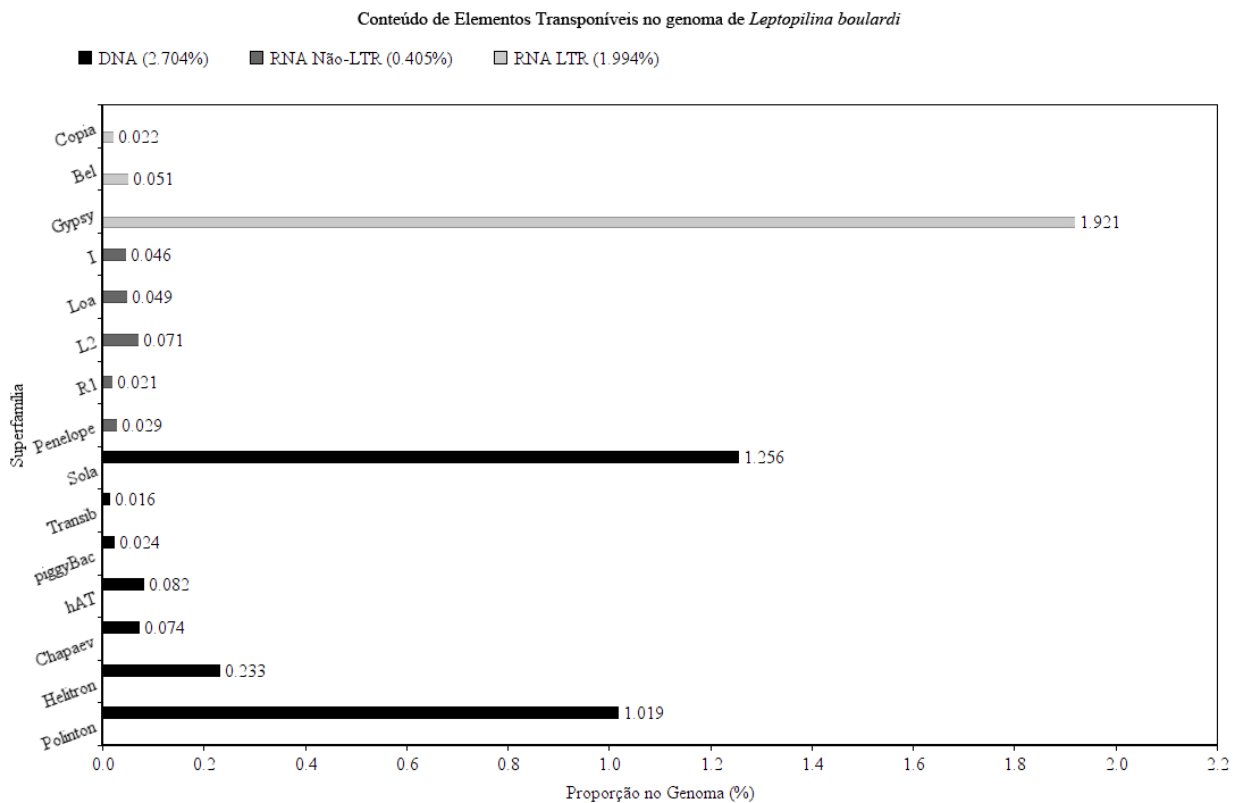
Foi utilizada a ferramenta PhyML do *software* SeaView para construção de árvores com máximo verossimilhança (*Maximum-Likelihood Tree*) utilizando suporte de ramo baseado em teste de aproximação da taxa de verossimilhança (aLRT), ao contrário do suporte de ramo *bootstrap* que é baseado pela repetição na formação de determinados agrupamentos, o suporte de ramo aLRT avalia três possibilidades de formação de ramo (a melhor possibilidade, a segunda melhor, e a possibilidade de não ocorrer determinado agrupamento) baseado na similaridade das sequências, e utiliza estes valor para calcular o suporte de ramo, dessa forma, sem a necessidade de reconstrução de centenas de árvores, a ferramenta PhyML apresenta uma velocidade muito superior de análise e uma demanda de poder computacional muito menor que análises baseadas em *bootstrap*⁴². As sequências que não possuíam grupo externo na literatura foram submetidas á distribuição de ramos *MindPointRoot*, que verifica o melhor modelo de agrupamento baseado na distância média entre todos os ramos da árvore.

5 Resultados

Foram identificadas 15 superfamílias diferentes de TEs através da análise com a ferramenta RepeatExplorer, indicando que TEs representam cerca de 5,19% do genoma de *L. bouleari* (Figura 2). Sendo os elementos de Tipo 1, que utilizam DNA como intermediário para transposição, os mais representativos, representando 2,7% do genoma da vespa, dentro deste grupo, TEs das superfamílias Sola e Polinton são os mais abundantes, representando 1.25% e 1.01% respectivamente. Elementos de Tipo 2, que utilizam RNA como intermediário para transposição, representaram cerca de 2,5% do genoma, sendo os TEs da superfamília Gypsy os mais abundantes, representando 1,92%.

Foram construídos 11 bancos de dados, o que possibilitou a reconstrução da história evolutiva de TEs de 11 superfamílias: Polinton, Helitron, Chapaev, R1, L2, Loa, I, Penelope, Gypsy, BEL e Copia. Apesar da grande representatividade de TEs da superfamília Sola não foram encontradas ORFs nas sequências dos *contigs*, isto pode ocorrer devido a degenerações provocadas por mutações na sequência de TEs muito antigos, o mesmo ocorreu em *contigs* das superfamílias hAT, piggyBac e Transib.

Figura 2. Superfamílias de TEs e sua proporção no genoma *L. bouleari*, valores obtidos na ferramenta RepeatExplorer.



5.1 Polinton

Elementos desta superfamília são encontrados no genoma de protistas, fungos e animais, sendo caracterizados por um conjunto bem definido de proteínas necessárias para sua transposição, dessa forma sua estrutura é caracterizada por 4 ORFs, na seguinte ordem: DNA polimerase B, Integrase retroviral, Protease de cisteína e ATPase, além deste conjunto de proteínas sua sequência é composta por regiões TSDs de 6pb e regiões TIRs de 100~1000 pb resultando numa sequência de 15~20kb, sendo um dos maiores elementos transponíveis podendo codificar até 10 proteínas⁴³².

Foram identificados 12 *contigs* que representam TEs da superfamília Polinton e que apresentaram Integrase em sua estrutura (Tabela 1), os quais formaram dois clados distintos (Figura 3), indicando a existência de dois táxon distintos dentro da superfamília Polinton, tendo estes divergido evolutivamente de sequências que vieram a ser classificadas como Polinton de *Nasonia vitripennis*, uma vespa parasitoide da família *Pteromalidae*.

TABELA 1

Características das sequências correspondentes à elementos da superfamília Polinton presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados Encontrados no CDD	Tamanho da ORF (AA)
CL2Contig46	7809	rve, Parvo_coat, GAT_1, katE, EpsG, ND2, Srg, ZF, GPCR, glyco_rpt, CHROMO, PHA02517, YadH,	308
CL2Contig8	4842	rve, CHROMO, Pox_A32, PTZ00440, ND5, EpsG, PMT_2, GPCR	308
CL2Contig67	4219	DNA_pol_B, DUF, Endonuclease_7	1276
CL2Contig31	4180	DNA_pol_B, PolBc, COG4267, EpsG, glyco_rpt, DUF	1127
CL2Contig315	4158	DNA_pol_B, Endonuclease_7, DUF	1276
CL2Contig2	3530	DNA_pol_B2, CHROMO, EpsG	771
CL2Contig1	3494	DNA_pol_B2, CHROMO, Sulfolob_CbsB	888
CL2Contig197	3494	DNA_pol_B, CHROMO, PHA, sulfolob_CbsB	888
CL2Contig25	3303	DNA_pol_B, DUF, ND6, O-ag_pol, glyco_rpt, EpsG	795
CL2Contig58	3035	DNA_pol_B, EpsG	957
CL2Contig1535	2919	DNA_pol_B, EpsG	958
CL2Contig287	2900	DNA_pol_B	800
CL2Contig1347	2674	DNA_pol_B, POLBc, TEL1, PHA, DUF	786
CL2Contig3	2590	DNA_pol_B2, EpsG, MFS, GPCR, DUF, ND5, Borrelia_orfA	732
CL2Contig182	2494	DNA_pol_B, ND2, ND5, ND6, DUF, EpsG, VanZ, PSN, SLC5, Gpos, glyco_rpt, YadH, MFS, NosY,	699
CL2Contig49	2423	rve, CHROMO, PHA, EpsG, glyco_rpt	308
CL2Contig5	2375	rve, CHROMO, PHA, ND5, 7TM	308
CL2Contig61	2294	rve, CHROMO, MSP7_C, 7TM-GPCR, glyco_rpt	307
CL2Contig1494	2271	rve, CHROMO, PHA	338
CL2Contig36	2182	rve, EpsG, Ndc1_Nup, 7TM_GPCR	157
CL2Contig1489	2171	DNA_pol_B, Endonuclease_7, EpsG, O-ag_pol_Wzy, glyco_rpt	701
CL2Contig336	2135	DNA_pol_B, EpsG, ND4, ND6	612
CL2Contig1046	1991	rve, CHROMO, Ndc1, ND5, 7TM, EpsG, ND2, PHA	157
CL2Contig20	1956	rve, CHROMO, ND5, EpsG	165
CL2Contig33	1939	rve, CHROMO, HNHc, 7TM_GPCR	308

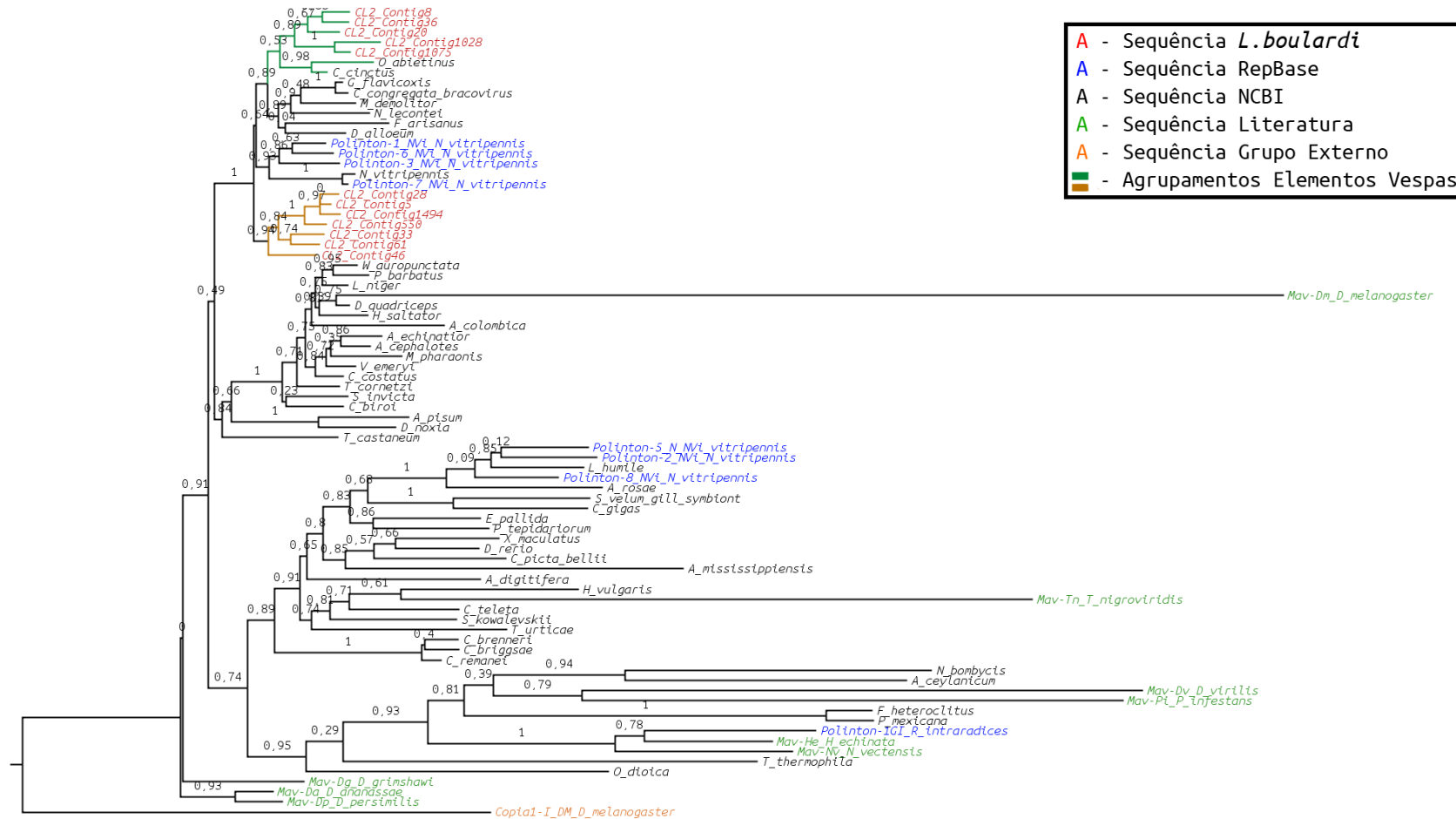
Continuação Tabela 1

CL2Contig19	1869	DNA_pol_B, Endonuclease_7	507
CL2Contig15	1832	DNA_pol_B2, CHROMO	444
CL2Contig1075	1830	rve, CHROMO, PHA, ND2, EpsG, ND5, glyco_rpt, mplasa, Tra5	307
CL2Contig484	1806	DNA_pol_B, ND4	499
CL2Contig44	1612	DNA_pol_B	528
CL2Contig501	1590	AAA, ND5, ND6, EpsG	181
CL2Contig201	1560	DNA_pol_B, PRK	422
CL2Contig1028	1502	rve, glyco_rpt, PHA, DltB, O-ag_pol_wzy	224
CL2Contig28	1493	rve, PHA00430, Molybdopterin	166
CL2Contig483	1474	rve, CHROMO, ND5, PHA	307
CL2Contig1256	1446	DNA_pol_B, PHA, POLBc	473
CL2Contig1011	1444	AAA, Pox, Mplasa, PTZ, HsdS, Gluco_trans, ND5, ND6, EpsG, 7TM, glyco_rpt	258
CL2Contig1631	1379	DNA_pol_B, Endonuclease	448
CL2Contig550	1289	rve, CHROMO, PHA	306
CL2Contig442	1247	DNA_pol_B, TRAM	400
CL2Contig1518	1176	DNA_pol_B, Endonuclease	350
CL2Contig1214	1070	DNA_pol_B, RPA, PHA	333
CL2Contig444	1060	rve, CHROMO, HNHc, 7TM	308

Fonte: Do Autor.

Legenda: rve: integrase core domain; Parvo_coat: Parvovirus coat protein; GAT_1: Type 1 glutamine amidotransferase; katE: Catalase; EpsG: Glycosyl Transferases; ND: NADH dehydrogenase subunit; Srg: chemoreceptor; ZF: Zinc finger; GPCR: G Protein Coupled Receptor; glyco_rpt: oligosaccharide repeat unit polymerase; CHROMO: Chromatin Organization Modifier; PHA02517: putative transposase; YadH: multidrug transport system; Pox_A32: Poxvirus A32 protein; PTZ00440: Reticulocyte binding protein; PMT_2: Dolichyl-phosphate-mannose-protein mannosyltransferase; DNA_pol_B: DNA polymerase family B; DUF: Protein of Unknown function; Endonuclease_7: Recombination Endonuclease; POLBc: DNA polymerase type B; COG4267: Uncharacterized membrane protein; Sulfolob_CbsB: Cytochrome b558/566; PHA: Polyhydroxyalkenoate synthesis regulator protein; O-ag_pol: O-Antigen polysaccharide polymerase. TEL1: Phosphatidylinositol kinase; MFS: Major Facilitator superfamily; VanZ: Glycopeptide antibiotics resistance; PSN: Presenilin signal peptide peptidase; NosY: ABC-type transport system; Ndc1_Nup: Nucleoporin protein; , HNHc: HNH nuclease; Tra5: Transposase; PRK: Phosphoribulokinase; Mplasa: helix-rich Mycoplasma protein; HsdS: Restriction Endonuclease S; Gluco_Trans: Glucosyl transferase; TRAM: Unknown function. Sequências com fundo cinza foram utilizadas na filogenia.

Figura 3. Análise filogenética da superfamília Polinton. Sequências vermelhas representam integrases de elementos presentes no genoma de *L. boulandi*; Sequências pretas representam integrases (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam integrases de elementos retirados do banco de dados RepBase; Sequências verdes representam integrases de elementos descritos na literatura; Sequência laranja representa transcriptase reversa de elemento Copia utilizado como grupo externo. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice A.



5.2 Helitron

Conhecidos como elementos de círculo rolante, estes elementos foram descritos primeiramente em genoma de plantas (*Arabidopsis thaliana* e *Oryza sativa*) e em nematódeo (*Caenorhabditis elegans*)⁴⁴ e até o momento foram identificados em vários organismos, de protistas à mamíferos, representando uma das mais importantes superfamílias de elementos transponíveis, parecendo ter um papel importante na evolução de genomas hospedeiros, onde frequentemente capturam genes dos hospedeiros podendo super-expressar esse gene devido ao seu mecanismo de transposição (que cria várias cópias do elemento no genoma) ou utilizar este gene na sua transposição⁴⁵. Sua característica principal se baseia numa região conhecida como RepHel (Rep do inglês *Replication initiator* e Hel fazendo alusão ao domínio da enzima Helicase), podendo conter diferentes domínios em sua sequência como CPR (Protease de cisteína), EN (*Apurinic*) e ZF (*Zinc finger*) e RPA (*Protein Replication A*), este último que codifica uma proteína similar a *Helicase* em algumas espécies⁴⁵.

Foram identificados 9 *contigs* que representam TEs da superfamília Helitron e que apresentaram Helicase em sua estrutura (Tabela 2), destes, 7 foram selecionados para construção da árvore filogenética, devido à similaridade entre suas Helicases, formando 3 clados distintos na árvore filogenética (Figura 4). Ficando evidente a formação de um clado apenas com elementos de *L. bouvardi*, um clado com elemento de *L. bouvardi* e um elemento ainda não caracterizado de *N. vitripennis* e um terceiro clado com elemento de *L. bouvardi* e elementos ainda não caracterizados de duas vespas parasitoides, *Diachasma alloeum* e *Fopius arisanus*.

TABELA 2

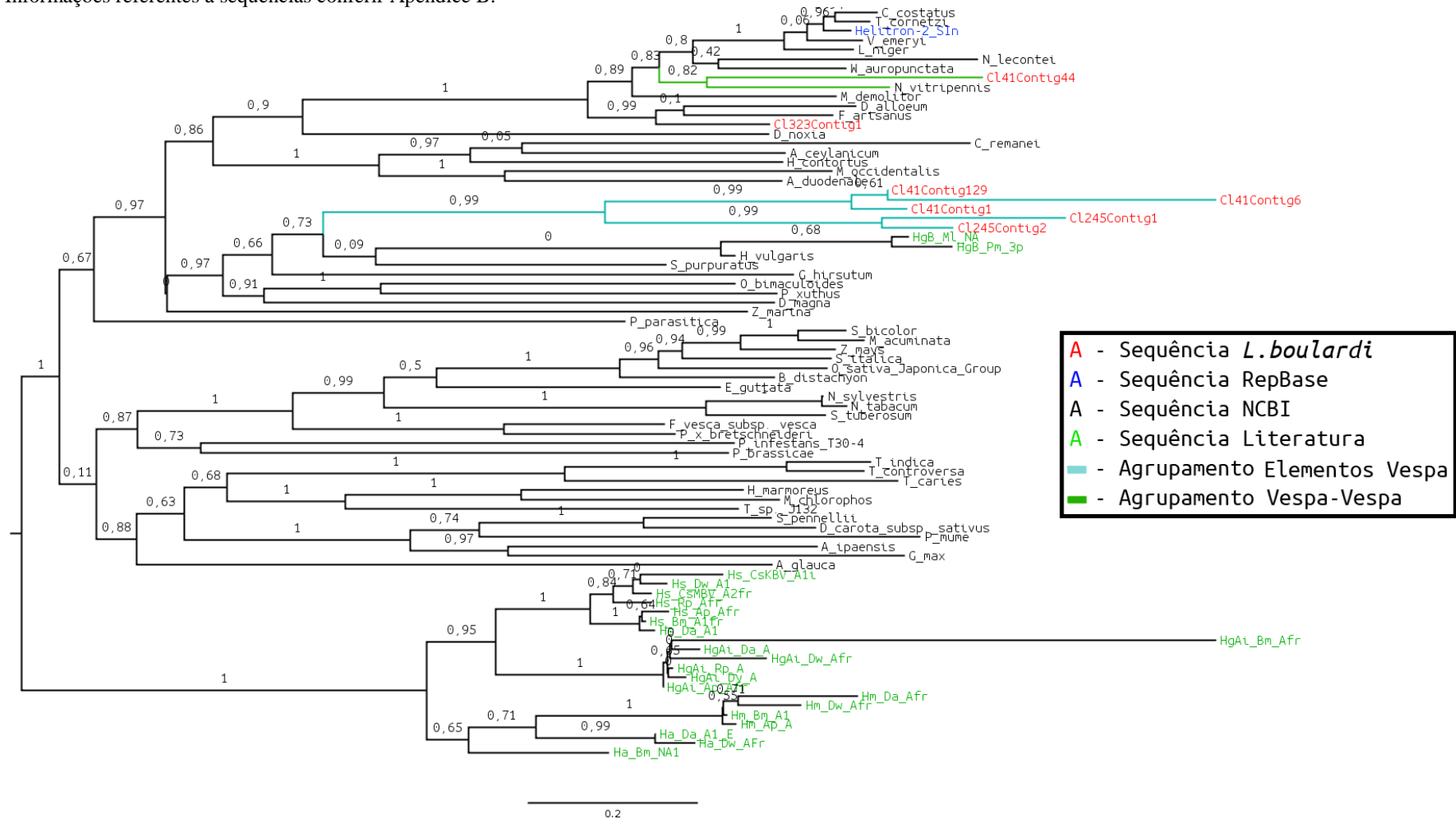
Características das sequências correspondentes à elementos da superfamília Helitron presentes no genoma de *Leptopilina bouvardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL41Contig1	1852	Helitron_like_n	608
CL41Contig6	567	Helitron_like_n	67
CL41Contig129	435	Helitron_like_n	131
CL41Contig44	369	Helitron_like_n	90
CL215Contig1	1590	PIF1	460
CL215Contig221	704	Herpes_Helicase	84
CL245Contig1	2550	Helitron_like_n, ND2, ND6, DUF, EpsG, glyco_rpt, Borrelia_orfA	326
CL245Contig2	519	Helitron_like_n	116
CL323Contig1	5255	Helitron_like_n, Herpes_Helicase, P-loop_NTPase	510
CL323Contig59	699	Herpes_Helicase, Hsm3, P-loop_NTPase	181

Fonte: Do Autor.

Legenda: Helitron_like_n: Helitron helicase domain at N-terminus; PIF1: PIF1-Like Helicase; Herpes-Helicase: Helicase from Herpes Virus; ND: NADH dehydrogenase subunit; DUF: Unknown Function; EpsG: Glycosyl Transferases; glyco_rpt: oligosaccharide repeat unit polymerase; Borrelia_orfA: protein encoded by Borrelia species with unknown function; P-loop_NTPase: P-loop containing Nucleoside Triphosphate Hydrolases; Hsm3: yeast proteosome chaperone . Sequências com fundo cinza foram utilizadas na filogenia.

Figura 4. Análise filogenética da superfamília Helitron. Sequências vermelhas representam helicases de elementos presentes no genoma de *L. boulandi*; Sequências pretas representam helicases (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam helicases de elementos retirados do banco de dados RepBase. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice B.



Fonte: Do Autor.

5.3 Chapaev

Representam uma superfamília relativamente nova de elementos transponíveis, identificados pela primeira vez no ano de 2007; sua estrutura é caracterizada por duas regiões TSDs de 4pb, duas regiões TIRs bem conservadas e duas ORFs, uma de zinc finger e outra de transposase⁴⁶, como em grande parte dos elementos transponíveis de Tipo 1 as transposases de *Chapaev* são caracterizadas pela presença de um motivo conservado “DDE” no seu domínio catalítico⁴⁷. A biologia da evolução destes elementos ainda é pouco conhecida, sabe-se que podem ter ocorrido eventos de transferência horizontal destes elementos, dado que são encontrados em espécies de mamíferos, répteis, peixes com mandíbulas, lampreias, insetos e bracovírus de inseto possuindo uma identidade de sequência nucleotídica que varia de 79 a 99%⁴⁸.

Apenas 1 *contig* que representa TEs da superfamília Chapaev e que apresentou zinc finger em sua estrutura foi identificado (Tabela 3), formando um clado com um elemento ainda não descrito da formiga *Vollenhovia emeryi* (Figura 5).

TABELA 3
Características das sequências correspondentes à elementos da superfamília Chapaev presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL231Contig1	2573	ZF	597

Fonte: Do Autor.

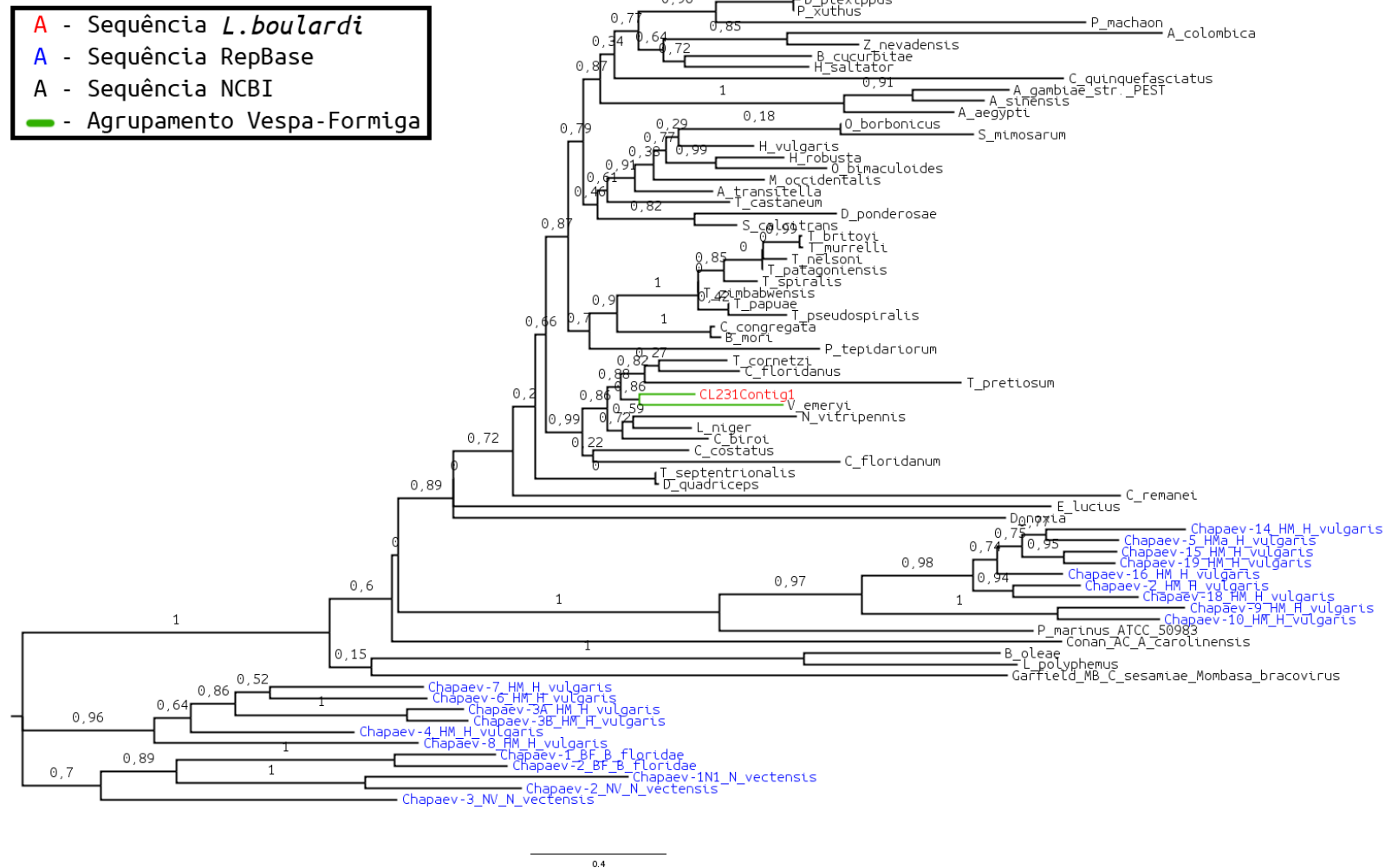
Legenda: ZF: Zinc Finger. Sequência com fundo cinza foi utilizada na filogenia.

5.4 R1

Uma superfamília de retrotransposons não-LTR, elementos desta superfamília se encontram inseridos em genes de RNA ribossomal porção 28S de artrópodes, e parecem estar presentes em várias linhagens de insetos desde a origem dos artrópodes⁴⁹. Inserções de TEs desta superfamília inibem a transcrição do gene 28S, sendo um eficiente inativador desta unidade gênica⁵⁰.

Oito *contigs* que representam os TEs da superfamília R1 foram identificados, destes, 6 apresentaram Transcriptase Reversa em sua estrutura (Tabela 4), formando um clado irmão de um clado formado pelo elemento SART presente em pulgão da ervilha (*Acyrtosiphon pisum*) (Figura 6).

Figura 5. Análise filogenética da superfamília Chapaev. Sequência vermelha representa *zinc finger* de elementos presentes no genoma de *L. boulandi*; Sequências pretas representam *zinc finger* (anotadas, previstas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam *zinc finger* de elementos retirados do banco de dados RepBase. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos WAG. Informações referentes à sequências conferir Apêndice C.



Fonte: Do Autor.

TABELA 4
Características das sequências correspondentes à elementos da superfamília R1 presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL72Contig131_singlet	992	RT_like	251
CL72Contig2	1081	RT_like	334
CL72Contig3	854	Endonuclease	200
CL72Contig7	5102	Endonuclease_R1, RT_like, RVT_1, PTZ	1008
CL72Contig82_singlet	437	RT_like	131
CL72Contig85_singlet	870	RT_like	268
CL149Contig1	1360	Endonuclease_R1, zf-CCCH	217
CL149Contig3	2970	Endonuclease_R1, RT_like, RVT_1	735

Fonte: Do Autor.

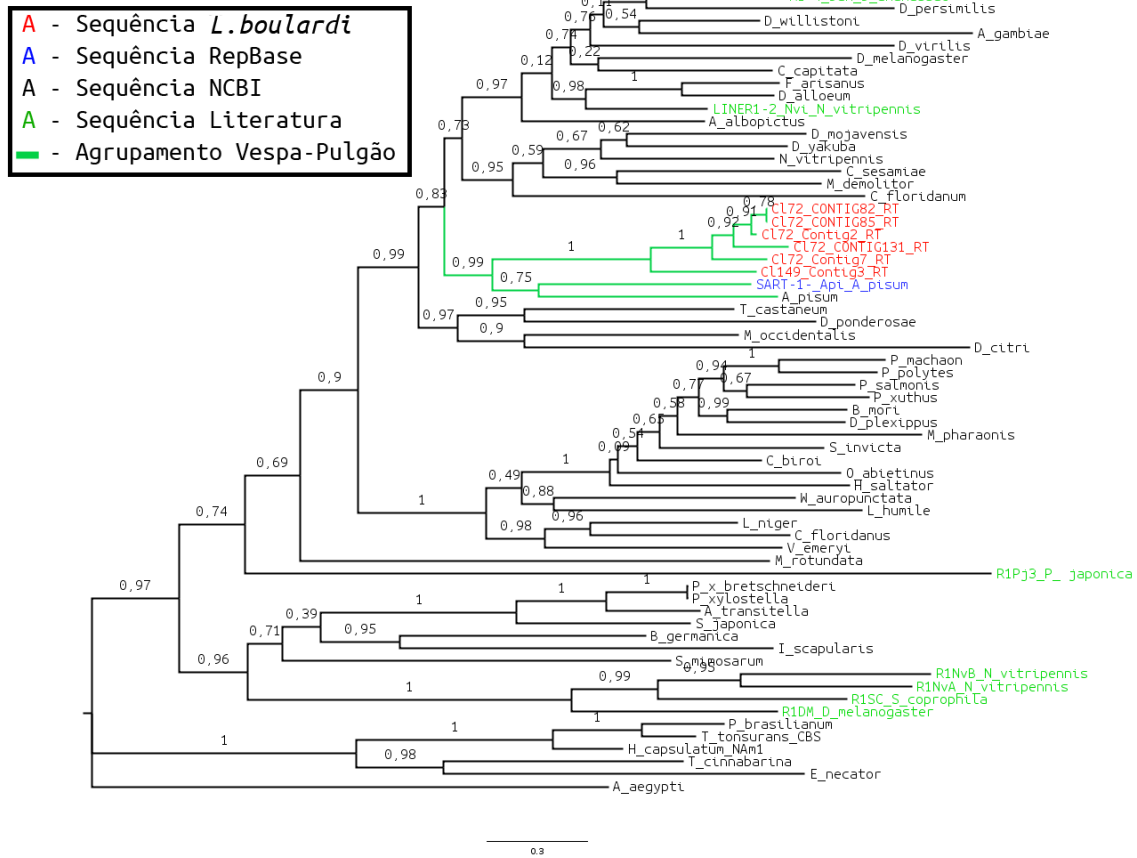
Legenda: RT_like: Reverse Transcriptase; RVT_1: Reverse Transcriptase; PTZ: Actin provisional; zf-CCCH: Zinc Finger; Endonuclease_R1: Endonuclease related in Retrotransposon R1. Sequência com fundo cinza foram utilizadas na filogenia.

5.5 Penelope

Muitos eventos de HTT envolvendo TEs da superfamília Penelope (PLEs, do inglês *Penelope-like Elements*) já foram identificados em mosca da fruta (*Drosophila melanogaster*)⁵¹⁻⁵². Duas linhagens de PLEs foram encontradas em eucariotos. A primeira linhagem apresenta dois domínios principais em sua estrutura, um domínio para transcriptase reversa e outro para endonuclease. A segunda linhagem apresenta apenas um domínio para transcriptase reversa⁵³. Análises filogenéticas indicaram que os domínios de transcriptase reversa de ambas as linhagens apresentam uma relação muito próxima com domínios de telomerasas, enzimas que são responsáveis pela estabilidade dos telômeros em cromossomos de organismos eucariotos⁵⁴.

Apenas 1 *contig* representando TEs da superfamília Penelope apresentou Transcriptase reversa em sua estrutura (Tabela 5), formando um clado com o elemento Penelope-1 de *N. vitripennis* e com um elemento ainda não identificado de *Trichogramma pretiosum*, uma vespa parasitóide (Figura 7).

Figura 6. Análise filogenética da superfamília R1. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. bouhardi*; Sequências pretas representam transcriptase reversa (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase; Sequências verdes representam transcriptase reversa de elementos descritos na literatura. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice D.



Fonte: Do Autor.

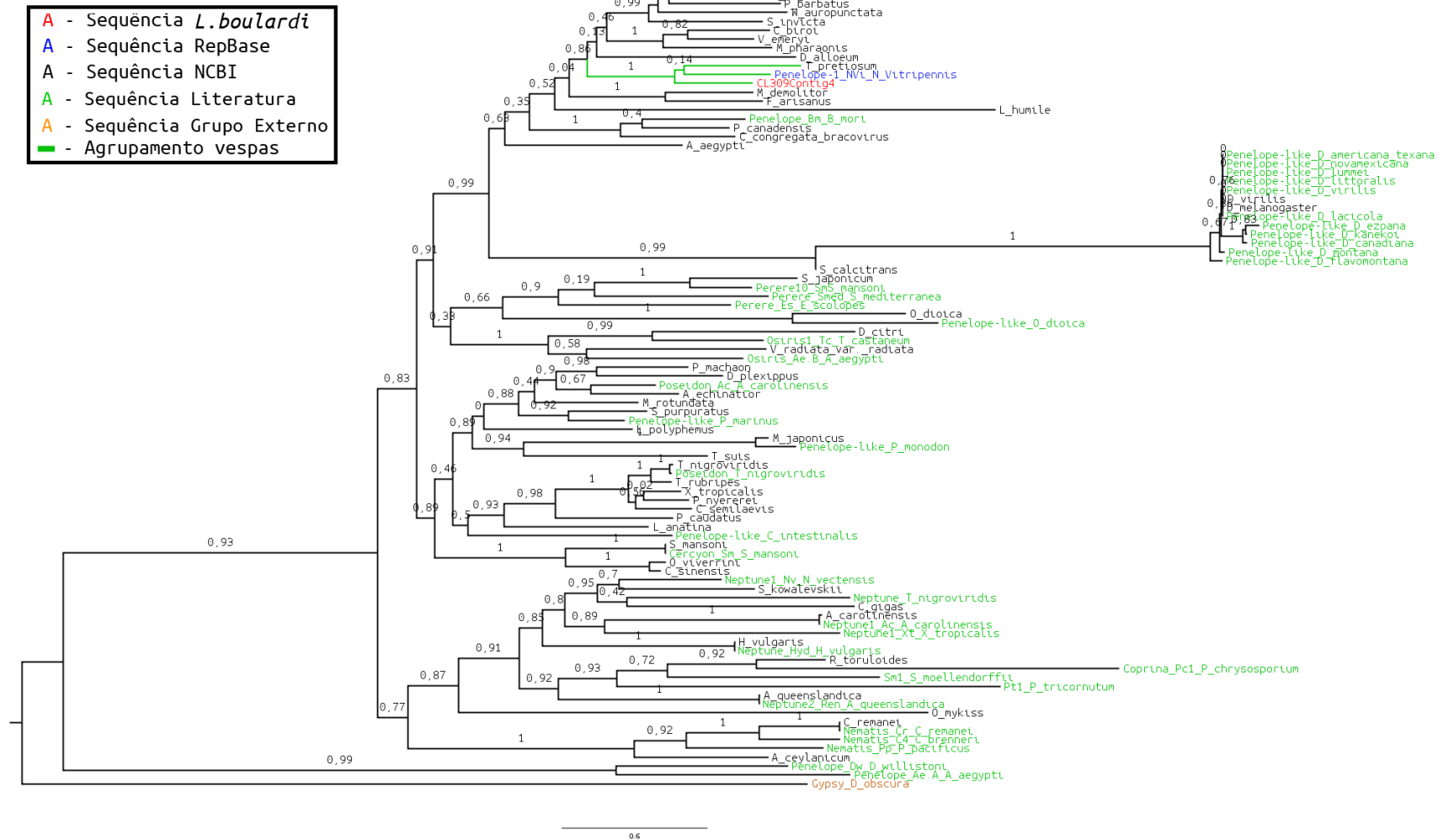
TABELA 5
Características das sequências correspondentes à elementos da superfamília Penelope presentes no genoma de *Leptopilina bouhardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL309Contig4	802	RT_like	255

Fonte: Do Autor.

Legenda: RT_like: Reverse Transcriptase. Sequência com fundo cinza foram utilizadas na filogenia.

Figura 7. Análise filogenética da superfamília Penelope. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. bouvardi*; Sequências pretas representam transcriptase reversa (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase; Sequências verdes representam transcriptase reversa de elementos descritos na literatura, Sequência laranja representa transcriptase reversa de Gypsy usado como grupo externo. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice E.



Fonte: Do Autor.

5.6 Loa

Elementos desta superfamília foram identificados pela primeira vez em *Drosophila silvestris*, em 1992⁵⁵, e desde então vem sendo identificado em invertebrados marinhos, aracnídeos, e em diversas famílias da ordem insecta¹⁹. Apresentam uma estrutura que pode variar de 5 a 8 kb, contendo duas ORFs, a primeira que contém domínio GAG, para expressão de capsídeo viral e a segunda que contém domínio para expressão de transcriptase reversa, não possui sequências terminais invertidas, assim como outros retroelementos sem LTR, porém, na sua extremidade 3' possui uma trinca de nucleotídeos TAA repetidos em tandem⁵⁵.

3 *contigs* representando TEs da superfamília Loa foram identificados, destes 2 apresentaram domínio para transcriptase reversa (Tabela 6), sendo que um destes elementos agrupou com elementos ainda não caracterizados de duas vespas parasitóides (*N. vitripennis* e *Trichogramma pretiosum*) (Figura 8).

TABELA 6

Características das sequências correspondentes à elementos da superfamília Loa presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL172Contig1	878	EEP	217
CL172Contig5	756	RT_Like, EEP	108
CL172Contig6	2533	RT_Like, RVT_1, zf-RVT, RNase_HI	372

Fonte: Do Autor.

Legenda: EEP: Exonuclease-Endonuclease-Phosphatase; RT_like: Reverse Transcriptase; RVT_1: Reverse Transcriptase; zf-RVT: Zinc-binding in Reverse Transcriptase; Rnase_HI: Ribonuclease H. Sequências com fundo cinza foram utilizadas na filogenia.

5.7 L2

Tem sido proposto que TEs da superfamília L2 divergiram de elementos da superfamília CR1⁵⁶. Apresentam duas ORFs em sua estrutura, as quais apresentam domínio para Endonuclease (ORF1) e Transcriptase Reversa (ORF2). Elementos desta superfamília são encontrados em diversos organismos (moluscos, artrópodes, equinodermatas, peixes, anfíbios, répteis e mamíferos)⁵⁷.

2 *singlets* representando TEs da superfamília L2 foram identificados, ambos apresentaram domínio para transcriptase reversa em sua estrutura (Tabela 7), formando um clado apenas de elementos de *L. boulardi* (Figura 9).

TABELA 7.

Características das sequências correspondentes à elementos da superfamília L2 presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL481Contig13_singlet	2092	RT_like	596
CL481Contig21_singlet	1006	RT_like	250

Fonte: Do Autor.

Legenda: RT_like: Reverse Transcriptase. Sequência com fundo cinza foram utilizadas na filogenia.

5.8 I

Elementos desta superfamília apresentam de 4 a 8 repetições de trincas nucleotídicas TAA em sua extremidade 3', possui tamanho médio de 5.4kb apresentando duas ORFs em sua estrutura, as quais apresentam domínio para Zinc Finger (ORF1) e Transcriptase Reversa (ORF2)⁵⁸. Elementos da superfamília I vem sendo relatados em diversas espécies de fungos e mamíferos, e principalmente em organismos da classe *Insecta*¹⁹.

Apenas 1 *contig* foi identificado representando TEs da superfamília I e apresentando domínio para transcriptase reversa em sua estrutura (Tabela 8), não formando clado específico com nenhum outro alinhamento, porém, foi identificada ancestralidade entre este elemento de *L. boulandi* e elemento I-2 de *N. vitripennis* (Figura 10).

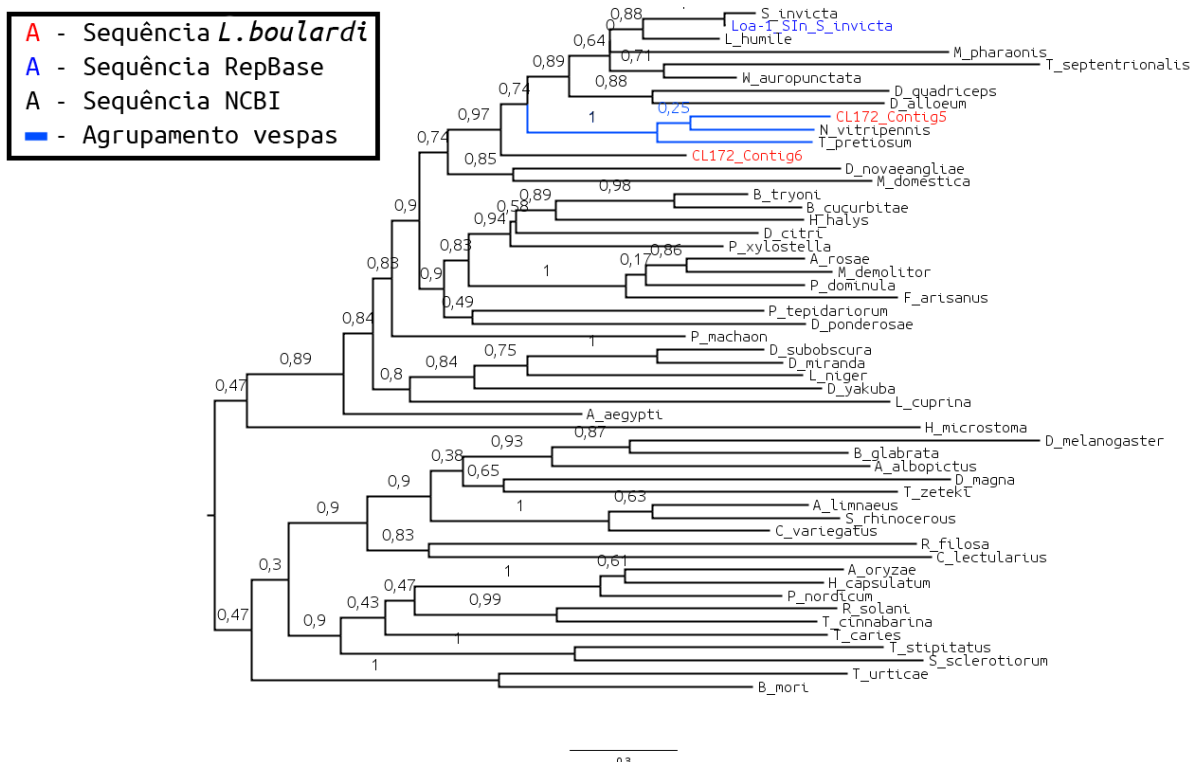
TABELA 8
Características da sequência correspondente ao elemento da superfamília I presente no genoma de *Leptopilina boulandi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL190Contig1	5960	RT_like, RNase_HI, RVT_1, EEP, PTZ	1258

Fonte: Do Autor.

Legenda: EEP: Exonuclease-Endonuclease-Phosphatase; RT_like: Reverse Transcriptase; RVT_1: Reverse Transcriptase; RNase_HI: Ribonuclease H. Sequência com fundo cinza foram utilizadas na filogenia.

Figura 8. Análise filogenética da superfamília Loa. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. boulandi*; Sequências pretas representam transcriptase reversa (anotadas, previstas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase; Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice F.



Fonte: Do Autor.

Figura 9. Análise filogenética da superfamília L2. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. bouvardi*; Sequências pretas representam transcriptase reversa (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice G.

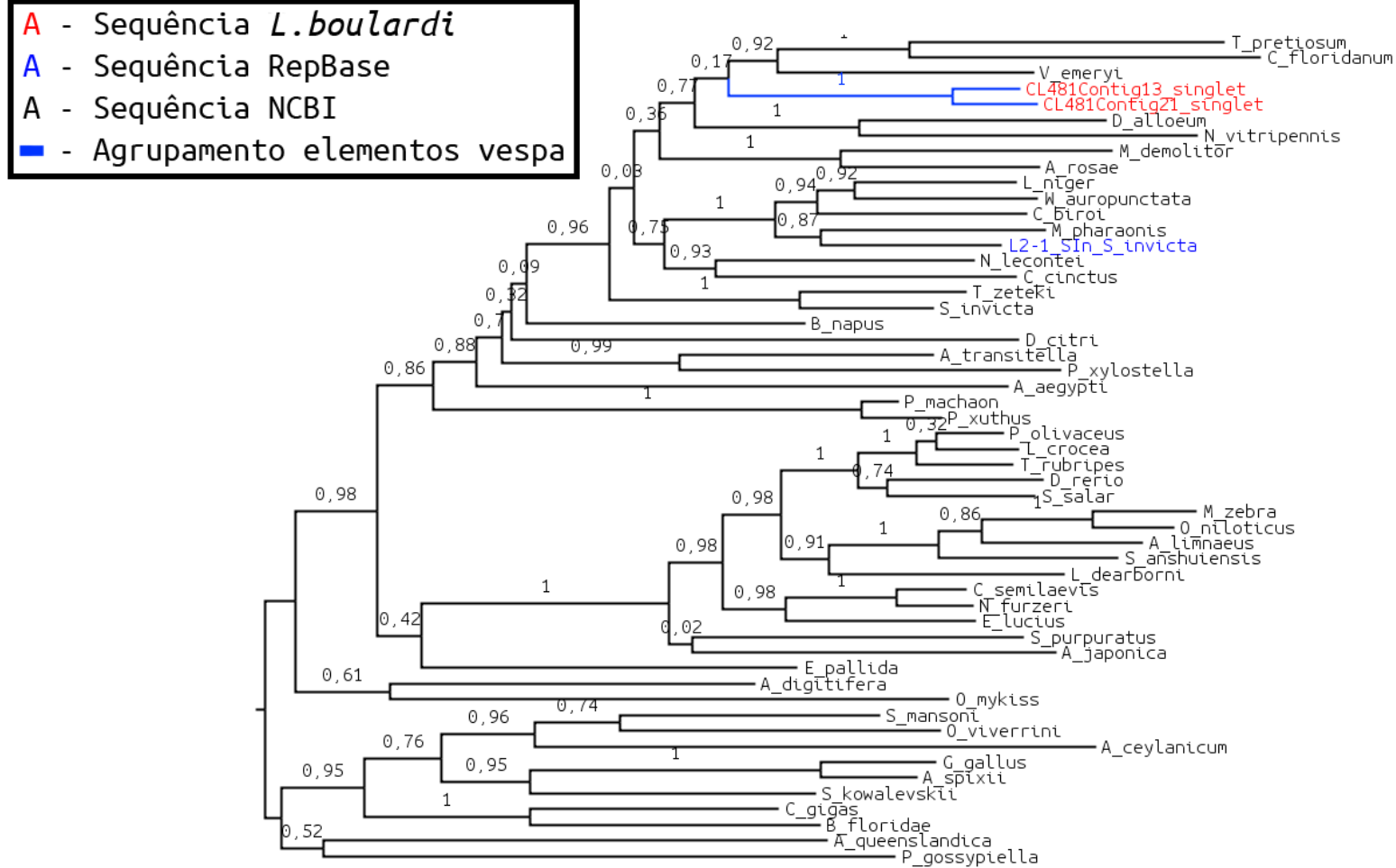
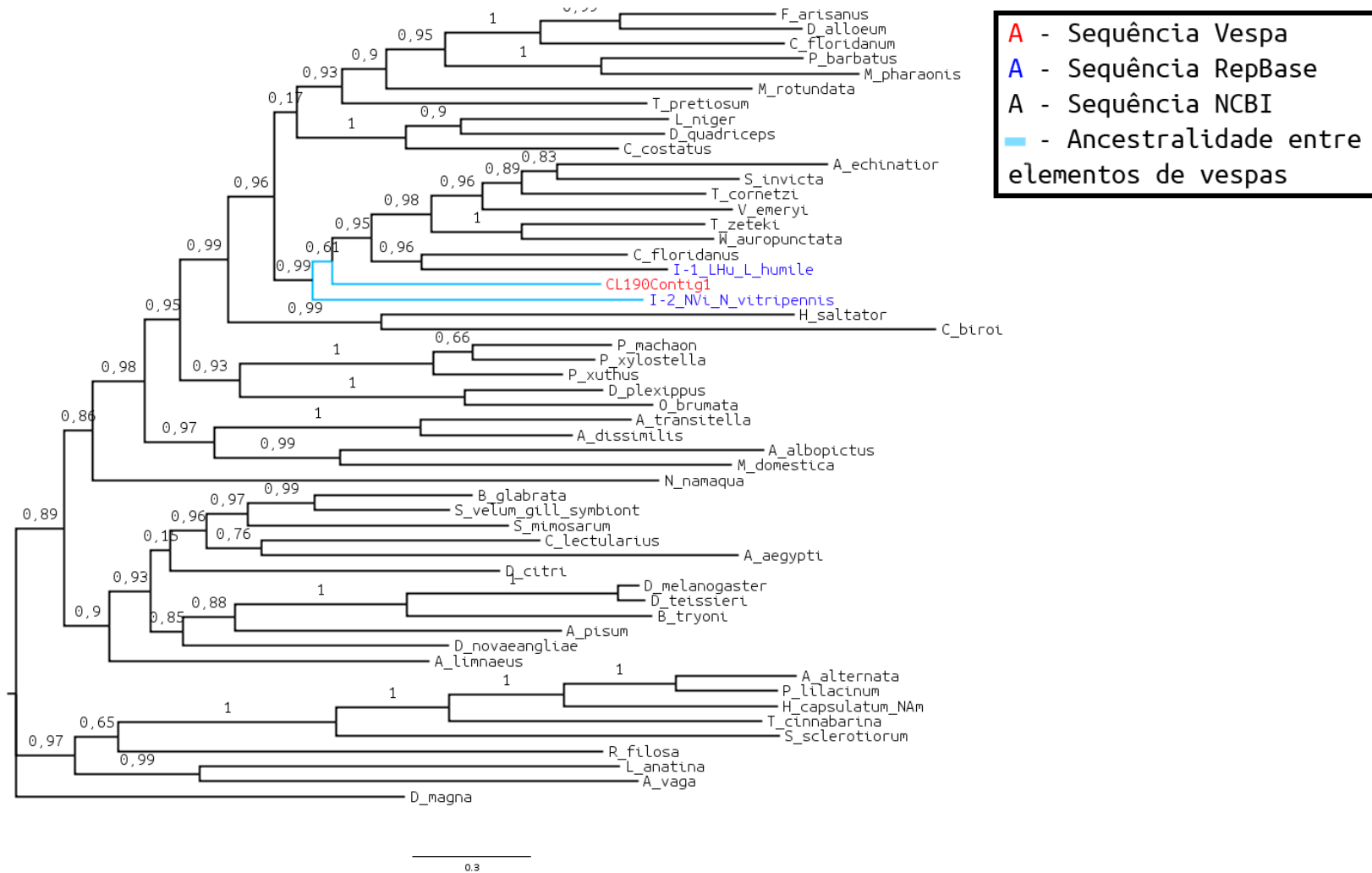


Figura 10. Análise filogenética da superfamília I. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. bouleari*; Sequências pretas representam transcriptase reversa (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice H.



Fonte: Do Autor.

5.9 Gypsy

Elementos desta superfamília apresentam duas ORFs bem definidas em sua estrutura, GAG e POL. A ORF GAG codifica uma proteína para replicação, POL codifica 4 proteínas Protease, Transcriptase Reversa, Integrase e RNase H, nesta ordem⁵⁹. Estes elementos estão presentes em praticamente todas as espécies de plantas superiores, desempenhando papéis importantes na evolução do genoma desses organismos⁶⁰, são abundantes em artrópodes e estão envolvidos em diversos processos de HTT em organismos deste filo⁶.

54 *contigs* representando TEs da superfamília Gypsy foram identificados, destes 41 apresentaram ORF em sua estrutura (Tabela 9). Dentre os 7 clados formados ocorreram agrupamentos com elementos ainda não identificados de vespa parasitoide (*Microplitis demolitor*), diferentes espécies de formigas (*Harpegnathos saltator*, *Lasius niger*), abelha (*Megachile rotundata*) e com elementos identificados de *N. vitripennis* (Gypsy-18-I, Gypsy-5, Gypsy-38, Gypsy-27-I, Gypsy-23-I, Gypsy-19-I e Gypsy-33-I), de *Pogonomyrmex barbatus*, formiga-vermelha, (Gypsy-9) e *Camponotus floridanus*, formiga-touro, (Gypsy-8) (Figura 11).

5.10 BEL

Elementos desta superfamília estão presentes apenas em genomas de metazoários, até o momento não identificados em mamíferos, o que sugere que eles surgiram mais tarde em processos evolutivos de organismos eucariotos ou que foram perdidos, ou ainda não identificados, em outras linhagens de eucariotos⁶¹. Estes elementos geralmente possuem tamanho entre 4.2 e 10kb, apresentando uma estrutura de ORFs similar à elementos da superfamília Gypsy e Copia⁶².

3 *contigs* que representam TEs apresentaram ORFs em sua estrutura (Tabela 10) formando três clados com elementos já descritos de *S. invicta* (Bel-3_SI-I e Bel-7-SI-I) e de *D. melanogaster* (Retroelement Roo) (Figura 12).

TABELA 9

Características das sequências correspondentes à elementos da superfamília Gypsy presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL11Contig57_singlet	4816	RT_LTR, RNase_HI, rve, RVT_1	1238
CL11Contig77_singlet	2069	RT_LTR, RNase_HI, RVT_1	678
CL11Contig13_singlet	1968	rve, CARD	313
CL13Contig1	8939	RT_LTR, RNase_HI, RVT_1, pepsin_retropepsin, DSRM, PTZ	618
CL13Contig10	8758	RT_LTR, RNase_HI, RVT_1, rve, DSRM, zf-H2C2	1236
CL13Contig2	7746	RT_LTR, RNase_HI, RVT_1, pepsin_retropepsin	1321
CL13Contig11	3479	RT_LTR, RNase_HI, RVT_1, rve	619
CL13Contig3	3354	RT_LTR, RNase_HI, RVT_1	1005
CL13Contig4	1930	RT_LTR, RNase_HI, RVT_1	497
CL13Contig112_singlet	1029	RT_like, RNase_HI, rve, DSRM	499
CL23Contig1	10970	RT_LTR, RNase_HI, RVT-1, rve, retropepsin_like, PTZ	1341
CL38Contig3	4101	RT_LTR, RNase_HI, rve, RVT_1	731
CL38Contig1	3625	RNase_H, rve, Tra5, DUF	306
CL38Contig5	3411	RT_LTR, RNase_HI, RVT_1	1093
CL38Contig166_singlet	1180	RT_LTR, RNase_H, RVT_1	380
CL51Contig2	5342	RNase_HI, rve, DUF	813
CL51Contig1	4079	rve, Tra5	372
CL51Contig3	2842	rve, Tra5, DUF	387
CL51Contig39_singlet	1255	rve, Tra5, ZF	387
CL51Contig5	1062	rve, ZF	308
CL56Contig2	5898	RT_LTR, RNase_HI, retropepsin_like, RVT_1, PHA	440
CL56Contig1	4706	RT_LTR, RNase_HI, retropepsin_like, RVT_1	1222
CL61Contig1	4969	RT_LTR, RNase_HI, retropepsin_line, RVT_1, Retrotrans_gag	841
CL61Contig7	1807	RNase_HI, rve	552
CL66Contig17_singlet	6663	RT_LTR, RNase_HI, pepsin_retropepsin, rve, RVT_1, PTZ	1526

Continuação Tabela 9.

CL66Contig1	4570	RT_LTR, RNase_HI, rve, retropepsin, RVT_1	780
CL66Contig6	2917	RNase_HI, rve	593
CL66Contig8	1063	RT_LTR, RVT_1	276
CL117Contig2	3366	RT_like, RNase_HI, rve	932
CL117Contig1	3005	RT_like, RNase_HI, rve	663
CL122Contig21_singlet	4551	RT_LTR, RNase_HI, RVT_1, pepsin_retropepsin	894
CL122Contig5	3242	rve, DSRM	606
CL122Contig7	2051	RT_like, pepsin_retropepsin	349
CL122Contig3	1359	RT_LTR, RNase_HI, rve	1103
CL181Contig1	6635	RT_like, RNase_H, rve, Tra5	437
CL196Contig3	5551	RT_LTR, RNase_HI, RVT-1, rve, retropepsin_like, Retrotrans_gag	942
CL260Contig3	5188	RT_LTR, RNase_HI, RVT-1, retropepsin_like	1545
CL260Contig1	3038	RT_LTR, RNase_HI, RVT-1, retropepsin_like	747
CL298Contig1	5819	RT_LTR, RNase_HI, RVT_1, rve, pepsin_retropepsin, DUF	1172
CL298Contig7_singlet	1233	RNase_HI, rve, RT_LTR	354
CL331Contig1	4826	RT_LTR, RNase_HI, RVT_1, retropepsin_like, DSRM	636
CL333Contig13_singlet	2500	RT_LTR, RNase_HI, retropepsin_like, RVT_1	534
CL333Contig5_singlet	2474	RT_LTR, retropepsin_like, RVT_1	480
CL333Contig7_singlet	2371	RT_LTR, RNase_HI, RVT_1, CACIF	650
CL367Contig2	3179	RT_LTR, RNase_HI, rve, Tra5	388
CL367Contig34_singlet	2135	RT_LTR, retropepsin_like, pepsin_retropepsin, RVT_1	693
CL369Contig3	1600	rve, ZF	430
CL39Contig46_inglet	7862	RT_LTR, RNase_HI, rve, RVT_1, Gypsy_protein	1250
CL397Contig1	4404	RT_LTR, RNase_HI, rve, RVT_1, ZF	695
CL447Contig2	4167	RT_LTR, RNase-HI, rve, retropepsin, RVT_1	1085
CL451Contig1	3957	rve, Tra5, DUF	423
CL456Contig3	1194	rve, ZF	352

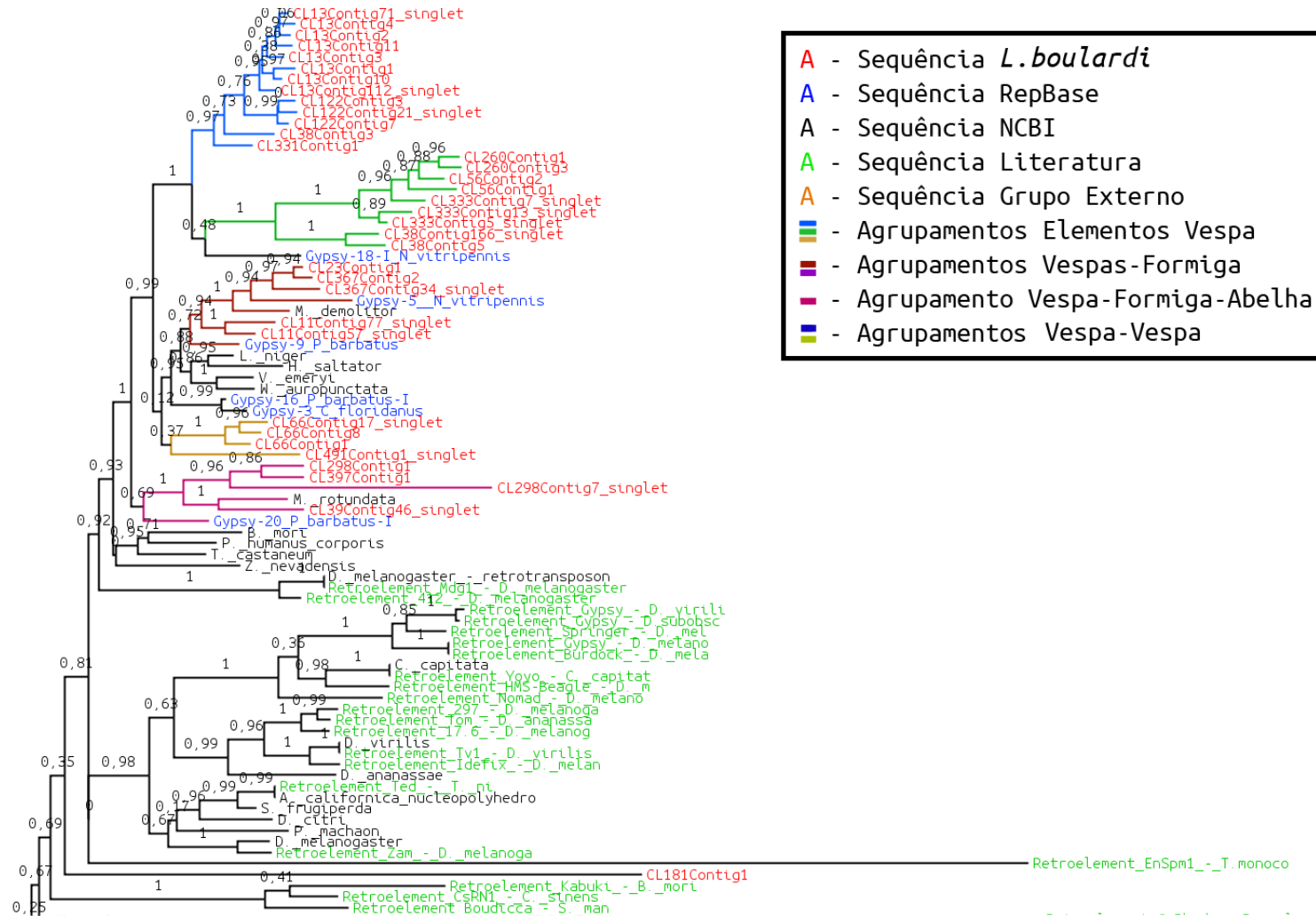
Continuação Tabela 9.

CL461Contig1	3915	rve, PotE	543
CL491Contig1_singlet	4651	RT_LTR, RNase_HI, RVT_1, pepsin_retropepsin, PTZ	966

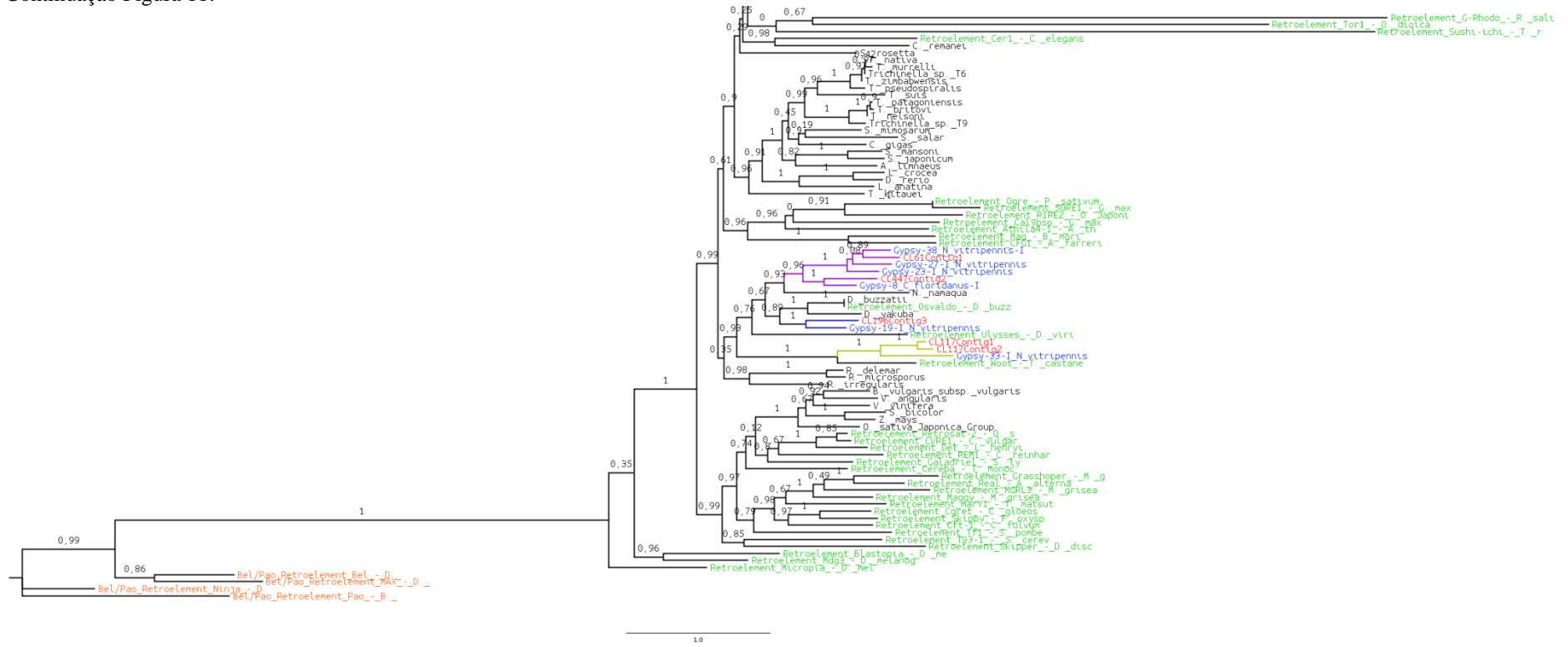
Fonte: Do Autor.

Legenda: RT_like: Reverse Transcriptase; RVT_1: Reverse Transcriptase; zf-RVT: Zinc-binding in Reverse Transcriptase; RNase_HI: Ribonuclease H; rve: Integrase; CARD: Caspase Recruitment domain; pepsin_retropepsin: Cellular and retroviral pepsin-like aspartate proteases; DSRM: Double-stranded RNA binding motif; PTZ: Actin provisional; zf-H2C2: Zinc-binding couple in Histone; Tra5: Transposase; DUF: Unknown Function; Retrotrans_gag: Gag or capsid-like protein from LTR transposon; CACIF: DNA-binding transcriptional activator CaiF; ZF: Zinc Finger; PotE: Amino acid transporter. Sequência com fundo cinza foram utilizadas na filogenia.

Figura 11. Análise filogenética da superfamília Gypsy. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. bouleari*; Sequências pretas representam transcriptase reversa (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase; Sequências verdes representam transcriptase reversa de elementos descritos na literatura, Sequência laranja representa transcriptase reversa de Bel usado como grupo externo. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice I.



Continuação Figura 11.



Fonte: Do Autor.

TABELA 10.
Características das sequências correspondentes à elementos da superfamília Bel presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF(AA)
CL449Contig1	6045	RT_like, Peptidase_A17, rve, pepsin_retropepsin, Viral_NABP	1185
CL498Contig2	4882	RT_like, Peptidase_A17, pepsin_retropsin, DUF	340
CL343Contig1	4174	RT_like, Peptidase_A17	202

Fonte: Do Autor.

Legenda: RT_like: Reverse Transcriptase; Peptidase_A17: Pao retrotransposon peptidase; rve: Integrase; pepsin_retropepsin: Cellular and retroviral pepsin-like aspartate proteases; Viral_NABP: Viral nucleic acid binding; DUF: Unknown Function. Sequência com fundo cinza foram utilizadas na filogenia.

5.11 Cópia

TEs desta superfamília foram identificados primeiramente em genoma de *D. melanogaster*, a estrutura destes elementos possui um tamanho aproximado de 5 kb⁶³, tendo em sua estrutura duas ORFs, GAG e POL, da mesma forma que em outros LTR, Gypsy e Bel¹⁶. Além de sua ampla diversidade em espécies do gênero *Drosophila* estes elementos estão presentes na maioria das divisões de plantas superiores⁶⁴. Esta superfamília de TEs vem sendo considerada modelo em estudos de HTT, devido a fortes evidências de transferências de transposons entre espécies de plantas⁶⁵ e entre espécies de animais⁶⁶.

Apenas 1 *contig* representou TEs da superfamília Cópia apresentando Transcriptase Reversa em sua estrutura (Tabela 11), este elemento formou um clado com um elemento ainda não identificado de *N. vitripennis* e um já identificado (Cópia-23-Nvi-I) (Figura 13).

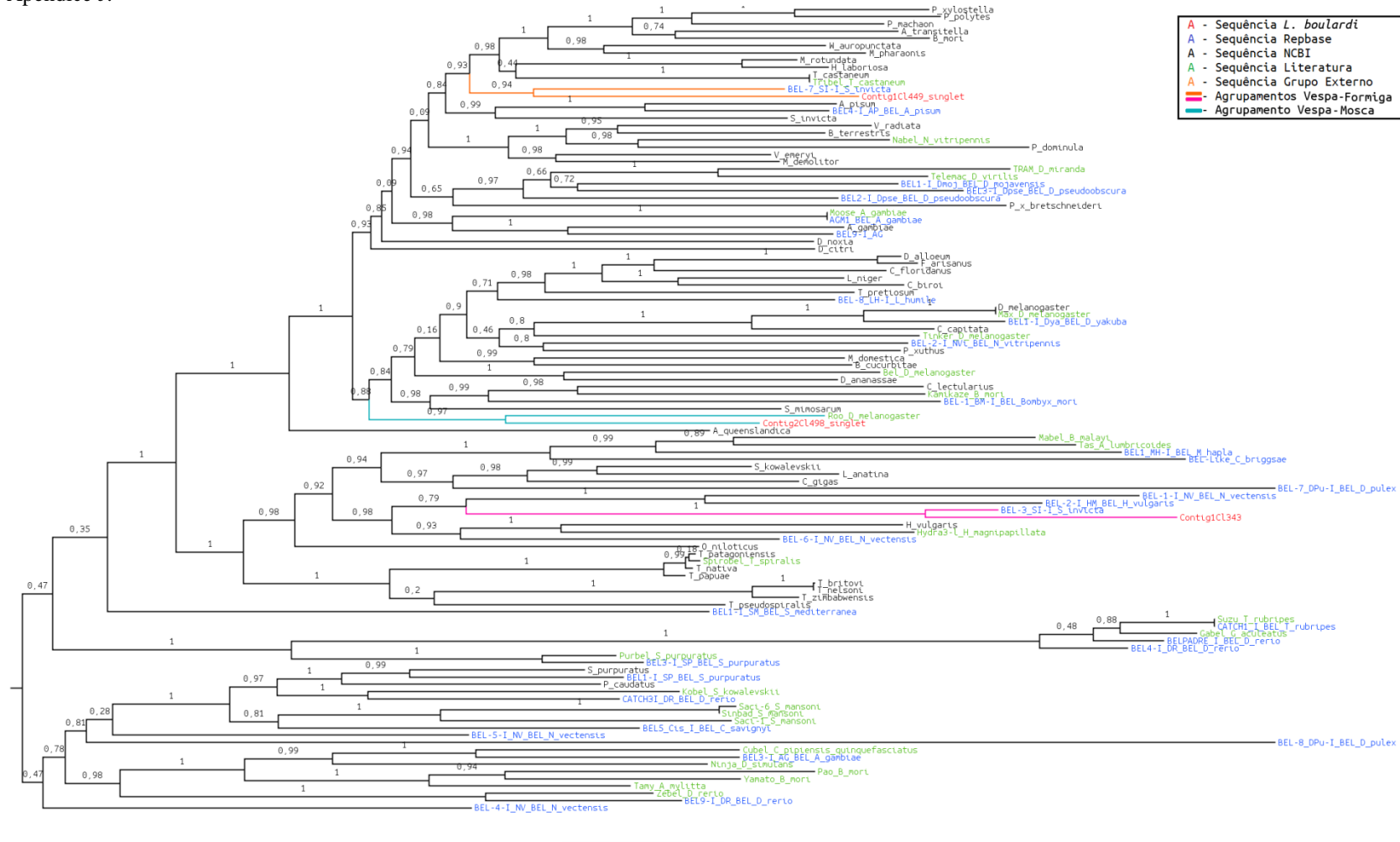
TABELA 11
Características das sequências correspondentes à elementos da superfamília Cópia presentes no genoma de *Leptopilina boulardi*.

Contig	Tamanho do Contig (pb)	Domínios Conservados encontrados no CDD	Tamanho da ORF (AA)
CL1Contig38	4815	RVT_2, RNase_HI, gag_pre-integrase, Cas9, PRK, BglG, glyco_rpt, UPF, ABC2, 7TM	576

Fonte: Do Autor.

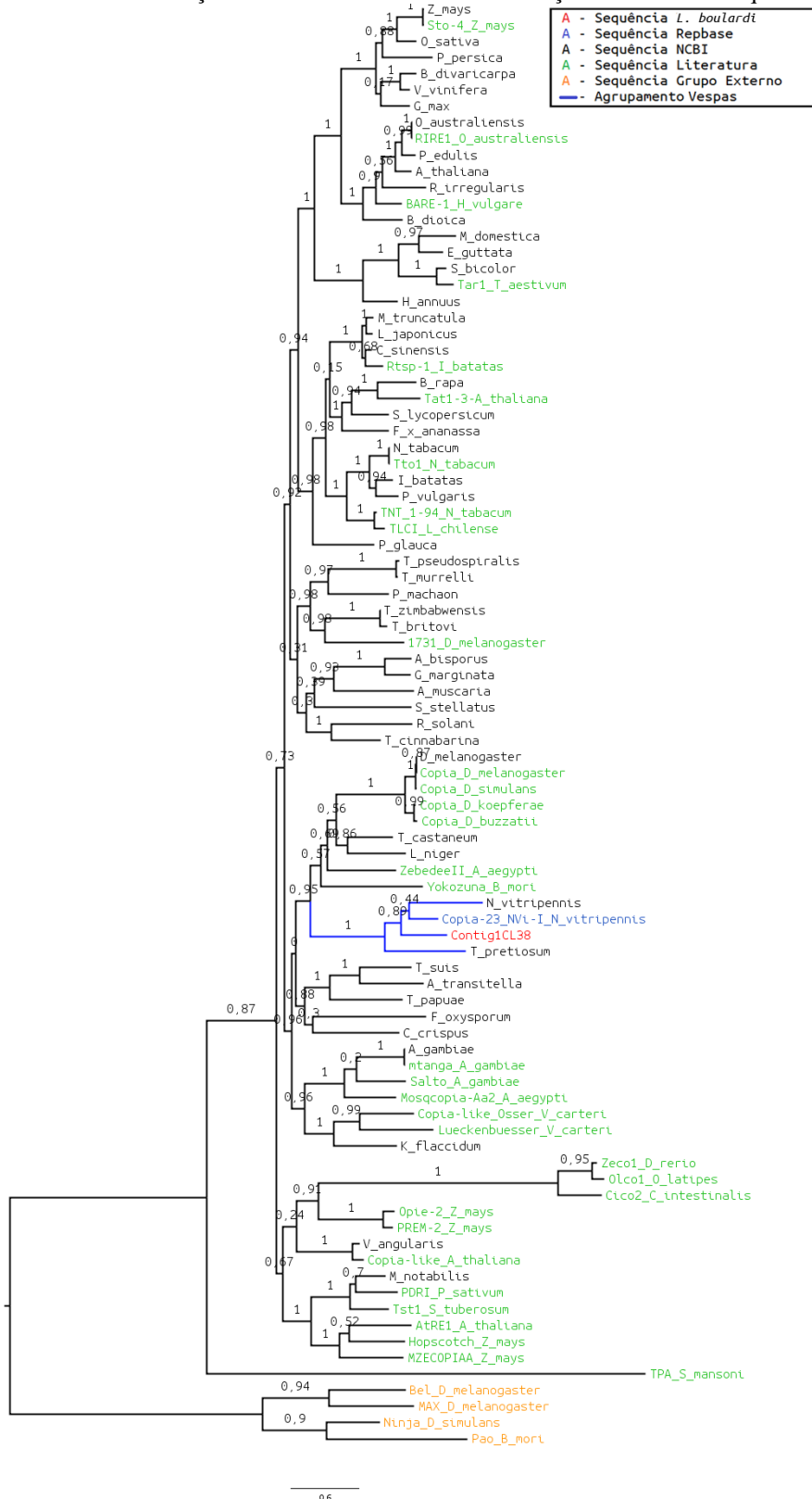
Legenda: RVT_2: Reverse Transcriptase; RNase_HI: Ribonuclease H; Cas9: CRISPR/CAS System; PRK: Phosphoribulokinase; BglG: Transcriptional antiterminator; glyco_rpt: oligosaccharide repeat unit polymerase; UPF: Uncharacterized Protein Family; ABC2: Membrane ABC Transporter; 7TM: Transmembrane Receptor. Sequência com fundo cinza foram utilizadas na filogenia.

Figura 12. Análise filogenética da superfamília BEL. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. bouhardi*; Sequências pretas representam transcriptase reversa (anotadas, preditas ou hipotéticas) provenientes do banco de dados NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase; Sequências verdes representam transcriptase reversa de elementos descritos na literatura. Valores representam suporte de ramo aLRT. Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice J.



Fonte: Do Autor.

Figura 13. Análise filogenética da superfamília Copia. Sequências vermelhas representam transcriptase reversa de elementos presentes no genoma de *L. bouhardi*; Sequências pretas representam transcriptase reversa (anotadas, preditas ou hipotéticas) provenientes do NCBI; Sequências azuis representam transcriptase reversa de elementos retirados do banco de dados RepBase; Sequências verdes representam transcriptase reversa de elementos descritos na literatura. Sequências laranjas representam transcriptase reversa de Bel usado como grupo externo. Valores representam suporte de ramo aLRT Extensão de ramos é proporcional à distância aminoacídica. Modelo de substituição de aminoácidos LG+G. Informações referentes à sequências conferir Apêndice K.



Fonte: Do Autor.

6 Discussão

O método de identificação e caracterização de sequências mostrou-se eficiente, quando comparado aos resultados de Ortiz *et al*⁶⁷, o qual identificou 8 superfamílias já descritas na literatura (Mariner, hAT, Helitron, EnSPM, Transib, Copia, Gypsy, Jockey e CR1). Uma vez que foi possível identificar um maior número de superfamílias de TEs, a divergência na diversidade de superfamílias entre os dois estudos pode ser explicada pelo método de análise da ferramenta RepeatExplorer, que considera apenas 0,1% dos *clusters* formados (descritos como *Top Clusters*), dessa forma, não identificando famílias com número muito baixo de cópias no genoma.

Dentre os TEs autônomos que apresentaram sequência proteica em sua estrutura e foram submetidos as análises filogenéticas, 26 ETs agruparam com ETs já descritos na literatura, 52 agruparam com ETs ainda não descritos.

Pode-se evidenciar a existência de diversos agrupamentos de TEs de vespas parasitas e formigas nas análises filogenéticas das superfamílias Chapaev, Gypsy e Copia. Estudos de relações ecológicas relatam que existem relações de predação entre formigas que habitam figueiras e vespas polinizadoras e parasitas⁶⁸, vespas que parasitam formigas⁶⁹, também são relatadas relações de mutualismo onde vespas parasitóides mimetizam a morfologia de formigas⁷⁰ e de vespas que mimetizam odores de formigas para evitar reconhecimento e realizar processos de parasitismo⁶⁸. De acordo com a ferramenta TimeTree⁷¹ vespas e formigas divergiram evolutivamente há mais de 150 milhões de anos.

Porém, não existe nenhum relato na literatura entre *L. bouvardi* e os organismos da família *Formicidae*, dessa forma, são necessários mais estudos para esclarecer estes agrupamentos de elementos transponíveis entre vespas e formigas.

Possivelmente ocorreu um evento de HTT envolvendo TE da superfamília BEL e elemento Roo (Superfamília BEL) de *Drosophila melanogaster*, devido ao agrupamento observado na filogenia desta superfamília, relatos na literatura descrevem relações de parasitismo entre *L. bouvardi* e *D. melanogaster*⁷².

Agrupamentos de TEs de diferentes espécies de vespas parasitas podem ser explicados por processos de co-evolução destes elementos e de seus genomas hospedeiro.

7 Considerações Finais

O *pipeline* desenvolvido neste trabalho mostrou-se eficiente tanto na caracterização dos TEs que conformam o genoma de *L. bouvardi* quanto na reconstrução da história evolutiva dos TEs identificados, representando assim uma nova abordagem que oferece economia de tempo e dinheiro quando aplicada em pesquisas de caracterização e análise evolutiva de TEs em organismos não modelo.

Embora análises filogenéticas sejam um método muito preciso para identificar relações de parentesco entre diferentes organismos, mais estudos, como por exemplo filogenia cruzada, distribuição desigual de TEs ou similaridade de sequências, devem ser realizados para testar a hipótese da existência de processos de HTT envolvendo vespas parasitas e formigas.

REFERÊNCIAS

- 1 MCCLINTOCK, B. **Controlling elements and the gene.** Cold Spring Harb Symp Quant Biol 21: 197-216. 1956;
- 2 MCCLINTOCK, B. **The association of mutants with homozygous deficiencies in *Zea mays*.** Genetics 26: 542-57. 1941;
- 3 MCCLINTOCK, B. **The stability of broken ends of chromosomes in *Zea mays*.** Genetics 26: 234-282. 1941;
- 4 MCCLINTOCK, B. **The fusion of broken ends of chromosomes following nuclear fusion.** Proc Natl Acad Sci USA 28: 458-463. 1942;
- 5 MCCLINTOCK, B. **Mutable loci in maize.** Year B Carnegie Inst Wash 47:155-169. 1948;
- 6 DOTTO, B. R.; CARVALHO, E. L.; SILVA, A. F.; et al, **HTT-DB: Horizontally transferred transposable elements database.** Bioinformatics 17: 2915-291. 2015;
- 7 ALZOHAIRY, A. M.; GYULAI, G.; JANSEN, R. K.; et al. **Transposable elements domesticated and neofunctionalized by eukaryotic genomes.** Plasmid 69: 1-15. 2012;
- 8 CHIMPANZEE SEQUENCING AND ANALYSIS CONSORTIUM. **Initial sequence of the chimpanzee genome and comparison with the human genome.** Nature 437: 69-87. 2005;
- 9 LANDER, E. S.; LINTON, L. M.; BIRREN, B.; et al. **Initial sequencing and analysis of the human genome.** Nature 409: 860-921. 2001;
- 10 RHESUS MACAQUE GENOME SEQUENCING AND ANALYSIS CONSORTIUM. **Evolutionary and biomedical insights from the rhesus macaque genome.** Science 316: 222-234. 2007;
- 11 SCHNABLE, P. S.; WARE, D.; FULTON, R. S.; et al. **The B73 maize genome: complexity, diversity, and dynamics.** Science 326: 1112-1115. 2009;
- 12 KAPITONOV, V. V.; JURKA, J.; **RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons.** PLoS Biology 3: 998-1011. 2005;
- 13 LANGDON, T.; SEAGO, C.; MENDE, M.; et al. **Retrotransposon evolution in diverse plant genomes.** Genetics 156: 313-325. 2000;
- 14 MENG, Q; CHEN, K.; MA, L.; et al; **A systematic identification of Kolobok superfamily transposons in *Trichomonas vaginalis* and sequence analysis on related transposases.** Journal of Genetics and Genomics 38: 63-70. 2011;

- 15 FINNEGAN, D. J. **Eucariotic transposable elements and genome evolution.** Trends Genet 5: 103-107. 1989.
- 16 WICKER, T.; SABOT, F.; HUAN-VAN, A.; et al. **A unified classification system for eukaryotic transposable elements.** Nature Reviews 8: 973-982. 2007;
- 17 KAPTONOV, V. V.; JURKA, J. **A universal classification of eukaryotic transposable elements implemented in Repbase.** 2008. Nature Reviews 9: 411-412. 2008;
- 18 NOVAK, P.; NEUMANN, P.; PECH, J.; et al. **RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next generation sequence reads.** Bioinformatics 29: 792-793. 2013;
- 19 JURKA, J.; KAPITONOV, V. V.; PAVLICEK, A.; et al. **Repbase Update, a database of eukaryotic repetitive elements.** Cytogenet. Genome Res 110: 462–467. 2005;
- 20 MIZUUCHI, K. **Transpositional recombination: mechanistic in-sights from studies of Mu and other elements.** Annu Rev Biochem 61: 1011-1051. 1992;
- 21 CHANDLER, M.; MAHILLON, J. **Insertion Sequences Revisited.** Mobilome DNA II 15: 305-306. 2002;
- 22 LUAN, D. D.; KORMAN, M. H.; JAKUBCZAK, J. L.; et al. **Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition.** Cell 72: 595-605. 1993;
- 23 BIEDLER, J.; TU, Z. **Non-LTR retrotransposons in the african malaria mosquito, *Anopheles gambiae*: unprecedented diversity and evidence of recent activity.** Mol. Biol. Evol. 20: 1811–1825. 2003;
- 24 GREGORY, T. R.; NICOL, J. A; TAMM, H; et al. **Eukaryotic genome size databases.** Nucleic Acids Res 35: 332-D338. 2007;
- 25 HANRAHAN, S. J.; JOHNSTON J. S. **New genome size estimates of 134 species of arthropods.** Chromosome Res 19: 809-823. 2011;
- 26 THOMAS, C. A. **The genetic organization of chromosomes.** Annu Rev Genet 5: 237-256. 1971.
- 27 MAUMUS, F.; FISTON-LAVIER, A. S.; QUESNEVILLE, H. **Impact of transposable elements on insect genomes and biology.** Current Opinion in Insect Science 7: 30-39. 2015;
- 28 KELLEY J. L.; PEYTON J. T.; FISTON-LAVIER A. S.; et al. **Compact genome of the Antarctic midge is likely an adaptation to an extreme environment.** Nat Commun 5: 1-8. 2014;
- 29 WANG, X.; FANG, X.; YANG, P.; et al. **The locust genome provides insight into swarm formation and long-distance flight.** Nat Commun 5: 1-9 2014;

- 30 CLARK, A. G.; EISEN, M. B.; SMITH, D. R.; et al. **Evolution of genes and genomes on the *Drosophila* phylogeny.** Nature 450: 203-218. 2007;
- 31 CARARETO, C. M. A.; MONTEIRO-VITORELLO, C. B.; SLUYS, M. V.; **Elementos de Transposição: Diversidade, evolução aplicações e impacto no genoma dos seres vivos.** São Paulo: SBG, 2015. 196p.
- 32 THE NASONIA GENOME WORKING GROUP. **Functional and Evolutionary Insights from the Genomes of Three Parasitoid Nasonia Species.** Science 15: 343–348. 2010;
- 33 RIZKI, R. M.; RIZKI, T. M. **Parasitoid virus-like particles destroy *Drosophila* cellular immunity.** Proc Natl Acad Sci USA 87: 8388–8392. 1990;
- 34 ROUTH, A.; DOMITROVIC, T.; JOHNSON, J. E. **Host RNAs, including transposons, are encapsidated by a eukaryotic single-stranded RNA virus.** Proc Natl Acad Sci USA 109: 1907–1912. 2012;
- 35 HERNIOU E. A.; HUGUET, E.; THÉZÉ J.; et al. **When parasitic wasps hijacked viruses: genomic and functional evolution of polydnviruses.** Philos Trans R Soc B 12: 368-381. 2013;
- 36 LABROSSE, C.; CARTON, Y.; DREZEN, J. M. *et al.* **Active supression of *D. melanogaster* immune response by long gland products of the parasite wasp *Leptopilina bouleardii*.** Journal of Insect Physiology 49: 513-22. 2003;
- 37 HUANG, X.; MADAN A. **CAP3: A DNA sequence assembly program.** Genome Res 9: 868-877. 1999;
- 38 MARCHLER, B. A.; BRYANT S.H. **CD-Search: protein domain annotations on the fly.** Nucleic Acids Res 32: 327-331. 2004;
- 39 ROMBE, I. T.; SYKES, K. F.; RAYNER S.; et al. **ORF-FINDER: a vector for high-throughput gene identification.** Gene 9: 33-41. 2002;
- 40 KATOH, K.; STANDLEY, D. M.; **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** Mol Biol Evol 4: 772-780. 2013;
- 41 DARRIBA, D.; TABOADA, G. L.; DOALLO, R. *et al*; **ProtTest 3: fast selection of best-fit models of protein evolution.** Bioinformatics 8: 1164-1165. 2011;
- 42 GUINDON, S.; GASCUEL, O.; **PhyML: A simple, fast, and accurate algorithm to estiamate large phylogenies by maximum likelihood.** Systematic Biology 5: 696-704. 2003;
- 43 KAPITONOV, V. V.; JURKA, J. **Self-synthesizing DNA transposons in eukaryotes.** Proc. Natl. Acad. Sci. USA 103: 4540–4545. 2006.
- 44 KAPITONOV, V. V.; JURKA, J. **Rolling-circle transposons in eukaryotes.** Proc. Natl. Acad. Sci. U. S. A. 98: 8714–8719. 2001;

- 45 KAPITONOV, V. V.; JURKA, J. **Helitrons on a roll: eukaryotic rolling-circle transposons**; Review *TRENDS in Genetics* 23: 521-529. 2007;
- 46 KAPITONOV, V. V.; JURKA, J. **Chapaev—a novel superfamily of DNA transposons**. *Rebase Rep.* 7: 774–781. 2007
- 47 YUAN, Y. W.; WESSLER, S. R. **The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies**. *Proc Natl Acad Sci USA.* 108: 7884–7889. 2011;
- 48 ZHANG, H. H.; FESCHOTTE, C.; HAN, M. J.; et al. **Recurrent Horizontal Transfer of *Chapaev* Transposons in Diverse Invertebrate and Vertebrate Animals**. *Genome Biol. Evol.* 6: 1375–1386. 2014;
- 49 YE, J.; PÉREZ-GONZÁLEZ C. E.; EICKBUSH D. G.; et al **Competition between R1 and R2 transposable elements in the 28S rRNA genes of insects**. *Cytogenet Genome Res* 110: 299–306. 2005;
- 50 PÉREZ-GONZÁLEZ C. E.; EICKBUSH, T. H. **Dynamics of R1 and R2 Elements in the rDNA Locus of *Drosophila simulans***. *Genetics* 158: 1557-1567. 2001;
- 51 EVGEN'EV, M.; ZELEN'TSOVA, H.; MNJOIAN, L.; et al. **Invasion of *Drosophila virilis* by the Penelope transposable element**. *Chromosoma* 109: 350-357. 2000;
- 52 MORALES-HOJAS, R.; VIEIRA, C. P; VIEIRA J. **The evolutionary history of the transposable element Penelope in the *Drosophila virilis* group of species**. *J Mol Evol* 63: 262-273. 2006
- 53 LIN, X.; FARID, N.; CASOLA, C. **An Ancient Trans-Kingdom Horizontal Transfer of *Penelopelike* Retroelements from Arthropods to Conifers**. *Genome Biology and Evolution Advance Access* 8: 1252-1266. 2016;
- 54 ARKHIPOVA, I. R.; PYATKOV, K. I.; MESELSON, M.; et al. **Retroelements containing introns in diverse invertebrate taxa**. *Nat Genet* 33:123-124. 2003;
- 55 FELGER, I.; HUNT, J. A.; **A non-LTR retrotransposon from the Hawaiian *Drosophila*: the Loa element**. *Genetica* 85: 119-130. 1992;
- 56 MALIK, H. S.; BURKE W. D.; EICKBUSH T. H.; **The age and evolution of non-LTR retrotransposable elements**. *Mol Biol Evol.* 6: 793-805. 1999;
- 57 LOVSIN, N.; GUBENSEK, F.; KORDI, D.; **Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia**. *Mol Biol Evol.* 12: 2213-2224. 2001;
- 58 CHABOISSIER, M. C.; FINNEGAN, D.; BUCHETON, A.; **Retrotransposition of the I factor, a non-long terminal repeat retrotransposon of *Drosophila*, generates tandem repeats at the 3' end**. *Oxford University Press* 13: 2467-2472. 2000;
- 59 LEE, S.; KIM, N. S. **Transposable Elements and Genome Size Variations in Plants**. *Genomics Inform* 12: 87-97. 2014;

- 60 JEUNG, J. U.; CHO, S. K.; LEE, S. J.; et al. **Characterization of Ty3-gypsy-like Elements in *Hibiscus syriacus***. *Mol Cells* 19: 318-327. 2005;
- 61 CHAUX, N.; WAGNER, A. **BEL/Pao retrotransposons in metazoan genomes**. *BMC Evol Biol* 11: 154-160. 2011;
- 62 LORENS, C.; FUTAMI, R.; COVELLI, L.; et al. **The Gypsy Database (GyDB) of mobile genetic elements: release 2.0**. *Nucleic Acids Res* 39: 70-74. 2011;
- 63 FLAVELL, A. J.; RUBY, S. W.; TOOLE, J. J.; et al. **Translation and developmental regulation of RNA encoded by the eukaryotic transposable element copia**. *Proc Natl Acad Sci USA* 12: 7107-7111. 1980;
- 64 FLAVELL, A. J.; DUNBAR, E.; RANDERSON, S. R.; et al. **Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants**. *Nucleic Acids Res.* 20: 3639–3644. 1991;
- 65 VOYTAS, D. F.; UMMINGS, M. P. C.; KONIECZNY, A.; et al, **Copia-like retrotransposons are ubiquitous among plants**. *Proc. Natl. Acad. Sci. USA* 89: 7124–7128. 1992;
- 66 GONZALEZ, P.; LESSIOS, H. A. **Evolution of sea urchin retroviral-like (SURL) elements: evidence from 40 Echinoid species**. *Mol. Biol. Evol.* 16: 938–952. 1999;
- 67 ORTIZ, M. F.; WALLAU, G. L.; GRAICHEN, D. A. S. *et al.*; **An evaluation of the ecological relationship between *Drosophila* species and their parasitoid wasps as an opportunity for horizontal transposon transfer**. *Mol Genet Genomics* 290: 67-78. 2015;
- 68 WANG, B.; GENG, X. Z.; MA, L.B.; et al. **A trophic cascade induced by predator y ants in a fig-fig wasp mutualism**. *J Anim Ecol.* 83: 1149-1157. 2014;
- 69 PÉREZ-LACHAUD, G.; BARTOLO-REYES, J. C.; QUIROA-MONTALVÁN, C. M.; et al. **How to escape from the host nest: imperfect chemical mimicry in eucharitid parasitoids and exploitation of the ants' hygienic behavior**. *J Insect Physiol* 75: 63-72. 2015;
- 70 MALCICKA, M.; BEZEMER, M.; VISSER, B.; et al. **Multi-trait mimicry of ants by a parasitoid wasp**. *Scientific Reports* 5: 8043-8039. 2015;
- 71 HEDGES S. B.; DUDLEY, J.; KUMAR, S. **TimeTree: a public knowledge-base of divergence times among organisms**. *Bioinformatics* 22: 2971-2972. 2006;
- 72 HITTA, M.; ESPAGNE, E.; LEMEUNIER, F.; **Mapping candidate genes for *Drosophila melanogaster* resistance to the parasitoid wasp *Leptopilina boulardi***. *Genet Res.* 2: 81-91. 2006.

APÊNDICE

APÊNDICE A - Informações referentes às sequências utilizadas na filogenia da superfamília Polinton.

REPBASE

Elemento	Organismo	Superfamília
Polinton-6_Nvi	<i>Nasonia vitripennis</i>	Polinton (Maverick)
Polinton-8_Nvi	<i>Nasonia vitripennis</i>	Polinton (Maverick)
Polinton-7_Nvi	<i>Nasonia vitripennis</i>	Polinton (Maverick)
Polinton-1_Nvi	<i>Nasonia vitripennis</i>	Polinton (Maverick)
Polinton-3_Nvi	<i>Nasonia vitripennis</i>	Polinton (Maverick)
Polinton-2_Nvi	<i>Nasonia vitripennis</i>	Polinton (Maverick)
Polinton-5_Nvi	<i>Nasonia vitripennis</i>	Polinton (Maverick)
Polinton-1RI	<i>Rhizophagus intraradices</i>	Polinton (Maverick)
Copia1-I_Dm	<i>Drosophila melanogaster</i>	Copia

LITERATURA

Elemento	Organismo	Código Acesso NCBI
Mav-Da	<i>Drosophila ananassae</i>	AAPP01019595.1
Mav-Dg	<i>Drosophila grimshawi</i>	AAPT01000720.1
Mav-Dp	<i>Drosophila permisilis</i>	AAIZ01001788.1
Mav-Dv	<i>Drosophila virilis</i>	AANI01015985.1
Mav-Dm	<i>Drosophila melanogaster</i>	AC016917.2
Mav-He	<i>Hydractinia echinata</i>	CO540554
Mav-Pi	<i>Phytophthora infestans</i>	AC147544.2
Mav-Tn	<i>Tetraodon nigroviridis</i>	BX908814.1
Mav-Nv	<i>Nematostella vectensis</i>	AC151588

BLASTP *Leptopilina bouleari*

Organismo	CL2Contig5 Código Acesso NCBI	Classificação NCBI
<i>Glyptapanteles flavicoxis</i>	ACE75265.1	Integrase
<i>Diachasma alloenum</i>	XP_015123511.1	Proteína Não Caracterizada
<i>Microplitis demolitor</i>	XP_014299741.1	Proteína Não Caracterizada
<i>Lasius niger</i>	KMQ83563.1	Integrase
<i>Vollenhovia emeryi</i>	XP_011869582.1	Proteína Não Caracterizada
<i>Tribolium castaneum</i>	EFA12198.1	Proteína Não Caracterizada
<i>Fopius arisanus</i>	XP_011312904.1	Integrase Predita
<i>Neodiprion lecontei</i>	XP_015515569.1	Proteína Não Caracterizada
<i>Wasmannia auropunctata</i>	XP_011684314.1	Proteína Não Caracterizada
<i>Orussus abietinus</i>	XP_012274833.1	Proteína Não Caracterizada
<i>Acromyrmex echinator</i>	XP_011067074.1	Proteína Não Caracterizada
<i>Cotesia congregata bracovirus</i>	CBZ06034.1	Integrase Hipotética
<i>Cerapachys biroi</i>	XP_011343510.1	Proteína Não Caracterizada
<i>Caenorhabditis remanei</i>	XP_003097570.1	Proteína Não Caracterizada
<i>Caenorhabditis briggsae</i>	XP_002648085.1	Proteína Não Caracterizada

<i>Caenorhabditis brenneri</i>	EGT60159.1	Integrase Hipotética
<i>Nasonia vitripennis</i>	XP_008205080.1	Proteína Não Caracterizada
<i>Harpegnathos saltator</i>	XP_011147013.1	Proteína Não Caracterizada
<i>Cephus cinctus</i>	XP_015602860.1	Proteína Não Caracterizada
<i>Dinoponera quadriceps</i>	XP_014486516.1	Proteína Não Caracterizada
<i>Exaiptasia pallida</i>	KXJ05594.1	Putative Transposon Protein
<i>Tetranychus urticae</i>	XP_015785484.1	Proteína Não Caracterizada
<i>Solenopsis invicta</i>	XP_011172557.1	Proteína Não Caracterizada
<i>Chrysemys picta bellii</i>	XP_008174408.1	Proteína Não Caracterizada
<i>Cyphomyrmex costatus</i>	KYM94337.1	Proteína Hipotética
<i>Atta cephalotes</i>	XP_012056381.1	Proteína Não Caracterizada
<i>Trachymyrmex cornetzi</i>	KYN22396.1	Hypotetical Protein
<i>Xiphophorus maculatus</i>	AAA93301.1	Integrase
<i>Pogonomyrmex barbatus</i>	XP_011635951.1	Proteína Não Caracterizada
<i>Solemya velum gill symbiont</i>	WP_043118747.1	Integrase Hipotética
<i>Parasteatoda tepidariorum</i>	XP_015903177.1	Proteína Não Caracterizada
<i>Crassostrea gigas</i>	XP_011424688.1	Proteína Não Caracterizada
<i>Monomorium pharaonis</i>	XP_012543416.1	Proteína Não Caracterizada
<i>Acyrtosiphon pisum</i>	XP_003243422.1	Proteína Não Caracterizada
<i>Oikopleura dioica</i>	AAS21408.1	Putative Integrase
<i>Athalia rosae</i>	XP_012253782.1	Proteína Não Caracterizada
<i>Diuraphis noxia</i>	XP_015374318.1	Proteína Não Caracterizada
<i>Saccoglossus kowalevskii</i>	XP_006813063.1	Proteína Não Caracterizada
<i>Hydra vulgaris</i>	XP_012561978.1	Proteína Não Caracterizada
<i>Capitella teleta</i>	ELT99049.1	Proteína Hipotética
<i>Atta colombica</i>	KYM76697.1	Proteína Hipotética
<i>Acropora digitifera</i>	XP_015763003.1	Proteína Não Caracterizada
<i>Tetrahymena thermophila</i>	AAG23743.1	Integrase
<i>Alligator mississippiensis</i>	XP_014466910.1	Proteína Não Caracterizada
<i>Nosema bombycis</i>	ABE26655.1	Integrase
<i>Fundulus heteroclitus</i>	XP_012708440.1	Proteína Não Caracterizada
<i>Poecilia mexicana</i>	XP_014867560.1	Proteína Não Caracterizada
<i>Ancylostoma ceylanicum</i>	EYC05399.1	Proteína Hipotética
<i>Danio rerio</i>	XP_009301155.1	Proteína Não Caracterizada
<i>Linepithema humile</i>	XP_012216150.1	Proteína Não Caracterizada

APÊNDICE B - Informações referentes às sequências utilizadas na filogenia da superfamília Helitron.

REPBASE

Elemento	Organismo	Superfamília
Helitron-2_SIn	<i>Solenopsis invicta</i>	Helitron

LITERATURA

Elemento	Organismo	Código Acesso NCBI
Ha_Da_A1_E	<i>Drosophila ananassae</i>	AAPP01019267.1
HgAi_Dy_A	<i>Drosophila yakuba</i>	AAEU02004714.1
HgAi_Rp_A	<i>Rhodnius prolixus</i>	ACPB01050589.1
HgAi_Da_A	<i>Drosophila ananassae</i>	AAPP01018495.1
HgAi_Bm_Afr	<i>Bombyx mori</i>	AADK01001589.1
Hs_Bm_A1fr	<i>Bombyx mori</i>	AADK01008449.1
Hs_Ap_Afr	<i>Acyrtosiphon pisum</i>	ABLF01048867.1
Hs_Dw_A1	<i>Drosophila willistoni</i>	AAPP01018364.1
Hs_Rp_Afr	<i>Rhodnius prolixus</i>	ACPB01059866.1
Hs_Da_A1	<i>Drosophila ananassae</i>	AAPP01018364.1
Hs-CsMBV_A2fr	<i>Cotesia sesamiae Mombasa bracovirus</i>	EF710637.1
Hm_Ap_A	<i>Acyrtosiphon pisum</i>	AC202211.4
Hm_Bm_A1	<i>Bombyx mori</i>	AADK01000792.1

BLASTP *Leptopilina bouvardi*

Organismo	Código de Acesso NCBI	Classificação NCBI
<i>Wasmannia auropunctata</i>	XP_011685465.1	Proteína Predita
<i>Cyphomyrmex costatus</i>	KYM96549.1	Helicase
<i>Diachasma alloeum</i>	XP_015116448.1	Proteína Predita
<i>Trachymyrmex cornetzi</i>	KYN14600.1	Helicase
<i>Lasius niger</i>	KMQ86963.1	Proteína Hipotética
<i>Vollenhovia emeryi</i>	XP_011859152.1	Proteína Predita
<i>Nasonia vitripennis</i>	XP_016842917.1	Proteína Predita
<i>Microplitis demolitor</i>	XP_008554267.1	Proteína Predita
<i>Fopius arisanus</i>	XP_011313360.1	Proteína Predita
<i>Neodiprion lecontei</i>	XP_015509564.1	Proteína Predita
<i>Strongylocentrotus purpuratus</i>	XP_001185162.2	Proteína Predita
<i>Papilio xuthus</i>	XP_013168531.1	Proteína Predita
<i>Hydra vulgaris</i>	XP_012557420.1	Proteína Predita
<i>Metaseiulus occidentalis</i>	XP_003746054.1	Proteína Predita
<i>Caenorhabditis remanei</i>	XP_003097452.1	Proteína Hipotética
<i>Gossypium hirsutum</i>	XP_016734092.1	Proteína Predita
<i>Zostera marina</i>	KMZ62956.1	Proteína Hipotética
<i>Ancylostoma duodenale</i>	KIH55324.1	Proteína Hipotética
<i>Haemonchus contortus</i>	CDJ90379.1	Proteína Não Caracterizada
<i>Plasmodiophora brassicae</i>	CEO98944.1	Proteína Hipotética
<i>Ancylostoma ceylanicum</i>	EPB76918.1	Proteína Hipotética
<i>Erythranthe guttata</i>	XP_012847176.1	Proteína Predita
<i>Octopus bimaculoides</i>	XP_014784237.1	Proteína Predita
<i>Mycena chlorophos</i>	GAT45315.1	<i>Transcriptional Factor B3</i>
<i>Nicotiana glauca</i>	XP_009803328.1	Proteína Predita

<i>Oryza sativa</i>	XP_015650422.1	Proteína Predita
<i>Sorghum bicolor</i>	XP_002448970.1	Proteína Predita
<i>Arachis ipaensis</i>	XP_016166002.1	Proteína Predita
<i>Musa acuminata</i>	ABF70031.1	Helicase
<i>Zea mays</i>	XP_008650174.1	Proteína Predita
<i>Fragaria vesca subsp. vesca</i>	XP_011465780.1	Proteína Predita
<i>Daphnia magna</i>	KZS08403.1	Helicase
<i>Brachypodium distachyon</i>	XP_010233128.1	Proteína Predita
<i>Tilletia indica</i>	OAJ05841.1	Proteína Hipotética
<i>Solanum tuberosum</i>	XP_015167070.1	Proteína Predita
<i>Setaria italica</i>	KQL29454.1	Proteína Hipotética
<i>Pyrus x bretschneideri</i>	XP_009365391.1	Proteína Predita
<i>Absidia glauca</i>	SAM01481.1	Proteína Hipotética
<i>Diuraphis noxia</i>	XP_015377552.1	Proteína Predita
<i>Nicotiana tabacum</i>	XP_016460125.1	Proteína Predita
<i>Hypsizygus marmoreus</i>	KYQ37551.1	Helicase
<i>Glycine max</i>	KRH57698.1	Proteína Hipotética
<i>Termitomyces sp. J132</i>	KNZ82329.1	Helicase
<i>Solanum pennellii</i>	XP_015057654.1	Proteína Predita
<i>Parasitella parasitica</i>	CEP10635.1	Proteína Hipotética
<i>Daucus carota subsp. sativus</i>	XP_017245330.1	Proteína Predita
<i>Prunus mume</i>	XP_016649779.1	Proteína Predita
<i>Phytophthora infestans T30-4</i>	XP_002998261.1	Helicase
<i>Tilletia caries</i>	OAJ10885.1	Proteína Hipotética
<i>Tilletia controversa</i>	OAJ31089.1	Proteína Hipotética

APÊNDICE C - Informações referentes às sequências utilizadas na filogenia da superfamília Chapaev.

REPBASE

Elemento	Organismo	Superfamília
Chapaev-1N1_NV	<i>Nematostella vectensis</i>	Chapaev
Chapaev-1_BF	<i>Branchiostoma floridae</i>	Chapaev
Chapaev-2_BF	<i>Branchiostoma floridae</i>	Chapaev
Chapaev-2_NV	<i>Nematostella vectensis</i>	Chapaev
Chapaev-3A_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-3B_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-3_NV	<i>Hydra vulgaris</i>	Chapaev
Chapaev-4_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-6_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-7_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-8_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-9_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-5_HMa	<i>Hydra vulgaris</i>	Chapaev

Chapaev-2_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-19_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-18_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-16_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-15_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-14_HM	<i>Hydra vulgaris</i>	Chapaev
Chapaev-10_HM	<i>Hydra vulgaris</i>	Chapaev

LITERATURA

Elemento	Organsimo	Código de Acesso NCBI
Garfiel_MB	<i>Cotesia sesamiae bracovirus</i>	EF710639
Conan_AC	<i>Anolis carolinensis</i>	AAWZ02011613
BLASTP <i>Leptopilina boulandi</i>	CL231Contig1	
Organismo	Código Acesso NCBI	Classificação NCBI
<i>Trachymyrmex cornetzi</i>	KYN13106.1	Proteína Hipotética
<i>Lasius niger</i>	KMQ85949.1	Proteína Hipotética
<i>Amyelois transitella</i>	XP_013197360.1	Proteína Não Caracterizada
<i>Cyphomyrmex costatus</i>	KYM94643.1	Proteína Hipotética
<i>Helobdella robusta</i>	XP_009023975.1	Proteína Não Caracterizada
<i>Trichinella zimbabwensis</i>	KRZ05510.1	Proteína Hipotética
<i>Cotesia congregata</i>	CAR82249.1	Proteína Hipotética
<i>Papilio xuthus</i>	XP_013164143.1	Proteína Não Caracterizada
<i>Cotesia sesamiae</i>	AGO14414.1	Proteína Hipotética
<i>Hydra vulgaris</i>	XP_012562008.1	Proteína Não Caracterizada
<i>Cerapachys biroi</i>	EZA62162.1	Proteína Hipotética
<i>Bactrocera cucurbitae</i>	XP_011180126.1	Proteína Não Caracterizada
<i>Metaseiulus occidentalis</i>	XP_003740020.1	Proteína Não Caracterizada
<i>Octopus bimaculoides</i>	XP_014785903.1	Proteína Não Caracterizada
<i>Camponotus floridanus</i>	EFN69642.1	Proteína Hipotética
<i>Dinoponera quadriceps</i>	XP_014478525.1	Proteína Não Caracterizada
<i>Vollenhovia emeryi</i>	XP_011859702.1	Proteína Não Caracterizada
<i>Harpegnathos saltator</i>	XP_011144139.1	Proteína Não Caracterizada
<i>Bombyx mori</i>	XP_004929732.1	Proteína Não Caracterizada
<i>Trichinella papuae</i>	KRZ73025.1	Proteína Hipotética
<i>Trichinella nelsoni</i>	KRX13342.1	Proteína Hipotética
<i>Copidosoma floridanum</i>	XP_014212974.1	Proteína Não Caracterizada
<i>Zootermopsis nevadensis</i>	KDR11799.1	Proteína Hipotética
<i>Trichogramma pretiosum</i>	XP_014221807.1	Proteína Não Caracterizada
<i>Trichinella pseudospiralis</i>	KRX87925.1	Proteína Hipotética
<i>Trichinella spiralis</i>	KRY36154.1	Proteína Hipotética
<i>Trichinella britovi</i>	KRY46233.1	Proteína Hipotética
<i>Trichinella murrelli</i>	KRX33543.1	Proteína Hipotética
<i>Tribolium castaneum</i>	EFA13536.1	Proteína Hipotética
<i>Nasonia vitripennis</i>	XP_001602271.3	Proteína Não Caracterizada
<i>Dendroctonus ponderosae</i>	ENN71795.1	Proteína Hipotética

<i>Parasteatoda tepidariorum</i>	XP_015919051.1	Proteína Não Caracterizada
<i>Trachymyrmex septentrionalis</i>	KYN41547.1	Proteína Hipotética
<i>Oryctes borbonicus</i>	KRT81892.1	Proteína Hipotética
<i>Danaus plexippus</i>	EHJ66079.1	Proteína Hipotética
<i>Culex quinquefasciatus</i>	XP_001844961.1	Proteína Não Caracterizada
<i>Papilio machaon</i>	XP_014356480.1	Proteína Não Caracterizada
<i>Bactrocera oleae</i>	XP_014100700.1	Proteína Não Caracterizada
<i>Stegodyphus mimosarum</i>	KFM57282.1	Proteína Hipotética
<i>Stomoxys calcitrans</i>	XP_013113917.1	Proteína Não Caracterizada
<i>Limulus polyphemus</i>	XP_013790331.1	Proteína Não Caracterizada
<i>Perkinsus marinus</i>	XP_002775426.1	Proteína Não Caracterizada
<i>Esox lucius</i>	XP_012988426.1	Proteína Não Caracterizada
<i>Aedes aegypti</i>	XP_001654679.1	Proteína Não Caracterizada
<i>Anopheles sinensis</i>	KFB47037.1	<i>PA-Like Protein</i>
<i>Anopheles gambiae str. PEST</i>	XP_001688699.1	Proteína Não Caracterizada
<i>Diuraphis noxia</i>	XP_015372174.1	Proteína Não Caracterizada
<i>Trichinella patagoniensis</i>	KRY08036.1	Proteína Hipotética
<i>Caenorhabditis remanei</i>	XP_003107288.1	Proteína Não Caracterizada
<i>Atta colombica</i>	KYM81997.1	Proteína Hipotética

APÊNDICE D - Informações referentes às sequências utilizadas na filogenia da superfamília R1.

REPBASE

Elemento	Organismo	Superfamília
R1-4_DAn	<i>Drosophila ananassae</i>	R1
LINER1-2_NVi	<i>Nasonia vitripennis</i>	R1
SART-1_Api	<i>Ancylostoma pisum</i>	R1

LITERATURA

Elemento	Organismo	Código de Acesso NCBI
R1DM	<i>Drosophila melanogaster</i>	140023
R1Sc	<i>Sciara coprophila</i>	L00945
R1NvA	<i>Nasonia vitripennis</i>	L00940
R1NvB	<i>Nasonia vitripennis</i>	L00941
R1Pj3	<i>Popillia japonica</i>	L00944

BLASTP *Leptopilina boulardi*

Organismo	Código Acesso NCBI	Classificação NCBI
<i>Tribolium castaneum</i>	BAK38646.1	Proteína Não Nomeada
<i>Anopheles gambiae</i>	BAC57906.1	<i>Reverse Transcriptase</i>
<i>Drosophila melanogaster</i>	CAB99192.1	<i>Reverse Transcriptase</i>
<i>Ceratitidis capitata</i>	XP_012160474.1	Proteína Não Caracterizada
<i>Lasius niger</i>	KMQ90166.1	<i>Reverse Transcriptase</i>
<i>Piscirickettsia salmonis</i>	WP_047927781.1	Proteína Hipotética
<i>Bombyx mori</i>	BAC57926.1	<i>Reverse Transcriptase</i>

<i>Papilio xuthus</i>	BAC06454.1	<i>Reverse Transcriptase</i>
<i>Metaseiulus occidentalis</i>	XP_003740993.1	Proteína Não Caracterizada
<i>Danaus plexippus</i>	EHI67685.1	<i>Reverse Transcriptase</i>
<i>Papilio machaon</i>	XP_014356048.1	Proteína Não Caracterizada
<i>Vollenhovia emeryi</i>	XP_011860276.1	Proteína Não Caracterizada
<i>Camponotus floridanus</i>	XP_011251459.1	Proteína Não Caracterizada
<i>Wasmannia auropunctata</i>	XP_011707944.1	Proteína Não Caracterizada
<i>Drosophila yakuba</i>	XP_015044786.1	Proteína Não Caracterizada
<i>Dendroctonus ponderosae</i>	ENN78512.1	Proteína Hipotética
<i>Cotesia sesamiae</i>	AGO14420.1	<i>Reverse Transcriptase</i>
<i>Diachasma alloeum</i>	XP_015123442.1	Proteína Não Caracterizada
<i>Harpegnathos saltator</i>	XP_011142303.1	Proteína Não Caracterizada
<i>Stegodyphus mimosarum</i>	KFM62077.1	Proteína Hipotética
<i>Acyrtosiphon pisum</i>	XP_008181354.1	Proteína Não Caracterizada
<i>Cerapachys biroi</i>	XP_011332975.1	Proteína Não Caracterizada
<i>Diaphorina citri</i>	XP_008479311.1	Proteína Não Caracterizada
<i>Amyelois transitella</i>	XP_013194811.1	Proteína Não Caracterizada
<i>Microplitis demolitor</i>	XP_014298515.1	Proteína Não Caracterizada
<i>Solenopsis invicta</i>	XP_011156700.1	Proteína Não Caracterizada
<i>Blattella germanica</i>	AKU04648.1	<i>Polyprotein</i>
<i>Pyrus x bretschneideri</i>	XP_009351357.1	Proteína Não Caracterizada
<i>Plutella xylostella</i>	XP_011558883.1	Proteína Não Caracterizada
<i>Fopius arisanus</i>	XP_011309785.1	Proteína Não Caracterizada
<i>Metaseiulus occidentalis</i>	XP_003739696.1	Proteína Não Caracterizada
<i>Linepithema humile</i>	XP_012222696.1	Proteína Não Caracterizada
<i>Orussus abietinus</i>	XP_012280817.1	Proteína Não Caracterizada
<i>Monomorium pharaonis</i>	XP_012543488.1	Proteína Não Caracterizada
<i>Nasonia vitripennis</i>	XP_008205084.1	Proteína Não Caracterizada
<i>Aedes aegypti</i>	AAB65093.1	<i>Lian-Aa1 retrotransposon protein</i>
<i>Histoplasma capsulatum NAM1</i>	XP_001537962.1	Proteína Não Caracterizada
<i>Megachile rotundata</i>	XP_012144078.1	Proteína Não Caracterizada
<i>Drosophila persimilis</i>	XP_002014107.1	Proteína Não Caracterizada
<i>Trametes cinnabarina</i>	CDO77910.1	Proteína Não Caracterizada
<i>Copidosoma floridanum</i>	XP_014212973.1	Proteína Não Caracterizada
<i>Ixodes scapularis</i>	EEC03448.1	<i>Reverse Transcriptase</i>
<i>Aedes albopictus</i>	AAQ01583.1	<i>Reverse Transcriptase</i>
<i>Penicillium brasilianum</i>	CEJ62916.1	<i>Reverse Transcriptase</i>
<i>Drosophila virilis</i>	XP_015023737.1	Proteína Não Caracterizada
<i>Saturnia japonica</i>	BAB21515.1	TRASDJ
<i>Drosophila mojavensis</i>	XP_015015679.1	Proteína Não Caracterizada
<i>Papilio polytes</i>	XP_013139951.1	Proteína Não Caracterizada
<i>Trichophyton tonsurans</i>	EGD99960.1	Proteína Hipotética
<i>Penicillium roqueforti FM164</i>	CDM38421.1	<i>Probable transposable element</i>
<i>Drosophila willistoni</i>	XP_015033996.1	Proteína Não Caracterizada

Erysiphe necator

KHJ34379.1

Proteína Hipotética

APÊNDICE E - Informações referentes às sequências utilizadas na filogenia da superfamília Penelope.

REPBASE

Elemento	Organismo	Superfamília
Penelope-1	<i>Nasonia vitripennis</i>	Penelope

LITERATURA

Elemento	Organismo	Código de Acesso NCBI
Penelope-like	<i>Drosophila virilis</i>	AAA92124
Penelope-like	<i>Drosophila americana texana</i>	AY999940.1
Penelope-like	<i>Drosophila novamexicana</i>	AY999957.1
Penelope-like	<i>Drosophila lummei</i>	AY999958.1
Penelope-like	<i>Drosophila littoralis</i>	AY999956.1
Penelope-like	<i>Drosophila lacicola</i>	AY999947.1
Penelope-like	<i>Drosophila ezoana</i>	AY999946.1
Penelope-like	<i>Drosophila kanekoi</i>	AY999950.1
Penelope-like	<i>Drosophila canadiana</i>	AY999944.1
Penelope-like	<i>Drosophila montana</i>	AY999945.1
Penelope-like	<i>Drosophila flavomontana</i>	AY999960.1
Penelope_Pm	<i>Bombyx mori</i>	BAAB01101440
Perere10_Sm	<i>Schistosoma mansoni</i>	BN000801
Perere_Smed	<i>Schmidtea mediterranea</i>	AY066937.1
Perere_ES	<i>Euprymna scolopes</i>	AY149468.1
Penelope-like	<i>Oikopleura dioica</i>	AY634216.1
Osiris1_Tc	<i>Tribolium castaneum</i>	AAJJ01005571
Osiris_Ae	<i>Aedes aegypti</i>	XM_011495317.1
Penelope-like	<i>Petromyzon marinus</i>	AY577942.1
Penelope-like	<i>Penaeus monodon</i>	AF077579.1
Poseidon	<i>Tetraodon nigroviridis</i>	AJ621377.1
Cercyon_Sm	<i>Schistosoma mansoni</i>	BK000685
Neptune	<i>Tetraodon nigroviridis</i>	AJ621374
Poseidon_Tet	<i>Tetraodon nigroviridis</i>	AJ621377
Coprina_Pc1	<i>Phanerochaete chrysosporium</i>	AADS01000564
Penelope_Dw	<i>Drosophila willistoni</i>	AAQB01010750
Penelope_Ae.A	<i>Aedes aegypti</i>	AAGE02009673
Neptune_1_Xt	<i>Xenopus tropicalis</i>	AC149155.3
BLASTP <i>Leptopilina bouvardi</i>	CL309Contig4	
Organismo	Código Acesso NCBI	Classificação NCBI
<i>Trichogramma pretiosum</i>	XP_014221249.1	Proteína Não Caracterizada
<i>Vollenhovia emeryi</i>	XP_011871696.1	Proteína Não Caracterizada
<i>Cerapachys biroii</i>	XP_011336716.1	Proteína Não Caracterizada
<i>Microplitis demolitor</i>	XP_008553167.1	Proteína Não Caracterizada

<i>Diachasma alloeum</i>	XP_015121934.1	Proteína Não Caracterizada
<i>Wasmannia auropunctata</i>	XP_011705916.1	Proteína Não Caracterizada
<i>Fopius arisanus</i>	XP_011303305.1	Proteína Não Caracterizada
<i>Camponotus floridanus</i>	EFN74754.1	Proteína Hipotética
<i>Lasius niger</i>	KMQ93684.1	Proteína Hipotética
<i>Pogonomyrmex barbatus</i>	XP_011646862.1	Proteína Não Caracterizada
<i>Polistes canadensis</i>	XP_014614262.1	Proteína Não Caracterizada
<i>Solenopsis invicta</i>	XP_011171623.1	Proteína Não Caracterizada
<i>Aedes aegypti</i>	XP_011171623.1	Proteína Não Caracterizada
<i>Monomorium pharaonis</i>	XP_012526439.1	Proteína Não Caracterizada
<i>Drosophila virilis</i>	AAL14979.1	<i>Reverse transcriptase</i>
<i>Drosophila melanogaster</i>	AAR86939.1	<i>Reverse transcriptase</i>
<i>Trichuris suis</i>	KFD45899.1	Proteína Hipotética
<i>Opisthorchis viverrini</i>	XP_009165178.1	Proteína Não Caracterizada
<i>Lingula anatina</i>	XP_013411770.1	Proteína Não Caracterizada
<i>Cotesia congregata</i>	CCQ71227.1	Proteína Hipotética
<i>Xenopus tropicalis</i>	XP_012820560.1	Proteína Não Caracterizada
<i>Schistosoma japonicum</i>	CAX83714.1	<i>Reverse transcriptase</i>
<i>Schistosoma mansoni</i>	DAA00890.1	<i>Reverse transcriptase</i>
<i>Megachile rotundata</i>	XP_012143822.1	Proteína Não Caracterizada
<i>Ancylostoma ceylanicum</i>	EYB95037.1	Proteína Hipotética
<i>Marsupenaeus japonicus</i>	BAM35674.1	<i>Reverse transcriptase</i>
<i>Priapulid caudatus</i>	XP_014676727.1	Proteína Não Caracterizada
<i>Danaus plexippus</i>	EHJ74365.1	<i>Putative ankyrin</i>
<i>Linepithema humile</i>	XP_012235687.1	Proteína Não Caracterizada
<i>Cynoglossus semilaevis</i>	XP_008333575.1	Proteína Não Caracterizada
<i>Tetraodon nigroviridis</i>	CAF94542.1	Proteína não nomeada
<i>Anolis carolinensis</i>	XP_008109886.1	Proteína Não Caracterizada
<i>Pundamilia nyererei</i>	XP_005747422.1	Proteína Não Caracterizada
<i>Takifugu rubripes</i>	AAK58879.1	<i>Reverse transcriptase</i>
<i>Vigna radiata var. radiata</i>	XP_014523126.1	Proteína Não Caracterizada
<i>Stomoxys calcitrans</i>	XP_013099882.1	Proteína Não Caracterizada
<i>Strongylocentrotus purpuratus</i>	XP_003731556.1	Proteína Não Caracterizada
<i>Diaphorina citri</i>	XP_008484571.1	Proteína Não Caracterizada
<i>Amphimedon queenslandica</i>	XP_011407404.1	Proteína Não Caracterizada
<i>Limulus polyphemus</i>	XP_013772595.1	Proteína Não Caracterizada
<i>Saccoglossus kowalevskii</i>	XP_006811186.1	Proteína Não Caracterizada
<i>Crassostrea gigas</i>	XP_011415401.1	Proteína Não Caracterizada
<i>Acromyrmex echinator</i>	XP_011069268.1	Proteína Não Caracterizada
<i>Hydra vulgaris</i>	XP_012566254.1	Proteína Não Caracterizada
<i>Oncorhynchus mykiss</i>	CDQ83180.1	Proteína não nomeada
<i>Caenorhabditis remanei</i>	XP_003101932.1	Proteína Hipotética
<i>Rhodospiridium toruloides</i>	CDR48584.1	<i>Reverse transcriptase</i>
<i>Papilio machaon</i>	XP_014364293.1	Proteína Não Caracterizada

Oikopleura dioica

CBY10203.1

Proteína não nomeada

APÊNDICE F - Informações referentes às sequências utilizadas na filogenia da superfamília Loa.

REPBASE**Elemento**

Loa-I

Organismo*Solenopsis invicta***Superfamília**

Loa

LITERATURA

Dados não encontrados

BLASTP Leptopilina boulardi**CL190Contig1****Organismo****Código Acesso NCBI****Classificação NCBI***Wasmannia auropunctata*

XP_011698268.1

Proteína Não Caracterizada

Nasonia vitripennis

XP_008202330.1

Proteína Não Caracterizada

Linepithema humile

XP_012234564.1

Proteína Não Caracterizada

Solenopsis invicta

XP_011168851.1

Proteína Não Caracterizada

Parasteatoda tepidariorum

XP_015923113.1

Proteína Não Caracterizada

Trichogramma pretiosum

XP_014229935.1

Proteína Não Caracterizada

Papilio machaon

XP_014360265.1

Proteína Não Caracterizada

Plutella xylostella

XP_011556288.1

Proteína Não Caracterizada

Diaphorina citri

XP_008482174.1

Proteína Não Caracterizada

Bactrocera tryoni

AMS38365.1

Proteína Hipotética

Aedes aegypti

AAB65093.1

*Reverse Transcriptase**Halyomorpha halys*

XP_014289988.1

Proteína Não Caracterizada

Fopius arisanus

XP_011313427.1

Proteína Não Caracterizada

Polistes dominula

XP_015173372.1

Proteína Não Caracterizada

Drosophila subobscura

AAB92394.1

*Reverse Transcriptase**Drosophila miranda*

S37765

Proteína Hipotética

Diachasma alloeum

XP_015118148.1

Proteína Não Caracterizada

Dufourea novaeangliae

XP_015437811.1

Proteína Não Caracterizada

Lasius niger

KMQ87436.1

Transcriptase reversa

Microplitis demolitor

XP_008546293.1

Proteína Não Caracterizada

Athalia rosae

XP_012252290.1

Proteína Não Caracterizada

Bactrocera cucurbitae

XP_011194809.1

Proteína Não Caracterizada

Drosophila yakuba

XP_015054647.1

Proteína Não Caracterizada

Lucilia cuprina

KNC30999.1

Proteína Hipotética

Dendroctonus ponderosae

ERL95993.1

Proteína Hipotética

Dinoponera quadricaps

XP_014476493.1

Proteína Não Caracterizada

Musca domestica

XP_005181784.1

Proteína Não Caracterizada

Tetranychus urticae

XP_015793894.1

Proteína Não Caracterizada

Trachymyrmex zeteki

KYQ57869.1

Proteína Hipotética

Cyprinodon variegatus

XP_015260458.1

Proteína Não Caracterizada

Bombyx mori

BAB21511.1

Tras3

Rhizoctonia solani

CUA67582.1

Proteína Não Caracterizada

Aedes albopictus

KXJ70294.1

Proteína Não Caracterizada

<i>Trachymyrmex septentrionalis</i>	KYN41602.1	Proteína Não Caracterizada
<i>Reticulomyxa filosa</i>	ETO08846.1	Proteína Hipotética
<i>Daphnia magna</i>	KZS20285.1	Proteína Não Caracterizada
<i>Talaromyces stipitatus</i>	XP_002488430.1	Proteína Não Caracterizada
<i>Monomorium pharaonis</i>	XP_012536890.1	Proteína Não Caracterizada
<i>Tilletia caries</i>	OAI96484.1	Proteína Hipotética
<i>Drosophila melanogaster</i>	CAC16871.1	<i>Pol Polyprotein</i>
<i>Aspergillus oryzae</i>	XP_001826621.2	Proteína Não Caracterizada
<i>Trametes cinnabarina</i>	CDO78032.1	Proteína Hipotética
<i>Hymenolepis microstoma</i>	CDS31781.2	Reverse Transcriptase
<i>Sinocyclocheilus rhinoceros</i>	XP_016388918.1	Proteína Não Caracterizada
<i>Biomphalaria glabrata</i>	ABN58714.1	<i>Pol-like Polyprotein</i>
<i>Penicillium nordicum</i>	KOS41014.1	Proteína Hipotética
<i>Cimex lectularius</i>	XP_014258194.1	Proteína Não Caracterizada
<i>Histoplasma capsulatum</i>	XP_001539168.1	Proteína Não Caracterizada
<i>Austrofundulus limnaeus</i>	XP_013870655.1	Proteína Não Caracterizada
<i>Sclerotinia sclerotiorum</i>	XP_001592153.1	Proteína Não Caracterizada

APÊNDICE G - Informações referentes às sequências utilizadas na filogenia da superfamília L2.

REPBASE

Elemento	Organismo	Superfamília
L2-1_SIn	<i>Solenopsis invicta</i>	L2

LITERATURA

Dados não encontrados

BLASTP *Leptopilina bouleari*

CL190Contig1

Organismo	Código Acesso NCBI	Classificação NCBI
<i>Vollenhovia emeryi</i>	XP_011859007.1	Proteína Não Caracterizada
<i>Lasius niger</i>	XP_011859003.1	Proteína Não Caracterizada
<i>Diachasma alloeum</i>	XP_015119647.1	Proteína Não Caracterizada
<i>Microplitis demolitor</i>	XP_014294909.1	Proteína Não Caracterizada
<i>Cephus cinctus</i>	XP_015605528.1	Proteína Não Caracterizada
<i>Cerapachys biroi</i>	XP_011337487.1	Proteína Não Caracterizada
<i>Neodiprion lecontei</i>	XP_015524184.1	Proteína Não Caracterizada
<i>Trichogramma pretiosum</i>	XP_014224251.1	Proteína Não Caracterizada
<i>Nasonia vitripennis</i>	XP_008205593.1	Proteína Não Caracterizada
<i>Copidosoma floridanum</i>	XP_014212890.1	Proteína Não Caracterizada
<i>Papilio machaon</i>	KPJ06072.1	<i>Putative Reverse Transcriptase</i>
<i>Athalia rosae</i>	XP_012269497.1	Proteína Não Caracterizada
<i>Wasmannia auropunctata</i>	XP_011685479.1	Proteína Não Caracterizada
<i>Diaphorina citri</i>	XP_008475834.1	Proteína Não Caracterizada
<i>Amyelois transitella</i>	XP_013185517.1	Proteína Não Caracterizada
<i>Aedes aegypti</i>	AAZ15237.1	<i>Reverse Transcriptase</i>
<i>Exaiptasia pallida</i>	KXJ18334.1	<i>Putative Reverse Transcriptase</i>

<i>Papilio xuthus</i>	KPJ04628.1	<i>Putative Reverse Transcriptase</i>
<i>Acropora digitifera</i>	XP_015761190.1	Proteína Não Caracterizada
<i>Danio rerio</i>	BAE46429.1	<i>Reverse Transcriptase</i>
<i>Lycodichthys dearboni</i>	ADJ80991.1	<i>Reverse Transcriptase</i>
<i>Larimichthys crocea</i>	XP_010743720.1	Proteína Não Caracterizada
<i>Strongylocentrotus purpuratus</i>	XP_792312.2	Proteína Não Caracterizada
<i>Takifugu rubripes</i>	AAD19348.1	<i>Reverse Transcriptase</i>
<i>Maylandia zebra</i>	XP_014263144.1	Proteína Não Caracterizada
<i>Anguilla japonica</i>	BAD72127.1	<i>Reverse Transcriptase</i>
<i>Oreochromis niloticus</i>	XP_013123065.1	Proteína Não Caracterizada
<i>Cynoglossum semilaevius</i>	XP_008333619.1	Proteína Não Caracterizada
<i>Plutella xylostella</i>	XP_011550139.1	Proteína Não Caracterizada
<i>Salmo salar</i>	AKP40998.1	<i>Reverse Transcriptase</i>
<i>Paralichthys olivaceus</i>	AAN15747.1	<i>Reverse Transcriptase</i>
<i>Monomorium pharaonsis</i>	XP_012524115.1	Proteína Não Caracterizada
<i>Austrofundulus limnaeus</i>	XP_013877416.1	Proteína Não Caracterizada
<i>Nothobranchius furzeri</i>	XP_015798087.1	Proteína Não Caracterizada
<i>Crassostrea gigas</i>	XP_011446877.1	Proteína Não Caracterizada
<i>Ancylostoma ceylanicum</i>	EYC03347.1	Proteína Hipotética
<i>Pectinophora gossypiella</i>	AEM44817.1	<i>Reverse Transcriptase</i>
<i>Saccoglossus kowalevskii</i>	XP_006825930.1	Proteína Não Caracterizada
<i>Oncorhynchus mykiss</i>	CDQ78441.1	Proteína Não Nomeada
<i>Schistosoma mansoni</i>	CAJ00240.1	<i>Reverse Transcriptase</i>
<i>Solenopsis invicta</i>	XP_011173443.1	Proteína Não Caracterizada
<i>Esox lucius</i>	XP_012988288.1	Proteína Não Caracterizada
<i>Opisthorchis viverrini</i>	XP_009162237.1	Proteína Não Caracterizada
<i>Branchiostoma floridae</i>	AAL40415.1	<i>Reverse Transcriptase</i>
<i>Trachymyrmex zeteki</i>	KYQ60541.1	Proteína Hipotética
<i>Gallus gallus</i>	AAA49027.1	<i>Reverse Transcriptase</i>
<i>Amphimedon queenslandica</i>	XP_011408069.1	Proteína Não Caracterizada
<i>Sinocyclocheilus anshuiensis</i>	XP_016341263.1	Proteína Não Caracterizada
<i>Acanthochelys spixii</i>	BAA88337.1	<i>Reverse Transcriptase</i>
<i>Brassica napus</i>	XP_013733568.1	Proteína Não Caracterizada

APÊNDICE H - Informações referentes às sequências utilizadas na filogenia da superfamília I.

REPBASE

Elemento	Organismo	Superfamília
I-1_Lhu	<i>Linepithema humile</i>	I
I-2_Nvi	<i>Nasonia vitripennis</i>	I

LITERATURA

Dados não encontrados

BLASTP Leptopilina boulardi CL190Contig1

Organismo	Código Acesso NCBI	Classificação NCBI
<i>Trichogramma pretiosum</i>	XP_014235377.1	Proteína Não Caracterizada
<i>Lasius niger</i>	KMQ92689.1	<i>Pol-like protein</i>
<i>Papilio machaon</i>	XP_014366482.1	Proteína Não Caracterizada
<i>Solenopsis invicta</i>	XP_011171611.1	Proteína Não Caracterizada
<i>Camponotus floridanus</i>	EFN62498.1	<i>RNA-directed DNA polymerase</i>
<i>Papilio xuthus</i>	XP_013166190.1	Proteína Não Caracterizada
<i>Wasmannia auropunctata</i>	XP_011687072.1	Proteína Não Caracterizada
<i>Vollenhovia emeryi</i>	XP_011863445.1	Proteína Não Caracterizada
<i>Plutella xylostella</i>	XP_011567766.1	Proteína Não Caracterizada
<i>Pogonomyrmex barbatus</i>	XP_011646738.1	Proteína Não Caracterizada
<i>Amyelois transitella</i>	XP_013183181.1	Proteína Não Caracterizada
<i>Copidosoma floridanum</i>	XP_014214776.1	Proteína Não Caracterizada
<i>Biomphalaria glabrata</i>	ABN58714.1	<i>Pol-like protein</i>
<i>Monomorium pharaonis</i>	XP_012523452.1	Proteína Não Caracterizada
<i>Stegodyphus mimosarum</i>	KFM64962.1	<i>RNA-directed DNA polymerase</i>
<i>Fopius arisanus</i>	XP_011313084.1	Proteína Não Caracterizada
<i>Dinoponera quadriceps</i>	XP_014471567.1	Proteína Não Caracterizada
<i>Danaus plexippus</i>	EHJ74035.1	<i>Pol-like protein</i>
<i>Harpegnathos saltator</i>	XP_011152156.1	Proteína Não Caracterizada
<i>Solemya velum gill symbiont</i>	KHF24557.1	Ribonuclease HI
<i>Diaphorina citri</i>	XP_008487215.1	Proteína Não Caracterizada
<i>Dufourea novaeangliae</i>	XP_015436995.1	Proteína Não Caracterizada
<i>Daphnia magna</i>	KZS09240.1	Proteína Não Caracterizada
<i>Cimex lectularius</i>	XP_014258531.1	Proteína Não Caracterizada
<i>Aedes albopictus</i>	KXJ68559.1	Proteína Hipotética
<i>Trachymyrmex zeteki</i>	KYQ59915.1	Proteína Hipotética
<i>Drosophila melanogaster</i>	AAA70222.2	<i>Reverse Transcriptase</i>
<i>Diachasma alloeum</i>	XP_015111524.1	Proteína Não Caracterizada
<i>Bactrocera tryoni</i>	AMS38373.1	Proteína Hipotética
<i>Athetis dissimilis</i>	ALM26252.1	<i>Gustatory receptor 2</i>
<i>Cyphomyrmex costatus</i>	KYM94053.1	<i>RNA-directed DNA polymerase</i>
<i>Trachymyrmex cornetzi</i>	KYN15796.1	Proteína Hipotética
<i>Megachile rotundata</i>	XP_012153232.1	Proteína Não Caracterizada
<i>Drosophila teissieri</i>	AAA74495.1	Proteína Desconhecida
<i>Acyrtosiphon pisum</i>	XP_016658134.1	Proteína Não Caracterizada
<i>Musca domestica</i>	XP_005175263.1	Proteína Não Caracterizada
<i>Nuttalliella namaqua</i>	AHN53432.1	<i>Exonuclease-Endonuclease</i>
<i>Operophtera brumata</i>	KOB77161.1	<i>Pol-like protein</i>
<i>Aedes aegypti</i>	AAF20019.1	<i>Pol-like protein</i>
<i>Cerapachys biroi</i>	XP_011349882.1	Proteína Não Caracterizada
<i>Sclerotinia sclerotiorum 1980</i>	XP_001586737.1	Proteína Hipotética
<i>Histoplasma capsulatum NAM1</i>	XP_001537028.1	Proteína Não Caracterizada
<i>Alternaria alternata</i>	BAI44754.1	<i>Pol-like protein</i>

<i>Lingula anatina</i>	XP_013416028.1	Proteína Não Caracterizada
<i>Reticulomyxa filosa</i>	ETO08846.1	Proteína Hipotética
<i>Adineta vaga</i>	ACZ73627.1	<i>Reverse Transcriptase</i>
<i>Trametes cinnabarina</i>	CDO77910.1	Proteína Hipotética
<i>Acromyrmex echinator</i>	XP_011049768.1	Proteína Não Caracterizada
<i>Austrofundulus limnaeus</i>	XP_013884305.1	Proteína Não Caracterizada
<i>Purpureocillium lilacinum</i>	OAQ58464.1	<i>Reverse Transcriptase</i>

APÊNDICE I - Informações referentes às sequências utilizadas na filogenia da superfamília Gypsy.

REPBASE

Elemento	Organismo	Superfamília
Gypsy-18-I	<i>Nasonia vitripennis</i>	Gypsy
Gypsy-5	<i>Nasonia vitripennis</i>	Gypsy
Gypsy-30-I	<i>Nasonia vitripennis</i>	Gypsy
Gypsy-27-I	<i>Nasonia vitripennis</i>	Gypsy
Gypsy-23-I	<i>Nasonia vitripennis</i>	Gypsy
Gypsy-19-I	<i>Nasonia vitripennis</i>	Gypsy
Gypsy-33-I	<i>Nasonia vitripennis</i>	Gypsy
Gypsy-9	<i>Pogonomyrmex barbatus</i>	Gypsy
Gypsy-16-I	<i>Pogonomyrmex barbatus</i>	Gypsy
Gypsy-20-I	<i>Pogonomyrmex barbatus</i>	Gypsy
Gypsy-3	<i>Camponotus floridanus</i>	Gypsy
Gypsy-8-I	<i>Camponotus floridanus</i>	Gypsy
Bel	<i>Drosophila melanogaster</i>	Bel
Max	<i>Drosophila melanogaster</i>	Bel
Ninja	<i>Drosophila simulans</i>	Bel
Pao	<i>Bombyx mori</i>	Bel

LITERATURA

Elemento	Organsimo	Código de Acesso NCBI
17.6	<i>Drosophila melanogaster</i>	130405
297	<i>Drosophila melanogaster</i>	130398
412	<i>Drosophila melanogaster</i>	130583
Athila4-1	<i>Arabidopsis thaliana</i>	7267363
Blastopia	<i>Drosophila melanogaster</i>	415798
Boudicca	<i>Schistosoma mansoni</i>	51235680
Burdock	<i>Drosophila melanogaster</i>	1905852
Calypso	<i>Glycine max</i>	6671126
Cer	<i>Caenorhabditis elegans</i>	557718
Cereba	<i>Triticum monococcum subsp. aegilopoides</i>	115392332
CFG1	<i>Azumapecten farreri</i>	124055238
Cft-1	<i>Cladosporium fulvum</i>	280492
Cgret	<i>Colletotrichum gloeosporioides</i>	10946132

CsRN1	<i>Clonorchis sinensis</i>	12958095
Del	<i>Lilium henryi</i>	226407
Grasshopper	<i>Magnaporthe grisea</i>	538067
Gypsy	<i>Drosophila melanogaster/Drosophila virilis</i>	1905852/157587
CAEV	<i>Drosophila melanogaster</i>	14039785
Idefix	<i>Drosophila melanogaster</i>	4165194
Kabuki	<i>Bombyx mori</i>	7243249
Mag	<i>Bombyx mori</i>	84801
Maggy	<i>Magnaporthe grisea</i>	522302
marY1	<i>Tricholoma matsutake</i>	5002510
Mdg1 and Mdg3	<i>Drosophila melanogaster</i>	2133676/1542877
MGRL3	<i>Magnaporthe grisea</i>	12698634
Micropia	<i>Drosophila melanogaster</i>	1030731
Nomad	<i>Drosophila melanogaster</i>	3169259
Ogre	<i>Pisum sativum</i>	34761728
Osvaldo	<i>Drosophila buzzatii</i>	4539021
Real	<i>Alternaria alternata</i>	6683624
REM1	<i>Chlamydomonas reinhardtii</i>	29423677
Retrosat-2	<i>Oryza sativa indica group</i>	4680179
RIRE2	<i>Oryza sativa japonica group</i>	5902445
Skypper	<i>Dictyostelium discoideum</i>	3135479
Skyppy	<i>Fusarium oxysporum</i>	2133292
Springer	<i>Caenorhabditis elegans</i>	14010621
Sushi-ichi	<i>Takifugu rubripes</i>	4378025
Ted	<i>Trichoplusia ni</i>	102939
Tf	<i>Schizosaccharomyces pombe</i>	173477
Tom	<i>Drosophila ananassae</i>	422418
Tor	<i>Oikopleura dioica</i>	49035804
Tv1	<i>Drosophila virilis</i>	3493214
Ty	<i>Saccharomyces cerevisiae</i>	146291076
Ulysses	<i>Drosophila virilis</i>	103519
Woot	<i>Tribolium castaneum</i>	805077
Yoyo	<i>Ceratitidis capitata</i>	1402848
Zam	<i>Drosophila melanogaster</i>	2791289
DmeGypV	<i>Drosophila melanogaster Gypsy virus</i>	M12927
DsuGypV	<i>Drosophila subobscura Gypsy virus</i>	X72390
DviGypV	<i>Drosophila virilis Gypsy virus</i>	M38438
CVRE1	<i>Chlorella vulgaris</i>	AB033237.1
SORE1	<i>Glycine max</i>	AB033271.1
BLASTP <i>Leptopilina boulandi</i>	CI368Contig1	
Organismo	Código Acesso NCBI	Classificação NCBI
<i>Lasius niger</i>	KMQ89174.1	Reverse Transcriptase
<i>Vollenhovia emeryi</i>	XP_011859426.1	Proteína Não Caracterizada

<i>Tribolium castaneum</i>	EFA13518.1	Proteína Hipotética
<i>Microplitis demolitor</i>	XP_014299028.1	Proteína Não Caracterizada
<i>Pediculus humanus corporis</i>	XP_002431501.1	Proteína Não Caracterizada
<i>Megachile rotundata</i>	XP_012151024.1	Proteína Não Caracterizada
<i>Autographa californica nucleopolyhedrovirus</i>	AAA92249.1	<i>Reverse Transcriptase</i>
<i>Drosophila melanogaster</i>	AAX28844.1	Reverse Transcriptase
<i>Diaphorina citri</i>	XP_008481775.1	Proteína Não Caracterizada
<i>Papilio machaon</i>	XP_014358561.1	Proteína Não Caracterizada
<i>Bombyx mori</i>	XP_012547467.1	Proteína Não Caracterizada
<i>Drosophila ananassae</i>	CAA80824.1	<i>Pol Poliprotein</i>
<i>Wolbachia endosymbiont of D. ananassae</i>	EAL58294.1	<i>Pol Poliprotein</i>
<i>Retrotransposon mgd1 D. melanogaster</i>	S70430	Proteína Hipotética
<i>Drosophila virilis</i>	AAC33318.1	<i>Pol Poliprotein</i>
<i>Wasmannia auropunctata</i>	XP_011686344.1	Proteína Não Caracterizada
<i>Zootermopsis nevadensis</i>	KDR23651.1	<i>Pol Poliprotein</i>
<i>Trichinella sp. T6</i>	KRX54395.1	<i>Pol Poliprotein</i>
<i>Stegodyphus mimosarum</i>	KFM56688.1	<i>Pol Poliprotein</i>
<i>Salpingoeca rosetta</i>	EGD79431.1	Proteína Hipotética
<i>Trichinella murrelli</i>	KRX34984.1	Pol Poliprotein
<i>Trichinella nativa</i>	KRZ47896.1	<i>Gag-pol polyprotein precursor</i>
<i>Trichinella zimbabwensis</i>	KRZ00760.1	<i>Gag-pol polyprotein precursor</i>
<i>Ceratitidis capitata</i>	AAC28743.1	<i>Pol Poliprotein</i>
<i>Drosophila yakuba</i>	XP_015044798.1	Proteína Não Caracterizada
<i>Rhizophagus irregularis</i>	EXX71174.1	<i>Gag-pol polyprotein precursor</i>
<i>Spodoptera frugiperda</i>	ADK33996.2	<i>Reverse Transcriptase</i>
<i>Trichinella nelsoni</i>	KRX12461.1	<i>Pol Poliprotein</i>
<i>Trichinella pseudospiralis</i>	KRZ25120.1	<i>Gag-pol polyprotein precursor</i>
<i>Trichuris suis</i>	KFD56336.1	Proteína Hipotética
<i>Sorghum bicolor</i>	XP_002437407.1	Proteína Não Caracterizada
<i>Nuttalliella namaqua</i>	AHN53424.1	Reverse Transcriptase
<i>Rhizopus delemar</i>	EIE92562.1	Proteína Hipotética
<i>Crassostrea gigas</i>	XP_011439224.1	Proteína Não Caracterizada
<i>Schistosoma mansoni</i>	CCD75793.1	Domínio KRAB-A
<i>Trichinella sp. T9</i>	KRX51502.1	<i>Pol Poliprotein</i>
<i>Zea mays</i>	XP_008676375.1	Proteína Não Caracterizada
<i>Trichinella britovi</i>	KRY50278.1	Pol Poliprotein
<i>Thelohanellus kitauei</i>	KII61633.1	Pol Poliprotein
<i>Drosophila buzzatii</i>	CAB39733.1	<i>Reverse Transcriptase</i>
<i>Beta vulgaris subsp. vulgaris</i>	AFK13856.1	<i>Retrotransposon Ty3 Protein</i>
<i>Oryza sativa Japonica Group</i>	AAX95863.1	Proteína Não Caracterizada
<i>Vitis vinifera</i>	XP_010657629.1	Proteína Não Caracterizada
<i>Rhizopus microsporus</i>	CEG84463.1	Proteína Hipotética
<i>Trichinella patagoniensis</i>	KRY07437.1	<i>Pol Poliprotein</i>

<i>Salmo salar</i>	XP_013996221.1	Proteína Não Caracterizada
<i>Schistosoma japonicum</i>	CAX83694.1	Gag-pol polyprotein precursor
<i>Harpegnathos saltator</i>	XP_011139166.1	Proteína Não Caracterizada
<i>Caenorhabditis remanei</i>	XP_003101825.1	Proteína Hipotética
<i>Larimichthys crocea</i>	KKF20219.1	Gag-pol polyprotein precursor
<i>Austrofundulus limnaeus</i>	XP_013881994.1	Proteína Não Caracterizada
<i>Vigna vulgaris</i>	KOM49767.1	Proteína Hipotética
<i>Lingula ananatina</i>	XP_013416379.1	Proteína Não Caracterizada
<i>Danio rerio</i>	XP_009293363.1	Proteína Não Caracterizada

APÊNDICE J - Informações referentes às sequências utilizadas na filogenia da superfamília Bel.

REPBASE

Elemento	Organismo	Superfamília
BEL4-I_AP	<i>Acyrtosiphon pisum</i>	BEL
AGM1	<i>Anopheles gambiae</i>	BEL
BEL9-I_AG	<i>Anopheles gambiae</i>	BEL
BEL-1_BM-I	<i>Bombyx mori</i>	BEL
BEL5_Cis	<i>Ciona savignyi</i>	BEL
CATCH3I_DR	<i>Danio rerio</i>	BEL
BELPADRE	<i>Danio rerio</i>	BEL
BEL9-I_DR	<i>Danio rerio</i>	BEL
BEL4-I_DR	<i>Danio rerio</i>	BEL
BEL-8_DPu	<i>Daphnia pulex</i>	BEL

LITERATURA

Elemento	Organismo	Código Acesso NCBI
Bel	<i>Drosophila melanogaster</i>	733532
Cubel	<i>Culex pipiens quinquefasciatus</i>	170058572
Gabel	<i>Gasterosteus aculeatus</i>	83921752
Kobel	<i>Saccoglossus kowalevskii</i>	154426342
Mabel	<i>Brugia malayi</i>	170588120
Max	<i>Drosophila melanogaster</i>	21214751
Moose	<i>Anopheles gambiae</i>	3347855
Nabel	<i>Nasonia vitripennis</i>	156542142
Ninja	<i>Drosophila simulans</i>	1255594
Pao	<i>Bombyx mori</i>	972522
Purbel	<i>Strongylocentrotus purpuratus</i>	115954126
Roo	<i>Drosophila melanogaster</i>	27368142
Saci-6	<i>Schistosoma mansoni</i>	67625726
Sinbad	<i>Schistosoma mansoni</i>	46142833
Spirobel	<i>Trichinella spiralis</i>	109716041
Suzu	<i>Takifugu rubripes</i>	23194331
Tamy	<i>Antheraea mylitta</i>	33327732

Tas	<i>Ascaris lumbricoides</i>	453579
Tribel	<i>Tribolium castaneum</i>	86577607
Zebel	<i>Danio rerio</i>	38304119
Hydra3-1	<i>Hydra magnipapillata</i>	204801980
Telemac	<i>Drosophila virilis</i>	AF009439.1
BEL-Like	<i>Caenorhabditis briggsae</i>	AC084491.1
Kamikaze	<i>Bombyx mori</i>	AB042120.2
Yamato	<i>Bombyx mori</i>	AB042122.1
Saci-1	<i>Schistosoma mansoni</i>	BK004068.1

BLASTP *Leptopilina bouleardi*

CL449Contig1

Organismo	Código Acesso NCBI	Classificação no NCBI
<i>Solenopsis invicta</i>	751232603	Proteína Não Caracterizada
<i>Vollenhovia emeryi</i>	795018778	Proteína Não Caracterizada
<i>Megachile rotundata</i>	805812684	Proteína Não Caracterizada
<i>Monomorium pharaonis</i>	826416235	Proteína Não Caracterizada
<i>Acyrtosiphon pisum</i>	641676838	Proteína Não Caracterizada
<i>Wasmannia auropunctata</i>	780665970	Proteína Não Caracterizada
<i>Papilio machaon</i>	943951315	Proteína Não Caracterizada
<i>Amyelois transitella</i>	913306016	Proteína Não Caracterizada
<i>Plutella xylostella</i>	768412919	Proteína Não Caracterizada
<i>Tribolium castaneum</i>	EFA13455.1	Proteína Hipotética
<i>Microplitis demolitor</i>	939683130	Proteína Não Caracterizada
<i>Anopheles gambiae</i>	68697272	<i>Bel12_AG transposon polyprotein</i>
<i>Bombyx mori</i>	827559819	Proteína Não Caracterizada
<i>Stegodyphus mimosarum</i>	675368973	<i>Retrovirus Pol polyprotein</i>
<i>Musca domestica</i>	557785303	Proteína Não Caracterizada
<i>Papilio xuthus</i>	910317007	Proteína Não Caracterizada
<i>Vigna radiata var. radiata</i>	951069360	Proteína Não Caracterizada
<i>Diaphorina citri</i>	662216680	Proteína Não Caracterizada
<i>Lasius niger</i>	861626638	<i>Gag-pol polyprotein precursor</i>
<i>Trichogramma pretiosum</i>	936703174	Proteína Não Caracterizada
<i>Camponotus floridanus</i>	752866226	Proteína Não Caracterizada
<i>Drosophila melanogaster</i>	21214752	<i>Gag-pol polyprotein precursor</i>
<i>Drosophila ananassae</i>	145220605	<i>Putative Gag-pol polyprotein</i>
<i>Papilio polytes</i>	909549808	Proteína Não Caracterizada
<i>Bombus terrestris</i>	808146855	Proteína Não Caracterizada
<i>Diachasma alloeum</i>	970905213	Proteína Não Caracterizada
<i>Fopius arisanus</i>	755978057	Proteína Não Caracterizada
<i>Cerapachys biroi</i>	759051719	Proteína Não Caracterizada
<i>Amphimedon queenslandica</i>	340380172	Proteína Não Caracterizada
<i>Diuraphis noxia</i>	985395262	Proteína Não Caracterizada
<i>Polistes dominula</i>	972207485	Proteína Não Caracterizada
<i>Cimex lectularius</i>	939240731	Proteína Não Caracterizada

<i>Pyrus x bretschneideri</i>	694451718	Proteína Não Caracterizada
<i>Trichinella britovi</i>	954381582	Proteína Hipotética
<i>Trichinella nelsoni</i>	954208981	Proteína Hipotética
<i>Ceratitidis capitata</i>	807044362	Proteína Não Caracterizada
<i>Trichinella zimbabwensis</i>	954471660	Proteína Hipotética
<i>Trichinella pseudospiralis</i>	954517053	Proteína Hipotética
<i>Trichinella papuae</i>	954598236	<i>Integrase-like Gypsy</i>
<i>Trichinella patagoniensis</i>	954331330	Proteína Hipotética
<i>Trichinella nativa</i>	954525302	Proteína Hipotética
<i>Bactrocera cucurbitae</i>	751479437	Proteína Não Caracterizada
<i>Habropoda laboriosa</i>	915654958	Proteína Hipotética
<i>Crassostrea gigas</i>	762108606	Proteína Não Caracterizada
<i>Lingula anatina</i>	919063124	Proteína Não Caracterizada
<i>Saccoglossus kowalevskii</i>	585657935	Proteína Não Caracterizada
<i>Strongylocentrotus purpuratus</i>	390356455	Proteína Não Caracterizada
<i>Priapulul caudatus</i>	957852880	Proteína Não Caracterizada
<i>Hydra vulgaris</i>	828202504	Proteína Não Caracterizada
<i>Oreochromis niloticus</i>	542181818	Proteína Não Caracterizada
<i>Danio rerio</i>	688538986	Proteína Não Caracterizada

APÊNDICE K - Informações referentes às sequências utilizadas na filogenia da superfamília Copia.

REPBASE

Elemento	Organismo	Superfamília
Copia-23_NVi-I	<i>Nasonia vitripennis</i>	Copia

LITERATURA

Elemento	Organsimo	Código de Acesso NCBI
Sto-4	<i>Zea mays</i>	4234851
BARE-1	<i>Hordeum vulgare</i>	M94470
TarI	<i>Triticum aestivum</i>	AB008772
Rtsp-I	<i>Ipomoea batatas</i>	AB162659
Tto-1	<i>Nicotiana tabacum</i>	1167522
Tnt1	<i>Nicotiana tabacum</i>	20044
1731	<i>Drosophila melanogaster</i>	8700
ZebedeeII	<i>Aedes aegypti</i>	Z86116
Yokozuna	<i>Bombyx mori</i>	4239842
Mtanga	<i>Anopheles gambiae</i>	8029447
Salto	<i>Anopheles gambiae</i>	AAL55241
MosqCopia-Aa2	<i>Aedes aegypti</i>	AY009101
Copia-like_Osser	<i>Volvox carteri</i>	X69552
Zeco1	<i>Danio rerio</i>	154726955
Olco1	<i>Oryzias latipes</i>	DQ913000
Cico1	<i>Ciona intestinalis</i>	DQ913003

Opie-2	<i>Zea mays</i>	AAL31657
PREM-2	<i>Zea mays</i>	U41000
Copia-like	<i>Arabidopsis thaliana</i>	AF287471
PDR1	<i>Pisum sativum</i>	AJ243359
Tst1	<i>Solanum tuberosum</i>	X52387
Hopscotch	<i>Zea mays</i>	AAA57005
MZECOPIA	<i>Zea mays</i>	M94481
Bel	<i>Drosophila melanogaster</i>	733531
Max	<i>Drosophila melanogaster</i>	21214751
Ninja	<i>Drosophila simulans</i>	1255594
Pao	<i>Bombyx mori</i>	972521
BLASTP <i>Leptopilina bouleardi</i>	CL368Contig1	
Organismo	Código Acesso NCBI	Classificação NCBI
<i>Nasonia vitripennis</i>	XP_008205663.1	Proteína Não Caracterizada
<i>Trichogramma pretiosum</i>	XP_014237094.1	Proteína Não Caracterizada
<i>Chondrus crispus</i>	XP_005717320.1	<i>Gag-Pol Retroviral</i>
<i>Boechera divaricarpa</i>	CDF77009.1	Proteína hipotética de elemento transponível
<i>Glycine max</i>	BAG72096.1	Reverse Transcriptase
<i>Arabidopsis thaliana</i>	AAC26250.1	Reverse Transcriptase
<i>Vitis vinifera</i>	CAN67821.1	Proteína Hipotética
<i>Solanum lycopersicum</i>	ABO36622.1	<i>Copia LTR rider</i>
<i>Oryza australiensis</i>	BAA22288.1	<i>Polyprotein</i>
<i>Bryonia dioica</i>	ADJ18449.1	<i>Gag-Pol</i>
<i>Oryza sativa</i>	ABF96216.1	Proteína Putativa de retrotransposon Copia
<i>Picea glauca</i>	AAF65309.1	<i>Reverse Transcriptase</i>
<i>Lasius niger</i>	KMQ90088.1	<i>Gag-Pol</i>
<i>Papilio machaon</i>	KPJ16606.1	Pol relatada de Retrovírus
<i>Zea mays</i>	AAD12998.1	<i>Pol</i>
<i>Tribolium castaneum</i>	XP_008194613.1	Proteína Não Caracterizada
<i>Klebsormidium flaccidum</i>	GAQ93561.1	Proteína Hipotética
<i>Brassica rapa</i>	XP_009107606.1	Proteína Não Caracterizada
<i>Nicotiana tabacum</i>	BAA11674.1	Proteína Não Caracterizada
<i>Phyllostachys edulis</i>	ADB85429.1	Proteína Putativa de retrotransposon
<i>Sorghum bicolor</i>	XP_002489303.1	Proteína Hipotética
<i>Medicago truncatula</i>	ACL97385.1	<i>Gag-Pol</i>
<i>Lotus japonicus</i>	ACL97387.1	<i>Gag-Pol</i>
<i>Fusarium oxysporum</i>	EMT69244.1	Proteína Hipotética
<i>Prunus persica</i>	AHL77759.1	<i>Polyprotein</i>
<i>Phaseolus vulgaris</i>	AER13172.1	<i>Gag-Pol</i>
<i>Amyelopsis transitella</i>	XP_013190478.1	Pol relatada de transposon TNT1-94
<i>Trichinella pseudospiralis</i>	KRY86795.1	Pol relatada de transposon TNT1-94
<i>Agaricus bisporus</i>	ACZ04926.1	<i>Polyprotein</i>
<i>Anopheles gambiae</i>	AAL56548.1	<i>Pol</i>

<i>Trichinella murrelli</i>	KRX42725.1	Pol relatada de transposon TNT1-94
<i>Citrus sinensis</i>	AFB73912.1	<i>Polyprotein</i>
<i>Morus notabilis</i>	XP_010113352.1	Pol relatada de transposon TNT1-94
<i>Vigna angularis</i>	KOM27543.1	Proteína Hipotética
<i>Helianthus annuus</i>	CAX68207.1	<i>Copia-Like retrotransposon</i>
<i>Trichuris suis</i>	KFD60665.1	Proteína Hipotética
<i>Fragaria ananassa</i>	ACS74199.1	<i>Gag-Pol</i>
<i>Amanita muscaria</i>	KIL57786.1	Proteína Hipotética
<i>Trichinella zimbabwensis</i>	KRZ01021.1	Pol relatada de transposon TNT1-94
<i>Ipomoea batatas</i>	AAV88069.1	Proteína Hipotética
<i>Rhizophagus irregularis</i>	EXX50149.1	<i>Gag-Pol</i>
<i>Trichinella papuae</i>	KRZ69067.1	Pol relatada de transposon TNT1-94
<i>Rhizoctonia solani</i>	KEP45491.1	<i>Gag-Pol</i>
<i>Erythranthe guttata</i>	XP_012856116.1	Proteína Não Caracterizada
<i>Trichinella britovi</i>	KRY46240.1	Pol relatada de transposon TNT1-94
<i>Drosophila melanogaster</i>	CAD27357.1	Proteína Hipotética
<i>Trametes cinnabarina</i>	CDO75139.1	Proteína Hipotética
<i>Galerina marginata</i>	KDR64888.1	Proteína Hipotética
<i>Malus domestica</i>	XP_008387315.1	Proteína Não Caracterizada
<i>Sphaerobolus stellatus</i>	KIJ49713.1	Proteína Hipotética