

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

Thayany Ferreira da Costa

Estudo do genoma mitocondrial de *Rhodnius prolixus* Stal, 1859

Rio de Janeiro

2019

Thayany Ferreira da Costa

Estudo do genoma mitocondrial de *Rhodnius prolixus* Stal, 1859

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Bioquímica, Instituto de Química, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Bioquímica.

Orientador: Rafael Dias Mesquita

Universidade Federal do Rio de Janeiro

2019

Aqui fica a ficha catalográfica

Thayany Ferreira da Costa

Estudo do genoma mitocondrial de *Rhdnius prolixus* Stal, 1859

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Bioquímica, Instituto de Química, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Bioquímica.

Aprovado em:

Dr. Rafael Dias Mesquita (orientador)

Dra. Bianca (Membro Interno)

Dr. Thiago Estevam Parente Martins (Membro Externo)

Dr. André (Suplente Externo)

Dedico este trabalho aos meus pais Zilda Ferreira e Walter Costa, as pessoas mais importantes da minha vida, quem amo incondicionalmente.

AGRADECIMENTOS

A conclusão desta dissertação me trás a memória os desafios que passei até chegar a este momento. Em cada um destes momentos tive ao meu lado pessoas que contribuíram para que a minha caminhada fosse mais leve e prazerosa. Logo, venho por meio deste texto, expressar parte da minha gratidão.

Agradeço ao Professor Rafael Mesquita pela orientação desde minha iniciação científica, e pela paciência e atenção durante todos esses anos. À professora Glória, por ser tão amável comigo, sempre disposta a ajudar.

A Ellen Dotson por ter sequenciado o genoma mitocondrial da colônia NIH e ao Professor Pedro Lagerblad por ter fornecido o transcriptoma da colônia UFRJ, fundamentais para este projeto.

Aos meus amigos do LBVD/BIOINFO, conheci pessoas incríveis ao longo desses anos, como Francisquinha, Priscila e Stephanie. Tenho muito a agradecer o Eloy, que sempre teve a paciência e disposição para me ajudar, seja para entender um script ou um desenho. Mas em especial agradeço a Bia Maria que sempre esteve ao meu lado, das noites que passamos juntas comentando quilos mortais até o auxílio para elaborar um projeto de doutorado.

Não posso deixar de agradecer aos membros da minha família que sempre me deram o suporte que eu precisava. As minhas tias que sempre me trataram como uma filha. Aos meus filhos caninos, que tornaram a minha vida mais leve e me ensinaram o amor incondicional. A minha irmã Simone e sobrinha Victória, que sempre me apoiaram e comemoraram minhas conquistas mesmo não entendendo o que faço. Aos meus amados pais, pelo amor incondicional, e por sempre acreditarem e me fazer acreditar que sou capaz de conquistar tudo o que desejar. Agradeço também a Camila, que sempre esteve ao meu lado me incentivando a ser uma pessoa melhor, que ouviu minhas angústias e me proporcionou muitas alegrias.

Epigrafe

RESUMO

A doença de Chagas é uma enfermidade causada pelo protozoário flagelado *Trypanosoma cruzi*. Insetos hemípteros da subfamília Triatominae, popularmente conhecidos como barbeiros, são os grandes responsáveis pela transmissão do parasita à humanos e outros animais. Dentre esses organismos, *Rhodnius prolixus* (Stal, 1859) se destaca como um dos vetores mais importantes de Chagas na América Latina e Central. O objetivo principal objetivo deste estudo foi obter mais conhecimento acerca desse inseto, através da montagem e análise do seu genoma mitocondrial. O genoma mitocondrial de *R. prolixus* colônia UFRJ foi completamente montado usando bibliotecas de cDNA do inseto e o mitogenoma de *R. prolixus* colônia NIH como molde. Para a montagem foram usados três abordagens diferentes, que envolveram o uso dos programas Trinity, SPAdes, Blast e Bowtie2, dentre outros. O mitogenoma de *R. prolixus* apresenta 37 genes codificantes para 13 proteínas, 2 rRNA e 22 para tRNA's. Com a montagem do mitogenoma UFRJ foi realizado uma árvore filogenética com as proteínas mitocondriais concatenadas de *Rhodnius prolixus* junto com outras espécies de Reduviídeos já sequenciados. Também foi feita a análise de expressão dos genes mitocondriais, usando o programa Salmon e Expander. A infecção pelo patógeno e a hematofagia parecem afetar a expressão dos genes mitocondriais. Como genomas mitocondriais são amplamente utilizados em estudos de taxonomia e biogeografia, e *R. prolixus* e *R. robustus* são espécies filogeneticamente muito próximas foi realizada uma comparação entre a sequência mitocondrial de *Rhodnius prolixus* UFRJ e sequências disponíveis na literaturade de *R. prolixus* e *R. robustus*. Além disso, foi feito a busca por polimorfismos presentes no genoma mitocondrial da colônia UFRJ, que pudessem ser úteis nos estudos de genética populacional. A busca por variantes identificou 25 SNP's e 2 Indels com potencial para serem usados como marcadores moleculares. Os resultados obtidos neste estudo produziram informações importantes acerca da biologia desse vetor, assim como para a filogenia do grupo que esse organismo está inserido.

ABSTRACT

Chagas disease is caused by *Trypanosoma cruzi*, a protozoan parasite transmitted to the mammalian hosts through the feces of blood-sucking triatomine bugs. *Rhodnius prolixus* is one of the most epidemiologically relevant vectors of Chagas disease in Latin and Central America. The aim of this study was to obtain more knowledge about this insect through the assembly of its mitochondrial genome. The UFRJ mitogenome was completely assembled using insect cDNA libraries and a NIH mitogenome to guide the assembly. For the assembly three different approaches were used, which involved the programs Trinity, SPAdes, Blast and Bowtie2, among others. The *R. prolixus* mitogenome has 37 genes coding for 13 proteins, 2 rRNAs and 22 for tRNAs. With the assembly of the UFRJ mitogenome, a phylogenetic tree was made with the concatenated mitochondrial proteins of *Rhodnius prolixus* along with other sequenced Reduviid species. Mitochondrial gene expression analysis was also performed using the Salmon and Expander program. Pathogen infection and haematophagy appear to affect mitochondrial gene expression. As mitochondrial genomes are widely used in taxonomy and biogeography studies, and *R. prolixus* and *R. robustus* are very closely related phylogenetically species, a comparison was made between the mitochondrial sequence of *Rhodnius prolixus* UFRJ and sequences available in the *R. prolixus* and *R. robustus*. In addition, a search was made for polymorphisms present in the mitochondrial genome of the UFRJ colony that could be useful in population genetic studies. The search for variants identified 25 SNPs and 2 Indels with potential to be used as molecular markers. The results obtained in this study produced important information about the biology of this vector, as well as for the phylogeny of the group that this organism is inserted.

LISTA DE FIGURAS

Figura 1: Esquema dos dois ciclos de transmissão do *T. cruzi* existentes.

Figura 2: Ciclo de vida do *Trypanosoma cruzi*.

Figura 3: Global da Doença de Chagas, casos relatados a WHO no período de 2006 a 2010.

Figura 4: Modos de infecção da Doença de Chagas no Brasil.

Figura 5: Distribuição dos triatomíneos de maior importância epidemiológica ao longo do território brasileiro, ano de notificação 2012 e 2016.

Figura 6: Distribuição aproximada das espécies da Subfamília Triatominae de maior relevância epidemiológica.

Figura 7: Distribuição geográfica de *R. prolixus* na América Central e do Sul.

Figura 8: Proteínas e RNAs presentes nas mitocôndrias, o esquema aponta a cooperação que existe entre a mitocôndria e o núcleo.

Figura 9: Esquema da replicação assimétrica no mitogenoma de *Drosophila*.

Figura 10: Visão geral da transcrição mitocondrial humano, processamento pós-transcricional e tradução de RNA.

Figura 11: Diagrama do genoma mitocondrial de insetos ancestrais.

Figura 12: Fluxograma usado para a montagem do genoma mitocondrial da colônia UFRJ.

Figura 13: Fluxograma usado para a visualização do genoma mitocondrial da colônia UFRJ.

Figura 14: Exemplo de controle de qualidade antes e após limpeza dos reads utilizados.

Figura 15: *Contigs* usados para a montagem do mitogenoma UFRJ.

Figura 16: Exemplo da visualização do alinhamento dos reads limpos do transcriptoma com o mitogenoma-UFRJ montado.

Figura 17: Cobertura do genoma mitocondrial colônia UFRJ.

Figura 18: Esquema circular do genoma mitocondrial de *Rhodnius prolixus*.

Figura 19: Mapa dos valores de RSCU para *R. prolixus*, *Triatoma infestans*, *T. dimidiata* e *T. rubrofasciata*.

Figura 20: Estrutura secundária predita dos RNAs transportadores do mitogenoma de *Rhodnius prolixus*, colônia UFRJ.

Figura 21: Esquema da região controle do mitogenoma de *R. prolixus*.

Figura 22: Alinhamento da unidade de repetição de 172 pb da região controle dos mitogenomas de *R. prolixus*.

Figura 23: Alinhamento da unidade de repetição de 85 pb da região controle dos mitogenomas de *R. prolixus*.

Figura 24: Alinhamento do trecho de 447 pb rico em G+C da região controle do mitogenoma de triatomíneos.

Figura 25: Alinhamento do trecho de 274 pb rico em A+T da região controle do mitogenoma de triatomíneos.

Figura 26: Alinhamento do trecho de 42 pb do trecho final da região controle do mitogenoma de triatomíneos.

Figura 27: Estrutura em formato de grampo no final da região controle do mitogenoma de *R. prolixus*.

Figura 28: Estrutura em formato de grampo formado no trecho de 274 pb na região controle do mitogenoma de *R. prolixus*, colônia UFRJ.

Figura 29: Árvore filogenética construída pelo método de máxima verossimilhança a partir das sequências mitocondriais concatenadas dos 13 genes codificadores de proteínas.

Figura 30: *Heatmap* da expressão dos CDS (codificantes de proteínas) mitocondriais de *Rhodnius prolixus*, colônia UFRJ.

Figura 31: Alinhamento múltiplo da região 610 a 690 das sequências de CYTB de *Rhodnius prolixus*.

Figura 32: Alinhamento múltiplo da região 610 a 690 das sequências de CYTB de

Rhodnius robustus.

Figura 33: Alinhamento múltiplo da região 670 a 900 das sequências de rRNAI de *Rhodnius prolixus*.

Figura 34: Alinhamento múltiplo da região 670 a 900 das sequências de rRNAI de *Rhodnius robustus*.

Figura 35: Árvore filogenética do gene mitocondrial citocromo b.

Figura 36: Histograma.

Figura 37: Porcentagem de transições e transversões que ocorreram nos SNPs do mitogenoma UFRJ.

Figura 38: Esquema das possíveis origens de replicação na região controle dos triatomíneos.

Figura 39: SNP presente no tRNAL2 do mitogenoma *Rhodnius prolixus*, colônia UFRJ.

LISTA DE TABELAS

Tabela 1: Comparação entre os genomas mitocondriais de triatomíneos totalmente sequenciados até o momento.

Tabela 2: Bibliotecas do transcriptoma de *Rhodnius prolixus*.

Tabela 3: Espécies da família Reduviidae que já possuem o mitogenoma disponível, usados na análise de filogenia..

Tabela 4: Lista das sequências de citocromo b (CYTB) de *Rhodnius prolixus* usadas neste estudo.

Tabela 5: Lista das sequências de citocromo b de *Rhodnius robustus* usadas neste estudo.

Tabela 6: Lista das sequências do rRNAI de *Rhodnius prolixus* usadas neste estudo.

Tabela 7: Lista das sequências do rRNAI de *Rhodnius robustus* usadas neste estudo.

Tabela 8: Dados das bibliotecas após a etapa de limpeza feita com o programa Cutadapt.

Tabela 9: Quantidade de *reads* após alinhamento com o mitogenoma montado.

Tabela 10: Anotação do genoma mitocondrial UFRJ de *Rhodnius prolixus*.

Tabela 11: Composição dos códons dos genes codificadores de proteínas em diferentes posições para os mitogenomas das duas linhagens de *R. prolixus*.

Tabela 12: Identidade das sequências de citocromo b (CYTB) de *R. prolixus* disponíveis na literatura comparados à sequência da colônia UFRJ.

Tabela 13: Identidade das sequências de citocromo b (CYTB) de *R. robustus* disponíveis na literatura comparados à sequência da colônia UFRJ.

Tabela 14: Identidade das sequências da subunidade maior do RNA ribossomal (rRNAI) de *R. prolixus* disponíveis na literatura comparados à sequência da colônia

UFRJ.

Tabela 15: Identidade das sequências da subunidade maior do RNA ribossomal (rRNAI) de *R. robustus* disponíveis na literatura comparados à sequência da colônia UFRJ.

Tabela 16: Número de variantes encontrados no mitogenoma-UFRJ, em cada biblioteca do transcriptoma.

Tabela 17: SNPs identificados em todas as amostras do mitogenoma-UFRJ.

Tabela 18: Substituições sinônimas ou não sinônimas nos genes codificadores de proteínas.

Tabela 19: Indels identificados em todas as amostras do mitogenoma-UFRJ.

SUMÁRIO

1. INTRODUÇÃO	15
1.1 Doença Tropicais Negligenciadas	Erro! Indicador não definido.
1.2 Doença de Chagas	176
1.3 Triatomíneos	2223
1.4 <i>Rhodnius prolixus</i>	2425
1.5 Genoma mitocondrial	287
2. OBJETIVO	35
2.1 Objetivo geral	35
2.2 Objetivos específicos	35
3. METODOLOGIA	36
3.1 Montagem do genoma mitocondrial de <i>Rhodnius prolixus</i> , colônia UFRJ	3637
3.2 Anotação e análise dos genomas mitocondriais	4243
3.3 Busca de mitogenomas, alinhamento múltiplo e construção da árvore filogenética	4344
3.4 Nível de expressão dos transcritos mitocondriais	4546
3.5 Comparação do mitogenoma de <i>R. prolixus</i> , colônia UFRJ, com sequências da literatura para as espécies <i>R. prolixus</i> e <i>R. robustus</i> .	45
3.6 Descoberta e caracterização de polimorfismos de nucleotídeo único, inserções e deleções	51
4. RESULTADOS	53
4.1 Montagem do genoma mitocondrial de <i>Rhodnius prolixus</i> , colônia UFRJ	53
4.2 Anotação e análise dos genomas mitocondriais	61
4.2.1 Genes codificadores de proteínas	64
4.2.2 RNAs não codificantes	66
4.2.3 Regiões não codificantes	6862
4.2.4 Origem de replicação	6962
4.3 Busca de mitogenomas, alinhamento múltiplo e construção da árvore filogenética	65
4.4 Nível de expressão dos transcritos mitocondriais	75
4.5 Comparação do mitogenoma de <i>R. prolixus</i> , colônia UFRJ, com sequências da literatura para as espécies <i>R. prolixus</i> e <i>R. robustus</i> .	76
4.6 Descoberta e caracterização de polimorfismos de nucleotídeo único, inserções e deleções	8676

5. DISCUSSÃO	9293
5.1 Montagem do genoma mitocondrial de <i>Rhodnius prolixus</i> , colônia UFRJ	92
5.2 Anotação e análise dos genomas mitocondriais	95
5.2.1 Genes codificadores de proteínas	97
5.2.2 RNAs não codificantes	98
5.2.3 Regiões não codificantes	98
5.2.4 Origem de replicação	100
5.3 Busca de mitogenomas, alinhamento múltiplo e construção da árvore filogenética	103
5.4 Nível de expressão dos transcritos mitocondriais	104
5.5 Comparação do mitogenoma de <i>R. prolixus</i> , colônia UFRJ, com sequências da literatura para as espécies <i>R. prolixus</i> e <i>R. robustus</i> .	106
5.6 Descoberta e caracterização de polimorfismos de nucleotídeo único, inserções e deleções.	110
6. CONCLUSÃO	114
7. REFERÊNCIAS BIBLIOGRÁFICAS	115114
8. Anexos	

1. INTRODUÇÃO

1.1 Doenças Tropicais Negligenciadas

A expressão doenças tropicais negligenciadas (DTN) é usada para classificar um grupo de doenças infectocontagiosas, sendo que atualmente 20 doenças estão incluídas nesta lista da Organização Mundial da Saúde (OMS). Elas possuem agentes etiológicos distintos, da mesma forma suas manifestações clínicas, métodos diagnósticos e estratégias de tratamento/prevenção/controlam variam de acordo com a singularidade de cada enfermidade (FEASEY *et al.*, 2010). O termo tropical é empregado para agrupá-las devido a sua predominância em países tropicais, porém elas também estão presentes em ambientes subtropicais. Ao todo as DTNs estão presentes em 149 países do mundo, elas se disseminam e perpetuam em regiões financeiramente desprovidas das Américas, África e Ásia afetando aproximadamente um bilhão de pessoas (VASCONCELOS, KOVALESKI E JUNIOR, 2016; MOLYNEUX, SAVIOLI E ENGELS, 2017). Em geral, as DTNs são enfermidades que afetam majoritariamente, mas não exclusivamente, indivíduos de baixa e média renda, que vivem em condições de vida precária, com acesso limitado a saúde e a estrutura sanitária e habitacional adequadas (ALMEIDA, ALMEIDA E RAMALHO, 2017). O termo "negligenciadas" é efeito da indiferença histórica devido o pouco apelo das DNTs diante de grandes indústrias farmacêuticas, sedes de pesquisas de renome e de responsáveis por promover políticas de saúde (FEASEY *et al.*, 2010).

Entretanto, nos últimos anos essa situação começou a mudar, as DNTs passaram a ter cada vez mais atenção no cenário mundial. A globalização está impondo ao mundo um ritmo migratório cada vez mais acelerado que tem gerado efeitos a nível local, regional e internacional (SANSOM, 2018) Os agentes etiológicos não estão isentos dessas mudanças, o movimento populacional, por exemplo, tem feito com que parasitas considerados endemicamente estáveis entrem em movimento (BOOTH; CLEMENTS, 2018). Logo, as DNTs não estão mais geograficamente restritas aos locais de baixa renda, existindo exemplos de diagnóstico em países não

endêmicos (NORMAN *et al.*, 2010). A consequência do aumento de visibilidade dado às doenças tropicais negligenciadas pode ser visto no crescente número de documentos discutindo esse tópico. A declaração do Milênio, assinada em 2000 pelos membros da Organização das Nações Unidas (ONU), estabeleceu metas para reduzir em 15 anos a pobreza extrema, dentre as metas está à eliminação de doenças que afetam milhões de pessoas ao redor o mundo, inclusive as DNT's (HOTEZ *et al.*, 2007). Em 2005 a OMS criou um departamento específico encarregado de abordar a situação dessas enfermidades, e o resultado foi à criação de um plano global de combate destas doenças, onde a OMS estabeleceu que o controle do hospedeiro intermediário e do vetor é algo imprescindível para a prevenção, controle, eliminação e erradicação de doenças tropicais negligenciadas. Cinco das vinte DTNs são causadas por patógenos cujos vetores são insetos hematófagos e que estão presentes na América Latina – doença de Chagas, dengue/chikungunya, filariose, leishmaniose e oncocercose.

1.2 Doença de Chagas

A doença de Chagas (DC) é uma enfermidade parasitária sistêmica causada pelo protozoário flagelado *Trypanosoma cruzi* (Chagas, 1909). A transmissão vetorial do *T. cruzi* é considerada o modo de maior relevância epidemiológica, insetos hematófagos da subfamília Triatominae, popularmente conhecidos como barbeiros, são os responsáveis por promovê-la (COSTA *et al.*, 2013). Este parasita desenvolve o seu ciclo de vida em, pelo menos, dois organismos, um hospedeiro vertebrado e nos invertebrados (triatomíneos), onde apresenta diferentes estádios evolutivos (JUNQUEIRA; GONÇALVES; MOREIRA, 2011). A princípio, qualquer vertebrado é capaz de suprir as demandas alimentares dos triatomíneos, mas no meio ambiente muitos mamíferos acabam servindo de reservatório para o *T. cruzi*. A princípio, a DC era considerada uma enzootia, por ela se manifestar somente em animais silvestres, pois eles serviam de alimento para triatomíneos também silvestres. Porém, tornou-se uma zoonose como consequência da invasão do ser humano a biótopos naturais, deste modo possibilitando o seu contato com esses insetos e uma transmissão “acidental”, pois, assim como outros mamíferos, também são suscetíveis ao barbeiro (Coura,

2003). Com o desmatamento dos ambientes naturais para dar espaço à agricultura, criação de gado e moradias humanas, animais silvestres, inclusive, algumas espécies de triatomíneos conseguiram eficazmente se adaptar a ecótopos artificiais. O que propiciou que esses locais se tornassem coabitados por vetores, reservatórios e humanos, e devido à suscetibilidade do homem ao parasita essa enfermidade tornou-se uma antropozoonose (Figura 1) (DIAS JÚNIOR, 2011).

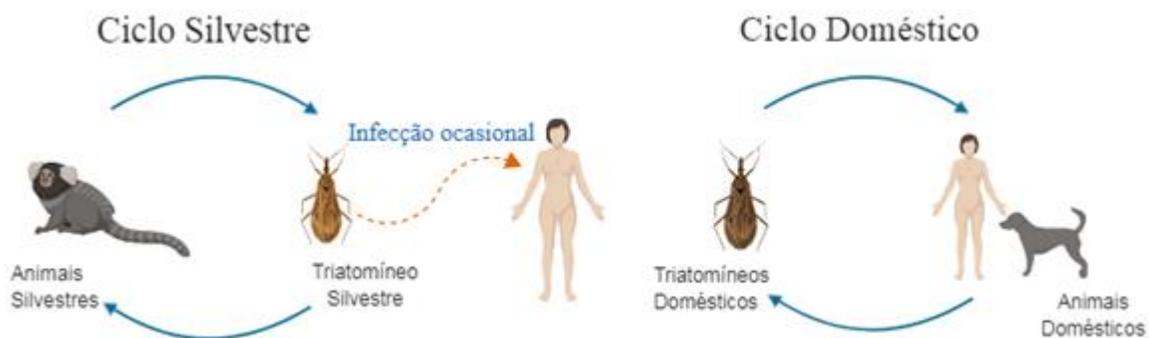


Figura 1 - Esquema dos dois ciclos de transmissão do *T. cruzi* existentes. No ciclo silvestre a transmissão do *T. cruzi* feita pelos triatomíneos ficaria restrita aos animais silvestres, mas a invasão do ser humano a esses espaços possibilitou a sua transmissão ocasional. Com o decorrer do tempo algumas espécies se adaptaram a viver em ambientes artificiais, chegando até a colonizar casas, isso tornou mais fácil à propagação da DC nos seres humanos.

Na transmissão vetorial o *T. cruzi* é transmitido pela excreta dos triatomíneos contaminados, isso acontece pois, em geral, eles defecam durante ou após o repasto e acabam liberando formas infectantes do parasito (Brasil, 2009). O ciclo se inicia quando o vetor adquire o parasita de um reservatório, normalmente, quando se alimentam de um mamífero. A forma infectante deste protozoário, a tripomastigota, circula no sangue do hospedeiro, o triatomíneo acaba a ingerindo ao se alimentar. Já no interior do barbeiro, o parasita se diferencia e duplica em epimastigota, em seu intestino médio, que após um tempo se diferencia em tripomastigotas metacíclicos na porção posterior do intestino. No próximo repasto, os barbeiros infectados, liberam os tripomastigotas metacíclicos presentes em seu intestino, que entram no corpo do hospedeiro quando ocorre o contato das fezes com a corrente sanguínea ou mucosas dos organismos.

Uma vez na corrente sanguínea a forma tripomastigota pode invadir ativamente diferentes células desse indivíduo ou então ser fagocitada por células do sistema imune do hospedeiro. O *T. cruzi* tem tropismo por células musculares, macrófagos e também por células epiteliais e fibroblastos, o que explica os sintomas clínicos apresentados pela DC. Quando invadem as células do hospedeiro, os tripomastigotas se diferenciam em amastigotas, responsáveis pela fase intracelular, que começam a se multiplicar por divisão binária. Ao se acumularem no interior das células, os amastigotas podem se diferenciar novamente em tripomastigotas, escapar de dentro da célula ao provocar lise celular e entrar em circulação novamente. Os tripomastigotas circulantes fecham o ciclo, já que podem ser ingeridos por um novo vetor ou infectar uma nova célula hospedeira (Figura 2) (WILKOWSKY, 2018; CDC, 2019). A maioria dos indivíduos infectados por *T. cruzi* carrega o parasito no sangue, tecidos e órgãos, durante toda a vida. Apesar desse modo ser o de maior relevância epidemiológica, não se deve desconsiderar os outros meios de transmissão da infecção chagásica como, transfusões de sangue; transplante de órgãos ou tecidos; via congênita; oral, pela ingestão acidental de alimento contaminado com o protozoário, e a considerada via de exceção, a acidental, decorrente por exemplo de acidentes laboratoriais (CARVALHO, 2018).

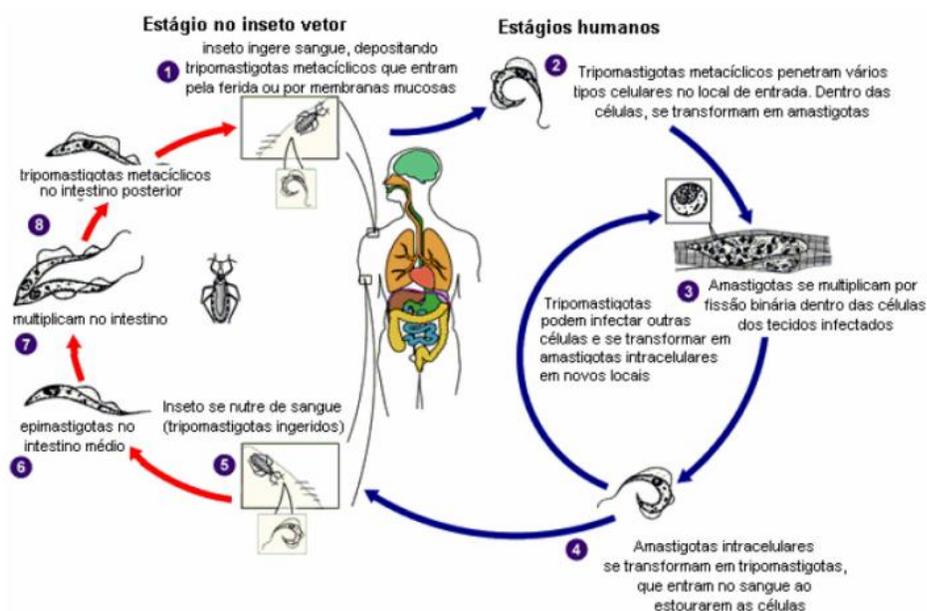


Figura 2 - Ciclo de vida do *Trypanosoma cruzi*. Fonte: Adaptado de (CDC, 2019).

Após a infecção pelo protozoário os hospedeiros podem ou não apresentar sintomas, pois essa enfermidade apresenta curso clínico bifásico, com uma fase aguda por vezes não identificada, podendo evoluir para a fase crônica. A maioria dos pacientes na fase aguda não apresentam sintomas ou manifestam sintomas inespecíficos (DIAS JÚNIOR, 2011). Durante a fase crônica, os pacientes também podem estar livres dos sintomas, mas alguns progridem para as formas clínicas da doença cardíaca, digestiva e cardiodigestiva. A gravidade dos casos está relacionada à cepa infectante, à via de transmissão e à existência de outras patologias concomitantes, que podem ser fatais se não diagnosticadas e não tratadas (AMARAL *et al.*, 2017). Após a contaminação, aproximadamente 30% dos pacientes infectados evoluem para a Cardiopatia Chagásica Crônica e em torno de 10% desenvolvem a forma digestiva da doença (AYO *et al.*, 2013; BESTETTI; RESTINI, 2014).

Apesar da DC ser primariamente limitada a 21 países do continente americano, devido a mobilidade populacional casos vêm sendo detectados em outros continentes (WHO, 2019a). A imigração de cidadãos latino-americanos contaminadas *por T. cruzi* para a América do Norte, Europa, Japão e Austrália, regiões sem experiência na prevenção, diagnóstico e tratamento da doença, viabilizaram a transmissão transfusional, congênita e por transplantes de órgãos em países não endêmicos (Figura 3). Casos de doença em viajantes que regressam de países latino-americanos e adoção de crianças naturais dessas regiões também propiciaram a expansão da distribuição dessa doença (PAHO, 2019).

Atualmente, a WHO estima que em torno de 8 milhões de pessoas no mundo todo estejam contaminadas pelo protozoário causador da DC, a grande maioria na América latina, e mais de 25 milhões estão correndo risco de infecção. Somando todas as formas de transmissão são relatados cerca de 56.000 novos casos de contágio por ano, que acabam gerando mais de 10.000 mortes devido a complicações da doença. No entanto, apesar de apresentar uma elevada carga de morbimortalidade em países da América latina, essa enfermidade ainda figura como uma doença tropical negligenciada.

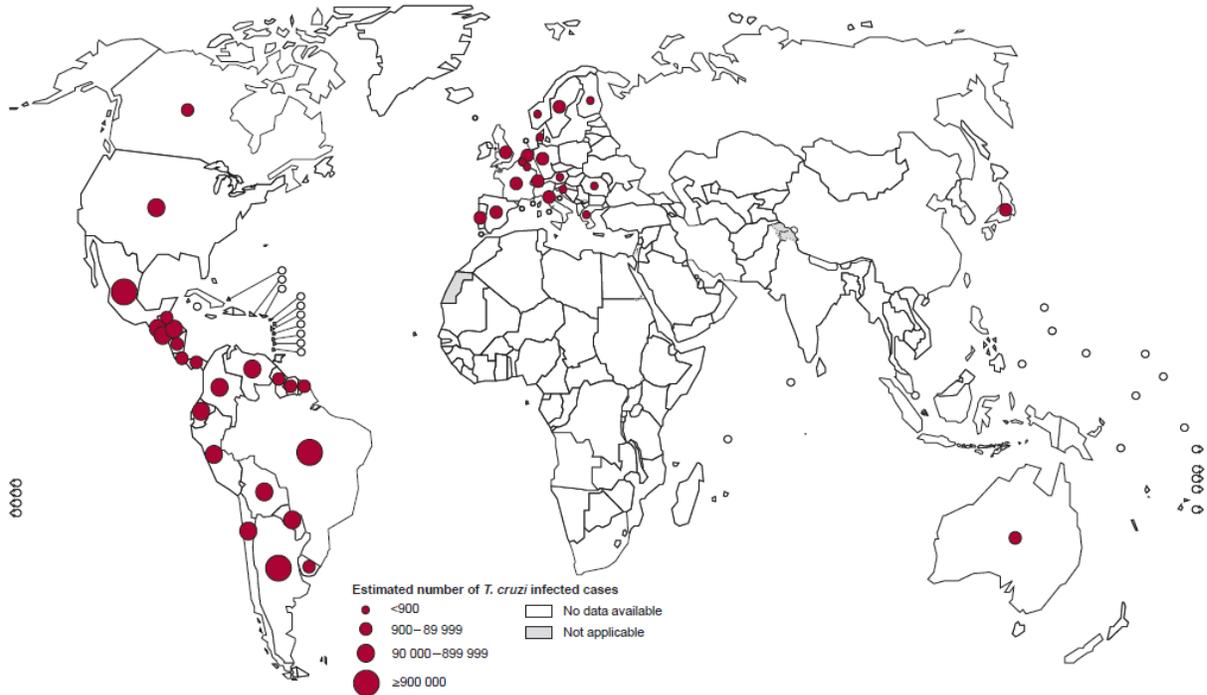


Figura 3 - Distribuição Global da Doença de Chagas, casos relatados a WHO no período de 2006 a 2010. Fonte: (WHO, 2019)

É necessário que haja mais ações efetivas para o melhoramento das medidas de prevenção e tratamento dessa patologia (WESTPHALEN; BISUGO; ARAÚJO, 2012). No presente, existe apenas dois fármacos disponíveis para o tratamento da DC, e eles são eficazes apenas na fase inicial da enfermidade, ainda não se dispõe de vacina para uso imediato (DIAS *et al.*, 2016).

Logo, a inexistência de um tratamento realmente competente contra a DC, que elimine ou evite o protozoário, e a grande chance de colonização e recolonização domiciliar pelos vetores, em áreas endêmicas, evidenciam a grande dificuldade no controle da doença. Diante desse cenário, a redução da propagação está intrinsecamente relacionada ao controle do vetor (DIAS *et al.*, 2016; PAVAN, 2013). Tanto que a resolução WHA 63.20 da sexagésima terceira Assembleia Mundial da Saúde em 2013 definiu como meta necessária para alcançar a erradicação da doença de Chagas, a interrupção da transmissão vetorial intradomiciliar na América latina até 2020 (WHO, 2019b). Sendo assim, é muito importante o desenvolvimento de novas

estratégias que possam controlar de maneira eficiente o agente transmissor.

No Brasil, hoje em dia, predominam os casos crônicos decorrentes da infecção por via vetorial em décadas passadas. Entretanto, ainda é relatado um número expressivo de pacientes com doença de chagas aguda, foram 2.330 casos confirmados no período de 2007 a 2017, destes 36 levaram a óbito. Entre os casos de DCA confirmados no Brasil, observou-se que a forma de transmissão oral foi a mais frequente. No entanto, vale ressaltar que quase 20% dos casos foram computados com a forma de transmissão ignorada ou sem preenchimento deste campo na ficha de notificação, sendo que 44,7% destes registros são do estado do Pará, estado com maior índice de transmissão vetorial do país, bem como 21,2% não informaram o estado de infecção. Apesar da ocorrência em menor proporção da transmissão vetorial (8,9%), vale ressaltar a persistência desta forma mesmo depois de 2006, ano em que o Brasil obteve a certificação de interrupção de transmissão por *Triatoma infestans* no País (Figura 4). Estes dados em conjunto reforçam a provável subestimativa dos casos relativos a esta forma de transmissão (SINAN, 2019). Além disso, outros países endêmicos ainda não conseguiram erradicar seu vetor principal, como a Colômbia com *Rhodnius prolixus* (WESTPHALEN; BISUGO; ARAÚJO, 2012).

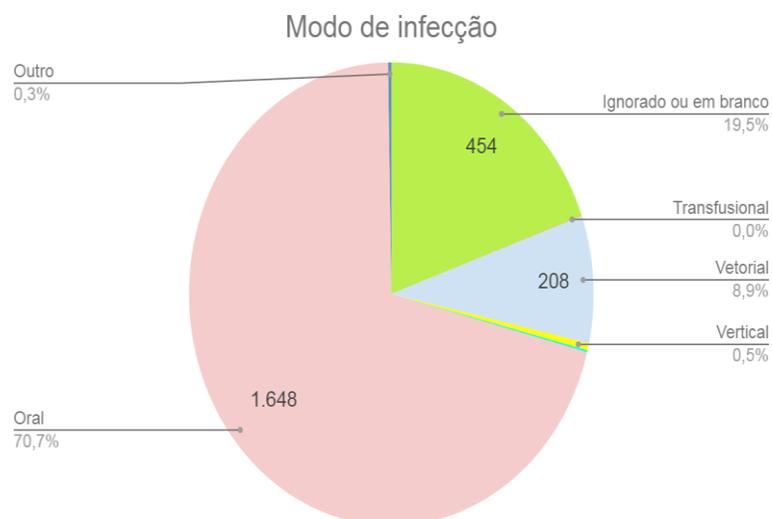


Figura 4 - Modos de infecção da Doença de Chagas no Brasil. Dados referentes ao período de 2007 a 2017. Persistência da transmissão vetorial mesmo após a interrupção de transmissão por *Triatoma infestans*, considerado principal vetor do país. Fonte: (SINAN, 2019).

Das mais de 150 espécies vetores conhecidas, 68 foram identificadas no Brasil (CARVALHO, 2018; BRASIL, 2009; Ministério da Saúde, 2019) (Figura 5). A momentânea exclusão da transmissão típica por *Triatoma infestans* não significa o controle efetivo da doença no país, já que outros triatomíneos tem potencial para ocupar o nicho ecológico deixado vago por essa espécie erradicada (CARVALHO, 2018).

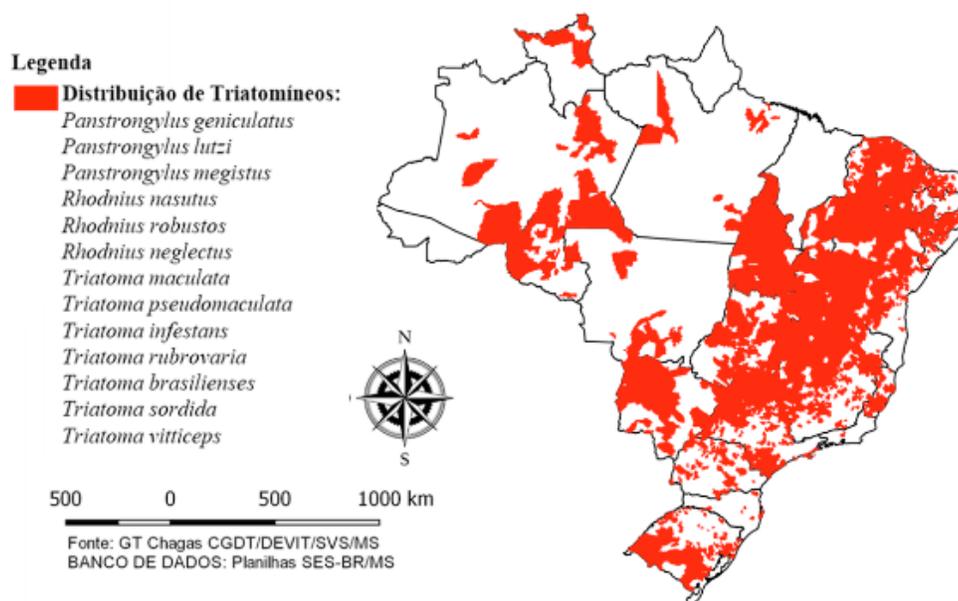


Figura 5 - Distribuição dos triatomíneos de maior importância epidemiológica ao longo do território brasileiro, ano de notificação 2012 e 2016. Fonte: (Ministério da Saúde, 2019).

1.3 Triatomíneos

Atualmente, a subfamília Triatominae, incluída na Ordem Hemiptera, Subordem Heteroptera, Família Reduviidae, é composta por 152 espécies ocorrentes e 2 extintas (VENDRAMI *et al.*, 2018). Os triatomíneos encontram-se amplamente distribuídos no continente americano, sua área de ocorrência vai desde o nordeste dos Estados Unidos até a Patagônia Argentina (Figura 6). Todavia, há relatos de espécies que habitam regiões tropicais na Ásia (GORLA; NOIREAU, 2017). Apesar de viverem em diferentes ecótopos, todos os triatomíneos são hematófagos, logo, potenciais vetores

do agente etiológico da DC. Até o momento, essa subfamília é agrupada em 5 tribos, Alberprosenini, Bolboderini, Cavernicolini, Rhodniini e Triatomini, que incluem 15 gêneros.

Ainda há controvérsias acerca da evolução dessa subfamília, em 1979 Lent e Wygodzinsky foram os primeiros a sugerir a monofilia do grupo com base na hematofagia e em características morfológicas que são adaptações associadas a seu hábito alimentar (LENT; WYGODZINSKY, 1979). Porém, devido a grande diversidade de habitats ocupados pelos triatomíneos, sua distribuição geográfica descontínua e ao fato de sua plasticidade morfológica, a hipótese de monofilia começou a ser contestada nos anos seguintes, surgindo a ideia da parafilia (SCHAEFER, 2003) e posteriormente da polifilia do grupo (SCHOFIELD; GALVÃO, 2009). Apesar disso, estudos recentes, com dados moleculares (WEIRAUCH; MUNRO, 2009) e morfológicos (WEIRAUCH, 2008) acharam evidências que suportam a hipótese original de monofilia criada por Lent e Wygodzinsky. Apesar desses valiosos esforços a filogenia do grupo ainda é uma questão delicada na literatura, já que mais tarde alguns estudos indicaram a parafilia de Triatominae em relação aos gêneros *Opisthacidius* e *Zelurus* (ZHANG *et al.*, 2016; JUSTI; GALVÃO, 2017; MONTEIRO *et al.*, 2018).

As duas Tribos mais conhecidas e bem estudadas desta subfamília são Triatomini e Rhodniini, isto por incluírem as espécies mais relevantes na transmissão para humanos. É evidente que o sucesso como vetor depende de determinadas características apresentadas pelas espécies, sem dúvida uma delas é a capacidade de domiciliação. Espécies capazes de habitar casas humanas, em geral, estão mais próximas da população e por isso, é mais provável que a transmissão vetorial seja proveniente de uma delas. Inclusive sabe-se que 12 barbeiros estão bem adaptados a habitações humanas, sendo estes considerados os principais vetores (DORN *et al.*, 2011). As espécies que são consideradas os principais vetores são: *Triatoma brasiliensis* e *Triatoma infestans* na América do Sul, *Triatoma dimidiata* e *Rhodnius prolixus* na América Central e norte da América do Sul.

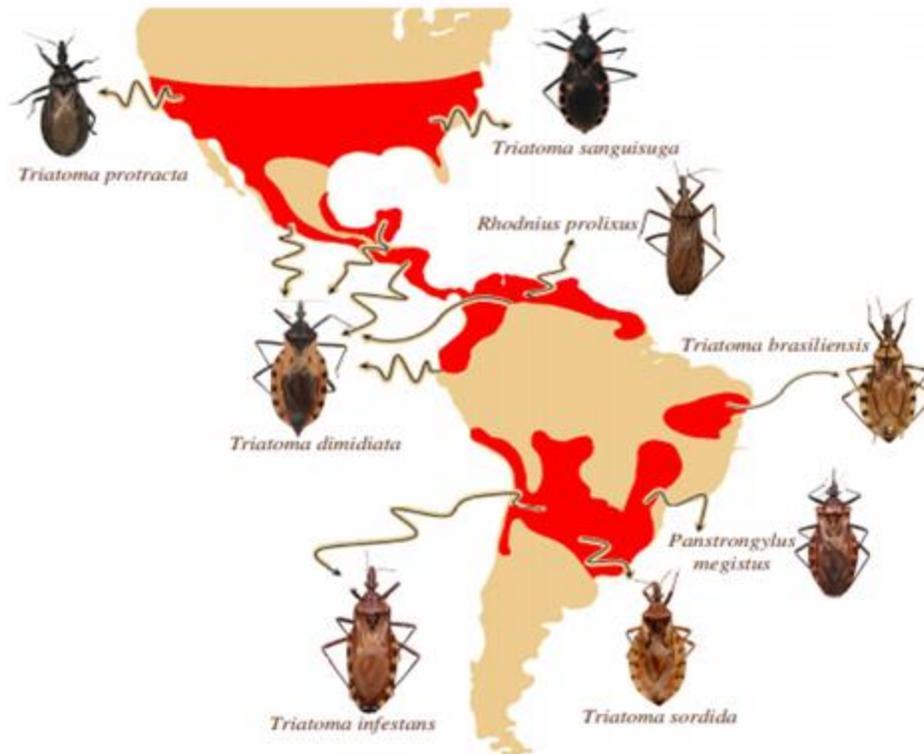


Figura 6 - Distribuição aproximada das espécies da Subfamília Triatominae de maior relevância epidemiológica. Fonte: (DORN *et al.*, 2011)

1.4 *Rhodnius prolixus*

O gênero *Rhodnius* é dividido em duas grandes linhagens evolutivas: os grupos *prolixus* e *pictipes* distribuídos a leste da Cordilheira dos Andes (também chamados de cis-andinos) e o grupo de *pallescens* a oeste dos Andes (trans-andino) (JUSTI; GALVÃO, 2017). O gênero tem ocorrência natural desde a América Central até o norte da Argentina, todavia a Amazônia é uma região que concentra uma alta riqueza de espécies. Esses vetores são encontrados com mais frequência em copas de palmeiras no ambiente silvestre, porém também podem ser encontrados eventualmente, no peridomicílio e no interior de habitações humanas.

Rhodnius prolixus (Stal, 1859) é um de seus membros mais conhecidos devido a sua importância epidemiológica, em grande parte consequência da sua adaptação a

habitações humanas, na Venezuela, Colômbia e em países da América Central (PAVAN, 2013) (Figura 7). Vale ressaltar também sua elevada capacidade reprodutiva, possuindo até três gerações anuais. Ademais além de transmitir *T. cruzi*, sabe-se que ele também é capaz de abrigar outro parasita, o *Trypanosoma rangeli* (Tejera, 1920), e ainda infecções mistas de *T. cruzi* e *T. rangeli*. Apesar desse protozoário flagelado não ser causador de doenças, a resposta imune induzida pela sua presença no hospedeiro leva a produção de anticorpos que podem gerar reações cruzadas com antígenos do *T. cruzi* interferindo no diagnóstico da doença de Chagas (BELINATO, 2014). Na literatura ele já foi registado em países como Bolívia, Brasil, Equador, Guiana Francesa, Guiana, Panamá e Suriname mas isso se deve a sua errônea identificação, sendo confundido com outro triatomíneos como por exemplo o *Rhodnius robustus* (GORLA; NOIREAU, 2017). É muito difícil discriminar *R. prolixus* e *R. robustus* baseado apenas em caracteres morfológicos, entretanto sua distinção ficou evidente após estudos filogenéticos empregando dados moleculares, no momento sabe-se que além de serem táxons diferentes, portanto com material genético distintos, *R. robustus* ainda abrange linhagens divergentes (MONTEIRO *et al.*, 2018). A correta identificação dessas espécies é de suma importância para o controle do transmissor da Doença de Chagas, já que táxons distintos podem apresentar capacidade vetorial diferente, a diversidade genética deles pode influenciar na sua adaptação como vetor, por exemplo na sua domesticação ou no desenvolvimento de resistência a pesticidas (GOURBIÈRE *et al.*, 2012).



Figura 7 - Distribuição geográfica de *R. prolixus* na América Central e do Sul. Fonte: (BELINATO, 2014).

Apesar dos avanços feitos no último século ainda há pouca informação genética acerca desses insetos. Até o momento, dos triatomíneos apenas *R. prolixus* tem o seu genoma sequenciado (MESQUITA *et al.*, 2015). Outros 3 membros da tribo Triatomini tiveram seu genoma mitocondrial publicado, *Triatoma dimidiata* (DOTSON; BEARD, 2001), *Triatoma infestans* (PITA *et al.*, 2017) e *Triatoma rubrofasciata* (DONG *et al.*, 2018). Logo, o estudo do genoma mitocondrial deste organismo não somente traz mais informações de um vetor com grande importância epidemiológica, como pode ser de grande utilidade para a resolução das questões filogenéticas não só do gênero *Rhodnius* como da subfamília Triatominae.

1.5 Genoma mitocondrial

As mitocôndrias são organelas citoplasmáticas, presentes na maioria das células eucarióticas, responsáveis pela geração de energia celular através da liberação de ATP (LAVROV; PETT, 2016). As mitocôndrias despertam a atenção dos cientistas a tempos, ainda na década de 60 foi descoberto que essas organelas apresentam genoma próprio (SMITH, 2016), a partir deste momento iniciou-se uma busca pelo sequenciamento desse material genético. Foi somente em 1981 que o primeiro genoma mitocondrial (mitogenoma) finalmente foi sequenciado por completo, o genoma mitocondrial humano (ANDERSON *et al.*, 1981). Posteriormente, outras espécies tiveram seus mitogenomas sequenciados, alguns de forma completa e outros apenas sequências parciais. Apenas quatro anos depois, em 1985, foi sequenciado o primeiro genoma mitocondrial de inseto, o da *Drosophila* sp. (SMITH, 2016).

Em geral, o mitogenoma dos insetos é formado por uma molécula circular de dupla fita, que possui em média de 15.000 a 18.000 pares de bases (CAMERON, 2014). Algumas diferenças ficam evidentes quando comparados ao genoma nuclear, nele há um empacotamento gênico denso, onde a maior parte dos nucleotídeos constitui uma sequência codificadora, logo, com poucas regiões não codificantes e também sem íntrons, além disso, os códons tem um uso mais flexível nessas organelas e o código genético é variante, ao menos quatro códons codificam aminoácidos diferentes dos realizados pelo código “universal” (ALBERTS *et al.*, 2010). Estruturalmente, esse cromossomo contém 37 genes; dos quais 13 são genes codificantes de proteínas (PCG - do inglês *Protein coding sequence*) estruturais e funcionais relacionadas ao transporte de elétrons e fosforilação oxidativa, dois são de RNA ribossomal (RNAr) e 22 são de RNA transportador (RNAt) envolvidos na tradução dos GCPs. Também está presente regiões não-codificadoras, como alguns pequenos espaçadores intergênicos (IGS – do inglês *InterGenic Spacer*) e uma região rica em A+T, denominada região de controle, por conter a origem de replicação e transcrição (CAMERON, 2014). Entretanto, apesar de ter uma maquinaria própria a mitocôndria ainda necessita de proteínas sintetizadas e transcritas no núcleo celular, que posteriormente serão traduzidas no citoplasma e importadas para o interior dessa

organela, como por exemplo, proteínas envolvidas na replicação e reparo do DNA (TAKAHASHI, 2010) (Figura 8).

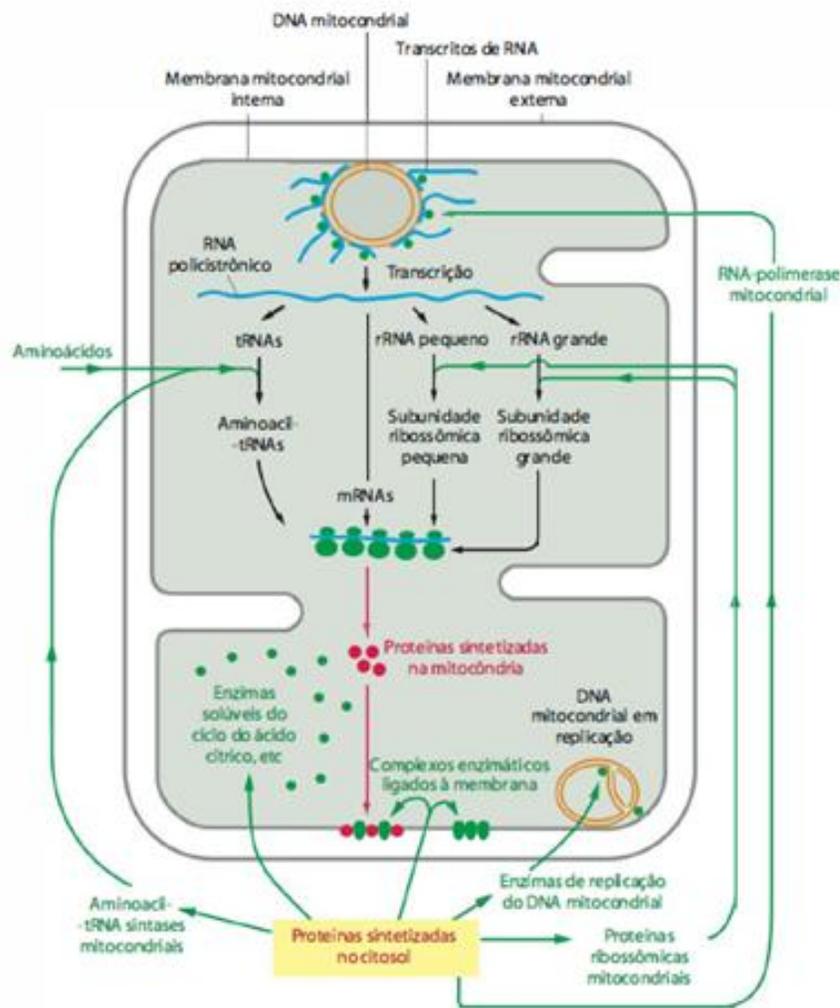


Figura 8 - Proteínas e RNAs presentes nas mitocôndrias, o esquema aponta a cooperação que existe entre a mitocôndria e o núcleo. O sistema, como um todo, só se mantém pela soma da presença de proteínas codificadas pelo genoma mitocondrial e de proteínas codificadas no núcleo e posteriormente importadas para o interior da mitocôndria. Fonte: (ALBERTS *et al.*, 2010).

O processo responsável por gerar energia nas mitocôndrias é chamado de fosforilação oxidativa, nele elétrons de alta energia, NADH e FADH, são transferidos ao longo da cadeia transportadora de elétrons para o oxigênio. O transporte desse elétron gera um gradiente de prótons que é usado para produzir ATP pela enzima ATP-sintase.

Até o momento, foram identificados 92 genes participando da estrutura desse processo, 13 são os PCGs codificados pelo mitogenoma e outros 79 são importados do núcleo celular. O complexo I, também chamado NADH-desidrogenase, é o maior dos complexos enzimáticos da cadeia respiratória, ele é composto por 44 cadeias polipeptídicas, das quais 14 enzimas constituem a sua subunidade central (NAD1, NAD2, NAD3, NAD4, NAD4I, NAD5 E NAD6 são provenientes do genoma mitocondrial já as outras 7 são importadas do núcleo). Mais 30 subunidades acessórias auxiliam a manter a estabilidade deste complexo. O complexo II é formado inteiramente por proteínas nucleares, são quatro subunidades. O complexo III, chamado citocromo c oxidoreductase, contém 11 subunidades, apenas 1 é codificada pelo mitogenoma, o citocromo b (CYTB). O complexo IV, ou citocromo c oxidase, consiste em 11 subunidades importadas do núcleo associadas a três subunidades codificadas pelo mitogenoma, Citocromo oxidase I (COXI), Citocromo oxidase II (COXII) e Citocromo oxidase III (COXIII). Por último, o complexo V, ATP-sintase, compreende 19 subunidades, 2 são codificadas pelo genoma mitocondrial (ATP6 e ATP8) (CHINNERY; HUDSON, 2013; ALBERTS *et al.*, 2010).

Apesar desse tráfego de proteínas a replicação do DNA das mitocôndrias não é restrito a fase S do ciclo celular, ou seja, é independente do tempo da replicação do genoma nuclear (CHINNERY; HUDSON, 2013). A replicação mitocondrial nos insetos foi mais estudada em espécies do gênero *Drosophila*. Nesses organismos, foram encontradas evidências que sugerem que a replicação ocorra de forma assimétrica (SAITO; TAMURA; AOTSUKA, 2005). O modelo de replicação assimétrico pressupõe que há uma origem de replicação específica para cada cadeia (OJ - Origem da fita *majority* (J) e ON- Origem da fita *minority* (N)), e que a síntese ocorre de forma unidirecional (Figura 9) (YASUKAWA; KANG, 2018). A síntese da fita *majority* é iniciada quando 97% da síntese da fita *minority* está concluída (GODDARD; WOLSTENHOLME, 1980). Logo, as origens das duas fitas estão localizadas a 97% de distância genômica. Este modelo de replicação faz com que a fita *minority*, que é utilizada como molde para a replicação da *majority*, fique exposta momentaneamente na forma de fita simples. As duas origens estão localizadas na região controle do mitogenoma das *Drosophilas*. SAITO; TAMURA; AOTSUKA (2005) mapearam

precisamente a extremidade 5' nas novas fitas sendo formadas pelas duas origens do mitogenoma de espécies de *Drosophila* (OJ está localizada a poucos nucleotídeos de distância do tRNAI, já ON está localizada mais próximo do rRNAs). Nesses insetos uma sequência de timinas está localizada *upstream* das origens, o que sugere que esta sequência de mononucleotídeos pode estar envolvida no reconhecimento da origem de replicação.

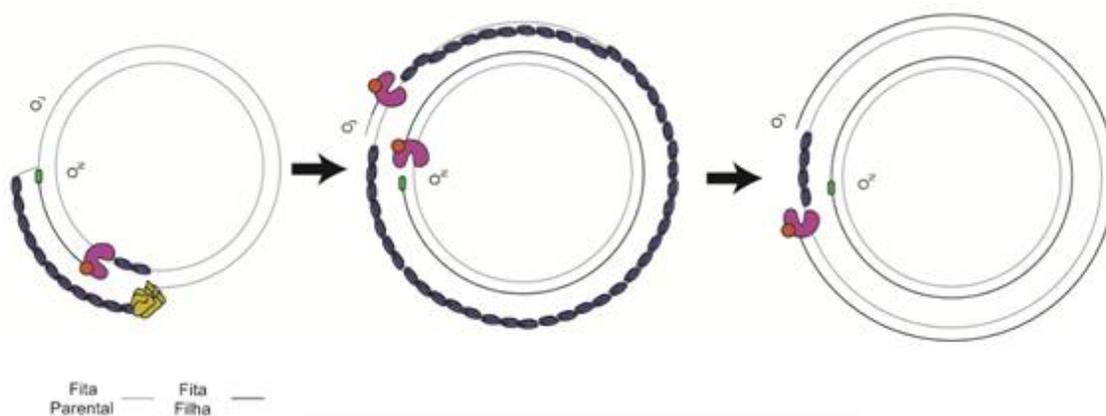


Figura 9 - Esquema da replicação assimétrica no mitogenoma de *Drosophila*. A síntese da fita *majority* (J) só é iniciada quando 97% da síntese da fita *minority* (N) está concluída. Fonte: (KAUPPILA, 2017)

Usando como referência a região da origem de replicação de *Drosophila*, o trecho de timinas foi encontrado no mitogenoma de insetos holometábolos, tais como *Tribolium castaneum* e *Bombyx mori* (SAITO; TAMURA; AOTSUKA, 2005). Entretanto, a cauda poli-T não foi encontrada no mitogenoma de *Locusta migratoria*, um membro da ordem Orthoptera, que teve a síntese identificada no mesmo estudo de SAITO; TAMURA; AOTSUKA (2005). Nos mitogenomas de outros hemimetábolos, assim como *L. migratoria*, poucos trechos de timinas foram encontrados na região que se especula estar as origens de replicação. Isto sugere que em *L. migratoria*, e em outros insetos que não tenham o trecho de timinas, um elemento diferente esteja envolvido na regulação da replicação do mitogenoma. Em *L. migratoria* uma estrutura em grampo está localizada 20 pb *upstream* de ON (SAITO; TAMURA; AOTSUKA, 2005). Esta

estrutura é um candidato promissor para a regulação pois a fita *heavy* (equivalente a fita *minority* de insetos), que contém a origem da fita *light*, apresenta um grampo no mitoDNA de *Homo sapiens* (GUSTAFSSON; FALKENBERG; LARSSON, 2016).

O mecanismo envolvido na transcrição do mitogenoma de insetos ainda é pouco conhecido e, infelizmente, poucos estudos foram feitos nesses organismos. Boa parte do conhecimento transcricional se baseia em teorias desenvolvidas para vertebrados. Historicamente, estudos mostravam a existência de três regiões promotoras no mitogenoma de *Homo sapiens*, uma para a fita *light* chamado LSP (LSP - do inglês *Light Single Promoter*) e duas para fita *heavy* (HSP1 e HSP2 - do inglês *Heavy Single Promoter*) (TERZIOGLU *et al.*, 2013). Porém, a existência de três promotores é um assunto bem controverso, alguns estudos sugerem a existência de apenas um promotor para fita *heavy* (*minor* em insetos), sendo eles HSP1 e HSP2 (ZOLLO; SONDEIMER, 2017). O debate sobre essa questão ainda existe, porém BLUMBERG *et al.* (2017) estudando transcritos nascentes em linhagens de células humanas, usando as técnicas PRO-Seq. e GRO-Seq., acharam evidências que suportam a existência e atividade dos três promotores. A transcrição nos mitogenomas é policistrônica, ou seja, produz um precursor de RNA que contém todos os genes controlados por um mesmo promotor (GUSTAFSSON; FALKENBERG; LARSSON, 2016). Logo, esse precursor deve ser processado para gerar moléculas individuais de mRNA, tRNA e rRNA (Figura 10) (D'SOUZA; MINCZUK, 2018). A edição deste precursor é guiado pela localização dos RNAs transportadores, os tRNAs que são codificados entre os genes são reconhecidos e removidos, assim os genes restantes de proteínas e os RNAs ribossomais podem ser processados para assumir sua forma funcional. Este mecanismo é conhecido como modelo de pontuação de tRNA (do inglês *tRNA punctuated model*) (OJALA *et al.*, 1981).

Em *Drosophila*, após a clivagem dos tRNAs os transcritos restantes recebem uma cauda poli-A. Neste inseto foram identificados 11 mRNAs com cauda poli-A, nove correspondem a mRNAs de um único gene, e dois são bicistrônicos (contêm pares de genes que não possuem um tRNA os separando, ATP8/ATP6 e NAD4I/NAD4)

(BERTHIER *et al.*, 1986; STEWART; BECKENBACH, 2009). Além disso, ainda foi detectado poliadenilação nos dois RNAs ribossomais (STEWART; BECKENBACH, 2009).

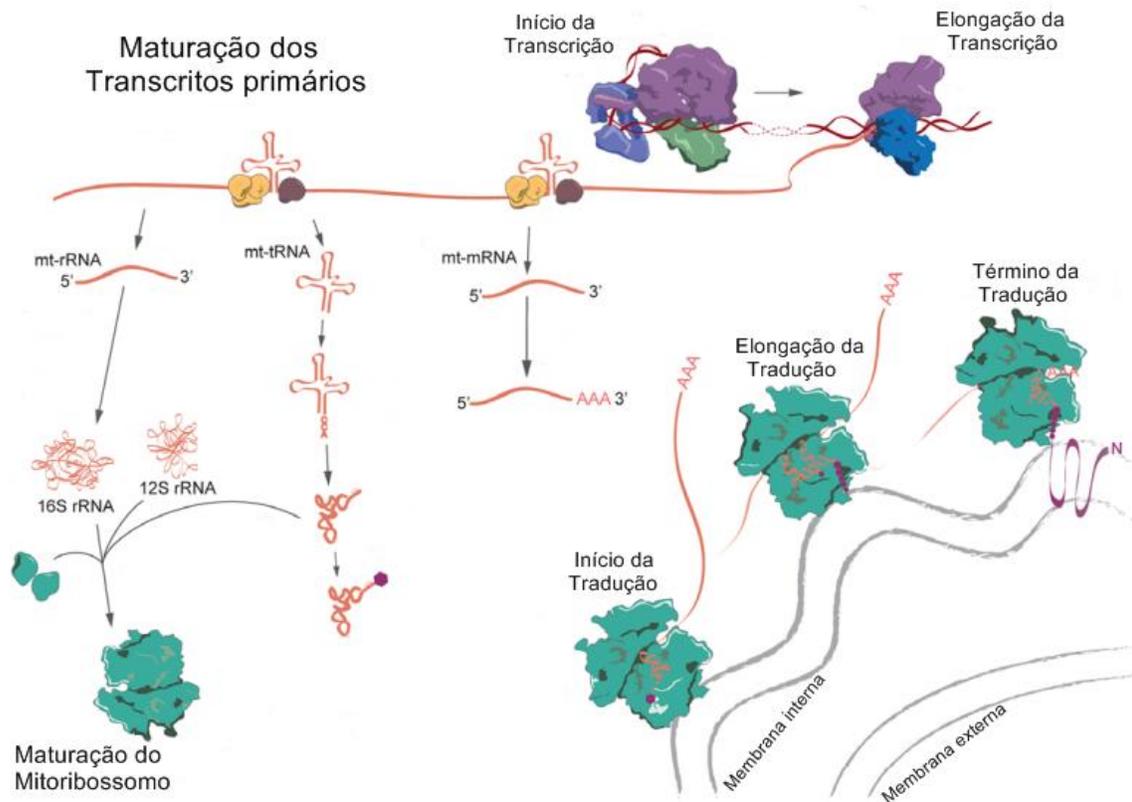


Figura 10 - Visão geral da transcrição mitocondrial humana, processamento pós-transcricional e tradução de RNA. O processamento do transcrito primário é guiado pela localização dos RNAs transportadores, conhecido como modelo de pontuação de tRNA. Após a clivagem dos tRNAs os RNAs mensageiros e RNAs ribossomais podem ser processados para assumir sua forma funcional.

O estudo dos mitogenomas com o decorrer do tempo contribuiu para uma ampla gama de campos científicos, sendo difícil detalhar todos eles. Para a medicina, por exemplo, informações provenientes desses estudos contribuíram para o diagnóstico e potencial tratamento de distúrbios genéticos associados às mitocôndrias (Smith, 2016). Além disso, para muitos táxons as sequências mitocondriais obtidas e estudadas constituem a pouca informação gênica disponível, anteriormente ausente. Mas, sem dúvida, seu uso é muito empregado em pesquisas de genética populacional e biologia evolutiva. Os genes mitocondriais são marcadores genéticos utilizados para muitos

fins, dentre eles pode-se citar a inferência filogenética, identificação de origem de espécies, filogeografia, análise da estrutura e dinâmica populacional e evolução molecular (ZHANG; HEWITT, 1997). Várias características contribuem para o sucesso do DNA mitocondrial como marcador molecular, tais como a ausência de recombinação nessa molécula, a herança não-mendeliana, ou seja, unicamente uniparental, ao contrário do genoma nuclear que possui herança biparental, alto número de cópias e expressão e altas taxas de evolução quando comparado ao DNA nuclear (em alguns casos chegando a ser até 10 vezes maior) (BARBOSA, 2018; NEVES, 2012).

Como dito anteriormente, há apenas três genomas mitocondriais de triatomíneos disponíveis na literatura; *Triatoma dimidiata* (DOTSON; BEARD, 2001), *Triatoma infestans* (PITA *et al.*, 2017) e *Triatoma rubrofasciata* (DONG *et al.*, 2018). Os mitogenomas seguem o arranjo e orientação de genes mitocondriais encontrado em insetos ancestrais (Figura 11) (CAMERON, 2014). A Tabela 1 apresenta informações comparativas sobre os genomas mitocondriais disponíveis na literatura.

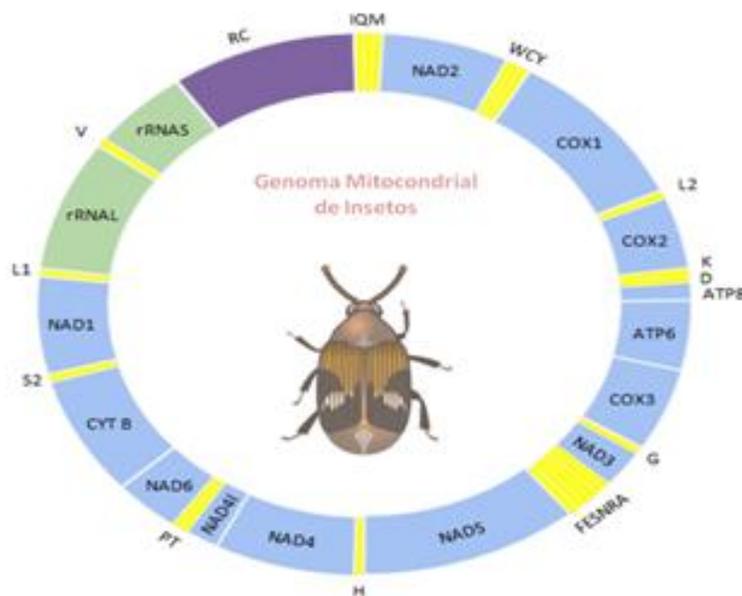


Figura 11 - Diagrama do genoma mitocondrial de insetos ancestrais. Os genes de RNA de transferência são indicados pelas abreviações IUPAC-IUB de uma letra para o seu aminoácido correspondente, coloridos em amarelo. Os genes que codificam proteínas estão em azul e os responsáveis pelo RNA ribossomal em verde. Em roxo está representada a região de controle (RC). Adaptado de (CAMERON, 2014).

Tabela 1 - Comparação entre os genomas mitocondriais de triatomíneos totalmente sequenciados até o momento.

Espécies de Triatomíneos	<i>T. dimidiata</i>	<i>T. infestans</i>	<i>T. rubrofasciata</i>
Tamanho (pb)	17.019	17.301	17.150
Conteúdo AT (%)	69,5	68,7	69,8
Total de genes	37	37	37
Genes para proteína	13	13	13
Genes para rRNA	2	2	2
Genes para tRNA	22	22	22
Número de acesso ao GenBank	AF301594	KY640305	MH934953

Entretanto, mesmo antes da publicação do mitogenoma de *T. dimidiata* alguns estudos já tinham sido realizados com sequências parciais, principalmente usando genes considerados marcadores moleculares. Em 1998 três marcadores mitocondriais (as subunidades ribossomais e COXI) foram utilizados para investigar a relação filogenética de oito membros da tribo Triatomini (GARCÍA; POWELL, 1998) e Stothard *et al.*, (1998) usou um fragmento da subunidade maior ribossomal para tentar desvendar a relação entre nove triatomíneos, incluindo três generos *Triatoma*, *Panstrongylus* e *Rhodnius*, *R. prolixus* estava representado. Pouco tempo depois (MONTEIRO *et al.*, (2000) publicou um artigo muito importante para a taxonomia dos membros do gênero *Rhodnius*, principalmente *R. prolixus* e *R. robustus*. Usando sequências da subunidade maior ribossomal e do citocromo b junto com um marcador nuclear, o estudo conseguiu inferir na filogenia dos Rhodniini e confirmar *R. robustus* como espécie distinta de *R. prolixus*. Este tema foi abordado novamente em estudos posteriores, como (MONTEIRO *et al.*, 2003) que empregou sequências do citocromo b de 26 populações de *R. prolixus* e *R. robustus*, que reforçou a hipótese levantada na pesquisa anterior, e também (FITZPATRICK *et al.*, 2008; PAVAN, 2013; NASCIMENTO *et al.*, 2019).

Até pouco tempo atrás, definir o mitogenoma de uma espécie desejada era um trabalho árduo e de alto custo, devido a etapas de clonagem e/ou as longas ampliações por PCR, que além de apresentarem elevado custo financeiro ainda exigiam muito tempo de preparação. Com o avanço da tecnologia dos sequenciadores de nova geração (NGS – do inglês *Next-Generation Sequencing*), tornou-se possível o uso de novas abordagens para identificar as sequências mitocondriais. Uma delas é o uso de mitogenomas, de espécies próximas, disponíveis na literatura como molde para reconstruir o genoma da espécie alvo, método de grande valia principalmente quando se trabalha com organismos não modelos. O alto número de cópias do genoma mitocondrial e seu elevado nível de expressão representa uma parcela considerável dos *reads* gerados em um sequenciamento NGS, o que permite que muitas vezes o mitogenoma quase total seja recuperado. A eficiência dessa técnica fica evidente ao observar genomas montados por essa abordagem (MOREIRA *et al.*, 2016; PITA *et al.*, 2017).

2. OBJETIVO

2.1 Objetivo geral

O objetivo deste trabalho foi estudar o genoma mitocondrial de *Rhodnius prolixus*

2.2 Objetivos específicos

- Anotação dos genes no genoma mitocondrial de *Rhodnius prolixus*, colônia NIH
- Montagem do genoma mitocondrial de *Rhodnius prolixus*, colônia UFRJ, utilizando o transcriptoma
- Anotação dos genes do genoma mitocondrial da colônia UFRJ.
- Identificação da expressão dos genes mitocondriais na colônia UFRJ
- Identificação da origem do mitogenoma da colônia UFRJ
- Avaliação de variantes no genomas mitocondrial da colônia UFRJ

3. Metodologia

3.1 Montagem do Genoma Mitocondrial de *Rhodnius Prolixus*, colônia UFRJ

A etapa inicial do estudo do genoma mitocondrial de *Rhodnius prolixus* consistiu na montagem do mitogenoma da colônia UFRJ. Para isso, as sequências mitocondriais foram buscadas nos transcriptomas disponíveis no laboratório utilizando um mitogenoma molde.

As bibliotecas usadas neste estudo foram provenientes do trabalho desenvolvido por colaboradores (CONTRERAS, 2015). O barbeiro *R. prolixus* mantido na colônia UFRJ/IBqM, cresceu a temperatura de 28°C com uma umidade relativa de 70%-80%. Neste estudo foram usadas as formas tripomastigotas e epimastigotas do protozoário *T. cruzi*, respectivamente cepa Y e CL-Brenner, para infectar ninfas de primeiro estágio em diferentes intervalos de tempo. Pode-se dividir os insetos em quatro grupos amostrais: o primeiro contém ninfas alimentadas com sangue; o segundo é de ninfas alimentadas e infectadas com a forma epimastigota; o terceiro são as ninfas alimentadas e infectadas com a forma tripomastigota; o último grupo é composto pelos insetos que não foram alimentados. Os grupos alimentados foram analisados 1, 4, 8 e 12 dias após a infecção feita nos grupos com o protozoário. Além disso, o grupo epimastigota e tripomastigota ainda é subdividido em dois, pois as ninfas foram infectadas com dois níveis de infecção por *T. cruzi*, 10^3 ou 10^7 parasitas por ml de sangue (Tabela 2). Foi feita uma triplicata para cada amostra, em cada réplica foi utilizado um conjunto de 50 ninfas. As bibliotecas foram sequenciadas pela tecnologia *Illumina* HiSeq2000.

A sequência escolhida como modelo foi o genoma mitocondrial *Rhodnius prolixus* da colônia NIH que foi previamente sequenciado por Sanger por colaboradores, que tem sua origem de uma colônia na Guatemala/MERTU.

Tabela 2 - Bibliotecas do transcriptoma de *Rhodnius prolixus*. A tabela contém informações acerca da quantidade e tamanho dos *reads* em cada condição, assim como do conteúdo % AT.

Amostra	Condição	Dia	Total reads	Tamanho do read	%AT
J	Jejum	-	6246450	101	59
			6726759	101	59
			6357987	101	60
S1	Sangue	1	7225473	101	61
			7231454	101	61
			6977832	101	62
S2		4	6831367	101	62
			6365569	101	62
			6917755	101	62
S3		8	6087647	101	61
			6886925	101	61
			6481183	101	62
S4		12	6563139	101	62
			6174558	101	62
			6628423	101	62
E1	Epimastigota 10 ³	1	7377261	101	61
			5812136	101	61
			6411395	101	61
E2		4	6688457	101	62
			7005241	101	62
			6177312	101	62
E3		8	5631649	101	62
			5641775	101	62
			5994291	101	62
E4		12	6025107	101	61
			6621110	101	62
			5888603	101	62
E5	Epimastigota 10 ⁷	1	6011569	101	61
			6593069	101	61
			6288728	101	61
E6		4	5801258	101	62
			6358742	101	61
			5878117	101	61
E7		8	6914199	101	61
			6595824	101	62
			7708321	101	61
E8		12	7253796	101	62
			6489302	101	63
			6462764	101	62

Tabela 2 - Bibliotecas do transcriptoma de *Rhodnius prolixus*. A tabela contém informações acerca da quantidade e tamanho dos *reads* em cada condição, assim como do conteúdo % AT. (Continuação).

Amostra	Condição	Dia	Total reads	Tamanho do read	%AT
T1	Tripomastigota 10 ³	1	6406586	101	61
			6565412	101	62
			6379660	101	62
T2		4	6792809	101	62
			6229216	101	62
			6403123	101	62
T3		8	6483184	101	63
			5641961	101	62
			7661809	101	62
T4		12	6816923	101	61
			6719875	101	62
			6300558	101	63
T5	Tripomastigote 10 ⁷	1	5892552	101	63
			5931355	101	62
			5227097	101	62
T6		4	6289071	101	62
			5967411	101	62
			5545812	101	62
T7		8	6210255	101	62
			5557271	101	62
			7288301	101	63
T8		12	6593399	101	62
			5877468	101	63
			5722239	101	62

A montagem do mitoDNA completo da colônia UFRJ foi realizada pela conciliação de 3 montagens diferentes (Figura 12). A primeira montagem (M1) consistiu simplesmente do conjunto de *contigs* mitocondriais presentes na montagem global do transcriptoma. A segunda e terceira estratégias (M2 e M3) foram empregadas para tentar cobrir toda a extensão do mitogenoma-NIH. M2 e M3 se basearam em montagens somente dos *reads* mitocondriais com os programas SPAdes e Trinity.

Primeira montagem (M1)

Os dados brutos do transcriptoma foram avaliados pelo programa FastQC versão 0.11.5 (ANDREWS, 2010). FastQC é um programa que gera um relatório sobre

a qualidade das sequências de um arquivo fastq, portanto, auxilia na decisão de estratégias que melhorem a qualidade dos dados gerados. De modo, que após essa avaliação pode-se determinar os melhores parâmetros para a limpeza das bibliotecas. Em seguida, os *reads* foram limpos usando o programa (Martin, 2011) Cutadapt versão 1.16 (MARTIN, 2011). A ferramenta Cutadapt foi empregada para remover os adaptadores das sequências produzidas pelo sequenciamento *illumina*, e aparar bases de baixa qualidade na extremidade 3' dos *reads* antes da remoção do adaptador, foram removidas bases com qualidade abaixo de 30. O nível máximo de tolerância de erros (mismatches, inserções e deleções) entre as sequências dos adaptadores e sequências presentes nos *reads*, para que uma sequência adaptadora seja efetivamente removida foi de 0.2. Ou seja, admitiu-se uma taxa de erro de 20%, para uma sequência ser removida ela poderia diferir em até 6 posições do adaptador *illumina*. Neste processo *reads* inferiores a 90 bp de comprimento também foram deletadas. Sucessivamente a este processo, as *reads* foram avaliadas pelo programa FastQC, desta vez para ter certeza da remoção dos contaminantes.

Os *reads* limpos do transcriptoma foram usadas como *input* no programa Trinity versão 2.2.0 (GRABHERR *et al.*, 2011) para serem montadas usando a configuração padrão. Os *contigs* montados, que são a união de dois ou mais *reads*, foram então transformados em um banco de dados usando a ferramenta makeblastdb, versão 2.2.31. Para formatar os *contigs* além de fornecer o arquivo fasta foi informado ao programa que o banco de dados seria de nucleotídeo. Após isso o transcriptoma ficou disponível para ser usado pela ferramenta BLAST (*Basic Local Alignment Search Tool*) (ALTSCHUL *et al.*, 1990).

A identificação dos mito-*contigs* na montagem M1 foi feita usando o algoritmo BLASTn versão 2.2.31 (ALTSCHUL *et al.*, 1990) para comparar a similaridade dos *contigs* da colônia UFRJ no mitogenoma da colônia NIH. O *hit* de menor identidade foi de 85%. Entretanto, houve uma seleção manual por posição, em caso de sobreposição completa de dois *contigs* o menor foi descartado, independente da identidade dos nucleotídeos. A identidade não foi levada em consideração neste primeiro momento, pois após a montagem foi realizada uma etapa de verificação. Ou seja, para a

montagem do mitogenoma foram usados os maiores *contigs* de M1, de forma a cobrir a maior porcentagem possível do mitogenoma. Os *contigs* selecionados (M1s) foram editados manualmente de acordo com o sentido e a região que se encontravam no mitogenoma.

Segunda e terceira montagens (M2 e M3)

Na nova estratégia os dados brutos da colônia UFRJ foram alinhados com o genoma mitocondrial da colônia NIH usando o programa Bowtie2 versão 2.2.6 (LANGMEAD; SALZBERG, 2012), isso permitiu que fossem filtrados somente os *reads* que mapearam com o genoma de referência. A princípio, foram criados os índices do mitogenoma NIH utilizando o comando *bowtie2-build*. O alinhamento feito foi do tipo *end-to-end*, o qual leva em consideração todos os nucleotídeos das sequências ao fazer a comparação. O *score-min* utilizado usa a função linear: $fx = -2 + -0,3 \times x$, onde x corresponde ao tamanho do *read*, que permite um erro de 5%. Os *reads* filtrados foram então alinhados usando o programa SPAdes (BANKEVICH *et al.*, 2012), um montador indicado para genomas, especialmente bacterianos, ele utiliza grafos de brujin para construir sequências consensos. Para a montagem dos *reads* brutos foi utilizado os parâmetros padrão deste programa. Os *contigs* gerados por esta ferramenta foram chamados de M2. Para a montagem do mitogenoma-UFRJ novamente houve uma seleção por posição. Foram selecionados manualmente apenas os *contigs* presentes em regiões deixados sem sobreposição por M1s. Logo, as sequências que foram selecionadas por essa abordagem foram chamadas M2s.

A terceira montagem M3 foi realizada com o Trinity, que é indicado para a montagem de transcriptomas. Este programa aplica três algoritmos em sequência (Inchworm, Chrysalis e Butterfly) para gerar os *contigs*, e assim como o SPAdes, utiliza grafos brujin para formar as sequências. O comando utilizado foi o padrão indicado pelo manual do programa. Novamente, os *contigs* foram selecionados por posição. Foram chamados de M3s os *contigs* localizados em regiões não cobertas por M1s e M2s.

As sequências geradas por essas três abordagens, M1s, M2s e M3s, foram analisadas e conciliadas manualmente, de acordo com a sua orientação, no programa BioEdit (HALL, 1999).

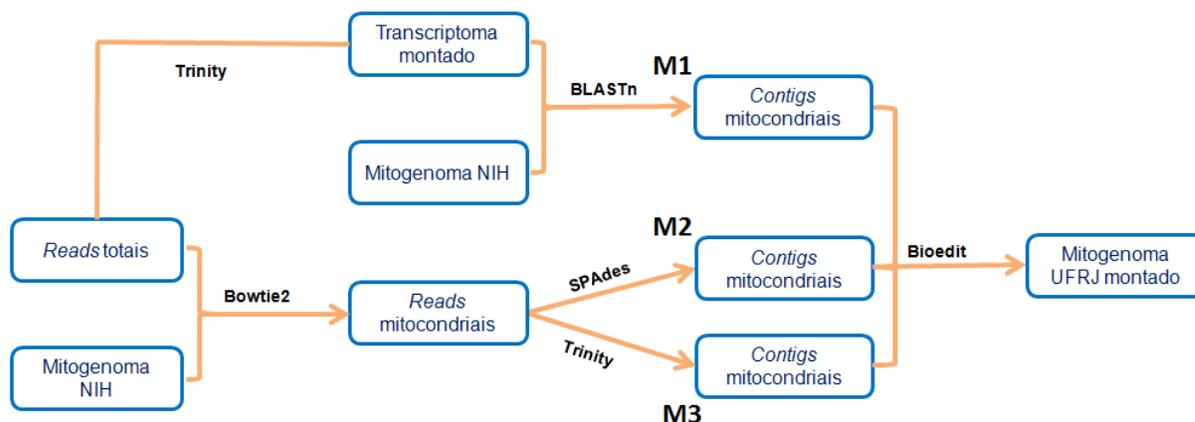


Figura 12 - Fluxograma usado para a montagem do genoma mitocondrial da colônia UFRJ. Dentro dos retângulos azuis estão indicados os arquivos de entrada de cada etapa, as setas estão acompanhadas dos programas usados.

Para saber o quanto do mitogenoma NIH cada estratégia conseguiu cobrir, os mito-*contigs* gerados pelas montagens M1, M2 e M3 foram alinhados separadamente com o mitogenoma NIH usando o programa Bowtie2 (LANGMEAD; SALZBERG, 2012). Novamente, o alinhamento foi do tipo *end-to-end* com *score-min* permitindo 5% de erro. Os arquivos gerados neste alinhamento, arquivos SAM, foram convertidos em um formato mais compacto o BAM, e posteriormente ordenados com o programa SAMTools versão 0.1.19 (LI *et al.*, 2009). SAMTools é uma ferramenta capaz de manipular arquivos SAM e BAM, e dentre outras funções, realiza conversão de formato, controle de qualidade, indexação e ordenação de arquivos. Isto possibilitou a visualização deste alinhamento de forma gráfica com o programa Ártemis versão 17.0.1 (CARVER *et al.*, 2011), evidenciando quais regiões do mitogenoma foram cobertas pelos *contigs* gerados por cada estratégia.

Para avaliar a montagem realizada, assim como, verificar a profundidade de sequenciamento de cada base no mitogenoma, os *reads* limpos do transcriptoma foram alinhados no mitogenoma montado (Figura 13). A ferramenta Bowtie2 (LANGMEAD; SALZBERG, 2012) foi utilizada para o alinhamento do tipo *end-to-end* e *score-min* permitindo 5% de erro. A ferramenta SAMTools foi usada mais uma vez para converter os arquivos de saída do alinhamento do formato SAM para o formato BAM e depois para SORTED. Este alinhamento foi então observado graficamente usando o programa Integrated Genome Viewer versão 2.6 (IGV) (JAMES *et al.*, 2017). Usando o IGV, a sequência do mitogenoma UFRJ foi checada manualmente para possíveis inconsistências na posição de um dado nucleotídeo. A sequência final do mitogenoma foi editada para conter em todas as posições o nucleotídeo mais frequente. Posteriormente, a cobertura dos *reads* limpos no mitogenoma UFRJ foi tabelado usando o software SAMTools (LI, 2011).

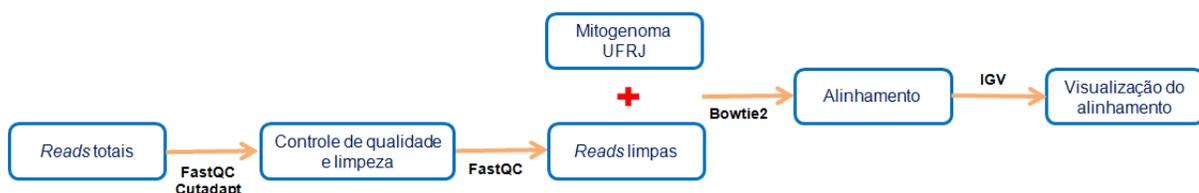


Figura 13 - Fluxograma usado para a visualização do genoma mitocondrial da colônia UFRJ. Dentro dos retângulos azuis estão indicados os arquivos de entrada de cada etapa, as setas estão acompanhadas dos programas usados. Após a visualização o mitogenoma foi editados para conter em todas as posições a base mais frequente.

3.2 Anotação e análise dos genomas mitocondriais

Após a montagem do genoma mitocondrial UFRJ, os mitogenomas das duas linhagens de *R. prolixus* foram anotados usando o software de anotação automática MITOS2 (BERNT *et al.*, 2013). Essa anotação foi verificada manualmente pela análise de resultados do BLASTX, que faz comparação entre uma sequência nucleotídica (mitogenoma) e um banco de dados proteico, e do tRNAscan (LOWE; CHAN, 2016), ferramenta utilizada para encontrar os tRNA's presentes numa dada sequência e que

também fornece suas estruturas secundárias. As repetições em tandem da região controle foram identificadas pelo programa Tandem Repeats Finder (BENSON, 1999).

A imagem circular do genoma foi feita usando o programa GenomeVX (CONANT; WOLFE, 2008). A composição de nucleotídeos e o uso relativo de códons sinônimos (RSCU do inglês - *Relative Synonymous Codon Usage*) foram feitas usando o programa CAIcal (PUIGBO *et al.*, 2008). RSCU é calculado pela razão da frequência observada e a frequência esperada de um códon. Por exemplo, dois códons codificam o aminoácido Tirosina (Y). Se o códon (UAU) for usado 100 vezes no genoma mitocondrial de um total de 120 tirosinas presentes o valor de RSCU para este códon será de $2 \cdot 100 / 120 = 1,67$. Os RSCU das duas colônias de *Rhodnius prolixus* e dos Triatomíneos da literatura (*T. dimidiata*, *T. infestans* e *T. rubrofasciata*) foram visualizados de maneira gráfica pelo programa CIMMiner (WEINSTEIN *et al.*, 1997).

Com o objetivo de encontrar domínios conservados na região controle, foi feito um alinhamento desse fragmento dos mitogenomas de *Rhodnius prolixus* e o dos três Triatomíneos da literatura (*T. dimidiata*, *T. infestans* e *T. rubrofasciata*) usando a ferramenta ClustalW versão 1.4 (THOMPSON *et al.* 1994) no programa Bioedit (HALL, 1999).

3.3 Busca de mitogenomas, alinhamento múltiplo e construção da árvore filogenética.

Uma busca foi realizada no NCBI a procura de mitogenomas dos Reduvidos já sequenciados. Foram encontrados 23 genomas completos de espécimes da família Reduviidae, as espécies e seus códigos de acessos estão organizados na (Tabela 3). *Apolygus lucorum* (HQ902161) e *Corythucha ciliados* (KC756280) foram escolhidos como grupo externo, pertencem respectivamente a família Miridae e Tingidae. As sequências codificantes (CDS) das 13 proteínas presentes no genoma mitocondrial dos genomas estudados nesta dissertação em conjunto com as desses insetos foram alinhadas pelo programa Muscle (EDGAR, 2004), usando a configuração padrão, e conferidas manualmente. O alinhamento foi feito por proteína, e posteriormente

concatenado usando o script catfasta2phyml (NYLANDER, 2015). A análise filogenética foi realizada usando a ferramenta RAxML (STAMATAKIS, 2014), baseada no método de máxima verossimilhança usando a matriz Jones Taylor Thornton (JTT; JONES *et al.*, 1992). A confiança dos nós da árvore foi testada por *Bootstrap* de 1000 réplicas. Após isso, a árvore foi então visualizada e editada usando o programa FigTree (RAMBAUT, 2012).

Tabela 3 - Espécies da família Reduviidae que já possuem o mitogenoma disponível, usados na análise de filogenia.

Família	Subfamília	Espécies	Acesso no GenBank	
Reduviidae	Ectrichodiinae	<i>Brontostoma colossus</i>	KM044501	
		<i>Agriosphodrus dohrni</i>	HM071001	
	Harpactorinae	<i>Scipinia horrida</i>	KY069967	
		<i>Rhynocoris incertis</i>	KY069966	
		<i>Velinus nodipes</i>	KY069961	
		Peiratinae	<i>Peirates arcuatus</i>	KF752445
			<i>Peirates atromaculatus</i>	KF913538
	<i>Peirates fulvescens</i>		KF913537	
	<i>Peirates lepturoides</i>		KF913541	
	<i>Peirates turpis</i>		KF913540	
	Reduviinae	<i>Sirthenea flavipes</i>	HQ645959	
		<i>Acanthaspis cincticrus</i>	KY069958	
		<i>Acanthaspis ruficeps</i>	KX505848	
		<i>Inara alboguttata</i>	KY069959	
		<i>Reduvius gregoryi</i>	KY069969	
		<i>Reduvius tenebrosus</i>	KC887529	
		<i>Tapeinus singularis</i>	KY069960	
	Salyavatinae	<i>Valentia hoffmanni</i>	FJ456952	
	Stenopodainae	<i>Canthesancus helluo</i>	KY069968	
		<i>Oncocephalus breviscutum</i>	KC887527	
	Triatominae	<i>Triatoma dimidiata</i>	AF301594	
<i>Triatoma infestans</i>		KY640305		
<i>Triatoma rubrofasciata</i>		MH934953		

3.4 Nível de expressão dos transcritos mitocondriais

Para verificar os níveis de expressão dos transcritos mitocondriais em cada condição do transcriptoma, os *reads* limpos do transcriptoma foram alinhados nos CDS do mitogenoma montado. O alinhamento foi feito usando novamente o programa Bowtie2 (LANGMEAD; SALZBERG, 2012), do tipo *end-to-end* e *score-min* permitindo 5% de erro. O programa Salmon versão 0.9.1 (PATRO *et al.*, 2017) foi usado para quantificar a abundância dos transcritos a partir do sequenciamento de RNA, foi usado os parâmetros padrão desta ferramenta. O Salmon faz a normalização dos *reads* em transcritos por milhão (TPM), que foi usado na análise de perfil de expressão.

Para a análise de perfil de expressão, uma tabela com as médias de TPM foi inserida no programa Expander (SHARAN; MARON-KATZ; SHAMIR, 2003). No Expander, foi usado a ferramenta *hierarchical clustering* para gerar os perfis de expressão dos genes mitocondriais. Esta ferramenta usa um algoritmo para calcular uma árvore de dendrogramas para todos os perfis de expressão.

3.5 Comparação do mitogenoma de *R. prolixus*, colônia UFRJ, com sequências da literatura para as espécies *R. prolixus* e *R. robustus*.

Com o intuito de elucidar a possível origem do mitogenoma de *R. prolixus* da colônia UFRJ foi realizada uma comparação com sequências nucleotídicas disponíveis na literatura. As sequências usadas foram referentes aos genes do citocromo b e da subunidade maior do RNA ribossomal de *R. prolixus* e de *R. robustus* (Tabelas 4, 5, 6 e 7). Essas sequências foram comparadas aos das colônias UFRJ através de um Blastn feito no próprio site do NCBI, nenhuma modificação foi feita na configuração padrão desta ferramenta.

Uma busca feita na subdivisão *Nucleotide* do site do NCBI usando os termos em inglês *Rhodnius prolixus* e *cytochrome b* revelou 36 sequências, elas estão listada na Tabela 4.

Tabela 4 - Lista das sequências de citocromo b (CYTB) de *Rhodnius prolixus* usadas neste estudo. A coluna “ID” contém o número de acesso das sequências do GenBank. A coluna “Região” informa o território onde o inseto foi coletado. A terceira coluna relata a referência bibliográfica.

ID	Região	Referência
AF045718	Colômbia, Bogotá	(Lyman <i>et al.</i> , 1999)
AF421339	Honduras, Francisco Morazan, Orica	(Monteiro <i>et al.</i> , 2003)
EF011726	Venezuela, Cojedes	
EF011723	Venezuela, Guárico, Ortiz	
EF011721	Venezuela, Guárico, San José Tiznados	
EF011716	Venezuela, Trujillo, Pampanito,	
EF043586	Guatemala, Tutucopote	
EF043579	Guatemala, Tituque	
EF043576	Guatemala, Las Palmas	
EF043588	Venezuela, Barinas	
EF043587	Venezuela, Barinas	
EF043585	Venezuela, Barinas	
EF043584	Venezuela, Barinas	
EF043583	Venezuela, Barinas	
EF043582	Venezuela, Portuguesa	
EF043581	Venezuela, Portuguesa	
EF043580	Venezuela, Barinas	
EF043578	Venezuela, Trujillo	
EF043577	Venezuela, Barinas	
KP126734	Colômbia, Casanare, Yopal	(Urbano <i>et al.</i> , 2015)
KP126733	Colômbia, Casanare, Yopal	
KP126732	Colômbia, Casanare, Yopal	
KP126731	Colômbia, Casanare, Yopal	
KP126730	Colômbia, Casanare, Yopal	
KP126729	Colômbia, Casanare, Yopal	
KP126728	Colômbia, Casanare, Yopal	
KP126727	Colômbia, Casanare, Yopal	
KP126726	Colômbia, Casanare, Yopal	
KP126725	Colômbia, Casanare, Yopal	
DQ118977	México, Chiapas	(Martínez <i>et al.</i> , 2006)
EU363822	Venezuela	(Harry <i>et al.</i> , 2008)
MH704764	Colômbia	(Nascimento <i>et al.</i> , 2019)
MH704763	Colômbia	
MH704762	Colômbia	
MH704761	Colômbia	
MH704760	Colômbia	

Uma nova busca na subdivisão *Nucleotide* do site do NCBI, dessa vez usando os termos em inglês *Rhodnius robustus* e *cytochrome b* revelou 75 sequências, que estão listada na Tabela 5.

Tabela 5 - Lista das sequências de citocromo b de *Rhodnius robustus* usadas neste estudo. A coluna “ID” contém o número de acesso das sequências do GenBank. A coluna “Região” informa o território onde o inseto foi coletado. A terceira coluna relata a referência bibliográfica.

ID	Região	Referência
AF421343	Brasil, Pará, Barcarena	(Monteiro <i>et al.</i> , 2003)
AF421342	Brasil, Pará, Itupiranga	
AF421341	Equador, Napo	
AF421340	Venezuela, Trujillo, Pampanito, La Concepcion	
FJ887793	Bolívia, Santa Cruz, El Torno	(Justi <i>et al.</i> , 2010)
EF071583	Brasil, Rondônia	(Maia da Silva <i>et al.</i> , 2007)
EF011728	Guiana Francesa, Roura, Bélizon	(Pavan; Monteiro, 2007)
EF011727	Brasil, Amazonas, Manaus	
EF011725	Tocantins, Tocantinópolis,	
EF011724	Brasil, Rondônia, Monte Negro	
EF011722	Brasil, Pará, Rio Mapuera	
EF011720	Brasil, Porto Velho	
EF011719	Venezuela, Trujillo, Candelaria	
EF011718	Venezuela, Trujillo, Candelaria	
EF011717	Venezuela, Portuguesa	
EF011715	Brasil, Pará	
EF011714	Brasil, Amazonas, Apuí	
EF011713	Brasil, Amazonas, Apuí	
EF011712	Guiana Francesa, Caiena	
EF011711	Brasil, Amazonas, Caruarí	
EF011710	Brasil, Amazonas, Caruarí	
EF011709	Brasil, Amazonas, Caruarí	
EF011708	Brasil, Amazonas, Balbina	
JX273158	Brasil, Amazonas, Novo Airão	
JN831970	Brasil, Amazonas, Médio Tajapós	(Quartier, et al., 2012)
JN831969	Brasil, Amazonas, Médio Tajapós	
JN831968	Brasil, Amazonas, Médio Tajapós	
JN831967	Brasil, Amazonas, Médio Tajapós	
JN831966	Brasil, Amazonas, Médio Tajapós	
JN831965	Brasil, Amazonas, Médio Tajapós	
JN831964	Brasil, Amazonas, Médio Tajapós	
JN831963	Brasil, Amazonas, Médio Tajapós	
JN831962	Brasil, Amazonas, Médio Tajapós	
JN831961	Brasil, Amazonas, Médio Tajapós	
JN831960	Brasil, Amazonas, Médio Tajapós	

Tabela 5 - Lista das sequências de CYTB de *Rhodnius robustus* usadas neste estudo. A coluna “ID” contém o número de acesso das sequências do GenBank. A coluna “Região” informa o território onde o inseto foi coletado. A terceira coluna relata a referência bibliográfica. (Continuação)

ID	Região	Referência
JN831960	Brasil, Amazonas, Médio Tajapós	(Quartier, et al., 2012)
JN831959	Brasil, Amazonas, Médio Tajapós	
JN831958	Brasil, Amazonas, Médio Tajapós	
JN831957	Brasil, Amazonas, Médio Tajapós	
JN831956	Brasil, Amazonas, Médio Tajapós	
JN831955	Brasil, Amazonas, Médio Tajapós	
JN831954	Brasil, Amazonas, Médio Tajapós	
JN831953	Brasil, Amazonas, Médio Tajapós	
EU363824	Guiana Francesa	(Harry <i>et al.</i> , 2008)
AF045717	Equador	(Lyman <i>et al.</i> , 1999)
KT805149	Bolívia, Ben, General José Ballivian, Yucumo	(Brenière <i>et al.</i> , 2017)
MF966306	Guiana Francesa, Saint Laurent du Maroni	(Barnabé <i>et al.</i> , 2018)
MF966305	Guiana Francesa, Saint Laurent du Maroni	
MF966304	Guiana Francesa, Rémire-Montjoly	
MF966303	Guiana Francesa, Saint Laurent du Maroni	
MF966302	Guiana Francesa, Saint Laurent du Maroni	
MF966301	Guiana Francesa, Matoury	
MF966300	Guiana Francesa, Matoury	
MF966299	Guiana Francesa, Matoury	
MF966298	Guiana Francesa, Saint Laurent du Maroni	(Barnabé <i>et al.</i> , 2018)
MF966297	Guiana Francesa, Mana	
MF966296	Guiana Francesa, Saint Laurent du Maroni	
MF966295	Guiana Francesa, Saint Laurent du Maroni	
MF966294	Guiana Francesa, Saint Laurent du Maroni	
MF966293	Guiana Francesa, Saint Laurent du Maroni	
MF966292	Guiana Francesa, Roura, Bélizon	
MF966291	Guiana Francesa, Caiena	
MF966290	Guiana Francesa, Caiena	
MF966289	Guiana Francesa, Matoury	
MF966288	Guiana Francesa, Antecum Pata	
MF966287	Guiana Francesa, Les Nouragues	
MF966286	Guiana Francesa, Saul	
MF966285	Guiana Francesa, Rémire-Montjoly	
MF966284	Guiana Francesa, Saint Laurent du Maroni	
MF966283	Guiana Francesa, Saint Laurent du Maroni	
MF966282	Guiana Francesa, Rémire-Montjoly	
MF966281	Guiana Francesa, Régina	
MF966280	Guiana Francesa, Rémire-Montjoly	
MF966279	Guiana Francesa, Roura, Bélizon	
MF966278	Guiana Francesa, Roura, Bélizon	
MF966277	Guiana Francesa, Roura, Bélizon	

A busca na subdivisão *Nucleotide* do site do NCBI agora usando os termos em

inglês *Rhodnius prolixus*, *ribosomal RNA* e *mitochondrial* revelou 13 sequências, destas apenas 8 eram mesmo mitocondriais e de *R. prolixus*, elas foram listadas na Tabela 6.

Tabela 6 - Lista das sequências do rRNAI de *Rhodnius prolixus* usadas neste estudo. A coluna "ID" contém o número de acesso das sequências do GenBank. A coluna "Região" informa o território onde o inseto foi coletado. A terceira coluna relata a referência bibliográfica. * indica a ausência desta informação no registro do banco de dados e na referência associada a ele. "LSHTM": Proveniente do insetário da London School of Hygiene and Tropical Medicine. "SNC": Proveniente do Servicio Nacional de Chagas, Córdoba, Argentina. "FMRP": Proveniente da Faculdade de Medicina Ribeirão Preto.

ID	Região	Referência
AF028747	*	(Stothard <i>et al.</i> , 1998)
AF045707	Colômbia, Bogotá	(Lyman <i>et al.</i> , 1999)
AF324519	*	(Garcia <i>et al.</i> , 2001)
EU822952	Colômbia	(Vendrami <i>et al.</i> , 2008)
EU822953	Colômbia	
EU822954	Colômbia	
EU827235	*	
EU827236	*	

A segunda busca para os genes do rRNAI na subdivisão *Nucleotide* do site do NCBI, essa vez usando os termos em inglês *Rhodnius robustus*, *ribosomal RNA* e *mitochondrial* apontou 33 sequências, destas 32 eram mesmo mitocondriais e de *R. prolixus*, elas foram listadas na Tabela 7.

Tabela 7 - Lista das sequências do rRNAI de *Rhodnius robustus* usadas neste estudo. A coluna "ID" contém o número de acesso das sequências do GenBank. A coluna "Região" informa o território onde o inseto foi coletado. A terceira coluna relata a referência bibliográfica.

ID	Região	Referência
AF045705	Equador	(Lyman <i>et al.</i> , 1999)
EU827206	Peru, Madre de Diós	(Vendrami <i>et al.</i> , 2008)
KT805173	Bolívia, Alto Beni, Yucumo	(Barnabé <i>et al.</i> , 2017)
MF966360	Guiana Francesa, Saint Laurent du Maroni	(Barnabé <i>et al.</i> , 2018)
MF966359	Guiana Francesa, Saint Laurent du Maroni	
MF966358	Guiana Francesa, Rémire-Montjoly	
MF966357	Guiana Francesa, Saint Laurent du Maroni	
MF966356	Guiana Francesa, Matoury	
MF966355	Guiana Francesa, Matoury	
MF966354	Guiana Francesa, Matoury	
MF966353	Guiana Francesa, Matoury	

Tabela 7 - Lista das sequências do rRNA de *Rhodnius robustus* usadas neste estudo. A coluna “ID” contém o número de acesso das sequências do GenBank. A coluna “Região” informa o território onde o inseto foi coletado. A terceira coluna relata a referência bibliográfica. (Continuação)

ID	Região	Referência
MF966352	Guiana Francesa, Saint Laurent du Maroni	(Barnabé et al., 2018)
MF966351	Guiana Francesa, Saint Laurent du Maroni	
MF966350	Guiana Francesa, Saint Laurent du Maroni	
MF966349	Guiana Francesa, Saint Laurent du Maroni	
MF966348	Guiana Francesa, Saint Laurent du Maroni	
MF966347	Guiana Francesa, Caiena	
MF966346	Guiana Francesa, Caiena	
MF966345	Guiana Francesa, Matoury	
MF966344	Guiana Francesa, Rémire-Montjoly	
MF966343	Guiana Francesa, Saint Laurent du Maroni	
MF966342	Guiana Francesa, Saint Laurent du Maroni	
MF966341	Guiana Francesa, Saint Laurent du Maroni	
MF966340	Guiana Francesa, Rémire-Montjoly	
MF966339	Guiana Francesa, Rémire-Montjoly	
MF966338	Guiana Francesa, Camopi	
MF966337	Guiana Francesa, Régina	
MF966336	Guiana Francesa, Rémire-Montjoly	
MF966335	Guiana Francesa, Roura, Bélizon	
MF966334	Guiana Francesa, Roura, Bélizon	
MF966333	Guiana Francesa, Roura, Bélizon	
MF966332	Guiana Francesa, Roura, Bélizon	

Após isso, foi feito um alinhamento no programa Bioedit (HALL, 1999) usando a ferramenta ClustalW (THOMPSON *et al.*, 1994) e todas as sequências de CYTB de *R. prolixus* (Literatura, UFRJ e NIH). Um segundo alinhamento continha todas as sequências CYTB de *R. robustus* mais as duas de *R. prolixus* (UFRJ e NIH). O mesmo foi feito para as sequências da subunidade maior do RNA ribossomal.

Em seguida, foi construída uma árvore filogenética com base nas sequências do gene do citocromo b (CYTB). Como grupo externo foram usadas as sequências do citocromo b de *Rhodnius pallescens* (FJ229359, JQ686675 e JQ686685), de *Rhodnius pictipes* (MF497699) e de *Rhodnius barreti* (JX273159 e JX273161). Sendo assim, todas as sequências do gene citocromo b foram alinhadas pelo programa Muscle

(EDGAR, 2004), usando configuração padrão, e conferidas manualmente. A análise filogenética foi realizada usando a ferramenta T-REX (BOC *et al.*, 2012), baseada no método de Neighbor-Joining (NJ) usando como modelo de distância 2-parâmetros de Kimura (KIMURA, 1980), que leva em consideração que transições e transversões ocorrem com frequências diferentes. A confiança dos nós da árvore foi testada por *Bootstrap* de 1000 réplicas. Após isso, a árvore foi então visualizada e editada usando o programa ITol versão 4.4.2 (LETUNIC; BORK, 2016).

3.5 Descoberta e caracterização de polimorfismos de nucleotídeo único, inserções e deleções.

A busca por polimorfismos de nucleotídeo único (SNP - do inglês *Single Nucleotide Polymorphism*), e inserções e deleções (Indels) foi feita para o mitogenoma de *Rhodnius Prolixus* colônia UFRJ. Para a descoberta dos SNPs e Indels desta colônia foi usado o alinhamento do mitogenoma montado e os *reads* limpos de cada biblioteca, já detalhado no tópico 3.1.

A identificação de variantes utilizou o programa Genome Analysis Toolkit (GATK) (VAN DER AUWERA *et al.*, 2013) e seguiu seu protocolo padrão (<https://software.broadinstitute.org/gatk/best-practices/workflow?id=11164>). Como pré-processamento o pacote Picard foi utilizado para preparar o arquivo de alinhamento para a chamada de variantes feita pelo GATK. A primeira ferramenta do Picard utilizada foi a MarkDuplicates, que identifica e marca *reads* duplicatas possivelmente presentes na amostra. As duplicatas surgem de produtos de PCR múltiplos da mesma molécula modelo que se ligam à célula de fluxo, ou seja, são sequências originárias de um único fragmento de DNA. Também são considerados *reads* duplicadas quando um único *cluster* de amplificação é detectado incorretamente como múltiplos *clusters* pelo sensor óptico do instrumento de sequenciamento, estes são chamados de duplicatas óticas. A persistência deste artefato geraria problemas na especificidade da busca por variantes, pois um erro de sequenciamento poderia ser adicionado na análise se fosse lido mais de uma vez estando duplicado, aumentando o número de variantes falsos positivos.

Esta situação foi atenuada marcando os *reads* duplicados para evitar a interferência deste artefato nos dados. Em seguida, a ferramenta *AddOrReplaceReadGroups* foi usada para inserir dados de identificação nos arquivos de alinhamento. Após isso foi usado o *CreateSequenceDictionary*, onde é criado um arquivo que funciona como um dicionário de sequência a partir de uma sequência de referência fornecida no formato FASTA. O passo *ReorderSam* ordena um arquivo BAM em relação a uma referência fornecida, utilizando a indexação gerada na etapa anterior.

Após isso, a busca por variantes foi iniciada pelo programa GATK, o algoritmo *HaplotypeCaller* chamou simultaneamente *SNPs* e *Indels*, que foram armazenados em um arquivo em formato VCF. A ferramenta *VariantFiltration* foi então utilizada para selecionar somente os *SNPs* e *Indels* que passassem por um filtro de alta qualidade baseado nos critérios indicados pelo protocolo do GATK.

Após isso a ferramenta VCFtools (DANECEK *et al.*, 2011) foi empregada para extrair informações sobre os arquivos VCFs, como selecionar dentre as variantes que passaram pelo filtro de qualidade, somente as presentes em todas as amostras. O VCFtools também calculou o número de *SNPs* e *Indels* que passaram pelo filtro e a taxa de transição/transversão. O programa SIFT (*Sorting Intolerant From Tolerant*) (SIM *et al.*, 2012), foi usado para verificar o efeito dos *SNPs* encontrados em genes codificadores de proteínas. O SIFT gera um *score* para trocas não sinônimas, levando em consideração a posição e o aminoácido alterado, e dados organismos próximos disponíveis publicamente. O SIFT infere se a substituição irá afetar a função da proteína dado a conservação dos aminoácidos nos organismos. Um SIFT *score* acima de 0,05 indica que a mutação seria tolerável para a proteína, enquanto um *score* menor ou igual a 0,05 é classificado como não tolerável, pois a troca de aminoácidos pode influenciar no seu desempenho.

4. RESULTADOS

4.1 Montagem do Genoma Mitocondrial de *Rhodnius Prolixus*, colônia UFRJ

A primeira etapa desenvolvida para a montagem do mitogenoma UFRJ consistiu na limpeza das bibliotecas do transcriptoma. A avaliação com o FastQC mostrou que era necessário remover sequências de baixa qualidade e de adaptadores. Isto pois alguns *reads* possuíam qualidade abaixo do ideal (valor de phred => 30), e ainda percebe-se a presença do adaptador *illumina universal* na região terminal dos *reads*, como pode ser visto na Figura 14A, indicado pela seta.

Após a etapa de limpeza com o Cutadapt, dos 403.835.894 *reads* totais nos transcriptomas, restaram 363.200.709, que representam 89,9% (Tabela 8).

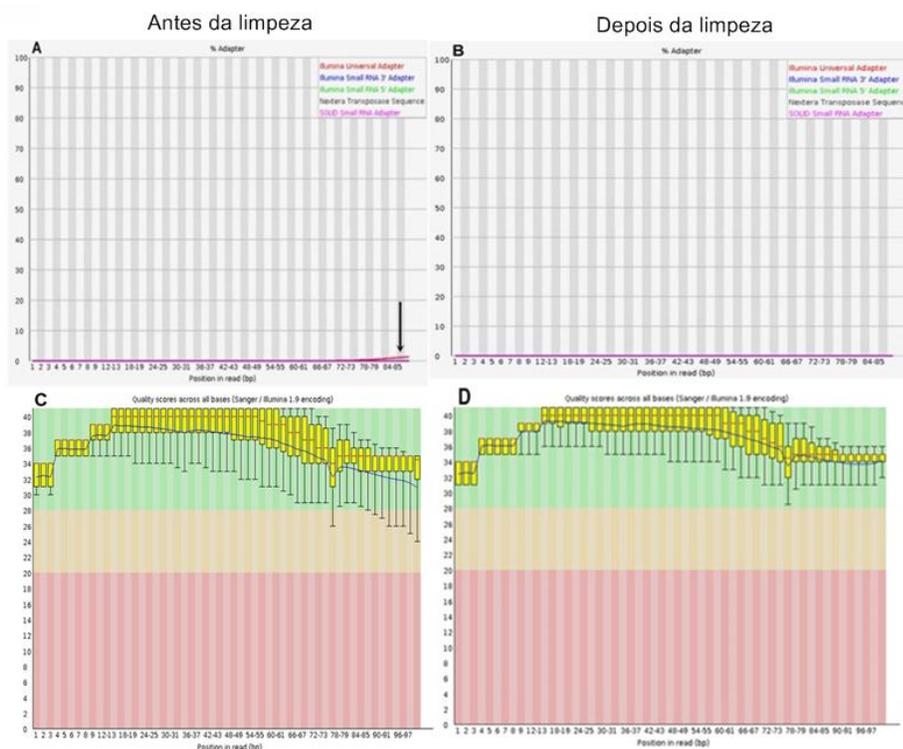


Figura 14 - Exemplo de controle de qualidade antes e após limpeza dos *reads* utilizados. A biblioteca da primeira replicata da condição jejum foi analisada antes (A e C) da etapa de limpeza e depois (B e D). Os painéis A e B mostram a busca de adaptadores, sugerindo a presença do adaptador universal *illumina* (seta, painel A). Os painéis C e D mostram a análise de qualidade das bases dessa biblioteca ao longo do read, após a limpeza (D) pode-se verificar a melhora da qualidade principalmente no final do *read*.

Tabela 8 - Dados das bibliotecas após a etapa de limpeza feita com o programa Cutadapt. A primeira coluna indica o número da amostra. A segunda coluna informa as condições usadas para as bibliotecas, e a terceira o dia que as amostras foram analisadas após a alimentação. Na quarta coluna está representado o total de *reads* limpos para cada replicata. Na quinta coluna foi inserido o comprimento máximo e mínimo dos *reads* após a limpeza. Na sexta coluna se encontra a porcentagem de AT de cada replicata após a limpeza. E por fim, na última coluna é mostrado a porcentagem de *reads* que permaneceram após a limpeza em relação aos *reads* brutos.

Amostra	Condição	Dia	Total reads	Tamanho do read	%AT	Aproveitamento %
J	Jejum	-	5523098	90-101	59	88,40%
			5936519	90-101	59	88,30%
			5675094	90-101	61	89,30%
S1	Sangue	1	6564395	90-101	62	90,90%
			6549892	90-101	62	90,60%
			6315560	90-101	62	90,50%
S2		4	6163174	90-101	62	90,20%
			5762913	90-101	62	90,50%
			6204662	90-101	62	89,70%
S3		8	5415458	90-101	62	89,00%
			6101262	90-101	62	88,60%
			5790747	90-101	62	89,30%
S4		12	5917787	90-101	63	90,20%
			5554467	90-101	63	90,00%
			5953778	90-101	62	89,80%
E1	Epimastigota 10 ³	1	6607270	90-101	62	89,60%
			5225737	90-101	62	89,90%
			5745676	90-101	62	89,60%
E2		4	6007055	90-101	62	89,80%
			6308220	90-101	62	90,10%
			5532546	90-101	63	89,60%
E3		8	5018172	90-101	63	90,30%
			5032089	90-101	63	90,20%
			5374771	90-101	63	90,70%
E4		12	5464463	90-101	63	90,20%
			5980738	90-101	63	90,60%
			5302799	90-101	62	90,20%
E5	Epimastigota 10 ⁷	1	5419580	90-101	62	90,20%
			5941188	90-101	62	90,10%
			5670715	90-101	62	89,40%
E6		4	5265627	90-101	62	90,80%
			5717044	90-101	62	89,90%
			5307645	90-101	62	90,30%

Tabela 8 – Dados das bibliotecas após a etapa de limpeza feita com o programa Cutadapt. A primeira coluna indica o número da amostra. A segunda coluna informa as condições usadas para as bibliotecas, e a terceira o dia que as amostras foram analisadas após a alimentação. Na quarta coluna está representado o total de *reads* limpos para cada replicata. Na quinta coluna foi inserido o comprimento máximo e mínimo dos *reads* após a limpeza. Na sexta coluna se encontra a porcentagem de AT de cada replicata após a limpeza. E por fim, na última coluna é mostrado a porcentagem de *reads* que permaneceram após a limpeza em relação aos *reads* brutos. (Continuação).

Amostra	Condição	Dia	Total reads	Tamanho do read	%AT	Aproveitamento %	
E7	Epimastigota 10 ⁷	8	6171904	90-101	62	89,30%	
			5915994	90-101	62	89,70%	
			6909053	90-101	62	89,60%	
E8		12	6567915	90-101	63	90,50%	
			5747934	90-101	64	88,60%	
			5824881	90-101	63	90,10%	
T1	Tripomastigota 10 ³	1	5679802	90-101	63	88,70%	
5901890			90-101	62	89,90%		
5713318			90-101	62	89,60%		
T2		4	6130267	90-101	63	90,20%	
			5616659	90-101	63	90,20%	
			5751894	90-101	62	89,80%	
T3		8	5858270	90-101	63	90,40%	
			5074105	90-101	63	89,90%	
			6885627	90-101	62	89,90%	
T4		12	6171795	90-101	62	90,50%	
			6077195	90-101	62	90,40%	
			5706089	90-101	63	90,60%	
T5		Tripomastigote 10 ⁷	1	5346356	90-101	63	90,70%
				5353878	90-101	63	90,30%
				4708276	90-101	62	90,10%
T6	4		5659195	90-101	63	90,00%	
			5405581	90-101	62	90,60%	
			5012681	90-101	62	90,40%	
T7	8		5605480	90-101	63	90,30%	
			5013574	90-101	63	90,20%	
			6607344	90-101	63	90,70%	
T8	12		5944710	90-101	63	90,20%	
			5325575	90-101	63	90,60%	
			5163326	90-101	62	90,20%	
Total:			363200709	90-101	62	89,93%	

A montagem *de novo* do transcriptoma com o programa Trinity (denominada M1) gerou 137.440 *contigs*, com valor N50 de 20.558. O comprimento médio dos *contigs* foi de 749 nucleotídeos, entretanto, o maior *contig* gerado foi de 18.756 pb e o menor foi de 201 pb.

Foram identificados 47 *contigs* mitocondriais presentes no transcriptoma-UFRJ montado, que cobriu 70,6% do mitoDNA-NIH. O maior mito-*contig* identificado nessa abordagem foi de 3.321 pb, ao passo que o menor de 207 pb. Como no caso de sobreposição completa de dois *contigs* o menor foi descartado, restaram 18 *contigs*, esse conjunto foi chamado de montagem M1s (Figura 15).

Os *contigs* M1s não cobriram sete regiões do mitogenoma-NIH. As lacunas têm respectivamente, 134, 1.496, 1.520, 1.211, 135, 218 e 119 nucleotídeos. Dessa forma, outras montagens, M2 e M3, foram implementadas para cobrir toda região do mitogenoma-NIH.

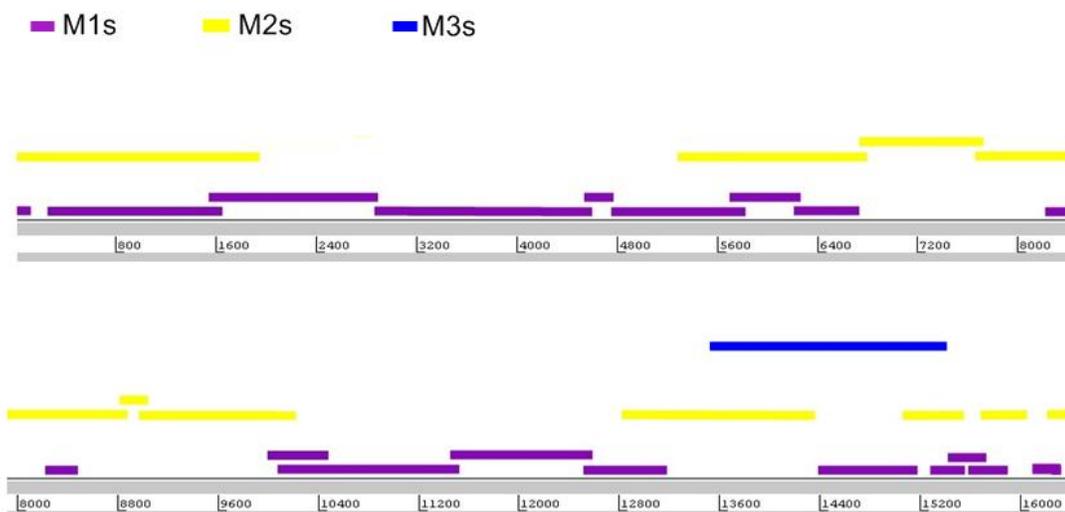


Figura 15 - Contigs usados para a montagem do mitogenoma UFRJ. O eixo X representa o mitogenoma-NIH. Os *contigs* da montagem M1s estão representados em roxo, da M2s em amarelo e da M3s em azul.

Do número total de *reads* presentes no transcriptoma da colônia UFRJ (403.835.894), mais de 78 milhões (78.019.631) foram identificados como *mitoreads* devido ao seu alinhamento com o mitogenoma-NIH. Estes *reads* foram oriundos de

todas as bibliotecas presentes no transcriptoma e utilizados para as montagens posteriores do mitogenoma-UFRJ.

A montagem *de novo* feita pelo programa SPAdes (chamada M2) gerou 1.115 *contigs*, que cobriram 88,9% do modelo. O maior *contig* gerado teve 2.101 pb, o menor contou com apenas 56 pb. Foram selecionados 11 *contigs* desta montagem, chamados de M2s (Figura 15) que estavam compreendidos nas lacunas deixadas por M1s. Com isso mais 29,37% do mitogenoma-NIH foi coberto.

Já a montagem *de novo* feita pelo Trinity gerou 211 *contigs*, que cobriram 82,93%. O maior *contig* gerado teve 1.581 nucleotídeos, e o menor 201 nucleotídeos. Novamente os *contigs* foram organizados por posição, e nesta montagem foi buscado *contigs* que tivessem alinhado na lacuna deixada por M1s e M2s. Apenas um *contig* foi selecionado e chamado de M3s (Figura 15). Com o uso do *contig* M3s o mitogenoma-NIH foi 100% coberto.

A montagem completa do genoma mitocondrial da colônia UFRJ foi possível pela sobreposição dos *contigs* gerados por essas três ferramentas. Ela foi realizada manualmente no programa Bioedit (HALL, 1999) e direcionada para a conferência final de nucleotídeos.

O alinhamento do genoma mitocondrial da colônia UFRJ com os *reads* limpos do transcriptoma possibilitou uma avaliação da montagem realizada. Dos *reads* do transcriptoma que passaram pelos filtros usados na etapa de limpeza (remoção de adaptador e *phred* 30), mais de 77 milhões (77.314.259) alinharam com o mitogenoma montado, o que corresponde a 19,14% do total de *reads* do transcriptoma (Tabela 9).

Tabela 9 - Quantidade de reads após alinhamento com o mitogenoma montado. A primeira coluna indica o nome da amostra. A segunda coluna informa as condições usadas para as bibliotecas, e a terceira o dia que as amostras foram analisadas após a alimentação. Na quarta coluna está representado o total de reads limpos para cada replicata. Na quinta coluna foi inserido o valor de reads que alinharam com o mitogenoma montado. E por fim, na última está a porcentagem de aproveitamento dos reads.

Amostra	Condição	Dia	Total reads	Reads alinhados	Aproveitamento %
J	Jejum	-	5523098	3158939	18,44%
			5936519		
			5675094		
S1	Sangue	1	6564395	3483557	17,93%
			6549892		
			6315560		
S2		4	6163174	3490528	19,25%
			5762913		
			6204662		
S3		8	5415458	4038845	23,34%
			6101262		
			5790747		
S4		12	5917787	3860875	22,16%
			5554467		
			5953778		
E1	Epimastigota 10 ³	1	6607270	3270337	18,60%
			5225737		
			5745676		
E2		4	6007055	3515251	19,70%
			6308220		
			5532546		
E3		8	5018172	3955522	25,64%
			5032089		
			5374771		
E4		12	5464463	4121246	24,61%
			5980738		
			5302799		
E5	Epimastigota 10 ⁷	1	5419580	2779660	16,32%
			5941188		
			5670715		
E6		4	5265627	2892833	17,75%
			5717044		
			5307645		
E7		8	6171904	4310757	22,69%
			5915994		
			6909053		
E8		12	6567915	5172490	28,51%
			5747934		
			5824881		

Tabela 910 - Quantidade de reads após alinhamento com o mitogenoma montado. A primeira coluna indica o número da amostra. A segunda coluna informa as condições usadas para as biblioteca, e a terceira o dia que as amostras foram analisadas após a alimentação. Na quarta coluna está representado o total de reads limpos para cada replicata. Na quinta coluna foi inserido o valor de reads que alinharam com o mitogenoma montado. E por fim, na última está a porcentagem de aproveitamento dos reads. (Continuação).

Amostra	Condição	Dia	Total reads	Reads alinhados	Aproveitamento %
T1	Tripomastigota 10 ³	1	5679802	3403376	19,68%
			5901890		
			5713318		
T2		4	6130267	3456835	19,75%
			5616659		
			5751894		
T3		8	5858270	4125780	23,16%
			5074105		
	6885627				
T4	12	6171795	4950750	27,57%	
		6077195			
		5706089			
T5	Tripomastigote 10 ⁷	1	5346356	3208095	20,82%
			5353878		
			4708276		
T6		4	5659195	3372861	20,98%
			5405581		
			5012681		
T7		8	5605480	3372861	20,98%
			5013574		
	6607344				
T8	12	5944710	3372861	20,98%	
		5325575			
		5163326			
Total:			363200709	77314259	19,14%

Usando o programa IGV, a sequência do mitogenoma UFRJ foi checada para possíveis inconsistências na presença de nucleotídeos pouco frequentes. De modo que a sequência final foi editada para conter em todas as posições o nucleotídeo mais frequente (Figura 16).



Figura 16 - Exemplo da visualização do alinhamento dos *reads* limpos do transcriptoma com o mitogenoma-UFRJ montado. No eixo X está a sequência de nucleotídeos que compõem o mitogenoma UFRJ. Os blocos coloridos mostram que ao menos dois nucleotídeos são bem representados para dada posição. Na posição 842 do mitogenoma-UFRJ, o nucleotídeo citosina está presente em 13.763 *reads*, enquanto a timina está em 11.084 *reads*.

O genoma mitocondrial completo de *R. prolixus* somou 16.454 pb. No genoma mitocondrial da colônia da UFRJ apresenta um conteúdo de A e T de 69,66%, contendo 41,52% Adenina (A), 28,14% Timina (T), 19,48% Citosina (C) e 10,85% é Guanina (G). Já o mitogenoma-NIH a composição A e T foi ligeiramente menor de 69,63%, a constituição do genoma é de 41,42% é A, 28,21% é T, 19,46% C e 10,91% é G.

Além disso, o alinhamento do mitogenoma-UFRJ com os *reads* limpos mostrou que estes cobriam completamente o mitoDNA (Figura 17). As regiões onde estão localizados os genes codificadores de proteínas foram os que tiveram maior cobertura, em contrapartida, as menores foram nos genes que codificam os RNA's transportadores.

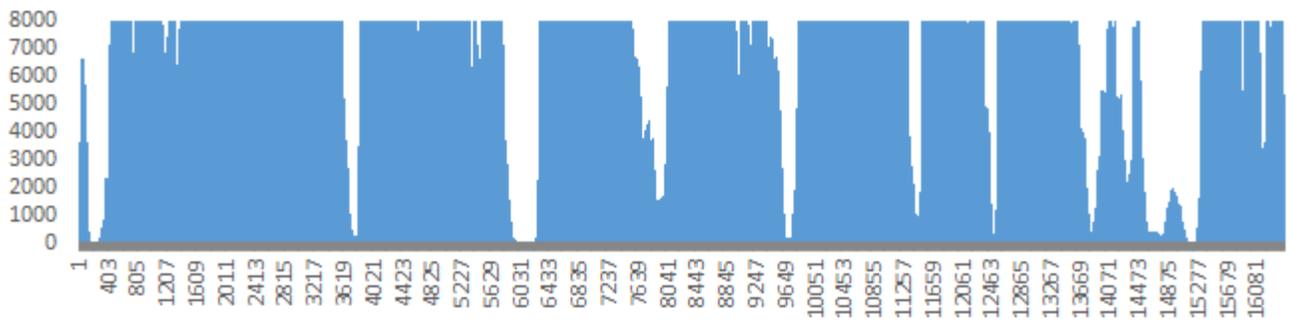


Figura 17 - Cobertura do genoma mitocondrial colônia UFRJ. Esta cobertura é referente ao alinhamento dos *reads* limpos do transcriptoma com o mitogenoma-UFRJ montado. As regiões com pouca cobertura foram identificadas como sendo de RNAs transportadores. O eixo x representa o mitogenoma-UFRJ. O eixo y representa número de *reads*. Nesta imagem o eixo y foi limitado a 8.000 reads para melhor visualização.

4.2 Anotação e análise dos genomas mitocondriais de *Rhodnius Prolixus*

O mitogenoma completo de *R. prolixus* contém 13 genes codificadores de proteínas, sendo eles, sete subunidades do complexo NADH desidrogenase (NAD1-6 e NAD4L), três subunidades da citocromo c oxidase (COX1-3), duas do complexo ATP sintase (ATP6 e 8) e o gene do citocromo b (CYTB). Além disso, também inclui dois genes para RNA ribossômico (rRNAI e rRNAs) e 22 para RNA transportadores, que são importantes para a tradução das proteínas mitocondriais. As informações a respeito da 1) disposição dos genes, assim como 2) o códon de início e 3) de parada dos genes codificadores de proteínas, e 4) o anticódon dos tRNAs foram inseridas na Tabela 10. Todos os genes, em conjunto com a região de controle, localizada após o gene que codifica a subunidade menor do rRNA estão ilustrados na Figura 18.

Tabela 10 - Anotação do genoma mitocondrial UFRJ de *Rhodnius prolixus*. A primeira coluna diz respeito ao gene identificado, as abreviações do tRNAs seguem o código de letras IUPAC-IUB. A coluna "Fita" evidencia se o gene está localização na fita *majority* (J) ou *minority* (N). A coluna "Códons" mostra os códons de início e de parada para os genes codificadores de proteínas somente. A coluna "Anticódon" mostra o anticódon e sua localização somente para os genes de RNAs transportadores.

Gene	Posição	Fita	Tamanho (pb)	Códons (início/parada)	Anticódon
tRNAQ	59-127	N	69	-	TTG, 31-33
tRNAM	127-194	J	68	-	CAT, 31-33
NAD2	195-1190	J	996	ATT/T(aa)	-
tRNAW	1189-1254	J	66	-	TCA, 31-33
tRNAC	1247-1308	N	62	-	GCA, 32-34
tRNAY	1309-1374	N	66	-	GTA, 32-34
COX1	1377-2912	J	1536	ATG/TAG	-
tRNAL2	2913-2979	J	67	-	TAA, 31-33
COX2	2980-3657	J	678	ATT/TAA	-
tRNAK	3664-3734	J	71	-	CTT, 32-34
tRNAD	3734-3798	J	65	-	GTC, 32-34
ATP8	3800-3961	J	162	ATG/TAA	-
ATP6	3955-4614	J	660	ATG/TAG	-
COX3	4626-5411	J	786	ATG/TA(a)	-
tRNAG	5411-5473	J	63	-	TCC, 30-32
NAD3	5474-5827	J	354	ATT/TAA	-
tRNAA	5837-5898	J	62	-	TGC, 30-32
tRNAR	5898-5962	J	65	-	TCG, 29-31
tRNAN	5968-6034	J	67	-	GTT, 31-33
tRNAS1	6034-6101	J	68	-	GCT, 26-28
tRNAE	6103-6167	J	65	-	TTC, 31-33
tRNAF	6170-6236	N	67	-	GAA, 33-35
NAD5	6237-7937	N	1700	ATT/TAA	-
tRNAJ	7938-8001	N	64	-	GTG, 32-34
NAD4	8002-9333	N	1332	ATG/TAG	-
NAD4I	9327-9611	N	284	ATA/TAA	-
tRNAT	9623-9686	J	64	-	TGT, 32-34
tRNAP	9687-9751	N	65	-	TGG, 31-33
NAD6	9730-10248	J	519	ATT/TAA	-
CYTB	10248-11381	J	1134	ATG/TAA	-
tRNAS2	11385-11453	J	69	-	TGA, 31-33
NAD1	11531-12457	N	927	ATA/TAA	-
tRNAL1	12464-12528	N	65	-	TAG, 30-32
rRNAL	12529-13785	N	1257	-	-
tRNAV	13786-13855	N	70	-	TAC, 32-34
rRNAS	13856-14658	N	803	-	-
tRNAI	16449-61	J	67	-	GAT, 32-34

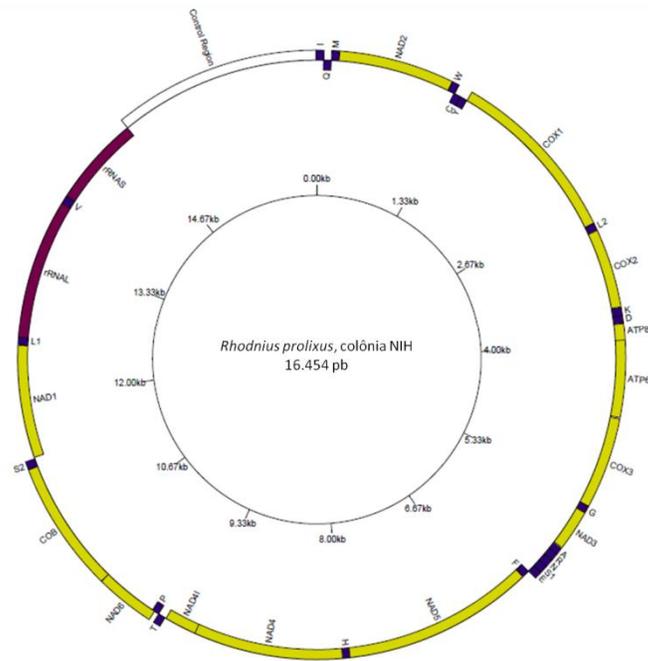
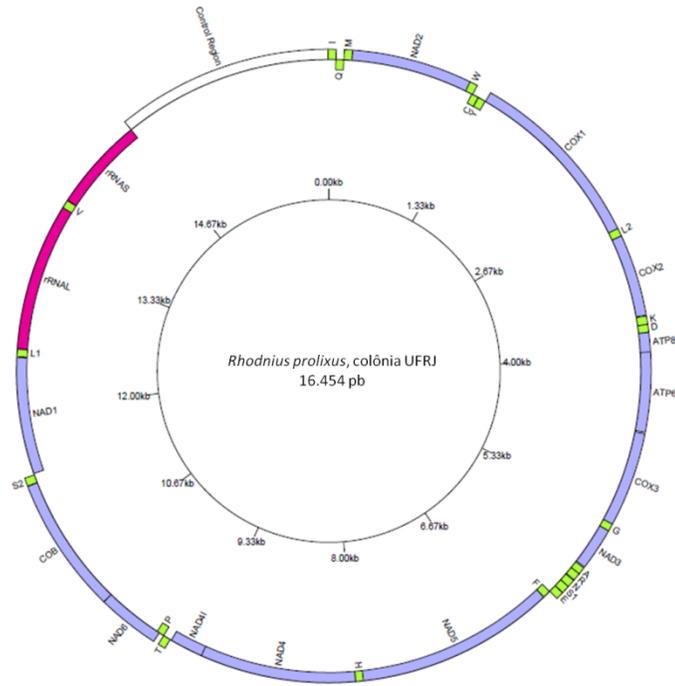


Figura 18 - Esquema circular do genoma mitocondrial de *Rhodnius prolixus*. Na parte superior está o mitogenoma da colônia UFRJ. Os genes localizados na fita *minority* estão representados voltados para o interior do mitogenoma e os da fita *majority* para o exterior. Os genes codificadores de proteína estão marcados em azul, os tRNAs em verde e os RNAs ribossomais em rosa. A região controle está em branco. Os tRNA's seguem o código de uma letra IUPAC-IUB. Na parte inferior se encontra o mitogenoma da colônia NIH. Os genes e tRNAs estão dispostos da mesma maneira. Entretanto, nesta colônia os genes codificadores de proteína estão marcados em amarelo, os tRNA estão em preto e os RNA's ribossomais em roxo. A região controle está em branco.

4.2.1 Genes codificadores de proteínas

Nos genes codificadores de proteínas dos mitogenomas de invertebrados, os códons de início canônicos são ATN, GTG e TTG (CAMERON *et al.*, 2012). No genoma de *R. prolixus* todos os 13 genes codificadores de proteínas (PCGs) usam códons de início canônico, mas apenas ATN foi encontrado. A maior parte das sequências de codificação foram iniciadas com o códon ATG, isso ocorreu em COXI, ATP6, ATP8, COXIII, NAD4, E CYTB. Já NAD2, COXII, NAD3 e NAD6 foram iniciados por ATT, NAD1 e NAD4I por ATA e NAD5 por ATC. Em relação aos códons de parada, foram encontrados dois tipos: TAA (presente em ATP8, COXII, COXIII, CYTB, NAD1, NAD2, NAD3, NAD5, NAD4I e NAD6), TAG (em ATP6, COXI e NAD4). Neste quesito não houve diferença entre os mitogenomas das colônias.

No mitogenoma da colônia UFRJ os quatro códons mais usados nos PCGs foram, ATT (Ile, 7,75%), TTA (Leu, 6,37%), TTT (Fen, 6,26%), e ATA (Met, 5,26%). Em decorrência disso, os aminoácidos mais comuns foram leucina (14,55%), isoleucina (10,27%), fenilalanina (8,89%) e metionina (6,94%). Esse padrão se manteve na colônia NIH, onde os códons mais usados foram ATT (Ile, 7,70%), TTA (Leu, 6,50%), TTT (Fen, 6,26%), e ATA (Met, 5,37%). Sendo assim, os aminoácidos mais presentes nas proteínas mitocondriais foram leucina (14,53%), isoleucina (10,27%), fenilalanina (8,89%) e metionina (7%).

Ao considerar os valores do uso relativo de códons sinônimos, ficou notável um uso assimétrico dos códons (Figura 19).

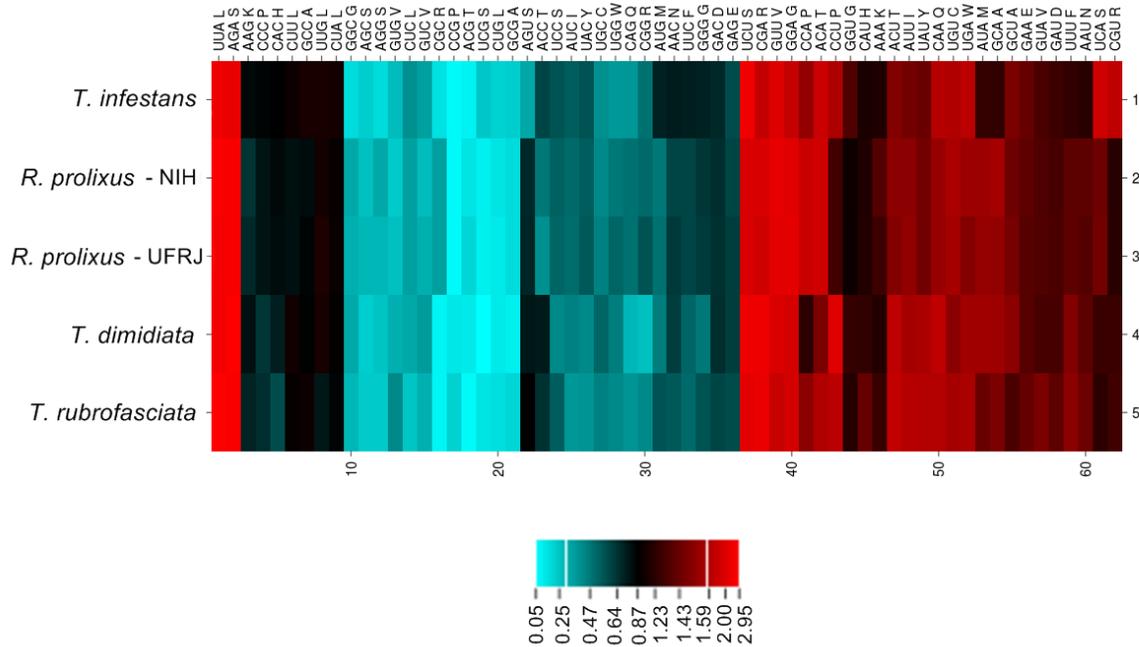


Figura 19 - Mapa dos valores de RSCU para *R. prolixus*, *Triatoma infestans*, *T. dimidiata* e *T. rubrofasciata*. Na parte superior estão listados os códons e os respectivos aminoácidos codificados por eles. No eixo Y estão listadas as espécies - *Triatoma infestans*, *R. prolixus* NIH, *R. prolixus* UFRJ, *T. dimidiata* e *T. rubrofasciata*. A preferência pelo uso dos códons fica evidenciada pelo uso das cores. Códons com os maiores valores de RSCU aparecem em cores de preto a vermelho.

Como pode ser visto na Figura 19, em *Rhodnius prolixus* houve uma inclinação para uso dos códons ricos em AT (eles apresentam os valores mais altos de RSCU). De modo geral, os triatomíneos considerados apresentaram um padrão de uso de códons relativamente parecido. Ainda há uma preferência por códons que terminem com adenina ou timina (Tabela 11). Nota-se que os códons com os valores de RSCU mais baixos foram os que possuem uma citosina e/ou guanina na segunda e terceira posição.

Tabela 11 - Composição dos códons dos genes codificadores de proteínas em diferentes posições para os mitogenomas das duas linhagens de *R. prolixus*. As porcentagens foram feitas levando em consideração o número total de códons (3690).

Mitogenoma NIH			
	1º posição	2º posição	3º posição
G	19,70%	15,80%	9,40%
A	32,60%	19,20%	37,10%
T	33,10%	46,20%	38,80%
C	14,60%	18,80%	14,70%

Mitogenoma UFRJ			
	1º posição	2º posição	3º posição
G	19,81%	15,74%	9,89%
A	32,55%	19,21%	36,80%
T	33,03%	46,26%	38,64%
C	14,61%	18,78%	14,66%

4.2.2 RNAs não codificantes

Os 22 RNA's transportadores típicos dos mitogenomas de invertebrados foram encontrados para as duas linhagens de *R. prolixus*. O tamanho deles varia entre 62 e 71 pb de comprimento. Todos os genes de tRNA tem sequência adequada e podem formar estrutura secundária em forma de trevo (característica), como pode ser visto na Figura 20. A sequência de todos os tRNAs foi idêntica para as duas linhagens.

O gene da subunidade maior do RNA ribossomal (rRNAI - *large subunit ribosomal RNA*) tem 1256 pb de comprimento, com conteúdo A+T de 74,86% nas duas linhagens de *R. prolixus*. Enquanto, o gene da subunidade menor do RNA ribossomal (rRNAs - *small subunit ribosomal RNA*) apresenta 803 pb de comprimento, com um conteúdo A + T de 71,98% para ambas colônias. Ambos estão com sequência completa.

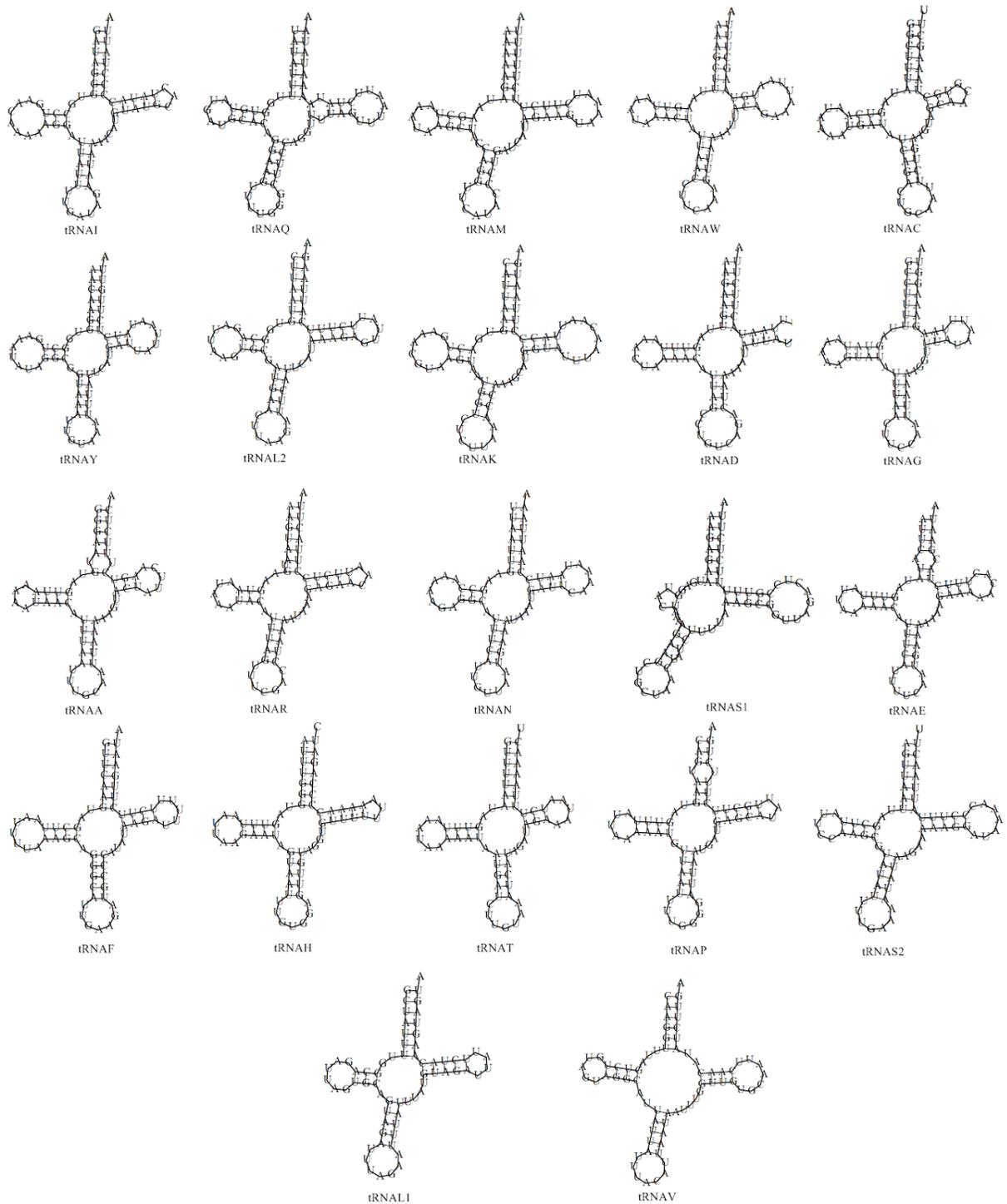


Figura 20 - Estrutura secundária predita dos RNAs transportadores do mitogenoma de *Rhodnius prolixus*, colônia UFRJ. Os nomes dos tRNAs segue nomenclatura do código de uma letra IUPAC-IUB. A estrutura secundária dos RNAs transportadores do mitogenoma da colônia NIH se encontram no anexo 1-a.

4.2.3 Regiões não codificantes

Além da região controle, outros 13 espaçadores intergênicos não-codificantes (IGS - do inglês *InterGenic Spacers*) foram identificados no mitogenoma de *Rhodnius prolixus*. A maioria dos IGS são curtos, contando com apenas de 1 a 11 pb. Entretanto, o maior IGS encontrado foi entre os genes tRNAS2 e NAD1 com 77 pb. Esta região intergênica não possui repetições em *tandem* e nem quadros abertos de leitura (ORF - do inglês, *Open Reading Frame*).

A região controle (CR - do inglês, *Control Region*), também chamada de região rica em A+T, é flanqueada pelos genes do rRNAs de um lado e pelo tRNAI do outro e apresenta 1.791 pb de comprimento. Essa região pode ser dividida em quatro trechos diferentes (Figura 21). O primeiro trecho (de 447 pb) está localizada após o gene rRNAs, é rico em G+C e termina em uma seqüência de 12 guaninas. Após as guaninas vem o segundo trecho, de 274 pb de comprimento, rico em A + T. O terceiro trecho (de 1028 pb) é composto por 2 conjuntos de repetições em *tandem* contendo 5 repetições de 172 pb e 3 repetições de 85 pb. A unidade repetitiva de 172 pb tem uma semelhança de 94% entre as duas colônias (Figura 22). Enquanto a unidade de 85 pb tem identidade de 98% entre as duas colônias (Figura 23). O trecho final da CR de 42 pb está imediatamente antes do gene de tRNAI. com 11 pares de nucleotídeos.

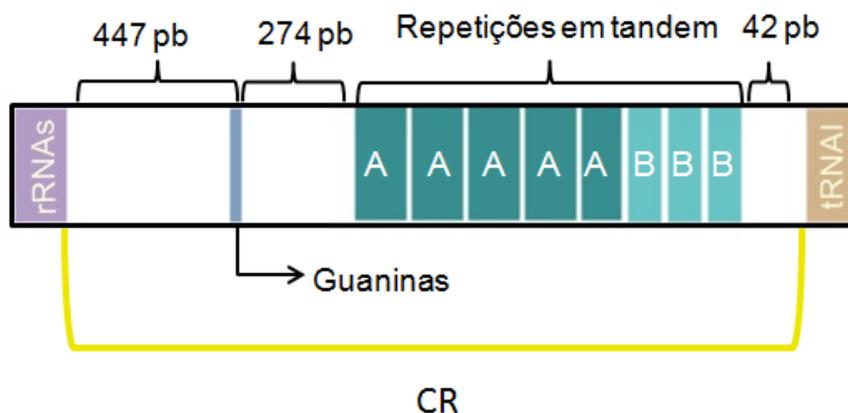


Figura 21 - Esquema da região controle do mitogenoma de *R. prolixus*. Esta região é flanqueada pelos genes do rRNAs (em roxo) e pelo tRNA (marrom). O primeiro trecho (447pb) é rico em G + C terminando com 12 guaninas. O trecho de 274 pb é rico em A+T. A região de repetição possui 5 repetições de 172pb (a) e 3 de 85pb (b). A região de 42pb representa o final da região de controle. Este esquema é válido para os mitogenomas de ambas as linhagens de *R. prolixus*.

```

      10      20      30      40      50      60      70      80
Colônia_UFRJ  GAAATAAGCTCTCGGTACCTTAAATACCTCTAAGCCATTTAGCGACAAGAAAAACCCGGAATCACTAACAAAAAACACAA
Colônia_NIH   ...GG.....GTCCC.CC..G.....C.....T.....

      110     120     130     140     150     160     170
Colônia_UFRJ  AAAGAAGGATACAAAAATGGACACATCCACAATAACAAGATTTCTACAAAACGAACAACTTGTATTTTAG
Colônia_NIH   .....

```

Figura 22 - Alinhamento da unidade de repetição de 172 pb da região controle dos mitogenomas de *R. prolixus*. Na linha "colônia_NIH" os pontos representam que aquela base é idêntica à acima.

```

      10      20      30      40      50      60      70      80
Colônia_UFRJ  --AATAATCCCAGGTGATTTAAACAACACCAGAATTTAATGTACCAATACTCATTATTTATTATAAACAATAATAAGATTCCATCCC
Colônia_NIH   CA.....T.....T.....A.--

```

Figura 23 - Alinhamento da unidade de repetição de 85 pb da região controle dos mitogenomas de *R. prolixus*. Na linha "colônia_NIH" os pontos representam que aquela base é idêntica à acima.

4.2.4 Origem de replicação

O alinhamento múltiplo do trecho de 447 pb (Figura 24), rico em G+C da região controle dos dois mitogenomas de *R. prolixus* em conjunto com a região equivalente dos mitogenomas de *Triatoma dimidiata* (AF301594), *Triatoma rubrofasciata* (MH934953) e *Triatoma infestans* (KY640305) mostrou três blocos conservados que podem estar relacionados a origem de replicação da fita *minority* (ON). Entretanto, o alinhamento do trecho de 274 pb (Figura 25), rico em A+T, também da região controle não mostrou nenhum bloco conservado. Nenhum bloco rico em T (poli-T) foi identificado nos trechos de 447 e 274pb.

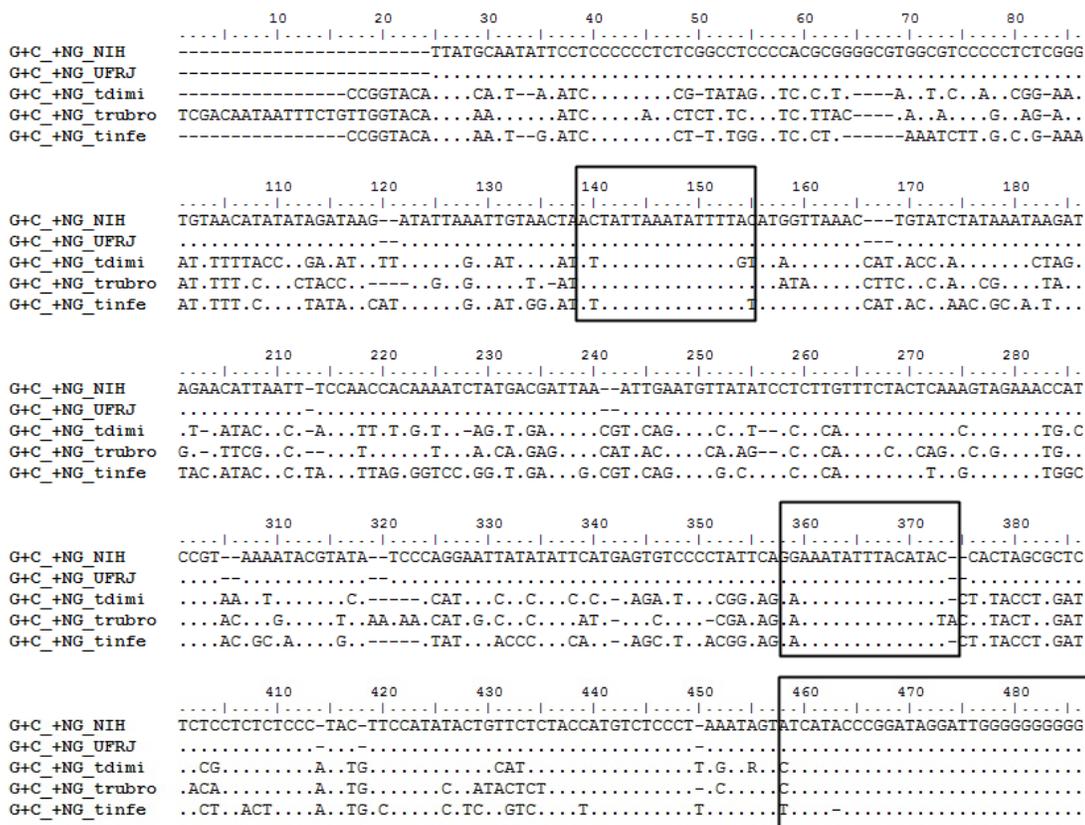


Figura 24 - Alinhamento do trecho de 447 pb rico em G+C da região controle do mitogenoma de triatomíneos. As sequências estão dispostas nesta ordem *Rhodnius prolixus* NIH, UFRJ, *Triatoma dimidiata*, *Triatoma rubrofasciata* e *Triatoma infestans*. Os retângulos indicam blocos conservados com potencial para serem a origem de replicação da fita *minority*.

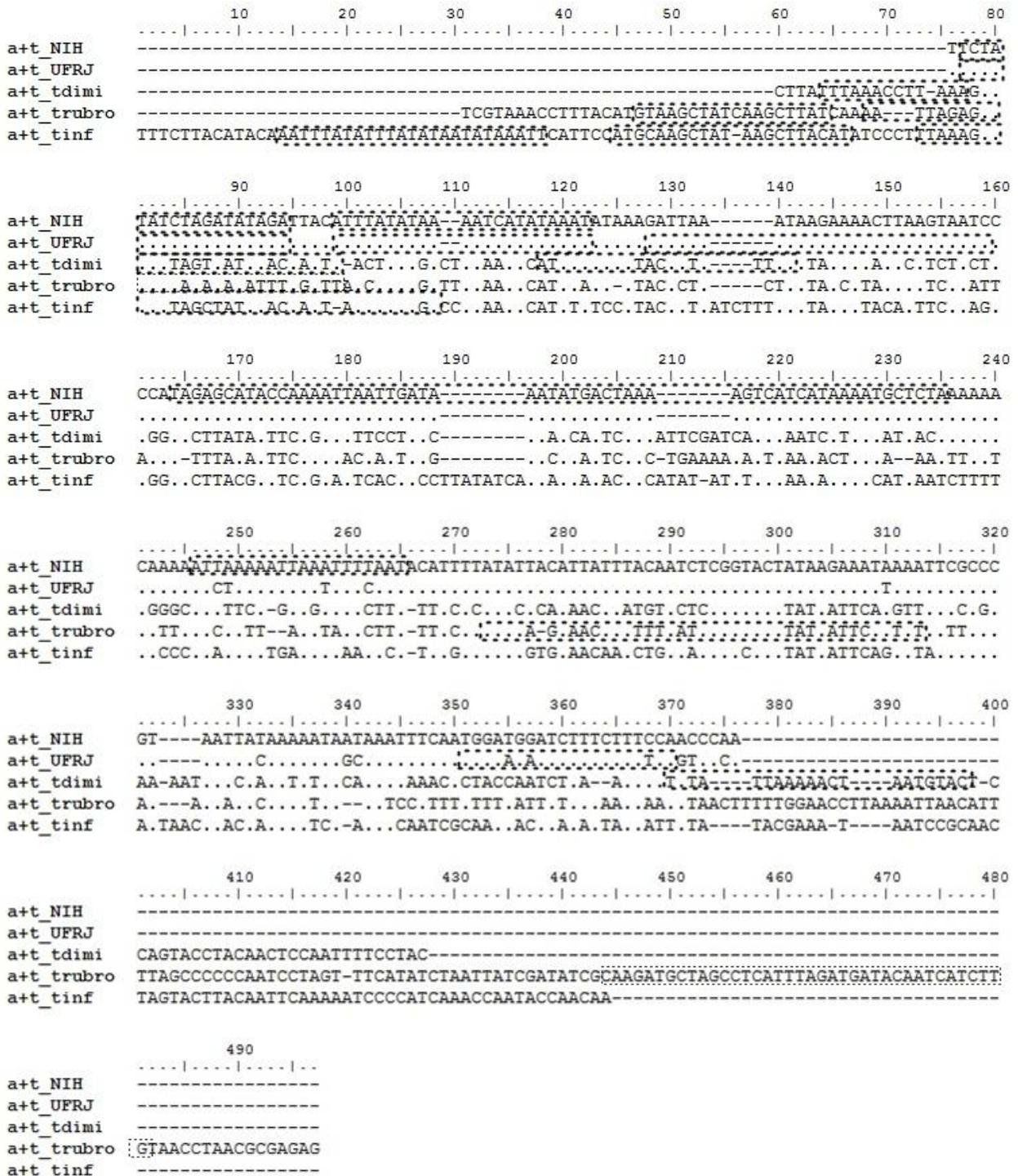


Figura 25 - Alinhamento do trecho de 274 pb rico em A+T da região controle do mitogenoma de triatomíneos. As sequências estão dispostas nesta ordem *Rhodnius prolixus* NIH, UFRJ, *Triatoma dimidiata*, *Triatoma rubrofasciata* e *Triatoma infestans*. Os retângulos com linhas tracejadas indicam grampos com potencial para serem a origem de replicação da fita *majority*.

O alinhamento do trecho de 42 pb localizado após as repetições em tandem mostrou a existência de um bloco poli-T na fita complementar (poli-A na figura), da posição 155 até a 163, em todos os triatomíneos analisados (Figura 26), que pode ser a origem de replicação da fita *majority*. Além disso, neste mesmo trecho foi identificada uma estrutura em grampo localizada entre as posições 164-170 e 171-177 - posições referentes aos mitogenomas *R. prolixus* UFRJ e NIH Figura 27. Este grampo contém o bloco poli-A citado logo reforçando que eles representam a origem de replicação da fita *majority*.

Todavia, ao menos três estruturas em grampo podem ser formadas no trecho de 274 pb tanto para os mitogenomas estudados nesta dissertação (Figura 28) como para os outros triatomíneos comparados (Figura 25).

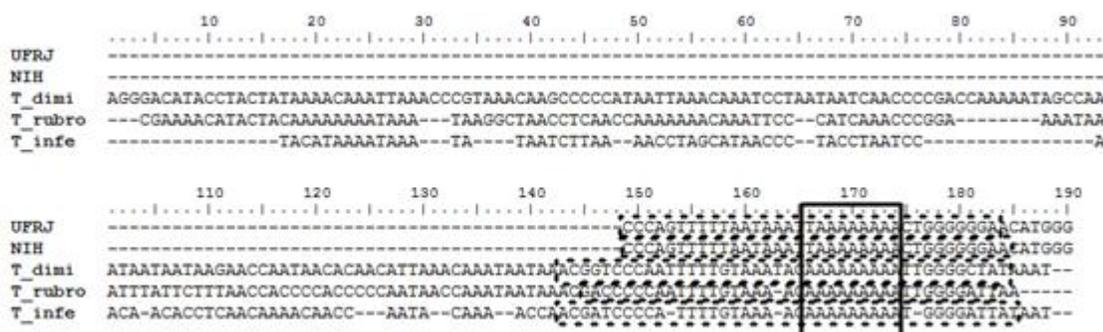


Figura 26 - Alinhamento do trecho de 42 pb do trecho final da região controle do mitogenoma de triatomíneos. As seqüências estão dispostas nesta ordem *Rhodnius prolixus* UFRJ, NIH, *Triatoma dimidiata*, *Triatoma rubrofasciata* e *Triatoma infestans*. O retângulo indica um bloco conservado com potencial para ser a origem de replicação da fita *majority*. Os retângulos com linhas tracejadas indicam grampos também com potencial para serem a origem de replicação da fita *majority*.

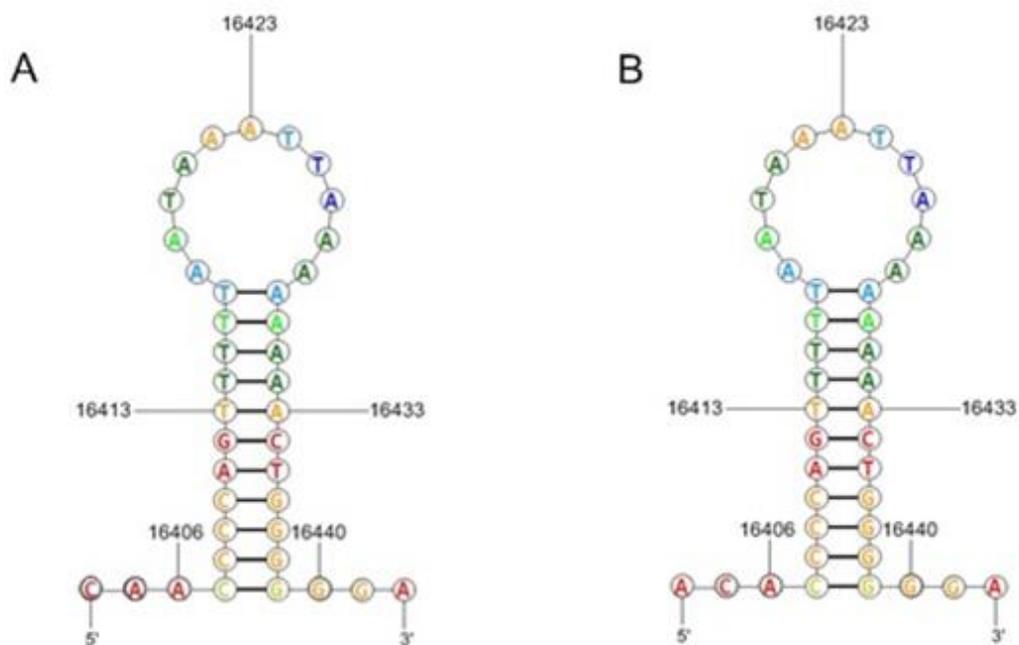


Figura 27 - Estrutura em formato de grampo no final da região controle do mitogenoma de *R. prolixus*. Em A está presente o grampo da colônia UFRJ em B o da colônia NIH. A única diferença encontrada entre eles é na posição 16.405 (adenina no UFRJ e citosina no NIH).

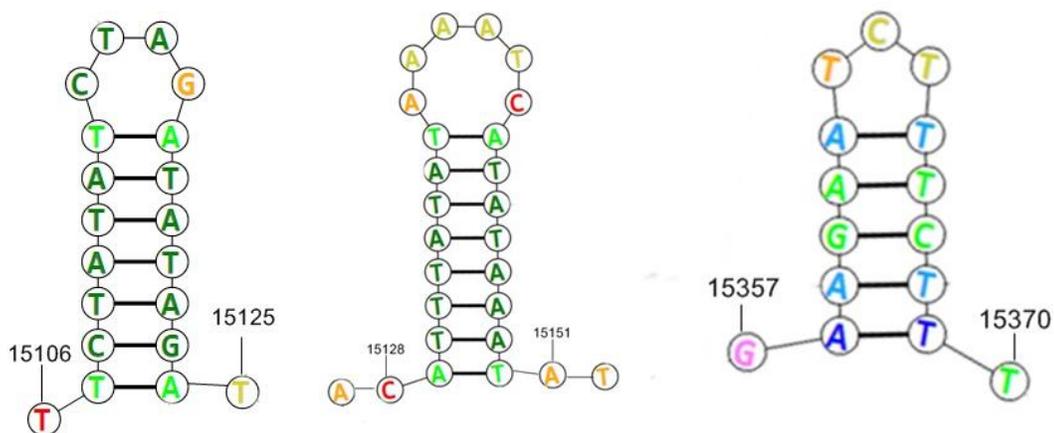


Figura 18 - Estrutura em formato de grampo formado no trecho de 274 pb na região controle do mitogenoma de *R. prolixus*, colônia UFRJ.

4.3 Busca de mitogenomas, alinhamento múltiplo e construção da árvore filogenética

A busca de mitogenomas de organismos próximos identificou 23 organismos com o mitogenoma completamente montado (Tabela 3). As buscas consideraram a família dos reduvídeos.

A reconstrução e análise filogenética geradas pelo método de máxima verossimilhança, utilizando as sequências mitocondriais concatenadas dos genes codificadores de proteínas, produziu uma árvore enraizada Figura 29 com duas espécies (representado por *A. lucorum* e *C. ciliate*). Em geral os valores de suporte dos nós foram altos, o menor deles foi 59. Como esperado as duas linhagens de *R. prolixus* foram agrupadas com suporte de 100. Todas as subfamílias presentes foram corretamente agrupadas, incluindo a subfamília *Triatominae* contendo não só os dois *R. prolixus* como também os três triatomas analisados. A exceção foi *Valentia hoffmanni* que foi agrupada com os membros da subfamília *Reduviinae*.

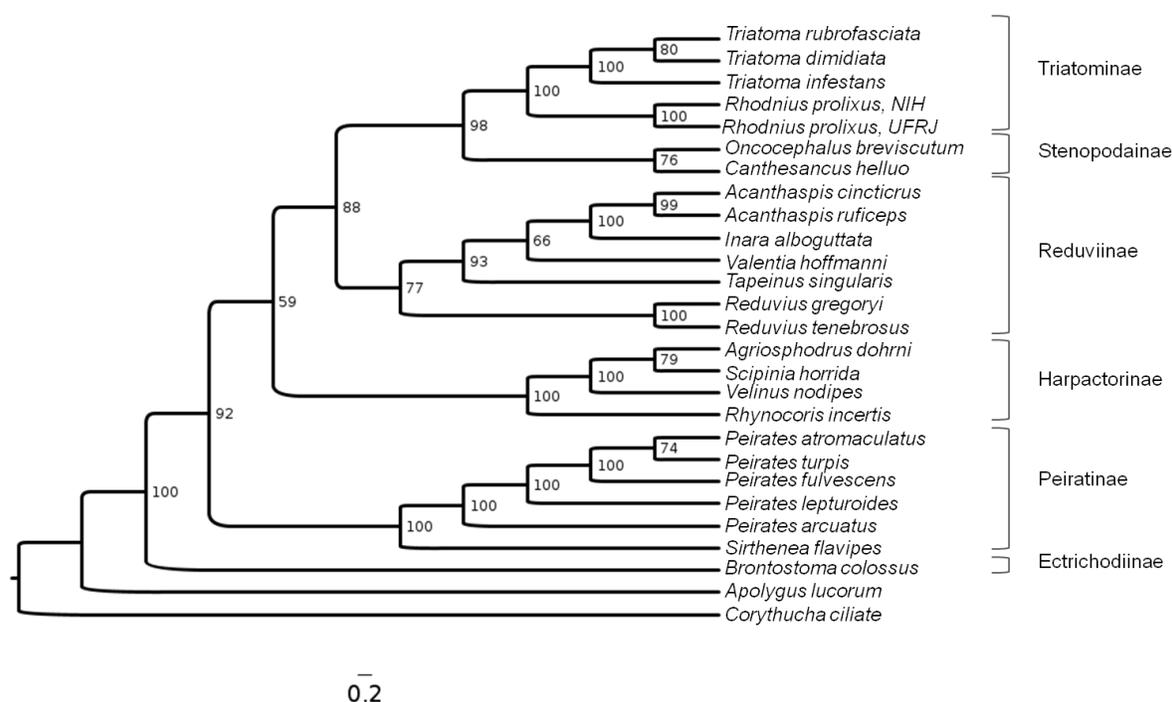


Figura 29 - Árvore filogenética construída pelo método de máxima verossimilhança a partir das sequências mitocondriais concatenadas dos 13 genes codificadores de proteínas. Foram usados todos os membros de Reduviidae sequenciados até o momento e dois grupos externos, *Apolygus lucorum* e *Corythucha ciliate*.

4.4 Nível de expressão dos transcritos mitocondriais.

O alinhamento das sequências codificantes de proteína (CDS) mitocondriais da colônia UFRJ e os *reads* limpos do transcriptoma permitiu a visualização da expressão destes genes (Figura 30) nas diferentes condições do transcriptoma. Dos *reads* do transcriptoma que passaram pelos filtros usados na etapa de limpeza (remoção de adaptador e *phred* 30), mais de 35 milhões (35.320.719) alinharam com os CDS, o que corresponde a 9,72% do total de *reads* limpos do transcriptoma.

Os diferentes tratamentos parecem interferir no nível de expressão dos genes codificantes proteínas (Figura 30), podendo ser destacado 1) o aumento de expressão dos genes NAD6, ATP8 e NAD2 nos tratamentos com tripomastigotas de *T. cruzi*; 2) a diminuição de expressão dos genes rRNAS, rRNAL e NAD1 logo após a alimentação com sangue; 3) a diminuição dos genes COXII, NAD3 e CYTB ao longo dos dias após a alimentação com sangue e 4) um aumento de expressão de quase todos os genes do complexo I (NAD1, NAD4, NAD4L, NAD5 e NAD6 - exceção NAD2) com o tratamento com epimastigotas (na maior concentração de 10^7 e no maior tempo de 12 dias). É importante destacar no resultado de expressão gênica (Figura 30) a diferença de perfil de expressão entre genes que estão adjacentes e com mesmo sentido de transcrição em uma mesma fita (sem ter nenhum tRNA nem rRNA entre eles). Estes grupos incluem na fita *majority* CYTB e NAD6; e ATP8, ATP6 e COXIII; e na fita *minority* somente NAD4I e NAD4.

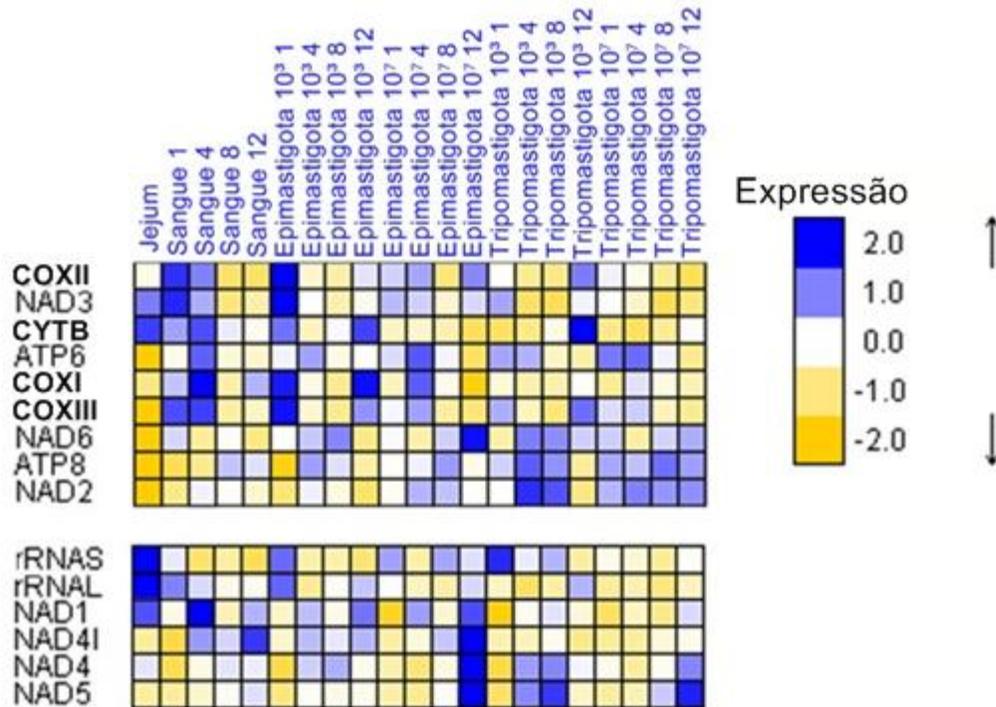


Figura 30 - Heatmap da expressão dos CDS (codificantes de proteínas) mitocondriais de *Rhodnius prolixus*, colônia UFRJ. Na parte superior estão as condições usadas no transcriptoma. Os genes foram organizados de modo a agrupar no bloco superior aqueles que encontram na fita *majority* e no bloco inferior os da fita *minority* do mitogenoma.

4.6 Comparação do mitogenoma de *R. prolixus*, colônia UFRJ, com sequências da literatura para as espécies *R. prolixus* e *R. robustus*.

A busca na literatura permitiu a identificação de dois genes para os quais já havia sequências depositadas tanto para a espécie *R. prolixus* quanto para *R. robustus*, sendo eles o Citocromo B (CYTB) e subunidade maior do RNA ribossomal (rRNAI). Os genes NAD1 e rRNAs possuíam poucas sequências disponíveis, assim foram desconsideradas nestas análises. As 36 sequências do gene CYTB de *R. prolixus* obtidas mostraram, no geral, identidade de cerca de 99% (Tabela 12), a maior diferença encontrada foi de 1,74% contra a sequência MH704760, de origem colombiana. Apenas uma sequência apresentou identidade de 100%, originária de Las Palmas, Guatemala. Por outro lado as 75 sequências de *R. robustus* se mostraram bem menos idênticas, sendo que a sequência mais parecida com o CYTB do mitogenoma UFRJ apresentou 97,13% de identidade (Tabela 13).

Tabela 12 - Identidade das sequências de citocromo b (CYTB) de *R. prolixus* disponíveis na literatura comparados à sequência da colônia UFRJ. A coluna “ID” contém os números de acesso das sequências do GenBank. A coluna “região” informa o território onde o inseto foi coletado. A coluna “Início” e “Fim” são referentes, respectivamente, a posição de início e fim da sequência em relação a CYTB de *Rhodnius prolixus* UFRJ. A coluna “Identidade” é a que mostra a identidade entre as sequências. A sequência CYTB de *Rhodnius prolixus* UFRJ tem 1.132 nucleotídeos.

ID	Região	Início	Fim	Identidade
AF045718	Colômbia, Bogotá	439	837	398/399(99%)
AF421339	Honduras, Francisco Morazan, Orica	340	1002	661/663(99%)
EF011726	Venezuela, Cojedes	340	1002	661/663(99%)
EF011723	Venezuela, Guárico, Ortiz	340	1002	661/663(99%)
EF011721	Venezuela, Guárico, San José Tiznados	340	1002	662/663(99%)
EF011716	Venezuela, Trujillo, Pampanito,	340	1002	661/663(99%)
EF043586	Guatemala, Tutucopote	469	883	412/415(99%)
EF043579	Guatemala, Tituque	469	883	412/415(99%)
EF043576	Guatemala, Las Palmas	469	883	415/415(100%)
EF043588	Venezuela, Barinas	469	883	413/415(99%)
EF043587	Venezuela, Barinas	469	883	413/415(99%)
EF043585	Venezuela, Barinas	469	883	412/415(99%)
EF043584	Venezuela, Barinas	469	883	413/415(99%)
EF043583	Venezuela, Barinas	469	883	414/415(99%)
EF043582	Venezuela, Portuguesa	469	883	413/415(99%)
EF043581	Venezuela, Portuguesa	469	883	410/415(99%)
EF043580	Venezuela, Barinas	469	883	413/415(99%)
EF043578	Venezuela, Trujillo	469	883	413/415(99%)
EF043577	Venezuela, Barinas	469	883	414/415(99%)
KP126734	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126733	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126732	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126731	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126730	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126729	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126728	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126727	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126726	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
KP126725	Colômbia, Casanare, Yopal	340	1002	662/663(99%)
DQ118977	México, Chiapas	501	803	312/313(99%)
EU363822	Venezuela	474	875	395/402(98%)
MH704764	Colômbia	340	998	657/659(99%)
MH704763	Colômbia	340	998	658/659(99%)
MH704762	Colômbia	340	998	657/659(99%)
MH704761	Colômbia	340	998	657/659(99%)
MH704760	Colômbia	340	998	648/659(98%)

Tabela 13 - Identidade das seqüências de citocromo b (CYTB) de *R. robustus* disponíveis na literatura comparados à seqüência da colônia UFRJ. A coluna “ID” contém os números de acesso das seqüências do GenBank. A coluna “região” informa o território onde o inseto foi coletado. A coluna “Início” e “Fim” são referentes, respectivamente, a posição de início e fim da seqüência da literatura em relação a de *Rhodnius prolixus* UFRJ. A coluna “Identidade” é a que mostra a identidade entre as seqüências. A seqüência CYTB de *Rhodnius prolixus* UFRJ tem 1.132 nucleotídeos.

ID	Região	Início	Fim	Identidade
AF421343	Brasil, Pará, Barcarena	340	1002	624/663(94%)
AF421342	Brasil, Pará, Itupiranga	340	1002	629/663(95%)
AF421341	Equador, Napo	340	1002	624/663(94%)
AF421340	Venezuela, Trujillo, Pampanito, La Concepcion	340	1002	643/663(97%)
FJ887793	Bolívia, Santa Cruz, El Torno	322	965	605/644(94%)
EF071583	Brasil, Rondônia	340	1002	622/663(94%)
EF011728	Guiana Francesa, Roura, Bélizon	340	1002	624/663(94%)
EF011727	Brasil, Amazonas, Manaus	340	1002	624/663(94%)
EF011725	Tocantins, Tocantinópolis	340	1002	626/663(94%)
EF011724	Brasil, Rondônia, Monte Negro	340	1002	625/663(94%)
EF011722	Brasil, Pará, Rio Mapuera	340	1002	623/663(94%)
EF011720	Brasil, Porto Velho	340	1002	622/663(94%)
EF011719	Venezuela, Trujillo, Candelaria	340	1002	644/663(97%)
EF011718	Venezuela, Trujillo, Candelaria	340	1002	643/663(97%)
EF011717	Venezuela, Portuguesa	340	1002	619/663(93%)
EF011715	Brasil, Pará	340	1002	628/663(95%)
EF011714	Brasil, Amazonas, Apuí	340	1002	622/663(94%)
EF011713	Brasil, Amazonas, Apuí	340	1002	620/663(94%)
EF011712	Guiana Francesa, Caiena	340	1002	624/663(94%)
EF011711	Brasil, Amazonas, Caruarí	340	1002	622/663(94%)
EF011710	Brasil, Amazonas, Caruarí	340	1002	624/663(94%)
EF011709	Brasil, Amazonas, Caruarí	340	1002	626/663(94%)
EF011708	Brasil, Amazonas, Balbina	340	1002	625/663(94%)
JX273158	Brasil, Amazonas, Novo Airão	340	1002	642/663(97%)
JN831970	Brasil, Amazonas, Médio Tajapós	427	882	424/456(93%)
JN831969	Brasil, Amazonas, Médio Tajapós	427	882	426/456(93%)
JN831968	Brasil, Amazonas, Médio Tajapós	427	882	425/456(93%)
JN831967	Brasil, Amazonas, Médio Tajapós	427	882	427/456(94%)
JN831966	Brasil, Amazonas, Médio Tajapós	427	882	423/456(93%)
JN831965	Brasil, Amazonas, Médio Tajapós	427	882	426/456(93%)
JN831964	Brasil, Amazonas, Médio Tajapós	427	882	427/456(94%)
JN831963	Brasil, Amazonas, Médio Tajapós	427	882	425/456(93%)
JN831962	Brasil, Amazonas, Médio Tajapós	427	882	426/456(93%)
JN831961	Brasil, Amazonas, Médio Tajapós	427	882	425/456(93%)
JN831960	Brasil, Amazonas, Médio Tajapós	427	882	426/456(93%)
JN831959	Brasil, Amazonas, Médio Tajapós	427	882	423/456(93%)
JN831958	Brasil, Amazonas, Médio Tajapós	427	882	424/456(93%)

Tabela 13 - Identidade das sequências de citocromo b (CYTB) de *R. robustus* disponíveis na literatura comparados à sequência da colônia UFRJ. A coluna “ID” contém os números de acesso das sequências do GenBank. A coluna “região” informa o território onde o inseto foi coletado. A coluna “Início” e “Fim” são referentes, respectivamente, a posição de início e fim da sequência da literatura em relação a de *Rhodnius prolixus* UFRJ. A coluna “Identidade” é a que mostra a identidade entre as sequências. A sequência CYTB de *Rhodnius prolixus* UFRJ tem 1.132 nucleotídeos. (Continuação).

ID	Região	Início	Fim	Identidade
JN831957	Brasil, Amazonas, Médio Tajapós	427	882	426/456(93%)
JN831956	Brasil, Amazonas, Médio Tajapós	427	882	424/456(93%)
JN831955	Brasil, Amazonas, Médio Tajapós	427	882	424/456(93%)
JN831954	Brasil, Amazonas, Médio Tajapós	427	882	426/456(93%)
JN831953	Brasil, Amazonas, Médio Tajapós	427	882	425/456(93%)
EU363824	Guiana Francesa	460	875	386/416(93%)
AF045717	Equador	439	837	373/399(93%)
KT805149	Bolívia, Ben, General José Ballivian, Yucumo	352	945	556/594(94%)
MF966306	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966305	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966304	Guiana Francesa, Rémire-Montjoly	358	1008	614/651(94%)
MF966303	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966302	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966301	Guiana Francesa, Matoury	358	1008	614/651(94%)
MF966300	Guiana Francesa, Matoury	358	1008	613/651(94%)
MF966299	Guiana Francesa, Matoury	358	1008	614/651(94%)
MF966298	Guiana Francesa, Saint Laurent du Maroni	358	1008	615/651(94%)
MF966297	Guiana Francesa, Mana	358	1008	614/651(94%)
MF966296	Guiana Francesa, Saint Laurent du Maroni	358	1008	613/651(94%)
MF966295	Guiana Francesa, Saint Laurent du Maroni	358	1008	613/651(94%)
MF966294	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966293	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966292	Guiana Francesa, Roura, Bélizon	358	1008	614/651(94%)
MF966291	Guiana Francesa, Caiena	358	1008	613/651(94%)
MF966290	Guiana Francesa, Caiena	358	1008	612/651(94%)
MF966289	Guiana Francesa, Matoury	358	1008	614/651(94%)
MF966288	Guiana Francesa, Antecum Pata	358	1008	615/651(94%)
MF966287	Guiana Francesa, Les Nouragues	358	1008	614/651(94%)
MF966286	Guiana Francesa, Saul	358	1008	615/651(94%)
MF966285	Guiana Francesa, Rémire-Montjoly	358	1008	614/651(94%)
MF966284	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966283	Guiana Francesa, Saint Laurent du Maroni	358	1008	614/651(94%)
MF966282	Guiana Francesa, Rémire-Montjoly	358	1008	614/651(94%)
MF966281	Guiana Francesa, Régina	358	1008	614/651(94%)
MF966280	Guiana Francesa, Rémire-Montjoly	358	1008	613/651(94%)
MF966279	Guiana Francesa, Roura, Bélizon	358	1008	614/651(94%)
MF966278	Guiana Francesa, Roura, Bélizon	358	1008	614/651(94%)
MF966277	Guiana Francesa, Roura, Bélizon	358	1008	614/651(94%)

Quanto às sequências do rRNAI, poucas diferenças foram encontradas entre o mitogenoma UFRJ e os dados de *R. prolixus* da literatura, apenas uma sequência não teve identidade de 100%, a sequência AF028747 apresentou identidade de 97,89% (Tabela 14). Para *R. robustus* a situação foi inversa (Tabela 15), onde quase a totalidade das sequências apresentaram identidade de cerca de 98% e apenas uma apresentou 100% (EU827206).

Tabela 14 - Identidade das sequências da subunidade maior do RNA ribossomal (rRNAI) de *R. prolixus* disponíveis na literatura comparados à sequência da colônia UFRJ. A coluna "ID" com os números de acesso do GenBank. A coluna "Região" informa o território onde o inseto foi coletado. A coluna "Início" e "Fim" são referentes, respectivamente, a posição de início e fim da sequência da literatura em relação a de *Rhodnius prolixus* UFRJ. A coluna "Identidade" é a que mostra a identidade entre as sequências. * indica a ausência desta informação no registro do banco de dados e na referência associada a ele. "LSHTM": Proveniente do insetário da London School of Hygiene and Tropical Medicine. "SNC": Proveniente do Servicio Nacional de Chagas, Córdoba, Argentina. "FMRP": Proveniente da Faculdade de Medicina Ribeirão Preto. A sequência rRNAI de *Rhodnius prolixus* UFRJ tem 1.256 nucleotídeos.

ID	Região	Início	Fim	Identidade
AF028747	*	199	481	278/284(98%)
AF045707	Colômbia, Bogotá	252	624	373/373(100%)
AF324519	*	118	592	475/475(100%)
EU822952	Colômbia	312	627	316/316(100%)
EU822953	Colômbia	312	627	316/316(100%)
EU822954	Colômbia	312	627	316/316(100%)
EU827235	*	312	627	315/316(100%)
EU827236	*	312	627	315/316(100%)

Tabela 15 - Identidade das sequências da subunidade maior do RNA ribossomal (rRNAI) de *R. robustus* disponíveis na literatura comparados à sequência da colônia UFRJ. A coluna "ID" com os números de acesso do GenBank. A coluna "Região" informa o território onde o inseto foi coletado. A coluna "Início" e "Fim", respectivamente, a posição de início e fim da sequência da literatura em relação a de *Rhodnius prolixus* UFRJ. A coluna "Identidade" é a que mostra a identidade entre as sequências. A sequência rRNAI de *Rhodnius prolixus* UFRJ tem 1.256 nucleotídeos.

ID	Região	Início	Fim	Identidade
AF045705	Equador	252	625	362/374(96%)
EU827206	Peru, Madre de Diós	312	627	316/316(100%)
KT805173	Bolívia, Alto Beni, Yucumo	285	581	290/298(97%)
MF966360	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966359	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966358	Guiana Francesa, Rémire-Montjoly	301	596	290/296(98%)
MF966357	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)

Tabela 15 - Identidade das sequências da subunidade maior do RNA ribossomal (rRNAI) de *R. robustus* disponíveis na literatura comparados à sequência da colônia UFRJ. A coluna “ID” com os números de acesso do GenBank. A coluna “Região” informa o território onde o inseto foi coletado. A coluna “Início” e “Fim”, respectivamente, a posição de início e fim da sequência da literatura em relação a de *Rhodnius prolixus* UFRJ. A coluna “Identidade” é a que mostra a identidade entre as sequências. A sequência rRNAI de *Rhodnius prolixus* UFRJ tem 1.256 nucleotídeos.(Continuação).

ID	Região	Início	Fim	Identidade
MF966356	Guiana Francesa, Matoury	301	596	290/296(98%)
MF966355	Guiana Francesa, Matoury	301	596	290/296(98%)
MF966354	Guiana Francesa, Matoury	301	596	290/296(98%)
MF966353	Guiana Francesa, Matoury	301	596	290/296(98%)
MF966352	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966351	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966350	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966349	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966348	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966347	Guiana Francesa, Caiena	301	596	290/296(98%)
MF966346	Guiana Francesa, Caiena	301	596	290/296(98%)
MF966345	Guiana Francesa, Matoury	301	596	290/296(98%)
MF966344	Guiana Francesa, Rémire-Montjoly	301	596	290/296(98%)
MF966343	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966342	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966341	Guiana Francesa, Saint Laurent du Maroni	301	596	289/296(98%)
MF966340	Guiana Francesa, Rémire-Montjoly	301	596	290/296(98%)
MF966339	Guiana Francesa, Rémire-Montjoly	301	596	290/296(98%)
MF966338	Guiana Francesa, Camopi	301	596	290/296(98%)
MF966337	Guiana Francesa, Régina	301	596	290/296(98%)
MF966336	Guiana Francesa, Rémire-Montjoly	301	596	290/296(98%)
MF966335	Guiana Francesa, Roura, Bélizon	301	596	290/296(98%)
MF966334	Guiana Francesa, Roura, Bélizon	301	596	290/296(98%)
MF966333	Guiana Francesa, Roura, Bélizon	301	596	290/296(98%)
MF966332	Guiana Francesa, Roura, Bélizon	301	596	290/296(98%)

Para ilustrar as comparações dos genes CYTB e rRNAI com os dados da literatura, uma região representativa presente em todas as sequências da literatura foi alinhada com as sequências da literatura tanto para *R. prolixus* quanto para *R. robustus*. A comparação de CYTB do mitogenoma UFRJ e NIH, entre os nucleotídeos 610 e 690, ilustra uma grande identidade de ambos em comparação com os dados de *R. prolixus* (Figura 31) enquanto contra *R. robustus* (Figura 32) mostra posições

nucleotídeos consistentemente diferentes. Quanto ao gene rRNAI, entre os nucleotídeos 670 a 900, pode-se observar o mesmo padrão, ou seja, uma maior identidade na comparação contra *R. prolixus* (Figura 33) e posições completamente diferentes contra *R. robustus* (Figura 34). A sequência EU827206, coletada em Madre de Dios no Peru, é idêntica ao citocromo b do mitogenoma-UFRJ (Tabela 15 e Figura 34), sendo uma exceção aos resultados observados para as outras sequências de *R. robustus*.

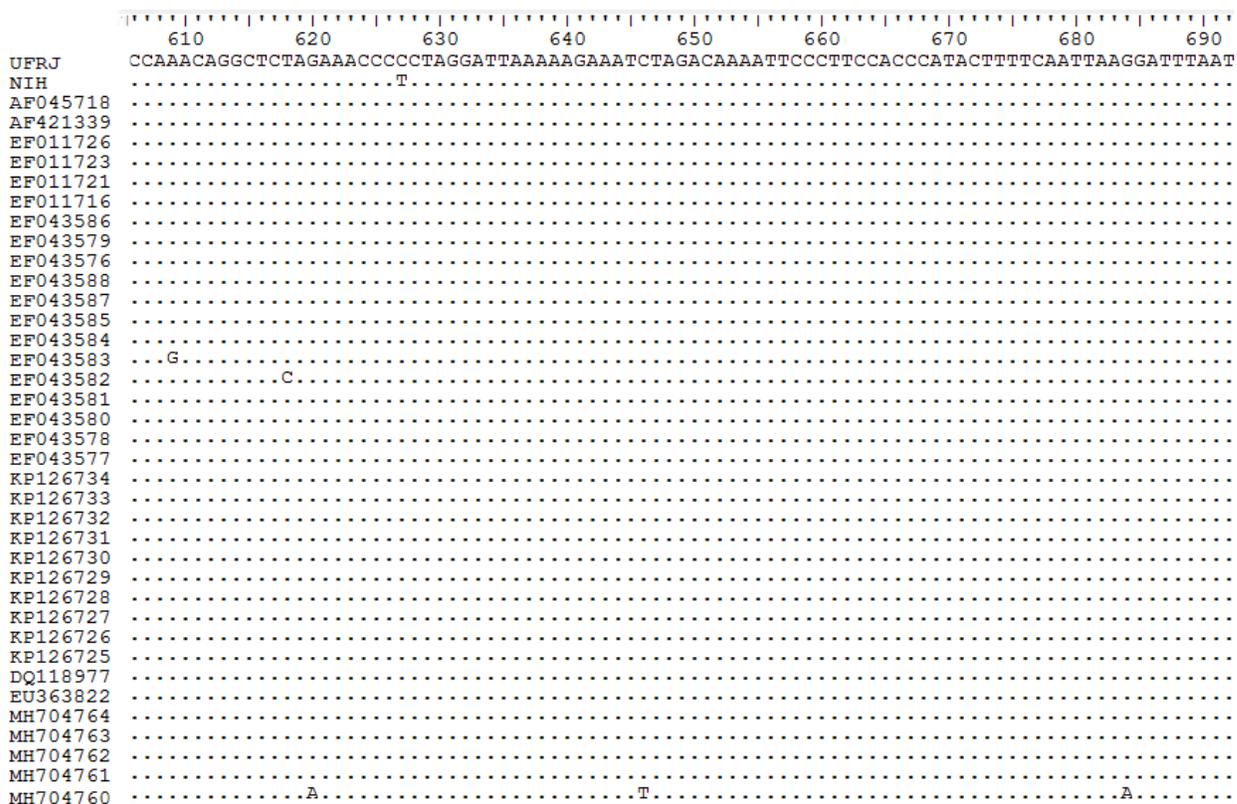


Figura 31 - Alinhamento múltiplo da região 610 a 690 das sequências de CYTB de *Rhodnius prolixus*. Os números de acesso do GenBank das sequências estão listados à esquerda. Os pontos representam nucleotídeos idênticos aos presentes na sequência UFRJ. "UFRJ" e "NIH" representam a CYTB dos mitogenomas de *R. prolixus* das colônias UFRJ e NIH.

	610	620	630	640	650	660	670	680	690
UFRJ	TACACCAAACAGGCTCTAGAAACCCCTAGGATTA	AAAAAGAAATCTAGACAAAATTC	CCCTCCACC	CATACTTTTCAATTAAGGATTTAAT					
NIH	T.....
AF421343	C.....	T.....	G.....	T.....	G.....
AF421342	C.....	T.....	G.....	T.....	G.....
AF421341	C.....	T.....	G.....	T.....	G.....
AF421340	G.....	TT.....
FJ887793	C.....	T.....	G.....	T.....	G.....
EF071583	CG.....	T.....	G.....	T.....	G.....
EF011728	C.....	T.....	G.....	T.....	G.....
EF011727	C.....	TT.....	G.....	T.....	G.....
EF011725	C.....	TT.....	G.....	T.....	G.....
EF011724	C.....	G.....	T.....	G.....
EF011722	C.....	TT.....	G.....	T.....	G.....
EF011720	C.....	G.....	T.....	G.....
EF011719	G.....	TT.....
EF011718	G.....	TT.....
EF011717	T.....	C.....	TT.....	G.....	T.....	G.....
EF011715	C.....	TT.....	G.....	T.....	G.....
EF011714	C.....	TT.....	G.....	T.....	G.....
EF011713	C.....	TT.....	G.....	T.....	A.....	G.....
EF011712	C.....	T.....	G.....	T.....	G.....
EF011711	C.....	T.....	G.....	T.....	G.....
EF011710	C.....	G.....	T.....	G.....
EF011709	C.....	G.....	T.....	G.....
EF011708	C.....	TT.....	G.....	T.....	G.....
JX273158	G.....	T.....	A.....	G.....
JN831970	C.....	TT.....	T.....	G.....
JN831969	C.....	T.....	T.....	G.....
JN831968	C.....	T.....	G.....	T.....	G.....
JN831967	C.....	T.....	G.....	T.....	G.....
JN831966	C.....	TT.....	G.....	T.....	G.....
JN831965	C.....	T.....	G.....	T.....	G.....
JN831964	C.....	TT.....	T.....	G.....
JN831963	C.....	TT.....	T.....	G.....
JN831962	C.....	TT.....	T.....	G.....
JN831961	C.....	TT.....	T.....	G.....
JN831960	C.....	TT.....	T.....	G.....
JN831959	T.....	C.....	TT.....	T.....	A.....	G.....
JN831958	C.....	TT.....	T.....	A.....	G.....
JN831957	C.....	TT.....	T.....	G.....
JN831956	C.....	TT.....	T.....	G.....
JN831955	C.....	TT.....	T.....	G.....
JN831954	C.....	TT.....	T.....	G.....
JN831953	C.....	TT.....	T.....	G.....
EU363824	C.....	T.....	G.....	T.....	G.....
AF045717	C.....	T.....	G.....	T.....	G.....
KT805149	C.....	T.....	G.....	T.....	G.....
MF966306	C.....	T.....	G.....	T.....	G.....
MF966305	C.....	T.....	G.....	T.....	G.....
MF966304	C.....	T.....	G.....	T.....	G.....
MF966303	C.....	T.....	G.....	T.....	G.....
MF966302	C.....	T.....	G.....	T.....	G.....
MF966301	C.....	T.....	G.....	T.....	G.....
MF966300	C.....	T.....	G.....	T.....	G.....
MF966299	C.....	T.....	G.....	T.....	G.....
MF966298	C.....	T.....	G.....	T.....	G.....
MF966297	C.....	T.....	G.....	T.....	G.....
MF966296	C.....	T.....	G.....	T.....	G.....
MF966295	C.....	T.....	G.....	T.....	G.....
MF966294	C.....	T.....	G.....	T.....	G.....
MF966293	C.....	T.....	G.....	T.....	G.....
MF966292	C.....	T.....	G.....	T.....	G.....
MF966291	C.....	T.....	G.....	T.....	G.....
MF966290	C.....	T.....	G.....	T.....	G.....
MF966289	C.....	T.....	G.....	T.....	G.....
MF966288	C.....	T.....	G.....	T.....	G.....
MF966287	C.....	T.....	G.....	T.....	G.....
MF966286	C.....	T.....	G.....	T.....	G.....
MF966285	C.....	T.....	G.....	T.....	G.....
MF966284	C.....	T.....	G.....	T.....	G.....
MF966283	C.....	T.....	G.....	T.....	G.....
MF966282	C.....	T.....	G.....	T.....	G.....
MF966281	C.....	T.....	G.....	T.....	G.....
MF966280	C.....	T.....	G.....	T.....	G.....
MF966279	C.....	T.....	G.....	T.....	G.....
MF966278	C.....	T.....	G.....	T.....	G.....
MF966277	C.....	T.....	G.....	T.....	G.....

Figura 32 - Alinhamento múltiplo da região 610 a 690 das sequências de CYTB de *Rhodnius robustus*. Os números de acesso do GenBank das sequências estão listados à esquerda. Os pontos representam nucleotídeos idênticos aos presentes na sequência UFRJ. "UFRJ" e "NIH" representam a CYTB dos mitogenomas de *R. prolixus* das colônias UFRJ e NIH.

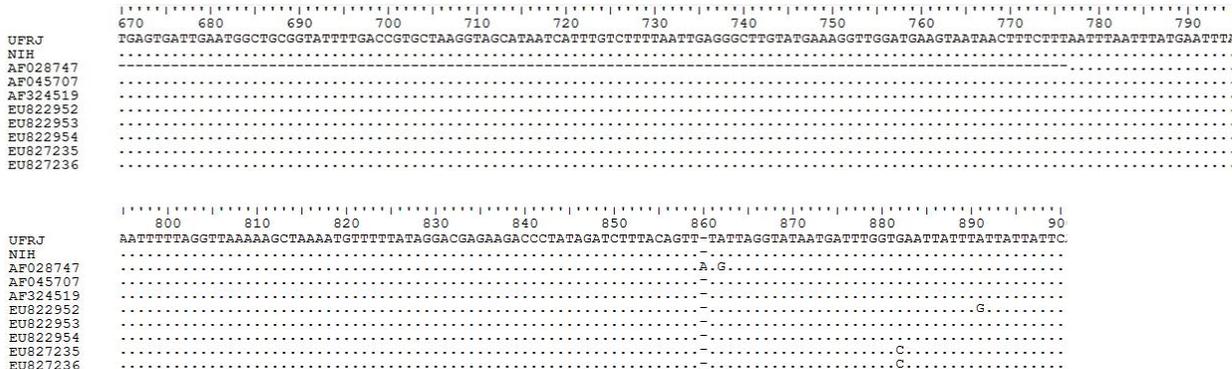


Figura 3319 - Alinhamento múltiplo da região 670 a 900 das sequências de rRNAI de *Rhodnius prolixus*. Os números de acesso do GenBank das sequências estão listados à esquerda. Os pontos representam nucleotídeos idênticos aos presentes na sequência UFRJ. "UFRJ" e "NIH" representam a CYTB dos mitogenomas de *R. prolixus* das colônias UFRJ e NIH.

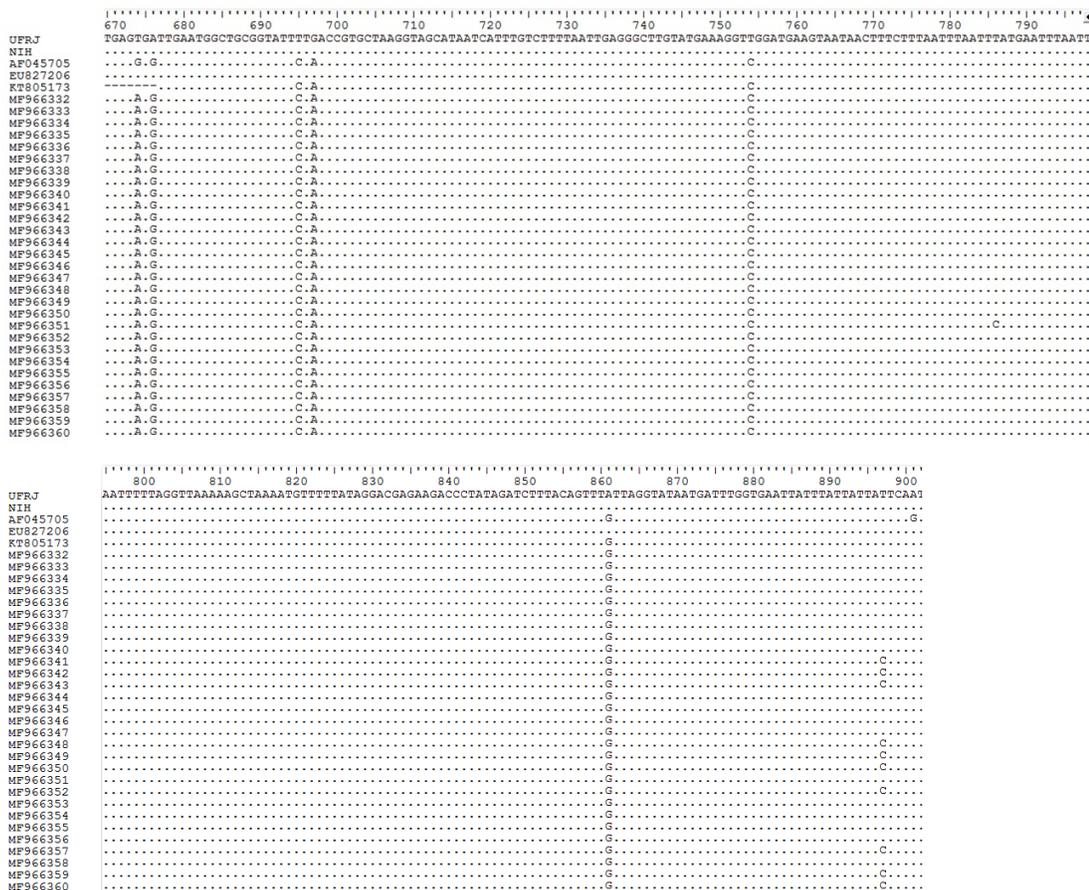


Figura 34 20- Alinhamento múltiplo da região 670 a 900 das sequências de rRNAI de *Rhodnius robustus*. Os números de acesso do GenBank das sequências estão listados à esquerda. Os pontos representam nucleotídeos idênticos aos presentes na sequência UFRJ. "UFRJ" e "NIH" representam a CYTB dos mitogenomas de *R. prolixus* das colônias UFRJ e NIH.

A árvore filogenética construída com base nas sequências do gene do citocromo b, pelo método de Neighbor-Joining (NJ) usando como modelo distância 2-parâmetros de Kimura (Kimura, 1980) produziu uma árvore enraizada com três espécies formando o grupo externo [representado por *Rhodnius pallescens* (FJ229359, JQ686675 e JQ686685) em vermelho, *Rhodnius pictipes* (MF497699) em azul claro e *Rhodnius barreti* (JX273159 e JX273161) em roxo], como pode ser visto na Figura 35. Em geral, os nós apresentam um bom valor de confiança. Todas as sequências identificadas como *R. prolixus* formaram um grupo único, com bom suporte. Já as amostras de *Rhodnius robustus*, formaram 5 grupos distintos, condizentes com as linhagens que são descritas na literatura. As linhagens de *Rhodnius robustus* I e V mostraram-se ser mais próximas de *R. prolixus* do que dos outros membros de sua própria espécie.

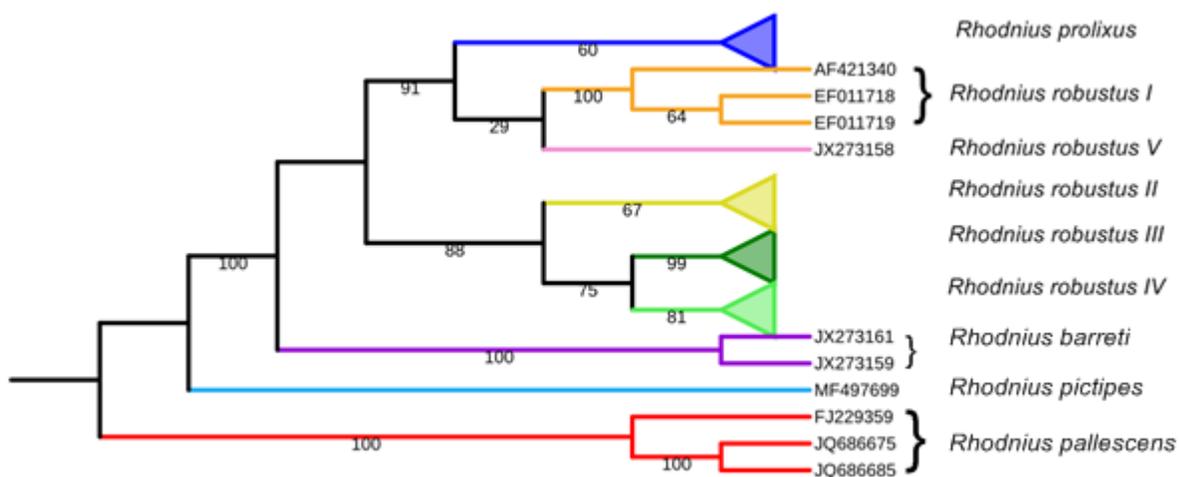


Figura 35 - Árvore filogenética do gene mitocondrial citocromo b. Esta árvore foi feita pelo método Neighbor-Joining, usando como modelo 2-parâmetros de Kimura, e suporte de 1000 réplicas de bootstrap. Como grupo externo estão em vermelho as sequências de *Rhodnius pallescens* (FJ229359, JQ686675 e JQ686685), em azul claro *Rhodnius pictipes* (MF497699) e em roxo *Rhodnius barreti* (JX273159 e JX273161). A árvore estendida se encontra no Anexo 2.

4.5 Descoberta e caracterização de polimorfismos de nucleotídeo único, inserções e deleções.

A busca por polimorfismos de nucleotídeo único (SNP - do inglês Single Nucleotide Polymorphism), e inserções e deleções (Indels) foi feita para o mitogenoma da colônia UFRJ. Para a descoberta dos variantes desta da colônia foi usado o alinhamento do mitogenoma montado com os reads limpos de cada biblioteca, resultado já mostrado anteriormente (Tópico 3.1). O programa GATK identificou 8.843 variantes neste mitogenoma, porém somente 1.554 passaram pelo filtro de qualidade do próprio programa, destes 1.498 foram considerados SNPs e 56 Indels. A tabela 16 contém a quantidade de variantes encontradas em cada condição do transcriptoma.

Tabela 1611 - Número de variantes encontrados no mitogenoma-UFRJ, em cada biblioteca do transcriptoma. A coluna “amostra” indica o número da biblioteca. A coluna “condição” informa o tratamento ou condição usado para a produção da biblioteca. A coluna “dia” informa o dia que as amostras foram analisadas após a alimentação. "SNP": polimorfismos de nucleotídeo único; e "INDEL": inserções e deleções.

Amostra	Condição	Dia	Antes da filtragem		Após a filtragem	
			SNP	INDEL	SNP	INDEL
J	Jejum	-	424	11	74	2
S1	Sangue	1	395	14	78	4
S2		4	382	12	72	6
S3		8	427	10	78	3
S4		12	427	13	18	2
E1	Epimastigote 10 ³	1	412	9	72	4
E2		4	388	12	81	2
E3		8	401	10	82	3
E4		12	422	11	73	2

Tabela 16 - Número de variantes encontrados no mitogenoma-UFRJ, em cada biblioteca do transcriptoma. A coluna “amostra” indica o número da biblioteca. A coluna “condição” informa o tratamento ou condição usado para a produção da biblioteca. A coluna “dia” informa o dia que as amostras foram analisadas após a alimentação. "SNP": polimorfismos de nucleotídeo único; e "INDEL": inserções e deleções. (Continuação).

Amostra	Condição	Dia	Antes da filtragem		Após a filtragem	
			SNP	INDEL	SNP	INDEL
E5	Epimastigota 10 ⁷	1	441	11	62	2
E6		4	403	10	67	2
E7		8	400	12	78	2
E8		12	415	15	68	3
T1	Tripomastigota 10 ³	1	407	12	72	2
T2		4	417	12	79	3
T3		8	399	14	79	3
T4		12	409	12	76	4
T5	Tripomastigota 10 ⁷	1	413	12	73	2
T6		4	412	12	62	2
T7		8	393	13	78	2
T8		12	409	10	76	2

Numericamente, a presença dos SNPs foi homogênea para todas as amostras do transcriptoma, sendo que os 1.498 SNPs encontrados ao todo compreendem, de fato, 144 posições únicas, que possuem resultados em mais de uma amostra. A Figura 36 indica o número de SNPs que foram encontrados em mais de uma amostra, sendo que apenas 25 desses SNPs foram identificados para todas as amostras, este grupo

está listado na Tabela 17. Em duas posições (4763 e 7267) dois nucleotídeos alternativos para cada posição foram identificados com frequência suficiente para serem considerados como SNPs.

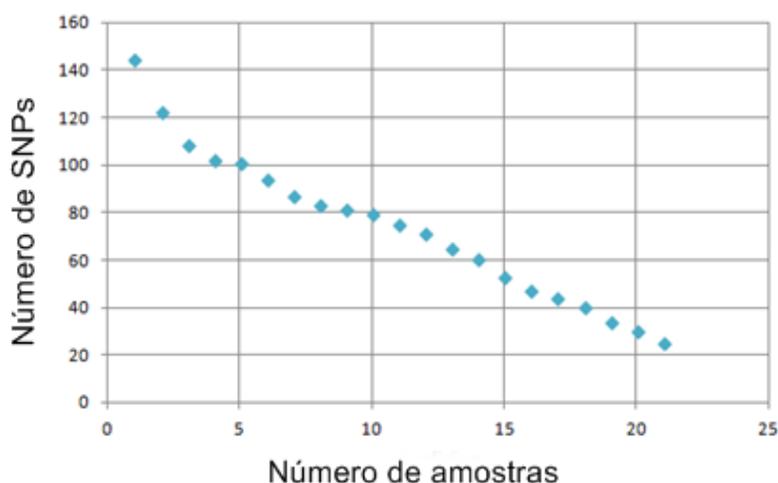


Figura 3621 - Histograma. Número de SNPs encontrados em apenas uma amostra (x=1) ou repetidos em mais de uma amostra.

Tabela 17 - SNPs identificados em todas as amostras do mitogenoma-UFRJ. "IGS": Espaçador intergênico; "POS": Posição do SNP; "REF": Nucleotídeo referência no mitogenoma UFRJ; "ALT": Nucleotídeo(s) alternativo(s).

Gene	POS	REF	ALT
NAD2	347	A	G
COXI	1529	T	C
tRNAL2	2931	G	A
COXII	2997	C	T
ATP6	4296	T	C
COXIII	4763	T	A
	4763	T	C
	4777	T	C
NAD5	6684	T	C
	6724	C	T
	7267	C	A
	7267	C	G
	7291	A	G
NAD4	8016	C	T

Gene	POS	REF	ALT
NAD4	8134	C	T
	8149	T	C
	8686	C	T
	8821	C	T
NAD6	10015	C	T
CYTB	11252	A	G
IGS	11503	C	T
NAD1	12341	A	G
rRNAI	13328	G	A
	13497	A	G
	13523	G	A
	13557	C	T
rRNAs	14192	G	A

As substituições mais frequentes nos SNPs identificados em todas as bibliotecas foram do tipo transição. Como pode ser visto na figura 37, 24 alterações (88%) representam transições (substituição de A-G mais C-T) e apenas 3 transversões (substituições de uma base purínica por uma pirimidina, ou vice-versa). A mudança mais frequente foi a troca de uma citosina por uma timina, encontrada 9 vezes (33,4%).

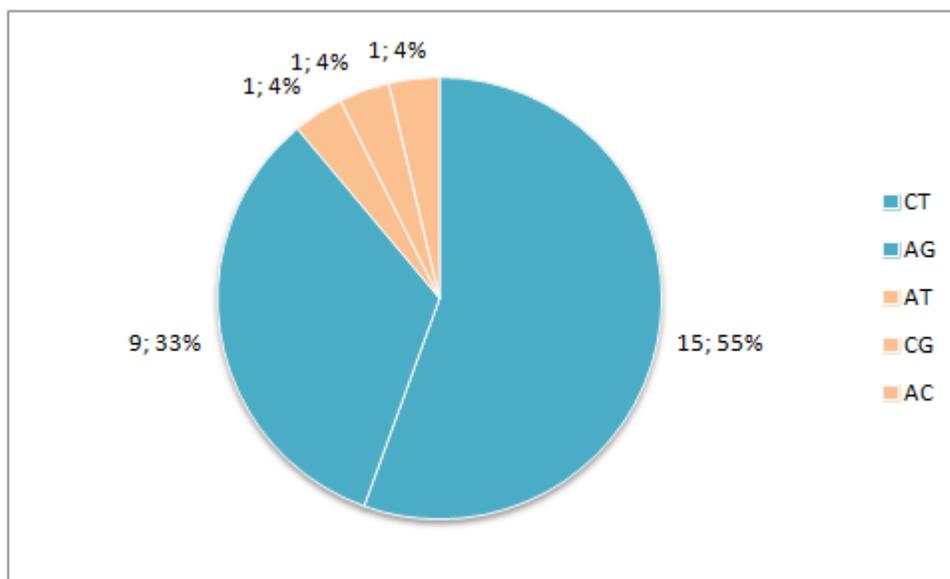


Figura 227 - Porcentagem de transições e transversões que ocorreram nos SNPs do mitogenoma UFRJ.

Das 25 posições identificadas como SNPs, 18 estão incluídas em regiões de genes codificadores de proteínas. Os SNPs que ocorrem nessas regiões podem ser classificados em sinônimos quando não há alteração no aminoácido (AA) presente na proteína ou não sinônimos quando há (Tabela 18).

Tabela 18 - Substituições sinônimas ou não sinônimas nos genes codificadores de proteínas. "AA": Aminoácido; "-": Não aplicável.

Gene	POS	AA original	AA alterado	SIFT score
NAD2	347	S	-	-
COXI	1529	I	-	-
COXII	2997	N	-	-
ATP6	4296	I	-	-
COXIII	4763	G	-	-
	4763	G	-	-
	4777	V	A	0.13
NAD5	6684	I	V	0.62
	6724	G	-	-
	7267	T	-	-
	7267	T	-	-
	7291	V	-	-
NAD4	8016	V	I	0.64
	8134	Q	-	-
	8149	L	-	-
	8686	L	-	-
	8821	L	-	-
NAD6	10015	L	F	0.44
CYTB	11252	N	S	0.43
NAD1	12341	G	-	-

Em 15 SNPs presentes na colônia UFRJ a mudança ocorreu no terceiro nucleotídeo presente no códon, o que resultou em mutações sinônimas, onde não houve alteração no aminoácido final. Entretanto, em 5 mudanças o SNP foi considerado não sinônimo. Na posição 4.777, região do gene de COXIII e 11.252, gene CYTB, a mudança foi no segundo nucleotídeo do códon o que resultou na alteração, respectivamente, de uma valina para uma alanina e uma asparagina para uma serina. Nas posições 6.684, gene NAD5, 8.016, gene NAD4, e 10.015, gene NAD6, a alteração foi no primeiro nucleotídeo do códon, e levou a substituição, respectivamente, de uma isoleucina por uma valina, uma valina por uma isoleucina e uma leucina por fenilalanina. A análise com o programa SIFT mostra a conservação da posição do aminoácido numa dada proteína levando em consideração, os dados de outros organismos próximos disponíveis na literatura. Valores SIFT score abaixo de 0,05 são

mudanças pouco comum, logo, são consideradas intoleráveis para a proteína, possivelmente afetando sua função. Quando há alteração na função da proteína os SNPs são classificados como deletérios, e quando não há alteração, são chamados de tolerados. Todos os SNPs não sinônimos no mitogenoma-UFRJ foram considerados toleráveis para as proteínas, coincidentemente as mudanças envolveram a troca de um aminoácido não polar por outro de mesma natureza (Tabela 19).

Quanto às inserções e deleções, foram identificados 18 *Indels* no mitogenoma de *Rhodnius prolixus* colônia UFRJ. Porém apenas 5 passaram pelo filtro de qualidade do GATK e desses somente 2 estão presentes em todas as amostras (Tabela 19).

Tabela 19 - Indels identificados em todas as amostras do mitogenoma-UFRJ. “POS”: Posição do Indel. “INDEL”: “IN” para inserção ou “DEL” uma deleção; “Seq.”: mostra a inserção ou deleção ocorrida na posição de referência, nucleotídeo em negrito é o alterado.

Gene	POS	INDEL	Seq.
rRNAs	14114	DEL	ATCCA
Região controle	15006	IN	GCTCC

O primeiro *indel* identificado na colônia UFRJ constitui uma deleção de duas citosinas. O segundo *indel* é uma inserção de uma citosina na região controle, região conhecida por ser variável. Este indel localiza-se no trecho de 437 pb da região controle, não se encontra em nenhum bloco apontado anteriormente como conservado entre os triatomíneos.

A maioria das regiões gênicas onde estão localizados os SNPs e *Indels* encontrados neste estudo não tem correspondência com as sequências de *Rhodnius robustus* disponíveis na literatura. Mesmo nos genes que tem sequências disponíveis, tais como CYTB, a região onde o SNP está localizado não é coberto pela sequência parcial.

5. DISCUSSÃO

O triatomíneo *Rhodnius prolixus* (Stal, 1859) é um dos vetores da Doença de Chagas mais conhecidos. Isso pode ser associado a sua grande relevância epidemiológica, sendo considerado o principal vetor na Venezuela, Colômbia e em países da América Central. Uma identificação precisa dos barbeiros é de suma importância para o controle da Doença de Chagas, pois táxons distintos apresentam uma capacidade vetorial variada (GOURBIÈRE *et al.*, 2012). Além disso, o controle do vetor continua sendo o mecanismo mais eficaz de combate a esta enfermidade (DIAS *et al.*, 2016). Sendo assim, informações acerca desse vetor são consideradas muito valiosas. Neste cenário se insere o uso do genoma mitocondrial para solucionar questões taxonômicas relacionadas a espécies filogeneticamente muito próximas. Alguns fatores levaram o mitogenoma a ser considerado um bom modelo de estudo, tais como, ser uma molécula de origem uniparental, estar presente em abundância nas células, sofrer mutações e ter um mecanismo de reparo pouco eficiente que contribui para que ocorra altas taxas de mutações (PAVAN, 2013).

A descrição do mitogenoma de *Rhodnius prolixus* pode ser muito útil nos estudos de filogenia, principalmente relacionado a espécies muito próximas como *R. robustus*, *R. pictipes*, *R. montenegrensis* e *R. nasutus*. Além disso, a caracterização do mitogenoma das duas colônias de *R. prolixus* pode não só ratificar a identidade das espécies mantidas nas colônias como para permitir a comparação delas com estudos de dispersão desta espécie ao longo do seu território.

5.1 Montagem do genoma mitocondrial de *Rhodnius prolixus*, colônia UFRJ.

A montagem denominada *M1* do genoma mitocondrial de *Rhodnius prolixus* da colônia UFRJ foi a que menos cobriu a sequência de referência, apesar de ter sido a que produziu o maior *mito-contig* (Figura 15). Esta montagem utilizou todos os *reads*, sem separação prévia dos mito-reads, contrastando com as abordagens *M2* e *M3* onde os mito-reads foram identificados e separados para serem montados pelo SPAdes e Trinity. Para os nossos dados, o uso dos maiores *contigs* das três estratégias foi

essencial para a total cobertura do genoma, pois houve regiões que só foram cobertas por *contigs* de uma das montagens. Como a M1 foi a primeira estratégia empregada neste estudo os seus *contigs* forneceram o delineamento inicial para o mitogenoma-UFRJ. Por este motivo, M2s e M3s foram utilizadas para completar o mitogenoma, chegando a 100% de cobertura. Com a evolução dos dados de sequenciamento muitos estudos tem utilizado as ferramentas de RNA-Seq. para a montagem de genomas mitocondriais, principalmente para espécies não modelo. MOREIRA, *et al.* (2015) montou o genoma mitocondrial de três espécies de peixes do gênero *Ancistrus*, utilizando dados de RNA-Seq. Neste estudo foi usado o transcriptoma montado pelo programa Trinity, para buscar por similaridades em bancos de dados de proteínas de humanos e de *zebrafish*. O transcriptoma montado também foi comparado ao mitogenoma sequenciado de uma espécie próxima. Com essa abordagem foi montado 99.2% do mitogenoma de *Ancistrus sp. 1*, 92.5% do mitogenoma *Ancistrus sp. 2a* e 94.7% do *Ancistrus sp. 2b*. De modo geral, as regiões não montadas foram referentes a tRNAs e uma porção do D-loop (similar a região controle dos insetos), locais normalmente menos cobertos por dados de RNA-Seq. Os genomas mitocondriais de *Amolops chunganensis* e *Quasipaa boulengeri* também foram montados usando dados de RNA-Seq (YUAN *et al.*, 2016). Neste caso, os *reads* foram limpos e montados pelos programas SOAPdenovo2 e Trinity e em seguida os *contigs* foram comparados com bancos de dados de sequências próximas. Novamente uma porção da região do D-loop não foi recuperada nestes organismos. Como, estas regiões normalmente são pouco cobertas, pode ser que a etapa de limpeza feita antes da montagem do transcriptoma, possa ter influenciado na montagem destes trechos. Dado que para os dados de *Rhodnius prolixus* uma limpeza mais rigorosa antes da etapa de montagem poderia ter afetado a montagem total do mitogenoma.

Como a montagem do mitogenoma não foi obtida por uma conciliação completa das montagens M1, M2 e M3 era possível que algumas posições, provavelmente onde existissem SNPs, poderiam conter um nucleotídeo que não fosse o mais abundante nos dados de sequenciamento. A conferência das bases do mitogenoma montado feita pelo IGV mostrou que de modo geral, existiam poucas posições onde o nucleotídeo mais representado não tinha sido usado. Sendo assim, o mitogenoma-UFRJ foi

manualmente editado para conter em todas as posições a base mais frequente. Ainda não houve relatos desta verificação na literatura, mesmo em *pipelines* que indicam programas mais indicados para montagem e anotação de mitogenomas (SCHOMAKER-BASTOS; PROSDOCIMI, 2018). Porém, este cuidado conferiu mais confiabilidade a montagem do genoma de *Rhodnius prolixus*.

Cabe ressaltar que a porcentagem de *reads* que alinharam antes e depois da limpeza em ambos os mitogenomas não variou muito. A porcentagem foi de aproximadamente 19,3% para os *reads* brutos e 19,1% para os *reads* limpos a pouca diferença encontrada pode estar relacionada a exclusão de transcritos menores, devido a limpeza feita pelo Cutadapt de *reads* com menos de 90 pares de bases (pb). Aliás, a quantidade de transcritos mitocondriais está dentro da margem esperada em dados de sequenciamento de RNA, em alguns trabalhos a presença de mitotranscritos no RNA-seq chega a 30% e outros varia entre 5% e 25% (YASUKAWA; KANG, 2018).

5.2 Anotação e análise dos genomas mitocondriais de *Rhodnius Prolixus*

O genoma mitocondrial completo das duas colônias de *R. prolixus* tem 16.454 pb de comprimento, portanto, menor do que os únicos três genomas de triatomíneos sequenciados até agora, o mitogenoma de *T. infestans* 17.301 pb (PITA *et al.*, 2017), o de *T. dimidiata* 17.019 pb (DOTSON; BEARD, 2001) e o de *T. rubrofasciata* 17.150 pb (DONG *et al.*, 2018).

O mitogenoma de *R. prolixus* contém, como a maioria dos reduviídeos já sequenciados (PITA *et al.*, 2017), 13 genes codificadores de proteínas, 22 RNAs transportadores, 2 RNAs ribossômicos e uma região controle (Figura 18). Desse total de genes, 23 estão localizados na fita *majority* (J) e 14 genes na *minority* (N). Até o momento, apenas o mitoDNA de *Brontostoma Colossus* foge dessa configuração, pois contém uma duplicação de 131 pb que inclui um gene completo codificante de um RNA transportador de arginina (KOCHER *et al.*, 2014). Como consequência, *Brontostoma Colossus* possui ao todo 23 tRNAs potencialmente funcionais.

Não só o número mas também a ordem de genes mitocondriais de *R. prolixus* segue o mesmo padrão encontrado na maioria dos genomas dos reduviídeos, como em *T. infestans* (PITA *et al.*, 2017). Inclusive, esse arranjo de genes é conservado no grupo dos insetos, e compartilhado com os crustáceos, outro táxon pertencente aos artrópodes (BOORE; LAVROV; BROWN, 1998). De modo geral, eventos evolutivos de rearranjo gênico nos mitogenomas é algo relativamente raro (ZHANG *et al.*, 2015), acontecendo somente em poucos grupos mais divergentes.

No mitogenoma da colônia UFRJ de *R. prolixus* a presença dos nucleotídeos A e T foi de 69,66%, praticamente idêntico ao da colônia NIH (69,63%). A pequena diferença encontrada entre as colônias aparenta ser resultado de mutações pontuais. O material genético presente na mitocôndria está constantemente exposto a reações de oxidação devido a presença de espécies reativas de oxigênio (ROS - do inglês *Reactive Oxygen Species*) (SHARMA; SAMPATH, 2019). Os ROS podem ser formados como subprodutos do metabolismo energético mitocondrial, devido a redução incompleta do oxigênio, e são capazes reagir diretamente com a molécula de DNA (SANTOS, 2015). Como o mitogenoma está localizado no lado interno da membrana mitocondrial interna, onde estão sendo gerados ROS, o mitoDNA acumula significativamente mais danos oxidativos do que o DNA nuclear (ROUBICEK; SOUZA-PINTO, 2017; GIAMPAZOLIAS; TAIT, 2016). O radical hidroxila, por exemplo, é capaz de causar modificações nas bases nitrogenadas e/ou no açúcar, através da adição de um grupamento -OH às ligações duplas das bases nitrogenadas, do deslocamento de um átomo de hidrogênio do grupamento metil da base nitrogenada timina ou ainda pela subtração desse átomo na ligação C-H da 2'-desoxiribose (SANTOS, 2015; CADET; WAGNER, 2013). Apesar do radical hidroxila ser capaz de afetar todas as bases nitrogenadas, a oxidação da guanina é a lesão mais abundante causada por ROS (GRUBER *et al.*, 2018). Na oxidação da guanina a hidroxila gera uma modificação no nitrogênio 7 e no carbono 8 do anel desta base, formando assim a 8-oxoguanina (8-oxoG) (GRUBER *et al.*, 2018). Essa base modificada, a 8-oxoG, é capaz de mimetizar funcionalmente a timina formando um par estável com a adenina, sendo assim, a presença desta modificação pode levar a um emparelhamento errôneo, substituindo um par C-G para um A-T (MOREIRA, 2017). Oxidações das bases podem ter contribuído

para a variedade existente entre as colônias de *Rhodnius prolixus*.

A composição AT% encontrada no mitogenoma das duas colônias de *Rhodnius prolixus* é relativamente parecida com a encontrada em *T. dimidiata* e *T. rubrofasciata*, e ligeiramente menor que a de *T. infestans*. Variações na presença e frequência dos nucleotídeos em mitogenomas estão habitualmente relacionadas a características intrínsecas à replicação assimétrica e unidirecional que ocorre no mitoDNA (GODDARD; WOLSTENHOLME, 1978). Como a hipótese mais aceita, até o momento, para a replicação do mitogenoma envolve a momentânea formação de uma fita simples, isso pode propiciar desaminações espontâneas, contribuindo dessa forma para alterações no emparelhamento de bases (REYES *et al.*, 1998). Estudos já mostraram que as taxas de desaminação dos nucleotídeos A e C são maiores no DNA de fita simples do que em DNAs de dupla fita (NIKOLAOU; ALMIRANTIS, 2005).

Saccone *et al.*, (1999) mostrou que, o mitoDNA de metazoários apresenta um desbalanceamento dos nucleotídeos A+C entre as fitas, estando representados com maior frequência na fita *majority*, o que de fato aconteceu para ambos os mitogenomas estudados. A explicação para esta característica envolve mecanismos de mutação por alteração química das bases. A desaminação é o processo que acarreta a perda de um grupo amino (-NH₂) da base nitrogenada de um nucleotídeo (CHATTERJEE; WALKER, 2017). Quando essa alteração química ocorre, por exemplo, em uma adenina, é gerada uma hipoxantina que emparelha-se com uma citosina ao invés de uma timina. Ou seja, a desaminação da adenina ocasiona a troca de um par de bases A-T por um par G-C (KIM, 2018). Quando a citosina perde um grupo amino ela se transforma em uracila, que é capaz de emparelhar com uma adenina, logo, essa reação resulta na troca de um par de bases C-G por um par T-A, esta desaminação é uma das mais frequentes nos genomas (KIM, 2018). Sendo assim, a exposição da fita *minority*, geraria um aumento de citocinas e diminuição de timinas, bem como aumento de adeninas e diminuição de guaninas na síntese de uma nova fita filha (*majority*).

Ainda relacionado a frequência de nucleotídeos é importante afirmar que existe um mecanismo de reparo capaz de identificar e corrigir lesões no genoma mitocondrial, porém toda a maquinaria envolvida neste processo é codificada pelo DNA nuclear

(SANTOS, 2015). Além disso, ele não se mostra tão eficiente pois apesar dele a taxa de mutação é bem mais alta no mitoDNA do que no DNA nuclear (REYES *et al.*, 1998).

5.2.1 Genes codificadores de proteínas

Nos genes codificadores de proteínas de genomas mitocondriais é comum a presença de códons de terminação incompletos. Isso acontece como consequência do processamento do transcrito primário com o modelo de pontuação de tRNA (OJALA, 1981) que após o processamento leva a códons de término com apenas T__ ou TA_. Então nos mitogenomas a poliadenilação, além de prolongar a sobrevivência dos RNAs mensageiros, completa os códons de término nos genes. Para os mitogenomas estudados os códons de parada dos genes NAD2 e COXIII se sobrepõe às extremidades 5' do tRNAW e tRNAG, respectivamente, necessitando de complementação pela adição da cauda poli-A (Tabela 10).

O valor de uso relativo de códons sinônimos (RSCU do inglês - *Relative Synonymous Codon Usage*) é um índice que avalia o uso dos códons, ou seja, o número de vezes que um códon aparece em um gene em relação ao número de ocorrências esperadas como se todos os códons fossem igualmente utilizados. Valores de RSCU igual a 1 representam códons com utilização média, variações neste valor, para mais ou menos, demonstram maior ou menor frequências. Logo valores maiores que 1 sugerem que houve seleção positiva, e menores que 1 sugerem que houve seleção negativa do códon em questão (UDDIN; CHAKRABORTY, 2016).

A figura 19 mostrou a tendência no uso dos códons ricos em AT nos genes codificantes de proteínas dos mitogenomas de *Rhodnius prolixus*. O mesmo ocorreu para os outros três triatomíneos *T. dimidiata*, *T. infestans* e *T. rubrofasciata* também analisados na mesma figura. Este padrão já foi observado em outros grupos de insetos, dentre eles *Hymenoptera*, *Lepidoptera* e *Coleoptera* (SUN *et al.*, 2009). Ainda foi perceptível uma preferência por códons que terminem com adenina ou timina, tanto em *Rhodnius prolixus* (Tabela 11 e Figura 19), quanto nos outros três triatomíneos (Figura 19).

Os quatro códons mais usados no mitogenoma de *Rhodnius prolixus*, tanto UFRJ quanto NIH, codificaram leucina, isoleucina, fenilalanina e metionina. Eles são aminoácidos apolares e hidrofóbicos que juntos representam mais de 40% do que é codificado no mitogenoma. Este fato pode-se dever a natureza das proteínas, já que a maioria delas está associada à membrana.

5.2.2 RNAs não codificantes

Os RNAs transportadores deste estudo apresentaram uma estrutura secundária típica. O tRNAS1 mostrou uma estrutura ligeiramente diferente dos demais, seu braço D formou uma alça simples, contendo a base diidrouridina. Este padrão já foi observado em outros insetos, como *Apis mellifera* (CROZIER; CROZIER, 1993), *Anopheles gambiae* (BEARD; HAMM; COLLINS, 1993) e *T. rubrofasciata* (DONG *et al.*, 2018).

5.2.3 Regiões não codificantes

No mitogenoma de *R. prolixus* foram encontradas 11 sobreposições gênicas, a maioria envolvendo poucos nucleotídeos, a maior com 21 entre tRNAP e NAD6 (Tabela 10). Também foram identificados 13 espaçadores intergênicos (IGS do inglês - *InterGenic Spacer*), o maior deles localizado entre os genes tRNAS2 e NAD1 com 78 pb. Essa região também foi encontrada no mitogenoma de outros insetos, como *Apis mellifera*, porém sua função ainda é desconhecida (CROZIER; CROZIER, 1993). Todavia, em algumas espécies de insetos, como em *T. dimidiata*, esse IGS contém uma fase de leitura não identificada (*ORF do inglês - Open reading frame*) codificando um peptídeo de 103 aminoácidos. Entretanto a literatura afirma que essa sequência não é similar a nenhuma outra depositada no *GenBank* (DOTSON; BEARD, 2001). Apesar desse IGS ter sido o maior encontrado no mitogenoma de *R. prolixus*, não houve evidências de ORF, nem de repetições em tandem. Além disso, esse espaçador foi o menor entre os triatomíneos: *T. dimidiata* (314 pb), *T. rubrofasciata* (182 pb) e *T.*

infestans (94 pb) (DOTSON; BEARD, 2001; DONG *et al.*, 2018; PITA *et al.*, 2017).

O mitogenoma de *Rhodnius prolixus* se mostrou bem compacto com poucas regiões não codificantes, se resumindo aos espaçadores intergênicos e a região controle. Em *R. prolixus* a região controle (CR do inglês - *Control Region*) apresenta 1.791 pb de comprimento, a menor dentre os triatomíneos sequenciados, o que colabora para seu genoma ser o menor até o momento. Essa porção do mitogenoma de insetos habitualmente é mais variável, exibindo altas taxas de substituição de nucleotídeos, inclusive o comprimento desta região pode variar entre espécies e/ou indivíduos (LI; LIANG, 2018). A região controle em *Rhodnius prolixus* está disposta da mesma maneira que nos outros triatomíneos, isto é, há uma clara divisão em quatro partes diferentes (DOTSON; BEARD, 2001).

O tamanho do primeiro trecho é relativamente parecido nesses insetos, em *R. prolixus* 447 pb, *T. dimidiata* 450 pb, *T. infestans* 450 pb e *T. rubrofasciata* 467 pb. Ela apresenta uma proporção maior de guaninas e citosinas do que o restante da sequência e em *R. prolixus* não houve diferença relevante entre as duas colônias. No final deste trecho foi identificada o bloco de guaninas característico também presente nos outros barbeiros (DOTSON; BEARD, 2001), entretanto, *T. infestans* e *T. rubrofasciata* possuem uma guanina a menos (PITA *et al.*, 2017; DONG *et al.*, 2018).

O segundo trecho, rico em A + T, tem 274 pb em *R. prolixus*, nas duas colônias. Esse trecho foi ligeiramente menor do que nos outros triatomíneos já sequenciados (*T. dimidiata* com 400 pb, *T. infestans* com 430 pb e *T. rubrofasciata* com 436 pb). Ao comparar as sequências deste trecho de todos estes triatomíneos, não foi observado nenhum bloco conservado.

O terceiro trecho, de 1028 pb contém as repetições em tandem. Em *Rhodnius prolixus* foram encontradas oito repetições, a primeira sequência apareceu cinco vezes (a última sequência ligeiramente menor que as demais) e a segunda três vezes. Entre as colônias de *R. prolixus* foi encontrada uma pequena diferença entre as sequências repetitivas. . Em *T. dimidiata*, já foram encontradas oito repetições tandem. Todavia, em *T. dimidiata* existem três sequências diferentes, a primeira sequência tem 82 bp, a

segunda tem 140 pb e é reproduzida cinco vezes e a terceira tem 173 pb e aparece duas vezes (DOTSON; BEARD, 2001). Já o trecho em tandem de *T. rubrofasciata* contém uma sequência de 152 pb cinco vezes repetida e outra de 198 pb que aparece duas vezes (Dong *et al.*, 2018). O tandem de *T. infestans* é o maior entre os barbeiros, apresenta doze repetições, um bloco de 125 pb está presente 8 vezes consecutivas e outro de 182 bp aparece três vezes de forma completa e uma quarta parcialmente (PITA *et al.*, 2017). Aparentemente, a variação neste trecho da região controle influenciou no tamanho total do mitogenoma, uma vez que foi o segmento com maior discrepância de comprimento nos barbeiros. O triatomíneo com a menor região em tandem também foi o que apresentou o menor mitogenoma, e isto também foi válido para o maior, *Rhodnius prolixus* e *Triatoma infestans* respectivamente. Esta relação foi vista também em outros hemípteros. Um estudo feito com 116 regiões controles, com espécies de 11 infraordens de Hemiptera, encontrou uma correlação positiva entre o comprimento total de repetições em tandem e a sequência da região controle (LI; LIANG, 2018).

O último trecho da região controle, tem apenas 42 pb em *Rhodnius prolixus*, a semelhança entre as duas colônias foi bem grande, poucos nucleotídeos divergiram. Este trecho é um pouco maior nas outras espécies da subfamília *Triatominae*. Em *T. infestans* tem 134 pb, em *T. rubrofasciata* 168 pb e *T. dimidiata* tem 188 bp. Ao comparar as sequências de todos esses triatomíneos, foi possível identificar uma conservação dos nucleotídeos que compõem esta porção.

5.2.4 Origem de replicação

A identificação e caracterização da região controle é de suma importância pois acredita-se que é onde possivelmente estejam localizadas as origens de replicação das duas fitas do mitogenoma (OJ - Origem da fita *majority* e ON- Origem da fita *minority*), pois é o que acontece em espécies do gênero *Drosophila*, *Tribolium castaneum* e *Bombyx mori* (SAITO; TAMURA; AOTSUKA, 2005). Nos hemípteros ainda não foi descrito na literatura as origens de replicação, nem a presença de blocos conservados que de fato possam caracterizar OJ e ou ON. Apesar da região controle aparentemente estar organizada de maneira conservada nos triatomíneos isso não se estende para

outros membros da ordem Hemiptera. As sequências que constituem a região controle podem ou não ser formadas por vários elementos estruturais, como trechos mononucleotídeos, repetições em tandem e regiões ricas em determinados nucleotídeos (LI; LIANG, 2018).

Baseado nisto foi realizada a busca por elementos conservados na região controle dos triatomíneos que possam dar indícios das origens de replicação nestes insetos. A partir do conhecimento obtido da origem replicativa de outros animais, como *Drosophila* e *Locusta migratoria* (SAITO; TAMURA; AOTSUKA, 2005), os elementos buscados foram ou sequências de timinas ou aquelas capazes de formar estruturas em forma de grampo. Sendo assim, foi feito o alinhamento múltiplo do primeiro, segundo e quarto trecho da região controle de *Rhodnius prolixus* (ambas as colônias), *T. dimidiata*, *T. rubrofasciata* e *T. infestans* (DOTSON; BEARD, 2001; DONG *et al.*, 2018; PITA *et al.*, 2017). A terceira porção, que contém as repetições em tandem, é muito variável nos barbeiros por isso aparentemente não mantém a conservação necessária para a manutenção deste mecanismo.

A busca pela origem de replicação da fita *minority* mostrou 3 possíveis candidatos no trecho de 447 pb rico em G+C do mitogenoma de triatomíneos. O primeiro bloco vai da posição 139 até 154, contendo 4 timinas seguidas. O segundo é menor, e vai da posição 360 até 373, só contém 3 timinas seguidas (Figura 24 e 38). Estes dois primeiros blocos foram selecionados por conter uma região conservada entre os triatomíneos, que é formada por adeninas e timinas. O último bloco está imediatamente antes da sequência de guaninas, apesar de não ter a sequência de timinas nem o grampo, ele apresentou 18 nucleotídeos conservados além das guaninas.

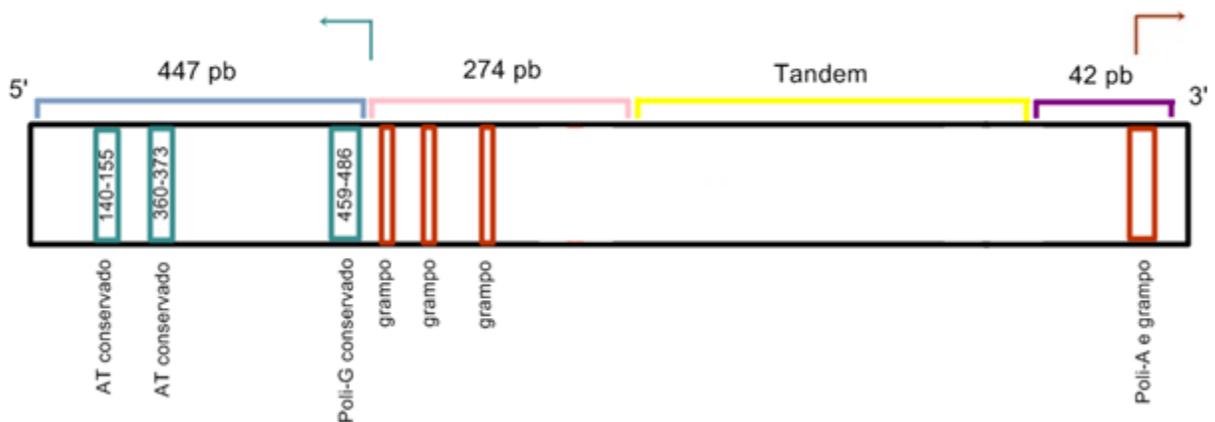


Figura 38 - Esquema das possíveis origens de replicação na região controle dos triatomíneos. Blocos relacionados a origem de replicação da fita *minority* (ON) estão representados em verde, enquanto os da fita *majority* (OJ) estão em vermelho. As ON e OJ mais prováveis estão indicadas com setas em verde e vermelho que apontam para o sentido da replicação. Todos os elementos nesta figura estão representados na fita *majority*.

A busca pela origem de replicação da fita *majority* (OJ), em *Rhodnius prolixus*, e nos outros triatomíneos analisados, mostrou que o quarto trecho (de 42 pb) possui um bloco com grande potencial para ser OJ. Este bloco com oito adeninas, que na fita complementar, *minority*, representam uma sequência de 8 timinas, foi identificado dentro de uma sequência capaz de formar uma estrutura em grampo (Figura 26 e 27). No gênero *Drosophila*, OJ se encontra nesta mesma região, a poucos nucleotídeos de distância do gene do tRNAI (SAITO; TAMURA; AOTSUKA, 2005), assim como em *R. prolixus*. Sendo assim, o grampo e/ou o poli-T, na fita *minority*, podem estar envolvidos na sinalização para o início da replicação da fita *majority*, representando OJ. Além disso, também foram achados outros três grampos no trecho de 274 pb rico em A+T, porém eles são de tamanhos variados e não houve uma conservação na sequência de nucleotídeos entre os triatomíneos. Isto reforça a ideia do trecho encontrado na região de 42 pb ser o mais promissor.

Em *homo sapiens*, e outros vertebrados, a origem de replicação da fita *light* (equivalente a fita *majority* de insetos) não se encontra na região controle e sim em um espaço inter-gênico (IGS) (FALKENBERG, 2018). Como dito anteriormente, o genoma mitocondrial de *Rhodnius prolixus* é bastante compacto, logo, tem IGS pequenos e em

pouca quantidade. Além disso, o maior espaçador IGS de *Rhodnius prolixus*, equivalente ao humano que contém a origem de replicação da fita *light*, não contém poli-T e nem forma grampos.

5.3 Busca de mitogenomas, alinhamento múltiplo e construção da árvore filogenética

Em relação a árvore filogenética das sequências codificadoras de proteínas mitocondriais (Figura 29), o agrupamento das duas linhagens de *R. prolixus* ocorreu como esperado. Porém como não há mais representantes disponíveis da Tribo Rhodiini não se pode inferir sobre a parafilia ou monofilia do grupo. As duas linhagens de *R. prolixus* formaram um grupo com os membros da tribo Triatomini, com suporte de 100%. Em relação tribo Triatomini, na árvore também não ficou evidente a monofilia desta tribo. As poucas amostras disponíveis para as duas tribos (Rhodiini e Triatomini) dificultam a solução das questões referente ao surgimento delas, assim como da subfamília Triatominae. Os apontamentos a respeito da evolução da subfamília Triatominae já foram extensamente debatidos na introdução deste trabalho.

Todos os triatomíneos e os membros de Stenopodainae (*Oncocephalus breviscutum* e *Canthesancus helluo*) foram agrupados em um clado com bom suporte, o que apoia que a ideia de Triatominae e Stenopodainae serem grupos irmãos. Essa proximidade entre os dois grupos já foi evidenciada por estudos baseados em características morfológicas e análises filogenéticas de seqüências mitocondriais e nucleares (CAMERON *et al.*, 2012; WEIRAUCH, 2008). As espécies pertencentes à subfamília Peiratinae, como esperado, formaram um clado com um bom suporte. Essa subfamília é grupo irmã de outras quatro subfamílias, Harpactorinae, Reduviinae, Stenopodainae e Triatominae (ZHANG *et al.*, 2016). O clado formado por essas subfamílias é grupo irmão de Ectrichodiinae, conhecidos por se alimentarem de miriápodes e isópodos (ZHANG *et al.*, 2016). Reduviinae aparenta ser uma subfamília polifilética, a árvore filogenética aponta uma relação entre *Valentia hoffmanni*, classificada na literatura na subfamília *Salyavatinae* (TRUONG *et al.*, 2007), e os dois

gêneros de Reduviinae (*Inara* e *Acanthaspis*) com suporte de nó de 66%. A relação entre Salyavatinae e Reduviinae já foi abordada em alguns trabalhos (ZHANG *et al.*, 2016).

5.4 Nível de expressão dos transcritos mitocondriais

Ainda há questões indefinidas acerca da transcrição nos mitogenomas de insetos. Um dos poucos consensos é que a transcrição gera um precursor policistrônico que, em seguida, precisa ser processado para que cada RNA se torne funcional (D'SOUZA; MINCZUK, 2018). Este processamento se inicia pela retirada dos tRNAs, modelo de pontuação de tRNA (do inglês tRNA punctuated model) (OJALA *et al.*, 1981).

O índice de expressão gerado neste estudo para as sequências codificantes do genoma mitocondrial de *Rhodnius prolixus* colônia UFRJ, não encontrou evidências para a expressão dos RNAs transportadores. Isto pode estar associado à modificação pós-transcricional que ocorre nessas moléculas, após a pontuação de tRNA, essas estruturas não são poliadeniladas (BERTHIER *et al.*, 1986). Uma vez que, um dos processos necessários para o sequenciamento do RNA é a construção da biblioteca de cDNA, onde um dos passos envolve a ligação a caudas poli-A presentes em RNAs maduros, isto impede a quantificação da expressão dessas moléculas. Entretanto, foi possível montar completamente o mitogenoma, o que incluiu a identificação dos tRNAs nos *reads* do transcriptoma, que pode ser consequência da visualização de transcritos primários ainda não completamente processados. Estes fatos podem ser reforçados pela figura 17 que mostra a baixa cobertura na região dos tRNAs. Por outro lado, os outros RNAs (mRNA e rRNA) recebem a cauda poli-A e não enfrentam a mesma situação (STEWART; BECKENBACH, 2009).

Os tratamentos utilizados parecem provocar mudanças no nível de expressão dos genes codificantes de proteínas mitocondriais. Dois efeitos majoritários foram observados causados pela infecção e alimentação. A infecção das ninfas de *Rhodnius prolixus* com tripomastigotas de *T. cruzi* geraram um aumento de expressão dos genes

NAD6 e NAD2, do complexo I, e ATP8, uma subunidade da ATP-sintase (Figura 30). Os complexos afetados pelo aumento da expressão de NAD6, NAD2 e ATP8 estão intrinsecamente ligados a produção de energia mas também podem levar a produção de radicais livres. O complexo I catalisa a oxidação do NADH e, como consequência dessa reação, transfere quatro prótons da matriz para o espaço intermembrana (CAVALCANTE, 2015). ECHTAY *et al.* (2002), salienta que o estado redox da cadeia transportadora de elétrons e a força próton motriz, gerada pelo transporte de prótons da matriz para o espaço intermembranas pelos complexos I, III e IV são preponderantes para a geração de ROS nas mitocôndrias. Já o complexo ATP-sintase é o local por onde os prótons bombeados para o espaço intermembrana retornam para a matriz mitocondrial, a favor do gradiente de concentração, e terminam com a produção de ATP. Desta forma o aumento da expressão dessas proteínas pode desempenhar um papel importante na resposta de *Rhodnius prolixus* contra a infecção pelo protozoário *T. cruzi*. Na literatura já foi mostrado que cepas de *Anopheles gambiae* refratárias ao protozoário *Plasmodium* apresentavam altos níveis sistêmicos de espécies reativas de oxigênio (ROS). Quando era realizada a suplementação dietética com antioxidantes, a capacidade desse mosquito de responder ao patógeno reduzia drasticamente (MOLINA-CRUZ *et al.*, 2008). Outro autor (DIAZ-ALBITER *et al.*, 2012), também mostrou a relação de ROS e de patógenos em insetos. O estudo feito com flebotomíneos da espécie *Lutzomyia longipalpis*, mostrou que ROS são prejudiciais ao parasito *Leishmania mexicana*, resultando na diminuição da população do patógeno. Logo, o aumento da expressão dessas proteínas mitocondriais pode ocasionar no aumento de espécies reativas de oxigênio (ROS), que podem ser uma tentativa de lidar com a carga de *T. cruzi*.

O mesmo pode estar ocorrendo no tratamento com epimastigotas, na maior concentração (10^7) e no maior tempo (12 dias), onde há um aumento de expressão de quase todos os genes do complexo I (NAD1, NAD4, NAD4L, NAD5 e NAD6). Entretanto, neste caso a resposta se deu somente na maior concentração e tempo de exposição sugerindo que a resposta possa ser consequência da intensidade de exposição. Existem outras situações onde a quantidade de microorganismos precisa ser controlada, por exemplo o microbioma intestinal normalmente é densamente

populado, entretanto, é necessário que haja um controle da microbiota existente. A produção de ROS é um dos mecanismo utilizados por insetos para o controle da microbiota intestinal (KWONG; MANCENIDO; MORAN, 2017). No próprio epitélio do intestino dos insetos há enzimas responsáveis pela produção de ROS. O aumento da expressão visto na infecção da maior concentração por epimastigotas, pode indicar que a indução do aumento de ROS produzido pelo complexo I seja parte do mecanismo de controle populacional de microorganismos que ocorre no trato intestinal(KWONG; MANCENIDO; MORAN, 2017).

A diminuição de expressão dos genes rRNAS, rRNAL e NAD1 logo após a alimentação com sangue pode ser uma estratégia para diminuir as espécies reativas de oxigênio, diminuindo a transcrição das proteínas dos complexos. Já é sabido que a alimentação hematofágica acaba induzindo um ambiente de estresse oxidativo em *Rhodnius prolixus* (VIJA-SUAREZ; ORTIZ; MOLINA, 2017). A hemoglobina, presente no sangue, é formada por quatro subunidades de globina ligadas ao grupamento prostético heme (SHINOTSUKA, 2014). Quando a hemoglobina é digerida no intestino médio do barbeiro, resulta na produção de grandes quantidades de heme, molécula citotóxica e pró-oxidante (GRAÇA-SOUZA *et al.*, 2006). O heme tem potencial de promover reações de formação de ROS, o que o torna tóxico quando livre (SHINOTSUKA, 2014). A diminuição dos genes NAD3, CYTB e COXII, participantes respectivamente dos complexos I, III e IV, ao longo dos dias após a alimentação com sangue sugere uma tentativa progressiva de diminuir a formação de ROS.

5.4 Comparação do mitogenoma de *R. prolixus*, colônia UFRJ, com sequências da literatura para as espécies *R. prolixus* e *R. robustus*.

O estudo de padrões ecológicos biogeográficos é de grande valia para organismos como os Triatomíneos. A identificação desses padrões pode fornecer fundamentos que auxiliem na definição de faixas geográficas que contenham espécies vetorais mais eficientes (ABAD-FRANCH *et al.*, 2009). Que por sua vez pode ser empregada para avaliar o risco de transmissão da Doença de Chagas, como já foi feito

para outras doenças (RIPOCHE *et al.*, 2018; MURRAY *et al.*, 2018). Para este propósito, a caracterização do local de ocorrência dos membros de um grupo, conhecimento acerca da origem evolutiva, assim como a correta identificação das espécies é algo extremamente benéfico.

A comparação feita entre as sequências geradas neste estudo e as de *Rhodnius prolixus* e *Rhodnius robustus* disponíveis na literatura foi feita para possivelmente elucidar a origem do mitogenoma de *R. prolixus* da colônia UFRJ. Devido a proximidade de *Rhodnius prolixus* com *Rhodnius robustus* este último também foi inserido nas análises. Na literatura já houve relatos de introgressão do material genético desses barbeiros, ou seja, a ocorrência do fluxo de genes entre essas espécies (FITZPATRICK *et al.*, 2008). Esta ocorrência é relevante pois algumas linhagens dos dois grupos conseguem cruzar e gerar descendentes, e devido a herança mitocondrial ser uniparental, o acervo genético mitocondrial de um organismo pode ter origem diferente da sequência nuclear.

As sequências usadas foram referentes aos genes do citocromo b e da subunidade maior do RNA ribossomal de *R. prolixus* e de *R. robustus*. As sequências de *Rhodnius prolixus* UFRJ, tanto do gene do citocromo b como da subunidade maior do RNA ribossomal, apresentaram identidade maior com as sequências de mesma espécie, oriundas da literatura. As diferenças encontradas foram proporcionalmente pequenas, e podem estar associadas a mutações pontuais. Além disso, a diferença encontrada está dentro da faixa de variação intraespecífica esperada para este grupo, alguns autores apontam máximo de 2% (PAVAN, 2013). Em contrapartida, ambos os genes CYTB e rRNAI deste estudo apresentaram uma identidade menor quando comparados com *R. robustus*. Contudo, vale ressaltar que três exemplares de *R. robustus* coletados em Trujillo na Venezuela aparentam ser menos diferentes de *R. prolixus* (UFRJ e NIH), o que pode indicar que a divergência evolutiva entre essas linhagens foi mais recente.

O alinhamento feito com o rRNAI mostrou ainda que possivelmente uma sequência, coletada em Madre de Dios no Peru e estabelecida em colônia (VENDRAMI *et al.*, 2008), está identificada equivocadamente. A identidade dela com a de *R. prolixus*

UFRJ foi de 100%, o que sugere que esta amostra está incorretamente identificada ou que houve uma introgressão do material genético de *R. prolixus* naquela colônia, como suposto pelos próprios autores.

A maior conservação do gene da subunidade maior do RNA ribossomal já era algo esperado, frente o CYTB. Por esta razão, o gene citocromo b é habitualmente escolhido para distinguir espécie filogeneticamente muito próximas. Enquanto o gene do rRNAI é mais indicado para separar espécies evolutivamente mais distantes, como por exemplo, de filos diferentes (PAVAN, 2013).

Como o gene do citocromo b tem uma taxa de mutação mais alta do que da subunidade maior do RNA ribossomal, o que o torna a melhor escolha entre as duas opções para distinguir espécie muito próximas, suas sequências foram escolhidas para a construção da árvore filogenética (Figura 35). O gênero *Rhodnius* é dividido em duas grandes linhagens evolutivas: os cis-andinos incluindo os grupos *prolixus* e *pictipes*, distribuídos a leste da Cordilheira dos Andes; e os trans-andinos com o grupo *pallescens*, a oeste dos Andes (trans-andino). Para análise da relação de *R. prolixus* e *R. robustus* nesta árvore filogenética foram escolhidos para grupo externo sequências de *Rhodnius pallescens*, *Rhodnius pictipes* e *Rhodnius barreti*, que até recentemente era classificado como *R. robustus* (ABAD-FRANCH *et al.*, 2013).

Os dados adquiridos pela construção da árvore filogenéticas do gene do citocromo b desses membros do gênero *Rhodnius* (Figura 35) ratificam o que havia sido constatado em trabalhos anteriores. O grupo externo da árvore enraizou formando duas linhagens evolutivas de *R. pallescens*, com um bom valor de suporte. A sequência FJ229359, proveniente de Santa fé, Veraguas, no Panamá, forma um clado separado das outras duas amostras de *R. pallescens*. O segundo clado é formado por espécimes coletados ao norte da Colômbia: JQ686675, oriunda de San Fernando, Bolívar; e JQ686685 proveniente de El carmen, Norte de Santander. A literatura já descreveu a existência de duas linhagens para *R. pallescens*, a sequência do primeiro clado, usado neste trabalho, faz parte da linhagem centro-americana (que agrupa indivíduos do oeste da Colômbia e oeste do Panamá), o segundo clado é descendente da linhagem colombiana (que inclui populações localizadas ao norte e ao sul da Colômbia)

(GÓMEZ-PALACIO *et al.*, 2012). Estudos apontam que a dispersão do ancestral dessas linhagens a partir do local de origem deste gênero provavelmente ocorreu no Mioceno (ABAD-FRANCH *et al.*, 2009). Outro membro do grupo externo é *Rhodnius pictipes* (MF497699), esta espécie faz parte de um complexo que leva seu nome e inclui ainda *R. amazonicus*, *R. brethesi*, *R. paraensis*, *R. stali* e *R. zeledoni* (JUSTI; GALVÃO, 2017). Ainda há controvérsias quanto a organização basal da tribo Rhodniini (MONTEIRO *et al.*, 2018), todavia, a sequência de *R. pictipes* utilizada neste trabalho, compartilhou mais traços com o complexo *prolixus*, como a separação das duas linhagens evolutivas já citadas anteriormente (JUSTI; GALVÃO, 2017). O último táxon, *Rhodnius barretti* (JX273159 e JX273161) na árvore ficou próximo do grupo formado por *R. prolixus* e *R. robustus*. De fato, estudos como o de Pavan, 2013, apontam para uma divergência mais recente desta espécie, datadas no Mioceno Superior/Plioceno.

Na árvore filogenética produzida neste estudo, *R. prolixus* e *R. robustus* formaram clados distintos, com um bom valor de suporte (88%). As espécies de *Rhodnius robustus* formaram claramente 5 grupos distintos. Os membros coletados na região do Orinoco (*R. robustus I*) formam um clado mais próximo de *R. prolixus* do que do restante das espécies de *R. robustus* (provenientes da região amazônica). Apesar da sequência *R. robustus V* ter sido agrupada com *R. robustus I*, sua posição na árvore permanece incerta. Segundo ABAD-FRANCH *et al.* (2013), ela é proveniente de um espécime mal conservado coletado no Brasil, e devido a qualidade da amostra não foi possível fazer a verificação do genoma nuclear. Logo, não foi feita a confirmação desta espécie, deste modo, neste espécime pode ter ocorrido introgressão do genoma mitocondrial. A quarta sequência proveniente da Venezuela, diferente dos anteriores, veio de uma coleta a leste da cordilheira dos andes, isto pode ter contribuído para esta sequência ter agrupado com o grupo amazônico *Rhodnius robustus IV*. Todas as sequências de *R. prolixus* foram agrupadas num único clado, com um valor de suporte razoável (60%). Os grupos formados no interior deste clado não apresentaram um valor de suporte adequado para validar a separação desta espécie em diferentes conjuntos, como ocorreu em *R. robustus*. Todavia, as duas colônias de *Rhodnius prolixus* empregadas neste estudo foram agrupadas dentro deste clado de espécies de *R. prolixus*. Estas observações reforçam o estudo de Monteiro *et al.* (2003), onde foi

evidenciado a distinção genética de *R. prolixus* e *R. robustus*, e ainda, que enquanto o primeiro formava um grupo monofilético, o segundo forma um complexo de espécies parafiléticos.

5.5 Descoberta e caracterização de polimorfismos de nucleotídeo único, inserções e deleções.

A diferenciação genética de espécies é uma ferramenta amplamente utilizada nos estudos taxonômicos. O uso de marcadores moleculares em membros da subfamília Triatominae, tem auxiliado na resolução de impasses taxonômicos (GARDIM *et al.*, 2014). Por exemplo, a distinção de *Rhodnius prolixus* e *R. robustus* como espécies distintas só foi confirmado após estudos baseados em dados moleculares. Os polimorfismos de nucleotídeos únicos (SNPs) têm sido utilizado como marcadores moleculares em vários estudos do gênero *Rhodnius* (PAVAN *et al.*, 2013), por exemplo buscando SNP nucleares com potencial para serem usados para separar *R. prolixus* do complexo de espécies formados por *R. robustus*.

Sendo assim, neste estudo foram utilizados dados oriundos de sequenciamento de RNA de segunda geração (*Illumina*) para descobrir variações presentes no mitogenoma de *Rhodnius prolixus* - UFRJ. Essas variações podem ser úteis na diferenciação de sequências mitocondriais entre *R. prolixus* e o complexo de espécies formados por *R. robustus*, assim como outros membros do gênero.

Dos 25 SNPs identificados no mitogenoma da colônia-UFRJ, 18 estão incluídos em regiões de genes codificadores de proteínas. Em 15 dos SNPs identificados o uso do nucleotídeo alternativo, representou uma mutação sinônima, o que não afeta o aminoácido final, entretanto em 5 houve a alteração do aminoácido, chamada de mutação não sinônima. Todos os SNPs oriundos de mutações não sinônimos do mitogenoma-UFRJ envolveram a troca de um aminoácido não polar por outro de mesma natureza. Logo, as mudanças foram consideradas toleráveis para as proteínas, os valores do SIFT score (Tabela 18) ratificaram isso. É interessante notar que todos os SNPs da colônia UFRJ estavam em códons para aminoácidos apolares. Isto sugere

a ideia de que os genes codificadores de proteínas do mitogenoma de *Rhodnius prolixus*, podem sofrer pressão seletiva positiva para o uso de códons de aminoácidos apolares/hidrofóbicos devido a localização transmembranar da maioria dessas proteínas. O único SNP identificado na região de um RNA transportador aparentemente não afeta o desempenho de sua função pois a mudança, não afeta os pareamentos responsável pela sua estrutura tridimensional e nem códon-anticódon, se localizando na alça D (Figura 39).

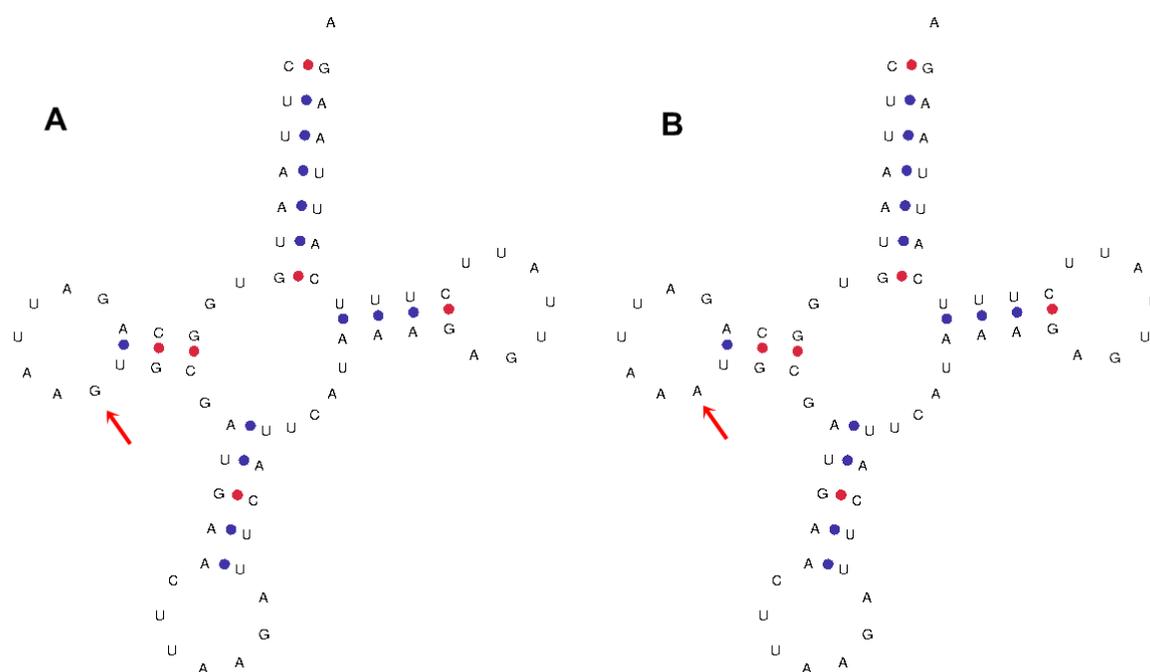


Figura 39 - SNP presente no tRNAL2 do mitogenoma *Rhodnius prolixus*, colônia UFRJ. A: tRNAL2 com nucleotídeo original. B: tRNAL2 com nucleotídeo alternativo. Setas vermelhas apontam o nucleotídeo.

Todos os 25 SNPs encontrados tem potencial para serem usados como marcadores moleculares para distinguir espécies de *Rhodnius*. Todavia, cabe ressaltar que devido ao modelo de pontuação de tRNA (do inglês *tRNA punctuated model*) (OJALA *et al.*, 1981) e a dificuldade de amplificação desses genes, o SNP no tRNAL2 dentre as opções é o menos indicado. A escolha entre os demais genes deverá ser feita pensando nos grupos a serem estudados, genes com altas taxas de mutação devem ser usados para espécies filogeneticamente próximas enquanto os com taxas

mais baixas são mais recomendados para diferenciar táxons mais distantes (Pavan, 2013).

Foram identificados dois indels na colônia UFRJ, uma deleção no gene do rRNAs e uma inserção na região controle. Aparantemente, os dois indels encontrados no mitogenoma UFRJ não afetam a função desempenhada pela região gênica onde estão localizados. Uma vez que variações no rRNAs já foram relatadas em organismos próximos a *Rhodnius prolixus*, tais como *Triatoma infestans* e *Drosophila pseudoobscura* (GARCÍA *et al.*, 2003; NOOR; LARKIN, 2000), e o indel da região controle não se encontra nos blocos conservados relatados anteriormente. Se um indel fosse relatado num gene codificador de proteína, isso poderia potencialmente afetar o desempenho da proteína.

Entretanto, a região onde foram encontrados os indels podem os tornar difíceis de serem usados como marcadores moleculares para distinguir espécies de *Rhodnius*. Isto, pois o gene que codifica o rRNAs é aparentemente bem conservado, por isto ele é mais indicado para distinguir espécies mais distantes (PAVAN, 2013). E apesar de alguns estudos defenderem o uso da região controle para comparação entre espécies, a literatura contém pouca informação desta porção do mitogenoma (LI; LIANG, 2018). Logo, isto dificultaria seu uso em trabalhos de classificação taxonômica.

As poucas sequências de *Rhodnius robustus* disponíveis na literatura não cobriram as regiões contendo SNPs e Indels no mitogenoma-UFRJ, logo, não foi possível verificar se esses variantes eram espécie-específicos.

6. CONCLUSÃO

- O genoma mitocondrial de *Rhodnius prolixus* está organizado de forma semelhante à maioria dos Reduviídeos já sequenciados. Esta espécie tem o menor genoma dentre os membros da subfamília Triatominae, mas contém os mesmos genes que os demais.
- O uso de códons do genoma mitocondrial de *Rhodnius prolixus* e dos triatomíneos não é homogêneo, mostrando códons preferenciais. Existe uma preferência por nucleotídeos A/T nas duas últimas posições dos códons.
- Os diferentes tratamentos utilizados no transcriptoma, infecção com *T. cruzi* e hematofagia, afetam a expressão das proteínas mitocondriais de *Rhodnius prolixus*.
- A sequência mitocondrial do mitogenoma UFRJ é mais similar às de *Rhodnius prolixus* da literatura.

7. REFERÊNCIAS

ABAD-FRANCH, F. *et al.* Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: A multi-scale appraisal of the tribe Rhodniini (Triatominae). **Acta Tropica**, v. 110, n. 2–3, p. 159–177, 1 maio 2009.

_____. *Rhodnius barretti*, a new species of Triatominae (Hemiptera: Reduviidae) from western Amazonia. **Memórias do Instituto Oswaldo Cruz**, v. 108, n. suppl 1, p. 92–99, 2013.

ALBERTS, B.; JOHNSON, A.; WALTER, P. **Biologia Molecular da Célula**. 5^a Ed ed. [s.l: s.n.].

ALMEIDA, THAYNARA SARMENTO OLIVEIRA DE; ALMEIDA, THASSIANY SARMENTO OLIVEIRA DE; RAMALHO, S. N. L. **Revista InterScientia**. [s.l: s.n.]. v. 5

ALTSCHUL, S. F. *et al.* **Basic local alignment search tool** *Journal of Molecular Biology* Academic Press, , 5 out. 1990. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0022283605803602?via%3Dihub>>. Acesso em: 4 jun. 2019

AMARAL, G. C. *et al.* DOENÇA DE CHAGAS: BREVE REVISÃO COM ÊNFASE NO TRATAMENTO. **Revista Corpus Hippocraticum**, v. 1, n. 1, 5 out. 2017.

ANDERSON, S. *et al.* Sequence and organization of the human mitochondrial genome. **Nature**, v. 290, n. 5806, p. 457–465, abr. 1981.

AYO, C. M. *et al.* Genetic susceptibility to Chagas disease: an overview about the infection and about the association between disease and the immune response genes. **BioMed research international**, v. 2013, p. 13, 2013.

BANKEVICH, A. *et al.* Original Articles SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. 2012.

BARBOSA, J. T. DE S. **Genômica comparativa e filogeográfica em espécies do gênero *Atta* (Hymenoptera: Formicidae)**. [s.l.] Centro de Ciências Agrárias, Programa de Pós-Graduação em Proteção de Plantas, 28 fev. 2018.

BEARD, C. B.; HAMM, D. M.; COLLINS, F. H. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. **Insect molecular biology**, v. 2, n. 2, p. 103–24, 1993.

BELINATO, M. R. F. **Efeito dos parasitos *Trypanosoma cruzi* e *Trypanosoma rangeli* em aspectos do fitness de *Rhodnius prolixus***. [s.l.] Centro de Pesquisas René Rachou, Fundação Oswaldo Cruz, Belo Horizonte, 2014.

BERTHIER, F. *et al.* RNA mapping on *Drosophila* mitochondrial DNA: precursors and template strands. **Nucleic acids research**, v. 14, n. 11, p. 4519–33, 11 jun. 1986.

BESTETTI, R. B.; RESTINI, C. B. A. Precordial chest pain in patients with chronic Chagas disease. **International journal of cardiology**, v. 176, n. 2, p. 309–14, 20 set. 2014.

BOORE, J. L.; LAVROV, D. V.; BROWN, W. M. Gene translocation links insects and crustaceans. **Nature**, v. 392, n. 6677, p. 667–668, abr. 1998.

BOOTH, M.; CLEMENTS, A. Neglected Tropical Disease Control - The Case for Adaptive, Location-specific Solutions. **Trends in parasitology**, v. 34, n. 4, p. 272–282, 1 abr. 2018.

BRASIL. **Guia de Vigilância Epidemiológica**. Brasília: [s.n.].

CADET, J.; WAGNER, J. R. DNA base damage by reactive oxygen species, oxidizing agents, and UV radiation. **Cold Spring Harbor perspectives in biology**, v. 5, n. 2, 1 fev. 2013.

CAMERON, S. L. *et al.* A mitochondrial genome phylogeny of termites (Blattodea: Termitoidae): Robust support for interfamilial relationships and molecular synapomorphies define major clades. **Molecular Phylogenetics and Evolution**, v. 65, n. 1, p. 163–173, 1 out. 2012.

_____. Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny. **Annual Review of Entomology**, v. 59, n. 1, p. 95–117, 7 jan. 2014.

CARVALHO, T. B. **Avaliação de SNPs (Single Nucleotide Polymorphisms) nas diferentes formas clínicas da doença de Chagas**. [s.l.] Universidade Estadual Paulista (UNESP), 23 fev. 2018.

CAVALCANTE, S. K. M. **Estudos comparativos da bioenergética mitocondrial do *Lophosilurus alexandri***. [s.l.] Instituto de Química e Biotecnologia, Programa de Pós-Graduação em Química e Biotecnologia, 25 jan. 2015.

CDC. **Parasites - American Trypanosomiasis**.

CHATTERJEE, N.; WALKER, G. C. Mechanisms of DNA damage, repair, and mutagenesis. **Environmental and molecular mutagenesis**, v. 58, n. 5, p. 235–263, 2017.

CHINNERY, P. F.; HUDSON, G. Mitochondrial genetics. **British medical bulletin**, v. 106, n. 1, p. 135–59, 2013.

CONTRERAS, H. D. P. **Genome wide analysis of the expression of the gut of *Rhodnius prolixus* after a blood meal and infection with the parasite *Trypanosoma cruzi***. [s.l.] UNIVERSIDADE FEDERAL DE RIO DE JANEIRO, 2015.

COSTA, M. *et al.* **DOENÇA DE CHAGAS: UMA REVISÃO BIBLIOGRÁFICA**. [s.l.] Shogakukan, 2013. v. 2

COURA, J. R. Tripanosomose , doença de Chagas. **Ciência e Cultura**, v. 55, n.

4, p. 30–33, 2003.

CROZIER, R. H.; CROZIER, Y. C. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. **Genetics**, v. 133, n. 1, p. 97–117, jan. 1993.

D'SOUZA, A. R.; MINCZUK, M. Mitochondrial transcription and translation: overview. **Essays in biochemistry**, v. 62, n. 3, p. 309–320, 2018.

DIAS, J. C. P. *et al.* II Consenso Brasileiro em Doença de Chagas, 2015. **Epidemiologia e Serviços de Saúde**, v. 25, n. 21, p. 1–10, jun. 2016.

DIAS JÚNIOR, J. O. **Avaliação das artérias carótidas de pacientes chagásicos acometidos por eventos isquêmicos cerebrais**. Belo Horizonte: [s.n.].

DIAZ-ALBITER, H. *et al.* The Journal of biological chemistry. **The Journal of Biological Chemistry**, v. 287, n. 28, p. 23995–24003, 2012.

DONG, L. *et al.* Complete Mitochondrial Genome of the Chagas Disease Vector, *Triatoma rubrofasciata*. **The Korean Journal of Parasitology**, v. 56, n. 5, p. 515–519, 31 out. 2018.

DORN, P. L. *et al.* Genetics of Major Insect Vectors. *In: Genetics and Evolution of Infectious Disease*. [s.l.] Elsevier, 2011. p. 411–472.

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, 1 jun. 2001.

ECHTAY, K. S. *et al.* Superoxide activates mitochondrial uncoupling protein 2 from the matrix side. Studies using targeted antioxidants. **The Journal of biological chemistry**, v. 277, n. 49, p. 47129–35, 6 dez. 2002.

FALKENBERG, M. Mitochondrial DNA replication in mammalian cells: overview of the pathway. **Essays in biochemistry**, v. 62, n. 3, p. 287–296, 2018.

FEASEY, N. *et al.* Neglected tropical diseases. **British Medical Bulletin**, v. 93, n. 1, p. 179–200, 1 mar. 2010.

FITZPATRICK, S. *et al.* Molecular Genetics Reveal That Silvatic *Rhodnius prolixus* Do Colonise Rural Houses. **PLoS Neglected Tropical Diseases**, v. 2, n. 4, p. e210, 2 abr. 2008.

GARCÍA, B. A. *et al.* **SHORT REPORT: VARIATION IN MITOCHONDRIAL 12S AND 16S RIBOSOMAL DNA SEQUENCES IN NATURAL POPULATIONS OF TRIATOMA INFESTANS (HEMIPTERA: REDUVIIDAE)**. [s.l.: s.n.]. Disponível em: <<http://www.ajtmh.org/docserver/fulltext/14761645/68/6/0680692.pdf?expires=1566167545&id=id&accname=guest&checksum=9EF5245740C7C4119E06CBA7EA7B287F>>. Acesso em: 18 ago. 2019.

GARCÍA, B. A.; POWELL, J. R. Phylogeny of Species of *Triatoma* (Hemiptera: Reduviidae) Based on Mitochondrial DNA Sequences. **Journal of Medical Entomology**, v. 35, n. 3, p. 232–238, 1 maio 1998.

GARDIM, S. *et al.* 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**, v. 23, p. 74–79, 1 abr. 2014.

GIAMPAZOLIAS, E.; TAIT, S. W. G. Mitochondria and the hallmarks of cancer. **FEBS Journal**, v. 283, n. 5, p. 803–814, mar. 2016.

GODDARD, J. M.; WOLSTENHOLME, D. R. Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. **Nucleic acids research**, v. 8, n. 4, p. 741–57, 25 fev. 1978.

GÓMEZ-PALACIO, A. *et al.* Morphometric and molecular evidence of intraspecific biogeographical differentiation of *Rhodnius pallescens* (HEMIPTERA: REDUVIIDAE: RHODNIINI) from Colombia and Panama. **Infection, Genetics and Evolution**, v. 12, n. 8, p. 1975–1983, 1 dez. 2012.

GORLA, D.; NOIREAU, F. Geographic distribution of Triatominae vectors in America. *In: American Trypanosomiasis Chagas Disease*. Second Edition ed. [s.l.] Elsevier, 2017. p. 197–221.

GOURBIÈRE, S. *et al.* Genetics and evolution of triatomines: from phylogeny to vector control. **Heredity**, v. 108, n. 3, p. 190–202, mar. 2012.

GRABHERR, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. **Nature biotechnology**, v. 29, n. 7, p. 644–52, 15 maio 2011.

GRAÇA-SOUZA, A. V. *et al.* Adaptations against heme toxicity in blood-feeding arthropods. **Insect Biochemistry and Molecular Biology**, v. 36, n. 4, p. 322–335, 1 abr. 2006.

GRUBER, D. R. *et al.* Oxidative damage to epigenetically methylated sites affects DNA stability, dynamics and enzymatic demethylation. **Nucleic acids research**, v. 46, n. 20, p. 10827–10839, 2018.

GUSTAFSSON, C. M.; FALKENBERG, M.; LARSSON, N.-G. Maintenance and Expression of Mammalian Mitochondrial DNA. **Annual Review of Biochemistry**, v. 85, n. 1, p. 133–160, 2 jun. 2016.

HOTEZ, P. J. *et al.* Control of Neglected Tropical Diseases. **New England Journal of Medicine**, v. 357, n. 10, p. 1018–1027, 6 set. 2007.

JUNQUEIRA, A.; GONÇALVES, T.; MOREIRA, C. Manual de capacitação na detecção de *Trypanosoma cruzi* para microscopistas de Malária e Laboratoristas da rede pública. p. 300, 2011.

JUSTI, S. A.; GALVÃO, C. The Evolutionary Origin of Diversity in Chagas Disease Vectors. **Trends in parasitology**, v. 33, n. 1, p. 42–52, 2017.

KAUPPILA, T. E. S. **Mitochondrial DNA Mutagenesis in Metazoa: From Phenotypes to Rescue**. [s.l.] Faculty of Mathematics and Natural Sciences, 2017.

KIM, J.-S. Precision genome engineering through adenine and cytosine base editing. **Nature Plants**, v. 4, n. 3, p. 148–151, 26 mar. 2018.

KOCHER, A. *et al.* Shotgun assembly of the assassin bug *Brontostoma colossus* mitochondrial genome (Heteroptera, Reduviidae). **Gene**, v. 552, n. 1, p. 184–194, 15 nov. 2014.

KWONG, W. K.; MANCENIDO, A. L.; MORAN, N. A. Immune system stimulation by the native gut microbiota of honey bees. **Royal Society Open Science**, v. 4, n. 2, p. 170003, fev. 2017.

LANGMEAD, B.; SALZBERG, S. L. Fast gapped-read alignment with Bowtie 2. **Nature Methods**, v. 9, n. 4, p. 357–359, 4 abr. 2012.

LAVROV, D. V.; PETT, W. Animal Mitochondrial DNA as We Do Not Know It: mt-Genome Organization and Evolution in Nonbilaterian Lineages. **Genome biology and evolution**, v. 8, n. 9, p. 2896–2913, 26 set. 2016.

LENT, H.; WYGODZINSKY, P. W. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. **Bulletin of the American Museum of Natural History**, p. 520, 1979.

LI, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, v. 27, n. 21, p. 2987–2993, 1 nov. 2011.

LI, K.; LIANG, A.-P. Hemiptera Mitochondrial Control Region: New Sights into the Structural Organization, Phylogenetic Utility, and Roles of Tandem Repetitions of the Noncoding Segment. **International journal of molecular sciences**, v. 19, n. 5, 26 abr. 2018.

MARTIN, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. **EMBnet.journal**, v. 17, n. 1, p. 10, 2 maio 2011.

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. 14936–41, 1 dez. 2015.

MINISTÉRIO DA SAÚDE. Doença de Chagas Aguda e distribuição espacial dos triatomíneos de importância epidemiológica, Brasil 2012 a 2016. **Boletim Epidemiológico**, v. 50, p. 10, 2019.

MOLINA-CRUZ, A. *et al.* Reactive Oxygen Species Modulate *Anopheles*

gambiae Immunity against Bacteria and *Plasmodium*. **Journal of Biological Chemistry**, v. 283, n. 6, p. 3217–3223, 8 fev. 2008.

MOLYNEUX, D. H.; SAVIOLI, L.; ENGELS, D. Neglected tropical diseases: progress towards addressing the chronic pandemic. **Lancet (London, England)**, v. 389, n. 10066, p. 312–325, 21 jan. 2017.

MONTEIRO, F. A. *et al.* Phylogeny and molecular taxonomy of the Rhodniini derived from mitochondrial and nuclear DNA sequences. **The American Journal of Tropical Medicine and Hygiene**, v. 62, n. 4, p. 460–465, 1 abr. 2000.

MONTEIRO, F. A. *et al.* Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. **Molecular ecology**, v. 12, n. 4, p. 997–1006, abr. 2003.

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.

MOREIRA, D. A. *et al.* The complete mitochondrial genome of *Corydoras nattereri* (Callichthyidae: Corydoradinae). **Neotropical Ichthyology**, v. 14, n. 1, 2016.

MOREIRA, D. A.; FURTADO, C.; PARENTE, T. E. The use of transcriptomic next-generation sequencing data to assemble mitochondrial genomes of *Ancistrus* spp. (Loricariidae). **Gene**, v. 573, n. 1, p. 171–175, 15 nov. 2015.

MOREIRA, V. R. **ANTIMONIATO DE MEGLUMINA (GLUCANTIME®) CAUSA DANOS AO DNA POR ESTRESSE OXIDATIVO E INDUZ SUPEREXPRESSÃO DE GENES ENVOLVIDOS NA DEFESA ANTIOXIDANTE E REPARO DO DNA**. [s.l.] Universidade Federal do maranhão, 2017.

MURRAY, K. A. *et al.* Pathogeography: leveraging the biogeography of human infectious diseases for global health management. **Ecography. A journal of space and time in ecology**, v. 41, n. 9, p. 1411–1427, 1 set. 2018.

NASCIMENTO, J. D. *et al.* Taxonomical over splitting in the *Rhodnius prolixus* (Insecta: Hemiptera: Reduviidae) clade: Are *R. taquarussuensis* (da Rosa *et al.*, 2017) and *R. neglectus* (Lent, 1954) the same species? **PLOS ONE**, v. 14, n. 2, p. e0211285, 7 fev. 2019.

NEVES, S. C. T. **Análises morfológicas, morfométricas e moleculares revelam uma nova espécie do gênero *Triatoma* do Estado do Tocantins, Brasil**. [s.l.] Centro de Ciências Agrárias, Programa de Pós-Graduação em Proteção de Plantas. Instituto Oswaldo Cruz, 2012.

NIKOLAOU, C.; ALMIRANTIS, Y. A study on the correlation of nucleotide skews and the positioning of the origin of replication: different modes of replication in bacterial species. **Nucleic acids research**, v. 33, n. 21, p. 6816–22, 2005.

NOOR, M. A. F.; LARKIN, J. C. A Re-evaluation of 12S Ribosomal RNA

Variability in *Drosophila pseudoobscura*. **Molecular Biology and Evolution**, v. 17, n. 6, p. 938–941, 1 jun. 2000.

NORMAN, F. F. *et al.* Neglected Tropical Diseases outside the Tropics. **PLoS Neglected Tropical Diseases**, v. 4, n. 7, p. e762, 27 jul. 2010.

OJALA, D.; MONTOYA, J.; ATTARDI, G. tRNA punctuation model of RNA processing in human mitochondria. **Nature**, v. 290, n. 5806, p. 470–474, abr. 1981.

PAHO. **Chagas disease**. Disponível em: <https://www.paho.org/hq/index.php?option=com_topics&view=article&id=10&Itemid=40743&lang=en>. Acesso em: 16 mar. 2019.

PAVAN, M. G. **Especiação em triatomíneos uma abordagem filogenética, biogeográfica e comportamental dos vetores de Chagas *Rhodnius prolixus* e *R. robustus* s.l. (Hemiptera: Reduviidae)**. [s.l.] Fundação Oswaldo Cruz, Instituto Oswaldo Cruz, 2013.

PAVAN, M. G. *et al.* A nuclear single-nucleotide polymorphism (SNP) potentially useful for the separation of *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). **Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases**, v. 14, p. 426–33, mar. 2013.

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.

REYES, A. *et al.* Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. **Molecular Biology and Evolution**, v. 15, n. 8, p. 957–966, 1 ago. 1998.

RIPOCHE, M. *et al.* Multi-Scale Clustering of Lyme Disease Risk at the Expanding Leading Edge of the Range of *Ixodes scapularis* in Canada. **International Journal of Environmental Research and Public Health**, v. 15, n. 4, p. 603, 27 mar. 2018.

ROUBICEK, D. A.; SOUZA-PINTO, N. C. DE. Mitochondria and mitochondrial DNA as relevant targets for environmental contaminants. **Toxicology**, v. 391, p. 100–108, 1 nov. 2017.

SACCONE, C. *et al.* Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. **Gene**, v. 238, n. 1, p. 195–209, 30 set. 1999.

SAITO, S.; TAMURA, K.; AOTSUKA, T. Replication origin of mitochondrial DNA in insects. **Genetics**, v. 171, n. 4, p. 1695–705, dez. 2005.

SANSOM, C. Neglected tropical diseases: securing sustainability. **The Lancet. Infectious diseases**, v. 18, n. 5, p. 502–503, 1 maio 2018.

SANTOS, V. T. DOS. **Estudo dos mecanismos moleculares do reparo de quebra de duplas fitas no DNA mitocondrial**. São Paulo: Biblioteca Digital de Teses e Dissertações da Universidade de São Paulo, 8 maio 2015.

SCHAEFER, C. W. Triatominae (Hemiptera: Reduviidae): Systematic questions and some others. **Neotropical Entomology**, v. 32, n. 1, p. 01–10, jan. 2003.

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.

SCHOMAKER-BASTOS, A.; PROSDOCIMI, F. mitoMaker: A Pipeline for Automatic Assembly and Annotation of Animal Mitochondria Using Raw NGS Data. 24 ago. 2018.

SHARMA, P.; SAMPATH, H. Mitochondrial DNA Integrity: Role in Health and Disease. **Cells**, v. 8, n. 2, 2019.

SHINOTSUKA, C. R. **Avaliação dos produtos da degradação da hemoglobina na lesão cerebral e mecanismos de proteção encefálica após a hemorragia intracraniana**. [s.l.] Fundação Oswaldo Cruz. Instituto de Pesquisa Clínica Evandro Chagas, 2014.

SINAN. **Doença de Chagas Aguda**. Disponível em: <<http://portalsinan.saude.gov.br/doenca-de-chagas-aguda>>. Acesso em: 16 mar. 2019.

SMITH, D. R. The past, present and future of mitochondrial genomics: have we sequenced enough mtDNAs? **Briefings in functional genomics**, v. 15, n. 1, p. 47–54, jan. 2016.

STEWART, J. B.; BECKENBACH, A. T. Characterization of mature mitochondrial transcripts in *Drosophila*, and the implications for the tRNA punctuation model in arthropods. **Gene**, v. 445, n. 1–2, p. 49–57, 15 set. 2009.

STOTHARD, J. R. *et al.* A preliminary survey of mitochondrial sequence variation in Triatominae (Hemiptera: Reduviidae) using polymerase chain reaction-based single strand conformational polymorphism (SSCP) analysis and direct sequencing. **Bulletin of Entomological Research**, v. 88, n. 5, p. 553–560, 10 out. 1998.

SUN, Z. *et al.* Comparison of base composition and codon usage in insect mitochondrial genomes. **Genes & Genomics**, v. 31, n. 1, p. 65–71, fev. 2009.

TAKAHASHI, M. M. **Sequenciamento e análise do genoma mitocondrial de *Gracilaria tenuistipitata* (Gracilariales, Rhodophyta)**. São Paulo: Instituto de Biociências, 21 maio 2010.

TERZIOGLU, M. *et al.* MTERF1 binds mtDNA to prevent transcriptional interference at the light-strand promoter but is dispensable for rRNA gene transcription regulation. **Cell metabolism**, v. 17, n. 4, p. 618–26, 2 abr. 2013.

TRUONG, X. L. *et al.* Notes on the subfamily Salyavatinae (Hemiptera:

Reduviidae) from Vietnam, with the description of a new genus. **Zootaxa**, v. 1615, n. 1, p. 1–20, 15 out. 2007.

UDDIN, A.; CHAKRABORTY, S. Codon usage trend in mitochondrial CYB gene. **Gene**, v. 586, n. 1, p. 105–114, 15 jul. 2016.

VASCONCELOS, R. S.; KOVALESKI, D. F.; JUNIOR, Z. C. T. **Saúde & Transformação Social Health & Social Change**. [s.l.] Univ. Federal de Santa Catarina, 2015. v. 6

VENDRAMI, D. P. *et al.* Análise das relações taxonômicas e sistemáticas entre espécies de triatomíneos (Hemiptera, Reduviidae) de colônias mantidas pelo Serviço Especial de Saúde de Araraquara, inferida de seqüências do 16S rDNA mitocondrial. **Revista Brasileira de Entomologia**, v. 52, n. 3, p. 455–462, set. 2008.

_____. Catalog of the entomologic collections of the Faculty of Public Health of the University of Sao Paulo - (2nd series ii): Triatominae (Hemiptera, Reduviidae). **Revista do Instituto de Medicina Tropical de São Paulo**, v. 60, n. 0, 28 maio 2018.

VIJA-SUAREZ, C. A.; ORTIZ, M. I.; MOLINA, J. Behavioral evidence of oxidative stress by hematophagy in the ocellar system of *Rhodnius prolixus* Stål, 1859 red-eyed mutants. **Journal of Insect Physiology**, v. 101, p. 31–38, 1 ago. 2017.

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.

WESTPHALEN, E. V. N.; BISUGO, M. DA C.; ARAÚJO, M. DE F. L. Aspectos epidemiológicos e históricos do controle da doença de Chagas no Continente Americano. **BEPA. Boletim Epidemiológico Paulista**, v. 9, n. 105, p. 18–35, 2012.

WHO. Chagas disease (American trypanosomiasis). 2019a.

_____. **WHO**. Disponível em: <https://www.who.int/neglected_diseases/mediacentre/WHA_63.20_Eng.pdf?ua=1>. Acesso em: 16 ago. 2019b.

WILKOWSKY, S. E. **Trypanosoma**. In: **Parasitic Protozoa of Farm Animals and Pets**. Switzerland: Springer International Publishing, 2018.

YASUKAWA, T.; KANG, D. An overview of mammalian mitochondrial DNA replication mechanisms. **Journal of biochemistry**, v. 164, n. 3, p. 183–193, 1 set. 2018.

YUAN, S. *et al.* Next-generation sequencing of mixed genomic DNA allows efficient assembly of rearranged mitochondrial genomes in *Amolops chunganensis* and

Quasipaa boulengeri. **PeerJ**, v. 4, p. e2786, 15 dez. 2016.

ZHANG, D.-X.; HEWITT, G. M. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. **Biochemical Systematics and Ecology**, v. 25, n. 2, p. 99–120, 1 mar. 1997.

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. 22177, 26 fev. 2016.

ZHANG, N.-X. *et al.* The Complete Mitochondrial Genome of *Delia antiqua* and Its Implications in Dipteran Phylogenetics. **PLOS ONE**, v. 10, n. 10, 1 out. 2015.

ZOLLO, O.; SONDHEIMER, N. Topological requirements of the mitochondrial heavy-strand promoters. **Transcription**, v. 8, n. 5, p. 307–312, 2017.

HALL, TA., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, vol. 41, p. 95-98.

8. ANEXOS

ANEXO 1 - RNAs não codificantes.

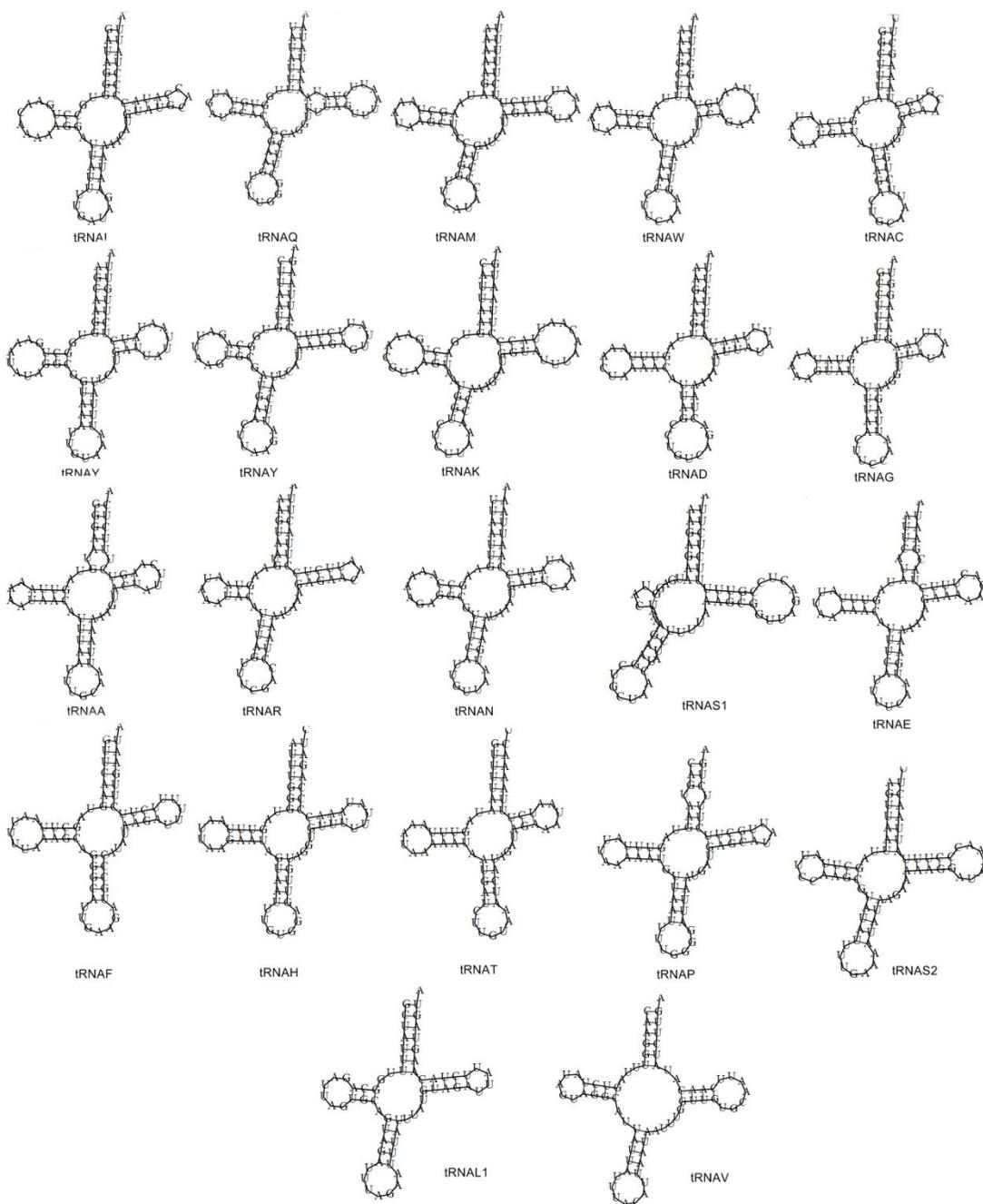


Figura 123 - Estrutura secundária predita dos RNAs transportadores do mitogenoma de *Rhodnius prolixus*, colônia NIH. Os nomes dos tRNAs segue nomenclatura do código de uma letra IUPAC-IUB.

ANEXO 2- Árvore filogenética do gene mitocondrial citocromo b (CYTB)

