

UNIVERSIDADE ESTADUAL DE CAMPINAS INSTITUTO DE BIOLOGIA

TIAGO BELINTANI

ESTUDO DE TRANSCRIPTOMAS POR RNA-SEQ EM TECIDOS DE CABEÇA E GLÂNDULA SALIVAR DE TRÊS ESPÉCIES DE *MEPRAIA* (HEMIPTERA, REDUVIIDAE, TRIATOMINAE)

> CAMPINAS 2022

TIAGO BELINTANI

ESTUDO DE TRANSCRIPTOMAS POR RNA-SEQ EM TECIDOS DE CABEÇA E GLÂNDULA SALIVAR DE TRÊS ESPÉCIES DE *MEPRAIA* (HEMIPTERA, REDUVIIDAE, TRIATOMINAE)

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para obtenção do Título de Doutor em Biologia Animal: "Relações Antrópicas, Meio Ambiente e Parasitologia".

Orientador: Prof. Dr. João Aristeu da Rosa

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO DR. TIAGO BELINTANI E ORIENTADO POR PROF. DR. JOÃO ARISTEU DA ROSA. Ficha catalográfica Universidade Estadual de Campinas Biblioteca do Instituto de Biologia Mara Janaina de Oliveira - CRB 8/6972

Belintani, Tiago, 1986B412e Estudos de transcriptomas por RNA-seq em tecidos de cabeça e glândulas salivares de três espécies de *Mepraia* (Hemiptera, Reduviidae, Triatominae) / Tiago Belintani. – Campinas, SP: [s.n.], 2022.
Orientador: João Aristeu da Rosa. Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.
1. Transcriptoma. 2. Filogenia. 3. Polimorfismo de nucleotídeo único. 4. RNA-seq. 5. DNA mitocondrial. I. Rosa, João Aristeu. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.<u>Informações para Biblioteca Digital</u>

Título em outro idioma: Study of transcriptomes by RNA-seq in head tissues and salivary glands of three species of *Mepraia* (Hemiptera, Reduviidae, Triatominae) Palavras-chave em inglês: Transcriptome Phylogeny Single nucleotide polymorphism RNA-seq DNA, Mitochondrial Área de concentração: Relações Antrópicas, Meio Ambiente e Parasitologia Titulação: Doutor em Biologia Animal Banca examinadora: João Aristeu da Rosa [orientador] Carlos Eduardo Almeida Carlos Congrains Castillo Danilo Ciccone Miguel Samira Chahad Ehlers Data de defesa: 10-06-2022 Programa de Pós-Graduação: Biologia Animal

Identificação e informações acadêmicas do(a) aluno(a) ORCID do autor: https://orcid.org/0000-0001-9485-1145 Currículo Lattes do autor: http://lattes.cnpq.br/8948388411348223

Campinas, 10 de junho de 2022.

COMISSÃO EXAMINARDORA

Prof. Dr. João Aristeu da Rosa.

Prof. Dr. Carlos Eduardo Almeida.

Dr. Carlos Congrains Castillo.

Prof. Dr. Danilo Ciccone Miguel.

Dra. Samira Chahad Ehlers

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica do aluno.

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Tese e na Secretaria do Programa de Pós-Graduação em Biologia Animal do Instituto de Biologia da Universidade Estadual de Campinas.

Dedicatória

Aos meus pais, irmãos e em especial a minha companheira Camila Azevedo. A todas as pessoas que de alguma forma tornaram esta tese possível em especial ao meu Tutor Prof. Dr. João Aristeu da Rosa por todos os ensinamentos. As pessoas que perdemos durante a pandemia.

AGRADECIMENTOS

Agradeço aos meu país que sempre me apoiaram e me deram condições para que eu pudesse concluir meus sonhos. Obrigado, por serem pessoas maravilhosas. Agradeço a minha companheira, Camila, pelo apoio, pela paciência, sem você não seria

possível.

Aos meus irmãos por sempre terem me apoiado.

Ao meu mentor, Prof. Dr. João Aristeu da Rosa. Agradeço por todos os ensinamos colaborações, conversas, muito obrigado. Sem o Doutor o estudo não seria possíve Obrigado por me ensinar como a ciência deve ser ética e útil para sociedade. Obrigado.

Ao meu tutor de bioinformática, Dr. Carlos Congrains Castillo. Por sua paciência, por todos os ensinamentos, discussões e por estar sempre disponível. Obrigado, sua ajuda foi essencial para a conclusão do projeto.

Ao Prof. Dr. Reinaldo Alvares Brito por ter me recebido em seu laboratório dando condições e liberdade para realizar os experimentos e especialmente pelo acesso a estrutura computacional. Muito Obrigado.

A Dra. Samira Chahad Ehlers, minha tutora e colaborada nas extrações de RNA. Obrigado por ser tão gentil e solícita.

A Dr. Heloisa Pinotti pela colaboração, amizade e apoio. Obrigado pela ajuda, principalmente nas extrações de RNA, pela paciência e por ser esta pessoa incrível. Agradeço aos colaboradores do Laboratório de Genética de Populações e Evolução da Universidade Federal de São Carlos em especial a Profa. Dra. Andréa Cristina Peripato, a Dra. Maria Isabel Camargo Isler Tscherne e a aluna de doutorado Edyane Moraes dos Santos.

Agradeço aos doutores Dr. Daniel Frias-Lasserre e Dr. Jader Oliveira pelas coletas dos espécimes utilizados no estudo. Obrigado.

A equipe do Laboratório de Parasitologia da Faculdade Ciências Farmacêuticas de Araraquara, em especial aos colegas Allef, André e Vinicius.

A Profa Mara Cristina Pinto do Laboratório de Parasitologia da Faculdade Ciências Farmacêuticas de Araraquara e toda sua equipe. Obrigado por toda ajuda ao longo dos últimos anos.

Aos membros da banca de defesa e aos suplentes. É muito gratificante ter pessoas tão qualificadas e preparadas para colaborar com minha defesa.

Agradecemos a Fundação de Amparo à Pesquisa do Estado de São Paulo, processo: 17/06460-4, que financiou o projeto.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

"Tudo o que você realmente precisa saber no momento é que o universo é muito mais complicado do que se poderia pensar" Douglas Adams

RESUMO

A doença de Chagas é uma antropozoonose negligenciada causada pelo protozoário *Trypanosoma cruzi*. A principal forma de transmissão é o contato com fezes ou urina de insetos hematófagos da subfamília Triatominae. Porém, o gênero Mepraia que conta com M. gajardoi, M. parapatrica e M. spinolai tem um destaque especial pois todas as três espécies têm ocorrências registradas somente para o Chile onde participam da dinâmica de transmissão da doença de Chagas. Este estudo investigou as diferenças genéticas interespecíficas de M. gajardoi, M. parapatrica e M. spinolai por meio de sequenciamento de larga escala, RNA-Seq. Por essa técnica foram obtidas 18 bibliotecas de RNA total de cabeças e glândulas salivares de ninfas machos de quinto estádio de M. gajardoi, M. parapatrica e M. spinolai. As bibliotecas foram utilizadas para montar 18 transcriptomas que tiveram seu perfil caracterizado. Além disso, a partir de transcriptoma comparativo foi possível determinar as características transcricionais de M. gajardoi, M. parapatrica e M. spinolai. Foram estimadas as taxas de divergências genéticas a partir de um conjunto de SNPs. Os SNPs foram utilizados para traçar as relações interespecíficas por meio de estimativas de árvores de espécies. A partir dos dados brutos do sequenciamento foram montados doze genomas mitocondriais por estratégia de referencia "isca". Os mitogenomas obtidos foram anotados, caracterizados e utilizados na reconstrução filogenética. Os mitogenomas forneceram filogenias coerentes com as disponíveis na literatura. Genes ortólogos foram descritos e utilizados para estimar a filogenia com as especies do estudo. As filogenias estimadas por meio de máxima verossimilhança e inferência bayesiana recuperam resultados com suporte elevado e permitiram delimitar as relações filogenéticas interespecíficas e intraespecíficas de Mepraia. A partir de dados de larga escala foi possível caracterizar transcriptoma de cabeças e glândulas salivares de ninfas de quinto estádio de Mepraia. Além da descrição completa do perfil transcricional foi possível determinar com clareza as relações evolutivas de Mepraia.

Palavras de chaves: Filogenômica, SNP, RNA-Seq, mtDNA, Transcriptoma.

ABSTRACT

Chagas disease is a neglected anthropozoonosis caused by the protozoan Trypanosoma cruzi. The main form of transmission is contact with feces or urine of hematophagous insects of the subfamily Triatominae. The main form of transmission is contact with feces or urine of hematophagous insects of the subfamily Triatominae. However, the genus Mepraia, which includes M. gajardoi, M. parapatrica and M. spinolai, has a special prominence because all three species have occurrences recorded only for Chile where they participate in the dynamics of Chagas disease. This study proposes to investigate the interspecific genetic differences of M. gajardoi, M. parapatrica and M. spinolai through large-scale RNA-Seq. By this technique, 18 libraries of total RNA from heads and salivary glands of fifth-stage male nymphs of M. gajardoi, M. parapatrica and M. spinolai were obtained. The libraries were used to assemble 18 transcriptomes that had their profile characterized. Furthermore, from a comparative transcriptome it was possible to determine the transactional characteristics of *M. gajardoi*, *M.* parapatrica and M. spinolai. Genetic divergence rates from SNP were described and determined. SNP were used to trace interspecific relationships by estimating species trees. Twelve mitochondrial genomes were assembled from the raw sequencing data using a "bait" reference strategy. The mitogenomes obtained were annotated, characterized, and used in the phylogenetic reconstruction. The mitogenomes obtained are consistent with other species of Triatominae, as well as animals. Mitogenomes provided phylogenies consistent with those available in the literature. The orthologous genes were determined and a phylogenomic study was performed with the sequences obtained. The phylogenies estimated through maximum likelihood and Bayesian inference retrieve results with high support and allowed to delimit the interspecific and intraspecific phylogenetic relationships of *Mepraia*. The results of this study show that from large-scale data it was possible to characterize transcriptome of heads and salivary glands of fifth instar Mepraia nymphs. In addition to the complete description of the transcriptional profile, it was possible to clearly determine the evolutionary relationships of Mepraia.

Keywords: Mepraia, Phylogenomics, SNP, RNA-Seq, mtDNA, Transcriptome

1. INTRODUÇÃO	13
1.1 Doença de Chaga	13
1.2 Subfamília Triatominae	14
1.3 Gênero Mepraia Mazza, Gajardo e Jorg (1940)	17
1.4 Ecoepidemiologia da doença de Chagas no Chile	22
1.5 Filogenia e especiação de <i>Mepraia</i>	22
1.6 Perspectivas e panorama geral do estudo de transcriptoma	26
1.7 Breves considerações sobre genômica e filogenômica	29
2. JUSTIFICATIVA	
3. OBJETIVOS	34
3.1 Objetivos específicos	34
4. MATERIAL E MÉTODOS	
4.1 Espécimes.	
4.2 Dissecação das três espécies de Mepraia	
4.3 Extração do RNA total	
4.4 Purificação do RNA	
4.5 Sequenciamento por RNA-Seq.	
4.6 Controle de qualidade e montagem das sequências	40
4.7 Alinhamento das Sequências.	41
4.8 Previsão de unigenes e avaliação da qualidade das montagens	41
4.9 Anotação funcional	42
4.10 Busca de Variantes	43
4.11 Inferência de ortologia e análise filogenética	44
4.12 Estimativa de árvores de espécies	45
4.13 Montagem, anotação e filogenia com mitogenomas.	47
4.14 Teste de introgressão.	49
5. RESULTADOS	

SUMÁRIO

5.1 Extração do RNA total	49
5.2 Determinação da integridade do RNA, Bibliotecas de cDNA e Controle de	
Qualidade de Dados	49
5.3. Montagem dos transcriptomas de novo	50
5.4. Anotação funcional	51
5.5 Inferência de ortologia e análise filogenética.	53
5.6 Estimativa de árvores de espécies	58
5.7 Montagem, anotação e filogenia de mitogenomas.	62
5.8 Teste de introgressão	72
6. DISCUSSÃO	72
7. CONCLUSÕES	82
8. REFÊRENCIAS BIBLIOGRÁFICAS	83
9. ANEXOS	107

1. INTRODUÇÃO

1.1 Doença de Chagas

A doença de Chagas é uma das principais doenças tropicais negligenciadas da atualidade (WHO, 2022). A enfermidade é endêmica nas Américas onde possui ampla distribuição, que se estende desde as regiões austrais da Argentina até o sul dos Estados Unidos (MURILLO-GODÍNEZ, 2018). A doença de Chagas ou tripanossomíase americana é causada pelo hemoflagelado *Trypanosoma cruzi* (Chagas, 1909), a transmissão ocorre principalmente pelo contato com fezes ou urina de insetos da subfamília Triatominae (LENT; WYGODZINSKY, 1979).

A doença foi descrita pelo médico e epidemiologista brasileiro Carlos Ribeiro Justiniano Chagas nos primeiros anos do século XX (CHAGAS, 1909; LENT; WYGODZINSKY, 1979). O médico brasileiro realizou um feito inédito para ciência, contando com perspicácia investigativa foi responsável por descrever o ciclo da enfermidade, assim como o agente etiológico. Também identificou o vetor e hospedeiros, além do ciclo epidemiológico e manifestações clínicas em humanos (CHAGAS 1909a, 1909b; GALVÃO, 2014). Desde a descrição da doença já se passaram mais de 100 anos e o contexto da doença continua relevante e merece atenção. Embora vários avanços no controle ou em tratamento à enfermidade tenham surgido, a infecção, hoje é uma das principais endemias da região Neotropical, responsável por um notável impacto social (WHO, 2020).

De acordo com a Organização Mundial da Saúde (OMS), estima-se que de 6 a 7 milhões de pessoas estejam infectadas com *T. cruzi* no mundo (WHO, 2020). Em consideração à mortalidade na América Latina, estima-se que anualmente 10.000 pessoas morram em decorrência da infecção chagásica (WHO, 2022). No Brasil, de acordo com a série histórica de 11 anos de Martins-Melo et al. (2016), entre o ano de 2000 e 2011 a infecção foi responsável pelas maiores taxas de mortalidade entre as doenças tropicais negligenciadas, superando outras zoonoses endêmicas da região (*e.g.* esquistossomose e leishmaniose). Neste mesmo estudo Martins-Melo *et al.*, (2016) mostram que além das elevadas taxas de mortalidade a doença está relacionada a índices de morbilidade que comprometem parte importante da vida do doente crônico.

A infecção é veiculada principalmente por via vetorial, ou seja, pelos insetos hematófagos da subfamília Triatominae, insetos popularmente conhecidos como: barbeiros, vinchucas, chupão, entre outros (MURILLO-GODÍNEZ, 2018). É importante ressaltar que *T. cruzi* possui outras vias de transmissão, como: transfusão de sangue, por via congênita ou

vertical, por alimentos infectados, acidentes laboratoriais e mais recentemente há discussões sobre a possibilidade de transmissão por meio de relações sexuais (ARAUJO et al., 2017; 2019). Cada via está ligada a mecanismos singulares que possibilitam a transmissão, como exemplo, na transfusão de sangue, a não triagem para *T*. cruzi é determinante, ou a via oral, potencializada pelas condições sanitárias e culturais relacionadas ao preparo de alimentos. Atualmente a via oral é relevante e merece atenção, essa via é documentada em diversos países da América do Sul, como: Argentina, Colômbia, Bolívia, Brasil, Equador e Venezuela (DIAS et al., 2008).

1.2 Subfamília Triatominae

A subfamília Triatominae Jeannel, 1919, compreende insetos hematófagos que podem veicular tripanosomatídeos dos gêneros *Trypanosoma (e.g. T. cruzi, Trypanosoma rangeli* Tejera, 1920) e *Blastocrithidia* Laird 1959 (e.g *Blastocrithidia triatomae* Cerisola et al. 1971) (D'ALESSANDRO, 1976; ROCHA; SILVA et al. 1977; LENT; WYGODZINSKY, 1979).

Os triatomíneos distribuem-se por toda região Neotropical, entre o sul da América do Norte, América Central e a por toda extensão da América do Sul (LENT e WYGODZINSKY, 1979). Também ocorrem na Ásia e Oceania onde não existem registros de transmissão de *T. cruzi*, condição diferente das Américas, onde os triatomíneos são responsáveis pela manutenção da enzootia chagásica na natureza (MONTEIRO et al., 2018).

Acredita-se que a associação de triatomíneos com humanos têm início na América Latina com a chegada do *Homo sapiens* Linnaeus, 1758 no continente (JUSTI et al., 2016). Evidências de 9.000 anos atrás corroboram a hipótese; lesões típicas das doenças foram descritas em múmias andinas pré-colombianos (ARAÚJO et al., 2009). Outro aspecto que enriquece a discussão é a evolução dos tripanosomatídeos. A origem evolutiva desses protozoários provavelmente teve início no grande continente anteriormente formado pela América do Sul, Austrália e Antártida, e depois divergindo a separação da América do Sul da Africa (HAAG; O'HUIGIN; OVERATH, 1998). No entanto, a história evolutiva entre tripanosomatídeos, vetor e hospedeiro gera discussões, não existe um consenso de qual interação ocorreu primeiro, cada evidência leva para hipóteses diferentes (ARAÚJO et al., 2009).

Triatominae conta com 157 espécies distribuídas em 18 gêneros (DALE; JUSTI; GALVÃO, 2021; ZHAO; GALVÃO; CAI, 2021). A primeira classificação sistematizada da subfamília resulta principalmente da revisão de Lent e Wygodzinsky (1979). A partir dos anos

2000 também foram publicadas revisões importantes de Triatominae (CARCAVALLO et al., 2000; GALVÃO et al., 2003).

A origem evolutiva dos triatomíneos é provavelmente monofilética e decorrente de adaptações em seus ancestrais reduvídeos predadores (MONTEIRO et al., 2018). A família Reduviidae é uma das maiores famílias da ordem Hemiptera (~6.800 espécies descritas) que são adaptadas a uma ampla gama de habitats onde são especializados em estratégias de captura de presas (HWANG; WEIRAUCH, 2012). No entanto, os triatomíneos possuem características bem diferentes dos demais reduvídeos, como, por exemplo, a hematofagia, associação com vertebrados e tripanosomatídeos. Sobre a evolução de Triatominae, Schofield e Galvão (2009) defendem uma transição gradativa do hábito predador para sugador ocorrendo de forma independente.

Fundamentado principalmente nas sinapomorfias de caracteres derivados de reduvídeos predadores, Lent e Wygodzinsky (1979) foram um dos primeiros a defender a origem monofilética de Triatominae. No entanto até o momento não há consenso, têm hipóteses monofiléticas (PATTERSON; GAUNT, 2010; WEIRAUCH, 2008; WEIRAUCH; MUNRO, 2009), polifiléticas (DE PAULA; DIOTAIUTI; SCHOFIELD, 2005; SCHOFIELD; GALVÃO, 2009) e parafilética (HWANG; WEIRAUCH, 2012) (Figura. 1).

A filogenia reconstruída por meio de cinco marcadores moleculares de Hwang e Weirauch (2012) mostra que Triatominae é próximo de reduvídeos dos gêneros *Zelurus* Hahn, 1826 e *Opisthacidius* Berg, 1879. As tribos Rhodniini Pinto, 1926 e Carvenicolini Usinger, 1944 formam um grupo irmão. Triatomini Jeannel, 1909 frequentemente forma um clado monofilético (JUSTI et al., 2014). Entanto a tribo possui grupos não naturais, ou seja, apresenta parafilia em alguns de seus gêneros. A parafilia da tribo Triatomini é bem evidenciada (HYPŠA et al., 2002; JUSTI et al., 2014), especialmente entre os gêneros *Triatoma* Laporte, 1832, *Eratyrus* Stål, 1859 e *Panstrongylus* Berg, 1879.

Como as relações taxonômicas não estão totalmente resolvidas, hipóteses são discutidas. Recentemente, Zhang *et al.*, (ZHANG et al., 2016) em estudo filogenômico (dados de transcriptoma) recuperaram uma filogenia que corrobora em partes o estudo de Hwang e Weirauch (2012). Na filogenia a subfamília Stenopodainae Amyot e Serville, 1843 e os gêneros *Zelurus* e *Opisthacidius* formam clado monofilético. Além disso, Triatominae é parafilético em relação a *Opisthacidius*, mas o clado recuperado possui um suporte muito baixo (<50%) limitando as discussões e necessita revisão.



Figura 1. Resumo das hipóteses filogenéticas publicadas de Reduviidae mostrando relações propostas de Triatominae com Reduviidae predadores específicos. Hipóteses monofiléticas (A, C, D, E e H), polifiléticas (B) e parafilética (F, G). Filogramas mostrando relação entre os reduvídeos e estratégia adotada. Adaptado de Monteiro *et al.*, (2018).

A tribo Triatomini é a mais diversa na subfamília, com oito gêneros, a seguir referidos: *Dipetalogaster* Usinger, 1939, *Eratyrus, Hermanlentia* Jurberg e Galvão, 1997, *Linshcosteus, Mepraia* Mazza, Gajardo e Jörg, 1940, *Nesotriatoma* Usinger, 1944, *Panstrongylus, Paratriatoma* Barber, 1938 *e Triatoma* Laporte, 1832. A filogenia mais completa até o momento, ao menos em relação ao número de espécies recupera filogenia com ótimo suporte, mostrando que a tribo é provavelmente dividida em três clados: 1) formado por *Panstrongylus, Nesotriatoma* e *T. tibiamaculata,* 2) com *Hermanlentia, Paratriatoma, Dipetalogaster, Linshcosteus, Triatoma* do Hemisfério Norte e 3) com *Eratyrus, Mepraia* e espécies de *Triatoma* do Hemisfério Sul (JUSTI et al., 2014).

De forma geral, as relações filogenéticas de Triatominae são incongruentes e dentre outros pontos possuem espécies "problemas" que não representam grupos naturais, como por exemplo, o caso de *Triatoma tibiamaculata* Pinto, 1926 espécie que frequentemente aparece próxima ao gênero *Panstrongylus* (HYPŠA et al., 2002; GARDIM et al. 2014, JUSTI et al., 2014a; BELINTANI et al., 2020). *Triatoma tibiamaculata* apresenta caracteres morfológicos "típicos" de *Triatoma*, contudo as filogenias revelam parentesco genético com *Panstrongylus*. Um novo estudo com elementos ultra conservados (UCE) e genes ribossomais corroboram as parafilias dos gêneros *Rhodnius*, *Triatoma* e *Panstrongylus*, além de mostrar a relação de *Triatominae* com reduvídeos predadores, *Opisthacidius Berg* (KIERAN et al., 2021).

A sistemática de triatomíneos fundamenta-se em características morfológicas, biológicas e biogeográficas, contudo atualmente estabeleceu-se uma tendência em utilizar as técnicas moleculares com intuito de aprimorar os estudos filogenéticos (ABAD-FRANCH e MONTEIRO, 2005). As filogenias de Triatominae merecem atenção devido ao contexto epidemiológico que as espécies estão inseridas (SCHOFIELD; GALVÃO, 2009).

1.3 Gênero Mepraia Mazza, Gajardo e Jorg (1940)

O gênero *Mepraia* compreende três espécies endêmicas do Chile, *Mepraia spinolai* (Porter, 1933), *Mepraia gajardoi* Frías, Henry e González, 1998, *Mepraia parapatrica* Frías, 2010 (Tabela 1). As três espécies podem transmitir o agente causal da doença de Chagas, especialmente *M. spinolai* que é mais associado a humanos (CALLEROS et al. 2010., FRÍAS et al., 2017). Atualmente existe grande interesse no estudo com espécies do gênero devido à importância vetorial e às particularidades desses triatomíneos, únicas, em comparação a outras espécies da subfamília. A recente eliminação de *T. infestans* de grandes áreas urbanas no Chile poderia explicar o aumento da *M. spinolai* em habitações humanas (FRÍAS et al. 2017).

A primeira espécie descrita, a princípio como *Triatoma spinolai* (Porter, 1934) foi *M. spinolai* (Figura. 2). Algumas publicações (LENT e WYGODZINSKY, 1979; LENT et al. 1994) consideram que a espécie foi descrita em 1934, entretanto a publicação original foi apresentada no ano de 1932 e publicada no ano de 1934 na Revista Chilena História Natural (FRÍAS et al., 2018). A espécie apresenta polimorfismo alar, com fêmeas micrópteras e machos micrópteros, braquípteros ou macrópteros. Em geral, a coloração é preta com marcações vermelhas ao longo do corpo (LENT e WYGODZINSKY, 1979). *Mepraia spinolai* é encontrado aproximadamente entre 26°S e 33°S, distribuído desde o nível do mar a altitudes de 3000m (FRÍAS et al., 2018). A espécie ocorre em ambiente silvestre, associado a pedras e

fendas nas rochas, também encontrada no peridomicílio humano e ocasionalmente junto a moradias humanas (GAJARDO-TOBAR, 1953). Diferente de outros triatomíneos, ativos durante a noite, *M. spinolai* possui maior atividade durante o dia (GAJARDO-TOBAR, 1953).

Animalia
Arthropoda
Insecta
Hemiptera
Reduviidae
Triatominae
Triatomini
Mepraia
M. gajardoi , M. parapatrica, M. spinolai
T. spinolai, T. chilena Usinger, 1939, T. porteri Neiva e Lent, 1940

Tabela 1. Classificação do gênero Mepraia.

Em 1940, o gênero *Mepraia* foi revalidado, quase simultaneamente dois novos gêneros foram propostos, o primeiro, *Mepraia* sugerido por Mazza, Gajardo e Jorg (1940) e o segundo *Triatomaptera* proposto por Neiva e Lent (1940), os autores sugerem o gênero para o que eles consideravam uma nova espécie: *Triatomaptera porteri* descrita baseada em machos e fêmeas micrópteras, além disso os autores sugeriram as mudanças de *T. spinolai* para o gênero proposto. Essa não foi a única sinonímia da espécie já que um ano antes Usinger (1939) descreveu uma espécie que denominou de *Triatoma chilena*. Lent e Wygodzinsky, (1979) consideram *T. chilena* e *Triatomaptera porteri* como sinonímias de *T. spinolai*. O gênero apenas foi revalidado em 1994 por Lent, Juberg e Galvão, considerando o polimorfismo alar e genitália masculina. No entanto ainda hoje existem autores que não concordam com a classificação do gênero (SCHOFIELD; GALVÃO 2009).

Inicialmente considerada como uma população de *M. spinolai* a segunda espécie descrita foi *M. gajardoi* (FRÍAS et al. 1998, JURBERG et al. 2002) (Figura. 3). A espécie apresenta diferenças cromossômicas substanciais e interfertilidade reduzida (essas duas espécies possuem capacidade de cruzar, mas a viabilidade da prole não sustenta o fluxo gênico). *Mepraia gajardoi* e *M. spinolai* apresentam o mesmo número de cromossomos sexuais (2n =20

autossomos, mais cromossomos X1X2Y nos machos) (FRIAS et al 1998, PANZERA et al., 1998, CALLEROS et al. 2010).



Figura 2. *Mepraia spinolai*. A. macho micróptero, B. fêmea micróptera. Fonte: https://www2.fcfar.unesp.br/#!/triatominae/subfamilia-triatominae/mepraia/mepraia-spinolai/

M. gajardoi é distribuída ao longo da costa norte do Chile, entre 18°S e 26°S, região semiárida do Chile (FRÍAS 2010). O polimorfismo alar típico do gênero ocorre também em *M. gajardoi*. As fêmeas são micrópteras, os machos são sempre alados braquípteros (FRÍAS, 2010). Seus habitats de ocorrência incluem rochas, ninho de aves e até cavernas. Comum ao gênero, também possui importância epidemiológica, ou seja, tem associação com humanos (SAGUA-FRANCO et al., 2000). Também é reportado atividade alimentar durante o dia, hábito não comum em Triatominae (FRÍAS, HENRY E GONZÁLEZ, 1998).



Figura 3. *Mepraia gajardoi*. A. macho, braquíptero, B. fêmea micróptera. Fonte: https://www2.fcfar.unesp.br/#!/triatominae/subfamilia-triatominae/mepraia/mepraia-gajardoi/

M. parapatrica (Figura 4) a última espécie do gênero descrita possui distribuição ao longo da costa norte do Chile, região II de Antofagasta e região III do Atacama (FRÍAS, 2010). As fêmeas são micrópteras, enquanto os machos podem ser macrópteros (asas grandes) ou braquípteros (asas curtas). A espécie é facilmente confundido com o morfotipo braquíptero de *M. spinolai* e *M. gajardoi* e ao morfotipo macróptero de *M. spinolai* (FRÍAS et al., 2018).

As três espécies do gênero diferem quanto à morfologia, em especial: nos caracteres da genitália masculina e da genitália feminina, no cariótipo e diretamente na determinação do sexo, nos padrões de heterocromatina dos cromossomos sexuais e autossomos (FRÍAS e ATRIA 1998; CALLEROS et al. 2010, FRÍAS, 2010). Recentemente, Frias (2018) mostra que as três espécies diferem quanto à morfometria e quanto aos caracteres morfológicos do hemiélitro, assim mostram que apesar de similaridades as três espécies possuem diferenças substanciais quanto às estruturas alares.

Mepraia habitam regiões com características diferentes (CAMPOS -SOTO et al., 2013). No Norte predomina o clima desértico, caracterizado por poucas precipitações. Nas zonas central e Sul, predomina o clima temperado, onde se observam claramente as quatro estações do ano. O Sul, o regime de chuvas e frio predominam e duram o ano todo. O Chile apresenta também grande flora diversa. No Norte a vegetação é escassa e se destacam arbustos e cactos. No Sul, o clima se torna mais úmido, a vegetação fica mais densa e diversificada até culminar, no extremo sul, com florestas de carvalhos e araucárias.

M. gajardoi é encontrado em áreas costeiras do Norte, enquanto *M. spinolai* possui preferência pelos vales semiáridos e áreas costeiras (FRIAS et al., 1998). *Mepraia parapatrica* é encontrada na região costeira de Antofagasta com características climáticas e ecológicas intermediárias a áreas costeiras do Norte, onde habitam *M. gajardoi*.



Figura 4. *Mepraia parapatrica*. A. macho braquíptero, B. fêmea micróptera. Fonte:https://www2.fcfar.unesp.br/#!/triatominae/subfamilia-triatominae/mepraia/mepraia-parapatrica/

1.4 Ecoepidemiologia da doença de Chagas no Chile

Hoje no Chile, quatro espécies de insetos vetores da doença de Chagas são conhecidos (LAMBARRI et al., 2018). A espécie responsável pelo ciclo doméstico é *T. infestans*, e as espécies envolvidas no ciclo silvestre são *M. spinolai*, *M. gajardoi* e *M. parapatrica* (LAMBARRI et al., 2018).

T. infestans é encontrado em todas as regiões do Chile onde apresenta importância médica. É encontrado em habitats domésticos, mas focos de populações reprodutivas também foram detectados em habitats selvagens (BACIGALUPO et al., 2006; 2010). As espécies de *Mepraia* são associadas ao ciclo silvestre, no entanto, *M. spinolai*, comumente encontrado em habitats compostos por áreas rochosas, como pedreiras, ninhos de pássaros e cavernas de animais (APT, REYES, 1990) as vezes é também capturado em áreas peridomiciliares ou em habitações humanas (SCHENONE et al., 1989, FRÍAS et al., 2017). Estudos indicam que a *M. spinolai* possui uma ampla gama de espécies de vertebrados em sua dieta, característica que permite a manutenção da carga parasitaria de *T. cruzi*, além de permitir o contato com diferentes variantes do protozoário (LAMBARRI et al., 2018). Como relatado por Saaveda *et al.*, (2022) o perfil alimentar de *M. spinolai* tem sido avaliado por diferentes técnicas sorológicas, mantendo uma dieta baseada em roedores selvagens, coelhos, cabras, cães, gatos e humanos.

1.5 Filogenia e especiação de Mepraia

O gênero *Mepraia* descrito como monotípico possui atualmente três espécies endêmicas no Chile (CAMPOS-SOTO., 2020). Esses triatomíneos possuem estrutura populacional bem definida (CAMPOS-SOTO et al., 2010), diferenças morfológicas e morfométricas descritas (CAMPOS et al., 2011, FRÍAS et al., 2018) e filogenias com topologias coerentes com os clados descritos, apesar disso a história evolutiva do grupo merece atenção, pois alguns aspectos evolutivos do gênero ainda não são compreendidos (CAMPOS et al., 2015).

A hipótese de especiação por eventos vicariantes e possíveis mudanças cromossômicas são corroboradas por evidências geológicas e suportam os modelos evolutivos descritos para o gênero (FRIAS, 2010). A última elevação das Cordilheiras dos Andes durante o Mioceno há cerca de 20 milhões de anos pode ter sido determinante para a evolução de *Mepraia* (FRÍAS, 2010). O soerguimento teria causado a separação do grupo *"breyeri"*

ancestral [o grupo formado por *Triatoma breyeri* Del Ponte, 1929 e *Triatoma eratyrusiformis* (Bachmann 1999), compreendem um grupo relacionado com as espécies de *Mepraia* (FRÍAS, 2010; JUSTI et al., 2010)], que atualmente ocorre nas áreas secas da Argentina, ao longo da encosta oriental dos Andes. Além das modificações geográficas, alterações cromossômicas podem ter contribuído para o isolamento reprodutivo entre essas espécies (em especial entre *M. gajardoi* e *M. spinolai*) (FRIAS ET AL. 1998).

A especiação parapátrica (FRIAS, 2010) foi proposta como principal mecanismo para a especiação de *M. parapatrica*. Este processo de especiação conjuntamente com a teoria de vicariância é utilizado também para explicar as populações insulares da espécie [ao menos duas ilhas possuem populações de *M. parapatrica* (SAGA et al., 2000., RIVES-BLANCHARD et al., 2017)]. Populações insulares de *Mepraia* foram relatadas habitando a Ilha Pan de Azúcar a 1,8 km do continente (26° 9'S) na Região do Atacama (SAGA et al., 2000). Também existem registros na Ilha de Santa María na região de Antofagasta (RIVES-BLANCHARD et al., 2017). Duas hipóteses tentam explicar a origem biogeográfica dessas populações (CAMPOS-SOTO et al., 2020). A primeira hipótese fundamenta-se na dispersão passiva por aves marinhas (SAGA et al., 2000). A hipótese alternativa, explica a origem dessas populações por eventos vicariantes relacionados às mudanças no nível do mar (FRÍAS, 2010). Poucos estudos foram realizados com essas populações, pelo menos um deles procurou explorar as relações filogenéticas das populações insulares e continentais (CAMPOS-SOTO et al., 2020). No estudo a hipótese defendida para a presença de triatomíneos insulares é a dispersão.

A princípio os processos de especiação parapatrica ocorrem em habitats contíguos e espacialmente segregados, que a partir do equilíbrio do fluxo gênico e a forte seleção natural possibilitam a manutenção de populações segregadas. Em consequência esse tipo de especiação, que embora, em princípio, seja um processo evolutivo comum, talvez seja o modelo mais difícil de verificar (FELDER et al., 2021). As filogenias com incongruências (CALLEROS et al., 2013) ou até mesmo as relações entre as populações insulares e costeiras dificultam a compreensão dos eventos filogeográficos que modulam a evolução do gênero (SAGA et al., 2000, RIVES et al., 2017, CAMPOS-SOTO et al., 2020).

As mudanças cromossômicas são consideradas um aspecto chave para especiação de *Mepraia*, como discutido em FRÍAS (2010), o modelo parapátrico de especiação por heterose negativa parece ser coerente com os dados biogeográficos, biológicos e morfológicos. A heterose é um termo utilizado para descrever a superioridade de genótipos heterozigotos sobre seus genótipos parentais correspondentes (TIMBERLAKE, 2013). Na prática, os híbridos superam em desempenho quaisquer tipos parentais conhecidos e a superdominância e epistasia

também podem estar envolvidas no vigor híbrido (BEN-ARI; LAVI, 2012). No entanto não há unanimidade na literatura na interpretação da base biológica da heterose no processo de especiação, pois é difícil avaliar com clareza o processo biológico (NOTTER; SCHERF; HOFFMANN, 2013).

As primeiras filogenias com marcadores mitocondriais COI e Cytb recuperam topologias coerentes com os três clados descritos (CALLEROS et al., 2010). Nessas filogenias as espécies *M. gajardoi* e *M. spinolai* aparecem filogeneticamente próximas e *M. parapatrica,* a mais distante (CAMPOS et al., 2013). No entanto, os marcadores moleculares utilizados possuem taxa evolutivas muito semelhantes fato que dificulta a discussão da hipótese. Nesse mesmo estudo, as estimativas de origem do gênero calculadas por meio dos marcadores mitocondriais, mostram que o gênero provavelmente divergiu há 3.6 milhões de anos atrás (CAMPOS et al., 2013). Na estimativa *M. spinolai* aparece como espécie mais antiga, seguida de *M. gajardoi* e *M. parapatrica,* mas a natureza dos genes estudados não permitem fechar a discussão, necessitando de estudos complementares (CAMPO et al., 2013).

A mais recente filogenia publicada (CAMPOS-SOTO et al., 2020) recupera uma história evolutiva diferente das propostas anteriores (CAMPOS-SOTO et al., 2020). No entanto, corrobora a relação de ancestralidade do gênero mostradas em (CAMPOS et al., 2013), onde *M. spinolai* é a espécie ancestral e *M. gajardoi* e *M. parapatrica* são mais recentes.

M. gajardoi difere de M. spinolai, em caracteres morfológicos, morfométricos, biogeográficos e na determinação cromossômica do sexo [M. gajardoi (X1X2Y) e M. spinolai (X1X2Y1Y2) (FRÍAS, 1998)] além do possível isolamento reprodutivo (FRIAS, 1998., CAMPOS et al., 2011., FRIAS et al., 2018). Inicialmente o fluxo gênico entre essas duas espécies foi desconsiderado (FRIAS, 1998). A hipótese de isolamento foi proposta fundamentada pelos cariótipos diferentes e pelo isolamento geográfico, no entanto estudos recentes mostram que as barreiras geográficas propostas não são coerentes com as novas evidencias filogeográficas (CALLEROS et al., 2010). Genes mitocôndrias corroboram a filogeografia proposta por Frías et al., (1998), no entanto marcadores nucleares (ITSI, ITSII e 5.8S) recuperam topologias diferentes. Segundo Calleros et al., (2010), a discordância é resultado de processos de hibridação, introgressão ou retenção de polimorfismos ancestrais. A falta de evidências concretas sobre o fluxo gênico entre populações de M. gajardoi e M. spinolai, não permitem determinar se são espécies distintas, sob o critério biológico, mas podese considerar que ambas passaram por um processo de especiação recente, com diferenças morfológicas e citogenéticas, mas com poucas diferenças genéticas (CAMPOS et al., 2013). As regiões de contato entre esses triatomíneos podem manter o fluxo gênico, na literatura as regiões

de contato são conhecidas como zonas hibridas (FELDER et al., 2011). Portanto, mesmo que as taxas de sucesso de cruzamentos entre *M. gajardoi* e *M. spinolai* em condições laboratoriais sejam relativamente baixas [6-10% (FRÍAS, 1998)], não se pode descartar a hibridação por completo.

Ao contrário das dúvidas referidas as relações filogenéticas das espécies do gênero com as espécies ancestrais são bem definidas (JUSTI et al., 2014). As evidências mais robustas indicam relação filogenética com *T. breyeri* e *T. eratyusiformis* (JUSTI et al., 2014, KIERAN et al., 2021). As espécies de *Mepraia* juntas com *T. breyeri* e *T. eratyrusiformis* formam o complexo *"spinolai"* ou *"T. spinolai"*. O complexo é monofilético e apresenta proximidade filogenética com o complexo *T. infestans* (JUSTI et al., 2014, KIERAN et al., 2021). As relações monofiléticas entre os táxons do complexo *T. spinolai* corroboram a origem do grupo por vicariância (CAMPOS et al., 2013).

As relações com *T. breyeri* e *T. eratyrusiformis* levaram a hipóteses interessantes. Hypša *et al.*, (2002) reconstruindo filogenia com "*barcodes*" sugerem que as espécies *T. eratyrusiformis* e *T. breyeri* sejam sinonimizadas para *Mepraia*. No estudo, as espécies formam clados estáveis e bem suportados. Outros estudos sustentam também as relações filogenéticas de *T. eratyrusiformis* com *Mepraia* (MORENO et al., 2006; FRÍAS, 2010; CAMPOS et al., 2014, JUSTI et al., 2014). As primeiras revisões sistemáticas já classificavam as espécies *T. breyeri*, *T. eratyrusiformis* e *M. spinolai* (syn. *T. spinolai*) junto ao complexo *T. spinolai* (naquele momento o gênero *Mepraia* não era reconhecido), essa proximidade era fundamentada nas similaridades morfológicas e pela preferencias por áreas semiáridas da América do Sul.

Fortes evidências sustentam o gênero *Mepraia* como monotípico, mas as relações intraespecíficas não são conclusivas (CAMPOS-SOTO et al., 2019). Como descrito, filogenias com genes mitocondriais suportam as três espécies como linhagens congruentes (CAMPOS et al., 2013), no entanto, o status de *M. parapatrica* ainda é controverso (CAMPOS et al., 2015). A plasticidade típica de Triatominae, eventos de introgressão ou hibridação podem estar atuando de forma efetiva nessas espécies (CALLEROS et al., 2010). Campos *et al.*, (2015) em estudo com várias amostras de genes mitocondriais de *Mepraia* e *T. eratyrusiformis* mostram indícios de introgressão e recuperam filogenias incongruentes e sugerem novos estudos para elucidar os resultados encontrados.

Na subfamília Triatominae, o gênero *Mepraia* está inserido junto ao clado de espécies sul-americanas (HYPŠA et al., 2002; DE PAULA; DIOTAIUTI; SCHOFIELD, 2005). As filogenias de Triatomini normalmente recuperam dois clados bem definidos separados entre espécies sul e norte americanas (MONTEIRO et al., 2018). A compreensão de processos

microevolutivos de *Mepraia* podem ser extremamente úteis para compreender a irradiação de espécies para outros ambientes, processos de especiação, adaptação, além de corroborar com informações estratégicas para controle e vigilância entomológica, bem como para elucidar processos evolutivos em Triatominae

1.6 Perspectivas e panorama geral do estudo de transcriptoma

Uma abordagem moderna e efetiva para inferir a história evolutiva das formas de vida provém do sequenciamento de genes (SHENDURE et al., 2017). As tecnologias de sequenciamento tem avançado de maneira rápida, especialmente os métodos de sequenciamento de alto rendimento (do inglês: *High throughput sequencing - HTS*) (REUTER; SPACEK; SNYDER, 2015).

O genoma humano foi publicado em forma de rascunho no ano de 2001 (LANDER et al., 2001, VENTER et al., 2001). Essa façanha foi realizada com o sequenciamento de DNA Sanger, que era limitado em rendimento e alto custo; de fato, a primeira sequência do genoma humano foi estimada em 0,5 a 1 bilhão de dólares. Essas limitações reduziram o potencial do sequenciamento de DNA para outras aplicações, como o sequenciamento de genoma de pessoas. No entanto esses primeiros desafios foram gradativamente resolvidos, em primeiro, o foco no custo do sequenciamento, em segundo o ganho de produção ((REUTER; SPACEK; SNYDER, 2015). O caminho para diminuição dos custos e da melhoria de rendimento, foi principalmente liderado por plataformas HTS comercias, que diferem em detalhes, mas os métodos de sequenciamento são bem similares (METZER, 2010).

Os sequenciamentos de alto rendimento revolucionaram as análises transcriptômicas, possibilitando determinar o conjunto total de transcritos, incluindo mRNAs, RNAs não codificados e micro RNAs de uma célula (CARVER et al. 2012), além de permitir quantificar níveis de expressão de cada transcrito durante o desenvolvimento e sob condições adversas, como estresse térmico, infecções por patógenos entre outros (WANG et al 2009). O sequenciamento massivo fornece dados para a compreensão da interferência das condições ambientais na expressão gênica dos organismos (MARCHANT et al. 2015).

O transcriptoma ou transcritoma é o conjunto completo de transcritos em uma célula, e sua quantificação é um retrato gênico do organismo em determinado momento, ou seja, é específica para um estádio de desenvolvimento, sexo, ambiente ou condição fisiológica (WANG et al, 2009). O entendimento do transcriptoma é essencial para a interpretação dos complexos mecanismos genéticos e suas implicações nos organismos (CARVER et al, 2011).

De acordo com Lower *et al.*, (2017) o conteúdo "informativo" de um organismo é registrado no DNA de seu genoma e expresso por meio de transcrição. Aqui, o mRNA serve como uma molécula intermediária transitória na rede de informações, enquanto os RNAs não codificantes (ncRNA) desempenham diversas funções adicionais. Os ncRNA são moléculas que não são traduzidas em proteínas e por esse motivo durante muito tempo, suas funções biológicas não são totalmente esclarecidas, no entanto estudos relacionam essas moléculas a vários processos celulares ou até mesmo doenças como câncer e Alzheimer (MONGA e BANERJEE, 2019). Também se acredita que os ncRNAs mais conservados sejam fósseis moleculares, e seus papéis atuais permanecem principalmente na regulação do fluxo de informações do DNA para a proteína (JEFFARES, POOLE, PENNY, 1998).

O RNA-Seq é uma técnica de sequenciamento que utilizada bibliotecas de cDNA (DNA complementar) geradas a partir da conversão de transcritos obtidos de determinado tecido biológico que podem ser acessados por meio de abordagens de sequenciamento massivo (WANG, GERSTEIN e SNYDER, 2009). De forma geral o RNA-Seq facilita os estudos com transcritos, permitindo um olhar preciso sobre a transcrição alternativa de genes, modificações pré-transcricionais (pré-RNAm), fusões genicas, SNP (Polimorfismo de nucleotídeo único ou polimorfismo de nucleotídeo simples, em inglês (single nucleotide polymorphism) e mudanças na expressão gênica (MAHER et al., 2009). Anterior a tecnologia de RNA-Seq, a técnica comum era o "microarrays", técnica que possui limitações relevantes em comparação a nova tecnologia (KUKURBA e MONTGOMERY, 2015). As desvantagens dos microarrays incluem custo, grande número de designs de sonda baseados em sequências de baixa especificidade, e a falta de controle sobre o pool de transcritos analisados (JAKSIK et al., 2015). Outros pontos fracos dos microarrays são sua precisão e especificidade relativamente baixas, bem como a alta sensibilidade da configuração experimental a variações na temperatura de hibridização, a pureza, a taxa de degradação do material genético e à amplificação do processo que, juntamente com outros fatores, pode impactar as estimativas de expressão gênica LOCKHART et al., 1996; RAMSAY, 1998; STOUGHTON, 2005).

A transcriptômica tem sido aplicada para quantificação da expressão gênica, no diagnóstico de doenças, na identificação de genes, na filogenia e na elucidação de vias metabólicas que respondem a estresses ambientais diversos (LOWE et al., 2017). Também pode ser utilizada para identificação de genes alvos para controle biológico e na elaboração de catálogos de genes super ou sub expressos em determinado tecido (CHU e COREY, 2012). Outro método muito útil e amplamente explorado são os estudos com Polimorfismo de Nucleotídeo Único (SNP). Os SNPs é uma variação na sequência de DNA que afeta somente

uma nucleobase (WANG et al., 1998). Embora um determinado SNP possa não causar mudanças no produto gênico, algumas substituições são associados com doenças em humanos. (COLLINS; GUYER; CHAKRAVARTI, 1997; MARTH et al., 1999). Além do mais, os SNPs vem sendo utilizados para estudos evolutivos (MIKKELSEN et al., 2005; ZHANG; LI, 2005; LIU et al., 2008).

Na subfamília Triatominae os estudos com transcritoma envolvem principalmente sialotranscriptomas, termo utilizado para descrever o transcriptoma obtido das glândulas salivares (TOBIAS e LATORRE-ESTIVALIS, 2021). Os sialotranscriptomas são interessantes devido ao potencial farmacológico, entretanto com o acesso cada vez mais comum às técnicas de *HTS*, outras abordagens são exploradas, como as relacionadas à fisiologia do inseto (BUSSACOS et al, 2011; ASSUMPÇÃO et al, 2012; RIBEIRO et al, 2012) ou em estudos filogenômicos (ZHANG et al, 2016). Segundo Assumpção et al. (2014), as glândulas salivares de triatomíneos possuem diversidade de compostos anti-hemostáticos. Esses compostos são estão envolvidos no repasto sanguíneo, além de possuir uma enorme quantidade de transcritos que podem ser úteis para o estudo e desenvolvimento de novos fármacos.

Outra estrutura informativa e interessante para estudos com vetores e outros insetos são as cabeças. As estruturas encontradas na cabeça de insetos possuem genes sob seleção positiva que podem revelar características ecológicas relacionadas à percepção sensorial do ambiente. Em triatomíneos, os órgãos sensoriais são importantes mecanismos de sobrevivência (GALVÃO, 2014, p.71). As famílias de genes quimiossensoriais de insetos codificam proteínas que são cruciais para a localização do hospedeiro, localização do parceiro e oviposição (ANDERSSON et al., 2016). De acordo com YUVARAJ *et al.*, (2018) o sistema quimiossensorial periférico do inseto compreende receptores odorantes (ORs), receptores gustativos (GRs), receptores ionotrópicos (IRs), proteínas de ligação a odorantes (OBPs), proteínas quimiossensoriais (CSPs) e proteínas de membrana de neurônios sensoriais (SNMPs). Os quimiossensores presentes na cabeça são de extrema importância para a sobrevivência e reprodução dos insetos (YUVARAJ et al., 2018).

Em triatomíneos, como em outros insetos hematófagos, as glândulas salivares, possuem uma grande variedade de compostos anti-hemostáticos que são injetados no hospedeiro durante a alimentação. Portanto, esses organismos são capazes de neutralizar as respostas hemostáticas desencadeadas para evitar a perda sanguínea após a lesão de tecidos, pela agregação plaquetária, coagulação e vasoconstrição. O sialotranscritoma possuem uma diversidade de transcritos que podem revelar compostos ativos farmacologicamente ou genes que auxiliariam na compreensão dos mecanismos evolutivos desses organismos (ASSUMPÇÃO et al. 2012). A expressão e regulação de genes específicos na cabeça entre os sexos podem estar relacionadas a diferenciação sexual e a estados fisiológicos distintos, como observado em alguns genes de *Drosophila*. A expressão gênica ocorre no corpo gorduroso e provavelmente desempenham funções específicas em comportamento sexuais (FUJII e AMREIN 2002; CHANG et al. 2011).

Caracterizar e quantificar a expressão genica de diferentes tecidos, fornece informações sobre a regulação de genes e revela diversos detalhes da biologia do organismo. Ademais podem ajudar ampliar o conhecimento de genes anotados a partir das referências disponíveis. A análise do transcriptoma permite o estudo preciso e valioso e tem sido fundamental para compreensão de doenças, mecanismos de interação parasita-hospedeiro, na biologia de vetores, hospedeiros, no desenvolvimento de fármacos e na inferência filogenética.

1.7 Breves considerações sobre genômica e filogenômica

O conhecimento de como as espécies estão relacionadas, umas com as outras, sustenta grande parte da biologia evolutiva. Determinar as relações filogenéticas entre as espécies é um objetivo importante por si só e está subjacente ao nosso sistema de classificação filogenética (KAPPLI et al., 2020). Os estudos filogenômicos tem mudado drasticamente a compreensão das relações entre os seres vivos (GIRIBET, 2015; TELFORD, BUDD e PHILIPPE, 2015). Atualmente o consenso é que a filogenômica refere-se ao estudo das relações filogenéticas por dados de sequenciamento de larga escala (CANNON e KOCOT, 2016). No entanto, o termo originalmente refere-se ao estudo da evolução de famílias gênicas (EISER e FRASER, 2003). A filogenômica tem relação direta com o surgimento de tecnologias de sequenciamento de alto rendimento de transcriptomas ou de dados de genomas completos (CANNOT e KOCOT, 2016).

O sequenciamento de nova geração (do inglês, *Next-Generation Sequencing-* NGS) também chamado de Sequenciamento Massivo em Paralelo (do inglês, *Massive Parallel Sequencing*) ou apenas Sequenciamento de Alto Rendimento (do inglês, *high throughput sequencing* são diferentes nomes para um grupo de técnicas modernas de sequenciamento de DNA. O NGS junto com o aumento da capacidade computacional dos últimos 30 anos são os responsáveis pela difusão de estudos filogenômicos. Essas tecnologias permitem sequenciar e processar as informações contidas no DNA de forma muito mais rápida e barata quando comparado aos sequenciamentos de primeira geração (*e.g.*, Sanger) e nos últimos anos estão revolucionando o estudo da genômica e da biologia molecular (SHENDURE e JI, 2008). Apesar

de dispor de várias plataformas de sequenciamento de NGS elas diferem entre si, todos os NGS se baseiam no processamento paralelo massivo de fragmentos de DNA e em comparação com o sequenciador de eletroforese capilar, o NGS é muito superior, enquanto o primeiro gera 96 fragmentos por vez, o NGS geram bilhões de fragmentos ao mesmo tempo (WHANG et al., 2009).

Os protocolos NGS iniciam com o preparo de uma biblioteca, que consiste fundamentalmente na fragmentação do DNA inicial em fragmentos menores ligados a adaptadores (sequências nucleotídicas curtas e conhecidas) utilizados como *primers* universais para distribuir espacialmente os fragmentos de DNA em uma superfície solida (Figura 5, GOODWIN et al., 2016). Os fragmentos obtidos são amplificados por Reação em Cadeia de Polimerase (PCR) gerando *clusters* de sequências iguais, esse passo permite que a incorporação de cada base durante o processo de sequenciamento emita um sinal suficientemente forte para detecção pelo sistema óptico do equipamento. Após sucessivos ciclos de incorporações as sequências são recuperadas por programas específicos para cada sequenciador. O processo de sinalização e captura é chamado de *base calling* (WICK, 2019).

Embora os NGS proporcionem uma quantidade de dados muito informativos no aspecto biológico, os desafios maiores estão no processamento dos dados, tornando o trabalho de bioinformática essencial para o desenvolvimento dos métodos atuais (PABINGER et al., 2014). Um dos passos essenciais do processamento dos dados brutos, consiste na averiguação dos dados brutos gerados a partir da *base calling*, usualmente no formato Fastq (Figura 5). As informações contidas no arquivo de saída no formato Fastq permitem obter além dos dados brutos informações importantes sobre a qualidade do sequenciamento. O valor de qualidade é representado pelo *Phred score* (Q), valor que remete a possibilidade de determinada base obtida no processo do *base calling* estar errada (EWING e GREEN, 1998).

O protocolo atual para os estudos filogenômicos são atualizados constantemente (YOUNG e GILLUNG, 2020), mas geralmente tem início, a partir da obtenção do material genético [esse podendo ser DNA ou RNA] e no NGS, como os oferecidos pelas plataformas *Illumina Hiseq*. Atualmente, a *Illumina* oferece melhor custo por par de base e tornou-se a plataforma mais utilizada nos estudos filogenômicos (EGGER et al., 2015; CANNOT e KOCOT, 2016).



Construção de Biblioteca

Figura 5. Fluxo de trabalho geral para as plataformas de sequenciamento de nova geração utilizado nas plataformas *Illumina*. O fluxograma ilustra o procedimento adotado pelas plataformas de nova geração para sequenciar amostras de material genético. Adaptado de Xiaotu *et al.* (2019).

Os dados de genomas ou transcriptomas obtidos pelo NGS são processados por meio de métodos computacionais que auxiliam na montagem das leituras brutas, na avaliação da qualidade e na filtragem de sequências de baixa qualidade, na identificação de regiões codificadoras (CDS, do inglês *conding sequence*) e na categorização de vocabulários ontológicos *gene ontology*, também conhecidos como termos *GOs* (do inglês: *Gene Ontology*) (Figura 6). Os termos *GOs* são amplamente utilizados nas ciências ômicas e podem ser obtidos por meio de bioinformática (CANNOT e KOCOT, 2016).



Figura 6. Fluxo de análise bioinformática para montagem sem referência por meio de dados brutos obtidos por RNA-Seq. Os procedimentos padrões para tratamento e investigação de dados de sequeciamento do tipo RNA-Seq são realizados por meio de recursos computacionais. Adaptado de Conesa *et al.* (2016).

Também a partir dos dados de NGS, é possível descrever os genes ortólogos que são sequências homólogas originadas a partir de um gene do último ancestral comum entre as espécies, que geralmente possuem funções biológicas equivalentes do gene ancestral (JENSEN, 2001). Os grupos de genes ortólogos são uteis para os estudos filogenéticos estimados por meio de métodos de reconstrução filogenética generalizados como Máxima Verossimilhança (ML) ou inferência Bayesiana (IB). Filogenias podem utilizar genes ortólogos e parálogos para descrever a história evolutiva dos organismos (KOONIN; LIPMAN, 1997; ALTENHOFF; DESSIMOZ, 2009; TATUSOV; HELLMUTH et al., 2015).

Os insetos constituem uma vasta diversidade biológica e morfológica e compreendem o maior número de espécies conhecidas (STORK et al., 2015). Muitas das relações filogenéticas entre as espécies de insetos é controversa e de difícil resolução (MISOF et al., 2014). Um exemplo é a linhagem de Polyneoptera Martynov, 1938, pois as topologias das ramificações desse grupo de insetos são altamente controversas. O grupo atualmente

33

apresentam 11 Ordens que possuem relações filogenéticas inconclusivas (BEHURA, 2015). No entanto, por meio de abordagens filogenômicas novos *insights* começam a surgir para o grupo (WIPFLER et al., 2014).

A filogenômica é importante na entomologia, pois pode ajudar a entender melhor as relações evolutivas de um grupo que representa 75% de todos os animais descritos até o momento (BEHURA, 2015). Na plataforma do *National Center for Biotechnologic Information (NCBI)* encontramos 1.696 genomas disponíveis da classe Insecta Linnaeus, 1758 um número relativamente pequeno mais que tem aumentado nos últimos anos ("INSECTA - Genome -NCBI," Acesso: 08 de junho de 2022). Ao consultar a plataforma no que se refere à Ordem Hemiptera Linnaeus, 1758 encontra-se 82 genomas completos. Para a família Reduviidae temos apenas dois triatomíneos: *Triatoma infestans* Klug 1835 *e Rhodnius prolixus* Stål, 1859 ("Reduviidae - Genome - NCBI," Acesso: 08 de junho de 2022). A integração da filogenética com os dados do genoma ainda é pouco explorado no que se refere às espécies da subfamília Triatominae. Exceção é o trabalho conduzido por Zhang et al. (2016) que obteve dados de transcriptoma de algumas espécies de Triatominae.

2. JUSTIFICATIVA

A compreensão da ecoepidemiologia chagásica é complexa e mesmo após um século da descrição do ciclo da doença e vários avanços no controle e vigilância, a enfermidade é relevante. A doença de Chagas destaca-se como as principais endemias na América do Sul, acometendo milhares de pessoas em idade produtiva. O conhecimento dos vetores envolvidos na enzootia podem auxiliar no controle epidemiológico e na previsão de modelos de vigilância mais efetivos. Além do mais, a identificação específica é de grande relevância epidemiológica, uma vez que possibilita o mapeamento da distribuição geográfica das populações de triatomíneos. Desse modo consegue-se visualizar áreas de risco de colonização e recolonização, e contribuir com iniciativas de controle vetorial.

Mepraia é um gênero com características únicas na subfamília Triatominae, por esse motivo, este estudo visa ampliar o conhecimento de *M. gajardoi*, *M. parapatrica* e *M. spinolai*. As características biológicas singulares dessas espécies serão uteis para auxiliar na compreensão da história evolutiva dos triatomíneos. As relações filogenéticas do gênero não são totalmente esclarecidas e, portanto, a obtenção de transcriptomas de glândulas salivares e da cabeça é importante, pois, além de permitir caracterizar os transcritos que podem estar envolvidos na hematofagia, os transcriptomas serão uteis para inferir com maior confiabilidade a filogenia do gênero. Para isso a técnica de RNA-Seq foi escolhida. A técnica permite revelar e quantificar um grande número de RNAs em uma amostra em um determinado momento. O RNA-Seq é um sequenciamento de alto rendimento que permite estudar com relativa acurácia: transcrições de genes alternativos, modificações pós-transcricionais, fusão gênica, mutações como SNPs e modificações na expressão gênica em condições pré-definidas. O RNA-Seq também tem grande potencial medicinal, pois permite traçar biomarcadores para indicações clínicas, inferir vias farmacológicas e fazer diagnósticos genéticos.

Por fim, tomando por base tais considerações, o presente estudo visa ampliar o conhecimento sobre esses triatomíneos que estão envolvidos na transmissão de *T. cruzi* ao homem nas regiões costeiras do Chile, não só para caracterizar os transcriptomas da cabeça e glândulas salivares de espécimes do gênero *Mepraia*, mas também investigar as relações filogenéticas de suas espécies.

3. OBJETIVOS

Caracterizar o transcriptoma de tecidos da cabeça e glândulas salivares de espécimes do gênero *Mepraia*, não apenas para sua caracterização, mas para compreender melhor as relações filogenéticas entre *M. gajardoi*, *M. parapatrica* e *M. spinolai*.

3.1 Objetivos específicos

- Coletar as três espécies estudadas.
- Extrair o RNA total de tecidos da cabeça e glândulas salivares de *M. spinolai, M. gajardoi* e *M. parapatrica*.
- Sequenciar e montar, pela primeira vez o transcriptoma de tecidos da cabeça e glândulas salivares de *M. spinolai, M. gajardoi e M. parapatrica.*
- Buscar diferenças que possam auxiliar no entendimento do processo de diferenciação entre as três espécies.
- Realizar estudo filogenômico a partir de dados obtidos por RNA-seq.
- Montar e anotar o mitogenoma das espécies.
- Inferir as relações filogenéticas através dos conjuntos de genes mitocondriais.

4. MATERIAL E MÉTODOS.

4.1 Espécimes.

Os espécimes utilizados para o estudo foram selecionados a partir de *M. gajardoi*, *M. parapatrica* e *M. spinolai* coletados em regiões do Chile (Figura. 7). As populações de *M. gajardoi* coletadas na região administrativa de Tarapacá (Região I) estavam aproximadamente 10km de distância entre si. Espécimes de *M. parapatrica* estavam 10 km de distância entre cada ponto e foram coletadas na região II (Antafogasta). *M. spinolai* possuem maior distância entre os pontos, aproximadamente 700 km correspondentes as regiões Metropolitana (Região de Santiago) e Atacama (Região III). Em média às distâncias entre os pontos de coleta das três espécies foram de 300km e como descrito anteriormente correspondem as três regiões administrativas chilenas.

Os estudos foram conduzidos a partir de espécimes capturados em regiões geográficas diferentes (*e.g.*, habitat, relevo), desse modo, para cada espécie estudada obteve-se dois pontos diferentes de captura (Tabela. 2). O número de triatomíneos coletados e a dificuldade de manutenção em laboratório, foi ponto crítico para decisão de utilizar ninfas machos de 5° estádio de desenvolvimento, pois essas foram coletadas em maior número, provavelmente pela sazonalidade do desenvolvimento das espécies (Tabela 3). O ciclo biológico das espécies de *Mepraia* não são totalmente descritos, portanto não foi possível determinar o período de desenvolvimento dos espécimes estudados. Mas como as distinções entre os estádios são bem características, pode-se determinar com certeza o estádio de desenvolvimento.



Figura 7. A. Recorte do mapa do Chile ilustrando as regiões de coletas de *M. gajardoi, M. parapatrica* e *M. spinolai*. B. Pontos de coletas de *M. gajardoi*. C. Pontos de coleta de *M. parapatrica*. D. Pontos de coleta *M. spinolai*
Locais	Datas	Regiões	Cidades	Localidades	Habitats	Coordenadas		Elevações	Espécies
1	12/03/18	Metropolitana	Colina	Cantera	Silvestre	S 33° 05'16.7"	W 70°40' 48.0"	755 m	M. spinolai
2	14/03/18	Atacama	Inca de Oro	Mina de ouro	Peridomicílio	S 26° 48.159'	W 069 57.040'	1916 m	M. spinolai
4	18/03/18	Tarapaca	Iquique	Caleta San Marcos	Silvestre	S 21° 08.004'	W 070° 07.606'	29 m	M. gajardoi
5	18/03/18	Tarapaca	Iquique	Caleta Rio Seco	Silvestre	S 21° 00.128'	W 070° 09.854'	16 m	M. gajardoi
6	18/03/18	Tarapaca	Iquique	Caleta San Marcos	Silvestre	S 21° 00.535'	W 070° 10.104'	23 m	M. gajardoi
7**	20/03/18	Antofagasta	Tal Tal	Caleta El medano	Peridomicílio	S 24° 53.571'	W 070° 31.506'	14 m	M. parapatrica
8**	22/03/18	Antofagasta	Tal Tal	Caleta Loreto	Silvestre	S 24° 50.062'	W 070° 32.447'	7 m	M. parapatrica
9**	22/03/18	Antofagasta	Tal Tal	Playa Chica	Silvestre	S 24° 45.196'	W 070° 33.403'	11 m	M. parapatrica

Tabela 2. Informações sobre as coletas de três espécies de Mepraia em quatro regiões do Chile.

** refere-se a novas regiões de distribuição anteriormente não conhecidas para espécie.

Tabela 3. Identificação de três espécies de *Mepraia* utilizadas para o estudo de transcriptoma.

Local *	Espécie	Estádio de desenvolvimento	Número	ambiente	Id da amostra
1	Mepraia spinolai	5° estádio	2	Silvestre	M. spinolai1 e M. spinolai2
2	Mepraia spinolai	5° estádio	4	Peridomicílio	M. spinolai3, M. spinolai4, M. spinolai5 e M. spinolai6
4	Mepraia gajardoi	5° estádio	2	Silvestre	M. garjadoi7, M. garjadoi8 e M. garjadoi9
6	Mepraia gajardoi	5° estádio	4	Silvestre	M. garjadoi10, M. garjadoi11 e M. garjadoi12
7	Mepraia parapatrica	5° estádio	3	Peridomicílio	M. parapatrica13, M. parapatrica14 e M. parapatrica15
8	Mepraia parapatrica	5° estádio	3	Silvestre	M. parapatrica14, M. parapatrica15 e M. parapatrica16

**locais descritos junto a tabela 2.*

4.2 Dissecação das três espécies de Mepraia

Para à aquisição dos tecidos da cabeça e da glândula salivar os espécimes foram mantidos previamente sem alimentação por no mínimo sete dias, desse modo, procura-se evitar contaminação das amostras com sangue do repasto. Os espécimes foram anestesiados em gelo e manipulados em ambiente estéril, a esterilização foi por tratamento com *RNase exterminator*® *BioAgency*. Todo o procedimento foi realizado em *RNA Holder*® *BioAgency* para evitar a degradação do RNA, tomando o cuidado em manter as amostras em gelo até o final do procedimento. Para a secção, as cabeças dos espécimes foram deslocadas na cervical, com o devido cuidado para manter a integridade das glândulas salivares. Após a secção as amostras de cabeças e glândulas salivares foram depositadas em 100 µl de *Trizol*® e congeladas a -80°C até o momento da extração.

4.3 Extração do RNA total.

Os RNAs totais foram obtidos de 18 cabeças e glândulas salivares, sendo 6 por espécie. A extração seguiu o protocolo de *Trizol* / clorofórmio modificado (CHOMCZYNSKI E MACKEY, 1995). As amostras foram mantidas no gelo durante todo o procedimento. As extrações individuais iniciaram-se por ação mecânica sobre o tecido (maceramento com pistilo) seguido da adição de volume de 1000 μ l de *Trizol*® (inicialmente 300 μ l e após o maceramento mais 700 μ l). Após a adição de 1000 μ l de *Trizol*® as amostram foram incubadas no gelo por 10 minutos e logo centrifugadas a 4°C a 10.000 g por 2 minutos. O sobrenadante de cada amostra foi transferido para micro tubo novo e adicionado 300 μ l de Clorofórmio: Álcool Isoamílico 24:1 (o método forma "fases" bem definidas permitindo a separação do material biológico de interesse). Após o processo de centrifugação as amostras passam rapidamente pelo vórtex, e voltam a ser centrifugadas a 4°C e 10.000 g durante 15 minutos. Essa etapa (lavagem) foi repetida quatro vezes, após a última lavagem a "fase aquosa" foi transferida para um novo micro tubo com 1 μ l de Glicogênio *Azure* (30 ng/ μ l).

Para a etapa de precipitação foi adicionado Isopropanol na proporção 1:1 (\cong 300 µl /amostra), a homogeneização foi por inversão (10X). As amostras homogeneizadas foram incubadas por 10 min em gelo. Após o período as amostras foram centrifugadas a 4 C^o em 12.000 g, por 10 min. O sobrenadante descartado e o *pellet* lavado por duas vezes com 500 µl EtOH 80%. O pellet foi seco a 60° C (*termo-blotting*) por 1 minutos e ressuspenso em 20 µl de

água DEPC pré-aquecida à 60° C.

4.4 Purificação do RNA.

As purificações dos RNAs totais seguiram protocolo adaptado de Cathala *et al* (1983). Assim, foi adicionado à amostra (de 20 µl) um volume de 13,8 µl de Cloreto de Lítio (LiCl) a 7.5M. Após incubação a -20°C, por 30 minutos, as amostras foram submetidas à centrifugação (4°C e 13.000 g por 15 minutos). Os sobrenadantes de cada amostra foram descartados e os *pellets* foram lavados com 1ml de EtOH 70% (\cong 4°C). Após as lavagens os *pellets* foram desidratados à 60° C (*termo-blotting*) e ressuspensos em 20 µl de água DEPC préaquecida a 60° C. A integridade do RNA foi visualmente triada em gel de agarose, quantificada em fluorômetro *Qµbit* enquanto as leituras *NanoDrop* asseguraram a escolha de amostras com índices próximos a 2.0 nas relações 260/230 e 260/280.

Por meio de interações intramoleculares as moléculas de RNA podem dobrar-se, alterando a estrutura secundária e afetando a migração das moléculas no gel. Para prevenir o enovelamento, utilizou-se um gel de agarose desnaturante à 0,8% (80 ml de tampão TBE + 0,64 mg de agarose) que respeitou as especificações: para cada 1 μ l de amostra foram adicionados 14 μ l de formamida (50 %) e 3 μ l de Buffer 5X. O tampão de corrida escolhido foi TBE 1X (100 ml TBE 10X + 900 ml de água DEPC). Os resultados foram visualizados em transluminador de UV.

4.5 Sequenciamento por RNA-Seq.

Previamente à construção das bibliotecas de cDNA as amostras tiveram a integridade do RNA avaliadas por eletroforese capilar pelo sistema *Agilent 2100 Bioanalyzer*, (Agilent Tecnologies®, protocolo padrão para RNA. O protocolo permite estimar a qualidade do RNA pelo índice RIN (do inglês *RNA Integrity Number*).

Bibliotecas de RNA-Seq foram construídas a partir de amostras das espécies estudadas, a quantidade de RNA total foi estimada utilizando o *TruSeq RNA Sample Prep kit* v2 (*Ilumina*) protocolo de acordo com as instruções do fabricante.

As bibliotecas agrupadas em 12 e corridas em uma lane de uma *flow cell* na plataforma *Illumina HiSeq2500* com 2x100 pb paired-end (Laboratório de Genômica Funcional Aplicada à Agricultura e Agroenergia, ESALQ-USP, Brasil).

4.6 Controle de qualidade e montagem das sequências.

Todas as *reads* "brutas" foram avaliadas por meio do aplicativo *FastQC* (ANDREWS, 2010). *FastQC* é um programa que fornece uma visão geral das *reads* e possibilita personalizar as estratégias de controle de qualidade, ou seja, permite aos usuários realizar várias verificações de controle de qualidade em dados de sequência bruta gerados por pipelines de sequenciamento de alto rendimento (ANDREWS, 2010).

Após as leituras no *FastQC* todas as *reads* foram aparadas e filtradas por qualidade *Phred* com auxílio do programa *Trimmomatic* (BOLGER et al. 2014). Em resumo: as *reads* foram avaliadas seguindo comprimento do conjunto de bases, assumindo um erro de 0,01, ou seja, um erro a cada 100 bases (qualidade *Phred* igual a 20) nos interiores das sequências, erro permitido 0,05 nas extremidades. Por meio do *Trimmomatic* também se verificou e removeu sequências adaptadoras. Esse programa analisa as duas *reads* do *pair-end* e descarta ambas as sequências se apenas uma não cumpre os parâmetros de filtragem estabelecidos. Essa característica é importante para o programa que monta os *contigs*, porque todas as leituras estão emparelhadas e permitem a produção de uma montagem mais eficaz, especialmente na ausência de um genoma de referência. As montagens seguiram a estratégia *de novo*, ou seja, sem genoma referência. Todas as montagens realizadas neste estudo foram realizadas com auxílio do pacote *Trinity* (GRABHERR et al. 2011).

O *Trinity*, representa um método para a reconstrução *de novo* eficiente e robusta de transcriptomas a partir de dados de RNA-seq. Em breve resumo: o *Trinity* combina três módulos de software independentes: *Inchworm, Chrysalis* e *Butterfly*, aplicados sequencialmente para processar grandes volumes de leituras de RNA-seq. *Trinity* particiona os dados de sequência em muitos gráficos de Bruijn individuais, cada um representando a complexidade transcricional em um determinado gene ou lócus e, em seguida, processa cada gráfico de forma independente para extrair isoformas de *splicing* e separar transcritos derivados de genes parálogos. Resumidamente, o processo funciona assim:

 Inchworm reúne os dados de RNA-Seq em sequências únicas de transcrições, geralmente gerando transcrições completas para uma isoforma dominante, mas depois relata apenas as porções únicas de transcrições emendadas *splicing* alternativos.

• *Chrysalis* agrupa os *contigs* do *Inchworm* em agrupamentos e constrói gráficos de Bruijn completos para cada agrupamento. Cada cluster representa a complexidade transcricional total de um determinado gene (ou conjuntos de genes que compartilham

sequências em comum). Crisálida então particiona o conjunto de leitura completo entre esses gráficos disjuntos.

 Butterfly então processa os gráficos individuais em paralelo, traçando os caminhos que as leituras e os pares de leituras tomam dentro do gráfico, relatando transcrições completas para isoformas emendadas alternadamente e separando as transcrições que correspondem a genes parálogos.

Neste estudo, para otimizar a montagem dos *contigs* os dados foram normalizados a partir das *reads* filtradas usando a ferramenta *in_silico_read_normalization.pl* que é parte do pacote do *Trinity* (filtra *reads* repetidas ou redundantes) (GRABHERR et al. 2011). A profundidade utilizada nessa etapa para evitar a redundância foi de até 60 repetições. Para essa montagem o tamanho mínimo adotado para os *contigs* foram 200 pb.

4.7 Alinhamento das Sequências.

Duas estratégias de mapeamento de leituras de RNA-Seq foram adotadas. A primeira estratégia foi alinhar as leituras com o genoma de referência de *T. infestans* (GCA_011037195.1). A segunda foi utilizando os transcritos obtidos neste estudo como referência, ou seja, por meio de estratégia "de novo". As montagens com índices de qualidade superior foram utilizadas para a construção de banco de dados de referência. Dessa forma obteve-se índices adequados para avaliar a qualidade das montagens. A estratégia foi adotada devido à falta de genomas de referência disponíveis nos bancos de dados genéticos. Até o momento não existem dados de espécies próximas a *Mepraia*.

Para evitar a redundância no mapeamento das 18 montagens deste estudo, as maiores isoformas por unigene detectado foram filtradas durante a montagem. Assim, obteve um transcrito por conjunto de isoformas. Para isso o programa *Bowtie2* (LANGMEAD e SALZBERG, 2012) foi utilizado. O *Bowtie2* permite comparar-se sequências biológicas. Confrontar os dados obtidos com sequências é importante para termos uma ideia da qualidade dos dados, sendo que os seus resultados obtidos oferecem índices importantes para avaliar as montagens do transcriptoma.

4.8 Previsão de unigenes e avaliação da qualidade das montagens

Para identificar regiões codificantes putativas dentro das sequências de transcrição geradas pelo montador Trinity foi realizado por um pipeline fundamentado no Transdecoder v5.4.0 (https://github.com/TransDecoder/TransDecoder/releases).

A predição das regiões codificantes (open reading frames "ORFs") para as montagens por espécie foi com o pipeline realizado em três etapas. Primeiro, TransDecoder.LongOrfs é usado para reter todos os CDS potenciais que codificam os peptídeos por mais de 100aa. Na segunda etapa, esses peptídeos são submetidos à ferramenta hmmscan incluída no pacote HMMER v.3.1b2 (EDDY, 2011) para busca de assinaturas de proteínas no banco de dados Pfam-A e NCBI Blastp 2.10.1 (CAMACHO et al., 2009) para pesquisar sequências semelhantes na base de dados não redundante do Genbank (nr), incluindo apenas proteínas de artrópodes. Na terceira etapa, o programa TransDecoder.Predict usa as informações produzidas pelas outras etapas para prever os CDS. A redundância dos CDS obtidos foi minimizada usando Cd-hit-est (FU et al., 2012) com um limite de similaridade de 0,99. Para obter o conjunto final de unigenes putativos, os transcritos com esses CDS foram filtrados usando as informações de montagem dos transcriptomas e apenas as isoformas com a expressão mais alta por componente foram mantidas. Para isso, as leituras de cada espécie foram mapeadas para o respectivo conjunto usando Bowtie2 (LANGMEAD e SALZBERG, 2012). Essas etapas foram executadas pelo script *align_and_estimate_abundance.pl* incluído no pacote Trinity. O nível de completude e redundância dos conjuntos brutos e filtrados de cada espécie foi avaliado por BUSCO (Benchmarking Universal Single-Copy Orthologs) (SIMÃO et al., 2015) usando o banco de dados Arthropoda como referência.

4.9 Anotação funcional

Uma pesquisa foi realizada para buscar sequências semelhantes às CDS em banco de dados de proteínas não-redundantes (*nr*) de artrópodes disponível no *GenBank* e para validar as montagens feitas por bioinformática, foi feito uma análise de homologia com o genoma de referência: *T. infestans* (GCA_011037195.1). Para isso foi utilizado o programa *NCBI Blastp* 2.10.1 (CAMACHO et al., 2009) adotando um limiar de *evalue* de 10⁻⁶.

Também se pesquisou por domínios de proteína conservados usando *InterProScan* 5.24-63.0 (JONES et al., 2014). A anotação funcional visando associar os termos da *Gene Ontology* com cada CDS foi realizada com o programa *Blast2GO* (CONESA et al. 2005). As anotações que foram obtidas pelo passo anterior foram analisadas por software *WEGO* (http://wego.genomics.org.cn/cgi-bin/weGO/index.pl) para exibir a distribuição dos termos *GO*

do nível 2 encontrados nos transcriptomas de cada uma das espécies e suas distribuições foram plotadas usando termos GO com frequências maiores que 1%.

4.10 Busca de Variantes.

Para minimizar e avaliar o efeito da referência na busca por variantes do tipo polimorfismo de nucleotídeo único ou do inglês *single polymorphism nucleotide (SNP)*, foram aplicadas duas estratégias para as referências: 1) o genoma de *T. infestans* (GCA_011037195.1) e 2) as montagens obtidas neste estudo. O *pipeline* está ilustrado na Figura 8. As leituras de cada biblioteca foram mapeadas para cada montagem usando *Bowtie2* (LANGMEAD & SALZBERG, 2012).

As leituras mapeadas (formato BAM) foram filtradas por *MarkDuplicates* implementada no programa *Picard* (<u>https://github.com/broadinstitute/picard/</u>). A seguir as sequências foram alinhadas e convertidas para o formato *mpileup* utilizando *SamTools* (LI et al., 2009) e a opção *samtools_view*. e filtradas com base na qualidade de mapeamento (>20) e qualidade PHRED >30 usado *Samtools mpileup* implementada no pacote Samtools. Os arquivos convertidos foram utilizados para a pesquisa de *SNP* utilizando o comando *mpileup2SNP* do *VarScan2* (KOBALDT, 2012). Para uma análise mais acurada a fim de evitar redundância de fragmentos, filtrou-se as isoformas que compartilham regiões semelhantes. Assim, pode-se obter um transcrito por cada conjunto de genes. Na análise foi utilizado uma cobertura mínima de 42, qualidade de mapeamento maior que 30 e qualidade *Phred* maior que 30.



Figura 8. Fluxo de trabalho adotado neste estudo para detectar variantes do tipo SNP de dados de sequenciamento do RNA-seq.

4.11 Inferência de ortologia e análise filogenética

A predição de ortologia dos CDS putativos foi realizada por meio do software Orthofinder v2.5.4 (EMMS e KELLY, 2019). O programa busca por ortogrupos e ortólogos, infere árvores gênicas enraizadas para todos os ortogrupos e identifica todos os eventos de duplicação de genes nessas árvores gênicas. Também como abordagem comparativa utilizouse a ferramenta BLASTN do NCBI (CAMACHO et al., 2009) com parâmetros configurados para o software MCL (VAN DONGEN, 2000). Para inferir as relações filogenéticas do gênero Mepraia na subfamília Triatominae recuperamos dados brutos de transcriptomas e genoma (Sequence Read Archive (SRA) - Genbank) de Panstrongylus megistus (SRR1304639), Psammolestes tertius, T. brasiliensis (ERR1276857), T. dimidiata (SRR2899821) T. infestans (SRR4427079), T. palladipennis (SRX2600752), T. rubrofasciata (SRR8466737), Rhodnius domesticus (SRR6750000) e R. nasustus (SRR6750002). Os dados obtidos do GenBank foram submetidos aos mesmos procedimentos descritos no Capítulo 4.7 desta tese, ou seja, controle de qualidade, trimagem, montagem, validação da montagem. A única exceção foi P. tertius (SRR6750002) que por ser um genoma teve a montagem realizada no software SPAdes v. version: 3.15.3 (PRJIBELSKI et al., 2020). Os ortólogos obtidos na busca realizada com BLASTN foram alinhados com MAFFT v.7.305b (KATOH; STANDLEY, 2013) e filtrados por Phyutility v.2.2.6 (SMITH; DUNN, 2008) e usados para inferir árvores filogenéticas usando RAxML v.8.2.9 (STAMATAKIS, 2014).

Os grupos de ortólogos recuperados Orthofinder v2.5.4 (EMMS e KELLY, 2019) obtidos dos proteomas foram convertidos em DNA com auxílio do *TranslatorX* (ABASCAL; ZARDOYA; TELFORD, 2010). Os conjuntos de nucleotídeos foram alinhados com auxílio do MAFFT v.7.305b (KATOH; STANDLEY, 2013) e foram avaliadas por dois métodos filogenéticos: a estimativa por máxima verossimilhança (ML) e inferência bayesiana (IB). A ML foi realizada usando uma otimização de máxima verossimilhança (ML) no IQ-TREE (NGUYEN et al., 2014). A inferência Bayesiana foi realizada no *BEAST* 2.5 (BOUCKAERT et al., 2018), se inferiu o modelo de substituição de nucleotídeos, intervalo da taxa de heterogeneidade e proporção de posições invariantes durante a análise MCMC com o pacote *bModelTest* (BOUCKAERT e DRUMMOND, 2017), com opção de divisão de transição-transversão e frequências empíricas. Executou-se amostragem de 10.000.000 gerações a cada 1.000 gerações e usou-se o *TRACER* (RAMBAULT et al., 2018) para confirmar a cobertura da cadeia (ou seja, tamanho de amostra efetivo > 200). *TreeAnnotator* (BOUCKAERT et al., 2018) foi usado para construir uma árvore de consenso e as árvores

iniciais de 10% foram descartadas como *burn-in*. As filogenias foram visualizadas e editadas no *Figtree* (RAMBAUT, 2012).

4.12 Estimativa de árvores de espécies

O foco de análise foram as relações entre as espécies de *Mepraia*, portanto foram inferidas árvores de espécies com *T. infestans* como grupo externo. Para isso um conjunto de SNPs foram obtidos a partir do mapeamento tendo como referência os transcriptomas obtidos nesse estudo. Assim, as leituras mapeadas foram filtradas pela ferramenta *MarkDuplicates* implementada no programa *Picard* (<u>https://github.com/broadinstitute/picard/</u>). A seguir as sequências foram alinhadas e convertidas para formato *mpileup* utilizando a ferramenta *SamTools* (LI et al., 2009). Os arquivos convertidos foram utilizados para a pesquisa de *SNP* utilizando o comando *mpileup2SNP* do *VarScan2* (KOBALDT, 2012). Para uma análise mais acurada a fim de evitar redundância de fragmentos, filtrou-se as isoformas que compartilham regiões semelhantes. Assim, pode-se obter um transcrito por cada conjunto de genes. Na análise foi utilizado uma cobertura mínima de 42, qualidade de mapeamento maior que 30 e qualidade *Phred* maior que 30.

Um conjunto de 60.685 SNPs foi obtido e utilizado para reconstruir filogenias sob três metodologias distintas: a primeira por meio do *SVDquartets* (CHIFMAN; KUBATKO, 2014) que é implementado no PAUP * v4a168. Nestas abordagens utilizamos SNP bi alélicos obtidos anteriormente [estes mapeados pela referência (transcritos obtidos neste estudo), filtrados para eliminar duplicatas, e acessados pelo *Varscans2* com descrito no capítulo 4.11] para construir uma matriz de *SNP* no formato nexus. A matriz gerada foi analisada no SVDquartets+PAUP*. A árvore foi gerada a partir de 100.000 quartetos e 1.000 réplicas (*bootstrap*). O modelo estimado pelo critério de informação de Akaike (AIC) foi TVM+G.

A segunda abordagem aplicada para estimar uma árvore de espécie foi no por meio do pacote *StarBEAST2* (HELED e DRUMMOND, 2010) implementado no BEAST2 (BOUCKAERT et al., 2018). O pacote permite aplicar um método totalmente Bayesiano de estimativa de árvore de espécies (HELED e DRUMMOND, 2010) e suporta a aplicação de um relógio relaxado à árvore da espécie. Isso permite estimar as taxas de substituição de espécies existentes e ancestrais em um modelo coalescente multiespécies. Foi inferido o modelo de substituição de nucleotídeos, intervalo da taxa de heterogeneidade e proporção de posições invariantes durante a análise MCMC com o pacote *bModelTest* (BOUCKAERT e DRUMMOND, 2017), com opção de divisão de transição-transversão e frequências empíricas. Executou-se amostragem de 10.000.000 gerações a cada 1.000 gerações por meio do *TRACER* (RAMBAULT et al., 2018) para confirmar a cobertura da cadeia (ou seja, tamanho de amostra efetivo> 200). *TreeAnnotator* (BOUCKAERT et al., 2018) foi usado para construir uma árvore de consenso e as árvores iniciais de 10% foram descartadas como *burn-in*. As árvores foram visualizadas no *Figtree* (RAMBAUT, 2012).

A terceira abordagem utilizada para inferir as linhagens foi por meio do pacote SNAPP (BRYANT et al., 2012) (do inglês SNP and AFLP Package for Phylogenetic analysis) implementado no BEAST2 (BOUCKAERT et al., 2018). O método permite inferir árvores de espécies e dados demográficos de espécies de marcadores bialélicos, como SNP (BRYANT et al., 2012). Como o BEAST, uma análise SNAPP não produz uma única árvore. Em vez disso, ele usa a cadeia de Markov Monte Carlo (MCMC) para gerar várias árvores (e os valores dos parâmetros correspondentes), cada um dos quais é uma amostra da distribuição posterior das árvores e parâmetros das espécies. As árvores e os valores dos parâmetros são produzidos como dois arquivos. Como em todas as análises MCMC, deve-se tomar cuidado ao interpretar essas saídas. Por exemplo, é necessário verificar a convergência e estar preparado para considerar árvores de múltiplas espécies plausíveis, ao invés de apenas uma. Executou-se amostragem de 10.000.000 gerações a cada 1.000 gerações e utilizou-se o TRACER (RAMBAULT et al., 2018) para confirmar a cobertura da cadeia (ou seja, tamanho de amostra efetivo> 200). TreeAnnotator (BOUCKAERT et al., 2018) foi usado para construir uma árvore de consenso e as árvores iniciais de 10% foram descartadas como burn-in. As árvores foram visualizadas no Figtree (RAMBAUT, 2012) e para gerar árvores de densidade essas árvores foram submetidas ao Densitree (BOUCKAERT, 2010).

Também se realizou a análise de delimitação de espécies sob um conceito de espécie filogenética. Para isso o servidor PTP (https://species.h-its.org/) uma implementação Bayesiana do modelo PTP para delimitação de espécies foi utilizado. A PTP é um modelo para delimitar espécies em uma árvore filogenética enraizada. No PTP, modela-se as especiações ou eventos de ramificação em termos de número de substituições (ZHANG et al., 2013). As árvores resultantes da análise gerada com *StarBEAST2* foram utilizadas como entrada para análise. A análise gera um valor de probabilidade posterior e as estimativas mais altas em um nó indica que todos os descendentes deste nó são mais propensos a ser de uma espécie (ZHANG ET AL., 2013). A análise seguiu: 500.000 MCMC gerações, desbaste = 100 e *burn-in* = 0,1 (ZHANG ET AL., 2013).

4.13 Montagem, anotação e filogenia com mitogenomas.

Atualmente o número de mitogenomas disponíveis é superior ao número de transcriptomas de Triatominae, por esse motivo, somadas pelas taxas evolutivas do genoma mitocondrial e pela possibilidade de explorar as relações filogenéticas entre triatomíneos, foram obtidos o genoma mitocondrial das três espécies de *Mepraia*. Para isso utilizou-se *reads* para recuperar os genomas mitocondriais, não havendo a necessidade de técnicas moleculares. Uma vez que o mtDNA animal tem uma taxa de evolução maior que os genes nucleares, representa um pilar da filogenética e da biologia evolutiva principalmente entre espécies relacionadas filogeneticamente. Além de que fatores como a ausência de processos recombinantes e a alta taxa de evolução destaca os mitogenomas como um dos principais marcadores utilizados para a realização de estudos evolutivos e taxonômicos de vertebrados e invertebrados (AVISE, 2009., KOLESNIKOV e GERASIMOV, 2012., SILVA et al., 2020).

Para obter o genoma mitocondrial (mtDNA) das espécies de *Mepraia* utilizou-se abordagem baseada em referência. Os mtDNA utilizados como referência foram: *Triatoma dimidiata* (17.019bp, NC_002609.1), *Triatoma infestans* (17.068 bp, NC_035547.1), *Triatoma longipennis* (15.703 bp, MT556658.1), *Triatoma recurva* (16.125 bp, MT556663.1) e *Triatoma dimidiata* (17.019bp, NC_002609.1). Foram utilizadas os transcriptomas de *M. gajardoi* (*M. gajadoi*7, *M. gajadoi*8, *M. gajadoi*9 e *M. gajardoi*10), *M. parapatrica* (*M. parapatrica*13, *M. parapatrica*14 e *M. parapatrica*15) e *M. spinolai* (*M. spinolai*3 e *M. spinolai*4).

As montagens foram realizadas seguindo o protocolo padrão do software MITObim disponível no https://github.com/chrishah/MITObimA. O protocolo consiste em usar um mtDNA próximo como "isca" para montagens de novo. Utilizou-se o MITObim 1.9.1 (HAHN; BACHMANN; CHEVREUX, 2013) para reconstruir os genomas mitocondriais e o mapeamento contra referências foi com programa **MIRA** 3.4.1.1 as 0 (https://www.chevreux.org/projects_mira.html). O MITObim é uma abordagem in silico para a reconstrução de genomas mitocondriais completos de organismos não-modelo diretamente a partir de dados de sequenciamento de última geração (HAHN; BACHMANN; CHEVREUX, 2013). Com estratégia de averiguação das montagens as anotações foram realizadas em duas plataformas distintas: MITOS (BERNT et al., 2013) e Mfannot disponível no domínio: https://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl. As anotações foram verificadas em completude e qualidade usando como referência os mitogenoma de T. dimidiata (DOTSON e BEARD, 2001) e T. infestans (PITA et al., 2017).

As estimativas de divergência genética foram calculados par a par no programa MEGA X (KUMAR et al., 2018). As análises foram realizadas usando o modelo de 2 parâmetros Kimura (KIMURA, 1980). A variação da taxa entre os locais foi modelada com uma distribuição gama (parâmetro de forma = 1). Também foi estimado o teste relativo de de Tajima (TAJIMA, 1989) tendo como input as 12 sequências de mtDNA no MEGA X (KUMAR et al., 2018).

Todas as posições ambíguas foram removidas para cada par de sequências (opção de deleção em pares).Para avaliar as relações filogenéticas de *Mepraia* com espécies de Triatominae recuperou-se todos os mtDNA disponíveis e assim obteve-se uma filogenia seguindo três abordagens: estimativa de máxima verossimilhança (ML), por inferência bayesiana (BI) e para avaliar relações filogenéticas estritas com *T. breyeri* e *T. eratyrusiformis* inferiu-se uma filogenia com dois genes mitocondriais, como descrito a seguir.

Primeiro a partir de busca no Genbank selecionou-se mtDNA (gêneros próximos) disponíveis na plataforma. A busca no *Genbank* resultou nos mitogenomas completos de *Panstrongylus rufotuberculatus* (NC_042682.1), *Triatoma infestans* (NC_035547), *Triatoma dimidiata* (NC_002609.1), *Triatoma longipennis* (MT556658.1), *Triatoma picturata* (MT556661.1), *Triatoma protacta* (MT556662.1), *Triatoma recurva* (MT556663.1), *Triatoma vitticeps* (MT556665.1). Além disso, para avaliar relações filogenéticas com espécies mais próximas, mas com poucas sequências disponíveis, foi construída uma árvore filogenética por inferência bayesiana utilizando os genes COI e Cyt B de *T. breyeri* e *T. eratyrusiformis*, espécies relacionadas com *Mepraia* [genes COI e Cyt b, dados particionados, modelo evolutiva para ambos GTR+G+I no *MrBayes* (RONQUIST et al., 2012)].

As 23 sequências foram alinhadas com MAFFT versão 7 (KATOH, 2013). Também se verificou a existência de potenciais códons de parada internos com o programa Mesquite versão 3.51 (MADDISON E MADDISON, 2018). A estimativa ML foi usando o IQTREE (NGUYEN et al., 2015) a razão de verossimilhança aproximada foi corrigida por modelo não paramétrica de Shimodaira-Hasegawa (SH-aLRT) (GUINDON et al., 2010) e estimado pelo suporte para cada nó com 1.000 réplicas. As árvores foram visualizadas e editadas no *Figtree* (RAMBAUT, 2012). A inferência bayesiana foi realizada no *MrBayes* (RONQUIST et al., 2012). A cadeia de Markov Monte Carlo (MCMC) foi usada para aproximar a posterior (METROPOLIS et al., 1953) com duas corridas de 4 cadeias MCMC por 10.000.000 gerações, com amostragem a cada 5.000 gerações. Árvores ótimas foram então amostradas a

cada 1.000 gerações para obter a árvore de consenso final e probabilidades posteriores associadas. As árvores foram visualizadas e editadas no *Figtree* (RAMBAUT, 2012).

4.14 Teste de introgressão.

Para testar sinais de hibridação entre as linhagens de *Mepraia*, aplicou-se o teste ABBA-BABA (PATTERSON et al., 2012). Seis amostras de cada espécie foram mescladas com a opção "*merge*" do *SamTools* (LI et al., 2009)". Foi utilizado *T. infestans* como grupo externo e como referência para mapeamento. Os testes foram realizados por meio do script Abbababa implementado por *ANGSD* (KORNELIUSSEN; ALBRECHTSEN; NIELSEN, 2014). O teste ABBA-BABA permite comparar a segregação do polimorfismo bialélicos de quatro amostras (H1, H2, H3 e H4), onde H1, H2 e H3 representam os grupos internos e H4 o grupo externo. A estimativa de introgressão foi realizada usando-se a estatística D, que foi calculada a partir dos desvios das proporções de alelos compartilhados entre H2 e H3 (ABBA) e H1 e H2 (BABA). Também se utilizou abordagem *jack-knife (boostrap)* para calcular o D corrigido para o viés, erro padrão estatística D e teste Z (SORAGGI, WIUF e ALBRECHTSEN, 2017).

5. RESULTADOS

5.1 Extração do RNA total.

Foi possível extrair o RNA total da cabeça e glândulas salivares de *M. gajardoi, M. parapatrica e de M. spinolai*. As amostras foram previamente avaliadas em concentração e quantidade e enviadas posteriormente para sequenciamento. A relação de amostras enviadas para sequenciamento e os dados de quantificação (*Nanodrop* e *Qubit*) estão sumarizados na tabela (Anexo C). Possíveis degradações foram avaliadas por gel de agarose desnaturante, a técnica permitiu visualizar a integridade das amostras (Anexo D). Os índices de qualidade foram adequados para o sequenciamento na plaforma *Illumina*.

5.2 Determinação da integridade do *RNA*, Bibliotecas de *cDNA* e Controle de Qualidade de Dados.

Para determinar a qualidade e integridade das amostras utilizou-se o sistema Agilent 2100 Bioanalyzer, (Agilent Tecnologies®, protocolo padrão para RNA). A avaliação permitiu

qualificar e quantificar cada amostra de *RNA*. Os parâmetros utilizados foram a concentração de RNA, a relação 28S/18S e o *RNA Integrity Number* – RIN, sendo esse último, o de maior importância para o sequenciamento. A concentração média de *RNA* foi de 69.44±9.1 e a razão 28S/18S foi de 0.01±0. A quantidade de *RNA* é importante para a realização do sequenciamento, portanto, no requisito concentração, todas as amostras corresponderam às exigências para o sequenciamento de nova geração. Quanto à razão 28S/18S, os valores foram relativamente inferiores para o esperado para eucariotos, mas o perfil é o esperado para amostras de insetos, devido a "exclusão de gap" (também chamada de "interrupção oculta") no rRNA 28S. Os dados estão sumarizados na tabela (Anexo E).

A eletroforese capilar apresentou *RNA* íntegro em todas as amostras (anexo G), sem rastros abaixo das bandas, indicando que não houve degradação e nem contaminação por DNA genômico, é preciso estabelecer a integridade dos materiais a serem utilizados.

O valor de RIN foi calculado por um algoritmo que analisa os resultados obtidos e atribui uma pontuação para as amostras, que varia de 1 a 10. Para amostras obtidas de tecidos de insetos, o valor de RIN igual ou superior a 8 é considerado ótimo e as amostras foram utilizadas para análises posteriores. O valor obtido do RIN foi 7.1 ± 0.3 em um grupo amostral de 18 (Tabela 2). Portanto, o valor de RIN, obtido nas amostras do presente estudo, foi satisfatório considerando a material de extração. Os valores gerados pelo *Bioanalyzer* em conjunto com os dados de quantificação (*Nanodrop e Qubit*) e o gel desnaturante garantiram que as amostras de RNA estavam adequadas para preparação das bibliotecas de *cDNA*.

Por meio da plataforma *Illumina HiSeq2500* com *2x100 pb paired-end* obteve-se 18 amostras, sendo seis réplicas por espécies. As bibliotecas de RNA-Seq foram construídas para cada amostra. Essas subsequentemente utilizadas para a montagem *de novo* de bancos de RNA-Seq.

Para a busca por sequências específicas, é preciso que as bibliotecas de cDNA, gravadas em formato *FASTQ*, sejam avaliadas e filtradas. A filtragem e o preparo das bibliotecas de cDNA é necessário pois adaptadores, primers precisam ser removidos, além de trechos de baixa qualidade. Esses adaptadores podem representar um problema real para o a realização das próximas etapas do RNA-Seq.

5.3. Montagem dos transcriptomas de novo.

As 18 bibliotecas de RNA (6 réplicas por técnicas por espécie) geraram um total de 267.766.381 x 2 *reads* de 100 pb que representam em média 15 milhões por biblioteca (2x

reads). As bibliotecas das 6 réplicas geraram 93.415.430, 91.093.659 e 83.257.292 x 2 reads para *M. spinolai, M. gajardoi* e *M. parapatrica*, respectivamente (Anexo H).

Esses *reads* gerados foram submetidos a filtragem e manteve-se apenas as sequências com um escore mínimo de qualidade *Phred* 20, ou seja, selecionou-se regiões com qualidades superiores a 20 (chances de erro estimadas de 1 em 100). O escore ampara o reconhecimento e remoção de sequências de baixa qualidade e permite a manutenção de sequências precisas. Depois que todos os arquivos estão preparados é possível iniciar a busca pelas sequências de interesse em nossa biblioteca de cDNA. Após a filtragem obteve-se um total de 239.843.935 pair-end-reads que representa em média 14 milhões de *reads* (Anexo H). Cada biblioteca teve em média em 10% do total aparado. Assim, após a filtragem resultaram 83.343.041, 81.994.244 e 74.506.650 x 2 *reads de M. gajardoi* e *M. parapatrica* e *M. spinolai,* respectivamente. Os percentuais de CG nas bibliotecas foram acima de 37% com a grande maioria entre 38 e 39%. As descrições detalhadas *reads* brutas e filtradas podem ser acessadas na tabela (anexo H).

As montagens *de novo* foram realizadas seguindo a funções "*default*" do montador *Trinity*. As descrições das 18 montagens podem ser encontradas na tabela (Anexo h). As amostras apresentam índices semelhantes de N50, conteúdo CG, comprimento, entre outros parâmetros descritivos (Anexo I).

As avaliações de completude com o BUSCO realizada a partir de ortólogos conservados de Artrópodes (2510 grupos conservados), permitiu identificar ortólogos completos compartilhados com o banco de dados de Artrópodes. Os percentuais de ortólogos completos entre as réplicas de *M. spinolai* 83.4% a 90.1% e de 2 a 8% de fragmentados, entre as réplicas de *M. gajardoi* foram identificados 71 a 92 % ortólogos completos e de 2 a 7% de grupos fragmentados e *M. parapatrica* teve percentuais de 88 a 91.9% de grupos BUSCO completos e 3 a 4 % fragmentados (anexo J). Copias duplicadas foram encontradas em todas as montagens brutas, no entanto a redundância em unigenes foi filtrada.

5.4. Anotação funcional

Os dados de todas as predições de CDS e das anotações funcionais estão sumarizadas na tabela (Anexo J). As predições de CDS com a ferramenta *Transdecoder* a partir dos bancos de dados do PFAM e do *Swiss-Prot* encontrou um número médio de 20.785 CDS (máximo de 24.103 e mínimo de 17.881) nas réplicas de *M. spinolai*, as réplicas de *M. gajardoi* obteve 22.955 CDS (máximo 27527 e mínimo de 15774, no entanto a maioria esteve acima de

20.000), e por fim *M. parapatrica* teve a média de 22.028 CDS (máximo de 24.175 e mínimo 19.857). A busca por proteínas redundantes por BLASTP tendo como referência sequências de Arthropoda resultou em média 12.366, 13.305 e 13.726 para *M. spinolai*, *M. gajardoi* e *M. parapatrica*, respectivamente. No entanto entre as réplicas de *M. spinolai* amostras atingiram mais de 16 mil matches, já as réplicas de *M. gajardoi* obtiveram máxima de 19.461 e *M. parapatrica* com um pouco menos, 17 mil matches aproximados.

Ao buscar por CDS no banco *InterPro*, que permite classificar as famílias proteicas identificadas nas montagens recuperou-se 12.038 famílias proteicas nas montagens de *M. spinolai* (máximo de 16.010 e mínimo de 8.824), 14.145 nas montagens de *M. gajardoi* (máximo de 18.570 e mínimo de 7.273, no entanto a maioria foi acima de 11.000) e *M. parapatrica* foram 12.885 (máxima de 15.275 e mínimo de 11.800) a média das famílias recuperadas para suas réplicas.

As anotações funcionais preditas com auxílio do *Blast2GO* e compiladas pelo WEGO forneceu informações detalhadas sobre os grupos proteicos (Anexo L). A predição recupera nas 18 montagens de transcriptoma em média 10.200 genes anotados entre as réplicas das espécies. Nas montagens os termos de ontologia funcionais foram mais representativos, seguido dos termos biológicos e celulares. Em média nas 18 amostras descreveu-se 8.684 GO termos funcionais, 5.139 biológicos e 2.736 celulares (Anexo L).

As CDS anotadas neste estudo foram distribuídas em 45 categorias ontológicas preditas pelo *WEGO* (https://wego.genomics.cn/) que representam três classes: Processo biológico (n=16), Função molecular (n=14) e Componente Celular (n=15) (Figura S3). Em todas as espécies o processo biológico referente à atividade catalítica e ligação foram os termos GO mais abundantes, na função molecular: regulação biológica, metabolismo, processos celulares e sistema imunológico foram mais abundantes, já nos componentes celulares: respostas a estímulos e sinalização foram mais abundantes, como é esperado, já que os transcritos estudados provem de tecido cefálico, onde ficam os principais órgãos sensoriais dos insetos. Outras categorias ontológicas podem ser visualizadas com detalhes na figura (Anexo M)

5.5 Inferência de ortologia e análise filogenética.

A busca por ortologia com um conjunto de 10 espécies resultaram em mais de 500 mil genes (média de 20 mil por espécie) e 487.628 ortogrupos (em média 20 mil por sequência) com percentuais médios de 97,4% de genes ortólogos entre as sequências das espécies de *Mepraia*. Todos os dados estão sumarizados na tabela (Anexo M). *Mepraia gajardoi* e *M. parapatrica* compartilham um número maior de genes ortólogos em relação *M. spinolai*. Entre todas as espécies os menores índices foram vistos em *P. tertius* seguido de *R. prolixus*. Além do mais as espécies de Rhodniini apresentam a maior quantidade de genes específicos, resultado esperado, pois são espécies filogeneticamente distantes (Anexo N).

Junto à busca por ortologia foi estimada uma árvore de espécie enraizada pelo algoritmo MSA, utilizado junto às buscas com o Orthofinder (Figura 9). O método recupera filogenia com suporte *boostrap* elevado (>90) onde os clados das espécies de *Mepraia* estão bem definidos. *Mepraia gajardoi* e *M. parapatrica* formam um clado monofilético separando *M. spinolai*. Aspecto interessante da árvore é a parafilia de *P. megistus* com o gênero *Triatoma* (Figura 9). A topologia recuperada são coerentes com as filogenias moleculares de Triatominae (JUSTI et al., 2014., MONTEIRO et al., 2018). Para melhor mostrar as relações entre as espécies de *Mepraia* e os demais gêneros reconstruiu-se uma árvore de genes a partir de um ortogrupos (Figura 10). De maneira resumida, três clados foram recuperados, o clado azul agrupa em maioria genes de espécies de *Mepraia*. E por fim, o clado verde, além de menor agrupam genes compartilhados entre a maior parte das espécies.



800.0

Figura 9. Árvore inferida pelo Método de arvore de espécies de alinhamento de Sequências Múltiplas (MSA). Nesta análise foram utilizados um conjunto de genes de cópia única de 10 espécies de Triatominae. As cores indicam o suporte *bootstrap*, sendo vermelho suporte acima de 95%.





Figura 10. Filogenia de Genes recuperada de 700 genes ortólogos de 10 espécies de Triatominae selecionados aleatoriamente. Cada cor representa um clado. O clado roxo apresenta maior número de genes de *Mepraia*, Azul e verde genes compartilhados em Triatominae.

A estimativa de ML a partir de 10 sequências de conjuntos de ortólogos de cópia única alinhados foi realizada a partir de uma matriz de 24 espécies alinhadas com o tamanho de 505.749bp. De forma resumida foram recuperados um filograma formado por dois grandes clados: o primeiro composto por espécies da Tribo Rhodniini (*P. tertius, R. robustos* e *R. prolixus*) e o segundo com espécies de Triatominii (*Panstrongylus, Mepraia, Rhodnius* e *Triatoma*) (Figura 11). Os dois clados apresentam parafilias, *P. megistus* em relação as espécies de *Triatoma* e *P. tertius* em relação as espécies de *Rhodnius*. Na filogenia estimada por ML o gênero *Mepraia* é monofilético (Figura 10). Na filogenia *M. gajardoi* e *M. parapatric*a estão próximas, *M. spinolai* forma um clado independente (Figura 11).

A inferência bayesiana recupera clados bem resolvidos em relação à estimativa ML, no entanto *P. megistus* e *Triatoma* mantem-se parafiléticos (BPP. 100, Figura 12). A tribo Rhodniini e Triatominii recuperadas como clados monofiléticos são coerentes com outras filogenias (BPP. 100, Figura 12). Como na estimativa ML, *M. gajardoi* e *M. parapatrica* (BPP. 100, Figura 12) possuem relação filogenética mais estrita em relação a espécie *M. spinolai*. As populações silvestres de *M. spinolai* (*M. spinolai* 1 e *M. spinolai* 2) e *M. parapatrica* (*M. parapatrica* 13 e *M. parapatrica* 14) divergem das populações coletadas em ambiente peridomiciliares, com exceção de MP12 que forma um clado com espécimes coletados em ambiente peridomiciliares. Todas as populações de *M. gajardoi* 3ão silvestres, mas a *M. gajardoi* 7 e *M. gajardoi* 8 formam clado independente de *M. gajardoi* 10, *M. gajardoi* 11 e *M. gajardoi* 20 em 2000 populações de Iquique na região de Tarapacá.



Figura 11. Filogenia estimada por máxima verossimilhança (ML) com genes ortólogos de cópia única de 10 espécies de Triatominae. A árvore foi estimada por um algoritmo estocástico implementado ao software IQ-TREE. A estimativa de ML a partir de 10 sequências de conjuntos de ortólogos de cópia única alinhados foi realizada a partir de uma matriz de 24 espécies alinhadas com o tamanho de 505.749 bp. Os valores nos nós são relativos ao *bootstrap* não paramétrico.



Figura 12. Filogenia com 10 matrizes alinhadas de sequências de DNA de ortólogos de cópia única estimada por inferência bayesiana pelo método de cadeia de Markov Monte Carlo (MCMC). Os valores de probabilidade superior podem ser visualizados entre os nós dos ramos.

5.6 Estimativa de árvores de espécies

Neste estudo recuperou-se árvores de espécies por meio de métodos diferentes e as topologias recuperadas foram coerentes, ou seja, as topologias foram semelhantes. Os 19 conjuntos de 60.085 SNPs agrupados por espécies (6 réplicas de cada espécie de *Mepraia* e 1 grupo externo) foram utilizados para determinar linhagens a partir do modelo coalescente. O *SVDquartets* produziu a topologia final baseada em quartetos que recupera, uma filogenia com suporte robusto (Figura 13). As estimativas recuperam filograma no qual *M. spinolai* é a linhagem ancestral de *M. gajardoi* e *M. parapatrica* (Figura 13).



Figura 13. Inferência de espécie-árvore realizada um conjunto de SNPs de 4 espécies de Triatominae. A espécie *T. infestans* foi utilizada como grupo externo para avaliar as relações de *Mepraia*. Os valores de suportes de quarteto estimados por SVAquartets são mostrados junto aos ramos.

A filogenia pelo modelo multiespécies coalescente recuperado com um conjunto de SNPs foi estimada com auxílio do pacote STARBEAST implementado no BEAST2 recupera, topologia enraizada com *T. infestans* como grupo externo e que mostra uma taxa de probabilidade posterior acima de 90 (Figura 14). As espécies *M. gajardoi* e *M. parapatrica* apresentam maior parentesco e *M. spinolai* aparece com linhagem ancestral. Entre as réplicas, as espécies de *M. gajardoi* são bem relacionados entre si, no entanto *M. spinolai1* e *M. spinolai2*, réplicas de *M. spinolai*, estão separadas das demais, o mesmo percebido entre as réplicas de *M. parapatrica* formada por trios de táxons separados (Figura 14).

As topologias geradas com a estimativa SNAPP mostram coerência com os resultados das demais análises (Figura 15). Para mostrar com clareza a topologia das linhagens recuperadas espelha-se o dendrograma sob uma árvore de densidade (Figura 14). A figura permite visualizar com clareza as áreas onde muitas árvores concordam em topologia e comprimento de ramos aparecem como áreas altamente coloridas, enquanto áreas com pouca concordância aparecem como teias. Nos resultados mostra-se a ancestralidade de *M. spinolai* e a relação de proximidade entre *M. gajardoi* e *M. parapatrica*. No entanto também é possível visualizar incongruências entre as topologias espelhadas (Figura 15).

As espécies putativas presentes na árvore gerada pelo modelo multiespécies coalescente com SNPs. foram avaliadas por ML e IB sob modelo PTP de delimitação de espécies (modelo PTP) e produziram resultados semelhantes (Figura 16). Foi possível descrever linhagens putativas coerentes com as espécies descritas (Figura 16). Todos os ramos em vermelho indicam uma linhagem única. Os resultados recuperam três clados sustentados pelos valores de probabilidade posterior (figura 16).



Figura 14. Inferência de espécie-árvore realizada com um conjunto de SNPs de quatro espécies de Triatominae. A espécie *T. infestans* foi utilizada como grupo externo para avaliar as relações de *Mepraia*. As topologias sobrepostas foram estimados pelo modelo multiespécies coalescente (*STARBEAST*) implementado no BEAST2;



Figura 15. Inferência de espécie-árvore reconstruída por um conjunto marcadores bialélicos do tipo SNPs de quatro espécies de Triatomiane. A espécie *T. infestans* foi utilizada como grupo externo para avaliar as relações de *Mepraia*. O algoritmo utilizado por *SNAPP* implementa um modelo coalescente completo, mas usa um novo algoritmo para integrar todas as árvores genéticas possíveis, em vez de amostrá-las explicitamente. As topologias sobrepostas foram mostram a convergência entre as topologias recuperadas.



Figura 16. Filograma gerado a partir do algoritmo de Delimitação de espécies por estimativa ML com base no Processo de Árvore de Poisson (PTP). A máxima verossimilhança e a inferência bayesiana produziram resultados idênticos, pelo motivo mostra o resultado da inferência bayesiana. Números em cada nó são probabilidades posteriores dos táxons internos formando uma espécie. Os ramos vermelhos indicam táxons que devem ser considerados como parte de uma mesma linhagem.

5.7 Montagem, anotação e filogenia de mitogenomas.

Por meio de dados de sequenciamento RNA-Seq da cabeça e glândulas salivares os mitogenomas de *M. gajardoi, M. parapatrica* e *M. spinolai* foram obtidos. Neste estudo montou-se quatro mitogenomas por espécie, aqui apresenta-se os mitogenomas que resultaram em melhores montagens, os demais foram utilizados no estudo filogenético.

Como utilizou-se a estratégia baseado em referência, as referências *Triatoma dimidiata* (17.019 pb, NC_002609.1) e *Triatoma infestans* (17.068 pb, NC_035547.1) foram responsáveis pelas melhores montagens.

As montagens resultaram em 12 genomas com tamanho que variam entre 15 e um pouco mais que 17kb (Tabela 6). A dimensão do mtDNA deste estudo são coerentes com outros genomas de espécies de triatomíneos (DOTSON; BEARD, 2001; PITA et al., 2017; DONG et al., 2018; NAGUILERA-URIBE et al., 2020). As três espécies de *Mepraia* possuem 13 regiões codificantes de proteínas, consistente com os mtDNA de animais (BOORE, 1999) e triatomíneos (DOTSON E BEARD, 2001; PITA et al., 2017; DONG et al., 2018; ZHAO et al., 2019). Alguns genomas montados são maiores que o mtDNA de *T. infestans* (17.301bp) e *T. dimidiata* (17.019 bp), bem como, outros são menores, mas consistentes com tamanho da maioria do mtDNA de outros animais (~16kb), ocorrendo com poucas exceções (BOORE, 1999). Outra característica dos mtDNA foram A-T tendenciosos (57.41 a 66.76%) (Tabela 4).

Amostras	Tamanho	reads	mx.cov.	av.cov	GC%	A+T
M. spinolai1	17529	1564229	298069	10999.78	29.62	66,15%
M. spinolai2	15975	1096992	452552	9157.03	29.85	64,13%
M. spinolai3	16941	1716780	577372	13561.60	30.30	62,50%
M. spinolai4	17195	1821226	450774	13886.08	29.92	62,10%
M. gajardoi7	17762	858105	242684	5924.86	29.26	61.13%
M. gajardoi8	15898	940275	301135	7237.08	29.31	66.76%
M. gajardoi9	16398	1146516	264059	9346.00	29.19	62,51%
M. gajardoi10	17173	1099046	270813	7744.17	29.32	63,03%
M. parapatrica15	17492	901222	328730	6499.46	29.33	58,85%
M. parapatrica16	15813	711225	202299	5723.61	29.65	61,87%
M. parapatrica17	16270	1240510	378017	10096.90	29.15	57,41%
M. parapatrica18	17094	742086	257044	5470.32	29.01	57,41%

Tabela 4. Dados descritivos dos mitogenomas de *M. gajardoi*, *M. parapatrica e M. spinolai*.

Não foram encontrados rearranjos nem alterações nos anticódons de tRNA entre os táxons examinados, e seu arranjo gênico é semelhante aos genomas mitocondriais utilizados como referência (DOTSON e BEARD, 2001., PITA et al., 2017). A direção dos genes dentro dos genomas mitocondriais, determinada pela sua presença na cadeia leve ou pesada, foi a mesma para todos os taxa, com a maioria dos genes sendo codificados na cadeia pesada. Um

exemplo de um dos arranjos gênicos aqui reunidos para o grupo de estudo é mostrado nas figuras 17, 18 e 19.



Figura 17. Mapa do genoma mitocondrial de *M. gajardoi*. As abreviaturas para os genes são as seguintes: COI, COII e COIII referem-se às subunidades do citocromo oxidase, CytB refere-se ao citocromo B e ND1-6 refere-se aos componentes da NADH desidrogenase. As setas indicam a orientação e a ordem dos genes. Seta azul: tRNA, verde rRNA e setas vermelhas: regiões codificantes



Figura 18. Mapa do genoma mitocondrial de *M. parapatrica*. As abreviaturas para os genes são as seguintes: COI, COII e COIII referem-se às subunidades do citocromo oxidase, CytB refere-se ao citocromo B e ND1-6 refere-se aos componentes da NADH desidrogenase. As setas indicam a orientação e a ordem dos genes. Seta azul: tRNA, verde rRNA e setas vermelhas: regiões codificantes.



Figura 19. Mapa do genoma mitocondrial de *M. spinolai*. As abreviaturas para os genes são as seguintes: COI, COII e COIII referem-se às subunidades do citocromo oxidase, CytB refere-se ao citocromo B e ND1-6 refere-se aos componentes da NADH desidrogenase. As setas indicam a orientação e a ordem dos genes. Seta azul: tRNA, verde rRNA e setas vermelhas: regiões codificantes.

O mitogenoma de *M. gajardoi* possui 17.138 bp, é formado por 13 genes codificadores de proteínas,18 tRNAs, 2 rRNA e uma grande região controle (Figura 17). No processo de anotação não foram encontrados os genes: trne, trnh, trni, trnl2, trnv. *Mepraia parapatrica* possui um genoma mitocondrial de 17.081 bp constituído por 13 genes codificadores de proteínas, 16 tRNAs, 2 rRNA e região controle (Figura 18). Não foram encontrados os genes: trnt, trnv. O genoma de *M. spinolai* é composto por 13 sequências codificadoras de proteína, 22 tRNAs, 2 rRNAs região controle (Figura 19). Não foram encontrados os genes: trnf, trnp, trnq, trnt, trnv.

As estimativas de divergência genética foram obtidas a partir dos mtDNA das três espécies. Os 12 mitogenomas de *Mepraia*, mostram um percentual de 85% de regiões conservadas. As estimativas de divergência genética interespecífica mostra um percentual médio de 0.09 entre todas as regiões. As maiores divergências foram estimadas para *M. spinolai*. Quando comparado com *M. gajardoi* o valor de divergência genética foi 0.13, quando comparado a *M. parapatrica* foi de 0.11. A distância entre *M. gajardoi* e *M. parapatrica*, é relativamente menor (0.09).

O teste de taxa relativa de Tajima (TAJIMA, 1993) executado no MEGAX (KUMAR et al., 2018) com *M. parapatrica* como grupo externo, pode ser acessado na tabela 5. A igualdade da taxa evolutiva entre as sequências *M. spinolai1* e *M. gajardoi7* com a sequência *M. parapatrica15* usada como grupo externo no teste de taxa relativa de Tajima (TAJIMA, 1989). A estatística do teste $\chi 2$ foi de 47,96 (P = 0,00010 com 1 grau[s] de liberdade) P-valor menor que 0,05 é frequentemente usado para rejeitar a hipótese nula de taxas iguais entre linhagens. Essa análise envolveu três sequências de nucleotídeos. Todas as posições contendo lacunas e dados ausentes foram eliminadas (opção de exclusão completa).

Configuração	regiões		
Sítios idênticos em todas as três sequências	11482		
Locais divergentes em todas as três sequências	67		
Diferenças únicas na Sequência M. spinolai	715		
Diferenças únicas na Sequência M. gajardoi	476		
Diferenças únicas na Sequência M. parapatrica	341		

Tabela 5. Resultados do teste de Tajima de três sequências

A reconstrução filogenética utilizou 23 espécies a partir de um conjunto de dados de 15.008 bp e foram estimadas por ML e por IB. Ambas as análises tiveram como foco as relações entre as espécies de *Mepraia* e a relações com as espécies do gênero *Triatoma* com mtDNA disponíveis.

A reconstrução por ML gerou um filograma com suporte *boostrap* de ML adequado, além de ter topologia congruente às filogenias moleculares de Triatominae (Justi et al., 2016, Figura 20). O filograma enraizada com *P. rufotuberculatus* resulta em dois clados: o primeiro clado com suporte relativamente baixo agrupa triatomíneos "norte-americanos" (como são conhecidos os triatomíneos relacionados geograficamente e filogeneticamente) (Figura 20). Neste clado *T. rubida* é a espécie mais externa, *T. protacta* aparece como espécie irmã de *T. rubrofasciata*. Como em outras filogenias *T. picturata, T. phyllosoma, T. longipennis* e *T. recurva* forma o complexo *phyllosoma* + *T. dimidiata* (JUSTI et al., 2016). *Mepraia* forma um clado monofilético (suporte de 100) (Figura 19). Este clado é próximo de outras espécies sulamericanas (*T. infestans* e *T. vitticeps*, suporte 100). *M. spinolai* forma dois clados (*M. spinolai* 1+ *M. spinolai* 2 e *M. spinolai* 3 e *M. spinolai* 4, suporte de 100) (Figura 20). A amostra *M. gajardoi* a que diverge mais entre as amostras de *M. gajardoi*. *M. parapatrica* também com dois clados (*M. parapatrica* 4 e *M. parapatrica* 2 e *M. parapatrica* 3, suporte 100).

A reconstrução filogenética por meio de IB, recupera topologias consistentes com as filogenias de Triatominae, além de taxas elevadas de suporte bayesiano (Figura 21). Como na ML tem-se uma clara divisão entre triatomíneos norte americanos e sul-americanos. As relações filogenéticas descritas na literatura entre esses dois clados são bem descritas (Monteiro et al., 2019). No clado norte americano, o complexo *phylosoma* bem suportado (BPP 0.89). As espécies *T. protacta* e *T. rubida*, aparecem como espécies próximas (BPP 100). No clado sul-americano, *Mepraia* é monofilético (BPP 100) e pode-se ver a proximidades entre *M. gajardoi* e *M. parapatrica*, em contrapartida *M. spinolai* forma um clado único (BPP 100). As relações intraespecíficas são diferentes às descritas para ML, especialmente para as sequências de *M. spinolai* e *M. parapatrica* (Figura 21).

A inferência bayesiana por meio dos genes mtDNA COI e Cyt b recupera topologia similar as outras filogenias apresentadas neste estudo (Figura 22). A taxa elevada probabilidade permite inferir que *T. breyeri* e *T. eratyriformis* são ancestrais do gênero *Mepraia*, que é monofilético (BPP 100, Figura 22). As relações interespecíficas de *Mepraia* se mantém semelhantes às demais filogenias deste estudo, pois *M. gajardoi* é táxon próximo de *M. parapatrica* e ambas divergem de *M. spinolai* que forma um clado único (BPP 100). As relações

entre as espécies de *Triatoma* são claras. A filogenia enraizada com *P. rufotuberculatus*, recupera dois clados, que separam as espécies sul e norte americanas. O complexo *phyllosoma* é monofilético e insere *T. rubida* e *T. protacta*.



Figura 20. Filogenia de máxima verossimilhança (ML). Estimativa realizada com 23 sequencias com 13 genes mitocondriais de três espécies de *Mepraia*, dez de *Triatoma* e uma de *Panstrongylus*. Os valores nos nós são relativos ao *Bootstrap* de ML.



Figura 21. Filogenia estimada por inferência bayesiana Estimativa realizada com 23 sequencias com 13 genes mitocondriais de três espécies de *Mepraia*, dez de *Triatoma* e uma de *Panstrongylus*. Os valores nos nós são relativos à probabilidade bayesiana posterior. A barra lateral ilustra a graduação dos ramos a partir dos valores de suporte.



Figura 22. Filogenia estimada por inferência bayesiana. Estimativa realizada com 26 sequências de genes mitocôndrias (COI e Cyt b) de três espécies de *Mepraia*, doze de *Triatoma* e uma de *Panstrongylus*. Os valores nos nós são relativos à probabilidade bayesiana posterior. A barra lateral ilustra a graduação dos ramos a partir dos valores de suport

5.8 Teste de introgressão

O teste de ABBA-BABA para todas as combinações possíveis exibem sinais introgressões. O teste ABBA BABA baseia-se em contagens de sítios no conjunto de dados que correspondem aos padrões de genótipos ABBA e BABA, ou seja, dadas três populações de grupo interno (*M. gajardoi, M. parapatrica* e *M. spinolai*) e um grupo externo (*T. infestans*). Os resultados mostram que *M. spinolai* e *M. parapatrica* apresentam sinais de introgressão. Da mesma maneira que as espécies *M. gajardoi* e *M. spinolai* apresentam sinais de introgressão. As espécies *M. gajardoi* e *M. parapatrica* não apresentaram sinais de introgressão (Tabela. 2).

 Tabela 6. Teste ABBA-BABA entre as linhagens de Mepraia.

H1	H2	2 H3 nABBA nBA		nBABA	Dstat	jackEst	SE	Z	
MG	MP	MS	74121	53822	0.1586566	0.1586566	0.003583537	44.38481	
MG = M. gajardoi									
MP = M. parapatrica									

MS = M. spinolai

6. DISCUSSÃO

Nesse estudo foi possível obter e caracterizar 18 transcriptomas de cabeças e glândulas salivares de três espécies de *Mepraia* em condições iguais de desenvolvimento, sexo e padrões semelhantes de temperatura, estresse e inanição de alimento. Tais condições permitiram comparar o perfil transcricional das três espécies, além de permitir inferir as relações filogenéticas de *M. gajardoi, M. parapatrica* e *M. spinolai*. No estudo não se avaliou os transcritos de cada tecido (cabeça e glândulas) separadamente. A escolha foi adotada pela falta de um genoma para mapear os transcritos com a referência.

O Chile é um país que ocupa uma longa estreita faixa costeira, que igualmente a 21 países sul-americanos é afetado pela doença Chagas (TAPIA-GARAY et al., 2018; WHO, 2021). O país não possui fauna triatomínica diversa, com apenas quatro espécies: *T. infestans* e três espécies de *Mepraia* (CAMPOS et al., 2011). Entende-se que não é uma diversidade representativa, tal como a do Brasil que possui mais de 60 espécies de triatomíneos (FRÍAS, 2010; CAMPOS et al., 2011; GALVÃO., 2014). Mas em compensação ao pequeno número de
espécies, em território chileno encontra-se um dos gêneros mais singulares e curiosos da subfamília, o gênero *Mepraia* (FRÍAS et al., 2018). O gênero possui características únicas em Triatominae, além de estar relacionado a transmissão de *T. cruzi* ao homem (TAPIA-GARAY et al., 2018). À vista disso, nesta tese apresenta-se pela primeira vez o transcriptoma de cabeças e glândulas salivares de *M. gajardoi*, *M. parapatrica* e *M. spinolai* obtidos por sequenciamento RNA-seq.

No momento o número de genomas completos disponíveis de Triatominae não é tão expressivo (MESQUITA et al., 2015; PITA et al., 2017), mas em comparação os transcriptomas são relativamente mais explorados, especialmente os sialotranscriptomas (CARVALHO et al., 2017; GONÇALVES et al., 2017; SANTIAGO et al., 2018; BRITO et al., 2019). O número maior de transcriptomas estudados, possibilitou que esse estudo dispusesse de dados de qualidade para estudos de genômica comparada. Em vista disso, como é comum nos estudos genômicos todas as métricas de qualidade dos dados brutos foram verificadas e permitiram estabelecer diretrizes de tratamento dos dados brutos, como a remoção de adaptadores inseridos no sequenciamento, bem como a remoção de *reads* de baixa qualidade (CONESA et al., 2016). Não existem "*pipelines*" canônicos para a análise de transcritos, os estudos geralmente aplicam estratégias personalizadas para as necessidades que surgem devidas aos desafios comuns da área (CONESA et al., 2016). Independente do desenho experimental a obtenção de "*reads*" de qualidade e adequadamente filtradas é parte importante para análises a jusante (GARBER et al., 2011).

Neste estudo buscou-se minimizar as filtragens excessivas, pois a estratégia pode levar a perda de regiões importantes para este tipo de estudo (CONESA et al., 2016). Para isso foi estabelecido um percentual médio de 10% de *reads* filtradas por biblioteca (Tabela S3, anexo). A estratégia de filtragem foi estabelecida a partir da análise de qualidade de cada conjunto de dado, por fim, personalizando cada estratégia de filtragem, mantendo ao menos 90% dos dados brutos. A trimagem é ponto crítico e suas implicações no trato de dados de larga escala é frequentemente discutida (DEL FABBRO et al., 2013). Ao filtrar pode-se aumentar a qualidade e a confiabilidade da análise, com ganhos simultâneos em termos de tempo de execução e recursos computacionais necessários, em contrapartida pode ocorrer a perda de dados importantes para as descrições genômicas (DEL FABBRO et al., 2013). Em estudos genômicos, métrica importante é o conteúdo GC (refere-se à proporção de guanina e citosina) e por mais que seja variável entre as diferentes espécies, valores elevados podem indicar regiões do genoma ricos em genes, já que são abundantes em regiões codificadores de proteínas (FUREY; HAUSSLER, 2003). As *reads* das 18 bibliotecas obtidas neste estudo exibem valores

acima de 37% de conteúdo CG (Tabela S3, anexo). O percentual de CG em Triatominae é variável, mas não diferem muito entre si. As taxas descritas para as reads deste estudo são similares a outras descrições em Triatominae: *P. lignarius* 42.3%, *R. prolixus* 36.8%, *T. infestans* 38.7% e *T. dimidiata* 36.8% (dados *Genbank*).

As 18 bibliotecas geraram um grande número de *reads* de 100 bp, dessas *M.* gajardoi gerou um número menor e *M. spinolai* e *M. parapatrica* tiveram médias semelhantes, entretanto analisando as 18 bibliotecas independentemente os valores não diferem nas métricas gerais (Tabela S3, anexo). Após a trimagem as *reads* com qualidade *Phred* acima de 20, foram mantidas. Após a trimagem passaram a possuir menores chances de erros de sequenciamento ou contaminantes do sequenciamento, ou seja, *reads* com melhores qualidade, aptos para montagem. Um índice de qualidade *Phred* é uma medida de qualidade de identificação de nucleotídeos que oferece um índice de confiabilidade para cada nucleobase (FUREY; HAUSSLER, 2003). Quando se assume índices de qualidade acima de 20, espera-se uma acurácia de 99%. As *reads* deste estudo obtiveram índices *Phred* pós trimagem \geq 35, isto é, a probabilidade de chamada de base incorreta 1 em 1000 (99.9%). As *reads* filtradas foram montadas por estratégia *de novo*, ou seja, sem referência.

As novas tecnologias permitem uma sondagem profunda e eficiente de dados genômicos a partir de "montadores" que não dependem de um genoma de referência, como o *Trinity* (GRABHERR et al., 2011), escolhido para este estudo. De acordo com Grabherr *et al.*, (2011) o montador *Trinity* possuir uma boa capacidade de montar transcrições presentes nos dados, também relatando isoformas de *splicing* alternativo e transcrições de genes recentemente duplicados. A estratégia de montagem resultou em 18 transcriptomas que exibem índices de qualidade adequados aos sugeridos pela literatura e para este tipo de estratégia de montagem (CONESA et al., 2016). As estimativas recuperadas foram similares entre as três espécies de *Mepraia*.

Os valores N50, métrica comumente utilizada como referência para qualificar montagens foram acima de 1300 pb em todas as amostras (Tabela S4, anexo). No entanto a métrica N50 é questionável e discutida para avaliar a completude de transcriptomas (MARTIN; WANG, 2011; HÖLZER; MARZ, 2019). Aceita-se que os transcriptomas são diferentes de conjuntos de genomas e o comprimento de *contigs* são inerentes aos transcritos (GERSTEIN; SNYDER, 2009; MARTIN; WANG, 2011; CORNEY, 2013). Como comparação quantitativa as métricas N50 de montagens de transcriptomas de Triatominae disponíveis figuram entre 1.000 e 2.000 bp (ASSUMPÇÃO et al., 2011; MARCHANT et al., 2016; DE CARVALHO et al., 2017; BRITO et al., 2019).

Como a utilização da métrica de N50 para avaliar montagens de transcriptomas é questionável, neste estudo utilizou-se a ferramenta BUSCO para auxiliar na avaliação das montagens, para isso foi utilizado um banco de dados de Arthropoda (Anexo 5). O programa BUSCO é um programa baseado em referência frequentemente utilizado para avaliar a qualidade de montagens de novo (RANA et al. 2016). O programa busca alinhar todos os transcritos às proteínas altamente conservadas dentro do conjunto de dados permitindo uma avaliação quantitativa das montagens de transcritos e genomas, estimando com eficiência a completude da anotação com base em sequências evolutivamente relacionadas (SIMÃO et al., 2015). Os resultados obtidos neste estudo foram promissores, pois a análise a partir de 2510 grupos de ortólogos de artrópodes obteve-se percentuais em maioria acima de 90%, mesmo havendo índices abaixo, esses foram superiores a 80%, portanto pode-se concluir que as montagens possuem um boa completude. A média de dados fragmentados foram próximos à 4%, destoando pouco entre as montagens. Foram recuperadas regiões anotadas duplicadas, mas todas as montagens tiveram as redundâncias filtradas. Em relação ao percentual de grupos BUSCO faltantes, era esperado, pois a montagem corresponde a apenas tecidos cefálicos e das glândulas salivares de triatomíneos.

A predição de regiões codificantes (CDS) oferece a oportunidade de integrar os genes putativos recuperados de transcriptomas e atribuir suas funções (ZHUO et al., 2001). Neste estudo por meio do banco de dados *PFam* e *SwissProt* foi possível recuperar em média 21 mil CDS, com montagens exibindo valores acima de 24 mil (Anexo H). A média de CDS completos foram acima de 12 mil. A busca no banco de dados de proteínas não redundantes (nr) do *NCBI* com a ferramenta *Blastp* e nos bancos *CDD*, *Pfam*, *ProSitePatterns* acessados pelo programa *Interproscan* recuperou número de CDS semelhantes para ambas as montagens. Importante ressaltar que até o presente momento ainda não está disponível um banco de dados restrito para Triatominae nem para Reduviidae aspecto que compromete a comparação com outros dados.

A distribuição de dados de ontologia a partir da busca pelo *Blast2GO* (CONESA et al., 2005) exibe uma relação de semelhança entre os termos "*gene ontology*" (*GO terms*) entre *M. gajardoi, M. parapatrica* e *M. spinolai*. Essa metodologia oferece uma plataforma adequada para pesquisa de genômica funcional em espécies não-modelo (CONESA et al., 2005). O percentual médio de termos GO é maior em *M. gajardoi*, e relativamente equalizado entre *M. parapatrica* e *M. spinolai*, provavelmente o maior número de regiões anotadas para *M. gajardoi* provem do número maior de unigenes montados. *Gene Ontology* é um banco de dados utilizado nas análises genômicas e tem o intuito de fornecer um vocabulário controlado de termos

definidos que representam as propriedades do produto do gene. Esses termos estão separados em três domínios: Componente Celular; Função Molecular e Processo Biológico são os "unigenes" relacionados à função em determinada fase de unidades de vida: células, tecidos, órgãos e organismos (ASHBURNER et al. 2000; HARRIS et al., 2008). Nos transcriptomas deste estudo os termos biológicos foram mais abundantes para todas as amostras, seguindo da função celular e funções moleculares. Em média foram descritos 8.684 termos relacionados a função moleculares, 5.139 relacionados a processos biológicos e 2.736 componentes celulares.

Neste estudo, um dos focos foram os transcritos de tecido cefálicos devido as suas implicações biológicas, comportamentais e seu potencial farmacológicos, em consequência do tecido escolhido o número de vocabulários ligados a estímulos e sinalização foram expressivos (Figura S3, anexo). Foi possível anotar termos de proteínas de ligação de odorantes (do inglês: Odorant-binding proteins (OBPs)). Os OBPs são utilizados para se referir a uma grande família de proteínas de insetos que são excepcionais em número, abundância e diversidade (SUN; XIAO; CARLSON, 2018). Embora a função das OBPs como um todo não esteja bem estabelecida, acredita-se que elas atuem como transportadores de odor, entregando as moléculas odoríferas aos receptores olfativos na membrana celular dos neurônios sensoriais (VOGT; PRESTWICH; LERNER, 1991). O entendimento da função das OBPs especialmente de insetos vetores é de extrema importância e pode implicar em estratégias de controle ou determinação de capacidade vetorial. Também foram descritos termos ligados a sinalização de feromônios (id: GO:OOO5549), esses termos descrevem genes ligados a sinalização intraespecíficas, tanto no envio ou no recebimento do estímulo. Sugere-se explorar melhor esses dados para melhorar a compreensão em relação a diversidade de termos GO e a relação desses termos entre as espécies de Mepraia e suas implicações biológicas.

A diversidade genética é uma medida de variação genética dentro de cada espécie, ou entre populações, que permitem elucidar processos evolutivos e biológicos (WRIGHT, 2005). Neste contexto estimou-se a diversidade genética a partir dos conjuntos de SNPs obtidos do mapeamento dos transcriptomas de cabeça e glândulas salivares. *Mepraia* e suas espécies não possuem genomas disponíveis ou transcriptomas completos, por esse motivo neste estudo adotou-se a estratégia de busca de variantes por meio do genoma de *T. infestans* e utilizando as próprias montagens como referência. As estimativas foram expressivamente diferentes quando comparadas às duas estratégias de busca de variantes (*SNP calling*). Utilizando a referência de *T. infestans* [espécie filogeneticamente distante de *Meprai*a, entretanto a mais próxima disponível (JUSTI et al., 2014)] foi recuperado um número de 187.030 posições variantes de SNPs, em contrapartida utilizando as montagens deste estudo como referência o número de SNPs foi mais de 60 mil bases polimórficas entre as três espécies avaliadas.

As divergências genéticas calculadas a partir de um conjunto de SNPs mostram que entre as três espécies estudadas, *M. spinolai* é a que mais diverge, e por mais que *M. gajardoi* e *M. parapatrica* apresentem menores divergências interespecíficas, são significativas. As taxas de substituição não sinônimas são relativamente maiores em todas as amostras estudadas. Outro aspecto interessante na descrição da diversidade de SNPs foi a menor diversidade intraespecíficas de *M. gajardoi* (PI: 0,02381), ou seja, as populações estudadas são mais relacionadas entre si, isso quando comparadas as outras espécies e amostras. *Mepraia parapatrica* e *M. spinolai* possuem maior diversidade intraespecíficas, sendo: PI: 0,10284 e PI: 0,15691, respectivamente. As variações intraespecíficas de *M. spinolai* mostram divergências entre as populações silvestres (*M. spinolai*1 e *M. spinolai*2) das coletas em peridomicílio (*M. spinolai*3, *M. spinolai*4, *M. spinolai*5 e *M. spinolai*6). As redes de haplótipos construídas neste estudo facilitam e ilustram as divergências entre as espécies, além do mais, as relações intraespecíficas de *M. spinolai* para as demais espécies, além do mais, as relações intraespecíficas de *M. spinolai* para as demais espécies.

A busca por grupos de ortólogos resultou em um número elevado de genes compartilhados em sua história evolutiva (>97%).

Outra estratégia adotada para avaliar as relações filogenéticas entre as espécies de *Mepraia* foi a reconstrução de uma filogenia com um conjunto de ortólogos, que recupera uma topologia que correlaciona as espécies/genes desse gênero com as espécies de outros gêneros acessados neste estudo. A estimativa ML e IB recuperam duas árvores filogenéticas com suporte elevado, no entanto a melhor resolução foi obtida pela inferência bayesiana. Em relação à resolução, estudos mostram diferenças relevantes entre métodos não-paramétricos e probabilidades posterior bayesiana em relação ao suporte dos ramos (ERIXON et al., 2003). Entanto, a confiança estatística resultante das probabilidades posteriores na análise bayesiana pode ser superestimada em relação aos valores de *bootstra*p da ML (SUZUKI, GLAZKO, NEI et al., 2002). Porém, mesmo que empiricamente, as probabilidades posteriores bayesiana apresentam ramos com suporte maiores quando comparada a métodos não paramétricos correspondentes (KAROL et al., 2001). Estudos avaliando as duas estratégias mostram valores superiores em clados recuperados pela inferência bayesiana em relação ao ML *boostrap*, mas os autores consideram que erros na reconstrução filogenética podem ser comuns, portanto, corroborando para inferências errôneas (ERIXON et al, 2003).

A primeira espécie de *Mepraia* foi descrita e classificada ao gênero *Triatoma* (Porter, 1934). *Mepraia gajardoi* foi a segunda espécie descrita a partir de caracteres morfológicos e citogenéticos (Frias, Henry & González, 1998). Mais recentemente, *M. parapatrica* (FRÍAS, 2010) foi descrita com uma proposta interessante de especiação, pois a espécie possui distribuição fronteiriça com as duas outras espécies, essas características dentre outras foram determinantes para a sugestão de um processo de especiação parapátrica. Quando Frías (2011) descreveu *M. parapatrica* o autor sugeriu que *M. gajardoi* fosse a espécie ancestral e *M. spinolai* a evolutivamente mais recente. Entretanto os estudos recentes apontam para outro caminho (CAMPOS-SOTO et al., 2013; CAMPOS-SOTO et al., 2020).

O primeiro estudo sistemático do gênero (CALLEROS et al., 2010) avaliou várias populações ao longo do território de distribuição conhecido para *Mepraia*. Neste estudo, a filogenia reconstruída através de marcadores mitocondriais e nucleares recupera duas linhagens ["uma linhagem é composta por exemplares da Região I mais setentrional e representa *M. gajardoi*. O outro inclui amostras do Sul III, IV e das Regiões Metropolitanas, e representa *M. spinolai*" (CALLEROS et al., 2010). Para ilustrar melhor as regiões de distribuição de *Mepraia*, ver Frias et al., 2010, 2018], no entanto a população descrita no estudo como da região II não apresenta clara relação com as duas linhagens descritas, os autores explicam a divergência da população da região II destacando processos de introgressão do mtDNA ou a retenção de polimorfismos ancestrais para explicar a topologia incongruente.

A população da região II, não definida no estudo, possivelmente podiam ser espécimes de *M. parapatrica*, os resultados foram importantes para a proposta de uma origem hibrida para essa espécie. No entanto os critérios geográficos sugeridos por Frias et al., (1998) para especiação de *Mepraia* deve ser revisto (CALLEROS et al., 2010). Um estudo posterior usando dois genes mitocondriais encontrou três clados congruentes com as espécies descritas e ao contrário das hipóteses anteriores estimou que *M. spinolai* fosse a espécie ancestral do gênero (CAMPOS et al., 2013). A origem de *Mepraia* não era totalmente definida, no entanto os resultados apresentados nesta tese, fornecem novos *insights* para compreensão das relações filogenéticas.

Todas as estimativas filogenéticas em nível de espécies reconstruídas neste estudo por diferentes métodos e tiveram *T. infestans* como grupo externo. *Triatoma infestans* foi escolhida para o estudo, pela relação filogenética com as espécies de *Mepraia* e pelo genoma disponível (JUSTI et al., 2014). As filogenias foram recuperadas com conjuntos de SNPs (>60 mil bp) estimados por estratégia de quartetos, coalescente multiespécies e modelos bayesianos. Independentemente do método adotado as filogenias de espécies recuperadas tiveram suporte robusto e corroboram a recente hipótese filogenética, *M. spinolai* ancestral, *M. gajardoi* e *M. parapatrica* recentes e estreitamente relacionadas.

A delimitação de espécies inferida a partir das filogenias resultantes da ML/IB recuperada sustenta três linhagens coerentes com as espécies descritas de *Mepraia*. O processo coalescente de multiespécies é um modelo de processo estocástico que descreve as relações genealógicas de uma amostra de sequências de DNA obtidas de várias espécies (DEGNAN e ROSENBERG, 2009). Representa a aplicação da teoria coalescente ao caso de várias espécies. O coalescente multiespécies resulta em casos em que as relações entre as espécies para um gene individual (a árvore do gene) podem diferir da história mais ampla da espécie (a árvore da espécie).

Aspecto desafiador para elucidar a história evolutiva de *Mepraia* resultam das incongruências entre as filogenias disponíveis (CALLEROS et al., 2010; CAMPO et al., 2013; CAMPOS-SOTO et al., 2021). Como são comuns filogenias com genes mitocondriais (MORITZ; DOWLING; BROWN, 2014) e em consideração a relação ao custo computacional em analisar todo transcriptoma das cabeças e glândulas salivares, neste estudo montou-se, anotou-se e reconstruiu filogenia a partir de dados do genoma mitocondrial de *M. gajardoi, M. parapatrica e M. spinolai* junto com 11 espécies da subfamília com mtDNA disponível.

A estratégia adotada para montagem foi por meio de "isca ("*bait*") e mapeamento interativo a partir de dados de leituras de RNA-Seq (HANG et al., 2013). O montador MITObim (HANG et al., 2013) é capaz de reconstruir genomas mitocondriais sem a necessidade de um genoma de referência da espécie alvo, baseando-se apenas em informações do genoma mitocondrial de táxons mais distantes. Neste estudo utilizou-se quatro mtDNA (descritos nos métodos) como isca, que foram mapeados pelo programa MIRA 3.4.1.1. Essa estratégia permitiu obter 12 genomas mitocondriais sendo quatro montagens por espécies.

Os mtDNA montados são congruentes com as descrições Triatominae (BEARD, 2001; PITA et al., 2017; DONG et al., 2018; AGUILERA-URIBE et al., 2020). Todas as montagens resultaram em 13 regiões codificantes, consistentes com mtDNA de animais (BOORE, 1999). A maior parte dos mitogenomas disponíveis de Triatominae, possuem entre 15 e 17 kb de bases, as montagens deste estudo mantem comprimento similar. A variação de comprimento entre os mtDNA deve-se provavelmente pela falta de algumas regiões intergénicas (DNA espaçador) e ao tamanho da região de controle (AVISE et al., 1987; FLOOK et al., 1995).

Um problema encontrado nas montagens deste estudo é ausência de RNAs transportadores (tRNA). Os tRNAs transportadores não foram numericamente compatíveis com

as outras descrições de mtDNA, pois comumente são encontrados 22 tRNAs, com exceção de *M. spinolai* (DOTSON e BEARD, 2001; PITA et al., 2017; DONG et al., 2018; ZHAO et al., 2019). Como descrito, as montagens de *M. gajardoi* e *M. parapatrica* obtiveram 18 e 16 tRNAs não compatíveis com mtDNA de animais (BOORE, 1999). Provavelmente o número menor de tRNAs provém de problemas técnicos da montagem, da estratégia de mapeamento ou pela característica biológica desse tipo de RNA. Os tRNAs são moléculas curtas e abundantes necessárias para traduzir as informações genéticas em sequências de proteínas e podem ser reguladas diferentemente em tecidos durante o estádio de desenvolvimento (SALINAS-GIEGÉ et al., 2015). De acordo, com Behrens et al (2021) a regulação dos níveis de tRNA e seu papel fisiológico permanecem subestimados devido à falta de métodos precisos e de resolução para quantificação de tRNA.

Nas últimas décadas o mtDNAs foram utilizados em estudos filogenéticos, se mostrando úteis para delimitar as relações filogenéticas, inferir a estrutura populacional, definir fluxo gênico, na avaliação de processos de hibridação, entre outros eventos biológicos (MORITZ; DOWLING; BROWN, 2014). Na subfamília Triatominae algumas filogenias foram reconstruídas e fundamentadas em genes mitocondriais (HYPSA et al., 2002; GARDIM, 2013; JUSTI et al., 2014; BELINTANI et al., 2020). Neste estudo, avaliou-se as relações filogenéticas por meio do mtDNA totais de 14 espécies de Triatominae. A estimativa ML teve resolução diferente da inferência bayesiana, no entanto ambas são coerentes com as filogenias da subfamília, ao menos em relação aos clados (HYPSA et al., 2002; GARDIM, 2013; JUSTI et al., 2014). Como o foco são as espécies de Mepraia, foi possível mostrar as relações das três espécies desse gênero por meio de genes mitocondriais. Neste estudo o gênero foi recuperado como monofilético, tendo M. spinolai como espécie mais distante em relação ao gênero. Mepraia gajardoi e M. parapatrica como espécies mais relacionadas. As filogenias com mtDNA mostram com clareza as relações entre as espécies acessadas neste estudo. Como em outras filogenias, as espécies sul-americanas formam clado independente da norte americanas (Justi et al., 2014, Monteiro et al., 2018). Além disso, foi possível mostrar a relação entre espécies do complexo T. phylossoma (T. longipennis, T. phylossoma, T. picturata, T. recurva + T. dimidiata). A relação de T. rubida e T. protacta é exibida na filogenia reconstruída por BI. Apesar das descrições de incongruências em filogenias com mtDNA os dados obtidos recuperam filogenia congruentes com a literatura (CAMPOS et al., 2013; CAMPOS-SOTO, et al., 2021).

A história evolutiva de *Mepraia* é contada pela sua relação com as espécies endêmicas da argentina: *T. breyeri* e *T. eratyrusiformis* (FRIAS, 2010). Essas duas espécies

estão separadas de *Mepraia* pela cadeia de montanhas andinas (LENT E WYGODZINSKY, 1979). Filogenias suportam um agrupamento monofilético para essas cinco espécies (CAMPOS et al., 2013, JUSTI et al., 2014). Este grupo monofilético é conhecido como complexo "*spinolai*" (MONTEIRO et al., 2018). As espécies de *Mepraia e T. eratyrusiformis* teriam se originado de uma população ancestral que divergiu devido ao soerguimento da Cordilheira dos Andes (FRIAS, 2010; MONTEIRO et al., 2018). As relações de *Mepraia* com *T. eratyrusiformis* também é suportada pelas similaridades morfológicas e morfométricas ((MORENO; GORLA; CATALÁ, 2006). Os resultados apresentados corroboram as relações entre *T. breyeri* e *T. eratyrusiformis* e as espécies de *Mepraia*, entanto a ausência de dados para essas espécies limitam as conclusões.

Devido à natureza ligada dos genes do mtDNA, espera-se que diferentes genes apresentem resultados semelhantes. A incongruência filogenética usando genes de mtDNA pode resultar de processos como heteroplasmia, integração nuclear de genes mitocondriais, erros de polimerase, contaminação e recombinação (CAMPOS et al., 2015., BOYCE; ZWICK; AQUADRO, 1989). O teste de introgressão desse estudo com dados de RNA-Seq mostram sinais de introgressão entre *M. parapatrica* e *M. spinolai* e *M. gajardoi* e *M. spinolai*. Estes resultados são coerentes com as discussões de fluxo gênico entre as populações (FRIAS et al., 2010., CAMPOS et al., 2015). Mesmo que as distâncias geográficas entre as populações estudadas sejam relativamente grandes, sinais de introgressão nessas amostras corroboram a existência de zonas de hibridação, ou seja, a necessidade de um número amostral é essencial para testar essa hipótese. Além do mais não existe um consenso sobre eventos de hibridação.

Campos-Soto *et al.*, (2016), sugerem mecanismos de isolamento reprodutivos consolidados entre *M. gajardoi* e *M. spinolai*. Os resultados de Campos-Soto *et al.*, (2016) mostram que embora a primeira geração (F1) seja viável, os autores descrevem diferentes mecanismos de isolamento que corroboram a inviabilidade reprodutiva. Em contrapartida, os resultados apresentados aqui mostram uma estrutura populacional coerente com as linhagens propostas por CALLEROS *et al.*, (2010) e CAMPOS-SOTO et al., 2020. Em resumo *M. gajardoi* pode estar delimitado apenas na região I, *M. spinolai* provavelmente esteja amplamente distribuído por diversas regiões (Região II, III, região Metropolita e IV) algumas delas compartilhadas com as demais espécies. No entanto, a região II merece atenção (FRÍAS et al., 1998, 2010., CALLEROS et al., 2010), pois pode ser uma possível zona de hibridação entre populações *Mepraia*.

A região II é discutida, Frias *et al.*, (2010) na descrição de *M. parapatrica*, considera a região como o principal ambiente para essa espécie, bem como, as regiões insulares

próxima à costa (CAMPOS-SOTO et al., 2020). Mas é importante considerar que as descrições de introgressão mostradas neste estudo e na literatura entre *M. spinolai* e *M. parapatrica* merecem atenção. A morfologia e a morfometria agregam à validade taxonômica de *M. parapatrica* (FRIAS et al., 1998, 2018.; CAMPO-SOTO et al., 2010), entretanto à plasticidade morfológica típica de Triatominae (DUJARDIN et al., 1999) e possíveis processos de introgressão/hibridização atuando dentro de *Mepraia* (CALLEROS et al., 2010) mantém a controvérsia sobre a validade do táxon.

O abrangente estudo filogenético apresentando é coerente com os estudos anteriores e fornecem novas informações sobre as relações filogenéticas de *Mepraia*. O número amostral pequeno e limitado principalmente pelo custo do RNA-seq, somado às dificuldades para obtenção de espécimes de todas as regiões, nossos resultados não permitem a resolução da história evolutiva do gênero, mas são informações relevantes para a construção do conhecimento científico da subfamília Triatominae, em especial para os triatomíneos sulamericanos.

7. CONCLUSÕES

Este estudo caracterizou pela primeira vez o transcriptoma de cabeças e glândulas salivares de M. gajardoi, M. parapatrica e M. spinolai por meio de sequenciamento RNA-seq. Os dados obtidos foram úteis para transcriptômica comparativa, estudos populacionais e filogenéticos. As filogenias reconstruídas por diferentes abordagens corroboram a monofilia do gênero e as relações filogenéticas de Mepraia com T. breyeri e T. eratyrusiformis. Mepraia gajardoi e M. parapatrica são mais relacionadas filogeneticamente e distantes de M. spinolai. Mepraia spinolai compartilha sinais de introgressão com M. gajardoi e M. parapatrica. Os sinais de introgressão sugerem que os genes compartilhados por meio de hibridação ou retrocruzamento podem ter sido importantes para história evolutiva do grupo. Pela primeira vez caracterizou-se o sialotranscriptoma de Mepraia. Foi possível montar o genoma mitocondrial das três espécies de Mepraia, bem como reconstruir filogenias que permitiram ampliar o número de marcadores mitocondriais conhecidos para esses triatomíneos, além de colaborar para resolução da filogenia do grupo. As espécies de Mepraia estão distribuídas por ambientes silvestres e peridomiciliares no Chile. Os dados aqui apresentados são significados para ciência, pois contribuem para a compreensão da história evolutiva dos vetores da doença de Chagas, doença que afeta milhões de pessoas em todo mundo.

8. REFÊRENCIAS BIBLIOGRÁFICAS

ABAD-FRANCH, F; MONTEIRO, F.A. Molecular research, and the control of Chagas disease vectors. Anais da Academia Brasileira de Ciências v. 77, n.3, p.437-454, 2005.

AGUILERA-URIBE, M. et al. Phylogeny of the North-Central American clade of bloodsucking reduviid bugs of the tribe Triatomini (Hemiptera: Triatominae) based on the mitochondrial genome. Infection, Genetics and Evolution, v. 84, p. 104373, 1 out. 2020.

ALTENHOFF, A. M.; DESSIMOZ, C. Phylogenetic and Functional Assessment of Orthologs Inference Projects and Methods. PLOS Computational Biology, v. 5, n. 1, p. e1000262, jan. 2009.

ANDREWS, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ ARAÚJO, A. et al. Paleoparasitology of chagas disease - A review. Memorias do Instituto Oswaldo Cruz, v. 104, n. SUPPL. 1, p. 9–16, 2009.

APT W, REYES H. Algunos aspectos de la enfermedad de Chagas en Latinoamérica. Parasitología al Día. 1990; 14:23-40

ARAUJO, P. F. et al. Sexual transmission of American trypanosomiasis in humans: A new potential pandemic route for chagas parasites. Memorias do Instituto Oswaldo Cruz, v. 112, n. 6, p. 437–446, 2017.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., HARRIS, M. A. Tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics, 25(1), 25-29, 2000.

ASSUMPÇÃO, T. C. F. et al. Insight into the salivary transcriptome and proteome of *Dipetalogaster maxima*. Journal of Proteome Research, v. 10, n. 2, p. 669–679, 4 fev. 2011.

ASSUMPCAO, T.C.F; EATON, D.P; PHAM, V.M; FRANCISCHETTI, I.M.B; AOKI, V; HANS, G. An insight into the sialotranscriptome of *Triatoma matogrossensis*, a kissing bug associated with fogo selvagem in South America. American Journal Tropical Medicine Hygiene, v. 86, n. 6, p. 1005-14, Jun 2012.

AVISE, J.A., ARNOLD, J., BALL, R.M., BERMINGHAM, E., NEIGEL, J.E., REEB, C.A., SAUNDERS, N.C. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology, Evolution, and Systematics, 18: 489–522, 1997.

AVISE, J. C. Phylogeography: retrospect and prospect. Journal of Biogeography, 36(1), 3–15, 2009. https://doi.org/10.1111/J.1365-2699.2008.02032.X

BELINTANI, T. et al. Phylogenetic and phenotypic relationships of the *Triatoma sordida* subcomplex (Hemiptera: Reduviidae: Triatominae). Acta Tropica, v. 212, p. 105679, 1 dez. 2020.

BACIGALUPO A, SEGURA JA, GARCIA A, HIDALGO J, GALUPPO S, CATTAN PE. Primer hallazgo de vectores de la enfermedad de Chagas asociados a matorrales silvestres en la Region Metropolitana, Chile. Revista Médica de Chile. 2006; 134:1230-1236

BACIGALUPO A, TORRES-PEREZ F, SEGOVIA V, GARCIA A, CORREA JP, MORENO L, ET AL. Sylvatic foci of the Chagas disease vector Triatoma infestans in Chile: Description of a new focus and challenges for control programs. Memórias do Instituto Oswaldo Cruz. 2010;105(5):633-641

BEN-ARI, G.; LAVI, U. Marker-assisted selection in plant breeding. Plant Biotechnology and Agriculture, p. 163–184, 2012.

Encyclopedia of Biodiversity | ScienceDirect. Disponível em: <https://www.sciencedirect.com/referencework/9780123847201/encyclopedia-ofbiodiversity>. Acesso em: 27 mar. 2022.

BERNT, M. et al. MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics and Evolution, v. 69, n. 2, p. 313–319, nov. 2013. BOLGER, A. M., LOHSE, M., e USADEL, B. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics, 170, 2014.

BOORE, J. L. Animal mitochondrial genomes. Nucleic Acids Research, v. 27, n. 8, p. 1767–1780, 1 abr. 1999.

BOUCKAERT R, VAUGHAN TG, BARIDO-SOTTANI J, DUCHENE S, FOURMENT M, GAVRYUSHKINA A, et al. BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. bioRxiv. 2018.

BOUCKAERT RR, DRUMMOND AJ. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evolutionary Biology. 2017;17(1):42. pmid:28166715

BOYCE, T. M.; ZWICK, M. E.; AQUADRO, C. F. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics, v. 123, n. 4, p. 825–836, 1 dez. 1989.

BRITO, R. N. et al. Transcriptome-based molecular systematics: *Rhodnius montenegrensis* (Triatominae) and its position within the *Rhodnius prolixus-Rhodnius robustus* cryptic-species complex. Parasites and Vectors, v. 12, n. 1, p. 305, 17 jun. 2019.

BREHNS, A.; RODSCHINKA, G., NEDIALKOVA, D. High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. Molecular Cell., 81,8, p.1802-1815. <u>https://doi.org/10.1016/j.molcel.2021.01.028</u>

CALLEROS, L., F. PANZERA, M. D. BARGUES, F. A. MONTEIRO, D. R. KLISIOWICZ, M. A. ZURIAGA, S. MAS-COMA, AND R. PÉREZ. Systematics of *Mepraia* (Hemiptera-Reduviidae): cytogenetic and molecular variation. Infection Genetics Evolution 10: 221–228, 2010.

CAMACHO, C. et al. BLAST+: architecture and applications. BMC Bioinformatics, v. 10, n. 1, p. 421, 2009.

CAMPOS, R. et al. Phylogenetic relationships of the spinolai complex and other triatomini based on mitochondrial DNA sequences (Hemiptera: Reduviidae). Vector-Borne and Zoonotic Diseases, v. 13, n. 1, p. 73–76, 1 jan. 2013.

CAMPOS, R. et al. Wing shape differentiation of Mepraia species (Hemiptera: Reduviidae). Infection, Genetics and Evolution, v. 11, n. 2, p. 329–333, mar. 2011.

CANNON J.T., KOCOT K.M. (2016) Phylogenomics Using Transcriptome Data. In: Bourlat S. (eds) Marine Genomics. Methods in Molecular Biology, vol 1452. Humana Press, New York, NY. https://doi.org/10.1007/978-1-4939-3774-5_4

CARCAVALLO, R. U. et al. Phylogeny of the Triatominae (Hemiptera: Reduviidae). Proposals for taxonomic arrangements. Entomology and Vector, v. 7, 2000.

CARVER, T; HARRIS, S.R; BERRIMAN, M; PARKHILL, J; MCQUILLAN, J.A. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics, 15; 28(4): 464–469, 2012.

CATHALA, G. SAVOURET, J.F, MENDEZ, B, WEST, B.L, KARIN, M, MARTIAL, J.A, BAXTER, J.D. A method for isolation of intact, translationally active ribonucleic acid. DNA, v. 2, n. 4, p. 329-35, 1983.

CHAGAS, C. Nova espécie mórbida do homem, produzida por um Trypanozoma (Trypanozoma cruzi): nota prévia. Brazil-Médico, Rio de Janeiro, v.23, n.16, p.161. 1909a.

CHAGAS, C. Nova tripanozomiase humana: estudos sobre a morfologia e o ciclo evolutivo do Schizotrypanum cruzi n. gen., n. sp., ajente etiolojico de nova entidade morbida do homem. Memorias do Instituto Oswaldo Cruz, Rio de Janeiro, v.1, n.2, p.159-218. 1909b.

CHANG, L.H., CHEN, P., LIEN, M.T., HO, Y.H., LIN, C.M., PAN, Y.T., WEI, S.Y., HSU, J. Differential adhesion and actomyosin cable collaborate to drive Echinoid-mediated cell sorting. Development 138(17): 3803—3812, 2011.

CHIFMAN, J.; KUBATKO, L. Quartet inference from SNP data under the coalescent model. Bioinformatics, v. 30, n. 23, p. 3317-3324, 2014.

CHOMCZYNSKI, P.; MACKEY, K. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. Biotechniques, v. 19, n. 6, p. 942-5, Dec 1995. ISSN 0736-6205.

CHU, Y.; COREY, DR."RNA sequencing: platform selection, experimental design, and data interpretation". Nucleic Acid Therapeutics. 22 (4): 27-14, 2012 doi:10.1089/nat.2012.0367.

COLLINS, F. S.; GUYER, M. S.; CHAKRAVARTI, A. Variations on a theme: cataloging human DNA sequence variation. Science, v. 278, n. 5343, p. 1580–1581, 28 nov. 1997.

CONESA, A. et al. A survey of best practices for RNA-Seq data analysis. Genome Biology 2016 17:1, v. 17, n. 1, p. 1–19, 26 jan. 2016.

CONESA, A; GÖTZ, S; GARCÍA-GÓMEZ, J.M; TEROL, J; TALÓN, M; ROBLES. M. Blast2GO: a universal tool for annotation, visualization, and analysis in functional genomics research. Bioinformatics, v. 21, n. 18, p. 3674-6, Set, 2005.

CORNEY, D. C. RNA-Seq Using Next Generation Sequencing. Materials and Methods, v. 3, 27 ago. 2013.

COURA, J. R.; DIAS, J. C. P. Epidemiology, control and surveillance of Chagas disease: 100 years after its discovery. Memorias do Instituto Oswaldo Cruz, v. 104, n. SUPPL. 1, p. 31–40, 2009.

DALE, C.; JUSTI, S. A.; GALVÃO, C. *Belminus santosmalletae* (Hemiptera: Heteroptera: Reduviidae): New Species from Panama, with an Updated Key for *Belminus* Stål, 1859 Species. Insects 2021, Vol. 12, Page 686, v. 12, n. 8, p. 686, 30 jul. 2021.

D'ALESSANDRO, A. Biology of Trypanosoma (Herpetosoma) rangeli Tejera. 1920. Biology of Kinetoplastida. Edited by: Lumsden, W.H.R., Evans, D.A: London: Academic Press, 1: 327-493, 1976.

DE CARVALHO, D. B. et al. Differential transcriptome analysis supports *Rhodnius montenegrensis* and *Rhodnius robustus* (Hemiptera, Reduviidae, Triatominae) as distinct species. PLoS ONE, v. 12, n. 4, p. e0174997, 1 abr. 2017.

DE PAIVA, V. F. et al. A review of the taxonomy and biology of Triatominae subspecies (Hemiptera: Reduviidae). Parasitology Research, 2022, p. 1–14, 5 jan. 2022.

DE PAULA, A. S.; DIOTAIUTI, L.; SCHOFIELD, C. J. Testing the sister-group relationship of the Rhodniini and Triatomini (Insecta: Hemiptera: Reduviidae: Triatominae). Molecular Phylogenetics and Evolution, v. 35, 2005.

DEGNAN JH, ROSENBERG NA. "Gene tree discordance, phylogenetic inference and the multispecies coalescent". Trends in Ecology & Evolution. 24 (6): 332–40, 2009.

DEL FABBRO, C. et al. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. PLoS ONE, v. 8, n. 12, 23 dez. 2013.

DIAS, J. C. P.; SILVEIRA, A. C.; SCHOFIELD, C. J. The impact of Chagas disease control in Latin America - a review. Memorias do Instituto Oswaldo Cruz, v. 97, 2002.

DIAS, J. P. et al. Acute Chagas disease outbreak associated with oral transmission. Revista da Sociedade Brasileira de Medicina Tropical, v. 41, n. 3, p. 296–300, 2008.

DONG, L., MA, X., WANG, M., ZHU, D., FENG, Y., ZHANG, Y., WANG, J. Complete mitochondrial genome of the Chagas disease vector, *Triatoma rubrofasciata*. Korean Jornal Parasitology., 56 pp. 515-519, 2018. 10.3347/kjp.2018.56.5.515.

DORN, P.L, JUSTI, S.J; DALE, C; STEVENS, L; GALVÃO, C; LIMA-CORDÓN, R; MONROY, R. Description of Triatoma mopan sp. n. from a cave in Belize (Hemiptera, Reduviidae, Triatominae). ZooKeys. 775: 69-95. 2018.

DOTSON, E. M.; BEARD, C. B. Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. Insect Molecular Biology, v. 10, n. 3, p. 205–215, 2001.

DURAND, E. Y. et al. Testing for ancient admixture between closely related populations. Molecular Biology and Evolution, v. 28, n. 8, p. 2239-2252, 2011

EGGER B, LAPRAZ F, TOMICZEK B et al. A transcriptomic-phylogenomic analysis of the evolutionary relationships of fl atworms. Current Biology 25:1347–1353, 2015. doi: 10.1016/j. cub.2015.03.034

EISEN JA, FRASER CM (2003) Phylogenomics: intersection of evolution and genomics. Science 300:1706–1707. doi: 10.1126/ science.1086292

EMMS D.M. & KELLY S. STRIDE: Species Tree Root Inference from Gene Duplication Events, Molecular Biology and Evolution 34(12): 3267-3278, 2017.

EMMS, D.M., KELLY, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20:238, 2019.

ERIXON P, SVENNBLAD B, BRITTON T, OXELMAN B. "Reliability of Bayesian posterior probabilities and bootstrap frequencyes in phylogenetics". Systematic Biology. 52 (5): 665–73, 2003.

EWING B, GREEN P. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Research 8(3):186-94. 1998.

EXCOFFIER, L. G. LAVAL, AND S. SCHNEIDER. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47-50. 2005.

FEDER, J. L., HOOD, G. R., DOELLMAN, M. M., SCHULER, H., MILLER, L. A., TAIT, C., GLOVER, M. M., MEYERS, P. J. Speciation. Process of. Reference Module in Life Sciences, 2021. https://doi.org/10.1016/B978-0-12-822562-2.00025-6
FITCH, W. M. Distinguishing homologous from analogous proteins. Systematic Zoology, v. 19, n. 2, p. 99–113, 1970.

FRÍAS D, ATRIA J. Chromosomal variation, macroevolution and possible parapatric speciation in Mepraia spinolai (Porter) (Hemiptera:Reduviidae). Genetics, Molecular Biology 21: 179-184, 1998.

FRÍAS, D. A new species and karyotype variation in the bordering distribution of Mepraia spinolai (Porter) and Mepraia gajardoi Frías et al. (Hemiptera: Reduviidae: Triatominae) in Chile and its parapatric model of speciation. Neotropical Entomology, 39(4): 572-583, 2010.

FRÍAS, D. A., A. HENRY eC. GONZÁLEZ1998. *Mepraia gajardoi*: A new species of Triatominae (Hemiptera: Reduviidae) from Chile and its comparison with *Mepraia spinolai*. Revista Chilena de História Natural, 71: 177-188,1998.

FRÍAS, D.A., OLIVERIA. J., ROSA, J.A. Morphological description of Mepraia spp. females (Hemiptera: Reduviidae, Triatominae). Acta Tropica 190: 389–394, 2019.

FRÍAS-LASSERRE, D. A new species and karyotype variation in the bordering distribution of Mepraia spinolai (Porter) and *Mepraia gajardoi* Frías et al (Hemiptera: Reduviidae: Triatominae) in Chile and its parapatric model of speciation. Neotropical Entomology, v. 39, n. 4, p. 572–583, 2010.

FRÍAS-LASSERRE, D. et al. The Interspecific Hemelytra Differences Among *Mepraia* Species Males (Hemiptera, Reduviidae:Triatominae) in Chile. Journal of Medical Entomology, v. 55, n. 6, p. 1478–1485, 25 out. 2018.

FUJII, S., AMREIN, H. (2002). Genes expressed in the Drosophila head reveal a role for fat cells in sex-specific physiology. Embo Journal, 21(20):5353-5363.2002.

FUREY, T. S.; HAUSSLER, D. Integration of the cytogenetic map with the draft human genome sequence. Human Molecular Genetics, v. 12, n. 9, p. 1037–1044, 1 maio 2003.

GAJARDO-TOBAR, R. Algo más sobre *Mepraia spinolai* (Porter). Hemiptera, Triatomidae. Revista Chilena de Entomologia 3:117-125, 1953. GALVÃO, C., org. Vetores da doença de chagas no Brasil [online]. Curitiba: Sociedade Brasileira de Zoologia, 2014, 289 p. Zoologia: guias e manuais de identificação series. ISBN 978-85-98203-09-6. Available from SciELO Books.

GALVÃO, C. et al. A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. Zootaxa, v. 202, 2003.

GALVÃO, C. et al. A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. Zootaxa, v. 202, 2003.

GALVÃO, C; CARCAVALLO, R; ROCHA, D.S; JURBERG, J. Checklist of the current valid species of the subfamily Triatominae Jeannel, (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. Zootaxa, 202, p. 1-36, 2003. GARBER, M. et al. Computational methods for transcriptome annotation and quantification using RNA-seq. Nature Methods 2011 8:6, v. 8, n. 6, p. 469–477, 27 maio 2011.

GARDIM, C. R. C. A. D. T. M. S. D. A. R. C. J. R. Evolutionary relationships of the Triatoma matogrossensis subcomplex, the endemic Triatoma in central-western Brazil, based on mitochondrial DNA sequences. American Journal of Tropical Medicine, v. 89, n. 4, p. 766–774, out. 2013.

GARDIM, S., ALMEIDA, C. E., TAKIYA, D. M., OLIVEIRA, J., ARAÚJO, R. F., CICARELLI, R. M. B., & DA ROSA, J. A. Multiple mitochondrial genes of some sylvatic Brazilian Triatoma: Non-monophyly of the *T. brasiliensis* subcomplex and the need for a generic revision in the Triatomini. Infection, Genetics and Evolution, 23, 74–79, 2014. https://doi.org/10.1016/j.meegid.2014.01.024

GONÇALVES, L. O. et al. Insights from tissue-specific transcriptome sequencing analysis of *Triatoma Infestans*. Memorias do Instituto Oswaldo Cruz, v. 112, n. 6, p. 456–457, 2017. GRABHERR, M. G. et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nature biotechnology, v. 29, n. 7, p. 644, jul. 2011. GUARNER, J. Chagas disease as example of a reemerging parasiteSeminars in Diagnostic PathologyW.B. Saunders, 1 maio 2019. Acesso em: 19 jan. 2021

HAAG, J.; O'HUIGIN, C.; OVERATH, P. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. Molecular and Biochemical Parasitology, v. 91, n. 1, 1 mar. 1998.

HAAG, J.; O'HUIGIN, C.; OVERATH, P. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. Molecular and Biochemical Parasitology, v. 91, n. 1, 1 mar. 1998.

HAHN, C.; BACHMANN, L.; CHEVREUX, B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research, v. 41, n. 13, p. e129–e129, 1 jul. 2013.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In Nucleic acids symposium series, 41, pp. 95-98. London: Information Retrieval Ltd., c1979-c2000, 1999.

HARRIS, M. A. et al. The Gene Ontology project in 2008. Nucleic acids research, v. 36, n. Database issue, 1 jan. 2008.

HELED, J., & DRUMMOND, A. J. (2010). Bayesian Inference of Species Trees from Multilocus Data. Molecular Biology and Evolution, 27(3), 570–580. https://doi.org/10.1093/molbev/msp274

HELLMUTH, M. et al. Phylogenomics with paralogs. Proceedings of the National Academy of Sciences of the United States of America, v. 112, n. 7, p. 2058–2063, 17 fev. 2015.

HÖLZER, M.; MARZ, M. De novo transcriptome assembly: A comprehensive cross-species comparison of short-read RNA-Seq assemblers. GigaScience, v. 8, n. 5, p. 1–16, 1 maio 2019. HU, T.; BANZHAF, W. Nonsynonymous to Synonymous Substitution Ratio: Measurement for Rate of Evolution in Evolutionary Computation. Lecture Notes in Computer Science (including

subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), v. 5199 LNCS, p. 448–457, 2008.

HWANG, W. S.; WEIRAUCH, C. Evolutionary History of Assassin Bugs (Insecta: Hemiptera: Reduviidae): Insights from Divergence Dating and Ancestral State Reconstruction. PLoS ONE, v. 7, n. 9, p. e45523, 28 set. 2012a.

HYPSA, V. et al. Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a New World origin of the Asiatic clade. Molecular Phylogenetics and Evolution, v. 23, 2002.

INSECTA - Genome - NCBI. Disponível em: <https://www.ncbi.nlm.nih.gov/genome/?term=INSECTA>. Acesso em: 6 abr. 2021. JAKSIK, R. et al. Microarray experiments and factors which affect their reliability. Biology Direct, v. 10, n. 1, 3 set. 2015.

JENSEN, R. A. Orthologs and paralogs - We need to get it right (multiple letters). Genome Biology BioMed Central, 3 ago. 2001. Disponível em: <http://genomebiology.biomedcentral.com/articles/10.1186/gb-2001-2-8-interactions1002>. JEFFARES DC, POOLE AM, PENNY D. elics from the RNA world". Journal of Molecular Evolution. 46 (1): 18, 1999.

JUSTI, S. A. et al. Molecular phylogeny of Triatomini (Hemiptera: Reduviidae: Triatominae). Parasites and Vectors, v. 7, n. 1, p. 149, 31 mar. 2014.

KAPLI, P., YANG, Z., & TELFORD, M. J. Phylogenetic tree building in the genomic age. Nature Reviews Genetics 2020 21:7, 21(7), 428–444, 2020.

KAROL, K. G., R. M. MCCOURT, M. T. CIMINO, AND C. F. DELWICHE. The closest living relatives of land plants. Science. 294:2351–2353, 2001.

KATOH, K., ROZEWICKI, J., & YAMADA, K. D. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in bioinformatics, 20(4), 1160-1166, 2019.

KATOH, K.; STANDLEY, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology and Evolution, v. 30, n. 4, p. 772-780, 2013.

KIERAN, T. J. et al. Ultraconserved elements reconstruct the evolution of Chagas diseasevectoring kissing bugs (Reduviidae: Triatominae). Systematic Entomology, v. 46, n. 3, p. 725– 740, 1 jul. 2021.

KIMURA M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120, 1980.

KOBOLDT, D.C; ZHANG, Q; LARSON, D.E; SHEN, D; MICHAEL, D; MCLELLAN, LIN, L; MILLER, C.A; MARDIS, E.R; DING, L; WILSON, R.K. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Research, v. 22, n. 3, p. 568-76, Mar, 2012.

KOLESNIKOV, A. A., & GERASIMOV, E. S. (2013). Diversity of mitochondrial genome organization. Biochemistry (Moscow) 2012 77:13, 77(13), 1424–1435, 2013. https://doi.org/10.1134/S0006297912130020

KORNELIUSSEN, T. S.; ALBRECHTSEN, A.; NIELSEN, R. ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics, v. 15, n. 1, p. 356, 2014.

KOSAKOVSKY POND, S. L.; FROST, S. D. W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Molecular Biology and Evolution, v. 22, n. 5, p. 1208–1222, 1 maio 2005.

KOSAKOVSKY POND, S. L.; FROST, S. D. W.; MUSE, S. V. HyPhy: hypothesis testing using phylogenies. Bioinformatics (Oxford, England), v. 21, n. 5, p. 676–679, 1 mar. 2005. KROPF, S.P., SÁ, M.R., The discovery of Trypanosoma cruzi and Chagas disease (19081909): tropical medicine in Brazil. História, Ciências, Saúde-Manguinhos. Suppl 1:13- 34,;16, Jul, 2009.

LAMBARRI, M. C., CIFUENTES, A. C., AYALA, S., TAPIA-GARAY, V., & LILLO, D. C. Eco-Epidemiology of Chagas Disease in Chile. Chagas Disease - Basic Investigations and Challenges, 2018 https://doi.org/10.5772/INTECHOPEN.74744

LASSERRE, D. F. et al. Morphological description of *Mepraia* spp. females (Hemiptera: Reduviidae, Triatominae). Acta Tropica, v. 190, p. 389–394, 1 fev. 2019.

LEACHÉ, A. D.; OAKS, J. R. The Utility of Single Nucleotide Polymorphism (SNP) Data in Phylogenetics. https://doi.org/10.1146/annurev-ecolsys-110316-022645, v. 48, p. 69–84, 6 nov. 2017.

LEIGH JW, BRYANT D. popart: full-feature software for haplotype network construction. Methods in Ecology and Evolution. 2015;6(9):1110–6.

LENT, H., JURBERG, J., GALVÃO, C. Revalidação do gênero *Mepraia* Mazza, Gajardo e Jorg, 1940 (Hemiptera, Reduviidae, Triatominae). Memorias Instituto Oswaldo Cruz 89, 347–352, 1994.

LENT, H.; JURBERG, J.; GALVÃO, C. Revalidação do gênero Mepraia, Mazza, Gajardo & Jorg, 1940 (Hemiptera, Reduviidae, Triatominae). Memorias do Instituto Oswaldo Cruz, v. 89, n. 3, p. 347–352, set. 1994.

LENT H, WYGODZISNKY P. Revision of the Triatominae (Hemiptera, Reduviidae) and their significance as vectors of Chagas' disease. Bulletin of the American Museum of Natural History, 1979; 163:123-520

LI, W.H.; WU, C.I; LUO, C.C. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Molecular Biology and Evolution. v. 2, n. 2, p. 150-74, Mar, 1985.

LIBRADO P, ROZAS J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 2009;25. pmid:19346325

LIU, J. et al. Natural selection of protein structural and functional properties: A single nucleotide polymorphism perspective. Genome Biology, v. 9, n. 4, p. 1–17, 8 abr. 2008.

LOCKHART, D. J. et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nature Biotechnology, v. 14, n. 13, p. 1675–1680, 1996.

MADDISON, D.R. MADDISON. Mesquite: A Modular System for Evolutionary Analysis. Version 3.51. http://www.mesquiteproject.org, 2018.

MARCHANT, A. et al. Under-Expression of Chemosensory Genes in Domiciliary Bugs of the Chagas Disease Vector Triatoma brasiliensis. PLoS Neglected Tropical Diseases, v. 10, n. 10, 28 out. 2016.

MARCHAT, A; MOUGEL, F; MENDONÇA, V; QUARTIER, M; JACQUIN-JOLY, E; MARCILLA, A; BARGUES, M.D; RAMSEY, J.M; MAGALLONGASTELUM, SALAZAR-SCHETTINO, P.M; ABAD-FRANCH, F. The ITS-2 of the Nuclear rDNA as a Molecular Marker for Populations, Species, and Phylogenetic Relationships in Triatominae (Hemiptera: Reduviidae), Vectors Chagas Diseases. Molecular Phylogenetics Evolution. v.18, p. 136-142, 2002.

MARTH, G. T. et al. A general approach to single-nucleotide polymorphism discovery. Nature Genetics 1999 23:4, v. 23, n. 4, p. 452–456, dez. 1999.

MARTIN, J. A.; WANG, Z. Next-generation transcriptome assembly. Nature Reviews Genetics, v. 12, n. 10, p. 671–682, out. 2011.

MARTINS-MELO, F. R., RAMOS JR, A. N., ALENCAR, C. H., e HEUKELBACH, J. Trends and spatial patterns of mortality related to neglected tropical diseases in Brazil. Parasite epidemiology and control, 1(2), 56-65, 2016.

MAZZA, S., TOBAR, R.G., JÖRG, M.E. Mepraia novum genus de Triatomidae. Mepraia spinolai (Porter) 1933, nov. comb., redescipción de ♂ y descripción de ♀. Investigaciones sobre Triatomidae. Universidad de Buenos Aires. Misión de Estudios de Patología Regional Argentina (Jujuy) 44, 1–30. 1940.

MCDONALD, J. H. KREITMAN. "Adaptive protein evolution at the Adh locus in Drosophila" . Nature. 351 (6328): 652 654, 1991.

MESQUITA, R. D. et al. Genome of Rhodnius prolixus, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. Proceedings of the National Academy of Sciences of the United States of America, v. 112, n. 48, p. 14936–14941, 1 dez. 2015.

METROPOLIS, N., ROSENBLUTH, A.W., ROSENBLUTH, M.N., TELLER, A.H., TELLER, E. Equations of state calculations by fast computing machines. Journal of Chemical Physics. 21, 1087–1091, 1953 <u>https://doi.org/10.1063/1.1699114</u>

METZER, M.L. Sequencing technologies - the next generation. Nature Reviews Genetics., 11. pp. 31-46, 2010.

MIKKELSEN, T. S. et al. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature, v. 437, n. 7055, p. 69–87, 1 set. 2005.

MISOF, B. et al. Phylogenomics resolves the timing and pattern of insect evolution. Science, v. 346, n. 6210, p. 763–767, 7 nov. 2014.

MIRARAB, S. et al. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics, v. 30, n. 17, p. i541–i548, 1 set. 2014.

MONGA I, BANERJEE I. Computational Identification of piRNAs Using Features Based on RNA Sequence, Structure, Thermodynamic and Physicochemical Properties". Current Genomics. 20 (7): 508–518.20 (7): 508-518, 2019.

MONTEIRO, F. A. et al. Evolution, Systematics, and Biogeography of the Triatominae, Vectors of Chagas Disease. In: Advances in Parasitology. [s.l.] Academic Press, 2018. v. 99p. 265–344.

MORENO ML, GORLA D., CATALÁ S. Association between antennal phenotype, wing polymorphism and sex in the genus Mepraia (Reduviidae, Triatominae). Infection, Genetics Evolution 6:228–234, 2006.

MORENO, M. L.; GORLA, D.; CATALÁ, S. Association between antennal phenotype, wing polymorphism and sex in the genus Mepraia (Reduviidae: Triatominae). Infection, Genetics and Evolution, v. 6, n. 3, p. 228–234, 1 maio 2006.

MORITZ, C.; DOWLING, T. E.; BROWN, W. M. Grinnell Resurvey View project The Oz Mammals Genomics Initiative View project. 2014.

MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol., 61(3), 539-542.

MURILLO-GODÍNEZ, G. Chagas disease (American trypanosomiasis). Medicina Interna de Mexico, v. 34, n. 6, p. 959–970, 10 ago. 2018.

MURRELL, B. et al. Gene-Wide Identification of Episodic Selection. Molecular Biology and Evolution, v. 32, n. 5, p. 1365–1371, 1 maio 2015.

NEI M. AND GOJOBORI T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3:418-426, 1986. NEIVA, A; LENT, H. Estudos sobre Triatomideos do Chile: Interessante caso de provável polimorfismo. Memorias Instituto Oswaldo Cruz. 35, 1940.

NGUYEN L-T, SCHMIDT HA, VON HAESELER A, MINH BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution. 2014;32(1):268–74. pmid:25371430

NOTTER, D. R.; SCHERF, B.; HOFFMANN, I. Breeding of Animals. Encyclopedia of Biodiversity: Second Edition, p. 636–649, 1 jan. 2013.

OLIVEIRA, J, AYALA J. M, JUSTI S.A, DE ROSA JA, GALVÃO, C. Description of a new species of *Nesotriatoma* Usinger, 1944 from Cuba and revalidation of synonymy between

Nesotriatoma bruneri (Usinger, 1944) and N. flavida (Neiva, 1911)(Hemiptera, Reduviidae, Triatominae). Journal of vector ecology, v. 43, n. 1, p. 148-157, 2018.

PABINGER S, DANDER A, FISCHER M, et al. A survey of tools for variant analysis of nextgeneration genome sequencing data. Briefings in Bioinformatics. 2014 Mar; 15(2): 256–278. doi: 10.1093/bib/bbs086

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. Molecular Phylogenetics and Evolution, v. 56, n. 2, p. 608–621, 1 ago. 2010.

PATTERSON, N. J. et al. ancient admixture in human history. Genetics, 2012.

PITA, S. et al. Complete mitochondrial genome of Triatoma infestans (Hemiptera, Reduviidae, Triatominae), main vector of Chagas disease. Infection, Genetics and Evolution, v. 54, p. 158–163, 1 out. 2017.

POINAR, G. A primitive bug, Paleotriatoma metaxytaxa gen. et sp. nov. in med-Cretaceous amber from northern Myanmar. Cretaceous Research, 93: 90-97, 2019.

PORTER, C.E. Una Triatoma nueva chilena. Revista Chilena de História Natural, 37:192-193,1933, 1934.

PRICE, M.N., DEHAL, P.S., ARKIN, A.P. FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution, 2009 26:1641-1650, doi:10.1093/molbev/msp077.

PRJIBELSKI, A. et al. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics, v. 70, n. 1, p. e102, 1 jun. 2020.

RAJKUMAR, A.P; QVIST, P; LÁZARO, R; LESCAI, F; JU, J; NYEGAARD, M; MORS, O; BORGLUM, A.D; LI, Q; CHRISTENSEN, J.H. Experimental validation of methods for

differential gene expression analysis and sample pooling in RNA-Seq. BMC Genomics, 16:548, 2015.

RANA, S. B. et al. Comparison of De Novo Transcriptome Assemblers and k-mer Strategies Using the Killifish, Fundulus heteroclitus. PLOS ONE, v. 11, n. 4, p. e0153104, 1 abr. 2016. RAMBAUT A, DRUMMOND AJ, XIE D, BAELE G, SUCHARD MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic biology.

RAMSAY, G. DNA chips: State-of-the art. Nature Biotechnology, v. 16, n. 1, p. 40–44, 1998. 2018;67(5):901–4. pmid:29718447

Reduviidae-Genome-NCBI.Disponívelem:<https://www.ncbi.nlm.nih.gov/genome/?term=Reduviidae>.Acesso em: 18 abr. 2022.

Reduviidae-Genome-NCBI.Disponívelem:<https://www.ncbi.nlm.nih.gov/genome/?term=Reduviidae>.Acesso em: 18 abr. 2022.

REUTER, J. A.; SPACEK, D. V.; SNYDER, M. P. High-Throughput Sequencing Technologies. Molecular Cell, v. 58, n. 4, p. 586–597, 21 maio 2015.

ROCHA E SILVA, E.O. DA, SOUZA, J. M. P. DE, ANDRADE, J. C. R. DE, MELLO, C. DA S. AND FERREIRA, O. Preferência alimentar (entre sangue humano e ave) dos *Triatoma sordida* encontrados em casas habitadas da regiao norte do estado de São Paulo-Brasil. Revista de Saúde Pública, São Paulo, 11, 258-269, 1977.

ROZAS, J., FERRER-MATA, A., SÁNCHEZ-DELBARRIO, J. C., GUIRAO-RICO, S., LIBRADO, P., RAMOS-ONSINS, S. E., SÁNCHEZ-GRACIA, A. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution., 34(12), 3299-3302, 2017.

RIVES-BLANCHARD, N. et al. Trypanosoma cruzi over the ocean: Insular zones of Chile with presence of infected vector Mepraia species. Acta Tropica, v. 172, p. 229–231, 1 ago. 2017.

SAAVEDRA, M., BACIGALUPO, A., BARRERA, M. V., VERGARA, M. J., ÁLVAREZ-DUHART, B., MUÑOZ-SAN MARTÍN, C., SOLÍS, R., & CATTAN, P. E. Trypanosoma cruzi infection in the wild Chagas disease vector, Mepraia spinolai: Parasitic load, discrete typing units, and blood meal sources. Acta Tropica, 229, 10636, 2022. https://doi.org/10.1016/J.ACTATROPICA.2022.106365

SAGUA H, ARAYA ROJAS J, GONZÁLEZ CORTES J, NEIRA CORTES I. *Mepraia spinolai* na costa sudeste do Oceano Pacífico (Chile)—primeiro registro insular e padrão de alimentação na ilha Pan de Azúcar. Memorias do Instituto Oswaldo Cruz; 95: 167-170, 2000. 10.1590/s0074-0276200000200006

SALINAS-GIEGÉ, T.; GIEGÉ, R.; GIEGÉ, P. Trna Biology in Mitochondria. Internacional Journal of Molecular Science. 6(3): 4518–4559, 2015.

SANTIAGO, P. B. et al. Exploring the molecular complexity of *Triatoma dimidiata* sialome. Journal of proteomics, v. 174, p. 47, 1 mar. 2018.

SCHENONE H, ROJAS A. Algunos datos y observaciones pragmáticas en relación a la epidemiologia de la enfermedad de Chagas. Boletín Chileno de Parasitología. 1989; 44:66-86

SCHOFIELD, C. J.; GALVÃO, C. Classification, evolution, and species groups within the Triatominae. Acta Tropica, v. 110, n. 2–3, p. 88–100, 1 maio 2009.

SHENDURE, J. et al. DNA sequencing at 40: past, present and future. Nature, 550:7676, v. 550, n. 7676, p. 345–353, 11 out. 2017.

SILVA, F. S., CRUZ, A. C. R., DE ALMEIDA MEDEIROS, D. B., DA SILVA, S. P., NUNES, M. R. T., MARTINS, L. C., CHIANG, J. O., DA SILVA LEMOS, P., CUNHA, G. M., DE ARAUJO, R. F., DE OLIVEIRA MONTEIRO, H. A., & NUNES NETO, J. P. Mitochondrial genome sequencing and phylogeny of *Haemagogus albomaculatus*, *Haemagogus leucocelaenus*, *Haemagogus spegazzinii*, and *Haemagogus tropicalis* (Diptera: Culicidae). Scientific Reports 2020 10:1, 10(1), 1–13. <u>https://doi.org/10.1038/s41598-020-73790-x</u>

SIMÃO, F. A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, v. 31, n. 19, p. 3210–3212, 1 out. 2015.

SORAGGI, S.; WIUF, C.; ALBRECHTSEN, A. Improved D-statistic for low-coverage data. bioRxiv, 2017.

SMITH, M. D. et al. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Molecular biology and evolution, v. 32, n. 5, p. 1342–1353, 1 maio 2015.

SMITH-CALDAS, M. R. B. et al. Phylogenetic relationships among species of the fraterculus group (Anastrepha: Diptera: Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase I. Neotropical Entomology, v. 30, n. 4, p. 565-573, 2001.

STAMATAKIS, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, v. 30, n. 9, p. 1312-1313, 2014.

SUN, J. S.; XIAO, S.; CARLSON, J. R. The diverse small proteins called odorant-binding proteins. Royal Society Open Biology, v. 8, n. 12, 19 dez. 2018.

SUZUKI Y, GLAZKO GV, NEI M. "Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics". Proceedings of the National Academy of Sciences of the United States of America. 99 (25): 16138–43, 2002.

TAJIMA F. Simple methods for testing molecular clock hypothesis. Genetics 135:599-607, 1993.

TAJIMA F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123:585-595, 1989.

TAPIA-GARAY, V. et al. Assessing the risk zones of Chagas' disease in Chile, in a world marked by global climatic change. Memorias do Instituto Oswaldo Cruz, v. 113, n. 1, p. 24, 1 jan. 2018.

TATUSOV, R.; KOONIN, E.; LIPMAN, D. A genomic perspective on protein families. Science, v. 278, 1997.

TELFORD MJ, BUDD GE, PHILIPPE H (2015) Phylogenomic insights into animal evolution. Current Biology, 25:R876–R887. doi: 10.1016/j. cub.2015.07.060

TENG, L. et al. Identification of genes under positive selection reveals differences in evolutionary adaptation between brown-algal species. Frontiers in Plant Science, v. 8, p. 1429, 15 ago. 2017.

TOBIAS, N. J., & LATORRE-ESTIVALIS, J. M. Perspectives in Triatomine Biology Studies: "OMICS"-Based Approaches. Entomology in Focus, *5*, 557–592, 2021. https://doi.org/10.1007/978-3-030-64548-9_20

TIMBERLAKE, W. E. Heterosis. Brenner's Encyclopedia of Genetics: Second Edition, p. 451–453, 27 fev. 2013.

USINGER, R.L. The Triatominae of North and Central America and the West Indies and their public health significance. Washington., Public Health Bull, 288, 81 p. 1944.

VAN DONGEN, S. Graph clustering by flow simulation. 2000. (Ph.D thesis). University of Utrecht, Utrecht, Netherlands.

VENTE, J.C.; ADAMS, E.W.; MYERS., E.W.; LI, P.W.; MURAL, R.K.; SUTTON, G.G.; SMITH, H.O.; YANDELL, M.; EVANS, M.D.; HOLT, R.A. The sequence of the human genome Science, 291 (2001), pp. 1304-1351

VOGT, R. G.; PRESTWICH, G. D.; LERNER, M. R. Odorant-binding-protein subfamilies associate with distinct classes of olfactory receptor neurons in insects. Journal of Neurobiology, v. 22, n. 1, p. 74–84, 1 jan. 1991.

WANG, D. G. et al. Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. Science, v. 280, n. 5366, p. ,1077-1082, 15 maio 1998.

WANG, D. G. et al. Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. Science, v. 280, n. 5366, p. ,1077-1082, 15 maio 1998.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: A revolutionary tool for transcriptomics. Nature Reviews Genetics, v. 10, n. 1, p. 57–63, jan. 2009.

WANG, Z; GERSTEIN, M; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics, v. 10, n. 1, p. 57-63, 2009.

WEIRAUCH, C. Cladistic analysis of Reduviidae (Heteroptera: Cimicomorpha) based on morphological characters. Systematic Entomology, v. 33, n. 2, p. 229–274, 1 abr. 2008.

WEIRAUCH, C.; MUNRO, J. B. Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. Molecular Phylogenetics and Evolution, v. 53, n. 1, p. 287–299, 1 out. 2009.

WEIRAUCH, C.; MUNRO, J. B. Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. Molecular Phylogenetics and Evolution, v. 53, n. 1, p. 287–299, 1 out. 2009.

WHO | Neglected tropical diseases - Summary. WHO, 2016.

WHO, World Health Organization Chagas disease (American trypanosomiasis). Disponível em: http://www.who.int/mediacentre/factsheets/fs340/en/. Junho de 2017. Acesso em: 02 março de 2019.

WHO. Chagas disease (also known as American trypanosomiasis). Disponível em: https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis)). Acesso em: 23 jun. 2021.

WICK, R.R Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biology. 20(129), 2019. WIPfIER, B., BAI, M., SCHOVILLE, S., DALLAI, R., UCHIFUNE, T., MACHIDA, R. ET AL. Ice crawlers (Grylloblattodea) – thehistory of the investigation of a highly unusual group of insects. J Insect Biodiversity 2: 1–2 GIRIBET G (2015) New animal phylogeny: future challenges for animal phylogeny in the age of phylogenomics. Organisms Diversity & Evolution, 2015:1–8, 2014. doi: 10.1007/s13127-015-0236-4

WORLD HEALTH ORGANIZATION. Chagas disease (also known as American trypanosomiasis). Disponível em: https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis). Acesso em: 19 jan. 2021.

WRIGHT, A. F. Genetic Variation: Polymorphisms and Mutations. eLS, 23 set. 2005.

YOUNG, A. D.; GILLUNG, J. P. Phylogenomics — principles, opportunities and pitfalls of big-data phylogenetics. Systematic Entomology, v. 45, n. 2, p. 225–247, 1 abr. 2020.

YUVARAJ, J. K. et al. Antennal Transcriptome Analysis of the Chemosensory Gene Families From Trichoptera and Basal Lepidoptera. Frontiers in Physiology, v. 9, p. 1365, 27 set. 2018.

ZHANG, J. et al. Evolution of the assassin's arms: Insights from a phylogeny of combined transcriptomic and ribosomal DNA data (Heteroptera: Reduvioidea). Scientific Reports, v. 6, n. 1, p. 1–8, 26 fev. 2016.

ZHANG, J., KAPLI, P., PAVLIDIS, P., STAMATAKIS, A. A general species delimitation method with applications to phylogenetic placements. Bioinformatics. Oxford, England 29, 2869–2876, 2019. <u>https://doi.org/10.1093/bioinformatics/btt499</u>

ZHANG, L.; LI, W. H. Human SNP reveal no evidence of frequent positive selection. Molecular Biology and Evolution v. 22, n. 12, p. 2504–2507, dez. 2005.

ZHANG, Z. et al. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. Genomics, Proteomics & Bioinformatics, v. 4, n. 4, p. 259-263, 2006.

ZHAO, Y., JIANG, M., WU, Y., SONG, F., CAI, W., LI, H. Mitochondrial genomes of three kissing bugs (Reduviidae: Triatominae) and their phylogenetic implications. International Jornal Biological and Macromolecular, 134, pp. 26-42, 2019. 10.1016/j.ijbiomac.2019.05.020

ZHAO, Y.; GALVÃO, C.; CAI, W. *Rhodnius micki*, a new species of Triatominae (Hemiptera, Reduviidae) from Bolivia. ZooKeys, v. 1012, p. 71–93, 26 jan. 2021.

9. ANEXOS

Anexo A. Declaração referente a biotica e biossegurança.



COORDENADORIA DE PÓS-GRADUAÇÃO INSTITUTO DE BIOLOGIA Universidade Estadual de Campinas Caixa Postal 6109. 13083-970, Campinas, SP, Brasil Fone (19) 3521-6378. email: cpgib@unicamp.br



DECLARAÇÃO

Em observância ao §5º do Artigo 1º da Informação CCPG-UNICAMP/001/15, referente a Bioética e Biossegurança, declaro que o conteúdo de minha Tese de Doutorado, intitulada "*Estudo de transcriptomas por RNA-Seq em tecidos de cabeça e glândula salivar de três espécies de Mepraia (Hemiptera, Reduviidae, Triatominae).*", desenvolvida no Programa de Pós-Graduação em Biologia Animal do Instituto de Biologia da Unicamp, não versa sobre pesquisa envolvendo seres humanos, animais ou temas afetos a Biossegurança.

Assinatura:
Nome do (a) aluno(a): Tiago Belintani
Assinatura: Mitutat
199
Nome do(a) aluno(a): João Aristeu da Rosa
Data: 18 de junho de 2022

Anexo B. Declaração de direito autorais.

Declaração

As copias de artigos de minha autoria ou de minha coautoria, já publicados ou submetido para publicação em revistas cientificas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada Estudo de transcriptomas por RNA-Seq em tecidos de cabeça e glândula salivar de três espécies de *Mepraia* (Hemiptera, Reduviidae, Triatominae), não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 21 de julho de 2022

Assinatura: Nome do(a) autor(a): Tiago Belintani RG n.º 42.292.987.6

Assinatura:

Nome do(a) orientador(a): João Aristeu da Rosa RG n.º 4.473.881
Amostra	Concentração	Relação 260/280	Volume (ul)
	(ng/ul)		
M. spinolai1	219,2	1,82	10
M. spinolai2	112	2.27	10
M. spinolai3	496	1,92	10
M. spinolai4	412,8	2.08	10
M. spinolai5	232	1.83	10
M. spinolai6	880	1.90	10
M. gajardoi7	225,6	1.85	10
M. gajardoi8	123,2	1.64	10
M. gajardoi9	644,8	1.97	10
M. gajardoi10	316,8	1.78	10
M. gajardoi11	166,4	1.79	10
M. gajardoi12	1080	1.85	10
M. parapatrica13	419,2	1.80	10
M. parapatrica14	573,6	1.94	10
M. parapatrica15	572,6	1,97	10
M. parapatrica16	147,2	1.78	10
M. parapatrica17	222,4	1.82	10
M. parapatrica18	257,6	1.96	10

Anexo C. Resultados das quantificações de RNA total de cabeça e glândulas salivares de três espécies do gênero *Mepraia*.



Anexo D. Imagem do gel de agarose desnaturante à 0,8% referente as amostras referenciadas na tabela 4:

Amostras	RIN	Concentração RNA (ng/µl)	rRNA Ratio [28s / 18s]
M. spinolai1	7.10	44	0.1
M. spinolai2	6.80	48	0.0
M. spinolai3	7.50	43	0.0
M. spinolai4	6.50	52	0.0
M. spinolai5	8.10	45	0.0
M. spinolai6	8.50	138	0.0
M. gajardoi7	6.10	40	0.0
M. gajardoi8	7.20	46	0.0
M. gajardoi9	8.40	82	0.0
M. gajardoi10	7.50	18	0.0
M. gajardoi11	3.30	45	0.0
M. gajardoi12	6.50	106	0.0
M. parapatrica13	7.20	52	0.0
M. parapatrica14	8.60	152	0.1
M. parapatrica15	8.70	85	0.1
M. parapatrica16	6.60	38	0.1
M. parapatrica17	7.20	123	0.0
M. parapatrica18	7.10	93	0.0

Anexo E. Parâmetros analisados pelo *Bioanalyzer* nas amostras de *RNA* obtidos de cabeça e glândulas salivares de três espécies do gênero *Mepraia*.



Anexo F. Eletroforese capilar das 18 amostras realizada no *Bioanalyzer* (*AgilentTecnologies*®). (L – marcador de peso molecular, Amostras: 1 a 6 *M. spinolai*; 7 a 12 *M. gajardoi* e 13 a 18 *M. parapatrica*)

Amostras	Total reads	bp reads pós		bp final	%CG	Scores
			trimagem (%)			Phred
M. spinolai1	16.531.051	35-101	14.881.161 (90,02%)	50-100	38	>35
M. spinolai2	16.096.887	35-101	14.383.720 (89,36%)	50-100	39	>35
M. spinolai3	14.734.129	35-101	13.349.078 (90.6%)	50-100	38	>35
M. spinolai4	14.257.893	35-101	12.507.904 (87,73%)	50-100	38	>35
M. spinolai5	16.766.548	35-101	14.729.877 (87,85%)	50-100	37	>35
M. spinolai6	15.028.922	35-101	13.491.301 (89,77%)	50-100	42	>35
M. gajardoi7	14.611.211	35-101	13.478.433 (92.52%)	50-100	37	>35
M. gajardoi8	15.679.924	35-101	14.182.485 (80.45%)	50-100	39	>35
M. gajardoi9	15.426.308	35-101	13.792.068 (89.41%)	50-100	38	>35
M. gajardoi10	16.136.186	35-101	14.719.238 (91.22%)	50-100	38	>35
M. gajardoi11	15.207.913	35-101	13.770.948 (90.55%)	50-100	40	>35
M. gajardoi12	14.032.117	35-101	12.051.072 (85.88%)	50-100	39	>35
M. parapatrica13	13.760.003	35-101	12.419.145 (90.26%)	50-100	38	>35
M. parapatrica14	13.728.931	35-101	12.204.684 (88.90%)	50-100	42	>35
M. parapatrica15	15.261.069	35-101	13.600.328 (89.12%)	50-100	39	>35
M. parapatrica16	14.144213	35-101	13.044.393 (92.22%)	50-100	40	>35
M. parapatrica17	13.877.408	35-101	11.951.720 (86.12%)	50-100	39	>35
M. parapatrica18	12.485.668	35-101	11.286.380 (90.39%)	50-100	39	>35
Total	267.766.381		239.843.935			

Anexo G. Resumo estatístico das bibliotecas antes e pós filtragem por qualidade das três espécies de *Mepraia*.

Montagem	Montagem	Número	unigenes	N50	Contigs	Contigs	Contigs	Contigs	Mediana	Média	CG%
	total	de			maior que	maior	maior que				
		contings			1000	que 2000	10000				
M. spinolai1	66706337	77620	58879	1583	19485	7943	101	45837	420	859.4	36.37
M. spinolai2	63677017	78868	60510	1451	18429	7059	79	28033	403	807.39	36.49
M. spinolai3	52344476	66369	51657	1365	15463	5497	51	25747	416	788.69	36.67
M. spinolai4	39216469	53456	43432	1189	11100	3508	48	28927	401	733.62	36.82
M. spinolai5	35526329	48220	39396	1196	10421	3059	51	26507	409	736.76	36.28
M. spinolai6	43559280	57418	45674	1274	12503	4210	60	22303	408	758.63	37.05
M. gajardoi7	28286716	41659	34714	1047	8210	2112	18	22093	399	679.01	36.59
M. gajardoi8	63418101	74777	56383	1529	18412	7186	125	36976	422	848.1	36.85
M. gajardoi9	77385695	92618	68421	1507	22121	8671	126	40804	417	835.54	36.67
M. gajardoi10	69051473	80453	59866	1579	20206	8154	134	30503	416	858.28	36.71
M. gajardoi11	49377596	62882	48782	1331	14298	4856	89	27068	413	785.24	36.72
M. gajardoi12	65380145	77747	59166	1518	18961	7485	115	40618	418	840.93	36.63
M. parapatrica13	56180669	65171	50165	1535	16619	6468	87	36978	440	862.05	36.60
M. parapatrica14	45697924	57175	44615	1360	13437	4848	46	34807	422	799.26	37.16
M. parapatrica15	70906869	82619	61478	1581	20548	8447	118	25096	419	858.24	36.79
M. parapatrica16	52205189	64931	49975	1393	14645	5299	120	29518	410	804.01	37.91
M. parapatrica17	52051783	65675	49982	1367	15276	5262	86	40605	411	792.57	36.97
M. parapatrica18	53737452	64914	49637	1459	15331	5737	141	34927	415	827.83	37.40

Anexo H. Estatísticas das montagens de 18 transcriptomas de novo, das três espécies de Mepraia.

Anexo I. Avaliação de completude BUSCO para controle de montagem de transcriptoma do tecido de gandulas salivares de cabeça de três espécies do gênero *Mepraia*

Amostras	Completos	Copias únicas	Cópias duplicadas	BUSCOs	Não encontrados	Pesquisa total de
	BUSCOs	BUSCOs	BUSCOs	fragmentados	BUSCOs	grupos BUSCOs
M. spinolai1	2226 (90.3%)	1448 (57,7%)	818 (32.6%)	67 (2.7%)	177 (7%)	2510
M. spinolai2	2229 (88.8)	1441 (58.6%)	758 (758%)	90 (3.6%)	191 (7.6%)	2510
M. spinolai3	2159 (86%)	1489 (59.3%)	670 (26.7%)	110 (4.4%)	241 (9.6%)	2510
M. spinolai4	1926 (76.8%)	1460 (58.2%)	466 (18.6%)	174 (6.9%)	410 (16.3%)	2510
M. spinolai5	1986 (79.1%)	1447 (57.6%)	539 (21.5%)	145 (8%)	379 (15.1%)	2510
M. spinolai6	2094 (83.4%)	1513 (60.3%)	581(23.1%)	125 (5%)	291 (11.6%)	2510
M. gajardoi7	1800(71.7%)	1353(53.9%)	447(17.8%)	179(7.1%)	531(21.2%)	2510
M. gajardoi8	2262(90.1%)	1443(57.5%)	819(32.6%)	75(3%)	152(6%)	2510
M. gajardoi9	2298(91.6%)	1392(55.5%)	906 (36.1%)	60(2.4%)	152(6%)	2510
M. gajardoi10	2313 (92.2%)	1395(55.6%)	918(36.6%)	58(2.3%)	139(5.5%)	2510
M. gajardoi11	2174(86.6%)	1504(26.7%)	670(26.7%)	99(3.9%)	237(9.5%)	2510
M. gajardoi12	2269(90.4%)	1456(58.0%)	813(32.4%)	71(2.8%)	170(6.8%)	2510
M. parapatrica13	2214(88.2%)	1506(60%)	708(28.2%)	79(3.1%)	217(8.7%)	2510
M. parapatrica14	2146(89.3%)	1489(26.2%)	657(26.2%)	108(4.3%)	256(10.2%)	2510
M. parapatrica15	2307(91.9%)	1431(34.9%)	876(34.9%)	138(2.6%)	138(5.5%)	2510
M. parapatrica16	2188(87.1%)	1542(61.4%)	646(25.7%)	110(4.4%)	212(8.5%)	2510
M. parapatrica17	2192(87.3%)	1457(58%)	735(29.3%)	118(4.7%)	200(8%)	2510
M. parapatrica18	2180(86.8%)	1494(27.3%)	686(27.3%)	105(4.2%)	225(9%)	2510

Amostras	CDS	CDS completos	nr (Genbank)	Interproscan	Blast2GO
M. spinolai1	24.103	15.823	16.973	16.010	11.648
M. spinolai2	23.862	14.837	15.661	14.587	11.105
M. spinolai3	21.278	12.356	12.976	12.188	9.746
M. spinolai4	17.904	8.579	8.901	8.579	8.043
M. spinolai5	17.881	8.824	9.107	8.824	8.060
M. spinolai6	19684	10.161	10.580	10.033	8.925
M. gajardoi7	15.774	7.328	7.058	7.273	6.921
M. gajardoi8	23.657	14.977	16.081	14.977	11.209
M. gajardoi9	27.527	18.570	19.461	18.570	12.936
M. gajardoi10	25.759	17.009	8.457	17.009	12.310
M. gajardoi11	20.838	11.764	12.475	11.764	12.309
M. gajardoi12	24.175	15.272	16.296	15.275	11.398
M. parapatrica13	21.506	13.400	14.076	13.215	10.034
M. parapatrica14	19.857	10.900	11.156	10.900	9.215
M. parapatrica15	25.338	16.675	17.847	16.368	11.764
M. parapatrica16	21.646	11.997	12.646	11.800	10.111
M. parapatrica17	22.080	12.583	13.217	12.401	10.166
M. parapatrica18	21.743	12.626	13.413	12.626	10.164

Anexo J. Anotação de unigenes do transcriptoma de ninfas masculinas de 5º estádio de M. gajardoi, M. paraparica e M. spinolai.

* nr: banco de dados de proteínas não redundantes GenBank (nr), incluindo apenas proteínas de artrópodes.

				Termos GO		
Amostras	Genes	Genes anotados	Biológicos	Celulares	Funcionais	Totais
M. spinolai1	11.648	11.648	6.164	3.022	9.854	19.040
M. spinolai2	11.105	11.105	5.559	2.946	9.293	17.798
M. spinolai3	9.746	9.746	5.297	2.627	8.183	16.107
M. spinolai4	8.043	8.043	4.347	2.155	6.888	13.190
M. spinolai5	8.060	8.060	4.321	2.148	6.679	13.148
M. spinolai6	8.925	8.925	4.343	2.447	7.583	14.373
M. gajardoi7	6.921	6.921	3.437	1.848	5.767	11.052
M. gajardoi8	11.209	11.209	5.562	2.971	9.451	17.984
M. gajardoi9	12.936	12.936	6.363	3.408	10.922	20.693
M. gajardoi10	12.310	12.310	6.153	3.242	10.451	19.846
M. gajardoi11	12.309	12.309	6.153	3.242	10.450	19.845
M. gajardoi12	11.398	11.398	5.590	2.993	9.564	18.237
M. parapatrica13	10.034	10.034	4.911	2.683	8.515	16.109
M. parapatrica14	9.215	9.215	4.485	2.480	7.746	14.711
M. parapatrica15	11.764	11.764	5.822	3.211	10.023	19.056
M. parapatrica16	10.111	10.111	5.009	2.687	8.628	16.234
M. parapatrica17	10.166	10.166	5.015	2.726	8.814	16.355
M. parapatrica18	10.164	10.164	4.990	2.695	8.664	16.349

Anexo L. Sumario da busca de ontologia genica pelo Blast2GO compiladas pela plataforma WEGO (*Web Gene Ontology Annotation Plotting*) de três espécies de *Mepraia*



Anexo M. Termos GOs encontrados nos CDS das seis réplicas de M. gajardoi, M. parapatrica e M. spinolai.

Species	Número de genes	Número de genes em ortogrupos	Número de genes não atribuídos	Porcentagem de genes em ortogrupos	Porcentagem de genes não atribuídos	Ortogrupos contendo espécies	Porcentagem de ortogrupos cespécies	Número de ortogrupos específicos da espécie	Número de ortogrupos específicos de espécies	Porcentagem ortogrupos específicos de espécies
M. spinolai1	24343	24106	237	99.0	1.0	11221	57.6	24	52	0.2
M. spinolai2	23862	23647	215	99.1	0.9	11129	57.2	19	42	0.2
M. spinolai3	21278	21109	169	99.2	0.8	10569	54.3	18	21	0.2
M. spinolai4	18053	17942	111	99.4	0.6	9662	49.6	10	19	0.1
M. spinolai5	17974	17862	112	99.4	0.6	9670	49.7	8	19	0.1
M. spinolai6	19516	19395	121	99.4	0.6	10167	52.2	8	500	0.1
M. gajardoi 7	26139	25328	811	96.9	3.1	11463	58.9	70	263	1.0
M. gajardoi 8	21049	20859	190	99.1	0.9	10638	54.6	16	35	0.2
M. gajardoi 9	24175	23973	202	99.2	0.8	11290	58.0	14	35	0.1
M. gajardoi 10	15709	15587	122	99.2	0.8	8928	45.9	6	16	0.1
M. gajardoi 11	23955	23760	195	99.2	0.8	11243	57.7	12	32	0.1
M. gajardoi 12	27527	27262	265	99.0	1.0	11642	59.8	26	66	0.2
M. parapatrica13	21506	21331	175	99.2	0.8	10654	54.7	15	34	0.2
M. parapatrica15	20067	19916	151	99.2	0.8	10271	52.7	10	22	0.1
M. parapatrica16	24940	24672	268	98.9	1.1	11304	58.1	28	68	0.3
M. parapatrica17	22080	21886	194	99.1	0.9	10767	55.3	16	46	0.2
M. parapatrica18	22026	21838	188	99.1	0.9	10666	54.8	11	28	0.1
P. megistus	18442	17707	735	96.0	4.0	9420	48.4	190	493	2.7
P. tertius	12086	7209	4877	59.6	40.4	4173	21.4	171	657	4.1
R. prolixus	16522	20439	1659	92.5	7.5	10365	53.2	253	377	3.0
R. robustus	13438	20111	611	97.1	2.9	10660	54.7	127	308	1.8
T. pallidipennes	23296	22921	375	98.4	1.6	9842	50.5	122	367	1.6
T. dimidiata	20.722	15930	592	96.4	3.6	8610	44.2	126	114	1.9
T. infestans	13.438	12838	600	95.5	4.5	7913	40.6	49	263	0.8

Anexo N. Estatísticas da busca de ortólogos no Orthofinder por espécies de Triatoma