



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”

Marcelo Roman Jurado Castro

Dinâmica evolutiva de elementos de transposição nas espécies
irmãs *Rhodnius robustus* e *R. prolixus*

São José do Rio Preto
2019

Marcelo Roman Jurado Castro

Dinâmica evolutiva de elementos de transposição nas espécies
irmãs *Rhodnius robustus* e *R. prolixus*

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biociências, junto ao Programa de Pós-Graduação em Biociências, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: OEA/GCUB nº 001/2016

Orientadora: Prof.^a Dr.^a Claudia Marcia
Aparecida Carareto

São José do Rio Preto
2019

C355d Castro, Marcelo Roman Jurado

Dinâmica evolutiva de elementos de transposição nas espécies irmãs *Rhodnius robustus* e *R. prolixus* / Marcelo Roman Jurado Castro. -- São José do Rio Preto, 2019

58 f. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientadora: Claudia Marcia Aparecida Carareto

1. Genética molecular. 2. Evolução molecular. 3. *Rhodnius robustus*. 4. *Rhodnius prolixus*. 5. Elementos de transposição. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Marcelo Roman Jurado Castro

Dinâmica evolutiva de elementos de transposição nas
espécies irmãs *Rhodnius robustus* e *R. prolixus*

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biociências, junto ao Programa de Pós-Graduação em Biociências, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: OEA/GCUB n° 001/2016

Orientadora: Prof.^a Dr.^a Claudia Marcia Aparecida Carareto

Comissão Examinadora

Prof.^a Dr.^a Claudia Marcia Aparecida Carareto
UNESP – Câmpus de São José do Rio Preto
Orientadora

Prof.^a Dr.^a Rita Daniela Fernandez Medina
Fundação Getúlio Vargas – Rio de Janeiro

Prof. Dr. Marcos Túlio de Oliveira
UNESP – Câmpus de Rio Claro

São José do Rio Preto
22 de fevereiro de 2019

*“Dedico este trabalho a minha mãe,
mulher guerreira e perseverante”*

Agradecimentos

À minha orientadora, Prof^a Dr^a Claudia Marcia Aparecida Carareto, que me conduziu nesta trajetória para a realização deste trabalho, muito obrigado pela oportunidade, ensinamento e acima de tudo paciência.

Ao, Dr Clément Goubert da Universidade Cornell, pela colaboração nos análises, que foram de enorme importância para a execução deste trabalho.

Aos colaboradores, Dra. Cristina Vieira da Universidade de Lyon e ao Dr. Fernando A Monteiro do Instituto Oswaldo Cruz, Fiocruz, que fazem parte no desenvolvimento do presente trabalho.

Aos meus amigos de laboratório Cecília, Izabella, Camila, Joice, Luis Gustavo, Edoardo, Rubens, Lucas e William um excelente equipe.

À UNESP e o Programa de Pós-Graduação em Biociências por ter me acolhido como aluno do curso de mestrado.

À Organização dos Estados Americanos (OEA) junto ao Grupo Coimbra de Universidades Brasileiras (GCUB) pela concessão da bolsa de mestrado, sob o edital OEA/GCUB nº 001/2016.

Aos meus amigos-parceiros, irmãos em Cristo: Edmur, Alison, Kairo, Iuri, Caio e Diogo obrigado por sua companhia e atenção, por serem minha família enquanto estive longe.

Aos meus queridos irmãos Jhoselin e Brandon por todo o apoio, carinho e compreensão. Vocês foram parte da minha inspiração.

Aos meus queridos pais, Flora e Roman por terem me proporcionado o que eu necessitava para a vida: amor, educação e caráter. Vocês são os grandes responsáveis pela conclusão desta etapa, sou-lhes eternamente grato.

Acima de tudo a Deus o único que sempre permanece....

RESUMO

Os elementos de transposição (TEs) são sequências repetitivas amplamente distribuídas nos genomas que representam importantes fontes geradoras de variabilidade genética. As características genômicas derivadas de sua capacidade de transposição e de aumento do número de cópias são específicas para cada linhagem. Um fenômeno derivado dessa habilidade é o incremento repentino do número de cópias, denominado explosões de transposição, descrito em várias espécies para diversas famílias de TEs, dentre elas, a família *Mariner* (Transposon de DNA, Classe II) em *Rhodnius prolixus*. Foi sugerido que explosões de transposição da família *Mariner* contribuíram para o processo de especiação de *R. prolixus*. Esta peculiaridade motivou o estudo da composição do mobiloma em *R. robustus*, espécie irmã de *R. prolixus*, juntamente a reanotação dos TEs de *R. prolixus*, para comparação do mobiloma das duas espécies. Para isso, foram realizados sequenciamentos NGS (Next Generation Sequencing) dos genomas de dois tipos genéticos de *R. robustus* (II e III), montagem e anotação dos TEs utilizando a ferramenta *dnaPipeTE*. Da mesma forma, foram anotados os TEs de *R. prolixus* a partir das *reads* brutas do seu genoma montado. Os resultados mostraram alta semelhança na frequência de TEs nos três genomas (*R. prolixus*: 19%, *R. robustus* II: 23,5%, *R. robustus* III: 21,5%), constituídos principalmente de TEs de Classe II (~13,5%) da superfamília *Tc1-Mariner* (~5,6%) e de da Ordem LINE, da Classe I (~5,1%). A comparação dos três mobilomas revelou três características importantes para se entender a dinâmica evolutiva destes TEs: (i) compartilhamento de famílias que ocorrem tanto em alta como em baixa proporção entre genomas; (ii) compartilhamento de famílias que ocorrem em proporções dissimilares entre genomas, alta em um e baixa no outro, e (iii) não compartilhamento de determinadas famílias entre genomas. Identificou-se ainda sinais de explosões de transposição recentes da família *Mariner* em *R. robustus* II e III, indicando que esse fenômeno não foi exclusivo de *R. prolixus*, estando, portanto, não associado diretamente à especiação das duas espécies irmãs.

Palavras-chave: Elementos de transposição. Explosões de transposição. Elementos *Mariner*. Especiação. *Rhodnius robustus*. *Rhodnius prolixus*

ABSTRACT

Transposable elements (TEs) are widely distributed repetitive sequences in the genomes that represent important sources of genetic variability. The genomic characteristics derived from their capacity of transposition and increasing in copy number are specific to each lineage. A phenomenon derived from this ability is the sudden increase of one or several families of TEs, called bursts of transposition, described in several species for several families of TEs, among them the *Mariner* family (DNA Transposons, Class II) in *Rhodnius prolixus*. It has been suggested that transposition bursts of the *Mariner* family contributed to the speciation process of *R. prolixus*. This peculiarity motivated the study of the mobilome composition of *R. robustus*, together with its reannotation in *R. prolixus*, for comparison of the mobilome of the two species. In order to do this, we performed NGS (Next Generation Sequencing) sequencing of the genomes of two genetic types of *R. robustus* (II and III) and the assembly and annotation of the TEs using the pipeline *dnaPipeTE*. Likewise, the TEs of *R. prolixus* were annotated from the raw reads of its assembled genome. The results showed high similarity in the frequency of TEs in the three genomes (*R. prolixus*: 19%, *R. robustus* II: 23.5%, *R. robustus* III: 21.5%), consisting mainly of TEs from Class II (~13.5%) of the superfamily Tc1-Mariner (~5.6%) and of the LINE Order, Class I (~5.1%). The comparison of the three mobilomes revealed three important characteristics to understand the evolutionary dynamics of these TEs: (i) sharing of families that occur both in high and low proportion between genomes; (ii) sharing of families occurring in dissimilar proportions between genomes, high in one and low in the other, and (iii) non-sharing of a family between genomes. There are also signs of recent burst of transposition of the *Mariner* family in *R. robustus* II and III, indicating that this phenomenon was not exclusive to *R. prolixus*, therefore not directly associated with the speciation of the two sister species.

Keywords: *Transposable elements. Burst of transposition. Mariner elements. Speciation. Rhodnius robustus. Rhodnius prolixus.*

SUMÁRIO

1. INTRODUÇÃO	9
1.1 Classificação dos Elementos de Transposição	11
1.2. Elementos de transposição em insetos.....	13
1.3 Complexo <i>Rhodnius prolixus</i> - <i>R. robustus</i> : origem, diversificação e elementos de transposição	14
2. OBJETIVOS.....	18
2.1.Objetivos Específicos	19
3. MATERIAL E MÉTODOS	20
3.1. Sequenciamento dos genomas	21
3.2. Método <i>dnaPipeTE</i>	21
3.3. Pré-processamento das <i>reads</i> de <i>R. robustus</i>	23
3.4. Análises genômicas de <i>R. robustus</i> II, <i>R. robustus</i> III e das <i>reads</i> brutas de <i>R. prolixus</i>	23
3.5. Recuperação e quantificação das superfamílias e famílias de TEs.....	24
3.6. Comparação do mobiloma de <i>R. prolixus</i> , <i>R. robustus</i> II e <i>R. robustus</i> III.....	26
3.7. Investigação dos TEs no genoma montado de <i>R. prolixus</i> utilizando <i>dnaPipeTE</i>	26
4. RESULTADOS	27
4.1. Conteúdo repetitivo em <i>R. prolixus</i> , <i>R. robustus</i> II e <i>R. robustus</i> III	28
4.2. Comparação da proporção e dinâmica dos TEs entre <i>R. prolixus</i> e <i>R. robustus</i> ...	29
4.3. Comparação das proporções relativas de famílias TE compartilhadas entre <i>R. prolixus</i> e <i>R. robustus</i>	32
5. DISCUSSÃO	35
6. CONCLUSÕES	42
7. REFERÊNCIAS	44
8. ANEXOS	53
8.1. Tabelas Suplementares	54
8.2. Figuras Suplementares	57

1. INTRODUÇÃO

Os elementos de transposição (TEs) são sequências de DNA medianamente repetitivas capazes de se movimentarem de um local para outro no genoma. A porção do genoma composta por TEs, denominada mobiloma após Kooning et al. (2008), é bastante variável em eucariotos, podendo ser bastante alto, como em plantas, por exemplo, entre 60-85% nas gramíneas (PATERSON et al., 2009; SCHNABLE et al., 2009), embora menos abundantes que em plantas compõem entre 3-20% do genoma dos fungos (DABOUSSY; CAPY, 2003) e entre 3-45% do genoma dos metazoários (HUA VAN et al., 2005). A habilidade dos TEs se replicarem mais rapidamente do que o genoma hospedeiro e causarem efeitos deletérios quando se inserem em regiões codificantes, interrompem a fase de leitura aberta ou alteram os níveis de expressão gênica, justifica serem considerados como parasitas genômicos (BROOKFIELD 2005; GOODIER; KAZAZIAN, 2008; KIDWELL; LISCH 2001; FESCHOTTE; PRITHAM 2007). As inserções de TEs podem comprometer a integridade genômica também em regiões não codificantes, atuando como centro para a recombinação homóloga desigual, resultando em duplicações, deleções e translocações de segmentos genômicos, codificantes ou não (BATZER; DEININGER, 2002).

Apesar dos efeitos deletérios que acarretam em nível genômico e organizmico, os TEs exercem um papel fundamental na evolução dos seres vivos. Sua prevalência e mobilidade pode ocasionar alterações estruturais, como translocações, duplicações segmentais e deleções capazes de induzir profundas alterações genômicas, levando à sua contração ou expansão, além do estabelecimento de novas relações de ligação entre genes (KIDWELL; LISCH, 2001). Tem-se reportado, também, que os TEs desempenham papel importante na regulação de diversos processos celulares, na diversificação das famílias gênicas, sendo capazes de promover a transdução e amplificação de fragmentos de genes do genoma hospedeiro (VAN DE LAGEMAAT et al., 2003), bem como, de influenciar na expressão e função gênica, ampliando a variabilidade do repertório transcricional e proteico (CARARETO et al., 2013; LOPES et al., 2013). Devido à essas habilidades, os TEs são considerados poderosos agentes evolutivos. Assim, o controle da expressão e transposição dos TEs é fundamental para estabelecer um equilíbrio entre efeitos positivos e negativos, garantindo a viabilidade das funções do genoma nos organismos hospedeiros.

A regulação da atividade dos TEs ocorre por meio de mecanismos do silenciamento pós-transcricional, utilizando RNAs de interferência, e silenciamento transcricional, por meio

da metilação do DNA e modificação de histonas (GRANZOTTO; CRUZ, 2015). Tem-se observado que o estresse fisiológico, induzido por alterações climáticas ou a invasão de novos habitats, como também por agentes biológicos (parasitas) interrompe essa regulação.

A abundância dos TEs e suas características tais como distribuição genômica e rearranjos cromossômicos que podem acarretar, são frequentemente linhagem-específicas, e essa especificidade sugere que estes contribuíram para o processo de especiação, seja como causa ou efeito (BOHNE et al., 2008; MARINO-RAMIREZ et al., 2005). Um fenômeno relacionado com a multiplicação ou incremento repentino de um ou vários TEs, que pode provocar uma reconstrução genômica radical é o chamado “*burst of transposition*” (explosões de transposição). Esse fenômeno, que acontece em períodos curtos no ciclo evolutivo dos TEs, rompendo a estase populacional (HORNS; PETIT; HOOD, 2017), tem sido associado por alguns autores a grandes eventos evolutivos e gênese de novos grupos filogenéticos (DE BOER et al., 2007; PACE; FESCHOTTE, 2007; PRITHAM; FESCHOTTE, 2007; RAY et al., 2008; VOLFF et al., 2001; VOLFF; KORTING; SCHARTL, 2000). Diversos estudos mostram que as explosões de transposição representam uma reação habitual do genoma a mudanças rápidas em ambientes externos; contudo, estresse genômico, tais como a domesticação, a poliploidia, a hibridização interespecífica e intergenérica, ou mesmo mudanças nos sistemas de acasalamento são também associados às explosões de transposição (revisão em BELYAYEV, 2014).

Nas populações naturais sob a influência de uma ecologia incomum, as explosões de transposição são de especial interesse porque este fenômeno pode estar subjacente a eventos micro e macro - evolutivos que levam, em última instância, à geração de diversidade biológica. Uma questão importante, então, a ser respondida é se existem relações causais entre explosões de transposição e especiação. Uma das formas de se responder a tal questão consiste em anotar os elementos de transposição em pares de espécies que passaram recentemente por tais eventos. Neste trabalho, nos propomos a investigar o mobiloma (conjunto de TEs do genoma) de duas espécies irmãs - *Rhodinus prolixus* e *R. robustus* - que atendem a essas condições, pois resultam de um processo de especiação relativamente recente (1,4 milhões de anos, segundo Monteiro et al. 2003) e que atualmente têm uma ecologia bastante distinta.

1.1. Classificação dos elementos de transposição

A anotação do mobiloma de uma espécie requer utilizar um sistema de classificação destas sequências móveis. Dentre os diferentes sistemas de classificação já propostos, desde

Finnegan (1988), a proposta mais aceita, e adotada atualmente é a de Wicker et al. (2007), que utiliza um sistema hierárquico que inclui os níveis classe, subclasse, ordem, superfamília, família e subfamília. O primeiro nível divide os TEs em duas classes, como na classificação original de Finnegan, de acordo com o mecanismo de transposição utilizado, com (*Classe I*) ou sem (*Classe II*) a participação de uma molécula de RNA intermediária no processo de transposição, enquanto que o nível subclasse é usado para distinguir apenas elementos da *Classe II*, com base no tipo de clivagem do TE a partir de sua posição original, com clivagem da cadeia dupla (*Sub - Classe I*) ou não (*Sub - Classe II*). O nível ordem classifica os TEs pelo mecanismo de inserção, organização geral e enzimologia. As superfamílias dentro de uma ordem compartilham estratégias de replicação, mas são distinguidas pela estrutura dos domínios proteicos ou não codificantes; as famílias são definidas pela conservação de suas sequências e as subfamílias por agrupamento filogenético das sequências de uma família (WICKER et al., 2007). O sistema de classificação de Wicker et al. (2007) é descrito a seguir e também ilustrado na Figura 1.

Os elementos de *Classe I* (Retrotransposons) se replicam usando um transcrito de RNA como intermediário para transposição, utilizando a enzima transcriptase reversa. Nesta classe incluem-se cinco ordens, dentre elas uma ordem é denominada de Retrotransposons com *LTRs*, pois apresentam Repetições Terminais Longas (*LTR*: do Inglês, *Long Terminal Repeat*) e quatro incluem os retrotransposons sem *LTRs* (do Inglês: *non-LTRs*), denominadas como *LINE* (*Long Interspersed Nuclear Elements*), *SINE* (*Short Interspersed Nuclear Elements*), *PLE* (*Penelope-like*) e *DIRS* (*Dictyostelium intermediate repeat sequence*). Essas ordens incluem um número total de 17 superfamílias (***LTR***: *Copia*, *Gypsy*, *Bel-Pao*, *Retrovírus*, *ERV*; ***LINE***: *R2*, *RTE*, *Jockey*, *L1*, *I*; ***SINE***: *tRNA*, *7SL* e *5S*; ***PLE***: *Penelope*; ***DIRS***: *DIRS*, *Ngaro*, *VIPER*).

Os elementos da *Classe II* (Transposons de DNA) não requerem o passo da transcrição reversa para se integrarem ao genoma, e são divididos em duas subclasses: *Subclasse I* que realiza a transposição via mecanismo corta-e-cola, e *subclasse II* que realiza a transposição pelo mecanismo cópia-e-cola. Os elementos da *Subclasse I* codificam para uma enzima transposase que excisa o TE de seu sítio hospedeiro e o integra em um novo local no genoma, e os da *Subclasse II* realizam processo de transposição sem clivagem de cadeia dupla, envolvendo o deslocamento de apenas uma fita de DNA. A *Subclasse I* é composta pelas ordens *TIR* e *Crypton* e a *subclasse II* é constituída pelas ordens *Helitron* e *Maverick*. Essas subclasses incluem 12

superfamílias (**TIR**: *Tc1-Mariner*, *hAT*, *Mutator*, *Merlin*, *Transib*, *P*, *PiggyBac*, *PIF-Harbinger*, *CACTA*; **Crypton**: *Crypton*; **Helitron**: *Helitron* e **Maverick**: *Maverick*).

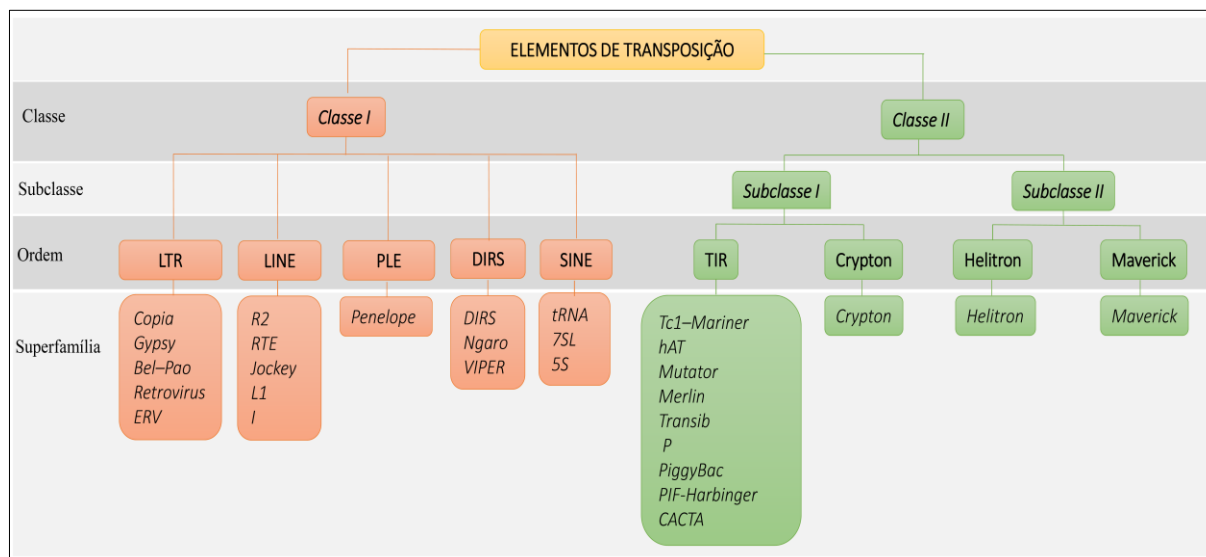


Figura 1. Sistema proposto por Wicker et al. (2007) para classificação dos elementos de transposição. LINE: long interspersed nuclear element; LTR: long terminal repeat; PLE: Penelope-like element; SINE: short interspersed nuclear element; TIR: terminal inverted repeat.

Os TEs podem ainda ser classificados em autônomos e não-autônomos. Aqueles que podem sintetizar as proteínas necessárias para sua mobilização são chamados de elementos autônomos, e aqueles elementos que não tem suas sequências codificadoras de proteínas necessárias para a transposição íntegras são chamados de não-autônomos, como os *SINEs* e *MITEs*. Embora sem os domínios codificadores de proteínas funcionais, esses elementos podem se mobilizar no genoma se as proteínas necessárias para a sua transposição forem fornecidas por outros elementos do genoma a eles relacionados (FESCHOTTE; JIANG; WESSLER, 2002).

1.2. Elementos de transposição em insetos

Existe uma ampla variação do conteúdo de TEs nos genomas de insetos, desde apenas 1% no mosquito não voador *Belgica antarctica* (KELLEY et al., 2014) a 65% no gafanhoto *Locusta migratoria* (WANG et al., 2014). De modo semelhante, a abundância das famílias dos TEs nos genomas também é amplamente variável. Os elementos de *Classe I* parecem ser os mais abundantes, embora haja variações na presença de cada ordem de TEs em diferentes

insetos. Por exemplo, os principais componentes do mobiloma de *Drosophila melanogaster* pertence à ordem *LTR*, seguindo-se os *LINEs* e em menor fração os transposons de DNA (SESSEGOLO et al., 2016). Em *Anopheles gambiae*, a ordem *LTR* é a mais prevalente, seguida dos *SINEs* e os *MITEs* (HOLT et al., 2002) (revisão em MAUMUS et al., 2015). No entanto, transposons *DNA* e *LINEs* são os elementos mais abundantes no pulgão-da-ervilha (*Acyrtosiphon pisum*) e da lagarta-do-cartucho (*Spodoptera frugiperda*) (revisão em MAUMUS et al., 2015). A abundância e distribuição dos TEs nos genomas de insetos sugere que estes são atores-chave na evolução de tais espécies. Portanto, sua descoberta é um passo crucial para uma compreensão mais profunda na biologia de insetos (MAUMUS et al., 2015).

A única espécie de triatomíneo com genoma sequenciado e mobiloma anotado é *R. prolixus* (MESQUITA et al., 2015). A análise genômica dessa espécie propiciou aumentar a compreensão sobre a genômica funcional, expressão gênica e também a identificação dos mecanismos de adaptação que podem envolver variações estruturais, incluindo a duplicação de genes ou a composição do mobiloma nessa espécie. O mobiloma de *R. prolixus* é constituído por aproximadamente 1.400 famílias de TEs, que correspondem uma pequena fração (5,6 %) do genoma de 702 Mb do inseto (MESQUITA et al., 2015). Esta porcentagem é muito menor em relação àquelas identificadas em outros insetos hematófagos, como *Aedes aegypti* (~ 50%) ou *Anopheles gambiae* (~ 20%), por exemplo, mas semelhante a outros anofelinos (2-11%) (NEAFSEY et al., 2015).

1.3. Complexo *Rhodnius prolixus* - *R. robustus*: origem, diversificação e elementos de transposição

A doença de Chagas, uma das doenças parasitárias relevantes na América Latina, é uma zoonose causada pelo protozoário *Trypanosoma cruzi*. Estima-se que entre seis a oito milhões de pessoas estejam infectadas em todo o mundo, pelo menos um milhão de pessoas no Brasil, sendo assim a doença de Chagas considerada uma das zoonoses mais desatendidas no mundo (https://www.paho.org/hq/index.php?option=com_topics&view=article&id=10&Itemid=40743&lang=en). Essa doença é transmitida por insetos vetores hematófagos em associação a seus diferentes ecótopos: silvestres (palmeiras, ninhos de aves, tocas de mamíferos), peridomicílio e no interior de casas (LENT; WYGODZINSKY, 1979). Os gêneros de maior prevalência epidemiológica pertencem a *Triatoma* e *Rhodnius* (subfamília Triatominae, família Reduviidae),

sendo este último o principal gênero transmissor da doença de Chagas na região amazônica (PAVAN et al., 2013; PAVAN; MONTEIRO, 2007).

O gênero *Rhodnius*, provavelmente surgiu nas florestas pan-amazônicas (região amazônica e região do Rio Orinoco) em associação com palmeiras e com a fauna habitante deste ecótopo (PATTERSON; GAUNT, 2010), está dividido em três complexos de espécies (DUJARDIN et al., 1998): (i) complexo de *R. pallescens* (*R. pallescens* e *R. ecuadoriensis*), (ii) complexo de *R. pictipes* (*R. pictipes*, *R. stali* e *R. brethesi*) e (iii) complexo de *R. robustus* (*R. prolixus*, *R. robustus*, *R. neglectus*, *R. nasutus*, *R. domesticus* e *R. neivai*). Foi proposto que o complexo *R. robustus* evoluiu por radiação adaptativa de um ancestral comum na região do Orinoco-Amazonas, representado atualmente por *R. pictipes* (SCHOFIELD, CJ e DUJARDIN J., 1999). Esse complexo diferenciou-se em duas regiões: ao sul, entre os cerrados do Brasil, originando as espécies *R. neglectus* e *R. nasutus*, e ao norte, nas planícies da Venezuela, a partir de um ancestral comum de *R. robustus* dando origem a *R. prolixus* que, posteriormente, adaptou-se a *habitat* doméstico. Essas duas espécies, *R. robustus* e *R. prolixus*, são estreitamente relacionadas, membros de um complexo de “espécies crípticas”, cuja validade taxonômica foi largamente questionada, em função de sua semelhança morfológica e isoenzimática (HARRY, 1993;1994). Diversos estudos revalidaram o *status* taxonômico de ambas espécies em face de diferenças quanto à adaptações ecológicas, marcadores genéticos (ABAD-FRANCH et al., 2013; MONTEIRO et al., 2003; PAVAN et al., 2013), morfometria tradicional e geométrica (VILLEGAS; FELICIANGELI; DUJARDIN, 2002) e padrões de atividade locomotora (PAVAN et al., 2016).

Um estudo de Monteiro e colaboradores (2003), utilizando uma sequência do gene mitocondrial citocromo b (*cyt b*), mostrou que não apenas *R. robustus* é bastante divergente em relação à *R. prolixus*, como apresenta uma notável heterogeneidade genética dentro da espécie, com alta divergência entre as populações geográficas representando um complexo parafilético de quatro linhagens crípticas (*R. robustus* I, II, III e IV), o que também foi confirmado por Pavan e Monteiro (2007) e Pavan et al. (2013). *Rhodnius robustus* I apresenta distribuição geográfica restrita em ecótopos silvestres, sendo encontrado apenas na bacia do Rio Orinoco, na Venezuela (PAVAN et al. 2013), em simpatria com *R. prolixus*. Interessantemente *R. robustus* I, está filogeneticamente mais proximamente relacionado a *R. prolixus* do que com as linhagens amazônicas (*R. robustus* II, III e IV), o que caracteriza a parafilia da espécie. Os tipos genéticos II, III e IV de *R. robustus* compreendem um grupo monofilético de linhagens

alopátricas distribuídas na região amazônica. *Rhodnius robustus* II é encontrado ao norte do estado de Rondônia e sul do estado do Amazonas (Brasil), norte do Equador e nordeste da Bolívia; *R. robustus* III se encontra no nordeste do Amazonas, norte do estado de Tocantins e este do estado do Pará (Brasil); e *R. robustus* IV é encontrado ao norte do Pará, norte do Amazonas, estado de Roraima (Brasil), norte da Venezuela (em simpatria com *R. prolixus*) e na Guiana Francesa (PAVAN et al. 2013). Embora *R. prolixus* esteja bem adaptado a domicílios humanos em muitos países de América Latina, populações selvagens dessa espécie estão presentes apenas na Venezuela e Colômbia.

O mobiloma de *R. prolixus*, embora relativamente pequeno comparado ao de outros insetos (5,6%, c.f. MESQUITA et al. 2015), apresenta uma característica interessante. Dentre os TEs de *Classe I*, na ordem *LTR*, que representam a menor fração do mobiloma (<1% no genoma) encontram-se apenas alguns representantes de três superfamílias (*Gypsy*, *Copia*, *BEL-Pao*). Ao contrário, os elementos sem *LTRs* são bastante diversos, consistindo de 10 superfamílias diferentes (*Jockey*, *Rose*, *LoA*, *Felyant*, *Totopi*, *CR-1*, *RTE*, *I*, *R1* e *Ingi*), compreendendo 1,4% no genoma. Mas, a característica notável do genoma dessa espécie, comparada aos outros insetos, é a preponderância dos elementos de *Classe II*, representando 75% de todos os TEs identificados, principalmente representada pelas superfamílias *Tc1-Mariner* e *MITEs* (aproximadamente 3% do conteúdo total no genoma). Muitas dessas sequências *Mariner* são *full-length* (sequências completas), com a sequência codificadora da transposase completa, sugerindo que estão ou estiveram ativas até recentemente.

Fernandez-Medina et al. (2016), que também participaram da anotação do mobiloma de *R. prolixus*, exploraram os membros da superfamília *Tc1-Mariner* e mostraram que a maior parte das sequências da família *Mariner* correspondem às sete subfamílias bem estabelecidas em insetos (*Irritans/Himar*, *Mellifera*, *Cecropia*, *Mauritiana*, *Vertumnana* e *Rosa*); contudo, descreveram duas subfamílias, que denominaram como *Prolixus1* e *Prolixus2*, que exibiam um elevado número de cópias. Em *Prolixus1*, os autores identificaram três clados: *Clado1* (com mais de 8.000 cópias) relacionado à subfamília *Irritans/Himar*, e *Clados II* e *III*, com 846 e 550 cópias genômicas, respectivamente. Os autores mostraram que esses clados sofreram expansões genômicas intensas, características de explosões de transposição. Os tempos relacionados a essas expansões foram datados: para o *Clado I*, a sequência ancestral comum mais recente (MRCA) foi datada em aproximadamente 1,32 milhões de anos (maa) e para os *Clados II* e *III* em aproximadamente 0,67 maa. O período mais antigo coincide àquele proposto para a

divergência entre *R. prolixus* e sua espécie irmã *R. robustus* (~ 1,4 maa) quando o antepassado comum *prolixus/robustus* vivia em diferentes habitats nas florestas do Rio Orinoco (MONTEIRO et al., 2003).

Diversas características de *R. prolixus*, tais como baixa diversidade genética (DUJARDIN et al., 1998; HARRY; GALINDEZ; CARIOU, 1992; LOPEZ; MORENO, 1995; MONTEIRO et al., 2003), possível resultado de gargalo da garrafa, somada à explosão de transposição observada em elementos *Mariner* (FERNANDEZ-MEDINA et al., 2016), são concordantes com um cenário de especiação fundamentado na “teoria dos refúgios” (MONTEIRO et al., 2003) e levaram a Fernandez-Medina et al. (2016) a formularem a hipótese de que as explosões de transposição identificadas nas subfamílias *Prolixus I* e *Prolixus II* poderiam ter se iniciado em decorrência de características ecológicas e populacionais da espécie durante sua origem. Os autores sugeriram ainda que a continuidade da expansão desses TEs poderia estar associada à crescente adoção de habitats domésticos, mais recentemente pela espécie, com a colonização das Américas pelo homem (MONTEIRO et al., 2003).

Contudo, uma hipótese alternativa à acima formulada seria que a explosões de transposição, teriam ocorrido antes da divergência *R. prolixus/R. robustus* e seriam compartilhadas pelas duas espécies. Para investigar essa hipótese propusemos o presente estudo, realizando a anotação do mobiloma de dois dos quatro tipos genéticos *R. robustus* (*R. robustus II* e *R. robustus III*), cujas amostras de DNA foram disponibilizadas para análises, buscando evidências da ocorrência ou não das explosões de transposição da família *Mariner*, na espécie *R. robustus*.

2. OBJETIVOS

O objetivo geral deste estudo foi caracterizar o mobiloma e identificar a ocorrência de explosões de transposição da família *Mariner* em *R. robustus* visando elucidar se essas explosões foram compartilhadas com sua espécie irmã *R. prolixus*.

2.1. Objetivos específicos

1. Anotar os TEs nos genomas sequenciados dos tipos genéticos II e III de *R. robustus* provenientes da Amazônia brasileira.
2. Caracterizar, classificar, quantificar, comparar e analisar evolutivamente os TEs anotados no genoma de *R. robustus* II e III;
3. Estimar o número de cópias dos elementos da família *Mariner*;
4. Reanotar os TEs de *R. prolixus* a partir das *reads* brutas do genoma sequenciado, seguindo a mesma metodologia que a utilizada para *R. robustus* II – III, e relacionar os presentes resultados com os obtidos na anotação inicial no genoma de *R. prolixus*.

3. MATERIAL E MÉTODOS

3.1. Sequenciamento dos genomas

A partir de amostras de DNA de linhagens de dois tipos genéticos de *R. robustus* cedidas por Fernando Araújo Monteiro (Laboratório de Epidemiologia e Sistemática Molecular, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, RJ), foram feitos sequenciamentos utilizando a tecnologia Illumina NGS de baixa cobertura *TruSeq Nano pair end*, pela plataforma GeT-PlaGe, Génomole Toulouse/Midi-pyrénées (França). Foram sequenciados os genomas dos tipos genéticos *R. robustus* – tipo II, a partir da mistura de DNA extraído de 3 indivíduos (3µg) da mesma população (coletados em Ouro Preto do Oeste, Rondônia, 10°43'S 62°15'W), e tipo III, a partir de mistura de DNA de dois indivíduos (3,7 µg) da mesma população (coleta em Rondon do Pará, Pará, 04°54'S 48°20'W). Também foi utilizado o conjunto das *reads* brutas do genoma público de *R. prolixus* (MESQUITA et al., 2015). Os genomas foram armazenados no servidor do laboratório de Evolução Molecular e também foram enviados os acessos para o colaborador Clément Goubert (*Department of Molecular Biology and Genetics, Cornell University, Ithaca, USA*), um dos autores da metodologia *dnaPipeTE*, que realizou a análise de qualidade das “*reads*” sequenciadas, e aplicou o *pipeline dnaPipeTE*. Os “*outputs*” produzidos foram disponibilizados, sendo estes o ponto de partida para as análises aqui apresentadas.

3.2. Método *dnaPipeTE*

dnaPipeTE consiste em um método *De Novo Assembly e Pipeline* de anotação de TEs (<https://lbbe.univ-lyon1.fr/-dnaPipeTE-.html>, GOUBERT et al., 2015) totalmente automatizado para montar e quantificar repetições de *reads* (leituras) genômicas. Para a montagem das sequências, o *dnaPipeTE* utiliza o programa *Trinity* (GRABHERR et al., 2011), originalmente projetado para utilização de dados produzidos por RNA-seq. Adicionado ao *Trinity*, o *pipeline* utiliza dois bancos de dados de TEs, o *Repbase* (versão 2014-01-31 downloaded de <http://www.girinst.org/>) e a biblioteca *TEFam* (<http://tefam.biochem.vt.edu/tefam/index.php>), e a ferramenta *RepeatMasker* (RM, versão open-4.0.5, <http://www.repeatmasker.org>).

Anotação das sequências repetitivas

O primeiro passo do *pipeline* do *dnaPipeTE* é a realização de amostragens das *reads* sequenciadas, utilizadas no formato *FASTQ*. Posteriormente, ocorre a montagem das *reads* de cada amostra, utilizando iterações do *Trinity*, formando *contigs* que representam as sequências repetitivas. Em um terceiro momento, os *contigs* montados são anotados utilizando a ferramenta *RepeatMasker* (Figura 2).

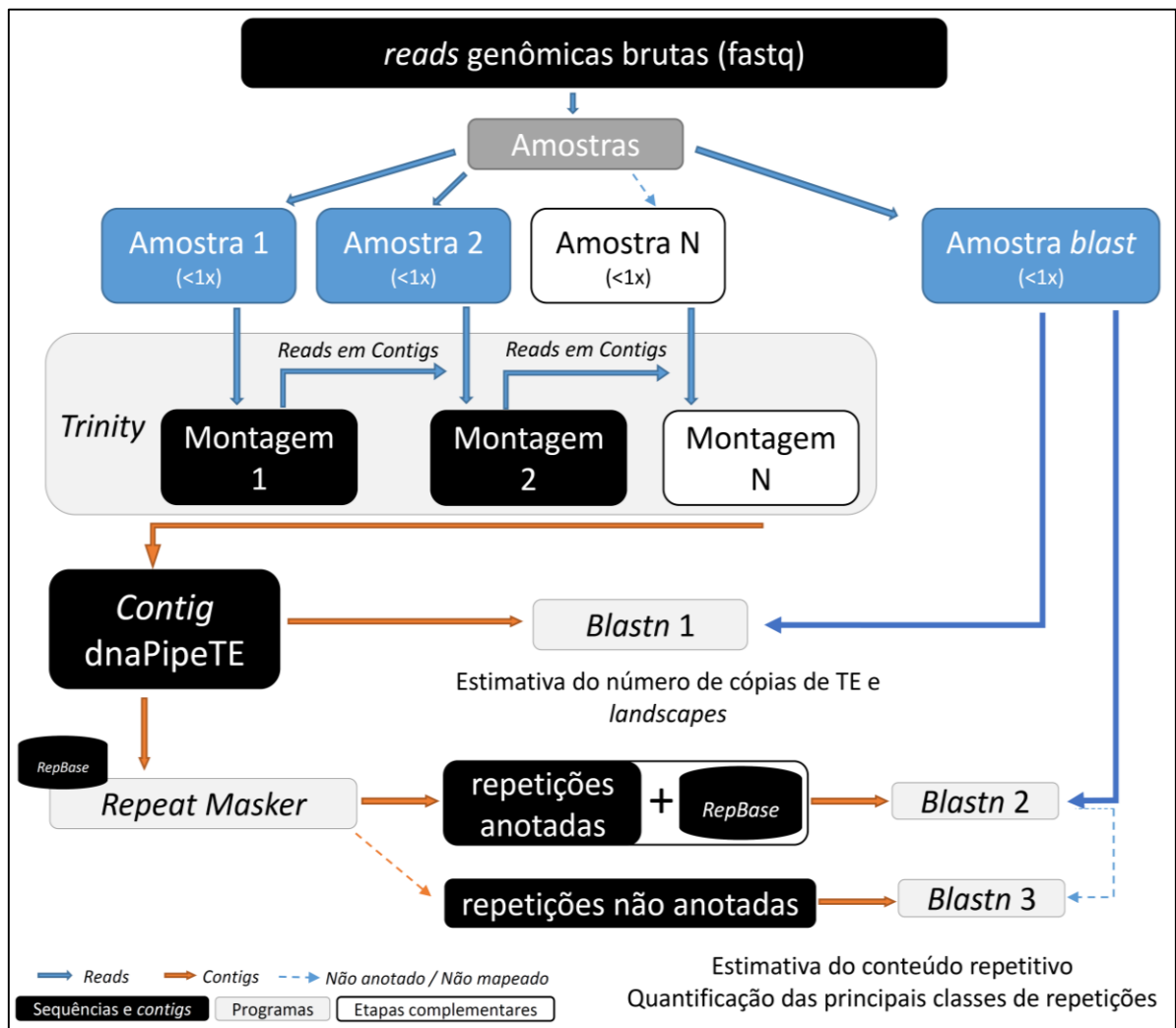


Figura 2. Pipeline do método *dnaPipeTE*. 1- Reads genômicas no formato *FASTQ* são amostradas. 2 - Montagem de repetições é realizada usando *Trinity*. 3- Para cada iteração, as *reads* previamente montadas são adicionadas à próxima amostra para melhorar a montagem das repetições. 4- Os *contigs* montados são anotados usando *RM*. 5 – Todos os *contigs* são comparados pelo *BLAST* para estimar a abundância de cada sequência obtida. 6 - Um segundo *BLAST* é realizado, e a mesma amostra é sucessivamente comparada contra os *contigs* anotados do *Repbase*. Fonte: Goubert et al. (2015).

A anotação dos *contigs* segue a regra 80-80-80 proposta por Wicker et al. (2007), em que as *queries* com 80% de cobertura em relação às sequências do banco de dados são armazenadas como “*full-length*” e as *queries* com *hits* inferiores a 80% são armazenadas como “*partial*” (sequências incompletas). Dos *contigs* não anotados pelo *Repeat Masker* somente a informação das ordens são retidas, de acordo com a classificação de Wicker et al. (2007): *Long Terminal Repeat (LTR)*, *Long Interspersed Element (LINE)*, *Short Interspersed Element (SINE)*, *DNA*, *Miniature Inverted-repeat Transposable Elements (MITEs)*, *Ribosomal RNA*, *low complexity*, and *simple/tandem repeats*. Para a identificação dos TEs a partir das bibliotecas de TEs (*Repbase* e *TEfam*) foram utilizados os parâmetros *default* do RM, pesquisa *slow* com programa NCBI BLAST (RMBLASTN, NCBI BLAST 2,2,23+), e somente o melhor *hit* foi mantido, como determinado pelo maior *score Smith–Waterman* proporcionado pelo RM.

Quantificação das repetições e estimativa da divergência

Para estimar a abundância de cada *contig* (número de cópias) e sua divergência, uma terceira amostra das *reads* é tomada e comparada ao conjunto de *contigs* montados e anotados pelo RM por meio de BLASTN (ASTCHUL et al., 1990). Para se obter uma estimativa total do conteúdo das repetições, essa amostra é também comparada aos *contigs* do *dnaPipeTE* anotados (BLASTN2). Finalmente, as *reads* não mapeadas são comparadas aos *contigs* não-anotados (BLASTN3) (Figura 2).

3.3. Pré-processamento das *reads* de *R. robustus*

As *reads* dos dois genomas de *R. robustus* (*R. robustus* II-III) foram “*trimadas*” usando o software *Urqt* (MODOLO; LERAT, 2015), com todos os parâmetros por *default*. As *reads* do DNA mitocondrial foram montadas com o software *NOVOplasty* (DIERCKXSENS; MARDULYN; SMITS, 2017), anotadas (usando como base de dados os alinhamentos de NCBI) utilizando *dnaPipeTE*, filtradas com o software *bowtie2* e removidas (LANGMEAD; SALZBERG, 2012). Foram ainda identificadas e descartadas possíveis sequências contaminantes provenientes de Bacteria, Archaea, vírus, fungos e microrganismos eucariotas usando o algoritmo *Kaiju* (MENZEL; NG; KROGH, 2016).

3.4. Análises genômicas de *R. robustus* II, *R. robustus* III e das *reads* brutas de *R. prolixus*.

Para as análises genômicas realizadas em *R. robustus* II e III foi considerado tamanho genômico estimado para *R. prolixus* (linhagem analisada neste estudo) de 733 MB (MESQUITA et al., 2015). Após análises preliminares utilizando coberturas de 0,25X, 0,50X e 1,0 X, foi escolhido o parâmetro 0,25X, que maximiza a montagem N50 para estes genomas (um procedimento realizado com *Trinity*). Para que os conjuntos de dados fossem homogêneos, também foi escolhida a cobertura 0,25 X para a análise das *reads* brutas da espécie referência *R. prolixus*.

3.5. Recuperação e quantificação das superfamílias e famílias de TEs

O *pipeline dnaPipeTE* realiza duas análises distintas de quantificação (Figura 2):

- 1) quantificação do total dos *contigs* montados (quantificação bruta). Para isto foi efetuado um mapeamento direto das *reads* da amostra 3 sobre os *contigs* anotados (BLASTN1).
- 2) estimativa do conteúdo de TEs, feita por sucessivos *blasts* uns com outros em duas etapas: (i) um segundo BLAST (BLASTN 2), em que a mesma amostra 3 foi sucessivamente blastada contra os *contigs* anotados somados às sequências de TEs do *Repbase*; e (ii) um terceiro BLAST (BLASTN 3) com os *contigs* não anotados para recuperar cópias que não tinham sido montadas e, desse modo, obter-se uma estimativa mais global do conteúdo repetitivo.

O *dnaPipeTE* gerou diversos *outputs* (saídas computacionais), resumidas no Quadro 1. A partir dos *outputs*, foram determinadas as proporções e distribuições comparativas nos três genomas, dos *contigs* anotados e quantificados, cruzando os dados contidos nestes *outputs* para as principais classes, superfamílias e famílias de TEs (classificação segundo WICKER et al., 2007):

a) Para comparar e calcular as proporções das principais classes de TEs foram utilizados os *outputs* “*Counts.txt*” e “*pieChart.pdf/png*”.

b) Para identificar as principais superfamílias de TEs e calcular as distribuições das proporções comparativas, foram utilizados os *outputs* “*Counts.txt*”, “*Trinity.fasta.out*”, “*reads_per_component_and_annotation*” e “*sorted.reads_vs_annotated.blast.out*”, dos quais os dados foram cruzados, filtramos e concatenados, gerando tabelas e gráficos comparativos.

c) Para identificar as famílias de TEs mais abundantes e estimar seus números de cópias de TEs, os dados foram agrupados a partir dos *contigs* gerados pelo *dnaPipeTE* (das saídas

"Trinity. Fasta", "reads_per_component_and_annotation" e "Counts.txt"), usando o programa CD-HIT-EST da plataforma CD-HIT (versão 4.6.1) (LI e GODZIK 2006). Para isso, foram estabelecidos alguns parâmetros com base no critério de Wicker et al. (2007) para agrupar pares de sequências com pelo menos 80% de similaridade na sequência mais curta alinhada, com um mínimo de 80% de identidade na sequência mais longa (parâmetros -aS 0,8 -c 0,8 -G 0 -g 1) (segundo GOUBERT, et al. 2015).

Quadro 1. Nome e descrição das saídas computacionais (*outputs*) geradas pela ferramenta *dnaPipeTE*.

Arquivo	Descrição
"Trinity. Fasta"	Este arquivo contém os <i>contigs</i> de <i>dnaPipeTE</i> , da última montagem realizada com <i>Trinity</i>
"reads_per_component_and_annotation"	Tabela com a contagem das <i>reads</i> e pares de bases alinhadas por <i>contig</i> do <i>dnaPipeTE</i> (do primeiro <i>BLAST</i>), assim como da melhor anotação do <i>RepeatMasker</i> .
"pieChart.pdf / png"	Gráfico com a proporção relativa das principais classes de repetições (incluindo os <i>blastn</i> 2 e 3)
"Bases_per_component.pdf / png"	Gráfico com o número de pares de bases alinhados em cada amostragem do <i>dnaPipeTE</i> (do <i>BLASTN</i> 1), ordenados pela proporção genômica do <i>contig dnaPipeTE</i> .
"landscapes.pdf"	Gráfico com a distribuição da idade dos TEs. Plotagem da distribuição da divergência entre as <i>reads</i> e <i>contigs</i> com os quais elas mapeiam.
"Trinity.fasta.out"	Output bruto do resultado do <i>RepeatMasker</i> (não classificado) das sequências <i>fasta</i> (<i>Trinity.fasta</i>) geradas.
"Counts.txt"	Contagem dos pares de bases da amostra alinhada para cada classe de TE.
"sorted.reads_vs_Trinity.fasta.blast.out"	Melhor hit por <i>reads</i> do <i>BLASTN</i> 1.
"sorted.reads_vs_annotated.blast.out"	Melhor hit por <i>reads</i> do <i>BLASTN</i> 2.
"sorted.reads_vs_unannotated.blast.out"	Melhor hit por <i>reads</i> do <i>BLASTN</i> 3.

*Estes arquivos contêm as saídas brutas do Trinity e os arquivos intermediários produzidos durante as etapas de montagem.

3.6. Comparação do mobiloma de *R. prolixus*, *R. robustus II* e *R. robustus III*

O mobiloma compartilhado pelos três genomas foi analisado por meio de três comparações (*R. prolixus* vs *R. robustus II*, *R. prolixus* vs *R. robustus III* e *R. robustus II* vs *R. robustus III*). Para isso, inicialmente os *contigs* gerados pelo *dnapipeTE* (*Trinity.fasta.out*) para cada análise foram agrupados em um só arquivo (exemplo: arquivo *.fas* de *contigs* de *R. prolixus* + *contigs R. robustus II*). A seguir, os *contigs* foram agrupados em *clusters* (agrupamentos) usando o programa CD-HIT-EST (versão 4.6.1) (LI e GODZIK 2006) utilizando os parâmetros citados em “c”. Posteriormente, os resultados da análise foram filtrados mantendo somente os *clusters* que agrupavam *contigs* de ambas espécies. Finalmente, para cada *cluster*, os *contigs* foram identificados usando o *output read_per_component_and_annotation* gerada por *dnaPipeTE*, que contém a anotação dos TEs, e com eles foram construídos gráficos de dispersão.

3.7. Investigação dos TEs no genoma montado de *R. prolixus* utilizando *dnaPipeTE*

Foi desenvolvida, ainda, uma estratégia para realizar buscas de sequências de TEs no genoma público montado de *R. prolixus* (ficheiro FASTA, WGS, 15X) obtido da plataforma VectorBase (www.vectorbase.org/organisms/rhodnius-prolixus) utilizando a metodologia *dnaPipeTE*. Para esse fim, o genoma montado de *R. prolixus* (702 Mb) foi picotado em pequenas *reads* usando o programa ART (HUANG et al., 2012), que é um simulador de sequenciamento Illumina (10X *single end* 150 bp *reads*), com os seguintes parâmetros: *-ss HS25 -l 150 -f 10 -ef*, isto a fim de simular *reads* de 150 pb HiSeq 2500, sem erros. A seguir, para a detecção dos TEs, foi utilizado *dnaPipeTE* para a análise das *reads* com os seguintes parâmetros: *genome_size 702.000.000, genome_coverage 0.25, sample_size 2, RM_t 0.2 -Trin_glue 2*. Esses parâmetros garantem duas iterações do montador *Trinity* com cobertura 0,25X, duas superposições destas são suficientes para unir os k-mer *contigs* durante o processo de montagem e anotação de TEs em que um limiar de 20% dos *contigs dnaPipeTE* deve ser coberto por um hit *RepeatMasker* para manter a anotação.

4. RESULTADOS

Os subconjuntos de sequências dos tipos genéticos de *R. robustus* II e III, juntamente com as *reads* já sequenciadas de *R. prolixus*, foram utilizados como entrada para análise pelo *dnaPipeTE*, seguindo os mesmos parâmetros nas três amostras (cobertura 0,25X, duas iterações de *Trinity* e um limiar de anotação de 0,4), revelando um panorama comparativo geral das proporções estimadas do conteúdo das diferentes famílias de TEs.

4.1. Conteúdo repetitivo em *R. prolixus*, *R. robustus* II e *R. robustus* III

A montagem do conteúdo repetitivo produziu 18.359 *contigs* com N50 de 341 pb para *R. prolixus*, 58.535 *contigs* com N50 de 325 pb para *R. robustus* II, e 53.981 *contigs* com N50 de 319 pb para *R. robustus* III. Foram anotadas 5.434 (*R. prolixus*), 9.406 (*R. robustus* II) e 9.778 (*R. robustus* III) sequências de TEs parciais, compreendendo 51, 57 e 56 superfamílias de TEs, respectivamente, de acordo com a classificação utilizada pelo Repbase.

A proporção do conteúdo genômico estimado pelo *dnaPipeTE* (Tabela Suplementar 1, Anexo 1) revela que: ~37,5% representam a porção repetitiva nos três genomas, que incluem TEs e outras repetições como: repetições simples de DNA, DNA satélite, repetições de baixa complexidade, rRNA e outros. Deste total, os TEs ocupam 19% (*R. prolixus*), 23,5% (*R. robustus* II) e 21,2% (*R. robustus* III), observando que nos três genomas os TEs mais abundantes são Transposons de DNA (*Classe II*), seguidos pelos retrotransposons *LINE* (*Classe I, non-LTR*) e *SINEs*, estando menos representados os *retrotransposons LTR* (Figura 3). Os números de pares de bases e as proporções genômicas de todas as sequências repetitivas presentes são apresentadas na Tabela Suplementar 1 (Anexo 1).

A anotação a partir do genoma montado de *R. prolixus* pelo *dnaPipeTE* identificou que 30,8% das *reads* correspondem à porção repetitiva de DNA, das quais 23,2% correspondem a sequências de TEs (Figura 3). Destes, a principal parcela corresponde aos Transposons DNA, seguida dos *LINEs* e *Helitrons*. Aproximadamente 5% do conteúdo repetitivo foi classificado como desconhecido, os quais podem também incluir sequências que não são TEs, assim como sequências de famílias gênicas.

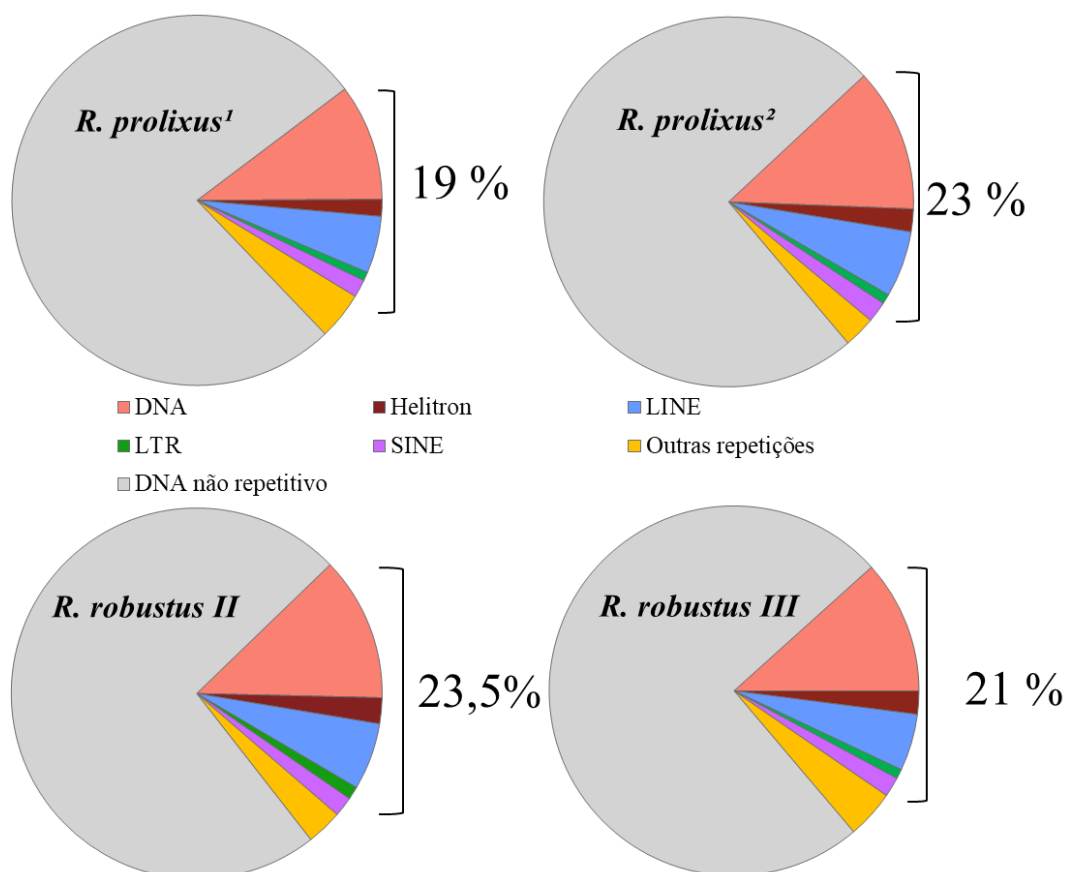


Figura 3. Proporção estimada das sequências repetitivas nos genomas de *R. prolixus*, *R. robustus* II e *R. robustus* III: **R. prolixus*¹ = anotação a partir das reads genômicas brutas; *R. prolixus*² = anotação a partir do genoma montado.

4.2. Comparação da proporção e dinâmica dos TEs entre *R. prolixus* e *R. robustus*.

Foram realizadas duas comparações entre os mobilomas anotados a partir de *reads* brutas de *R. prolixus* e de *R. robustus*: (i) proporções das superfamílias de TEs resultantes BLASTN 2, apresentadas na Tabela Suplementar 2 e Figura Suplementar 1 (Anexo 2) e Figura 4 e (ii) análise da divergência entre as superfamílias (Figura 5).

As proporções das superfamílias de TEs são altamente semelhantes nos três genomas. Os Retrotransposons *LTR* (*Classe I*) representam a menor fração, ocupando < 1% de porção genômica e < 4% em relação ao conteúdo dos TEs nos três genomas (Tabela Suplementar 2, Anexo 1). As superfamílias mais representativas dos Retrotransposons *LTR* são *Gypsy*, *Bel-Pao* e *Copia*. Por outro lado, os elementos *não-LTR* ocorrem em maior proporção nos três genomas, correspondendo a 6,5% a 7,5 % do conteúdo genômico e 30 a 34% do conteúdo de TEs. Os

clados/superfamílias mais representativos quanto ao conteúdo genômico são representantes da ordem LINE (*Jockey*, *I*, *LOA* e *RTE*) e os *SINEs* (*tRNA*) (Figura 4, Tabela Suplementar 2).

Notavelmente, a grande maioria dos TEs pertence à *Classe II*, variando de 11,5% a 14,8 % do conteúdo genômico e de aproximadamente 60% a 65% do total de TEs (Figura 4, Tabela Suplementar 2, Anexo 1). As superfamílias em maior proporção nos genomas são *Tc1/Mariner*, *hAT*, *Helitrons* e *DNA*. Na superfamília *Tc1/Mariner*, destaca-se a super-representação da família *Mariner* (Figura Suplementar 1, Anexo 2; Tabela Suplementar 3, Anexo 1) que corresponde a 3,6% de *R. prolixus*, 3,9% de *R. robustus* II e 3,1% de *R. robustus* III do conteúdo genômico, representado entre 14,3 a 25% do total de TEs.

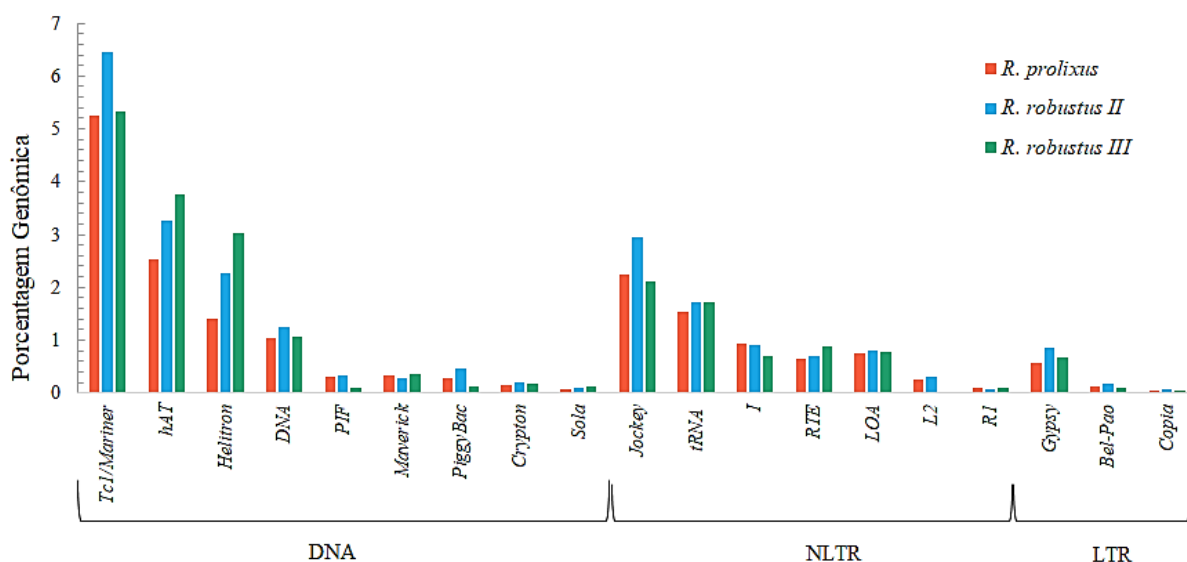


Figura 4. Proporção genômica das principais superfamílias de TEs nos genomas *Rhodnius*.

A Figura suplementar 1 (Anexo 2) juntamente com a Figura 6 apresentam as proporções das superfamílias de TEs em relação ao mobiloma total, e a distribuição da divergência das *reads* em relação aos *contigs dnaPipeTE* anotados. Essas análises foram também realizadas com os *contigs* de TEs obtidos a partir do genoma montado de *R. prolixus*. As duas figuras refletem alta semelhança na dinâmica evolutiva dos TEs nas quatro comparações. Dentre as *reads*, 63% foram anotadas como transposons de DNA (*Classe II*) constituindo o grupo de TEs mais abundante. Destes, a superfamília *Tc1-Mariner* é a mais abundante, variando de 24,7% (*R. robustus* III) a 27,9% (*R. prolixus*) do total de TEs. As *reads* da família *Mariner* é a mais prevalente do total de TEs (Figura suplementar 1, Anexo 2). As *reads* desta família são

altamente semelhantes aos *contigs* anotados (similaridade de sequência > 95%), sugerindo a existência de uma atividade dessa família e a ocorrência de explosões de transposição recentes.

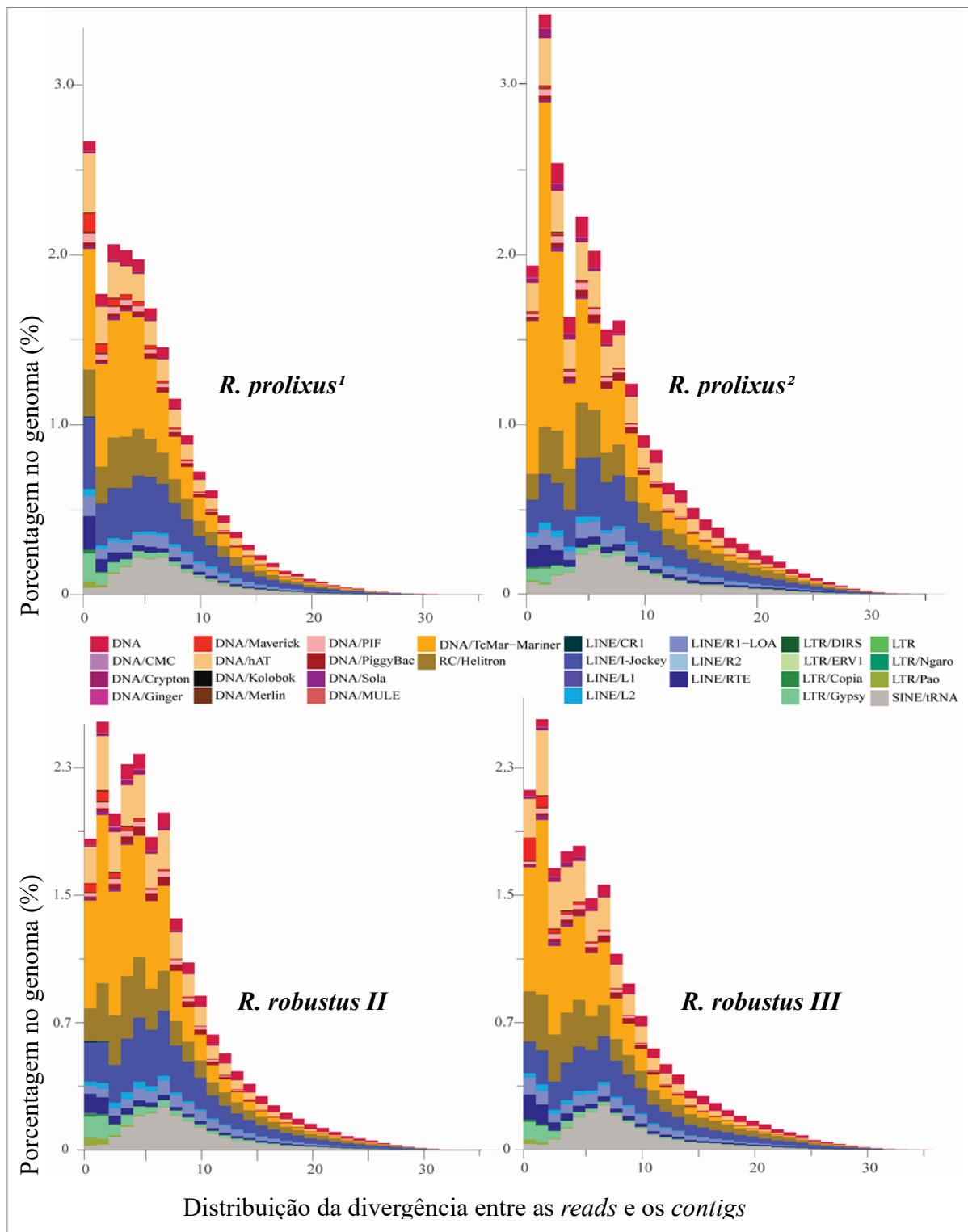


Figura 5. Idade relativa das superfamílias de TEs baseada na divergência das *reads* e relação aos *contigs*: **R. prolixus*¹ = a partir das *reads* genômicas brutas; *R. prolixus*² = a partir do genoma montado.

Destaca-se nas análises da idade relativa dos TEs com base na divergência das *reads* em relação aos *contigs* de TEs (Figura 5), que a divergência das *reads* está fortemente concentrada à esquerda do gráfico. Observa-se que a maior frequência dos Transposons de DNA apresenta divergência entre 1-5%, sugerindo que sofreram uma expansão recente, comparados aos *SINEs*, que mostram características de atividade mais antiga (maior frequência em torno de 5% de divergência).

Dentre os Transposons de DNA, a família *Mariner* apresenta *reads* altamente semelhantes aos *contigs* anotados (similaridade de sequência > 95%), sugerindo a existência de uma atividade dessa família e a ocorrência de explosões de transposição recentes. Já, os retrotransposons *LTR*, embora os menos abundantes, parecem ser em sua maior proporção muito recentes (divergência entre 0 e 2%). Os *LINEs*, por sua vez parecem ter mantido uma taxa constante de transposição durante a evolução recente, com a notável exceção do membro do *clado* *LINE/RTE*, que parece ter aumentado consideravelmente o número de cópias recentemente.

4.3. Comparação das proporções relativas de famílias TE compartilhadas entre *R. prolixus* e *R. robustus*

Concordantemente com os resultados anteriores, as análises comparativas par-a-par realizadas entre os três genomas do gênero *Rhodnius*, revelam altas porcentagens de compartilhamento entre as famílias de TEs: ~74% para *R. prolixus* - *R. robustus II*, ~76% para *R. prolixus* - *R. robustus III* e ~80% para *R. robustus III* - *R. robustus III*). Análises de correlação linear (em escala log10) mostram correlações positivas, altamente significativas, principalmente para as comparações *R. prolixus* - *R. robustus III* ($p < 0.001$) e *R. robustus II* - *R. robustus III* ($p < 0.001$) (Figura suplementar 2, Anexo 2).

Na Figura 6 são apresentados os resultados das análises de compartilhamento separadamente para cada *Classe* I e II. Como para a análise global, as correlações são todas positivas e significativas, indicando que a maior parte das famílias compartilhadas ocorrem em proporções similares nos três genomas. Contudo, para algumas famílias as proporções não são correspondentes, como abaixo destacado para as três comparações:

a) *R. prolixus* - *R. robustus* II: Algumas famílias apresentam-se com grandes diferenças entre ambos genomas, por exemplo as famílias *Jockey-2* com 0,12% em *R. prolixus* e 0,0005 % em *R. robustus* II e *I-5* com 0,4% em *R. prolixus* e 0,001% em *R. robustus* II, na *Classe I*; e a família *Helitron-2* com 0,01% em *R. prolixus* e 0,0001 % em *R. robustus* II, na *Classe II*. Integralmente, a maior parte das famílias compartilham a mesmas proporções genômicas (principalmente transposons de DNA e não-LTRs), dentre estas famílias se caracterizam aquelas cujo grau de compartilhamento é mais abundante, destacando-se *Mariner-1* e *Helitron2-N1* (compartilham ~1.6% entre os genomas). Contrariamente outras famílias são menos abundantes nos genomas, por exemplo *hAT-N33* (~0,0008%) e *Gypsy-58* (0,001%).

b) *R. prolixus* - *R. robustus* III: observa-se a tendência linear geral, com a maioria das famílias compartilhando as mesmas proporções de famílias de TEs, umas com valores mais altos, por exemplo a família *Mariner-1* (*R. prolixus*: 5,9% e *R. robustus* III: 4,6%) na *classe II* e outras com valores mais baixos, por exemplo a família *Jockey-2* (*R. prolixus*: 0,0003% e *R. robustus* III: 0,0004%) na *classe I*. Dentre as famílias da *Classe II* que ocorrem em proporções significativamente diferentes destacam-se a família *SMAR26*, com 0,66% em *R. prolixus* e apenas 0,001% em *R. robustus* III, a família *hAT-N3* com 0,26% em *R. robustus* III e 0,0001% em *R. prolixus*; e a família *Helitron-2* com 1,01% em *R. prolixus* e 0,03 % em *R. robustus* III. Na *Classe I* destacam-se as famílias *RTE-10* (*R. prolixus*: 0,0001% e *R. robustus* III: 0.006%) e *Gypsy-11* (*R. prolixus*: 0,01% e *R. robustus* III: 0,0001%).

c) *R. robustus* II - *R. robustus* III: como exemplo de famílias compartilhadas ocorrendo em alta proporção nos dois genomas destacam-se *Mariner-1* (5,5% em *R. robustus* II e 6 % em *R. robustus* III) e *Helitron-2N1* (1,5% em *R. robustus* II e 2 % em *R. robustus* III) na *classe II*, enquanto que ocorrem em ambos os genomas em baixas proporções as famílias *Nimb-4* (~0,0001%) e *I-5* (~0,0004%) na *classe I*. Igualmente às outras comparações, muitas famílias compartilham proporções significativamente diferentes como *Helitron-2N3* (1,03 % em *R. robustus* III e 0,05% em *R. robustus* II) na *classe II* e *Jockey-1* (0,05% em *R. robustus* II e 0.0001 % em *R. robustus* III) na *classe I*, entre outras.

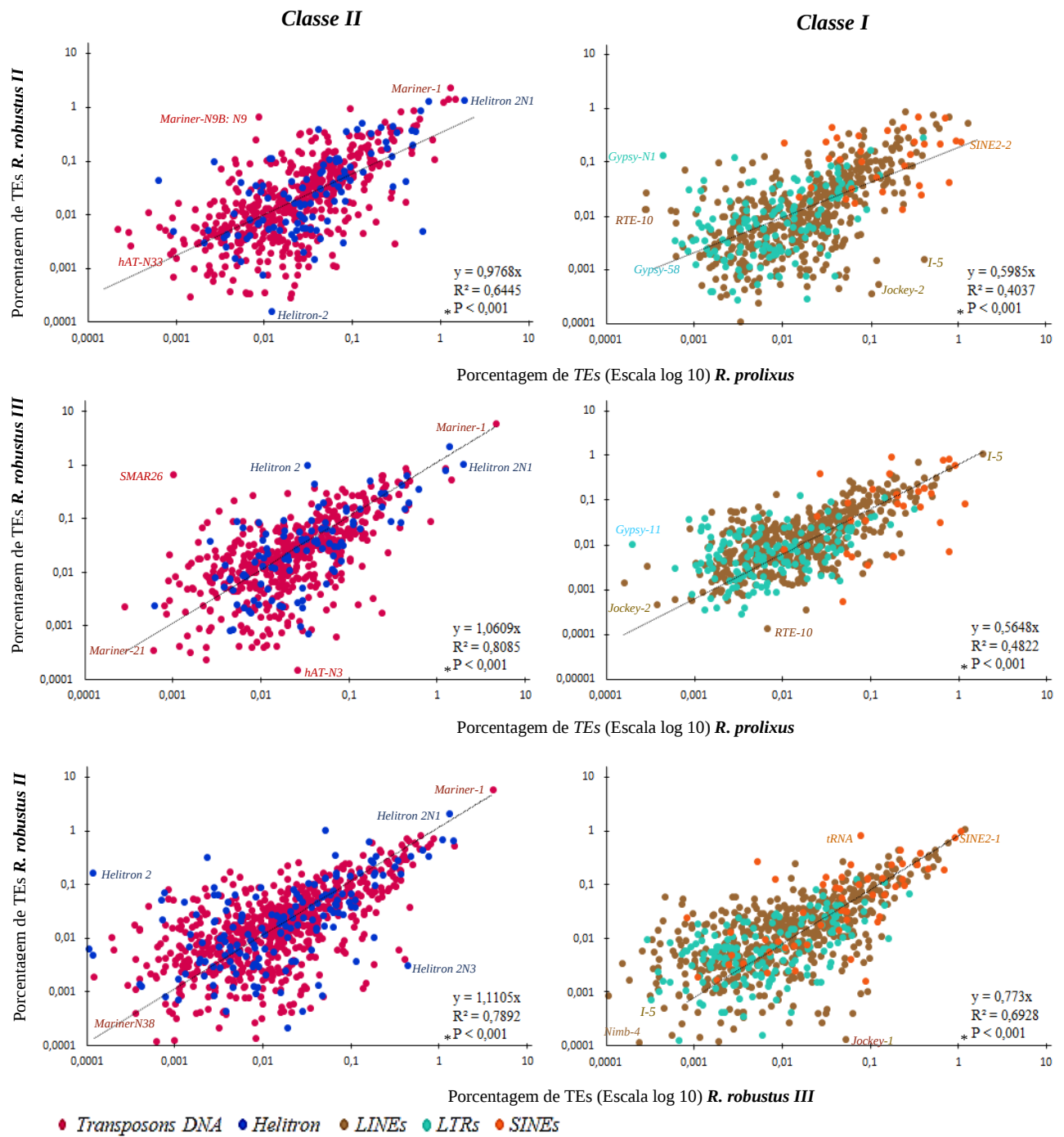


Figura 6. Comparação da proporção genômica relativa das famílias de TEs compartilhadas nas *Classe I* e *II* entre *R. prolixus*, *R. robustus II* e *R. robustus III*, em termos de porcentagem genômica (escala $\log_{10}$): *Correlação significativa (valor- $p \leq 0,01$)

5. DISCUSSÃO

Este estudo apresenta a primeira descrição do mobiloma de *R. robustus* e sua comparação em relação ao mobiloma de sua espécie irmã, *R. prolixus*. Para sua realização foram utilizadas as *reads* obtidas por sequenciamento de baixa cobertura (NGS) dos genomas de dois tipos genéticos de *R. robustus* (tipo II e III), como também as *reads* obtidas no sequenciamento que gerou o genoma montado de *R. prolixus* (MESQUITA et al., 2015). Esses dados foram utilizados para identificar as sequências de TEs por meio da ferramenta *dnaPipeTE*, um *pipeline* que permite uma análise rápida e completa dos TEs em dados brutos de NGS (*reads* não montadas), desenvolvido para a realização de estudos comparativos com insetos (GOUBERT et al., 2015). O *dnaPipeTE* traz vantagens em relação o *Repeat Explorer* (RE) (NOVAK et al., 2013), que também permite anotar e quantificar TEs a partir de dados brutos de NGS. Usando Trinity como um montador das sequências de TEs, o *dnaPipeTE* pode recuperar *contigs* de TEs maiores e realizar múltiplas iterações com amostras adicionais independentes. Enquanto o RE faz a anotação dos TEs somente nas *reads* amostradas, o *dnaPipeTE* pode anotar e quantificar as famílias TE com seus *contigs* e o número de *reads* mapeadas, permitindo identificação de maior número de TEs do que o RE, com uma diminuição substancial no tempo computacional, como demonstrado no estudo do genoma de *Aedes albopictus* (GOUBERT et al., 2015). Além disso, o *dnaPipeTE* fornece uma grande quantidade de *outputs* (tabelas de resumo, gráficos, conjuntos de dados ordenados) e é o primeiro método capaz de gerar uma representação da distribuição da divergência dos TEs, que representam a idade relativa dos TEs no genoma, sem montagem prévia do genoma. Finalmente, estudos comparativos entre anotações de TEs a partir de montagem genômica e de dados NGS validaram o método *dnaPipeTE* (GOUBERT et al., 2015; SESSEGOLO et al., 2016; DA SILVA et al., 2018; THOMAS-BULLE et al., 2018) dando segurança aos dados aqui reportados.

A análise do mobiloma de *R. robustus* e *R. prolixus* partiu do nosso questionamento se *R. robustus* também teria sofrido explosões de transposição de elementos *Mariner* reportados em *R. prolixus* (FERNANDEZ-MEDINA et al., 2016). Pela indisponibilidade de amostras de DNA, não foi possível sequenciar o genoma de *R. prolixus* pela metodologia NGS, por essa razão foram utilizadas as *reads* brutas do sequenciamento *Shotgun* (WGS) dessa espécie, de modo a se ter uma comparação da composição do mobiloma das duas espécies irmãs utilizando

a mesma metodologia de anotação dos TEs, já que metodologias diferentes podem produzir anotações diferentes, tanto qualitativa quanto quantitativamente (RIUS et al., 2016).

Os dados aqui apresentados revelam em primeira instância uma abundante quantidade de TEs nas duas espécies de *Rhodnius* (19 a 23,5%). Em *R. prolixus*, o conteúdo de TEs representa o triplo do valor previamente estimado (~5,6%) por Mesquita et al (2015) e Filée et al. (2015), a partir da mesma montagem genômica por sequenciamento WGS. Essa diferença poderia ser explicada tendo em conta as seguintes considerações: 1) são relatadas dificuldades de montagem de genomas sequenciados via WGS (que produz *reads* de 800 a 900 pb), resultando em subestimativas do conteúdo de TEs, quando o mobiloma é recente e altamente repetitivo (GREEN, 2002); 2) os resultados de nossas análises complementares, utilizando a estratégia de cortar o genoma montado de referência de *R. prolixus* em pequenas *reads*, usando um simulador de sequenciamento Illumina, e análise dessas *reads* pelo *dnaPipeTE* com cobertura de 0,25X, produziram resultados surpreendentes: o conteúdo de TEs no genoma montado de *R. prolixus* é de 23,2%. Isso significa que a mesma quantidade de TEs anotadas nas *reads* brutas de *R. prolixus* está presente no genoma montado, e é similar à anotada no genoma de sua espécie irmã *R. robustus*. Ademais, a distribuição divergência das *reads* em relação aos *contigs* de TEs foi bastante similar ao do genoma anotado pelo método *dnaPipeTE*. Esses resultados indicam uma possível subestimativa do conteúdo de TEs no genoma montado de *R. prolixus*, associada possivelmente ao método de anotação dos TEs, já que Mesquita et al. (2015) não utilizou uma abordagem *de novo*, e parte dos TEs podem não ter sido detectados.

Independentemente da quantidade total dos TEs nos genomas das duas espécies, as análises comparativas dos mobilomas evidenciam uma convergência no panorama dos TEs que os compõem, com uma distribuição similar tanto quantitativa como qualitativa. Tanto em *R. prolixus* (¹ e ²) como em *R. robustus* II e III, o mobiloma é predominantemente constituído por elementos de *Classe II* (~ 65%), dentre os quais as superfamílias *Tc1/Mariner* (~27%), *hAT* (~14%) e *Helitron* (~10%) são as mais abundantes, totalizando mais que 50% dos TEs. Também nos três genomas, os elementos de *Classe I* representam aproximadamente 35%, sendo as mais abundantes os retrotransposons sem LTRs da Ordem LINE (superfamílias Jockey: ~11%, Loa: ~3,5% e RTE: 1%) e *Gypsy* (~3,5%), superfamília da Ordem Retrotransposons com LTR. Estas proporções aproximam-se das publicadas por Mesquita et al. (2015) e Filée et al. (2015), em cujas anotações, dos 75% dos elementos *Classe II* que compõem o genoma de *R. prolixus*, 50%

pertencem às superfamílias *Tc1/Mariner*, *hAT* e *Helitron* e 25% pertencem à *Classe I*, e destes, 15,5% representam as superfamílias *Jockey*, *Loa*, *RTE* e *Gypsy*.

Um aspecto importante a ser considerado no presente trabalho é que não foram anotados os elementos não-autônomos conhecidos como *MITEs* (*Miniature Inverted-Repeat Transposable Elements*). Os *MITEs* são ausentes do banco de dados *Repbase*, dessa forma não foram anotados pelo *dnaPipeTE*. Os *MITEs* são Transposons de DNA não autônomos, numerosos em genomas de plantas e animais, pequenos (100-600 pb), flanqueados por TIRs e TSDs, ricos em repetições AT e sem transposase, provavelmente originados por deleções de Transposons de DNA autônomos (FESCHOTTE; SWAMY; WESSLER, 2003). De acordo com Mesquita et al. (2015), o genoma de *R. prolixus* abriga 0,5% de *MITEs*. Já Zhang et al. (2013) reportaram que os *MITEs* de *R. prolixus* derivam de cinco superfamílias de Transposons de DNA (*Sola*, *HAT*, *Ginger* e *Tc1/Mariner*). Também Filée et al. (2015), mediante anotação *de novo* de *MITEs* presentes no genoma dessa espécie identificaram oito *clusters* de sequências com características de *MITEs*, sendo que quatro desses eram associados a elementos da família *Mariner* presentes no genoma. Como famílias das superfamílias *Sola*, *HAT*, *Ginger* e *Tc1/Mariner* foram abundantes na anotação por nós realizada, é muito provável que parte dos TEs *Classe II* (~65%) anotados pelo *dnaPipeTE* sejam *MITEs*.

A prevalência dos elementos de *Classe II* no genoma das duas espécies de *Rhodnius* pode ser uma característica exclusiva deste complexo de espécies, gênero *Rhodnius*, ou até da família (Reduviidae), considerando que em outras famílias de insetos com mobilomas anotados, os elementos de *Classe I* são consideravelmente mais abundantes, tomando-se como representantes os genomas de *Bombix mori* (Bombycidae): 89%, espécies de *Drosophila* (Drosophilidae): 67% a 93%, *Tribolium castaneum* (Tenebrionidae): 87% e *Anopheles gambiae* (Culicidae): 72% (WANG et al. 2008; CLARK et al. 2007). Contudo, independente da proporção em que ocorrem nos genomas, se alta ou baixa, as superfamílias *Tc1/Mariner*, *hAT* e *Helitron* são altamente ubíquas, como demonstrado em estudos genômicos comparativos em 257 espécies de eucariotos, incluindo animais, plantas e fungos. Uma das explicações para tal ubiquidade é a alta taxa de transferência horizontal de elementos dessas superfamílias, em particular da superfamília *Tc1/Mariner* (WALLAU et al., 2014; FILÉE et al., 2015), que representa quase um terço dos genomas das duas espécies de *Rhodnius*.

Dentre as famílias da superfamília *Tc1/Mariner*, a mais representativa em *R. prolixus* e *R. robustus* II-III é a família *Mariner* (~17%), constituída principalmente de três clados (*Mariner-1_RPr*: ~11%, *Mariner-N9_RPr*: ~3% e *Mariner-N9_RPr*: ~1%) (Tabela Suplementar 3, Anexo 1) que representam mais de 90% dos integrantes da família. Além dessa prevalência, destaca-se a baixíssima ou nula divergência entre suas cópias, atributo de famílias de TE recentes ou ativas em que o número de cópias aumenta mais rapidamente do que o acúmulo de mutações nessas cópias (LERAT et al., 2011; STATON et al., 2012). Esta extrema similaridade entre as cópias da família *Mariner* foi reportada Fernandez-Medina et al. (2016) e Filée et al. (2015) em *R. prolixus*, que associaram esta baixíssima divergência e alto número de cópias a explosões de transposição. Além disso, Fernandez-Medina et al. (2016), em uma tentativa de entender a dinâmica de expansão nestes três clados, formulou a hipótese que as explosões de transposição poderiam estar associadas ao evento da especiação do ancestral *prolixus/robustus* originando a espécie *R. prolixus*. Esta foi a questão central proposta para ser investigada neste trabalho. Os resultados obtidos mostram claramente que as explosões de transposição da família *Mariner* evidenciadas em *R. prolixus* não são exclusivas dessa espécie, mas são compartilhadas com os dois tipos genéticos de sua espécie irmã *R. robustus* II – III, portanto não estão relacionadas à especiação de *R. prolixus*, mas sim é um atributo do ancestral das duas espécies.

Como anteriormente mencionado neste trabalho, as quatro espécies que formam o atual complexo *R. robustus* evoluíram a partir de um ancestral semelhante a *R. robustus* em duas regiões da América do Sul. Ao norte, nas planícies da Venezuela, evoluíram *R. prolixus* e *R. robustus*, a última constituída hoje por quatro tipos genéticos, sendo que o tipo 1, que ocorre na Venezuela compartilha com *R. prolixus* um ancestral comum mais recente do que com os tipos genéticos II, II e IV (MONTEIRO et al., 2003). Ao sul do Brasil, evoluíram as espécies *R. neglectus* e *R. nasutus*. As explosões de transposição evidenciadas em linhagens de *R. robustus* e *R. prolixus* respondem à questão proposta neste trabalho, mas levanta outro questionamento: as explosões de transposição também seriam compartilhadas por *R. neglectus* e *R. nasutus*, e neste caso, teriam ocorrido no ancestral desse complexo de espécies, ou seriam exclusivas do par *R. prolixus* / *R. robustus*. Novos estudos são necessários para ampliar nosso entendimento sobre o papel dos elementos *Tc1-Mariner* e das explosões de transposição na evolução destas espécies.

Além da predominância da família *Mariner*, os três genomas comparados neste trabalho revelam três características importantes para se entender a dinâmica evolutiva destes TEs: (i) compartilhamento de famílias que ocorrem tanto em alta (ex. *Helitron-2N1*, *Mariner-1*, ambas famílias da *Classe II*), como em baixa (ex. *Nimb*, *RTE*, Ordem *LINE Classe I*) proporção nos três genomas; (ii) compartilhamentos de famílias que ocorrem em proporção dissimilar entre genomas, alta em um e baixa no outro (ex. *Helitron-2* ou *Jockey-1*) e (iii) não compartilhamento de determinada família entre genomas, indicando sua ausência em um deles, ou frequência tão baixa que não pode ser amostrada; entre 20 a 27% dos TEs são exclusivos de um dos genomas (ex. *BEL-13*, é exclusivo no genoma de *R. robustus III* comparado a *R. prolixus* ou *Copia2-VV*, que é exclusivo no genoma de *R. robustus II* comparado a *R. robustus III*). Este panorama pode ser explicado mediante o “ciclo evolutivo dos TEs” (PINSKER et al., 2001), que pode ser caracterizado por fases que podem ser ilustradas pelas situações acima elencadas.

O ciclo inicia com a introdução de um TE em um genoma que não o abriga por meio de três origens principais: (i) origem *de novo*, por meio de eventos de mutação e de recombinação de sequências degradadas já presentes no genoma, (ii) por transferência horizontal (HT) mediada por um ou mais vetores, e (iii) por meio de hibridização introgressiva e poliespermia entre espécies aparentadas. A taxa de frequência de ocorrência de HT depende de vários fatores como a capacidade de invasão, características do genoma hospedeiro ou fatores ambientais (PINSKER et al., 2001). Uma característica interessante dos elementos *Mariner* é encontrarem-se envolvido em inúmeros eventos de HT (DOTTO et al., 2015). Wallau et al. (2016) reportaram eventos de HT de 24 linhagens diferentes de elementos *Mariner* caracterizados em genomas de espécies de *Drosophila*. Em *R. prolixus* Filée et al (2015) detectaram nove eventos de HT de elementos *Mariner*, a maioria envolvendo insetos (*Dendroctonus ponderosae*, *Scolia oculata*, *Bombus terrestris*, *Glossina pallipes* e *Drosophila*). Nessa sequência, confirmamos a aplicação dessa etapa essencial no início do ciclo evolutivo para os elementos *Mariner* que nos detectamos.

No ciclo evolutivo dos TEs, após sua origem, inicia-se uma fase de expansão no genoma, podendo ocorrer explosões de transposição. Nesse sentido, sendo *Mariner* um modelo de HT, e sabendo-se que após a transferência horizontal podem ocorrer “explosões de transposição” do elemento invasor, já que estando ausente daquele genoma não haverá mecanismos pré-estabelecidos de regulação que controle sua expansão genômica, as explosões de transposição de *Mariner*, identificadas neste e nos trabalhos anteriores, ilustram a dinâmica de expansão recente dos TEs nos genomas de *R. prolixus* e *R. robustus*.

A fase seguinte do ciclo é caracterizada por dois eventos: (i) redução crescente da mobilidade do elemento devido ao acúmulo de cópias defectivas, e (ii) estabelecimento de mecanismos epigenéticos de silenciamento desenvolvidos pelo genoma hospedeiro (PINSKER et al., 2001). E a fase final do ciclo, que pode durar milhões de anos, é denominada de senescência. Com a atividade transposicional regulada ou suprimida pelos mecanismos de silenciamento, a frequência das cópias funcionais diminui como resultado de mutações aleatórias, excisão, seleção purificadora e perdas estocásticas. O resultado da regulação e senescência atuando independentemente nos três genomas, já que *R. prolixus* e *R. robustus* são espécies diferentes e *R. robustus* II e *R. robustus* III constituem populações alopátricas divergentes (MONTEIRO et al., 2003), pode explicar as diferenças observadas na abundância de famílias de TEs nos três genomas de *R. prolixus*/*R. robustus* II/*R. robustus* III.

Em síntese, este trabalho mostra que a superfamília *Tc1-Mariner*, em geral, e a família *Mariner*, em particular, caracterizadas em *Rhodnius*, constituem excelente modelo para ilustrar o ciclo evolutivo dos TEs. As proporções de TEs anotados mostram que após as explosões de transposição, podem estar em curso diferentes pressões seletivas e eventos estocásticos que estão modulando as distribuições dos elementos *Mariner* em cada genoma. Concluimos, ainda, que este trabalho comprova que a ferramenta *dnaPipeTE*, baseada em anotação *de novo*, é eficiente para caracterizar a composição e dinâmica dos TEs em genomas sequenciados pela metodologia NGS, possibilitando inclusive detectar maior diversidade do mobiloma, em decorrência de amostrar diretamente as *reads* dos TEs. Acreditamos que os dados e as interpretações aqui fornecidas oferecerão uma base para estudos futuros, com o objetivo de compreender o papel desempenhado pelos elementos de transposição durante o processo de evolução e a adaptação das espécies.

5. CONCLUSÕES

1. O método *dnaPipeTE* é eficiente para caracterizar e comparar a composição e dinâmica genômica dos TEs, permitindo uma análise rápida e completa de genomas não montados sequenciados pela metodologia NGS;
2. Estimou-se uma abundante quantidade de TEs (19 %) no genoma de *R. prolixus*, mais do que o dobro da proporção previamente reportada (~5,6%) por Mesquita et al (2015);
3. A distribuição genômica dos TEs similar nos mobilomas de *R. prolixus* e de *R. robustus*, predominantemente constituídos por elementos de *Classe II* (~ 65%), é característica do complexo de espécies complexo *R. robustus*;
4. Baixa divergência das *reads* da família *Mariner* em relação aos *contigs* caracterizam explosões de transposição nesta família em *R. robustus* II e III, indicando que esse fenômeno não é exclusivo de *R. prolixus*, não estando portanto este evento associado diretamente à especiação desta espécie;
5. A comparação dos mobilomas das duas espécies *Rhodnius* apresenta características importantes para compreensão da dinâmica evolutiva dos TEs, como o (i) compartilhamento de famílias que ocorrem tanto em alta como em baixa proporção entre genomas; (ii) compartilhamento de famílias que ocorrem em proporções dissimilares entre genomas, alta em um e baixa no outro, e (iii) não compartilhamento de determinadas famílias entre genomas.

Os resultados obtidos neste trabalho proporcionam uma ampla visão sobre a dinâmica evolutiva dos TEs nos genomas *R. prolixus* e *R. robustus* fornecendo subsídios para compreendermos a evolução e adaptação destas espécies.

6. REFERÊNCIAS

- ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic local alignment search tool. **J Mol Biol**, v. 215, p. 403–410, 1990.
- ABAD-FRANCH, F.; PAVAN, M. G.; JARAMILLO, O. N.; PALOMEQUE, F. S.; DALE, C.; CHAVERRA, D.; MONTEIRO, F. A. *Rhodnius barretti*, a new species of Triatominae (Hemiptera: Reduviidae) from western Amazonia. **Mem Inst Oswaldo Cruz**, v. 108 Suppl 1, p. 92-9, 2013.
- BATZER, M. A.; DEININGER, P. L. Alu repeats and human genomic diversity. *Nat Rev Genet*, v. 3, n. 5, p. 370-9, May 2002.
- BELYAYEV, A. Bursts of transposable elements as an evolutionary driving force. **J Evol Biol**, v. 27, n. 12, p. 2573-84, Dec 2014.
- BOHNE, A.; BRUNET, F.; GALIANA-ARNOUX, D.; SCHULTHEIS, C.; VOLFF, J. N. Transposable elements as drivers of genomic and biological diversity in vertebrates. **Chromosome Res**, v. 16, n. 1, p. 203-15, 2008.
- BROOKFIELD, J.F. The ecology of the genome - mobile DNA elements and their hosts. **Nat Rev Genet**. v. 6, p. 128-36, 2005.
- CARARETO, C.M.A.; HERNANDEZ, E.H.; VIEIRA, C. Genomic regions harboring insecticide resistance-associated Cyp genes are enriched by transposable element fragments carrying putative transcription factor binding sites in two sibling *Drosophila* species. **Gene**, v. 537, p. 93-99, 2013.
- CLARK, A. G.; EISEN, M. B.; SMITH, D. R.; BERGMAN, C. M.; OLIVER, B.; MARKOW, T. A.; KAUFMAN, T. C.; KELLIS, M.; GELBART, W.; IYER, V. N.; POLLARD, D. A.; SACKTON, T. B.; LARRACUENTE, A. M.; SINGH, N. D. Evolution of genes and genomes on the *Drosophila* phylogeny. **Nature**, v. 450, n. 7167, p. 203-18, Nov 08 2007.
- DABOUSSI, M. J.; CAPY, P. Transposable elements in filamentous fungi. **Annu Rev Microbiol**, v. 57, p. 275-99, 2003.
- DA SILVA, A. F.; DEZORDI, F. Z.; LORETO, E. L. S.; WALLAU, G. L. *Drosophila* parasitoid wasps bears a distinct DNA transposon profile. **Mob DNA**, v. 9, p. 23, 2018.
- DE BOER, J. G.; YAZAWA, R.; DAVIDSON, W. S.; KOOP, B. F. Bursts and horizontal evolution of DNA transposons in the speciation of pseudotetraploid salmonids. *BMC Genomics*, v. 8, p. 422, Nov 16 2007.
- DIERCKXSENS, N.; MARDULYN, P.; SMITS, G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. **Nucleic Acids Res**, v. 45, n. 4, p. e18, Feb 28 2017.
- DOTTO, B. R.; CARVALHO, E. L.; SILVA, A. F.; DUARTE SILVA, L. F.; PINTO, P. M.; ORTIZ, M. F.; WALLAU, G. L. HTT-DB: horizontally transferred transposable elements database. **Bioinformatics**, v. 31, n. 17, p. 2915-7, Sep 1 2015.

- DUJARDIN, J. P.; MUNOZ, M.; CHAVEZ, T.; PONCE, C.; MORENO, J.; SCHOFIELD, C. J. The origin of *Rhodnius prolixus* in Central America. **Med Vet Entomol**, v. 12, n. 1, p. 113-5, Jan 1998.
- FERNANDEZ-MEDINA, R. D.; GRANZOTTO, A.; RIBEIRO, J. M.; CARARETO, C. M. Transposition burst of Mariner-like elements in the sequenced genome of *Rhodnius prolixus*. **Insect Biochem Mol Biol**, v. 69, p. 14-24, Feb 2016.
- FESCHOTTE, C.; JIANG, N.; WESSLER, S. R. Plant transposable elements: where genetics meets genomics. **Nat Rev Genet**, v. 3, n. 5, p. 329-41, May 2002.
- FESCHOTTE, C.; SWAMY, L.; WESSLER, S. R. Genome-wide analysis of mariner-like transposable elements in rice reveals complex relationships with stowaway miniature inverted repeat transposable elements (MITEs). *Genetics*, v. 163, n. 2, p. 747-58, Feb 2003.
- FESCHOTTE, C.; PRITHAM, E. J. DNA transposons and the evolution of eukaryotic genomes. **Annu Rev Genet**, v. 41, p. 331-68, 2007.
- FINNEGAN, D. J. Eukaryotic transposable elements and genome evolution. **Trends Genet**, v. 5, n. 4, p. 103-7, Apr 1989.
- FILÉE, J.; ROUAULT, J. D.; HARRY, M.; HUA-VAN, A. Mariner transposons are sailing in the genome of the blood-sucking bug *Rhodnius prolixus*. **BMC Genomics**, v. 16, p. 1061, Dec 15 2015.
- GOODIER, J. L.; KAZAZIAN, H. H.; JR. Retrotransposons revisited: the restraint and rehabilitation of parasites. **Cell**, v. 135, n. 1, p. 23-35, Oct 03 2008.
- GOUBERT, C.; MODULO, L.; VIEIRA, C.; VALIENTEMORO, C.; MAVINGUI, P.; BOULESTEIX, M. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). **Genome Biol Evol**, v. 7, n. 4, p. 1192-205, Mar 11 2015.
- GRABHERR, M. G.; HAAS, B. J.; YASSOUR, M.; LEVIN, J. Z.; THOMPSON, D. A.; AMIT, I.; ADICONIS, X.; FAN, L.; RAYCHOWDHURY, R.; ZENG, Q.; CHEN, Z.; MAUCELI, E.; HACHOEN, N.; GNIRKE, A.; RHIND, N.; DI PALMA, F.; BIRREN, B. W.; NUSBAUM, C.; LINDBLAD-TOH, K.; FRIEDMAN, N.; REGEV, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nat Biotechnol**, v. 29, n. 7, p. 644-52, May 15 2011.
- GRANZOTTO, A.; CRUZ, G. M. Q. Regulação de Elementos de Transposição: Mecanismos Epigenéticos de Silenciamento, Autorregulação e Ativação por Estresse em: CARARETO, C. M. A.; MONTEIRO-VITORELLO, C. B.; SLUYS, M. V. Elementos de Transposição Diversidade, evolução, aplicações e impacto os genomas dos seres vivos. Cap. 4, **Ed. Fiocruz**, 2015.
- GREEN, P. Whole-genome disassembly. **Proc Natl Acad Sci U S A**, v. 99, n. 7, p. 4143-4, Apr 2 2002.
- HARRY, M. Use of the median process of the pygophore in the identification of *Rhodnius nasutus*, *R. neglectus*, *R. prolixus* and *R. robustus* (Hemiptera: Reduviidae). **Ann Trop Med Parasitol**, v. 87, n. 3, p. 277-82, Jun 1993.

- HARRY M. Morphometric variability in the Chagas' disease vector *Rhodnius prolixus*. **Jpn J Genet**, v. 69, n. 3, p. 233-50, Jun 1994.
- HARRY, M.; GALINDEZ, I.; CARIOU, M. L. Isozyme variability and differentiation between *Rhodnius prolixus*, *R. robustus* and *R. pictipes*, vectors of Chagas disease in Venezuela. **Med Vet Entomol**, v. 6, n. 1, p. 37-43, Jan 1992.
- HOLT, R. A.; SUBRAMANIAN, G. M.; HALPERN, A.; SUTTON, G. G.; CHARLAB, R.; NUSSKERN, D. R.; WINCKER, P.; CLARK, A. G.; RIBEIRO, J. M.; WIDES, R.; SALZBERG, S. L.; LOFTUS, B.; YANDELL, M.; MAJOROS, W. H.; RUSCH, D. B.; LAI, Z.; KRAFT, C. L.; ABRIL, J. F.; ANTHOUARD, V.; ARENSBURGER, P.; ATKINSON, P. W.; BADEN, H.; DE BERARDINIS, V.; BALDWIN, D.; BENES, V.; BIEDLER, J.; BLASS, C.; BOLANOS, R.; BOSCH, D.; BARNSTEAD, M.; CAI, S.; CENTER, A.; CHATURVERDI, K.; CHRISTOPHIDES, G. K.; CHRYSTAL, M. A.; CLAMP, M.; CRAVCHIK, A.; CURWEN, V.; DANA, A.; DELCHER, A.; DEW, I.; EVANS, C. A.; FLANIGAN, M.; GRUNDSCHOBBER-FREIMOSER, A.; FRIEDLI, L.; GU, Z.; GUAN, P.; GUIGO, R.; HILLENMEYER, M. E.; HLADUN, S. L.; HOGAN, J. R.; HONG, Y. S.; HOOVER, J.; JAILLON, O.; KE, Z.; KODIRA, C.; KOKOZA, E.; KOUTSOS, A.; LETUNIC, I.; LEVITSKY, A.; LIANG, Y.; LIN, J. J.; LOBO, N. F.; LOPEZ, J. R.; MALEK, J. A.; MCINTOSH, T. C.; MEISTER, S.; MILLER, J.; MOBARRY, C.; MONGIN, E.; MURPHY, S. D.; O'BROCHTA, D. A.; PFANNKUCH, C.; QI, R.; REGIER, M. A.; REMINGTON, K.; SHAO, H.; SHARAKHOVA, M. V.; SITTER, C. D.; SHETTY, J.; SMITH, T. J.; STRONG, R.; SUN, J.; THOMASOVA, D.; TON, L. Q.; TOPALIS, P.; TU, Z.; UNGER, M. F.; WALENZ, B.; WANG, A.; WANG, J.; WANG, M.; WANG, X.; WOODFORD, K. J.; WORTMAN, J. R.; WU, M.; YAO, A.; ZDOBNOV, E. M.; ZHANG, H.; ZHAO, Q.; ZHAO, S.; ZHU, S. C.; ZHIMULEV, I.; COLUZZI, M.; DELLA TORRE, A.; ROTH, C. W.; LOUIS, C.; KALUSH, F.; MURAL, R. J.; MYERS, E. W.; ADAMS, M. D.; SMITH, H. O.; BRODER, S.; GARDNER, M. J.; FRASER, C. M.; BIRNEY, E.; BORK, P.; BREY, P. T.; VENTER, J. C.; WEISSENBAUGH, J.; KAFATOS, F. C.; COLLINS, F. H.; HOFFMAN, S. L. The genome sequence of the malaria mosquito *Anopheles gambiae*. **Science**, v. 298, n. 5591, p. 129-49, Oct 4 2002.
- HORNS, F.; PETIT, E.; HOOD, M. E. Massive Expansion of Gypsy-Like Retrotransposons in *Microbotryum Fungi*. **Genome Biol Evol**, v. 9, n. 2, p. 363-371, Feb 1 2017.
- HUANG, W.; LI, L.; MYERS, J. R.; MARTIN, G. T. ART: a next-generation sequencing read simulator. **Bioinformatics**, v. 28, n. 4, p. 593-4, Feb 15 2012.
- HUA-VAN, A.; LE ROUZIC, A.; MAISONHAUTE, C.; CAPY, P. Abundance, distribution and dynamics of retrotransposable elements and transposons: similarities and differences. **Cytogenet Genome Res**, v. 110, n. 1-4, p. 426-40, 2005.
- KELLEY, J. L.; PEYTON, J. T.; FISTON-LAVIER, A. S.; TEETS, N. M.; YEE, M. C.; JOHNSTON, J. S.; BUSTAMANTE, C. D.; LEE, R. E.; DENLINGER, D. L. Compact genome of the Antarctic midge is likely an adaptation to an extreme environment. **Nat Commun**, v. 5, p. 4611, Aug 12 2014.
- KIDWELL, M. G.; LISCH, D. R. Perspective: transposable elements, parasitic DNA, and genome evolution. **Evolution**, v. 55, n. 1, p. 1-24, Jan 2001.

- KOONIN, E. V.; WOLF, Y. I. Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. **Nucleic Acids Res**, v. 36, n. 21, p. 6688-719, Dec 2008.
- LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nat Methods**, v. 9, n. 4, p. 357-9, Mar 4 2012.
- LI, W.; GODZIK, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. **Bioinformatics**; 22:1658–1659, 2006.
- LENT, H.; WYGODZINSKY, P. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. **Bull Am Mus Nat Hist**, v. 163, p. 123-520, 1979.
- LERAT, E.; BURLET, N.; BIEMONT, C.; VIEIRA, C. Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. **Gene**, v. 473, n. 2, p. 100-9, Mar 1 2011.
- LOPEZ, G.; MORENO, J. Genetic variability and differentiation between populations of *Rhodnius prolixus* and *R. pallens*, vectors of Chagas' disease in Colombia. **Mem Inst Oswaldo Cruz**, v. 90, n. 3, p. 353-7, May-Jun 1995.
- LOPES, F. R.; JJINGO, D.; SILVA, C. R. Transcriptional Activity, Chromosomal Distribution and Expression Effects of Transposable Elements in Coffea Genomes. **Plos One**, v. 8, n. 11, p. e78931, 2013
- MAUMUS, F.; FISTON-LAVIER, ANNA-SOPHIE.; QUESNEVILLE, H. Impact of transposable elements on insect genomes and biology. **Current Opinion in Insect Science**, v.10, p. 35. 2015
- MARINO-RAMIREZ, L.; LEWIS, K. C.; LANDSMAN, D.; JORDAN, I. K. Transposable elements donate lineage-specific regulatory sequences to host genomes. **Cytogenet Genome Res**, v. 110, n. 1-4, p. 333-41, 2005.
- MENZEL, P.; NG, K. L.; KROGH, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. **Nat Commun**, v. 7, p. 11257, Apr 13 2016.
- MESQUITA, R. D.; VIONETTE-AMARAL, R. J.; LOWENBERGER, C.; RIVERA-POMAR, R.; MONTEIRO, F. A.; MINX, P.; SPIETH, J.; CARVALHO, A. B.; PANZERA, F.; LAWSON, D.; TORRES, A. Q.; RIBEIRO, J. M.; SORGINE, M. H.; WATERHOUSE, R. M.; MONTAGUE, M. J.; ABAD-FRANCH, F.; ALVES-BEZERRA, M.; AMARAL, L. R.; ARAUJO, H. M.; ARAUJO, R. N.; ARAVIND, L.; ATELLA, G. C.; AZAMBUJA, P.; BERNI, M.; BITTENCOURT-CUNHA, P. R.; BRAZ, G. R.; CALDERON-FERNANDEZ, G.; CARARETO, C. M.; CHRISTENSEN, M. B.; COSTA, I. R.; COSTA, S. G.; DANSA, M.; DAUMAS-FILHO, C. R.; DE-PAULA, I. F.; DIAS, F. A.; DIMOPOULOS, G.; EMRICH, S. J.; ESPONDA-BEHRENS, N.; FAMPA, P.; FERNANDEZ-MEDINA, R. D.; DA FONSECA, R. N.; FONTENELE, M.; FRONICK, C.; FULTON, L. A.; GANDARA, A. C.; GARCIA, E. S.; GENTA, F. A.; GIRALDO-CALDERON, G. I.; GOMES, B.; GONDIM, K. C.; GRANZOTTO, A.; GUARNERI, A. A.; GUIGO, R.; HARRY, M.; HUGHES, D. S.; JABLONKA, W.; JACQUIN-JOLY, E.; JUAREZ, M. P.; KOERICH, L. B.; LANGE, A. B.; LATORRE-ESTIVALIS, J. M.; LAVORE, A.; LAWRENCE, G. G.; LAZOSKI, C.; LAZZARI, C. R.; LOPES, R. R.; LORENZO, M. G.; LUGON, M. D.; MAJEROWICZ, D.; MARCET, P. L.; MARIOTTI, M.;

- MASUDA, H.; MEGY, K.; MELO, A. C.; MISSIRLIS, F.; MOTA, T.; NORIEGA, F. G.; NOUZOVA, M.; NUNES, R. D.; OLIVEIRA, R. L.; OLIVEIRA-SILVEIRA, G.; ONS, S.; ORCHARD, I.; PAGOLA, L.; PAIVA-SILVA, G. O.; PASCUAL, A.; PAVAN, M. G.; PEDRINI, N.; PEIXOTO, A. A.; PEREIRA, M. H.; PIKE, A.; POLYCARPO, C.; PROSDOCIMI, F.; RIBEIRO-RODRIGUES, R.; ROBERTSON, H. M.; SALERNO, A. P.; SALMON, D.; SANTESMASSES, D.; SCHAMA, R.; SEABRA-JUNIOR, E. S.; SILVA-CARDOSO, L.; SILVA-NETO, M. A.; SOUZA-GOMES, M.; STERKEL, M.; TARACENA, M. L.; TOJO, M.; TU, Z. J.; TUBIO, J. M.; URSIC-BEDOYA, R.; VENANCIO, T. M.; WALTER-NUNO, A. B.; WILSON, D.; WARREN, W. C.; WILSON, R. K.; HUEBNER, E.; DOTSON, E. M.; OLIVEIRA, P. L. Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. **Proc Natl Acad Sci USA**, v. 112, n. 48, p. 14936-41, Dec 01 2015.
- MODOLO, L.; LERAT, E. UrQt: an efficient software for the Unsupervised Quality trimming of NGS data. **BMC Bioinformatics**, v. 16, p. 137, Apr 29 2015.
- MONTEIRO, F. A.; BARRETT, T. V.; FITZPATRICK, S.; CORDON-ROSALES, C.; FELICIANGELI, D.; BEARD, C. B. Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. **Mol Ecol**, v. 12, n. 4, p. 997-1006, Apr 2003.
- NEAFSEY, D. E.; WATERHOUSE, R. M.; ABAL, M. R.; AGANEZOV, S. S.; ALEKSEYEV, M. A.; ALLEN, J. E.; AMON, J.; ARCA, B.; ARENSBURGER, P.; ARTEMOV, G.; ASSOUR, L. A.; BASSERI, H.; BERLIN, A.; BIRREN, B. W.; BLANDIN, S. A.; BROCKMAN, A. I.; BURKOT, T. R.; BURT, A.; CHAN, C. S.; CHAUVE, C.; CHIU, J. C.; CHRISTENSEN, M.; COSTANTINI, C.; DAVIDSON, V. L.; DELIGIANNI, E.; DOTTORINI, T.; DRITSOU, V.; GABRIEL, S. B.; GUELBEOGO, W. M.; HALL, A. B.; HAN, M. V.; HLAING, T.; HUGHES, D. S.; JENKINS, A. M.; JIANG, X.; JUNGREIS, I.; KAKANI, E. G.; KAMALI, M.; KEMPPAINEN, P.; KENNEDY, R. C.; KIRMITZOGLOU, I. K.; KOEKEMOER, L. L.; LABAN, N.; LANGRIDGE, N.; LAWNICZAK, M. K.; LIRAKIS, M.; LOBO, N. F.; LOWY, E.; MACCALLUM, R. M.; MAO, C.; MASLEN, G.; MBOGO, C.; MCCARTHY, J.; MICHEL, K.; MITCHELL, S. N.; MOORE, W.; MURPHY, K. A.; NAUMENKO, A. N.; NOLAN, T.; NOVOA, E. M.; O'LOUGHLIN, S.; ORINGANJE, C.; OSHAGHI, M. A.; PAKPOUR, N.; PAPATHANOS, P. A.; PEERY, A. N.; POVELONES, M.; PRAKASH, A.; PRICE, D. P.; RAJARAMAN, A.; REIMER, L. J.; RINKER, D. C.; ROKAS, A.; RUSSELL, T. L.; SAGNON, N.; SHARAKHOVA, M. V.; SHEA, T.; SIMAO, F. A.; SIMARD, F.; SLOTMAN, M. A.; SOMBOON, P.; STEGNIY, V.; STRUCHINER, C. J.; THOMAS, G. W.; TOJO, M.; TOPALIS, P.; TUBIO, J. M.; UNGER, M. F.; VONTAS, J.; WALTON, C.; WILDING, C. S.; WILLIS, J. H.; WU, Y. C.; YAN, G.; ZDOBNOV, E. M.; ZHOU, X.; CATTERUCCIA, F.; CHRISTOPHIDES, G. K.; COLLINS, F. H.; CORNMAN, R. S.; CRISANTI, A.; DONNELLY, M. J.; EMRICH, S. J.; FONTAINE, M. C.; GELBART, W.; HAHN, M. W.; HANSEN, I. A.; HOWELL, P. I.; KAFATOS, F. C.; KELLIS, M.; LAWSON, D.; LOUIS, C.; LUCKHART, S.; MUSKAVITCH, M. A.; RIBEIRO, J. M.; RIEHLE, M. A.; SHARAKHOV, I. V.; TU, Z.; ZWIEBEL, L. J.; BESANSKY, N. J. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. *Science*, v. 347, n. 6217, p. 1258522, Jan 2 2015.

- NOVAK, P.; NEUMANN, P.; PECH, J.; STEINHAISL, J.; MACAS, J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. **Bioinformatics**, v. 29, n. 6, p. 792-3, Mar 15 2013.
- NEUMANN, P.; KOBLIZKOVA, A.; NAVRATILOVA, A.; MACAS, J. Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. **Genetics** 173:1047–1056, mar 2006.
- PACE, J. K., 2ND; FESCHOTTE, C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. **Genome Res**, v. 17, n. 4, p. 422-32, Apr 2007.
- PATERSON, A. H.; BOWERS, J. E.; BRUGGMANN, R.; DUBCHAK, I.; GRIMWOOD, J.; GUNDLACH, H.; HABERER, G.; HELLSTEN, U.; MITROS, T.; POLIAKOV, A.; SCHMUTZ, J.; SPANNAGL, M.; TANG, H.; WANG, X.; WICKER, T.; BHARTI, A. K.; CHAPMAN, J.; FELTUS, F. A.; GOWIK, U.; GRIGORIEV, I. V.; LYONS, E.; MAHER, C. A.; MARTIS, M.; NARECHANIA, A.; OTILLAR, R. P.; PENNING, B. W.; SALAMOV, A. A.; WANG, Y.; ZHANG, L.; CARPITA, N. C.; FREELING, M.; GINGLE, A. R.; HASH, C. T.; KELLER, B.; KLEIN, P.; KRESOVICH, S.; MCCANN, M. C.; MING, R.; PETERSON, D. G.; MEHBOOB UR, R.; WARE, D.; WESTHOFF, P.; MAYER, K. F.; MESSING, J.; ROKHSAR, D. S. The Sorghum bicolor genome and the diversification of grasses. **Nature**, v. 457, n. 7229, p. 551-6, Jan 29 2009.
- PATTERSON, J. S.; GAUNT, M. W. Phylogenetic multi-locus codon models and molecular clocks reveal the monophyly of haematophagous reduviid bugs and their evolution at the formation of South America. **Mol Phylogenet Evol**, v. 56, n. 2, p. 608-21, Aug 2010.
- PAVAN, M. G.; CORREA-ANTONIO, J.; PEIXOTO, A. A.; MONTEIRO, F. A.; RIVAS, G. B. *Rhodnius prolixus* and *R. robustus* (Hemiptera: Reduviidae) nymphs show different locomotor patterns on an automated recording system. **Parasit Vectors**, v. 9, p. 239, Apr 27 2016.
- PAVAN, M. G.; MESQUITA, R. D.; LAWRENCE, G. G.; LAZOSKI, C.; DOTSON, E. M.; ABUBUCKER, S.; MITREVA, M.; RANDALL-MAHER, J.; MONTEIRO, F. A. A nuclear single-nucleotide polymorphism (SNP) potentially useful for the separation of *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). **Infect Genet Evol**, v. 14, p. 426-33, Mar 2013.
- PAVAN, M. G.; MONTEIRO, F. A. A multiplex PCR assay that separates *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). **Trop Med Int Health**, v. 12, n. 6, p. 751-8, Jun 2007.
- PINSKER, W.; HARING, E.; HAGEMANN, S.; MILLER, W. J. The evolutionary life history of P transposons: from horizontal invaders to domesticated neogenes. **Chromosoma**, v. 110, n. 3, p. 148-58, Jul 2001.
- PRITHAM, E. J.; FESCHOTTE, C. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. **Proc Natl Acad Sci USA**, v. 104, n. 6, p. 1895-900, Feb 6 2007.

- RAY, D. A.; FESCHOTTE, C.; PAGAN, H. J.; SMITH, J. D.; PRITHAM, E. J.; ARENSBURGER, P.; ATKINSON, P. W.; CRAIG, N. L. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. **Genome Res**, v. 18, n. 5, p. 717-28, May 2008.
- RIUS, N.; GUILLEN, Y.; DELPRAT, A.; KAPUSTA, A.; FESCHOTTE, C.; RUIZ, A. Exploration of the *Drosophila buzzatii* transposable element content suggests underestimation of repeats in *Drosophila* genomes. **BMC Genomics**, v. 17, p. 344, May 10 2016.
- SCHNABLE, P. S.; WARE, D.; FULTON, R. S.; STEIN, J. C.; WEI, F.; PASTERNAK, S.; LIANG, C.; ZHANG, J.; FULTON, L.; GRAVES, T. A.; MINX, P.; REILY, A. D.; COURTNEY, L.; KRUCHOWSKI, S. S.; TOMLINSON, C.; STRONG, C.; DELEHAUNTY, K.; FRONICK, C.; COURTNEY, B.; ROCK, S. M.; BELTER, E.; DU, F.; KIM, K.; ABBOTT, R. M.; COTTON, M.; LEVY, A.; MARCHETTO, P.; OCHOA, K.; JACKSON, S. M.; GILLAM, B.; CHEN, W.; YAN, L.; HIGGINBOTHAM, J.; CARDENAS, M.; WALIGORSKI, J.; APPLEBAUM, E.; PHELPS, L.; FALCONE, J.; KANCHI, K.; THANE, T.; SCIMONE, A.; THANE, N.; HENKE, J.; WANG, T.; RUPPERT, J.; SHAH, N.; ROTTER, K.; HODGES, J.; INGENTHRON, E.; CORDES, M.; KOHLBERG, S.; SGRO, J.; DELGADO, B.; MEAD, K.; CHINWALLA, A.; LEONARD, S.; CROUSE, K.; COLLURA, K.; KUDRNA, D.; CURRIE, J.; HE, R.; ANGELOVA, A.; RAJASEKAR, S.; MUELLER, T.; LOMELI, R.; SCARA, G.; KO, A.; DELANEY, K.; WISSOTSKI, M.; LOPEZ, G.; CAMPOS, D.; BRAIDOTTI, M.; ASHLEY, E.; GOLSER, W.; KIM, H.; LEE, S.; LIN, J.; DUJMIC, Z.; KIM, W.; TALAG, J.; ZUCCOLO, A.; FAN, C.; SEBASTIAN, A.; KRAMER, M.; SPIEGEL, L.; NASCIMENTO, L.; ZUTAVERN, T.; MILLER, B.; AMBROISE, C.; MULLER, S.; SPOONER, W.; NARECHANIA, A.; REN, L.; WEI, S.; KUMARI, S.; FAGA, B.; LEVY, M. J.; MCMAHAN, L.; VAN BUREN, P.; VAUGHN, M. W.; YING, K.; YEH, C. T.; EMRICH, S. J.; JIA, Y.; KALYANARAMAN, A.; HSIA, A. P.; BARBAZUK, W. B.; BAUCOM, R. S.; BRUTNELL, T. P.; CARPITA, N. C.; CHAPARRO, C.; CHIA, J. M.; DERAGON, J. M.; ESTILL, J. C.; FU, Y.; JEDDELOH, J. A.; HAN, Y.; LEE, H.; LI, P.; LISCH, D. R.; LIU, S.; LIU, Z.; NAGEL, D. H.; MCCANN, M. C.; SANMIGUEL, P.; MYERS, A. M.; NETTLETON, D.; NGUYEN, J.; PENNING, B. W.; PONNALA, L.; SCHNEIDER, K. L.; SCHWARTZ, D. C.; SHARMA, A.; SODERLUND, C.; SPRINGER, N. M.; SUN, Q.; WANG, H.; WATERMAN, M.; WESTERMAN, R.; WOLFGRUBER, T. K.; YANG, L.; YU, Y.; ZHANG, L.; ZHOU, S.; ZHU, Q.; BENNETZEN, J. L.; DAWE, R. K.; JIANG, J.; JIANG, N.; PRESTING, G. G.; WESSLER, S. R.; ALURU, S.; MARTIENSSEN, R. A.; CLIFTON, S. W.; MCCOMBIE, W. R.; WING, R. A.; WILSON, R. K. The B73 maize genome: complexity, diversity, and dynamics. **Science**, v. 326, n. 5956, p. 1112-5, Nov 20 2009.
- SESSEGOLO, C.; BURLET, N.; HAUDRY, A. Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. **Biol Lett**, v. 12, n. 8, Aug 2016.
- STATON, S. E.; BAKKEN, B. H.; BLACKMAN, B. K.; CHAPMAN, M. A.; KANE, N. C.; TANG, S.; UNGERER, M. C.; KNAPP, S. J.; RIESEBERG, L. H.; BURKE, J. M. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. **Plant J**, v. 72, n. 1, p. 142-53, Oct 2012.

- SCHOFIELD, C.J. and DUJARDIN, J. Theories on the evolution of *Rhodnius*. **Biol** v. 71 p. 183-197, Oct 1999.
- THOMAS-BULLE, C.; PIEDNOËL, M.; DONNART, T.; FILÉE, J.; JOLLIVET, D.; BONNIVARD, É. Mollusc genomes reveal variability in patterns of LTR-retrotransposons dynamics. **BMC Genomics**, v. 19, n. 1, p. 821, 2018/11/15 2018.
- VAN DE LAGEMAAT, L. N. et al. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. **Trends Genet**, v. 19, n. 10, p. 530–536, 2003.
- VILLEGAS, J.; FELICIANGLI, M. D.; DUJARDIN, J. P. Wing shape divergence between *Rhodnius prolixus* from Cojedes (Venezuela) and *Rhodnius robustus* from Merida (Venezuela). **Infect Genet Evol**, v. 2, n. 2, p. 121-8, Dec 2002.
- VOLFF, J. N.; KORTING, C.; MEYER, A.; SCHARTL, M. Evolution and discontinuous distribution of Rex3 retrotransposons in fish. **Mol Biol Evol**, v. 18, n. 3, p. 427-31, Mar 2001.
- VOLFF, J. N.; KORTING, C.; SCHARTL, M. Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. **Mol Biol Evol**, v. 17, n. 11, p. 1673-84, Nov 2000.
- WANG, S.; LORENZEN, M. D.; BEEMAN, R. W.; BROWN, S. J. Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. **Genome Biol**, v. 9, n. 3, p. R61, 2008.
- WANG, X.; FANG, X.; YANG, P.; JIANG, X.; JIANG, F.; ZHAO, D.; LI, B.; CUI, F.; WEI, J.; MA, C.; WANG, Y.; HE, J.; LUO, Y.; WANG, Z.; GUO, X.; GUO, W.; WANG, X.; ZHANG, Y.; YANG, M.; HAO, S.; CHEN, B.; MA, Z.; YU, D.; XIONG, Z.; ZHU, Y.; FAN, D.; HAN, L.; WANG, B.; CHEN, Y.; WANG, J.; YANG, L.; ZHAO, W.; FENG, Y.; CHEN, G.; LIAN, J.; LI, Q.; HUANG, Z.; YAO, X.; LV, N.; ZHANG, G.; LI, Y.; WANG, J.; WANG, J.; ZHU, B.; KANG, L. The locust genome provides insight into swarm formation and long-distance flight. **Nature Communications**, v. 5, p. 2957, 01/14/online 2014.
- WALLAU, G. L.; CAPY, P.; LORETO, E.; HUA-VAN, A. Genomic landscape and evolutionary dynamics of Mariner transposable elements within the *Drosophila* genus. **BMC Genomics**, v. 15, p. 727, Aug 27 2014.
- WALLAU, G. L.; CAPY, P.; LORETO, E.; LE ROUZIC, A.; HUA-VAN, A. VHICA, a New Method to Discriminate between Vertical and Horizontal Transposon Transfer: Application to the Mariner Family within *Drosophila*. **Mol Biol Evol**, v. 33, n. 4, p. 1094-109, Apr 2016.
- WICKER, T.; SABOT, F.; HUA-VAN, A.; BENNETZEN, J. L.; CAPY, P.; CHALHOUB, B.; FLAVELL, A.; LEROY, P.; MORGANTE, M.; PANAUD, O.; PAUX, E.; SANMIGUEL, P.; SCHULMAN, A. H. A unified classification system for eukaryotic transposable elements. **Nat Rev Genet**, v. 8, n. 12, p. 973-82, Dec 2007.
- ZHANG, H. H.; XU, H. E.; SHEN, Y. H.; HAN, M. J.; ZHANG, Z. The origin and evolution of six miniature inverted-repeat transposable elements in *Bombyx mori* and *Rhodnius prolixus*. **Genome Biol Evol**, v. 5, n. 11, p. 2020-31, 2013.

8. ANEXOS

8.1 ANEXO 1 – Tabelas suplementares

Tabela Suplementar 1. Número de pares de bases (pb) e proporção (%) das sequências repetitivas (TEs, repetições simples de DNA, DNA satélite, DNA de baixa complexidade, rRNA outros DNA) anotadas usando *dnaPipeTE* no genoma de *R. prolixus* e *R. robustus* II - III.

	<i>R. prolixus</i>		<i>R. robustus II</i>		<i>R. robustus III</i>	
	Pb	%	Pb	%	Pb	%
LTR	1,369,283	0,8	2,063,977	1,1	1,513,639	0,8
LINE	9,036,881	5,0	10,644,012	5,8	9,067,148	5,0
SINE	2,804,457	1,5	3,133,083	1,7	3,145,279	1,7
DNA	18,481,976	10,2	22,913,643	12,6	21,255,599	11,7
Helitron	2,688,281	1,5	4,118,237	2,3	3,704,583	2,0
rRNA	54,665	0,0	56,709	0,0	99,074	0,1
DNA de Baixa complexidade	5,668	0,0	8,794	0,0	17,928	0,0
DNA satélite	7,416,288	4,1	4,984,264	2,7	7,529,581	4,1
DNA repetição simples	62,168	0,0	93,037	0,1	41,712	0,0
Outros	1,131,288	0,6	555,406	0,3	565,611	0,3
Desconhecido	11,565,271	6,4	40,061,977	22,0	14,704,919	8,1
DNA de Cópia simples	126,746,195	69,9	93,465,475	51,3	120,448,752	66,1
Total	181,362,421	100,0	182,098,614	100,0	182,093,825	100,0

pb = Pares de bases alinhados; % = Porcentagem dos pares de bases alinhadas por TEs.

Tabela Suplementar 2. Distribuições das superfamílias nos genomas de espécies de *Rhodnius*.

Classe	Ordem	Superfamilias	<i>R.prolixus</i>			<i>R.robustus</i> II			<i>R.robustus</i> III												
			*pb	% TEs	% G	*pb	% TEs	% G	* pb	% TEs	% G										
CLASSE I	LTR	<i>Bel-Pao</i>	201352	0,59	0,111	300928	0,70	0,165	175680	0,45	0,096										
		<i>Copia</i>	98637	0,29	0,054	150535	0,35	0,083	86906	0,22	0,048										
		<i>Gypsy</i>	1045433	3,06	0,576	1570001	3,66	0,862	1220975	3,12	0,671										
		<i>LTR</i>	4111	0,01	0,002	4344	0,01	0,002	5740	0,01	0,003										
		<i>Ngaro</i>	2125	0,01	0,001	1619	0,00	0,001	2217	0,01	0,001										
		<i>EVR</i>	11241	0,03	0,006	21791	0,05	0,012	11463	0,03	0,006										
		<i>Outras</i>	4587	0,01	0,003	12562	0,03	0,007	3866	0,01	0,002										
		<i>DIRS</i>	1428	0,00	0,001	2197	0,01	0,001	1451	0,00	0,001										
	Sub total LTR		1,368,914	4,01	0,755	2,063,977	4,81	1,133	1,508,298	3,85	0,828										
	PLE	<i>Penelope</i>	32756	0,10	0,018	59125	0,14	0,032	35225	0,09	0,019										
		<i>Dong</i>	15534	0,05	0,009	8918	0,02	0,005	13833	0,04	0,008										
		<i>I</i>	1676573	4,91	0,924	1662201	3,88	0,913	1294364	3,31	0,711										
		<i>Jockey</i>	4061974	11,89	2,240	5349244	12,48	2,938	3823545	9,77	2,100										
		<i>L1</i>	33382	0,10	0,018	43402	0,10	0,024	148498	0,38	0,082										
		<i>L2</i>	440876	1,29	0,243	568084	1,32	0,312	1583	0,00	0,001										
		<i>LOA</i>	1358805	3,98	0,749	1440781	3,36	0,791	1402359	3,58	0,770										
		<i>R1</i>	159236	0,47	0,088	135272	0,32	0,074	160032	0,41	0,088										
		<i>R2</i>	5078	0,01	0,003	5428	0,01	0,003	12974	0,03	0,007										
		<i>RTE</i>	1195745	3,50	0,659	1261756	2,94	0,693	1627588	4,16	0,894										
		<i>CR1</i>	28858	0,08	0,016	40172	0,09	0,022	12194	0,03	0,007										
		<i>Tad</i>	0	0,00	0,000	26768	0,06	0,015	0	0,00	0,000										
		<i>CRE</i>	1070	0,00	0,001	1414	0,00	0,001	1018	0,00	0,001										
		<i>Outras</i>	3444	0,01	0,002	41491	0,10	0,023	1682	0,00	0,001										
		Sub-total não-LTR		9013331	26,38	4,97	10644056	24,83	5,84	8534895	21,80	4,69									
	SINEs	<i>tRNA</i>	2803303	8,21	1,546	3115727	7,27	1,711	3141039	8,02	1,725										
		<i>Outras</i>	1316	0,00	0,001	18637	0,04	0,010	397	0,00	0,000										
	Sub-total SINEs		2804619	8,21	1,55	3134364	7,31	1,72	3141436	8,02	1,73										
CLASSE II	TIR	<i>Tc1/Mariner</i>	9533907	27,91	5,257	11769349	27,45	6,463	9679802	24,73	5,316										
		<i>hAT</i>	4586181	13,42	2,529	5930663	13,83	3,257	6866612	17,54	3,771										
		<i>P</i>	7644	0,02	0,004	9288	0,02	0,005	7257	0,02	0,004										
		<i>PiggyBac</i>	517632	1,52	0,285	852261	1,99	0,468	223120	0,57	0,123										
		<i>Academ</i>	18926	0,06	0,010	20877	0,05	0,011	25048	0,06	0,014										
		<i>DNA</i>	1874250	5,49	1,033	2253378	5,26	1,237	1940122	4,96	1,065										
		<i>CMC</i>	74028	0,22	0,041	99206	0,23	0,054	96095	0,25	0,053										
		<i>Ginger</i>	20225	0,06	0,011	16173	0,04	0,009	31626	0,08	0,017										
		<i>Kolobok</i>	42928	0,13	0,024	63921	0,15	0,035	49243	0,13	0,027										
		<i>Merlin</i>	115636	0,34	0,064	92058	0,21	0,051	77697	0,20	0,043										
		<i>MULE</i>	38069	0,11	0,021	36997	0,09	0,020	46164	0,12	0,025										
		<i>PIF</i>	549999	1,61	0,303	594789	1,39	0,327	198132	0,51	0,109										
		<i>Sola</i>	150560	0,44	0,083	173582	0,40	0,095	211840	0,54	0,116										
		<i>Zator</i>	7337	0,02	0,004	12289	0,03	0,007	1868	0,00	0,001										
		<i>Zisupton</i>	0	0,00	0,000	0	0,00	0,000	1712	0,00	0,001										
		<i>Outras</i>	3972	0,01	0,002	121702	0,28	0,067	2685	0,01	0,001										
		Crypton	<i>Crypton</i>	251886	0,74	0,139	379108	0,88	0,208	338279	0,86	0,186									
			<i>Helitron</i>	2565502	7,51	1,415	4118237	9,61	2,262	5521963	14,11	3,032									
	<i>Maverick</i>		616445	1,80	0,340	489341	1,14	0,269	644424	1,65	0,354										
	Sub-total DNA		20,975,127	61,40	11,56	27,033,219	63,05	14,84	25,963,689	66,32	14,26										
	Total TEs			34161991			19			42875616			23,5			39148318			21,5		
	Outros (não TEs)			147200430			81			139222998			76,5			142945507			78,5		
Total do genoma			181,362,421			100			182,098,614			100			182,093,825			100			

* % G = Porcentagem genômica; % TEs = Porcentagem dos TEs; * pb = pares de bases alinhados (BLASTN 2)

Tabela Suplementar 3. Famílias de TEs mais abundantes identificadas (> 0.10 %) nos genomas de espécies de *Rhodnius*.

<i>R. prolixus</i>					<i>R. robustus II</i>					<i>R. robustus III</i>				
%G	%TE	N	Superfamília RM	Anotação RM	%G	%TE	N	Superfamília RM	Anotação RM	%G	%TE	N	Superfamília RM	Anotação RM
0.38	2.01	7108	RC/Helitron	Helitron-2N1	0.54	2.31	123675	DNA/TcMar-Mariner	Mariner-1*	1.18	5.92	14002	DNA/TcMar-Mariner	Mariner-1*
0.30	1.58	5626	DNA/TcMar-Mariner	Mariner-1*	0.44	1.91	44262	RC/Helitron	Helitron-2N1	0.45	2.27	5464	RC/Helitron	Helitron-2N1
0.30	1.55	5637	SINE/tRNA	SINE2-2	0.36	1.56	273628	DNA/TcMar-Mariner	Mariner-1*	0.28	1.40	3484	DNA/TcMar-Mariner	Mariner-1*
0.28	1.47	5275	LINE/I	I-5	0.34	1.46	174427	DNA/TcMar-Mariner	Mariner-1*	0.21	1.08	2723	LINE/I	I-5
0.28	1.45	5613	DNA/TcMar-Mariner	Mariner-1*	0.29	1.24	288162	DNA/TcMar-Mariner	Mariner-1*	0.20	1.04	2459	RC/Helitron	Helitron-2N1
0.25	1.29	4715	SINE/tRNA	SINE2-2	0.28	1.23	250592	DNA/TcMar-Mariner	Mariner-1*	0.19	0.99	2425	SINE/tRNA	SINE2-2
0.23	1.22	4555	DNA/TcMar-Mariner	Mariner-N9B*	0.26	1.14	72624	LINE/I	I-5	0.17	0.89	2108	DNA/TcMar-Mariner	Mariner-N9B*
0.23	1.21	4267	DNA/TcMar-Mariner	Mariner-1*	0.26	1.14	85525	SINE/tRNA	SINE2-1	0.16	0.84	2011	SINE/tRNA	SINE2-2
0.20	1.08	3953	DNA/TcMar-Mariner	Mariner-1*	0.22	0.96	279671	DNA/TcMar-Mariner	Mariner-1*	0.16	0.82	2254	SINE/tRNA	SINE2-1
0.20	1.07	3990	DNA/TcMar-Mariner	Mariner-1*	0.22	0.96	132097	DNA	DNA4-1	0.16	0.82	2063	RC/Helitron	Helitron-2N3
0.20	1.04	4168	DNA4-1	-	0.20	0.89	182993	RC/Helitron	Helitron-2N3	0.15	0.75	1788	DNA/hAT-Tip100	hAT-N27
0.17	0.91	3208	DNA/TcMar-Mariner	Mariner-1*	0.20	0.86	63156	DNA/TcMar-Tc1	Mariner-N17	0.15	0.74	1866	LINE/R1-LOA	Loa-2
0.16	0.86	3659	DNA/TcMar-Mariner	Mariner-1*	0.19	0.84	112036	DNA/TcMar-Mariner	Mariner-N9B*	0.14	0.73	1859	DNA/TcMar-Mariner	Mariner-N9*
0.16	0.83	3213	DNA4-1	-	0.18	0.78	77323	LINE/R1-LOA	Loa-2	0.14	0.72	2156	DNA4-1	-
0.16	0.82	3117	SINE/tRNA	SINE2-2	0.16	0.72	159441	SINE/tRNA	SINE2-1	0.13	0.67	1555	RC/Helitron	Helitron-2
0.15	0.78	2941	LINE/I-Jockey	Jockey-2	0.16	0.68	194765	DNA/TcMar-Mariner	Mariner-N9B*	0.13	0.67	1614	DNA/TcMar-Mariner	SMAR26
0.15	0.77	2810	SINE/tRNA	SINE2-2	0.15	0.68	167441	SINE/tRNA	SINE2-1	0.13	0.66	1667	DNA/TcMar-Mariner	Mariner-N11
0.14	0.73	2633	DNA/TcMar-Mariner	Mariner-1*	0.15	0.66	113816	DNA/TcMar-Mariner	Mariner-N9*	0.13	0.64	1487	DNA/Maverick	Polinton-3
0.13	0.69	2819	DNA/TcMar-Mariner	Mariner-1*	0.14	0.61	141776	RC/Helitron	Helitron-2N1	0.12	0.63	1540	DNA/TcMar-Mariner	SMAR26
0.13	0.67	2410	DNA/PIF-ISL2EU	ISL2EU-N1	0.14	0.60	200329	DNA	DNA4-1	0.12	0.62	1586	SINE/tRNA	SINE2-2
0.12	0.65	2525	SINE/tRNA	SINE2-1	0.13	0.58	47096	LINE/I-Jockey	Jockey-6	0.11	0.56	1322	DNA/TcMar-Tc1	Mariner-N39
0.12	0.65	2347	RC/Helitron	Helitron-2N3	0.13	0.57	30909	LINE/I-Jockey	Jockey-2	0.11	0.56	1347	DNA/TcMar-Tc1	DNA-2
0.12	0.62	2438	RC/Helitron	Helitron-2N1	0.13	0.56	69149	DNA/PIF-ISL2EU?	ISL2EU-N1	0.10	0.53	1266	DNA/TcMar-Mariner	Mariner-1*
0.12	0.62	2525	SINE/tRNA	SINE2-1	0.13	0.56	184500	RC/Helitron	Helitron-2N1	0.10	0.53	1331	RC/Helitron	Helitron-2N1
0.12	0.61	2428	SINE/tRNA	SINE2-1	0.12	0.53	153363	DNA/TcMar-Tc1	DNA-2	0.10	0.53	1289	RC/Helitron	Helitron-2N3
0.11	0.57	2105	RC/Helitron	Helitron-2N3	0.12	0.53	223695	DNA/TcMar-Mariner	Mariner-1*	0.10	0.51	1227	DNA/TcMar-Tc1	Mariner-N17
0.11	0.56	2083	RC/Helitron	Helitron-2N1	0.12	0.52	1480	LINE/I	I-5	0.10	0.48	1274	DNA/TcMar-Tc1	DNA-4
					0.11	0.50	125086	LINE/I-Jockey	Jockey-2					
					0.11	0.48	163160	LINE/I-Jockey	Jockey-1					

%G= Porcentagem genômico; %TE= Porcentagem dos TEs; N=Numero de copias estimado; **Superfamília RM**=Superfamília segundo RepeatMasker; **Anotação RM**= Anotação segundo RepeatMasker. : *clados mais representativos na família Mariner

8.2. ANEXO 2 - Figuras suplementares

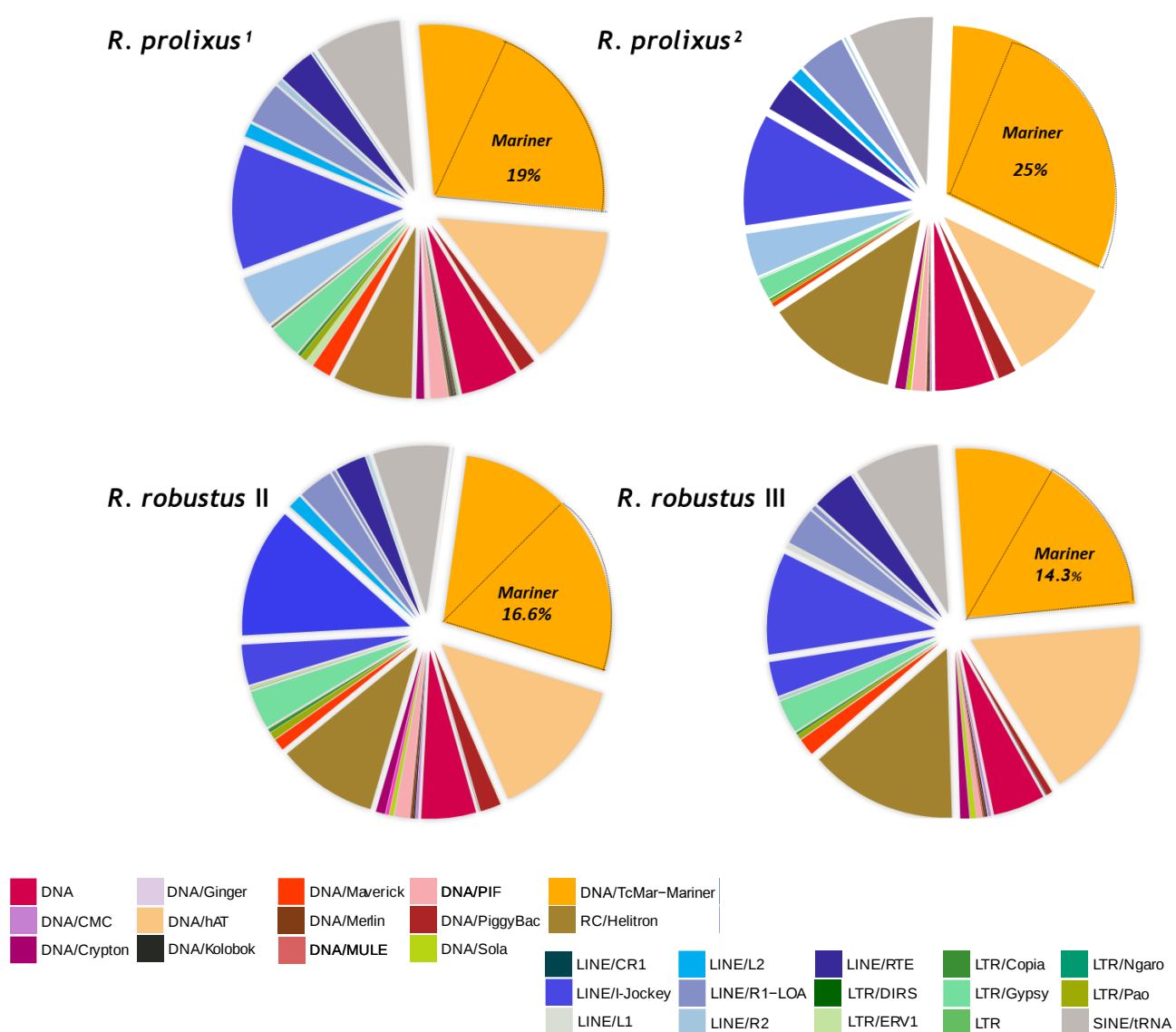


Figura suplementar 1. Proporção genômica relativa das principais famílias de TEs: * *R. prolixus*¹ = A partir das reads genômicas brutas; *R. prolixus*² = A partir do genoma montado

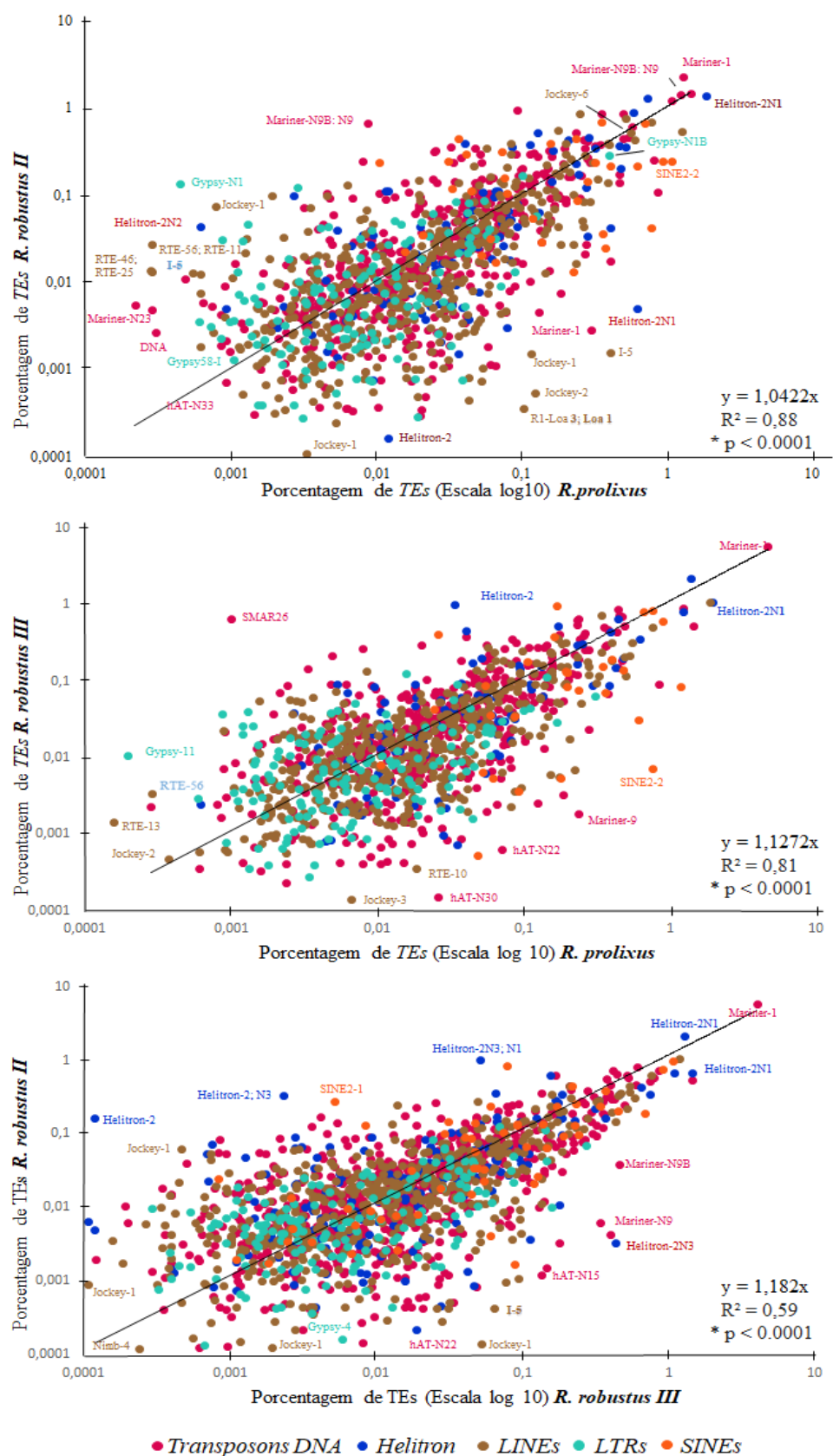


Figura Suplementar 2. Comparação da proporção genômica relativa das famílias de TEs compartilhadas entre *R. prolixus*, *R. robustus II* e *R. robustus III*, em termos de porcentagem genômica (escala log₁₀) : *Correlação significativa (valor-P ≤ 0,01)