

Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

Instituto Oswaldo Cruz

Pós-Graduação em Biologia Parasitária

Márcio Galvão Pavan

Especiação em triatomíneos: uma abordagem filogenética, biogeográfica

e comportamental dos vetores de Chagas *Rhodnius prolixus*

e *R. robustus* s.l. (Hemiptera: Reduviidae)

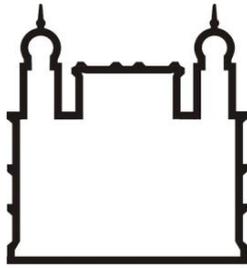
Tese apresentada ao Instituto Oswaldo Cruz  
como parte dos requisitos para obtenção do título  
de Doutor em Ciências.

Orientadores: Dr. Fernando Araújo Monteiro

Dr. Alexandre Afrânio Peixoto (*in memoriam*)

RIO DE JANEIRO

Maio de 2013



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **Instituto Oswaldo Cruz**

### **Pós-Graduação em Biologia Parasitária**

Márcio Galvão Pavan

Especiação em triatomíneos: uma abordagem filogenética, biogeográfica e comportamental dos vetores de Chagas *Rhodnius prolixus* e *R. robustus s.l.* (Hemiptera: Reduviidae).

Orientadores: Dr. Fernando Araújo Monteiro

Dr. Alexandre Afrânio Peixoto (*in memoriam*)

**Aprovado em: 03/05/2013.**

#### BANCA EXAMINADORA:

**Dra. Ana Carolina Paulo Vicente – Presidente da banca examinadora**

FIOCRUZ / Instituto Oswaldo Cruz / Laboratório de Biologia Molecular de Microorganismos

**Dr. Fernando Abad-Franch – Membro da banca examinadora e Revisor**

FIOCRUZ / Centro de Pesquisas Leônidas e Maria Deane / Laboratório de Biodiversidade em Saúde

**Dr. Pedro Cordeiro Estrela de Andrade Pinto – Membro da banca examinadora**

Universidade Federal da Paraíba / Centro de Ciências Exatas e da Natureza / Departamento de Sistemática e Ecologia

**Dr. Rafael Maciel de Freitas – 1º Suplente da banca examinadora**

FIOCRUZ / Instituto Oswaldo Cruz / Laboratório de Transmissores de Hematozoários

**Dr. Gonzalo José Bello Bentancor – 2º Suplente da banca examinadora**

FIOCRUZ / Instituto Oswaldo Cruz / Laboratório de AIDS e Imunologia Molecular

Rio de Janeiro, maio de 2013.

Ficha catalográfica a ser elaborada pela

Biblioteca Central de Manguinhos

Pavan, Márcio Galvão

Especiação em triatomíneos: uma abordagem filogenética, biogeográfica e comportamental dos vetores de Chagas *Rhodnius prolixus* e *R. robustus* s.l. (Hemiptera: Reduviidae).

Tese de doutorado em Biologia Parasitária, área de concentração em Genética Molecular e de Microorganismos

Instituto Oswaldo Cruz – FIOCRUZ

Rio de Janeiro, 2013.

1. *Rhodnius* 2. *Rhodnius prolixus* 3. *Rhodnius robustus* 4. biogeografia de vicariância 5. região amazônica 6. complexo de espécies 7. relógio circadiano 8. comportamento 9. atividade locomotora

Este trabalho foi realizado no Laboratório de Epidemiologia e Sistemática Molecular e no Laboratório de Biologia Molecular de Insetos, ambos do Instituto Oswaldo Cruz, Fiocruz. Foram utilizados recursos da Fundação Oswaldo Cruz e do CNPq.

“A nossa maior glória não reside no fato de nunca cairmos, mas sim em levantarmo-nos sempre depois de cada queda”.

Confúcio

Dedico esta tese aos meus pais, Evandro e Eliane,  
à minha esposa e à minha filha, Paula  
Andrea e Gabriela, e ao grande mestre  
Alexandre Peixoto (*in memoriam*).

## AGRADECIMENTOS

Agradeço aos meus pais, Evandro e Eliane, por toda dedicação, carinho, apoio e amor incondicional durante esse longo caminho. Sei o quanto foi difícil investir em minha formação e acreditar nos meus sonhos. Espero ter aproveitado todos os ensinamentos que a mim foram dados, para que um dia eu também seja um bom exemplo para minha filha.

Agradeço à minha esposa, Paula Andrea, por todo o carinho, compreensão, cumplicidade, perseverança, amor e pela paciência. Nesses três anos experimentamos os mais variados sentimentos, desde a extrema tristeza, com a perda de pessoas importantes nas nossas vidas, até a maior alegria que a vida pôde nos dar, nossa pequena Gabriela. Este nosso pequeno milagre enche os nossos corações de amor e a nossa casa de alegria. Que toda esta nossa luta seja revertida em força, união e valorização das pequenas conquistas. Esta tese também é sua.

Agradeço ao meu orientador, amigo e “pai científico” Dr. Fernando Monteiro, por estes quatro anos a mais de ensinamentos. Por mais que a gente esteja juntos há 10 anos, sempre aprendo uma nova lição, seja ela científica ou pessoal. Depois do grande susto que a vida lhe deu, fico extremamente feliz de tê-lo aqui para partilharmos nossas conquistas, alegrias e angústias. Sem dúvida você é um grande ser humano e mestre.

Tenho muito a agradecer ao meu outro orientador, Dr. Alexandre Peixoto, um pesquisador incrível, de grandes ensinamentos e de uma genialidade natural, sempre com muita simplicidade. Sem dúvida foi uma perda inestimável, não apenas para mim, mas principalmente para a ciência. Uma pena a vida nos tirar o privilégio de sua presença de uma maneira tão abrupta. Durante estes quatro anos, construí um sonho de trabalharmos juntos, de ter a chance de agregar seus conhecimentos aos problemas evolutivos e filogenéticos dos triatomíneos. Sinto falta de nossas conversas e de sua empolgação ao discutir os mínimos resultados. Mas tenho certeza que este seu legado será levado adiante por todos aqueles que tiveram contato com você. Resta a mim meu singelo muito obrigado.

Agradeço também ao meu “anjo da guarda”, minha aluna Jéssica Correa Antonio, que ao longo destes quatro anos foi incansável e perseverante. À Dra. Tatiana Peretolchina, do Laboratório de Geno-sistemática (Лаборатория геносистематики), Instituto Limnológico do Setor Siberiano da Academia Russa de

Ciência (Лимнологический институт Сибирского отделения Российской Академии наук). Mesmo tendo voltado para Sibéria, não deixou de lembrar do colega brasileiro e simplesmente disponibilizou o computador de seu laboratório na Rússia e seu computador pessoal para rodar mais de 90 gigabytes de informação desta tese.

À Dra. Karina Alessandra Morelli, responsável técnica da Plataforma de *DNA Barcoding* do Instituto Oswaldo Cruz, FIOCRUZ, pela ajuda inicial com os microssatélites e pelo apoio incondicional como profissional e amiga, seja dentro ou fora do ambiente de trabalho.

À Dra. Ana Cristina Bahia Nascimento, do Instituto de Bioquímica Médica da Universidade Federal do Rio de Janeiro, pela grande ajuda com a clonagem do ITS-2. Você foi incansável e muito amiga!

Ao Dr. Fernando Abad-Franch pela colaboração na coleta de Manaus (Amazonas) até Porto Velho (Rondônia), pela revisão criteriosa e minuciosa desta tese em “tempo recorde” e às sugestões relevantes que com certeza fará para a melhoria deste trabalho.

Aos demais membros da banca examinadora, Dra. Ana Carolina Paulo Vicente, Dr. Pedro Estrela, Dr. Rafael Maciel e ao Dr. Gonzalo Bello, por terem aceitado este grande desafio de receber a tese faltando pouco tempo para a defesa. Tenho certeza de que farão ótimas sugestões para o enriquecimento do trabalho.

À Carolina Dale e Dr. Cléber Galvão, do Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos, Instituto Oswaldo Cruz, FIOCRUZ, por terem gentilmente cedido as amostras de *R. robustus s.l.* de São Roberto (Maranhão) e de Abaetetuba (Pará); e ao Dr. Rafael Mesquita, do Laboratório de Bioinformática, Departamento de Bioquímica, da Universidade Federal do Rio de Janeiro, pela ajuda no isolamento dos genes de relógio no genoma de *R. prolixus*.

Aos técnicos Ricardo Mota, Diego Leite e ao motorista Júnior, do Centro de Pesquisas Leônidas e Maria Deane, FIOCRUZ, e ao mestre Walter Souza Santos, do Laboratório de doença de Chagas, do Instituto Evandro Chagas, pela colaboração e auxílio incansável na coleta de Manaus (Amazonas) até Porto Velho (Rondônia). Apesar de até termos corrido risco de vida, vocês souberam manter o alto astral e trabalhar com muito profissionalismo; ao Dr. Luiz Herman Soares Gil, do Instituto de Pesquisas em Patologias Tropicais de Rondônia, pelo auxílio nas coletas em Porto Velho (Rondônia); ao Dr. Aldo Valente e aos técnicos Nivaldo e Gilberto, do

Laboratório de doença de Chagas, do Instituto Evandro Chagas, pela colaboração e auxílio na coleta de Belém (Pará) até Palmas (Tocantins).

Ao Dr. Luiz Guilherme Bauzer, do Laboratório de Biologia Molecular de Insetos, do Instituto Oswaldo Cruz, FIOCRUZ, por ter se disponibilizado a ajudar na revisão da tese (mesmo que informalmente) e colaborar com comentários relevantes, mesmo estando na Alemanha fazendo seu pós-doc. Além disso, você e Luciana foram grandes amigos nos momentos mais difíceis.

Ao Dr. Gustavo Rivas, do Laboratório de Biologia Molecular de Insetos, do Instituto Oswaldo Cruz, FIOCRUZ, pelo auxílio nos experimentos de comportamento.

À todos os membros do Laboratório de Sistemática Bioquímica e “agregados”, usuários da Plataforma de *DNA Barcoding*, pela paciência e disponibilidade, muitas vezes deixando de fazer seus experimentos para que eu pudesse terminar esta tese; à Zezé do Laboratório de Doenças Parasitárias, por cuidar dos meus “filhos” triatomíneos na colônia.

À todos os amigos que me ajudaram nos momentos difíceis, sejam eles profissionais ou mesmo pessoais.

## INDICE

|   |    |
|---|----|
| RESUMO .....  | 1  |
| ABSTRACT .....  | 2  |
| 1. INTRODUÇÃO .....   | 3  |
| 1.1. Doenças tropicais negligenciadas - a doença de Chagas.....   | 3  |
| 1.2. A doença de Chagas na região amazônica .....   | 4  |
| 1.3. Os vetores da doença de Chagas .....   | 5  |
| 1.3.1. O gênero <i>Rhodnius</i> na região pan-amazônica.....  | 7  |
| 1.4. Técnicas moleculares para identificação de espécies crípticas de triatomíneos.....   | 8  |
| 1.4.1. Análise de microssatélites em estudos populacionais.....   | 9  |
| 1.4.2 Sequenciamento de DNA.....  | 10 |
| 1.4.2.1. Análise de sequências nucleares.....   | 11 |
| 1.4.2.2 Análise de sequências mitocondriais.....  | 13 |
| 1.4.3. Complexos de espécies crípticas de <i>Rhodnius</i> na região pan-amazônica .....   | 14 |
| 1.5. A biogeografia e a região pan-amazônica .....  | 16 |
| 1.5.1. Conceitos biogeográficos segundo os modelos dispersionista e vicariante .....  | 18 |
| 1.5.2. Hipóteses biogeográficas aplicadas à região amazônica .....  | 20 |
| 1.5.2.1. Hipótese de incursões marinhas .....   | 21 |
| 1.5.2.2. Hipótese de rios atuando como barreiras.....   | 22 |
| 1.5.2.3. Hipótese de refúgios.....  | 23 |
| 1.5.2.4 Outras hipóteses biogeográficas .....   | 24 |
| 1.5.3. Análises biogeográficas atuais.....  | 26 |
| 1.5.3.1. Biogeografia cladística .....  | 26 |
| 1.5.3.2. Relógio molecular e a filogeografia bayesiana.....   | 28 |
| 1.5.3.3. O método bayesiano aplicado à teoria da coalescência .....   | 31 |
| 1.5.4. Filogeografia de <i>R. prolixus</i> e <i>R. robustus s.l.</i> na região pan-amazônica .....  | 33 |
| 1.6. Ritmos circadianos .....   | 36 |
| 1.6.1. A genética dos ritmos circadianos .....  | 38 |
| 1.6.2. Ritmos circadianos em triatomíneos.....  | 44 |
| 2. JUSTIFICATIVA .....  | 47 |
| 3. OBJETIVOS .....  | 50 |
| 3.1. Objetivos gerais.....  | 50 |
| 3.2. Objetivos específicos .....  | 50 |
| 4. METODOLOGIA .....  | 52 |
| 4.1. Estratégia de coleta e obtenção das amostras.....  | 52 |
| 4.2. Extração de DNA e identificação dos espécimes.....   | 61 |
| 4.3. Amplificação pela PCR e sequenciamento do mtCytb .....   | 61 |
| 4.4. Amplificação pela PCR, clonagem e sequenciamento do ITS-2.....   | 62 |
| 4.5. Análise de microssatélites .....   | 63 |
| 4.6. Isolamento de novos marcadores moleculares no genoma de <i>R. prolixus</i> .....   | 66 |
| 4.6.1. Desenvolvimento de um método diagnóstico para <i>R. prolixus</i> e <i>R. robustus s.l.</i> .....   | 66 |
| 4.6.2. Identificação dos genes ortólogos do relógio circadiano de <i>R. prolixus</i> .....  | 66 |
| 4.6.3. Análise filogenética dos genes homólogos de <i>per</i> , <i>tim</i> e <i>cry</i> .....   | 68 |
| 4.6.4. Escolha das regiões gênicas dentro de cada gene do relógio circadiano a serem utilizadas como novos marcadores moleculares aplicados à sistemática de triatomíneos ..... | 68 |
| 4.6.5. Amplificação dos marcadores pela PCR e sequenciamento .....  | 69 |
| 4.7. Edição e alinhamento das sequências.....   | 70 |
| 4.8. Análises filogenéticas e de distância genética intra e interespecíficas.....   | 71 |
| 4.8.1. mtCytb .....   | 71 |

|  |            |
|--|------------|
| 4.8.2. ITS-2 .....   | 72         |
| 4.8.3. Árvore filogenética de espécies com base nas sequências de mtCytb e ITS-2.....  | 76         |
| 4.8.4. Reconstrução filogenética das sequências dos marcadores relacionados ao ritmo circadiano.....   | 77         |
| 4.8.5. Árvore filogenética de espécies com base nas sequências de mtCytb, ITS-2 e dos marcadores relacionados ao ritmo circadiano.....       | 78         |
| 4.9. Análises filogeográficas e biogeográficas.....  | 79         |
| 4.9.1. Calibrações das datações dos eventos cladogenéticos .....   | 79         |
| 4.9.2. Testes das hipóteses biogeográficas de “incursões marinhas” e “refúgios pleistocênicos” .....   | 81         |
| 4.9.3. Teste de hipóteses quanto à origem e dispersão da linhagem ancestral de <i>R. prolixus</i> e <i>R. robustus s.l.</i> .....            | 84         |
| 4.10. Análise da atividade locomotora de <i>R. prolixus</i> e <i>R. robustus s.l.</i> .....  | 91         |
| 4.10.1. Escolha e preparo das amostras e delineamento experimental .....   | 91         |
| 4.10.2. Análise dos dados de atividade locomotora.....   | 94         |
| <b>5. RESULTADOS.....</b>  | <b>95</b>  |
| 5.1. Coleta e identificação morfológica dos triatomíneos .....   | 95         |
| 5.2. Filogenia de <i>R. prolixus</i> e <i>R. robustus s.l.</i> e divergência molecular entre as sequências .....                             | 99         |
| 5.2.1. MtCytb.....   | 99         |
| 5.2.2. ITS-2 .....   | 105        |
| 5.2.3. Árvore filogenética de espécies com base nas sequências de mtCytb e ITS-2... ..   | 114        |
| 5.3. Análise de microsatélites .....   | 115        |
| 5.4. Isolamento de novos marcadores moleculares no genoma de <i>R. prolixus</i> e análises filogenéticas .....                               | 124        |
| 5.4.1. Marcador molecular diagnóstico para <i>R. prolixus</i> - <i>AmpG</i> .....  | 124        |
| 5.4.1. Genes do relógio circadiano .....   | 125        |
| 5.4.2. period .....  | 125        |
| 5.4.3. timeless .....  | 131        |
| 5.4.4. <i>cryptochrome1</i> e <i>cryptochrome2</i> .....   | 136        |
| 5.5. Filogenia de <i>R. prolixus</i> e <i>R. robustus s.l.</i> com base na análise concatenada dos marcadores nucleares e mitocondrial ..... | 144        |
| 5.6. Análises filogeográficas de <i>R. prolixus</i> e <i>R. robustus s.l.</i> .....  | 145        |
| 5.6.1. Hipótese do relógio molecular e datação das cladogêneses.....   | 145        |
| 5.6.2. Genealogia molecular e testes de expansão populacional súbita.....  | 149        |
| 5.6.3. Hipóteses biogeográficas de dispersão de <i>R. prolixus</i> e <i>R. robustus s.l.</i> nas regiões Amazônica e do Orinoco .....        | 156        |
| 5.7. Atividade locomotora de <i>R. prolixus</i> e <i>R. robustus</i> II.....   | 158        |
| <b>6. DISCUSSÃO .....</b>  | <b>164</b> |
| 6.1. Análises filogenéticas e identidades taxonômicas de <i>R. prolixus</i> e <i>R. robustus s.l.</i> ... ..                                 | 164        |
| 6.1.1. Análise filogenética do mtCytb .....  | 166        |
| 6.1.2. Identificação taxonômica utilizando o mtCytb .....  | 169        |
| 6.1.3. Análises filogenéticas segundo o modelo coalescente de especiação.....  | 170        |
| 6.1.4. Análise filogenética do ITS-2 .....   | 171        |
| 6.1.5. Análise filogenética dos genes relacionados ao relógio circadiano .....   | 175        |
| 6.2. Análise de microsatélites .....   | 180        |
| 6.3. Filogeografia de <i>R. prolixus</i> e <i>R. robustus s.l.</i> na região amazônica.....  | 185        |
| 6.3.1. Relógio molecular relaxado, calibração dos nós da árvore filogenética e datação dos eventos cladogenéticos.....                       | 186        |
| 6.3.2. Padrões populacionais e genealogia molecular.....   | 189        |
| 6.3.3. Hipóteses de origem e dispersão das linhagens .....   | 192        |
| 6.4. As cladogêneses da linhagem <i>R. robustus s.l.</i> no contexto biogeográfico da região pan-amazônica.....                              | 194        |
| 6.5. Análise locomotora de <i>R. prolixus</i> e <i>R. robustus</i> II .....  | 200        |

|   |     |
|---|-----|
| 6.6. Taxonomia molecular de <i>R. prolixus</i> e <i>R. robustus s.l.</i> .....                  | 203 |
| 6.6.1. AmpG .....   | 204 |
| 6.6.2. Utilização dos indels de timeless como marcador diagnóstico.....                         | 205 |
| 6.7. Utilidade dos novos marcadores nucleares na sistemática molecular de triatomíneos<br>..... | 205 |
| 7. CONCLUSÕES .....   | 207 |
| 8. ANEXO 1.....   | 209 |
| 9. ANEXO 2.....   | 224 |
| 10. REFERÊNCIAS BIBLIOGRÁFICAS.....   | 249 |

## INDICE DE FIGURAS

|   |     |
|---|-----|
| <b>Figura 1.1.</b> Distribuição estimada de <i>R. prolixus</i> e das linhagens crípticas de <i>R. robustus s.l.</i> .....   | 16  |
| <b>Figura 1.2.</b> Idades geológicas da Era Cenozóica (desde a Época do Mioceno). .....   | 21  |
| <b>Figura 1.3.</b> Alça regulatória principal, ou alça regulatória <i>per</i> . .....   | 40  |
| <b>Figura 1.4.</b> Alça regulatória secundária, ou alça regulatória <i>Clk</i> . .....  | 41  |
| <b>Figura 4.1.</b> Transecto percorrido (linha vermelha) por áreas contíguas de distribuição de <i>R. robustus</i> III (amarelo) e IV (roxo), com representação dos 18 pontos visitados. ....   | 54  |
| <b>Figura 4.2.</b> Transecto percorrido (linha vermelha) por áreas contíguas de distribuição de <i>R. robustus</i> II, III e IV, com representação dos 10 pontos visitados. ....  | 69  |
| <b>Figura 4.3.</b> Distribuição geográfica dos espécimes amostrados e sequenciados de <i>R. prolixus</i> , <i>R. robustus s.l.</i> , <i>R. barretti</i> , <i>R. pictipes</i> , <i>R. pallescens</i> I e II e <i>R. nasutus</i> . .... | 60  |
| <b>Figura 4.4.</b> Ganho e perda de caracteres desde o nascimento. ....   | 75  |
| <b>Figura 4.5.</b> Áreas biogeográficas da América do Sul e América Central. ....   | 87  |
| <b>Figura 4.6.</b> Hipóteses biogeográficas de dispersão testadas em uma macroescala. ....  | 89  |
| <b>Figura 4.7.</b> Hipóteses biogeográficas de dispersão testadas em uma mesoescala. ....   | 90  |
| <b>Figura 4.8.</b> Experimento de atividade locomotora. ....  | 93  |
| <b>Figura 4.9.</b> Monitores para os experimentos de atividade locomotora. ....   | 93  |
| <b>Figura 5.1.</b> Árvore filogenética do mtCytb. ....  | 101 |
| <b>Figura 5.2.</b> Gráfico das três coordenadas utilizadas na análise de PCA. ....  | 104 |
| <b>Figura 5.3.</b> Árvore filogenética consenso de ITS-2. ....  | 109 |
| <b>Figura 5.4.</b> Árvore filogenética de ITS-2 após análise de homologia das sequências. ....  | 113 |
| <b>Figura 5.5.</b> Árvore filogenética consenso de espécies obtida a partir da concatenação das sequências de mtCytb e ITS-2. ....  | 115 |
| <b>Figura 5.6.</b> Gráfico das três coordenadas utilizadas na análise de PCA dos <i>loci</i> de microssatélites analisados. ....  | 121 |
| <b>Figura 5.7.</b> Gráfico dos valores de verossimilhança para definição do número de agrupamentos (K). ....  | 123 |
| <b>Figura 5.8.</b> Gráfico da estruturação das quatro populações analisadas. ....   | 124 |
| <b>Figura 5.9.</b> Éxons (quadrados em ciano) e íntrons (linhas em cinza e ciano) do gene <i>per</i> predito. ....  | 127 |
| <b>Figura 5.10.</b> Árvore filogenética de NJ construída com base em sequências de aminoácidos dos ortólogos de PER em Insecta. ....  | 128 |
| <b>Figura 5.11.</b> Árvore filogenética de <i>per</i> . ....  | 130 |
| <b>Figura 5.12.</b> Árvore filogenética dos <i>indels</i> de <i>per</i> . ....  | 131 |
| <b>Figura 5.13.</b> Éxons (quadrados em ciano e amarelo) e íntrons (linhas em cinza e ciano) do gene ortólogo <i>tim</i> predito. ....  | 133 |
| <b>Figura 5.14.</b> Árvore filogenética de NJ dos genes ortólogos de TIM em Insecta. ...  | 134 |
| <b>Figura 5.15.</b> Árvore filogenética consenso de <i>tim</i> . ....   | 136 |
| <b>Figura 5.16.</b> Árvore filogenética dos <i>indels</i> de <i>tim</i> . ....  | 136 |
| <b>Figura 5.17.</b> Árvore filogenética (NJ; Dayhoff; 1000 replicações <i>bootstrap</i> ) construída com base em sequências de aminoácidos dos homólogos de CRY em Insecta. ....  | 138 |
| <b>Figura 5.18.</b> Éxons (quadrados em ciano e amarelo) e íntrons (linhas em cinza) do gene ortólogo <i>cry2</i> predito no genoma de <i>R. prolixus</i> . ....  | 139 |

|   |     |
|---|-----|
| <b>Figura 5.19.</b> Árvore filogenética do íntron 6 de <i>cry2</i> . .....  | 142 |
| <b>Figura 5.20.</b> Árvore filogenética do íntron 8 de <i>cry2</i> . .....  | 143 |
| <b>Figura 5.21.</b> Árvore filogenética dos <i>indels</i> do íntron 6 de <i>cry2</i> . .....  | 143 |
| <b>Figura 5.22.</b> Árvore filogenética dos <i>indels</i> do íntron 8 de <i>cry2</i> . .....  | 144 |
| <b>Figura 5.23.</b> Árvore filogenética consenso de espécies obtida a partir da concatenação das sequências de mtCytb, ITS-2, <i>per</i> , <i>tim</i> e <i>cry2</i> . .....   | 145 |
| <b>Figura 5.24.</b> Reconstrução filogenética bayesiana contendo as datações dos eventos vicariantes do complexo <i>R. robustus s.l.</i> e <i>R. prolixus</i> . .....   | 147 |
| <b>Figura 5.25.</b> Rede de haplótipos construída pelo algoritmo de <i>Median-joining</i> com base nas sequências de mtCytb de <i>R. prolixus</i> (azul) e <i>R. robustus</i> I (vermelho). ....                    | 149 |
| <b>Figura 5.26.</b> Rede de haplótipos construída pelo algoritmo de <i>Median-joining</i> com base nas sequências de mtCytb de <i>R. robustus</i> II. ....  | 150 |
| <b>Figura 5.27.</b> Rede de haplótipos construída pelo algoritmo de <i>Median-joining</i> com base nas sequências de mtCytb de <i>R. robustus</i> III e <i>R. robustus</i> n.sp. ....                               | 151 |
| <b>Figura 5.28.</b> Rede de haplótipos construída pelo algoritmo de <i>Median-joining</i> com base nas sequências de mtCytb de <i>R. robustus</i> IV. ....  | 152 |
| <b>Figura 5.29.</b> Mismatch-distribution das amostras de <i>R. prolixus</i> e <i>R. robustus</i> I-IV. ....  | 154 |
| <b>Figura 5.30.</b> Gráfico bayesiano <i>Skyride</i> mostrando que apenas <i>R. robustus</i> II apresenta indícios de expansão demográfica súbita. ....   | 155 |
| <b>Figura 5.31.</b> Atividade locomotora de ninfas de triatomíneos. ....  | 159 |
| <b>Figura 5.32.</b> Médias marginais estimadas do arco seno da raiz quadrada das proporções de atividade diurna nas condições LD 12:12. ....  | 161 |
| <b>Figura 5.33.</b> Médias marginais estimadas do arco seno da raiz quadrada das proporções de atividade diurna nas condições LD 12:12 em comparação com a presença e ausência de abrigo. ....                      | 162 |
| <b>Figura 5.34.</b> Diagrama de caixas dos períodos do relógio circadiano de 23 <i>R. prolixus</i> e 15 <i>R. robustus</i> II. ....   | 163 |
| <b>Figura 6.1.</b> Esquema das topologias obtidas a partir do sequenciamento dos genes mtCytb (A), ITS-2 (B), <i>period</i> (C); <i>timeless</i> (D); <i>cryptochrome2-I6</i> (E), e <i>cryptochrome2-I8</i> . .... | 167 |
| <b>Figura 6.2.</b> (A) Sistema Pebas, que supostamente teria ocorrido entre 23 e 10 Maa. (B) Sistema Acre, que supostamente teria sido formado entre 10 e 7 Maa. ....   | 195 |
| <b>Figura 6.3.</b> Árvore de máxima parcimônia consenso, utilizando como base 414-pb de mtCytb, 383-pb mtlstutRNA e 632-pb D2 28S RNA. Retirado de Monteiro et al. (2000). ....                                     | 198 |

## INDICE DE TABELAS

|  |     |
|--|-----|
| <b>Tabela 4.1.</b> Localidades visitadas durante o percurso pelo transecto representado na Figura 4.1A. ....   | 54  |
| <b>Tabela 4.2.</b> Localidades visitadas durante o percurso pelo transecto representado na Figura 4.2A. ....   | 55  |
| <b>Tabela 4.3.</b> Amostras de <i>Rhodnius</i> utilizadas no sequenciamento de sete marcadores. Os números abaixo de cada marcador utilizado (colunas entre “Área biogeográfica” e “Campo/Colônia”) representam o número de indivíduos provenientes de cada localidade / colônia que tiveram o fragmento em questão sequenciados. .... | 56  |
| <b>Tabela 4.4.</b> Características gerais dos 10 <i>loci</i> analisados. ....  | 64  |
| <b>Tabela 4.5.</b> Tipos de calibração utilizados nos nós das árvores filogenéticas. ....  | 81  |
| <b>Tabela 5.1.</b> Porcentagem de divergências genéticas (K-2p) interespecíficas (abaixo da diagonal fundo branco) e intraespecíficas (na diagonal, fundo cinza) das sequências do mtCyb de <i>R. prolixus</i> , <i>R. robustus s.l.</i> e <i>R. barretti</i> calculadas em comparações par-a-par. ....                                | 102 |
| <b>Tabela 5.2.</b> Índices de variabilidade e testes de neutralidade. ....   | 105 |
| <b>Tabela 5.3.</b> Sítios variáveis encontrados nas comparações feitas entre as sequências de ITS-2. ....  | 106 |
| <b>Tabela 5.4.</b> Porcentagem de divergência K2-p intraindividual das sequências de ITS-2 de <i>R. robustus</i> II, III e IV. ....  | 110 |
| <b>Tabela 5.5.</b> Porcentagem de divergências genéticas (K-2p) interespecíficas (abaixo da diagonal fundo branco) e intraespecíficas (na diagonal, fundo cinza) das sequências de ITS-2. ....   | 111 |
| <b>Tabela 5.6.</b> Porcentagem de divergências genéticas (K-2p) interespecíficas (abaixo da diagonal fundo branco) e intraespecíficas (na diagonal, fundo cinza) das sequências de ITS-2 que apresentaram porcentagens de divergência inferiores a 0,6% em comparações intraespecíficas. ....  | 112 |
| <b>Tabela 5.7.</b> Resumo da variabilidade dos 10 <i>loci</i> de microssatélites encontrada nas quatro populações de <i>R. robustus</i> III e n.sp. analisadas. ....   | 117 |
| <b>Tabela 5.8.</b> Frequência alélica encontrada em cada uma das quatro populações dos 10 <i>loci</i> de microssatélites analisados. ....  | 119 |
| <b>Tabela 5.9.</b> Valores do índice G calculados a partir de comparações par-a-par das frequências alélicas interpopulacionais de cada um dos 8 <i>loci</i> polimórficos analisados. ....   | 120 |
| <b>Tabela 5.10.</b> Teste Chi-Quadrado das frequências alélicas dos marcadores microssatélites das populações de <i>R. robustus</i> III, IV e sp. nov. ....  | 122 |
| <b>Tabela 5.11.</b> Valores dos índices de $R_{ST}$ (abaixo da diagonal) e $F_{ST}$ (acima da diagonal) calculados a partir de comparações par-a-par das frequências alélicas interpopulacionais dos oito <i>loci</i> polimórficos analisados. ....  | 123 |
| <b>Tabela 5.12.</b> Datação dos eventos vicariantes, em milhão de anos atrás (Maa) calculadas com base nas cinco diferentes estratégias de calibração dos nós das árvores. ....  | 148 |
| <b>Tabela 5.13.</b> Valores do Fator de Bayes (BF) obtidos a partir da comparação entre os valores de verossimilhança calculados pelos algoritmos de <i>pathlikelihood</i> e <i>stepping-stone</i> para os modelos de crescimento populacional constante e variável. ....  | 155 |
| <b>Tabela 5.14.</b> Valores marginais dos logaritmos de verossimilhança dos modelos biogeográficos em macroescala, calculados segundo o modelo harmônico e corrigidos pelo método de Bezier. ....  | 157 |

|  |     |
|--|-----|
| <b>Tabela 5.15.</b> Valores marginais dos logaritmos de verossimilhança dos modelos biogeográficos em mesoescala, calculados segundo o modelo harmônico e corrigidos pelo método de Bezier. .... | 158 |
| <b>Tabela 5.16.</b> Análise Four-way ANOVA, comparando a presença/ausência de abrigo, tempo de privação de alimento, espécie e estágio ninfal. ....  | 160 |

## RESUMO

As dificuldades para a interrupção definitiva da transmissão vetorial do *Trypanosoma cruzi* decorrem primariamente de eventos de reinfestação por populações silvestres de triatomíneos das casas previamente tratadas com inseticidas e, de maneira secundária, do contato esporádico do homem com vetores infectados que habitam o peridomicílio e o ambiente silvestre. A escrutinização das regiões potenciais de transmissão endêmicas ou hipoendêmicas depende da correta diagnose das espécies vetoras e dos seus limites de ocorrência geográfica. No caso de espécies crípticas, a identificação taxonômica deve ser feita principalmente por métodos moleculares de baixo custo e eficientes, que a partir dos quais seja possível gerar resultados simples de interpretar. A diagnose correta dos vetores é útil para a determinação regional dos riscos potenciais de transmissão vetorial da doença em seres humanos. Contudo, o desenvolvimento de novas estratégias de controle mais eficazes depende do entendimento acerca dos padrões biológicos, ecológicos e evolutivos que estão envolvidos nos processos de domiciliação e especiação desses vetores. Estes padrões estão inseridos em um contexto espacial e temporal complexo, que só pode ser compreendido quando estudos em diferentes escalas geográficas e históricas são realizados. Neste contexto, o complexo de espécies crípticas *R. prolixus* e *R. robustus s.l.* serve como um importante modelo, por ser um grupo bem estudado, de filogenia conhecida e ocorrência geográfica previamente estimada. Nesta tese foi avaliado o status taxonômico dessas espécies e as possíveis barreiras naturais ao fluxo gênico entre as linhagens determinadas com base em (1) análises filogenéticas, utilizando um marcador mitocondrial (Cytb) e cinco marcadores nucleares (ITS-2, *period*, *timeless*, *cryptochrome2-I6*, *cryptochrome2-I8*); (2) biogeográficas, utilizando duas calibrações geológicas do relógio molecular relaxado; (3) populacionais, analisando 10 *loci* de microssatélites; e (4) de comportamento, medindo as atividades locomotoras e estimando o período endógeno do relógio circadiano. Adicionalmente, foi identificado um marcador em um gene nuclear de cópia única (TPS165) que separa as espécies crípticas e de diferentes importâncias epidemiológicas: *R. prolixus* (importância primária) e *R. robustus s.l.* (importância secundária). Ao todo, foram analisados 254 espécimes de *R. prolixus* e *R. robustus s.l.* As amostras de *R. robustus* foram obtidas após duas coletas ao longo de transectos que representam áreas contíguas de ocorrência das linhagens crípticas *R. robustus* III e IV, e *R. robustus* II, III e IV, de acordo com o que se conhece sobre a distribuição destes táxons. *R. prolixus* é uma espécie distinta de *R. robustus s.l.* e, portanto, válida. *R. robustus s.l.* é um complexo parafilético de, pelo menos, sete linhagens, das quais três parecem representar três espécies distintas – *R. barretti*, *R. robustus* I e *R. robustus* V. A partir do resultado filogenético incongruente com base na topologia das árvores filogenéticas dos genes analisados das quatro linhagens restantes, *R. robustus* II, III, IV e sp. nov. representam um complexo de espécies *in status nascendi*. As populações ancestrais de *R. prolixus* e *R. robustus s.l.* sofreram cladogêneses decorrentes de eventos geológicos amazônicos ocorridos entre o Mioceno Superior e o Pleistoceno Médio (entre 7 e 0,5 Maa). Apesar de os resultados filogeográficos, genealógicos e populacionais serem em parte concordantes com a hipótese de que *R. prolixus* e *R. robustus s.l.* se originaram na região do Orinoco e se diferenciaram possivelmente devido às modificações naturais na região pan-amazônica promovidas pelas ‘incursões marinhas’ do Mioceno, as datações obtidas para os eventos cladogenéticos das linhagens são concordantes com os períodos glaciais em que as florestas tropicais úmidas teriam sido reduzidas a ‘refúgios’, durante o Plioceno e Pleistoceno.

## ABSTRACT

One of the current challenges for the interruption of *Trypanosoma cruzi* transmission relies primarily on re-infestation events of insecticide-treated houses by sylvatic populations of triatomines and secondarily on the sporadic contact between infected sylvatic triatomines and susceptible humans. The scrutinization of regions with endemic or hypoendemic transmission depends on accurate species identification and its geographical boundaries. In the case of cryptic species, devising of simple, efficient and cost-effective methods for their identification is needed. Biological, ecological and evolutionary processes that can be involved in vectorial domiciliation and speciation are also relevant for the development of innovative surveillance/control strategies. These processes are embedded in a complex spatial and temporal context, which can be appreciated when historical and geographical approaches are considered in different scales. In this context, the cryptic species complex *R. prolixus* and *R. robustus s.l.* serves as a good model, as it is an extensive-studied group, with well-known phylogeny, and its geographical limits were previous estimated. During this thesis the taxonomic status of these lineages were evaluated and also possible natural barriers for gene flow were determined, based on the following analyses: (1) phylogenetic, using one mitochondrial (Cytb) and five nuclear markers (ITS-2, *period*, *timeless*, *cryptochrome2-I6*, *cryptochrome2-I8*); (2) biogeographical, using a ‘relaxed’ molecular clock with two geological calibrations; (3) population genetics, analyzing 10 microsatellite loci; and (4) behavioral, measuring locomotor activities and estimating the circadian clock periods. Moreover, it was described a non-coding, single-copy nuclear DNA fragment that contains a single-nucleotide polymorphism (SNP) with the potential to distinguish the important domestic Chagas disease vector, *Rhodnius prolixus*, from members of the sylvatic *Rhodnius robustus* cryptic lineages complex. Two hundred and fifty four specimens of *R. prolixus* and *R. robustus s.l.* were obtained. *R. robustus s.l.* sampling were conducted along transects which cross areas of contiguous geographic distribution among *R. robustus* III and IV, and *R. robustus* II, III, and IV, based on the previous knowledge about lineages boundaries. *R. prolixus* is a paraphyletic assemblage of, at least, seven different lineages, three of which seem to be different species – *R. barretti*, *R. robustus* I, and *R. robustus* V. Phylogenetic results of *R. robustus* II, III, IV, and n.sp. showed incongruent topologies on gene trees, and thus they represent a complex of species *in status nascendi*. Ancestral populations of *R. prolixus* and *R. robustus s.l.* cladogenesis have probably occurred during geological events between Late Miocene and Middle Pleistocene (7-0.5 Mya). Although phylogeographical, genealogical, and population results partially agree with the predictions that *R. prolixus* and *R. robustus s.l.* were originated in the Orinoco biogeographical area and speciated during ‘marine incursions’ of the Miocene Epoch, ‘relaxed’ clock analyses showed recent splits estimations, which coincides with the glacial time that occurred the fragmentation of moist forests in ‘refuges’ during Late Pliocene and Early/Middle Pleistocene.

## 1. INTRODUÇÃO

### 1.1. Doenças tropicais negligenciadas - a doença de Chagas

As doenças tropicais negligenciadas estão presentes em regiões financeiramente desprovidas das Américas, África, Ásia e Oceania, afetando aproximadamente um bilhão de pessoas (Kline et al., 2013; OMS, 2010). Cinco das dezessete enfermidades inseridas neste contexto são causadas por patógenos cujos vetores são insetos hematófagos e estão presentes na América Latina – doença de Chagas, dengue, filariose, leishmaniose e malária. O plano global de combate destas doenças estabelece como imprescindível o desenvolvimento de novas estratégias para o controle vetorial, que deve ser precedido de um melhor entendimento da biologia destes insetos (OMS, 2007).

A estratégia utilizada nos últimos 20 anos para o controle dos vetores da doença de Chagas serviu de aprendizado para o desenvolvimento de métodos de vigilância mais eficazes. Embora a maioria dos triatomíneos (Hemiptera: Reduviidae) seja potencialmente capaz de se infectar naturalmente pelo agente etiológico *Trypanosoma cruzi*, os programas de controle vetorial concentraram seus esforços na erradicação dos vetores epidemiologicamente mais relevantes – os de comprovada adaptação aos domicílios humanos – como *Triatoma infestans* e *Rhodnius prolixus*, através da borrifação de inseticidas nas casas infestadas (Dias, 2009). Apesar da erradicação “virtual”<sup>1</sup> ou efetiva de *T. infestans* no Uruguai, Chile, Brasil, leste do Paraguai, sul do Perú e algumas partes da Argentina, e de *R. prolixus* no sul do México, Guatemala, Honduras, Nicarágua, El Salvador e Costa Rica (sumarizados em Abad-Franch et al., 2013), ainda se registram anualmente na América Latina 50.000 novos casos e 12.500 mortes (Rassi et al., 2010; Senior, 2007).

A principal dificuldade encontrada para se interromper definitivamente a transmissão da doença de Chagas está associada a eventos de reinfestação das casas previamente tratadas com inseticidas. Estes eventos podem ser resultado da migração de populações silvestres de espécies vetoras primárias, como também de espécies secundárias nativas, capazes de formar colônias no interior dos domicílios (Cecere et al., 2006; Fitzpatrick et al., 2008).

---

<sup>1</sup> Neste caso, eliminação “virtual” se refere ao fato da espécie não ter sido mais encontrada, o que logicamente não quer dizer que a espécie tenha sido erradicada no local (Abad-Franch et al. 2013).

Na região do Chaco da Bolívia, Paraguai e Argentina, populações silvestres e peridomésticas de *T. infestans* comumente reinfestam domicílios previamente tratados (Abad-Franch et al., 2010b; Ceballos et al., 2011; Quisberth et al., 2011), assim como fazem as de *R. prolixus* na Venezuela e Colômbia (Fitzpatrick et al., 2008; Guhl et al., 2009). Espécies de Triatominae como *Rhodnius pallescens*, *Rhodnius stali*, *Rhodnius neglectus*, *Rhodnius nasutus*, *Rhodnius ecuadoriensis*, *Triatoma maculata*, *Triatoma carrioni*, *Panstrongylus geniculatus* e *Panstrongylus herreri* têm a capacidade de colonizar ambientes artificiais em algumas regiões de suas áreas de ocorrência (Abad-Franch e Monteiro, 2007; Aguilar et al., 1999; Coura et al., 2002). Situações envolvendo eventos de reinfestação são mais complexas e seu controle requer vigilância contínua com participação das próprias comunidades e retratamento de qualquer novo foco doméstico que venha a ser detectado.

### 1.2. A doença de Chagas na região amazônica

O método tradicional de controle vetorial utilizando inseticidas é ineficaz em áreas onde a presença de vetores domiciliados é rara, como na região amazônica. Esta região abriga mais de 25 espécies vetoras da doença de Chagas pertencentes a nove gêneros e, com a exceção de algumas populações isoladas de *T. maculata*, *P. geniculatus*, *P. herreri*, e *R. stali*, a grande maioria dos vetores apresenta hábitos silvestres (Abad-Franch e Monteiro, 2007).

A ocupação desordenada do ambiente natural pode ser citada como o principal motivo de o ser humano ocasionalmente entrar em contato com estes vetores e se infectar. A construção de domicílios em regiões rurais e periurbanas próximas às florestas e o desmatamento seletivo no peridomicílio (*i.e.* não derrubada de palmeiras) favorecem invasões esporádicas das casas por vetores silvestres adultos que habitam palmeiras (Abad-Franch et al., 2010a). Como a incidência da doença é baixa e a transmissão neste contexto se caracteriza como sendo esporádica, pode-se dizer que a doença de Chagas na região amazônica é hipoendêmica; a prevalência da infecção por *T. cruzi* provavelmente está entre 1% e 4% (Aguilar et al., 2007).

As descobertas de focos de transmissão mais intensos na Amazônia, relacionados à exploração de recursos naturais e ingestão do protozoário em alimentos contaminados, são cada vez mais frequentes e traduzem a real importância do contato

entre vetores silvestres e seres humanos para a transmissão da doença. Nestes cenários epidemiológicos mais pontuais (que sobrepõem àquele hipoendêmico), extrativistas que chegam a passar meses no interior das florestas coletando fibras das palmeiras de piaçava (*Leopoldina piassaba*) são frequentemente atacados por *Rhodnius brethesi* infectados (Coura et al., 2002). Outro exemplo é o de famílias inteiras que são infectadas oralmente após a ingestão acidental do parasito em sucos contaminados (Valente et al., 1999). De 1968 a 2000, mais da metade dos casos agudos de doença de Chagas relatados na Amazônia brasileira (111 de um total de 205 casos) foram atribuídos a surtos de infecção oral. Entre 2002 e 2008, foram identificados outros 568 casos, representando mais de 75% do total neste período (Coura e Junqueira, 2012).

Frente a este panorama novo e complexo de transmissão, autoridades e pesquisadores perceberam a importância de se entender melhor a dinâmica populacional, a ecologia e o comportamento não apenas dos vetores primários, como também dos vetores silvestres (Abad-Franch e Monteiro, 2005). A correta identificação taxonômica das populações de triatomíneos envolvidas na transmissão da doença é o alicerce para o desenvolvimento de novas estratégias de controle e vigilância do vetor. O maior problema deriva do fato de que muitas destas espécies são crípticas (i.e. morfologicamente muito semelhantes). A correta identificação é importante, pois mesmo sendo morfologicamente semelhantes, elas podem apresentar capacidades vetoriais distintas (Pavan e Monteiro, 2007).

### 1.3. Os vetores da doença de Chagas

A subfamília Triatominae se insere na ordem Hemiptera e família Reduviidae (Lent e Wygodzinsky, 1979) e é composta pelas tribos Triatomini, Rhodniini, Cavernicolini, Bolboderini e Alberproseniini. Apenas as duas primeiras tribos incluem as principais espécies epidemiologicamente relevantes (*T. infestans*, *T. dimidiata*, *Triatoma brasiliensis brasiliensis*, *R. prolixus*, entre outras).

Triatominae é a única subfamília estritamente hematofágica pertencente a essa ordem. A maioria dos triatomíneos se alimenta preferencialmente em animais endotérmicos, embora existam casos em que também suguem o sangue de animais exotérmicos ou ainda a hemolinfa de invertebrados (Brenière et al., 2004; Christensen

e de Vasquez, 1981; Freitas et al., 2005; Salvatella et al., 1994; Sandoval et al., 2004). Estes insetos possuem desenvolvimento hemimetabólico com cinco estádios ninfais, além da fase adulta (Lent e Wygodzinsky, 1979).

A monofilia desta subfamília ainda é uma questão controversa na literatura. Hipóteses conflitantes apóiam a origem monofilética, polifilética e parafilética da hematofagia na família Reduviidae.

Em uma extensa revisão morfológica, Lent e Wygodzinsky (1979) analisaram 133 das 137 espécies de Triatominae (inseridas em 18 gêneros de seis tribos) e concluíram que esta subfamília é monofilética. Apesar de a maioria das espécies ser autóctone das Américas (120 ao todo), 13 são encontradas no sul e sudoeste da Ásia, e na Oceania. Contudo, nenhuma espécie autóctone foi observada na África.

Com o objetivo de solucionar este “enigma” biogeográfico referente à ausência de triatomíneos nativos na África, Schofield (1988) questionou a monofilia da subfamília. A partir de estudos ecológicos, foi postulado que múltiplas transições da predação à hematofagia ocorreram após eventos independentes de exploração de ninhos de invertebrados por diferentes linhagens de reduviídeos predadores e, portanto, uma origem polifilética deveria ser mais plausível.

Posteriormente, a monofilia de Triatominae foi reafirmada com base em novas análises de caracteres morfológicos, incluindo na topologia do cladograma de Triatominae táxons predadores da família Reduviidae (Weirauch, 2008), e análises moleculares (Hypsa et al., 2002; Patterson e Gaunt, 2010; Weirauch e Munro, 2009).

Recentemente, Hwang e Weirauch (2012) defenderam a parafilia de Triatominae em relação com o gênero *Opisthacidius* com base em análises moleculares (análise de sequências das regiões ribossomais 16S, 18S, 28S D2, 28S D3-D5 e do gene nuclear *wingless*). Esta foi a reconstrução filogenética de Reduviidae mais abrangente até o momento, na qual foram amostrados 178 táxons pertencentes a 18 subfamílias [seguida de Patterson e Gaunt (2010), que analisaram 85 táxons de 12 subfamílias].

A tribo Rhodniini está inserida na subfamília Triatominae e é composta pelo gênero parafilético *Rhodnius* (que possui 16 espécies formalmente descritas), e o gênero *Psammolestes* (três espécies), que se agrupa dentro do clado composto pelas

espécies do gênero *Rhodnius* (Galvão et al., 2003; Hypsa et al., 2002; Monteiro et al., 2002). Espécies nativas desses vetores podem ser encontradas em diferentes ecótopos silvestres, como em copas de palmeiras, ninhos de aves, tocas de mamíferos, e eventualmente, no peridomicílio e no interior de casas (Lent e Wygodzinsky, 1979).

### 1.3.1. O gênero *Rhodnius* na região pan-amazônica

Acredita-se que o estoque ancestral do gênero *Rhodnius* provavelmente teria surgido nas florestas pan-amazônicas (região amazônica somada à região do Orinoco) em associação com palmeiras e com a fauna habitante deste ecótopo (Patterson e Gaunt, 2010). Atualmente 11 das 16 espécies de *Rhodnius* ocorrem em todas as oito áreas biogeográficas (sensu Morrone, 2006) constituintes da região pan-amazônica (Amazônia + bacia do Orinoco) – regiões do Orinoco, Guiana, Napo, Imerí, Inambari, Rondônia, Pará e Belém (Abad-Franch et al., 2009).

O gênero *Rhodnius* apresenta duas grandes linhagens evolutivas principais, das quais as espécies atuais se originaram - a ‘linhagem *pictipes*’ e a ‘linhagem *robustus*’ (Abad-Franch e Monteiro, 2007). A ‘linhagem *pictipes*’ inclui cinco espécies da vertente oriental da Cordilheira dos Andes (‘grupo’ *cis*-Andino), na região pan-amazônica – *R. pictipes*, *R. stali*, *R. brethesi*, *R. amazonicus* e *R. paraensis*; e três da vertente ocidental dos Andes (‘grupo’ *trans*-Andino) – *R. pallescens*, *R. colombiensis* e *R. ecuadoriensis*. A ‘linhagem *robustus*’ compreende as espécies *R. prolixus*, *Rhodnius robustus*, *Rhodnius milesi*, *Rhodnius dalessandroi*, *R. nasutus*, *R. neglectus*, *Rhodnius domesticus*, *Rhodnius neivai*, *Psammolestes arthuri*, *Psammolestes tertius* e *Psammolestes coreodes*, sendo as seis primeiras espécies amazônicas. Como as espécies de *Rhodnius* da ‘linhagem *robustus*’ estão filogeneticamente mais próximas das espécies do gênero *Psammolestes* do que das demais espécies de *Rhodnius* da ‘linhagem *pictipes*’, pode-se dizer que este gênero é parafilético (Monteiro et al., 2002).

Estudos filogenéticos recentes apontam que algumas espécies do gênero *Rhodnius* são, na verdade, complexos de espécies crípticas. Um complexo de espécies da ‘linhagem *pictipes*’ e outro pertencente a ‘linhagem *robustus*’ foram descobertos a partir do sequenciamento e análise de um marcador nuclear, o segundo espaçador intergênico ribossomal (ITS-2), e de um fragmento do gene mitocondrial citocromo b

(mtCytb) (Monteiro et al., 2003; Pavan, 2009). O emprego de técnicas moleculares para identificação de espécies crípticas é imprescindível para a geração de resultados mais robustos e confiáveis.

#### *1.4. Técnicas moleculares para identificação de espécies crípticas de triatomíneos*

A sistemática, de acordo com Mayr (1969), é a área da biologia destinada à detecção e descrição de espécies (sistemática alfa), ao estudo do relacionamento filogenético (sistemática beta) e à análise da variabilidade genética em populações de uma ou mais espécies (sistemática gama). A sistemática molecular é o produto da sistemática e da genética molecular. Informações genéticas hereditárias podem ser úteis na identificação de espécies crípticas e na geração de conhecimento dos padrões naturais (como a determinação das regiões geográficas de ocorrência de uma espécie e variabilidade genética de suas populações), que por sua vez servirão de base para o entendimento de processos evolutivos (como a formação de barreiras naturais de isolamento reprodutivo, eventos de especiação ou ainda resistência vetorial aos inseticidas) (Gourbière et al., 2012). Existe ainda a possibilidade de usar a informação genética para inferir a história evolutiva do organismo e identificar as relações de parentesco em qualquer nível taxonômico.

A sistemática molecular foi marcada há quatro décadas por uma grande descoberta metodológica para análise de genótipos: a mobilidade eletroforética de enzimas, ou isoenzimas (Hubby e Lewontin, 1966). As isoenzimas são codificadas por diferentes alelos de um determinado *locus* e, por ser um marcador co-dominante, é possível visualizar três fenótipos distintos como resultado: homozigoto para um alelo, homozigoto para outro alelo, ou heterozigoto. Contudo, é esperado para qualquer par de espécies filogeneticamente muito próximas que o número de *loci* polimórficos seja baixo (e.g. Harry, 1993, 1994; Monteiro et al., 2009), uma vez que as isoenzimas revelam apenas uma pequena parte da variação do DNA. As variantes de uma mesma enzima são resultado de modificações nas sequências de aminoácidos que, por sua vez, podem alterar as propriedades físico-químicas da enzima. Quando essas propriedades físico-químicas são alteradas, a mobilidade eletroforética das isoenzimas é modificada. Portanto, substituições sinônimas ou modificações de

aminoácidos que não resultem na alteração de carga elétrica ou tamanho da molécula não podem ser detectadas nas análises de isoenzimas.

O segundo grande marco na sistemática molecular foi a descrição do método de amplificação do DNA *in vitro*, desenvolvido por Kleppe e colaboradores (1971) e posteriormente aprimorado por Kary B. Mullis (Saiki et al., 1986) – a reação em cadeia da polimerase (PCR). Com a utilização da PCR é possível amplificar determinados genes a partir de quantidades ínfimas de material biológico ou de amostras muito antigas ou mal preservadas. Durante as décadas de 80 e 90 a técnica de isoenzimas foi sendo substituída gradativamente por técnicas moleculares baseadas na PCR capazes de fornecer informações sobre o genótipo dos indivíduos analisados: RAPD (sigla em inglês que significa “DNA polimórfico amplificado ao acaso”), PCR-RFLP (sigla em inglês que significa “Polimorfismo de tamanho dos fragmentos de restrição”), análise de microssatélites e o sequenciamento de DNA. Particularmente estas duas últimas técnicas vêm sendo fundamentais para o avanço do conhecimento sobre os triatomíneos (e.g. Calleros et al., 2010; Fitzpatrick et al., 2008; Hypsa et al., 2002; Lyman et al., 1999; Marcilla et al., 2002; Monteiro et al., 2003; Monteiro et al., 2004).

#### 1.4.1. Análise de microssatélites em estudos populacionais

Estudos de genética de populações são úteis para a observação da distribuição da variabilidade alélica (e sua oscilação no tempo); em última análise, servirão para avaliar o papel de processos evolutivos, como mutação, deriva, seleção natural e migração, na geração de variabilidade genética e estruturação de populações naturais. Investigações desta natureza utilizam um arcabouço teórico e metodológico capaz de responder questões microevolutivas.

Microssatélites são *motifs* de dois a seis pares de bases, que se repetem sequencialmente (*in tandem*) no DNA nuclear (ex.: (GT)<sub>n</sub>, (CTA)<sub>n</sub>). Acredita-se que a variabilidade elevada no número de repetições seja decorrente do fenômeno conhecido como *DNA slippage* (ou “escorregada” da polimerase, que leva à incorporação ou deleção errônea desses *motifs* durante a replicação do DNA) e recombinação (*crossing-over* desigual ou conversão gênica) (Ellegren, 2004). Estas regiões evoluem mais rápido em até cinco ordens de grandeza que as demais regiões

dos genomas nuclear ou mitocondrial (taxas de substituição nucleotídica médias em insetos de  $10^{-6}$ - $10^{-2}$  por geração em comparação com  $10^{-9}$ - $10^{-8}$ , respectivamente).

Existem basicamente três modelos propostos para explicar a evolução dos *loci* de microssatélites: (1) envolvendo apenas um passo mutacional por vez – modelo SSM (*single step mutation*), de Kimura e Otha (1978); (2) com passos mutacionais múltiplos – IAM (*infinite allele model*), de Kimura e Crow (1964); e (3) o modelo dependente do tamanho alélico, onde alelos pequenos tendem a aumentar em número de repetições, enquanto alelos maiores tendem a diminuir, de Ellegren (2004). Embora ainda haja discussão sobre qual modelo evolutivo seria o mais apropriado, os microssatélites são hoje os marcadores mais utilizados em estudos populacionais, pois apresentam uma série de vantagens, como por exemplo: (1) são neutros; (2) possuem herança biparental (o que permite identificar possíveis híbridos); (3) são altamente polimórficos; (4) são co-dominantes; e (5) a metodologia usada admite a análise de vários *loci* de uma só vez.

O número de estudos publicados com triatomíneos utilizando microssatélites vem crescendo e já existem iniciadores para a amplificação de muitos *loci* para *R. pallenscens* (Harry et al., 1998), *T. dimidiata* (Anderson et al., 2002), *T. infestans* (Garcia et al., 2004; Marcet et al., 2006), *T. pseudomaculata* (Harry et al., 2008a), *T. brasiliensis* (Harry et al., 2009) e *R. prolixus* (Fitzpatrick et al., 2008; Harry et al., 2008b). Com a disponibilidade de iniciadores para tantos *loci*, trabalhos genético-populacionais podem ser privilegiados com a utilização de microssatélites como marcadores, o que leva à possibilidade de uma observação mais nítida do fluxo gênico entre populações.

#### 1.4.2 Sequenciamento de DNA

A técnica de sequenciamento de DNA pode ser considerada como a mais informativa para estudos em sistemática alfa, beta e gama. Por utilizar como matéria-prima o conteúdo genético propriamente dito, os nucleotídeos, o sequenciamento de DNA é objetivo e aplicável em qualquer nível taxonômico, desde que se escolha a região gênica mais adequada.

#### 1.4.2.1. Análise de sequências nucleares

O genoma nuclear pode ser dividido, de maneira simples, em íntrons e éxons. Os éxons são regiões codificantes de aminoácidos e são relativamente mais conservados entre os táxons, quando comparados aos íntrons ou ainda ao genoma mitocondrial (Avice, 2000). Esta menor taxa de mutação se deve a um mecanismo presente em regiões codificantes do núcleo (e ausente em mitocôndrias) de revisão e reparo (mecanismo de *proofreading*). Quando a DNA polimerase insere erroneamente um nucleotídeo na cópia de uma fita-molde de DNA, proteínas de “reparo” revisam os nucleotídeos inseridos em éxons e identificam os possíveis erros. Caso exista algum, exonucleases clivam a região complementar errada e a polimerase refaz o fragmento. Algumas dessas inserções nucleotídicas errôneas (que ocorrem aleatoriamente) permanecem nos éxons e são chamadas de mutações.

De acordo com o modelo de evolução neutra de Kimura (1968), a maioria das mutações que ocorre nos éxons é sinônima (ou silenciosa), ou seja, não resulta na alteração do aminoácido. Considerando as mutações não sinônimas, pode-se dizer que a maioria é desvantajosa (i.e. prejudica a aptidão do indivíduo) e, portanto, é removida da população por seleção natural purificadora. As demais mutações são neutras ou vantajosas. As neutras são a maioria e não interferem na aptidão do organismo e sua frequência oscila na população a cada geração. Já as vantajosas são selecionadas positivamente e podem aumentar em frequência na população.

O sequenciamento das regiões nucleares codificantes é utilizado, na maioria das vezes, em estudos filogenéticos de espécies mais distantes. Este foi o caso nas análises filogenéticas de insetos da família Reduviidae, onde foram utilizados genes ribossomais nucleares (18S e 28S) e éxons do gene *wingless* (Hwang e Weirauch, 2012; Patterson e Gaunt, 2010; Weirauch e Munro, 2009).

Contudo, existem ainda alguns genes que apresentam taxas mutacionais superiores àquelas encontradas nos genes descritos acima e que já foram utilizados em estudos filogenéticos para identificação de espécies crípticas. A aplicação de genes de relógio circadiano como marcadores moleculares já tem sido satisfatória em estudos de sistemática molecular de espécies incipientes ou filogeneticamente próximas (Araki et al., 2009; Bauzer et al., 2002a; Bauzer et al., 2002b; Mazzoni et al., 2006; Rona et al., 2010; Tauber e Kyriacou, 2005; Wheeler et al., 1991). A alta

variabilidade genética encontrada pode estar associada às suas funções no controle de ritmos circadianos ligados ao comportamento sexual e ao isolamento reprodutivo temporal (Sakai e Ishida, 2001; Tauber et al., 2003). O funcionamento do ciclo circadiano é explicado em maiores detalhes na seção 1.6.1. *A genética dos ritmos circadianos.*

Estudos evolutivos em *Drosophila* e em outros dípteros relacionados revelaram que algumas regiões do gene de relógio circadiano *period* (*per*) estão entre as sequências que mais rapidamente evoluem em insetos, e que esta alta taxa de mutação parece ter consequências funcionais importantes (Peixoto, 2002). Ortólogos de *per*, *timeless* (*tim*), *clock* (*Clk*) e *cycle* (*cyc*) têm sido muito úteis como marcadores moleculares nos estudos evolutivos de complexos de espécies de insetos vetores, tanto em flebotomíneos (Araki et al., 2009; Meireles-Filho et al., 2006b; Vigoder et al., 2010) como em mosquitos (Rona et al., 2009; 2010).

Os íntrons são regiões não codificantes do genoma e, por sofrerem menor pressão seletiva, são mais variáveis que os éxons. A análise de íntrons pode ser útil em estudos de taxonomia alfa, beta e até gama (Irimia e Roy, 2008; Ohshima e Yoshizawa, 2011). Contudo, existem ainda alguns íntrons que são altamente conservados, por estarem envolvidos em certas funções pós-transcricionais e podem não apresentar diferenças suficientes para separar espécies proximamente relacionadas (Gazave et al., 2007).

Regiões não codificantes ribossomais também são muito úteis em análises filogenéticas de espécies proximamente relacionadas. O primeiro e segundo espaçadores ribossomais internos (ITS-1 e ITS-2), localizados entre as subunidades ribossomais 18S e 28S já foram utilizados em triatomíneos para: (1) filogenia de espécies do gênero *Panstrongylus* (Marcilla et al., 2002) e *Mepraia* (Calleros et al., 2010), além dos complexos *phyllosoma* (Martinez et al., 2006) e *dimidiata* (Bargues et al., 2008; Dorn et al., 2009) de espécies crípticas; (2) filogeografia de *T. infestans* (Bargues et al., 2006); (3) taxonomia alfa de *R. stali* (Justi et al., 2010) e *T. phyllosoma* (Villalobos et al., 2011); e (4) genética de populações de *T. rubrovaria* na Argentina, Brasil e Uruguai (Pacheco et al., 2007) e *T. infestans* na Bolívia (Piccinali et al., 2011; Quisberth et al., 2011; Waleckx et al., 2011).

#### 1.4.2.2 Análise de sequências mitocondriais

O genoma mitocondrial não apresenta íntrons e é formado por um único cromossomo circular, que contém genes codificantes de proteínas estruturais e funcionais relacionadas ao transporte de elétrons e fosforilação oxidativa. Por não possuir um mecanismo de reparo do DNA eficiente, sua taxa de evolução pode ser até 10 vezes mais rápida que a do genoma nuclear.

Alguns genes mitocondriais, como Cytb, COI e COII, são muito utilizados na separação filogenética de espécies proximamente relacionadas, como já demonstrado em trabalhos com *R. prolixus* e *R. robustus s.l.* (Márquez et al., 2011; Monteiro et al., 2003), *T. rubida* e *T. recurva* (Pfeiler et al., 2006), espécies dos complexos *T. brasiliensis* (Mendonça et al., 2009; Monteiro et al., 2004), *R. pictipes* (Pavan, 2009) e espécies do gênero *Mepraia* (Calleros et al., 2010). Estes genes também já foram utilizados em análises populacionais de *T. sanguisuga* (de la Rúa et al., 2011) e *T. infestans* (Ceballos et al., 2011; Giordano et al., 2005).

Existem ainda genes mitocondriais mais conservados, como aqueles codificantes das subunidades ribossômicas 12S e 16S, que já foram utilizados em estudos filogenéticos do gênero *Triatoma* (Garcia e Powell, 1998), de espécies da tribo Rhodniini (de Paula et al., 2007), espécies da subfamília Triatominae (Hypsa et al., 2002) e ainda de insetos pertencentes à família Reduviidae (Gaunt e Miles, 2002; Hwang e Weirauch, 2012; Patterson e Gaunt, 2010; Weirauch e Munro, 2009).

O uso exclusivo de genes mitocondriais como marcadores em estudos de sistemática molecular apresenta algumas limitações, dentre as quais as mais importantes são (1) alta taxa de substituição nucleotídica; em análises contendo espécies filogeneticamente distantes, pode causar um acúmulo de substituições em uma mesma base (i.e. saturação) e consequente homoplasia; (2) a presença no genoma nuclear de pseudogenes mitocondriais (NUMTs), que pode levar a identificação errônea de espécies; e (3) possibilidade de introgressão do material genético mitocondrial de uma espécie em outra, por conta de cruzamentos interespecíficos (revisado em Mas-Coma e Bargues, 2009).

O problema de saturação do marcador mitocondrial pode ser evitado em análises filogenéticas contendo organismos filogeneticamente distantes se os nucleotídeos da terceira base dos códons forem excluídos, pois este é o sítio que

apresenta maiores taxas mutacionais (por conta da degeneração do código genético). Os NUMTs são resultado de translocações de fragmentos mitocondriais do genoma mitocondrial para o genoma nuclear. Uma vez integrados, os fragmentos não funcionais acumulam mutações livremente. Contudo, a presença de NUMTs pode se manifestar de diversas maneiras: presença de bandas inespecíficas nas eletroforeses em géis de agarose das amostras amplificadas pela PCR; presença de códons de parada ou que não codificam aminoácidos; bandas extras em perfis de restrição; posição filogenética anômala das amostras sequenciadas; e proporção significativa de ambiguidades na sequência (i.e. picos duplos em um mesmo sítio nos eletroferogramas), semelhantes àsquelas observadas quando há eventos de heteroplasmia (Hlaing et al., 2009). Especificamente em relação à terceira limitação (introgressão), a utilização de marcadores nucleares de herança biparental em concomitância com os mitocondriais é imprescindível na detecção de híbridos naturais.

A escassez de marcadores nucleares já descritos para triatomíneos que sejam úteis na separação de espécies filogeneticamente próximas dificulta a detecção de linhagens crípticas ainda desconhecidas ou possíveis áreas com a presença de híbridos naturais. O projeto genoma de *R. prolixus*, iniciado em 2006 sob a coordenação de Erwin Huebner, da Universidade de Manitoba, no Canadá, e financiamento dos *National Institutes of Health*, dos EUA, já está próximo ao fim (fase de anotação dos genes) e servirá como base, dentre muitas outras coisas, para o isolamento de novos marcadores nucleares que poderão ser úteis na geração de conhecimento acerca da biologia do animal e no desenvolvimento de novas estratégias de controle vetorial.

#### 1.4.3. Complexos de espécies crípticas de *Rhodnius* na região pan-amazônica

*Rhodnius pictipes* é a espécie do gênero cuja distribuição abrange a maior área na América do Sul (Bolívia, Brasil, Colômbia, Equador, Guiana Francesa, Guiana, Peru, Suriname e Venezuela) (Galvão et al., 2003; Lent e Wygodzinsky, 1979). Esta espécie apresenta grande similaridade cromática e morfológica com *R. stali* e em menor medida com *R. amazonicus*.

Utilizando um marcador mitocondrial (Cytb) e outro nuclear (ITS-2), Pavan (2009) descobriu que a espécie *R. pictipes* representa, na realidade, um complexo

parafilético de, pelo menos, quatro espécies crípticas. As três novas espécies foram encontradas em uma região andina (Sierra Nevada de Santa Marta, Magdalena) e nas terras baixas da Amazônia (*lowlands*), na Colômbia, em Rondônia, Brasil e em Beni, na Bolívia, e são filogeneticamente mais relacionadas com *R. stali* e *R. brethesi*, do que com *R. pictipes* s.s. *R. amazonicus* forma um grupo aparentemente mais basal em relação a estas espécies.

Um outro caso já bastante discutido na literatura se refere às espécies crípticas *R. prolixus* e *R. robustus*. Por muitos anos, a validade taxonômica desta segunda espécie foi questionada, em função de sua semelhança morfológica e isoenzimática com a primeira (Harry, 1993, 1994).

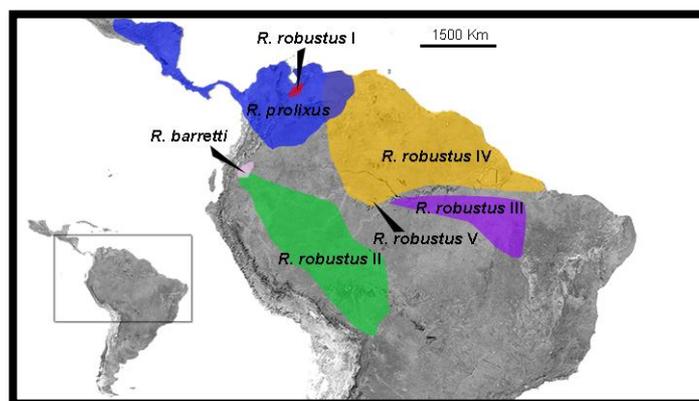
*Rhodnius prolixus* é atualmente um dos vetores epidemiologicamente mais relevantes da doença de Chagas na América Latina, particularmente na Venezuela, Colômbia e (recentemente) América Central (c.f. Hashimoto e Schofield, 2012), por ser capaz de colonizar eficientemente habitações humanas. *Rhodnius robustus* é considerado um vetor secundário, por estar restrito ao ambiente silvestre. Ocasionalmente esta espécie invade domicílios durante a noite (Felicangeli et al., 2002; Pavan e Monteiro, 2007), ou ainda cai acidentalmente em trituradores de frutas, causando microepidemias de transmissão oral, sobretudo na região norte do Brasil (Valente et al., 1999).

Diversos estudos foram realizados com o objetivo de solucionar o impasse taxonômico envolvendo as duas espécies. Os resultados obtidos a partir de análises de eletroforese de proteínas salivares (Soares et al., 1998), RAPD (Felicangeli et al., 2002; Garcia et al., 1998), morfometria tradicional e geométrica (Villegas et al., 2002) e sequenciamento de fragmentos de mtCytb e mt16S (Lyman et al., 1999) levaram à conclusão de que *R. robustus* é uma espécie válida e diferente de *R. prolixus*.

Com o objetivo de estudar as relações filogeográficas entre as duas espécies, Monteiro e colaboradores (2003) analisaram sequências de um fragmento de 663-pb do mtCytb e a região 28S D2 nrDNA de 26 populações provenientes de sete países latino-americanos e concluíram que não apenas *R. robustus* é diferente de *R. prolixus*, como ainda representa um complexo parafilético de quatro linhagens crípticas (*R. robustus* I, II, III e IV; Figura 1.1). Este complexo parafilético *R. robustus* s.l. pôde ser confirmado (juntamente com a identidade separada de *R. prolixus*) quando foi

utilizado um marcador nuclear, o ITS-2 (Lazoski, Pavan e Monteiro, dados não publicados). Para facilitar a identificação molecular deste complexo de espécies crípticas, Pavan e Monteiro (2007) desenvolveram um PCR-Multiplex para diferenciar as espécies simpátricas *R. prolixus* e *R. robustus* I, e ainda *R. robustus* II-IV, submetendo as amostras à uma PCR contendo iniciadores espécie-específicos que amplificam regiões do mtCytb de diferentes tamanhos e posteriormente à eletroforese em um gel de agarose 1,5%.

Abad-Franch e Monteiro (2005; 2007) relataram a presença de um terceiro táxon dentro do grupo monofilético composto por *R. prolixus* e *R. robustus* I (proveniente de uma localidade próxima de Manaus, e chamado de *R. robustus* V) e ainda de outro táxon que seria o mais basal da ‘linhagem *robustus*’, provenientes de áreas silvestres da Amazônia equatoriana. Posteriormente, este táxon mais basal foi encontrado em Puerto Asís, Colômbia, por Márquez e cols. (2011), identificado a partir do sequenciamento de 28S D2 nrDNA e de um fragmento de 663-pb do mtCytb. Portanto, até o momento são descritos na literatura seis linhagens crípticas de *R. robustus s.l.*



**Figura 1.1.** Distribuição estimada de *R. prolixus* e das linhagens crípticas de *R. robustus s.l.* Modificado de Pavan e colaboradores (2013).

### 1.5. A biogeografia e a região pan-amazônica

A biodiversidade amazônica sempre foi uma questão intrigante para os pesquisadores da história natural (cf. Hoorn et al., 2010). A alta variabilidade genética encontrada em *R. robustus s.l.* apenas exemplifica o que é observado com frequência

em outros táxons autóctones (e.g. Aleixo, 2004; Funk et al., 2007; Gamble et al., 2008; Hall e Harvey, 2002). Os fatores e processos naturais ocorridos ao longo do tempo geológico são responsáveis pelo padrão de diversidade observado hoje. Os cenários ecológicos foram modificados ao longo do tempo em decorrência de orogêneses, eventos hidro e orográficos e mudanças climáticas que afetaram linhagens ancestrais. A fragmentação dos habitats ou mesmo sua completa modificação teria ocasionado a fragmentação e/ou migração de populações ancestrais, promovendo a diversificação de populações e colocando espécies em perigo de extinção, ou mesmo ocasionando-a (Abad-Franch e Monteiro, 2007). Portanto, para entender a biologia de uma linhagem, as informações acerca de sua origem, vicariância e/ou dispersão devem ser adicionadas ao conhecimento de seus padrões ecológicos atuais, assim como o fator “tempo”, que deve ser levado em consideração no estudo filogenético. É neste contexto que a filogeografia e a biogeografia se encontram. Estudos macroevolutivos e macroecológicos, como a filogeografia, biogeografia (*i.e.* ecologia histórica) e a modelagem de nichos ecológicos são importantes para estimar a ecologia das espécies, seu potencial de alcance geográfico e, conseqüentemente, mapear o risco de infestação de novas áreas e de infecção por *T. cruzi* em humanos (e.g. Abad-Franch et al., 2009; Bargues et al., 2008; Coutinho et al., 2012; Gurgel-Goncalves et al., 2011; 2012; Monteiro et al., 2003; 2004; Torres-Perez et al., 2011).

A filogeografia é a área da ciência biológica na qual são analisados os princípios e os processos que delineiam a distribuição geográfica de linhagens genealógicas ao longo do tempo geológico (Avice, 2000). A partir destas análises, é possível associar dados climáticos, palinológicos e geológicos para entender os padrões biogeográficos observados na região de estudo. Nas subseções a seguir serão abordados alguns temas relevantes para a compreensão dos padrões e processos geradores da biodiversidade na região amazônica – os modelos de especiação vicariante e dispersionista que nortearam os princípios da biogeografia (1.5.1); as hipóteses biogeográficas vicariantes delineadas para explicar a geração da biodiversidade na região pan-amazônica (1.5.2); e as análises biogeográficas e filogeográficas atuais que têm como objetivo estimar os centros de origem das linhagens ancestrais, as rotas de dispersão das populações desta linhagem e o tempo geológico destes eventos (1.5.3). Ao fim desta seção é apresentado o conhecimento

filogeográfico já publicado acerca do complexo de espécies crípticas *R. robustus s.l.* e *R. prolixus* (1.5.4) – objetos de estudo desta tese.

#### *1.5.1. Conceitos biogeográficos segundo os modelos dispersionista e vicariante*

A influência de muitos conceitos evolutivos publicados por Charles Darwin em “A Origem das Espécies” (1859), como os de origem, dispersão e adaptação das espécies, e ancestralidade comum, foi extraordinária para o entendimento do fenômeno de especiação. Com base nestes conceitos, foram criadas as premissas que nortearam o modelo dispersionista de especiação (cf. Croizat et al., 1974): (1) qualquer táxon observado nos dias atuais deve provir de uma área geográfica restrita, o centro de origem de sua linhagem ancestral; (2) se uma linhagem se adapta ao ambiente, naturalmente a população se expande e parte de seus integrantes ultrapassa as barreiras deste centro de origem por migração, seja para localidades próximas ou distantes; e (3) o grau de sucesso migratório é diretamente proporcional à capacidade de dispersão de cada táxon.

A crítica principal a este modelo foi dirigida à premissa dos centros de origem, uma vez que não é possível testá-los experimentalmente. O centro de origem é definido empiricamente como a região dentro da distribuição geográfica de um determinado táxon onde se observa maior variabilidade morfológica ou genética. Contudo, estes centros de origem de um único táxon não oferecem informações para o estabelecimento de padrões mais gerais, sendo necessário para a análise biogeográfica o estabelecimento de congruência de vários outros táxons. Adicionalmente a ideia de que organismos atuais, sobretudo aqueles com ampla distribuição geográfica, teriam surgido a partir de um pequeno número de indivíduos em uma área geográfica restrita era considerada como improvável (Craw, 1978). Na realidade era pouco parcimonioso acreditar que estas pequenas populações teriam posteriormente migrado (ativa ou passivamente) para uma área geográfica muito mais extensa. A partir da década de 60, o dispersionismo foi perdendo forças e na década de 80 surgiu um novo modelo biogeográfico, com base no princípio de vicariância.

As ideias no modelo vicariante têm como base o modelo de especiação alopátrica para explicar a geração da biodiversidade. Ao contrário do modelo dispersionista, a principal premissa neste modelo é a de que distribuições geográficas

disjuntas de grupos-irmãos seriam uma consequência do surgimento (no passado) de uma barreira geográfica que teria isolado reprodutivamente populações de um estoque ancestral com ampla distribuição geográfica. A segunda premissa tem como base o princípio de uniformidade dos padrões filogenéticos de táxons presentes em uma mesma região (ou princípio de vicariância). Se os eventos geológicos de uma determinada área geográfica levaram uma linhagem à especiação, outros organismos desta região teriam sofrido repercussão similar, em termos de isolamento reprodutivo.

A biogeografia de vicariância substituiu gradualmente o modelo dispersionista em função de três grandes acontecimentos, segundo Wiley (1988): (1) a elaboração da hipótese panbiogeográfica e os princípios vicariantes de Croizat (1958); (2) o surgimento da teoria da tectônica de placas; e (3) a aplicação da sistemática filogenética às ideias de Croizat.

Um dos maiores desafios para os biogeógrafos até a metade do século passado era o de encontrar uma explicação plausível para a ocorrência de um mesmo táxon (ou organismos morfologicamente semelhantes) em áreas geográficas distantes, como em regiões tropicais do hemisfério sul e regiões temperadas do hemisfério norte. Os dispersionistas explicavam que tais distribuições eram resultado da dispersão progressiva da biota a partir de um centro de origem localizado no hemisfério norte, com três rotas independentes pela América do Sul, África e Indonésia (revisado em Craw, 1978).

A partir da premissa de que o tamanho e a distribuição dos continentes e oceanos no passado eram idênticos aos dos dias atuais, na hipótese de diversificação da biota segundo o modelo dispersionista foram feitas duas predições:

(1) a dispersão da biota sempre ocorre de regiões extensas para pequenas áreas (hipótese de dispersão “continente → ilhas”). Portanto, a hipótese “norte → sul” deriva do fato de o hemisfério norte ser territorialmente maior;

(2) a seleção natural era exercida mais fortemente em lugares com climas inóspitos. As alterações climáticas de aridez e frio no centro de origem localizado no Holártico teriam sido responsáveis pela origem de novas formas de vida. Por conseguinte, o hemisfério norte teria sido povoado por estes novos táxons e os grupos mais basais teriam migrado para o hemisfério sul.

A ascensão do conceito vicariante colocava em cheque a idéia de migração de organismos ancestrais bem adaptados. Ao postular a hipótese panbiogeográfica, Croizat (1958) seria o primeiro biogeógrafo a considerar que as barreiras geográficas eram peças-chave na evolução dos organismos – um princípio evolutivo antes negligenciado nesta área, o de especiação por vicariância.

Os conceitos de movimento das placas tectônicas e da deriva continental fortaleceram os pensamentos biogeográficos vicariantes acerca da geração da biodiversidade. Os princípios geológicos somados ao modelo vicariante resultariam na formulação de uma nova hipótese de origem e dispersão da biota no planeta. As linhagens ancestrais teriam ocorrido em uma grande área geográfica, a Pangéia, e suas populações teriam se separado com o rompimento da região em dois supercontinentes (Gondwana e Laurasia). Esta separação teria levado às populações a eventualmente especiareem em alopatria (revisado em Craw, 1978).

O princípio de vicariância se consolidou no pensamento biogeográfico a partir da criação da sistemática filogenética, o estudo da história de grupos monofiléticos no tempo e no espaço, levando em consideração eventos de cladogênese, anagênese, alopatria e simpatria (Wiley, 1988). Utilizando os princípios filogenéticos, finalmente foi possível testar hipóteses biogeográficas sob a luz da evolução, levando em consideração ancestralidade comum, eventos geológicos e tamanhos populacionais.

Os conceitos de especiação segundo os modelos dispersionista e vicariante ao longo dos anos foram aplicados em diferentes esferas da biogeografia. A biogeografia pode ser dividida em dois grandes níveis, em termos de escala, padrão e processos envolvidos na formação desses padrões: um nível histórico (biogeografia histórica), que se destaca pelo estudo do padrão de distribuição congruente de táxons distintos, e um nível ecológico (biogeografia ecológica), no qual é estudado o padrão de distribuição individual dos organismos (De Candolle, 1820 apud Avise, 2000).

### *1.5.2. Hipóteses biogeográficas aplicadas à região amazônica*

A bacia amazônica é a maior bacia hidrográfica do planeta, ocupando mais de oito milhões de quilômetros quadrados de área total, o que representa mais de um terço da América Latina (UNEP, 2004). Esta região abriga mais de 40.000 plantas,

2.000 espécies de aves e mamíferos e dois milhões e meio de espécies de insetos (De Queiroz e De Carvalho, 2008). Esta bacia, em conjunto com a região do Orinoco, forma a pan-amazônia.

A biogeografia histórica da pan-amazônia é amplamente estudada e dezenas de hipóteses biogeográficas foram propostas na tentativa de explicar a riqueza de espécies desta região. Nestas hipóteses prevalece o conceito vicariante de especiação. As hipóteses mais conhecidas são (1.5.2.1) a hipótese de ‘incurções marinhas’ (Räsänen et al., 1995), durante o Mioceno; (1.5.2.2) a hipótese de rios atuando como barreiras (Wallace, 1852), do Plioceno/Pleistoceno; e (1.5.2.3) a hipótese de ‘refúgios’ (Haffer, 1969), que teria ocorrido durante o Pleistoceno (Figura 1.2). Além destas, um grande número de outras suposições foram postuladas, na tentativa de elucidar as questões biogeográficas da região amazônica. Na seção (1.5.2.4) *outras hipóteses biogeográficas*, são citadas outras duas.

| Éon         | Era       | Período     | Época               | Idade         | Duração (Maa) |               |               |
|-------------|-----------|-------------|---------------------|---------------|---------------|---------------|---------------|
| Fanerozóico | Cenozóico | Quaternário | Recente ou Holoceno |               | 0,01 - 0      |               |               |
|             |           |             | Pleistoceno         | Superior      | 0,13 - 0,01   |               |               |
|             |           |             |                     | Médio         | 0,78 - 0,13   |               |               |
|             |           |             |                     | Inferior      | 1,81 - 0,78   |               |               |
|             |           | Terciário   | Neógeno             | Plioceno      | Superior      | Gelasiano     | 2,59 - 1,81   |
|             |           |             |                     |               | Médio         | Piacenziano   | 3,60 - 2,59   |
|             |           |             |                     |               | Inferior      | Zancleano     | 5,33 - 3,60   |
|             |           |             |                     | Mioceno       | Superior      | Messiniano    | 7,25 - 5,33   |
|             |           |             |                     |               |               | Tortoniano    | 11,61 - 7,25  |
|             |           |             |                     |               | Médio         | Serravalliano | 13,65 - 11,61 |
|             |           |             |                     |               |               | Langhiano     | 15,97 - 13,65 |
|             |           | Inferior    | Burdigaliano        | 20,43 - 15,97 |               |               |               |
|             |           |             | Aquitano            | 23,03 - 20,43 |               |               |               |

**Figura 1.2.** Idades geológicas da Era Cenozóica (desde a Época do Mioceno). **Maa** - Milhões de anos atrás (Maa). Retirado de Pavan (2009).

#### 1.5.2.1. Hipótese de incurções marinhas

A hipótese mais difundida sobre eventos vicariantes durante o Mioceno Médio e Superior, de 17 a 6,5 milhões de anos atrás (Maa), é a de ‘incurções marinhas’ (Hoorn, 1993, 1994; Lovejoy et al., 2006; Räsänen et al., 1995; Wesselingh et al., 2006). Foi postulada a partir de análises de depósitos sedimentares e de fósseis de organismos marinhos encontrados na Formação de Pebas. Esta Formação está

localizada entre as regiões norte e central dos Andes, mais precisamente nas terras baixas amazônicas entre o Perú e o Brasil, onde os principais afluentes do rio Amazonas foram formados.

A elevação do nível do mar durante o Mioceno Superior (entre 11,6 e 8 Ma), alcançando um máximo de até 100 metros acima do nível atual, teria conectado o Mar do Caribe à Bacia do Paraná, ocasionado alagamentos nas regiões amazônicas baixas (florestas de várzea), levando ao isolamento três grandes blocos de maior altitude: o Escudo Brasileiro, o Escudo das Guianas e a porção leste dos Andes (Colinvaux e De Oliveira, 2001; Hooghiemstra e Van der Hammen, 1998; Latrubesse et al., 2007; Lovejoy et al., 2006). Estoques teriam sido mantidos em isolamento nestes blocos, divergindo geneticamente e, eventualmente, especiando em alopatria. Segundo esta hipótese, estes blocos abrigariam os maiores centros de endemismo no cenário atual da região amazônica (Hoorn et al., 2010; Räsänen et al., 1995).

Para testar a hipótese de incursões marinhas como o evento vicariante responsável pela geração da biodiversidade pan-amazônica, os resultados obtidos a partir de populações neotropicais devem satisfazer as seguintes predições: (a) as linhagens provenientes de cada um dos três blocos de maior altitude devem ser monofiléticas entre si; (b) as populações ancestrais devem ter ocupado um dos blocos; e (c) populações mais derivadas devem se encontrar nas terras baixas da porção oeste amazônica (Conn e Mirabello, 2007).

#### *1.5.2.2. Hipótese de rios atuando como barreiras*

A hipótese de que os rios atuariam como barreira ao fluxo genético das populações presentes em margens opostas foi apresentada primeiramente por Alfred Russel Wallace (1852). O desenvolvimento do sistema de rios amazônicos foi iniciado após o evento de soerguimento dos Andes, durante o Mioceno Inferior, e teria sido extremamente relevante na geração da alta diversidade da biota amazônica, principalmente durante o Plioceno/Pleistoceno, período em que foi formado o maior rio da região, o Amazonas. Se os rios representam um papel importante na especiação de organismos e continuam sendo uma barreira para a dispersão, esta hipótese pode ser testada em populações atuais (Gascon et al., 2000; Patton e da Silva, 2005).

A melhor estratégia de amostragem para que as predições desta hipótese sejam testadas apropriadamente é a de analisar organismos presentes em ambas as margens de um grande rio amazônico, desde a sua nascente até a desembocadura (Conn e Mirabello, 2007; Gascon et al., 2000). A diferenciação genética encontrada, segundo este modelo, deverá ser gradual em ambas as margens, menor na nascente e maior na desembocadura, pois é neste sentido que os rios se alargam (Patton et al., 1994).

A divergência entre as populações presentes nas margens opostas das regiões de terra-firme (áreas não alagáveis) deverá ser maior do que entre populações que ocupam as áreas alagadas (várzea), uma vez que os sedimentos das margens opostas da primeira nunca se misturam, ao contrário do observado nas margens da segunda. No período de seca, as margens em áreas de terra-firme são separadas pela largura do rio somada à largura das florestas planas alagáveis e as margens das áreas de várzea separadas somente pela largura do rio. Durante o período de cheia, as margens em áreas de terra-firme são separadas pela largura do rio, ao passo que as margens em áreas de várzea são inundadas. Desta forma, as populações presentes nas margens das planícies alagadas devem apresentar maior fluxo gênico por consequência da erosão natural dos rios. A deposição de sedimentos nas margens com o passar do tempo modifica a posição do canal ao longo da planície aluvionar, transferindo blocos de várzea de uma margem para a outra (Gascon et al., 2000; Patton e da Silva, 2005).

#### *1.5.2.3. Hipótese de refúgios*

O período Quaternário foi marcado pelos ciclos glaciais e interglaciais ocasionados pela influência astronômica dos ciclos de Milankovitch (Bradley, 1999). Estes ciclos descrevem variações de incidência solar regionais ou planetárias causadas por mudanças graduais no eixo da Terra e em sua órbita em torno do Sol.

Na hipótese dos ‘refúgios’, Haffer (1969) e Vanzolini & Williams (1970) supõem que as glaciações pleistocênicas teriam ocasionado a contração das florestas úmidas em pequenas porções separadas por vegetações de savana. Estes ‘refúgios’ teriam mantido isoladas populações de diversas linhagens ancestrais, levando-as eventualmente à especiação. A suposição de contração das florestas teve como base análises paleobotânicas e geomorfológicas e a localização dos possíveis refúgios foi presumida a partir de dados empíricos resultantes da sobreposição da distribuição

geográfica de borboletas, plantas e pássaros (Haffer e Prance, 2002; Patton et al., 2000).

A hipótese dos ‘refúgios pleistocênicos’ como responsáveis pela geração da diversidade observada atualmente deve ser testada em espécies endêmicas, que apresentem pouca flexibilidade quanto ao tipo de hábitat que ocupam e que satisfaçam, em grande parte, às seguintes predições: (a) sinais de efeito ‘gargalo de garrafa’ (*bottleneck*) na população original que teria ocupado diferentes ‘refúgios’, associados às contrações e isolamento causados pelas flutuações climáticas (a população original pode ser detectada pela descendência dos subgrupos); (b) expansão populacional, associada ao fim ou ao aumento geográfico dos ‘refúgios’, nos períodos interglaciais; (c) variabilidade genética baixa dentro dos refúgios; e (d) baixa estruturação filogeográfica (Conn e Mirabello, 2007).

A existência de períodos glaciais (e interglaciais) é indiscutível na atualidade, mas o mesmo não pode ser dito quanto aos efeitos destes períodos na contração (e expansão) de florestas tropicais úmidas. Apesar de já terem sido publicados mais de 500 artigos discutindo aspectos geológicos e paleocológicos da região amazônica durante o período do Quaternário (Hooghiemstra e Van der Hammen, 1998), a presença de dados palinológicos contraditórios, associados à ausência de outros tipos de evidência, como dados paleoclimáticos ou de reconstrução da paleovegetação (Mayle et al., 2004; Turcq et al., 2002), exemplificam a dificuldade de se entender a biogeografia da região durante o Pleistoceno. O mais provável é que as florestas tropicais úmidas, como a Floresta Amazônica, tenham sido reduzidas pela metade e substituídas em parte por vegetações de montanha e por matas semi-decíduais nas terras baixas (Anhuf et al., 2006; Bush et al., 2004). A vegetação da região periférica possivelmente sofreu mudanças mais abruptas, formando ‘cinturões áridos’, cobertos por vegetação de savana (Bush et al., 2004; Van der Hammen e Hooghiemstra, 2000).

#### *1.5.2.4 Outras hipóteses biogeográficas*

Dezenas de outras hipóteses para a geração de diversidade da biota amazônica, além das mencionadas acima, tiveram como base a biogeografia de vicariância. Duas hipóteses em particular tiveram como ponto de partida a mesma premissa de elevação do nível do mar utilizada na criação da hipótese de ‘incursões marinhas’ do Mioceno

(Räsänen et al., 1995) para explicar a geração de diversidade de espécies endêmicas da Amazônia durante o Plioceno/Pleistoceno (Marroig e Cerqueira, 1997; Nores, 1999).

A primeira foi chamada de “hipótese da lagoa”, desenvolvida por Marroig e Cerqueira (1997). Dois anos após a publicação da hipótese de ‘incursões marinhas’ do Mioceno, os autores defenderam a hipótese, antes proposta por Klammer (1984), de formação de uma lagoa na região atual do Médio e Baixo Amazonas durante o Plioceno/Pleistoceno (2,5 a 0,75 Maa). Com base na análise de sedimentos fósseis de terrenos alagados no Médio-Amazonas, postularam que no tempo Piacenziano (2,5 Maa), teria ocorrido uma elevação no nível do mar de 180 m (a ‘grande incursão Fouratiana’), possivelmente provocada pelos ciclos de Milankovitch. Os Alpes Andinos teriam entrado em processo de degelo e esta água teria fluído pelas regiões baixas amazônicas até a região Central, formando uma lagoa (uma vez que essa é a região de menor altitude na Amazônia). De maneira alternativa, essa lagoa poderia representar ainda um alagamento provocado pelas incursões marinhas durante a incursão Fouratiana.

A segunda foi proposta por Nores (1999), com base em estudos empíricos com aves, que as ‘incursões marinhas’ teriam prosseguido por todo o Período Terciário na região amazônica e promovido ciclos de especiação alopátrica geradores da biodiversidade. Partindo desta premissa, Nores redesenhou como porções de terra todas as áreas geográficas acima de 100 metros de altitude. Desta forma, foi possível inferir que teriam existido duas grandes ilhas ao norte do Rio Amazonas e pequenas ilhas e arquipélagos no Escudo das Guianas e ainda na periferia oeste da bacia amazônica.

Ainda que em muitas análises filogeográficas uma única e exclusiva hipótese biogeográfica seja atribuída aos eventos de especiação de alguns organismos (e.g. Dick et al., 2004; Hayes e Sewlal, 2004; Opazo et al., 2006), tudo indica que a diversificação neotropical é muito complexa de ser explicada e que certamente uma única hipótese não será suficiente para explicar toda a diversidade da biota amazônica atual (Aleixo, 2004; Geurgas et al., 2008; Hoorn et al., 2010; Patel et al., 2011). Inferências cladísticas das regiões biogeográficas ocupadas por espécies endêmicas, assim como análises baseadas em polimorfismos de sequências de DNA podem ser

úteis na datação aproximada dos períodos de maior diversificação da biota e na identificação das hipóteses biogeográficas que melhor explicam cada período da história amazônica (Hoorn et al., 2010).

### 1.5.3. Análises biogeográficas atuais

As ideias presentes nos modelos dispersionista e vicariante foram agrupadas nas últimas três décadas em modelos coalescentes – a biogeografia cladística e a filogeografia. No primeiro modelo, a procura por padrões biogeográficos congruentes utilizando a cladística é enfatizada, sem que se imponha qualquer predição dispersionista ou vicariante *a priori* (Morrone e Crisci, 1995). A filogeografia se destina ao estudo dos princípios e processos, que podem ser indiretamente estimados pela distribuição geográfica das linhagens genealógicas.

Nas primeiras análises filogeográficas eram utilizados os relacionamentos evolutivos de táxons distintos, inferidos a partir de análises de marcadores mitocondriais (e.g. Avise et al., 1988; Ball et al., 1988). Estes padrões filogenéticos encontrados eram posteriormente confrontados com hipóteses biogeográficas de geração de biodiversidade para explicar os processos envolvidos na formação dos padrões. Os tempos de divergência entre os táxons eram calculados a partir do emprego do relógio molecular, com o objetivo de situar as cladogêneses no tempo geológico. Hoje em dia as análises para estimar o tempo geológico das cladogêneses foram aprimoradas e é possível adicionar dados fósseis, morfológicos, ou moleculares associados a modelos evolutivos. Estes dados temporais são muito úteis ainda na formulação de hipóteses coalescentes para inferir a dinâmica de populações no passado.

#### 1.5.3.1. Biogeografia cladística

A biogeografia cladística foi desenvolvida por Platnick e Nelson (1988) e se propõe a analisar “filogeneticamente” as áreas de ocorrência de espécies. De maneira simplificada, pode-se dizer que neste modelo o relacionamento filogenético entre os táxons é analisado para que seja possível estimar um cladograma de área. A partir da topologia resultante do estudo filogenético, o cladograma é construído pela simples substituição dos nomes dos táxons pelo nome das áreas de ocorrência.

O uso do modelo cladístico pode ser simples e eficiente, quando se analisa apenas táxons que sejam endêmicos de uma única área geográfica. Quando há um táxon presente em mais de uma área, torna-se necessário aplicar uma ou mais premissas não mutuamente exclusivas ao modelo: (1) táxons com larga distribuição geográfica se tornam uma sinapomorfia das áreas por eles habitadas e, portanto, estas regiões formarão um clado monofilético de “áreas irmãs”; (2) um táxon pode estar amplamente distribuído em áreas cujo relacionamento filogenético é mono ou parafilético; e/ou (3) cada ocorrência do táxon é tratada separadamente e estas áreas geográficas têm iguais chances de estar presente em qualquer parte do cladograma (revisadas em Morrone e Crisci, 1995).

Uma derivação deste modelo é o método de análise de parcimônia de endemicidade (PAE), criado por Rosen (1988). Este método foi criado em um contexto paleontológico, mas hoje é aplicado nas análises de localidades de ocorrência de táxons atuais (Morrone e Crisci, 1995). Nesta análise, as áreas são tratadas como táxons e os táxons, como caracteres binários (presença e ausência). O relacionamento das áreas geográficas na PAE é estimado com base no compartilhamento de táxons (quanto mais táxons compartilhados, maior a proximidade cladogenética das localidades). As regiões que apresentam um maior número de táxons podem ser consideradas os centros de origem das espécies (Harold e Mooi, 1994). O objetivo final desta análise é a interpretação da história de ocupação espacial pelos táxons ao longo do tempo.

Uma questão de extrema relevância se define quanto ao método utilizado para reconstrução dos cladogramas, seja no método cladístico, na PAE ou utilizando sequências de DNA. O método de máxima-parcimônia minimiza o número de passos mutacionais em uma reconstrução filogenética e, conseqüentemente, pode gerar erros quando as probabilidades de passos mutacionais entre os táxons são desiguais (*i.e.* não diferencia organismos que evoluam em uma taxa mais acelerada do que outros) (Cunningham et al., 1998). Isto significa que organismos que passaram por radiações adaptativas são frequentemente agrupados em um mesmo clado, mesmo não sendo filogeneticamente relacionados. Além disso, fontes importantes de incerteza associadas aos modelos empregados na máxima-parcimônia não são levadas em consideração, como por exemplo as incertezas referentes aos processos de dispersão de diferentes organismos (Ronquist, 2004).

### 1.5.3.2. Relógio molecular e a filogeografia bayesiana

Métodos probabilísticos (máxima-verossimilhança e inferência bayesiana) são os mais indicados para interpretar a história filogenética de organismos, por utilizar modelos explícitos de evolução dos caracteres (Drummond et al., 2002; Felsenstein, 1981; Pagel, 1999). Estes métodos são baseados em uma função que calcula a probabilidade que uma árvore pode produzir em dados observados.

Ao contrário da reconstrução filogenética por máxima-parcimônia, incorpora-se no método de máxima-verossimilhança as diferenças nos comprimentos dos ramos (estimadas na árvore filogenética) para o cálculo de probabilidades condicionais de caracteres ancestrais (Cunningham et al., 1998). Neste método, uma única topologia de árvore é assumida como verdadeira – aquela que maximiza a probabilidade de observar os dados, considerando um determinado modelo (Delsuc et al., 2005).

As probabilidades condicionais são generalizadas posteriormente nos métodos de reconstrução bayesianos de forma a obter os caracteres ancestrais sem a necessidade de fixar uma topologia com o tamanho dos ramos conhecido para especificar parâmetros exatos nos modelos de Markov (Lemey et al., 2009).

A estatística apresentada no modelo de máxima-verossimilhança se limita a calcular a probabilidade de se observar corretamente os dados considerando apenas uma única topologia de árvore como sendo a verdadeira. O grande problema reside na falta de conhecimento acerca da topologia correta na maioria das análises filogeográficas que utilizam dados reais. Já no modelo bayesiano, calcula-se a probabilidade da topologia encontrada ser correta, dadas as informações contidas nos elementos analisados e nas *prioris* (caracteres morfológicos ou sequências de DNA/RNA).

A topologia final da árvore obtida pela análise bayesiana é considerada como uma hipótese, com uma probabilidade associada à topologia encontrada. Ao contrário da máxima-verossimilhança, não se estabelece uma única topologia como sendo a correta e o acúmulo (ou perda) de informações acerca dos relacionamentos ancestrais depende da informação contida no conjunto de dados analisado (e.g. sequências de DNA). Esta vantagem infelizmente vem associada a um revés – a necessidade de

especificar distribuições de probabilidades *a priori* para todos os parâmetros estimados no modelo evolutivo escolhido, sendo que muitas destas *prioris* podem ser pouco intuitivas. Nestes casos, são utilizadas *flat priors*, que apresenta a mesma probabilidade para todos os eventos estipulados em um intervalo de confiança.

A utilização das *prioris* somadas ao conjunto de dados, sob outro ponto de vista, pode ser também vantajosa. Partindo-se do pressuposto de que o relacionamento filogenético entre os táxons não é diretamente observável na natureza e que os padrões migratórios dos táxons raramente são bem compreendidos, a abordagem estatística bayesiana, a partir das cadeias de Markov Monte-Carlo (MCMC) (Green, 1995; Hastings, 1970; Metropolis et al., 1953), integra padrões mais realistas às análises de relacionamento ao estimar certo grau de incerteza aos resultados (Lemey et al., 2009). Esta incerteza é traduzida em intervalos de confiança para cada parâmetro calculado. Desta forma é possível criar testes de hipóteses a partir do estabelecimento de valores *a priori* arbitrários para os parâmetros calculados na inferência bayesiana. Esta informação *a priori* pode ser a incorporação de dados fósseis, da datação de eventos vicariantes hipotéticos, ou ainda resultados obtidos em outros estudos independentes.

A acomodação estatística para a falta de conhecimento prévio (traduzida na inserção de intervalos de confiança) funciona como uma proteção contra vieses potenciais na análise, mas ao custo da redução no poder de inferir definitivamente as proposições (uma vez que não existem valores absolutos dos parâmetros inferidos). Esta afirmação não é necessariamente negativa se considerarmos que a ciência é constituída de verdades transitórias e de constante acúmulo de conhecimento. Adicionalmente, cabe ressaltar que qualquer inferência filogenética que inclua mais de 15 táxons é heurística e, portanto, não haverá a escolha da árvore mais provável (ou mais verossímil), uma vez que será impossível visitar todas as árvores possíveis.

Análises filogenéticas aplicadas à biogeografia (filogeografia) são úteis para interpretar historicamente as relações evolutivas dos táxons e confrontar estes resultados com as hipóteses biogeográficas vigentes. O emprego do fator “tempo” na filogenia procede da revolucionária proposta feita por Zuckerkandl e Pauling (1962), de que a evolução molecular ocorreria em uma taxa uniforme ao longo do tempo. Nesta proposta, formalmente declarada por Margoliash (1963) e chamada de

“hipótese do relógio molecular”, assume-se a mesma taxa evolutiva ao longo do tempo para todas as espécies, mesmo quando analisadas espécies filogeneticamente distantes.

O desenvolvimento das técnicas de sequenciamento de DNA e a formulação da teoria neutra de evolução molecular (Kimura, 1968) foram as peças-chave para calcular as taxas de mutação neutra dos organismos a partir das análises de mais marcadores moleculares (Kumar, 2005). Com o objetivo de estimar o tempo de divergências entre táxons, mesmo na ausência de indicadores fósseis, muitos pesquisadores compilaram dados de divergências entre espécies de uma mesma família ou classe e associaram a separação destas linhagens a eventos geológicos conhecidos, para a calibração de relógios moleculares fixos “universais” (Brower, 1994).

A hipótese do relógio molecular foi questionada pela comunidade científica e o debate acerca do tema se estendeu até a década de 90 (revisado em Kumar, 2005). O padrão geral observado em análises de relógio molecular é o de que espécies proximamente relacionadas apresentam taxas de substituição (de nucleotídeos ou aminoácidos) similares e seguem o modelo neutro de evolução, mas o mesmo não pode ser dito, na maioria das vezes, quando são comparados organismos filogeneticamente mais distantes (Li e Drummond, 2012). As taxas evolutivas podem variar consideravelmente entre espécies, pois dependem de uma combinação de fatores como tempo de geração, tamanho populacional, taxa metabólica e eficácia na reparação do DNA. Adicionalmente mutações benéficas ou deletérias podem aumentar ou diminuir, respectivamente, suas frequências na população de maneira mais acelerada que as mutações neutras (Pybus, 2006). Algumas linhagens podem ainda apresentar um aumento da taxa de substituição como consequência de eventos de radiação adaptativa (Benton, 1999). A prática da premissa de um relógio molecular fixo aplicada às diferentes linhagens em uma determinada topologia, portanto, pode gerar não apenas estimativas equivocadas dos tempos de separação das linhagens, como ainda topologias de árvores equivocadas (Drummond et al., 2006; Yoder e Young, 2000).

Nos últimos anos, diversos métodos têm sido desenvolvidos para minimizar (ou “relaxar”) a premissa do relógio molecular fixo (Ho et al., 2005). A hipótese do

relógio molecular relaxado parte do princípio de que as taxas evolutivas de diferentes organismos podem variar em uma topologia de árvore. Esta variação pode ser definida *a priori* a partir de diferentes calibrações dos nós das árvores filogenéticas (pelo emprego de modelos paramétricos para as estimativas dos tempos de divergência chamados de “calibrações suaves”), ou da estimativa de um intervalo para a taxa de mutação das linhagens, por milhão de anos (ao invés de um valor fixo aplicado no relógio estrito). Na proposta bayesiana mais recente de relógio molecular “relaxado” (Drummond et al., 2006), as taxas evolutivas de cada ramo são estimadas a partir de uma distribuição paramétrica, como normal logarítmica, exponencial, ou gama. Os parâmetros calculados (i.e. distribuição posterior) são estimados a partir do conjunto “dados + definições *a priori*”.

#### 1.5.3.3. O método bayesiano aplicado à teoria da coalescência

A teoria da coalescência foi formalizada por Kingman (1982) e nela é estabelecida a relação entre tamanho populacional e ancestralidade. Os indivíduos atuais que são filogeneticamente relacionados “coalescem”, num dado período de tempo, a um ancestral comum mais recente (MRCA). A partir da premissa de coalescência, é possível rastrear todos os possíveis alelos de um *locus*, compartilhados por todos os membros de uma população, para retrospectivamente encontrar o ancestral comum mais recente (Avise, 2000). A estimativa do tempo até o ancestral comum mais recente (TMRCA) pode ser feita com base na aplicação do relógio molecular (fixo ou relaxado) às sequências analisadas. Esta estimativa de tempo está relacionada ao tamanho populacional e ao tipo de herança do marcador genético utilizado. Quanto maior a população efetiva, mais remoto é o TMRCA (cf. Kingman 1982 para maiores detalhes). Caso o marcador analisado possua herança biparental (e.g. DNA nuclear), o TMRCA será quatro vezes mais antigo do que o encontrado em análises de marcadores de herança uniparental (e.g. DNA mitocondrial ou cromossomo Y em humanos), pois estes podem ser passados por ambos parentes para as próximas gerações de ambos os sexos (macho parental → fêmea da prole, macho parental → macho da prole, fêmea parental → macho da prole, e fêmea parental → fêmea da prole), diferentemente dos uniparentais, que permanecem apenas na linhagem de um dos sexos [mitocondrial – linhagem materna; cromossomo Y (e.g. seres humanos) – linhagem paterna].

O método bayesiano aplicado à teoria da coalescência foi um importante passo para o progresso das análises filogeográficas de espécies proximamente relacionadas ou populações de uma mesma espécie. A partir desta composição, foi possível estimar de maneira mais realista tamanhos populacionais atuais e antigos (Hudson, 1990) e, conseqüentemente, interpretar (ou hipotetizar) os eventos geológicos, ecológicos e populacionais com maior precisão (Heled e Drummond, 2008).

A informação que pode ser extraída dos dados sobre o tamanho populacional depende da acurácia com a qual a genealogia é inferida e do grau de extensão dos processos demográficos que o modelo coalescente pode acomodar (Atkinson et al., 2008). A acurácia da genealogia depende do método de substituição nucleotídica utilizado (i.e. se reflete a complexidade dos dados ou a simplifica) e da taxa de substituição do *locus* gênico analisado por geração (marcadores mais conservados refletirão poucos eventos de coalescência, se comparados aos marcadores de evolução rápida). Se, por exemplo, um determinado marcador apresenta uma taxa de substituição de um nucleotídeo a cada 100 mil anos, eventos inferiores a este tempo não são observáveis. Já o modelo coalescente depende intrinsecamente de um modelo matemático acurado.

Qualquer método coalescente que confira a capacidade de inferir a demografia histórica de um organismo requer um modelo demográfico. A utilização de modelos coalescentes de variação no tamanho de população ao longo da história (Donnelly e Tavaré, 1995; Griffiths e Tavaré, 1994) é imprescindível para se recriar dinâmicas populacionais mais realistas. Estes modelos matemáticos devem ser flexíveis o suficiente para se encaixar em muitos cenários demográficos, como, por exemplo, os de crescimento populacional constante, exponencial, logarítmico, entre outros. Os gráficos de horizonte (do inglês *Skyline Plot*) foram muito úteis como uma ferramenta de seleção do melhor modelo demográfico para um determinado conjunto de dados (Pybus e Rambaut, 2002; Pybus et al., 2000).

A grande limitação do modelo supracitado era a de que a inferência da história demográfica tinha como base a genealogia em si e não o conjunto de dados de sequências de DNA e, portanto, ignoravam o erro associado das estimativas filogenéticas. Cinco anos depois surgiu o modelo bayesiano de gráficos de horizonte (do inglês *Bayesian Skyline Plot*, ou BSP), com o objetivo de suprir esta fragilidade

(Drummond et al., 2005). Neste modelo foram adicionados intervalos de confiança para as estimativas de tamanho populacional efetivo para cada ponto no tempo, retrocedendo até o ancestral comum mais recente (calculados por amostragens das MCMC), a fim de representar o grau de incerteza do tempo coalescente (uma vez que a fórmula para o cálculo do TMRCA contabiliza o tamanho populacional efetivo) (Drummond et al., 2005). Recentemente este modelo foi aperfeiçoado e renomeado para gráfico bayesiano de vista aérea (adaptado do inglês *Bayesian Skyride*), com o qual é possível estimar ainda o número de possíveis eventos de contração ou expansão em uma população analisada (Heled e Drummond, 2008).

Os avanços nas análises populacionais (utilizando o método bayesiano) se estenderam ao ponto de possibilitar melhores inferências acerca de padrões complexos de migração populacional, levando em consideração direção de migração e flutuações de tamanho populacional (Beerli e Palczewski, 2010). Utilizando o modelo n-coalescente para múltiplas populações com migração diferencial (Hudson, 1990), é possível testar hipóteses de fluxo gênico histórico entre as populações, i.e., se a dinâmica populacional seguiu um padrão panmítico, de migração direcional a partir de um centro de origem, de fluxo gênico entre regiões adjacentes (*stepping-stones*), de completo isolamento ou ainda padrões mais complexos (Beerli e Palczewski, 2010). Como se trata de um modelo coalescente, sua aplicação pode ser útil no teste de hipóteses biogeográficas de origem e dispersão de linhagens ancestrais de uma espécie.

#### 1.5.4. Filogeografia de *R. prolixus* e *R. robustus* s.l. na região pan-amazônica

As espécies do gênero *Rhodnius* estão distribuídas em 27 províncias biogeográficas, dentre as quais se acredita haver pelo menos três centros de endemismo: a bacia do Orinoco, o sudeste amazônico e o Cerrado (Abad-Franch et al., 2009). Estes dois últimos centros parecem corresponder mais a antigas zonas de contato entre espécies que se tornaram simpátricas do que a centros de origem das linhagens (Abad-Franch et al., 2009).

As espécies de *Rhodnius* com maior abrangência na América Latina são *R. pictipes* e *R. robustus s.l.*, estando presentes em 36 e 27 ecorregiões, respectivamente (cf. Abad-Franch et al. 2009). *R. robustus* II é a linhagem do complexo *R. robustus* de espécies crípticas com maior distribuição conhecida – está presente em 13 ecorregiões.

Monteiro e colaboradores (2003) levantaram uma hipótese biogeográfica acerca deste padrão filogenético ao observarem a existência de quatro linhagens crípticas de *R. robustus s.l.* Os autores estimaram os tempos de divergência entre as espécies com base em um relógio molecular ‘universal’ fixo (2,3% de divergência entre sequências, para-a-par, por milhão de anos) estimado para insetos (Brower, 1994). O evento vicariante que teria separado a população ancestral da região do Orinoco (que posteriormente originou *R. prolixus* e *R. robustus* I) da população que originou as espécies amazônicas de *R. robustus s.l.* foi estimado em três milhões de anos atrás (datando do Plioceno). Já a diversificação de *R. prolixus* e *R. robustus* I e ainda das linhagens amazônicas teria ocorrido durante o Pleistoceno, há menos de 2 Maa. Como o padrão filogenético entre as espécies de *R. robustus s.l.* era descontínuo, com segregação geográfica dos haplótipos, tendo ocorrido durante o Plioceno Médio/Superior e Pleistoceno, Monteiro e colaboradores (2003) propuseram que o isolamento é compatível com a hipótese dos ‘refúgios’.

A descoberta inesperada de Monteiro e colaboradores (2003) de um complexo de quatro linhagens de *R. robustus* foi uma inestimável contribuição para a identificação de possíveis áreas biogeográficas de diversificação da tribo Rhodniini na região amazônica. Contudo, a hipótese de especiação vicariante de *R. prolixus* e *R. robustus s.l.* em refúgios pleistocênicos foi levantada pelos autores com base apenas nos tempos de separação entre as espécies calculados a partir da aplicação de um relógio molecular fixo. O problema que reside nesta análise é que as taxas evolutivas entre os táxons podem variar consideravelmente, em função de uma série de questões (previamente levantadas na seção 1.6.3.2) e, portanto, o relógio fixo não poderia ser aplicado. Adicionalmente, a estimativa de tempo dos eventos cladogenéticos obtida a partir da aplicação desse relógio fixo deve conter um erro associado grande. A calibração utilizada (Brower, 1994) foi calculada com base na divergência genética entre sete organismos pertencentes a quatro ordens da classe Insecta, que divergiram durante eventos geológicos distintos em cinco regiões diferentes do mundo. Neste

cálculo não foi levada em consideração a taxa de substituição do gene mtCytb, utilizado por Monteiro e cols. (2003) na análise filogeográfica de triatomíneos, e tampouco um táxon pertencente à ordem Hemiptera, ou um evento geológico amazônico vicariante foi utilizado na estimativa de divergência feita por Brower (1994).

No intuito de diminuir o erro associado à calibração ‘universal’, Pavan (2009) aplicou aos dados obtidos por Monteiro e cols. (2003) uma calibração semi-estrita do relógio molecular. Esta calibração semi-estrita gera um possível intervalo de tempo para os eventos cladogenéticos a partir da aplicação de duas estimativas fixas de divergência entre sequências por milhão de anos. Portanto, esses valores não variam nos diferentes nós da árvore e apenas associam uma incerteza ao resultado.

Com base em um evento geológico conhecido e bem datado (separação da península de Baja Califórnia do restante do México), Pfeiler e cols. (2006) calibraram um relógio molecular (1,1 a 1,8% por milhão de anos) utilizando a divergência entre sequências do mtCytb de duas subespécies de triatomíneos, *Triatoma rubida conchimiensis* e *T. rubida sonoriana*. Pavan (2009), a partir das novas datações estimadas, pôde concluir que os eventos de diversificação do complexo teriam ocorrido durante o Mioceno/Plioceno e que os padrões de divergência e distribuição das linhagens de *R. robustus s.l.* estariam mais de acordo com a hipótese biogeográfica de ‘incursões marinhas’.

A calibração utilizada por Pavan (2009) deve naturalmente representar melhor a taxa de divergência entre os táxons ao longo do tempo do que a utilizada por Monteiro e cols. (2003). Apesar do avanço no conhecimento biogeográfico das espécies obtidos até o momento, novos resultados acerca da diversidade de *R. robustus s.l.* na região amazônica devem ser incluídos e algumas questões relevantes merecem ser revisitadas para uma melhor compreensão dos processos bióticos e abióticos responsáveis pelos eventos de especiação observados. Seguem abaixo os principais questões que motivaram a reavaliação deste tema:

(1) a hipótese de que os táxons analisados por Monteiro e cols. (2003) evoluem em uma mesma taxa não foi testada em nenhuma das análises. Ao assumir uma taxa evolutiva constante entre as espécies (segundo a hipótese do ‘relógio fixo’), a topologia da árvore gerada pode não ser a melhor;

(2) não foram realizados quaisquer testes estatísticos de hipóteses que corroborem (ou não) a diversificação da linhagem em decorrência de ‘incursões marinhas’ ou ‘refúgios pleistocênicos’;

(3) não foram consideradas nas análises de Monteiro e cols. (2003) e Pavan (2009) duas linhagens descobertas recentemente, que são filogeneticamente relacionadas ao complexo (Abad-Franch e Monteiro, 2005; 2007; Márquez et al., 2011) e que podem auxiliar no entendimento da diversificação da linhagem;

(4) apesar de ter representado a maior amostragem de *R. prolixus* e *R. robustus* já publicada até aquele momento, Monteiro e cols. (2003) obtiveram poucos espécimes de cada uma das linhagens crípticas e a baixa amostragem de cada linhagem de *R. robustus s.l.* pode ter ocultado detalhes importantes acerca da divergência genética entre elas (*R. robustus* I – 7 espécimes; *R. robustus* II – 11 espécimes; *R. robustus* III – 10 espécimes; *R. robustus* IV – 15 espécimes);

(5) os limites da distribuição geográfica das linhagens não haviam sido visitados até o momento e estas regiões podem conter informações mais ricas sobre o processo de diversificação das espécies.

#### 1.6. Ritmos circadianos

Determinados aspectos do comportamento de insetos têm um papel fundamental na dinâmica de transmissão das doenças e do controle vetorial (Klowden e Zwiebel, 2004; Lazzari e Lorenzo, 2009; Pates e Curtis, 2005). O controle dos ritmos de atividade (Clements, 1999; Saunders, 2002) é um tema particularmente relevante para o entendimento de algumas questões epidemiológicas, pois estes ritmos afetam o tempo e o grau de contato dos vetores não só com seus hospedeiros (parâmetros importantes na transmissão de patógenos), como também com determinados tipos de inseticidas (Lazzari e Lorenzo, 2009). Estudos acerca do comportamento de espécies filogeneticamente próximas podem ainda contribuir para a identificação de processos evolutivos relacionados à especiação, tais como a formação de barreiras naturais de isolamento reprodutivo.

As oscilações diárias na fisiologia, no metabolismo e no comportamento dos insetos em resposta às mudanças periódicas no ambiente (causadas pelo período de

rotação da Terra) são controladas por um relógio endógeno (Abruzzi et al., 2011; Ceriani et al., 2002; Keegan et al., 2007; McDonald e Rosbash, 2001). Mesmo na ausência de variações externas, este relógio é capaz de manter esses ritmos por um período próximo a 24 horas, sendo por isso chamados de ritmos circadianos (i.e. “cerca de um dia”) (Marques e Menna-Barreto, 2003; Saunders, 2002).

O relógio circadiano apresenta um período próximo a 24 horas, quando as condições ambientais são constantes. A sincronização dos ritmos endógenos com o período diário é dependente de um ou mais indicadores ambientais. Estes indicadores são chamados coletivamente pelo termo alemão *zeitgebers*. Os *zeitgebers* mais estudados são as variações diárias de luminosidade e de temperatura (revisado em Boothroyd e Young, 2008; Glaser e Stanewsky, 2007).

As décadas de 50 e 60 foram marcadas pelo acúmulo de conhecimento sobre os *zeitgebers* luz e temperatura. Neste período anterior à genética molecular, os ritmos diários eram avaliados indiretamente pela observação da atividade locomotora de insetos (principalmente *Drosophila*) e mamíferos (Aschoff, 1960; Pittendrigh, 1958). Estes organismos eram submetidos a diferentes condições exógenas: (1) semelhantes às naturais (uma fase clara durante o dia e outra escura durante a noite); (2) sem a presença de indicadores ambientais (escuridão completa ou luz durante o dia todo); (3) com variação de temperatura ao longo do dia; ou (4) com aplicação de sinais súbitos de mudança de fase em um ambiente sem a presença de indicadores ambientais (perturbação das fases escura e clara subjetivas pela aplicação de pulsos luz e escuridão, respectivamente, por um curto período de tempo).

A influência das condições exógenas percebida no ritmo endógeno dos organismos estudados foi considerada como sendo variável. O período do relógio endógeno em condições semelhantes às naturais era sempre de 24 horas. Mesmo na ausência de indicadores ambientais, um oscilador interno seria capaz de manter o ritmo endógeno com um período de 20 a 28 horas. Quando havia perturbação no início da fase escura subjetiva dos espécimes de *Drosophila* com aplicação de pulsos de luz, o período do relógio se atrasava, i.e. tornava-se maior do que o normal, e o contrário ocorria quando a perturbação era feita no fim da fase escura subjetiva. Contudo, o mesmo resultado não era observado quando essa perturbação era feita com variações de temperatura (Pittendrigh, 1958).

Um modelo de funcionamento do relógio circadiano pôde ser proposto por Pittendrigh e cols. (1958) a partir das análises locomotoras de organismos. Os autores postularam que pelo menos dois osciladores internos seriam responsáveis pela sincronização do relógio com o período do dia, um oscilador principal, sensível à luz, e um oscilador secundário, ligado ao oscilador principal, sensível à temperatura. Na presença de outros osciladores, estes também seriam ligados ao oscilador principal. Na ausência de luz, os osciladores secundários funcionariam de acordo a manter o ritmo diário do organismo.

### 1.6.1. A genética dos ritmos circadianos

Nas décadas seguintes, a genética dos ritmos circadianos foi (e tem sido) extensivamente estudada na espécie-modelo *Drosophila melanogaster* (Hall, 2003). A partir de estudos de RT-PCR, Western Blot e cultura de células (Cyran et al., 2003; Hardin, 2005), foi possível observar que o marca-passo molecular do relógio deste inseto inclui alças regulatórias interligadas, que são responsáveis pelo controle dos ritmos relacionados ao comportamento, fisiologia e a outros processos biológicos (revisado em Allada e Chung, 2010; e Hardin, 2011). Estas alças são mantidas por retroalimentação transcricional, que consiste de proteínas que se ligam à sequências específicas de DNA e estimulam a transcrição de seus próprios repressores. Posteriormente, com o aprimoramento das técnicas de RNA de interferência, análise de expressão de genes e trangênese, foi possível desvendar maiores detalhes acerca destas alças regulatórias (revisado em Allada e Chung, 2010; e Hardin, 2011).

De maneira simplificada, pode-se dizer que nestas duas alças regulatórias os fatores de transcrição codificados pelos genes *Clock* (*Clk*) e *cycle* (*cyc*) formam um heterodímero que promove a ativação da transcrição de *period* (*per*), *timeless* (*tim*), *vri* (*vri*) e *PAR-domain protein 1* (*Pdp1*) (Cyran et al., 2003; Darlington et al., 1998; Glossop et al., 2003). O heterodímero PER/TIM inibe a função de CLK/CYC<sup>2</sup>, e as proteínas ativadora PDP1ε (resultante de um dos transcritos alternativos de *Pdp1*) e repressora VRI competem pelo mesmo sítio regulatório no promotor de *Clk* para controlar sua expressão.

---

<sup>2</sup> Por convenção, todos os genes do relógio circadiano são escritos em itálico, com letras minúsculas ou com apenas a primeira letra maiúscula (dependendo de como foi citada pela primeira vez na literatura). Todas as proteínas produzidas por estes genes são escritas com letras maiúsculas.

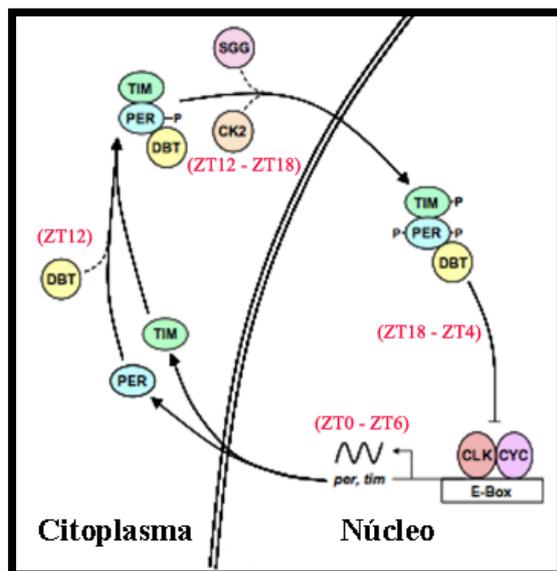
Na ausência de fatores ambientais, o período do ciclo nas alças regulatórias se aproxima das 24 horas em função do atraso nos processos de ativação transcricional e/ou repressão de heterodímeros por eventos de degradação dependentes de fosforilação de PER, TIM e CLK (Allada e Chung, 2010). O processo de fosforilação precede o desaparecimento da proteína e, portanto, determina a meia-vida das três proteínas citadas.

A alça regulatória principal, ou alça regulatória *per* (Figura 1.3), é considerada como sendo a mais importante, por controlar a ativação e repressão do heterodímero CLK-CYC. Durante o início e o meio do dia, entre ZT0 e ZT6 (do inglês “tempo *zeitgeber*”), o heterodímero CLK-CYC se liga nas regiões *E-box* (em complementaridade com a sequência CACGTG) dos promotores de *per*, *tim*, *vri* e *Pdp1*, promovendo a ativação da transcrição destes quatro genes (Cyran et al., 2003; Darlington et al., 1998; Glossop et al., 2003). No início da noite (ZT12), PER e TIM se acumulam no citoplasma e formam os complexos PER-TIM (juntamente com a proteína *doubletime*). Essa diferença de seis a doze horas entre a transcrição de *per* e *tim* no núcleo celular e o acúmulo de PER e TIM no citoplasma podem ser causados por dois motivos: (1) degradação dependente de fosforilação das proteínas – a proteína PER se associa no citoplasma à DOUBLETIME (DBT). Este complexo, na presença de TIM, é estável (TIM-PER-DBT), mas em sua ausência, PER-DBT é degradado. Este processo dificulta o acúmulo de PER e, conseqüentemente, a formação do complexo PER-TIM; ou (2) atraso na acumulação de PER no citoplasma em função de sua tradução ser regulada.

Entre o início e o meio da noite (ZT12-18), PER e TIM são fosforilados pela CASEÍNA-KINASE-2 (CK2) e pela SHAGGY (SGG), que conferem ao complexo a sinalização de endereçamento ao núcleo celular. Esse intervalo de tempo é grande, pois a sinalização nuclear e a fosforilação podem ser alteradas pela associação da fosfatase heterotrimérica serina/treonina-2A (PP2a) com PER e da proteína fosfatase-1 (PP1) com TIM.

Uma vez que o complexo TIM-PER-DBT é translocado para o núcleo [pode ser que PER-DBT entrem no núcleo dissociados de TIM (Yu et al., 2009)] entre o meio da noite e o início do dia (entre ZT18 e ZT4), este complexo se liga a CLK, promovendo sua fosforilação e inibindo a ligação de CLK-CYC na região *E-box* dos

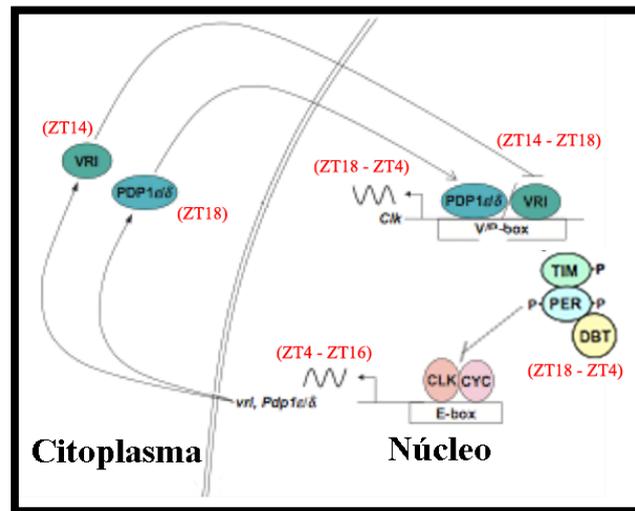
promotores de *per* e *tim*. Consequentemente, este heterodímero é reprimido de sua função de ativação transcricional. Ainda durante o início da manhã (entre ZT0 e ZT4), a luz promove a degradação de TIM (a ser comentada mais adiante). A liberação de TIM promove a fosforilação de PER por DBT, o que resulta na ligação posterior de uma ubiquitina ligase E3 chamada SLIMB à PER, que por sua vez o endereçará à degradação via proteassoma durante o meio do dia (ZT6).



**Figura 1.3.** Alça regulatória principal, ou alça regulatória *per*. Todos os nomes dos genes, elementos reguladores e das proteínas são definidos no texto. A linha dupla representa a membrana nuclear e a linha sinusoidal representa o RNAm. As setas sólidas representam a síntese, ligação e endereçamento do RNAm e das proteínas e as pontilhadas, atividades regulatórias. A linha bloqueada representa repressão das proteínas. P – sítio de fosforilação. Entre parêntesis em vermelho, o horário *zeitgeber* (ZT) correspondente a cada evento ilustrado. Modificada de Hardin (2011).

A alça regulatória secundária, ou alça regulatória *Clk* (Figura 1.4), é intimamente ligada à alça regulatória principal. O heterodímero CLK-CYC, além de ativar a transcrição de *per* e *tim* na alça regulatória principal, também se liga às regiões E-box dos promotores de *vri* e *Pdp1*  $\epsilon/\delta$  desde o início do dia até o início da noite (ZT4-ZT16) e promove também suas transcrições. A acumulação de VRI ocorre no início da noite (ZT14) e a de PDP1  $\epsilon/\delta$  somente quatro horas depois. VRI se liga à região promotora de *Clk*, chamada VRI/PDP1-boxes (V/P-boxes), para reprimir a transcrição de *Clk*. Entre ZT16 e ZT4, os níveis de CLK-CYC disponíveis estão baixos, por conta da repressão exercida por PER-DBT na alça regulatória principal. Este fato acarreta o declínio do RNAm de *vri* e, consequentemente, da proteína VRI.

PDP1  $\epsilon/\delta$  passa a atuar como um intensificador da transcrição de *Clk* entre o meio da noite e o início do dia (ZT18-ZT4). Uma vez que o complexo PER-DBT é degradado no meio do dia (ZT6), CLK-CYC deixa de ser reprimido na alça regulatória principal e o ciclo é reiniciado.



**Figura 1.4.** Alça regulatória secundária, ou alça regulatória *Clk*.

As setas e linhas seguem o mesmo padrão de representação da **Figura 1.3**. O complexo TIM-PER-DBT não pertence a esta alça regulatória (pertence àquela ilustrada na **Figura 1.3**) e está no esquema apenas para mostrar como é feita a inibição da ligação CLK-CYC. Modificado de Hardin (2011).

Existem indicativos de que a transcrição de *Clk* pode ocorrer sem a presença de um ativador (Hardin, 2011). No caso destes serem verdadeiros, a alça regulatória secundária não parece ser tão necessária para o funcionamento da principal. Contudo, a alça regulatória de *Clk* parece ter funções importantes na expressão de outros genes rítmicos, como o *takeout*, que regula o ritmo no comportamento de cópula em *Drosophila*.

A variação diária de luminosidade é captada em espécimes de *Drosophila* pelo sensor de luz azul CRYPTOCHROME (CRY) (Emery et al., 1998; Stanewsky et al., 1998), que na presença de luz interage com TIM (início da manhã). Esta interação induz a degradação de TIM via proteassoma, encerrando assim a repressão do heterodímero CLK/CYC. A captação de luz por CRY promove o arrastamento do ciclo para o período exato de 24 horas.

Em alguns insetos, como borboletas *Danaus plexippus* e *Antheraea pernyi*, e os mosquitos *Aedes aegypti* e *Culex quinquefasciatus*, existem duas formas do gene *cry* (Gentile et al., 2009; Zhu et al., 2005), uma (*cry1*) ortóloga do fotoreceptor de *Drosophila* e outra (*cry2*) que é ortóloga ao repressor transcricional encontrado em vertebrados (Yuan et al., 2007). Esta segunda forma é a única encontrada em alguns outros grupos de insetos, como abelhas *Apis mellifera* e besouros *Tribolium castaneum* (Rubin et al., 2006; Yuan et al., 2007). A história da origem e evolução de *cry1* e *cry2* (juntamente com os fotoreceptores de vertebrados) é intrincada e envolve, pelo menos, três eventos de duplicação e dois de perdas diferenciais de *cry1* e *cry2* (Yuan et al., 2007).

As questões levantadas por Pittendrigh e colaboradores (1958) nas análises locomotoras de *Drosophila*, em relação à perturbação das fases claras e escuras subjetivas com aplicação de escuridão e luz, respectivamente, foram respondidas mais de quatro décadas depois, a partir das análises genéticas dos ritmos circadianos. Quando se aplica luz no início da fase escura, há a degradação de TIM, que está se acumulando no citoplasma. Com isso, TIM demorará mais a se acumular e formar o complexo DBT-PER-TIM e conseqüentemente atrasará a fase do ciclo (i.e. o período do relógio neste ciclo será maior). Caso a luz seja aplicada no fim da noite, a degradação de TIM começará antes do previsto, o que causa o avanço de fase do ciclo (o período do relógio neste ciclo será menor) (Allada e Chung, 2010).

O funcionamento das alças regulatórias do relógio circadiano parece ser relativamente conservado entre insetos no que diz respeito às funções, arquitetura e à presença dos principais componentes (Allada e Chung, 2010). Contudo, a expressão do RNAm de alguns dos genes podem variar.

Estudo realizados com flebotomíneos da espécie *Lutzomyia longipalpis* (vetor da leishmaniose) mostraram que apesar de *per* e *tim* apresentarem picos de expressão semelhantes aqueles observados em *D. melanogaster*, *cyc* e *Clk* apresentaram padrões completamente distintos, sendo que o primeiro gene é expresso de forma rítmica (ao passo que em *D. melanogaster* é constitutiva) e o segundo com maior expressão no fim do dia (em *D. melanogaster*, no fim da noite/início da manhã) (Meireles-Filho et al., 2006a; Meireles-Filho et al., 2006b).

Um aspecto interessante em relação as diferenças de expressão dos principais genes do relógio circadiano foi observado nos últimos anos na comparação entre os mosquitos diurnos *Aedes aegypti* (vetor da dengue e febre amarela) e os mosquitos noturnos *Culex quinquefasciatus* (vetor da filariose e do vírus do oeste do Nilo). Apesar de *Ae. aegypti* e *Cx. quinquefasciatus* apresentarem um perfil de expressão da maioria dos genes semelhante, a expressão de *cry2* apresentou um padrão unimodal em *Cx. quinquefasciatus*, com maior expressão do gene na escotofase, e em *Ae. aegypti* uma expressão bimodal, com picos de expressão na escotofase e fotofase (Gentile et al., 2006; 2009). Esta diferença sugere que a regulação deste gene possa estar direta ou indiretamente relacionada com o padrão de atividade locomotora (diurno ou noturno) observado (Gentile et al., 2009).

Os efeitos da sincronização do relógio pela temperatura, comparados com os da sincronização pela luz, são pouco conhecidos em *Drosophila*, e muito menos em insetos hematófagos. Algumas importantes diferenças já foram observadas entre os dois processos, como o fato de cérebros dissecados de *Drosophila* serem sincronizados com ciclos de claro-escuro, mas não serem capazes de sincronizar seus ritmos de expressão gênica com ciclos de temperatura (Glaser e Stanewsky, 2007; Sehadova et al., 2009).

A interação dos *zeitgebers* luz e temperatura na sincronização e modulação dos ritmos de atividade em *Drosophila* é muito relevante e recentemente foi bem ilustrada em análises locomotoras realizadas em condições naturais (Vanin et al., 2012). Estas análises recentes ressaltam a importância de ainda se estudar a atividade locomotora dos organismos, mesmo depois de quase um século desde o seu início, para que seja possível posteriormente avaliar geneticamente o funcionamento do relógio circadiano. Essa afirmação é ainda mais relevante quando se trata de organismos cuja genética dos ritmos circadianos ainda não foi elucidada.

O relógio circadiano, além de controlar as oscilações diárias na atividade locomotora dos organismos, está envolvido na coordenação e no funcionamento apropriado de processos de desenvolvimento e reprodução em *Drosophila*, tais como aprendizado e memória, acasalamento, detoxificação, olfato, estado nutricional, eclosão de pupas, comportamento de corte e resposta imune a patógenos (Allada e Chung, 2010; McDonald e Rosbash, 2001). Nestes casos, os genes principais do

relógio circadiano também parecem ser atuantes. *per* e *tim* parecem estar envolvidos nos ritmos de aprendizado e memória (Lyons e Roman, 2009), eclosão de pupas (Myers, 2003) e acasalamento (Sakai e Ishida, 2001). *per* está envolvido ainda na resposta imune a patógenos (Allada e Chung, 2010) e em diferenças espécie-específicas no sons de corte (Kyriacou e Hall, 1986). O sensor de luz azul *cry* parece estar envolvido nos ritmos alimentar e antenal (captação de odores) (Krishnan et al., 2001; Sarov-Blat et al., 2000).

Os ritmos circadianos de acasalamento, sons de corte e olfato são imprescindíveis na interação e no reconhecimento intraespecífico. Os genes envolvidos nestes processos, portanto, podem ser muito úteis como marcadores moleculares em estudos que envolvam complexos de espécies incipientes de insetos vetores (Araki et al., 2009).

#### 1.6.2. Ritmos circadianos em triatomíneos

O ritmo circadiano em triatomíneos vem sendo estudado nas duas espécies epidemiologicamente mais relevantes, *T. infestans* e *R. prolixus*. Muitos processos biológicos destas espécies ligados à reprodução e à busca de alimento já foram relacionados ao controle por ritmos endógenos, como cruzamento e oviposição (Constantinou, 1979, 1984; Lazzari, 1991), dispersão (McEwen e Lehane, 1993) e orientação por gás carbônico e calor, no início da noite (Barrozo et al., 2004; Bodin et al., 2008).

Até o momento pouco se sabe acerca da genética do ritmo circadiano em triatomíneos. Somente através da avaliação das proteínas PER e TIM a partir de ensaios imunohistoquímicos foi observado que estas são expressas de maneira cíclica em *R. prolixus*, com picos de expressão na escotofase (Vafopoulou et al., 2010), assim como já observado em *Drosophila*, *L. longipalpis*, *Cx. quinquefasciatus* e *Ae. aegypti* (Allada et al., 1998; Gentile et al., 2009; Meireles-Filho et al., 2006b). Portanto, todos os processos biológicos supracitados que envolvem um ciclo circadiano foram estudados com base no ritmo de respostas comportamentais frente a diferentes *zeitgebers*.

Experimentos de atividade locomotora revelaram que os adultos de *R. prolixus* e *T. infestans* apresentam um padrão de atividade bimodal durante a noite (Constantinou, 1979; Lazzari, 1992). O pico de atividade no início da noite representa a saída destes insetos de seus abrigos em busca de alimento. Já o pico de atividade no fim da noite corresponde à busca destes adultos por um abrigo para passar o dia seguinte (Lorenzo e Lazzari, 1998). Durante o dia, os triatomíneos permanecem escondidos dentro de refúgios (e.g. frestas das paredes de habitações humanas, ninhos de árvores, tocas de mamíferos), em um estado de inatividade chamado acinésia (Lazzari e Lorenzo, 2009).

Aspectos do comportamento de triatomíneos, como os padrões de atividade e dispersão, comunicação intra e interespecífica e busca por hospedeiros poderão ser úteis para o desenvolvimento de novas estratégias de controle vetorial (Lazzari e Lorenzo, 2009). Uma das contribuições foi a recomendação de aplicar inseticidas no interior das casas no horário de maior atividade dos vetores. A sensibilidade dos triatomíneos adultos aos inseticidas é maior no período inicial da noite, pois este é o momento em que estão mais ativos e que permanecem mais tempo fora de seus abrigos. Os insetos adultos saem para buscar alimento e, após o repasto sanguíneo, suas cutículas abdominais ficam distendidas. Esta distensão do tecido facilitaria, portanto, a absorção dos componentes químicos ativos nos inseticidas (Lazzari e Lorenzo, 2009).

Os estádios juvenis de triatomíneos também apresentam um papel importante na transmissão da doença para humanos, principalmente quando pertencentes às espécies capazes de formar colônias no interior das casas, como *R. prolixus* e *T. infestans*. Apesar de serem ápteros e apresentarem baixa mobilidade em relação aos adultos (Vazquez-Prokopec et al., 2004), são igualmente capazes de transmitir o protozoário *T. cruzi*. Ainda carecem informações acerca do comportamento de triatomíneos durante os estádios juvenis. Nos poucos trabalhos que incluem a análise do ritmo circadiano de ninfas, os autores pressupõem que estes indivíduos apresentam o mesmo padrão noturno de atividade observado em adultos (e.g. Barrozo et al., 2004; Bodin et al., 2008).

Processos evolutivos como especiação e formação de barreiras naturais de isolamento tampouco foram analisados anteriormente em *Rhodnius*. Análises de

sinalização química, como feromônios, e de picos de atividade locomotora diária, por exemplo, podem ser relevantes na inferência de possíveis sinais de isolamento reprodutivo e temporal, respectivamente, entre espécies incipientes ou filogeneticamente próximas.

## 2. JUSTIFICATIVA

As dificuldades para a interrupção definitiva da transmissão da doença de Chagas decorrem de eventos de reinfestação das casas previamente tratadas com inseticidas e, de maneira secundária, do contato esporádico do homem com vetores infectados que habitam o peridomicílio e o ambiente silvestre. Este panorama epidemiológico complexo de endemismo e hipoendemismo motivou autoridades e pesquisadores a investir no conhecimento da dinâmica populacional, da ecologia e do comportamento não apenas dos vetores primários, como também dos vetores silvestres (Abad-Franch e Monteiro, 2005).

A escrutinização das regiões potenciais de transmissão endêmicas ou hipoendêmicas depende da correta diagnose das espécies vetoras nativas e dos seus limites de ocorrência geográfica. No caso de espécies crípticas, a identificação taxonômica deve ser feita preferencialmente por métodos moleculares de baixo custo e eficientes, cujos resultados sejam de simples interpretação.

Com este objetivo em mente, Pavan e Monteiro (2007) criaram um método diagnóstico de *PCR-multiplex*, que utiliza iniciadores espécie-específicos para a amplificação de fragmentos de tamanhos distintos do gene mitocondrial *Cytb*, outrora analisado por Monteiro e cols. (2003). Ao submeter as amostras amplificadas à eletroforese em um gel de agarose 1,5%, é possível diferenciar *R. prolixus* de *R. robustus* I e ainda *R. prolixus* das linhagens crípticas *R. robustus* II-IV.

Embora sua eficiência tenha sido comprovada na maior parte das áreas de distribuição desses táxons, pode haver problemas em regiões de possível simpatria. Devido a natureza matrilinear deste marcador, sua análise exclusiva pode gerar resultados equivocados onde *R. prolixus* e *R. robustus s.l.* co-ocorrem e ocasionalmente hibridizam, causando introgressão do DNA mitocondrial (Fitzpatrick et al., 2008). Para superar essa limitação, faz-se necessária a criação de um método molecular (complementar ao *PCR-multiplex* descrito) que utilize um marcador nuclear de cópia única.

Processos como os de domiciliação e de especiação de vetores envolvem interações ecológicas, biológicas, comportamentais e evolutivas complexas, estruturadas hierarquicamente no espaço e no tempo e, portanto, seu entendimento depende essencialmente de estudos em diferentes escalas geográficas e históricas. Os

dados extraídos nestes tipos de análises poderão ser úteis no desenho de novas estratégias de controle vetorial (Abad-Franch e Monteiro, 2007; Lazzari e Lorenzo, 2009).

A maior fonte de estímulo para o desenvolvimento desta tese adveio da constatação de que ainda existem muitas lacunas no conhecimento acerca dos padrões e processos geradores da diversidade encontrada em triatomíneos. O complexo de espécies crípticas *R. prolixus* e *R. robustus s.l.* foi utilizado como modelo, por ser um grupo bem estudado, de filogenia conhecida e ocorrência geográfica de cada espécie previamente estimada.

A investigação de populações de *R. prolixus* e do complexo de linhagens crípticas *R. robustus s.l.* na região pan-amazônica foi proposta na tentativa de elucidar questões geográficas, populacionais e biológicas geradoras da diversidade encontrada, como: (1) os limites geográficos de cada linhagem; (2) a validade das linhagens pertencentes ao complexo *R. robustus* como sendo espécies distintas; e (3) as possíveis barreiras geográficas e comportamentais ao fluxo gênico. Para alcançar estas metas, foram escolhidas regiões criteriosamente selecionadas que representam duas áreas contíguas de distribuição de três linhagens de *R. robustus* (II, III e IV), ao longo das quais foram realizados transectos de amostragem. Isso possibilitou a investigação minuciosa de áreas que são as mais ricas em informações sobre o processo de especiação em si: os gradientes ou zonas híbridas compreendidos em regiões de contato entre duas espécies (parapátricas).

As questões levantadas nesta tese referentes à identificação molecular de *R. prolixus* e *R. robustus s.l.* e ainda aos processos geradores da diversidade molecular observada foram respondidas a partir de quatro abordagens distintas: (1) identificação de um marcador molecular em um gene de cópia única que separe *R. prolixus* de *R. robustus s.l.*; (2) análises filogenéticas e filogeográficas das populações de *R. prolixus* e *R. robustus s.l.*, mediante sequenciamento de dois marcadores já empregados em triatomíneos (mtCytb e ITS-2), além do isolamento e sequenciamento de três genes nucleares relacionados ao comportamento alimentar, sexual e isolamento reprodutivo em *Drosophila* (ortólogos de *per*, *tim* e *cry*); (3) genética de populações, com base na análise de dez *loci* de microssatélites das populações simpátricas de diferentes linhagens de *R. robustus s.l.*; e (4) análise comportamental, a partir da determinação

dos picos de atividade locomotora e o período intrínseco do relógio circadiano de *R. prolixus* e *R. robustus s.l.*

### 3. OBJETIVOS

#### 3.1. Objetivos gerais

Avaliar o status taxonômico de *R. prolixus* e das linhagens pertencentes ao complexo *R. robustus s.l.* e determinar as possíveis barreiras naturais ao fluxo gênico entre as linhagens provenientes da região amazônica, a partir de análises filogenéticas, biogeográficas, populacionais e de comportamento.

#### 3.2. Objetivos específicos

- Determinar a presença de possíveis zonas híbridas nas regiões limítrofes das distribuições geográficas conhecidas das linhagens II, III e IV de *R. robustus*.
- Desenvolver um método molecular diagnóstico eficiente, cujo resultado seja de fácil interpretação, para a separação de *R. prolixus* e *R. robustus s.l.*, com base em um marcador nuclear.
- Quantificar a diversidade genética intra e interespecífica de *R. prolixus* e do complexo *R. robustus s.l.* a partir das sequências de mtCytb e ITS-2.
- Estimar o tempo de divergência entre *R. prolixus* e *R. robustus s.l.* e ainda entre as linhagens contidas no complexo *R. robustus* com base nas sequências do mtCytb, a fim de situar as possíveis cladogêneses no tempo geológico.
- Comparar os resultados filogeográficos com as diferentes previsões das hipóteses biogeográficas de geração da diversidade na região amazônica de ‘incurções marinhas’ e ‘refúgios pleistocênicos’.
- Realizar inferências sobre a origem e dispersão de *R. prolixus* e *R. robustus s.l.* na região amazônica, a partir do estudo filogenético, genealógico e coalescente das sequências do mtCytb.
- Analisar o marcador nuclear ITS-2 em concomitância com o mtCytb em regiões de possível simpatria ou peripatria entre espécies amazônicas de *R. robustus*, para a identificação de possíveis híbridos naturais.
- Avaliar a existência de fluxo gênico entre as linhagens amazônicas de *R. robustus* em regiões de simpatria, utilizando 10 *loci* de microssatélites.
- Isolar no genoma de *R. prolixus* os ortólogos de três genes ligados ao comportamento alimentar, sexual e isolamento reprodutivo descritos em *Drosophila* (*per*, *tim* e *cry*), para testá-los como possíveis marcadores

moleculares que diferenciem filogeneticamente complexos de espécies crípticas em Triatominae.

- Determinar os picos de atividade locomotora e o período intrínseco do relógio circadiano de *R. prolixus* e *R. robustus* II, para verificar possíveis diferenças no padrão de atividade entre as espécies.

## 4. METODOLOGIA

### 4.1. Estratégia de coleta e obtenção das amostras

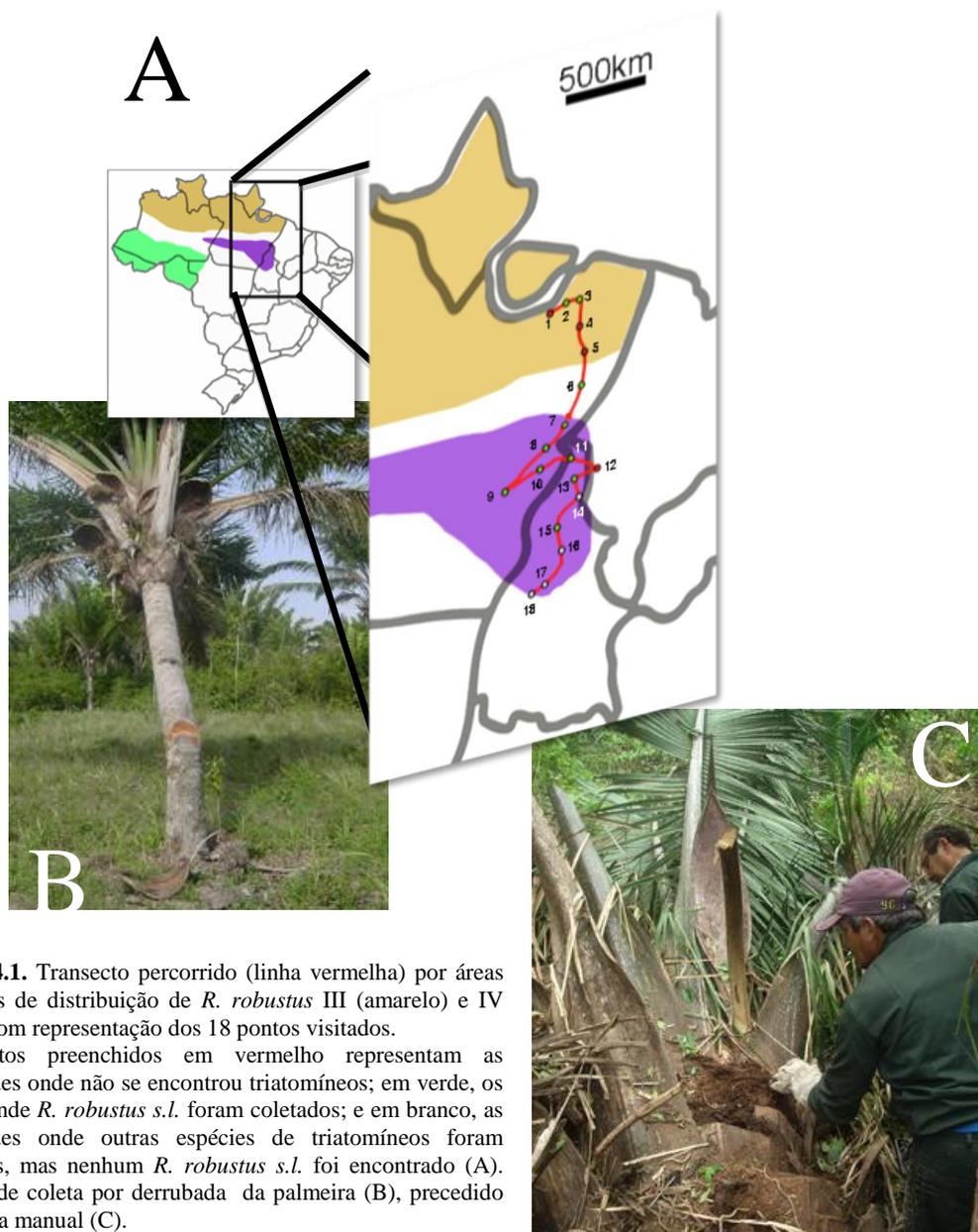
Foram realizadas duas coletas ao longo de transectos que representam áreas contíguas de ocorrência de *R. robustus* III e IV, e *R. robustus* II, III e IV, de acordo com o que se conhece sobre a distribuição destes táxons (Monteiro et al., 2003; Pavan e Monteiro, 2007). A primeira coleta foi realizada entre os dias 25 de outubro e 15 de novembro de 2009, e a segunda entre os dias 25 de setembro e 15 de outubro de 2010. As capturas e o transporte dos insetos vivos foram autorizados pelo IBAMA (SISBIO 20287-1). Todos os espécimes obtidos que sobreviveram à viagem foram mantidos nos insetários do Laboratório de Doenças Parasitárias e do Laboratório de Referência Nacional e Internacional em Taxonomia de Triatomíneos, localizados no Instituto Oswaldo Cruz, Fiocruz. Os espécimes mortos foram devidamente etiquetados quanto ao local de origem e acondicionados em tubos contendo álcool absoluto.

No primeiro transecto foram percorridos 2500 km (Figura 4.1), ao longo dos quais foram selecionados 18 pontos para a coleta de triatomíneos. A coleta foi iniciada em Belém (Pará) e terminou em Palmas (Tocantins). O método utilizado para captura dos insetos foi a derrubada de palmeiras *Attalea sp.* precedida da coleta manual (Diotaiuti et al., 1984). Em cada localidade foram amostradas de uma a cinco palmeiras, todas localizadas a menos de 600 m de distância das habitações humanas. O escore de matéria orgânica em decomposição (MOD) de cada palmeira foi determinado visualmente, como proposto por Abad-Franch e cols. (2005). A porcentagem de infestação das palmeiras (número de palmeiras positivas para a presença de triatomíneos dividido pelo número total de palmeiras amostradas), assim como a densidade de triatomíneos por palmeira amostrada foram determinados para cada ponto de coleta.

Em um segundo transecto com 1500 km de extensão foram visitados 10 pontos (Figura 4.2), iniciando em Manaus (Amazonas) e terminando em Porto Velho (Rondônia). O método utilizado para coleta de triatomíneos foi o de captura manual nas copas de palmeiras *Oenocarpus bataua* e *Attalea sp.* (Abad-Franch et al., 2010a). Este método consiste em subir com o auxílio de uma escada até a altura da copa da palmeira, fazer uma inspeção visual para a presença de triatomíneos, retirar alguns ramos da palmeira a partir da base (“limpeza da palmeira”), coletar uma parte da

matéria orgânica contida na copa e nas junções dos ramos da palmeira com o tronco até encher um saco plástico de 50 litros. Posteriormente este material é colocado em um lençol branco e checado visualmente para a presença de triatomíneos. Este processo foi repetido em cada palmeira de cinco a sete vezes. Todas as palmeiras escolhidas estavam situadas em ambiente silvestre e distavam de 30 a 100 m da rodovia. Assim como realizado no primeiro transecto, os ecores de MOD, a porcentagem de infestação das palmeiras por triatomíneos e a densidade de triatomíneos por palmeira amostrada foram determinados para cada ponto de coleta do segundo transecto.

Adicionalmente foram obtidos seis espécimes coletados em um estudo de caso de surto de doença de Chagas aguda com suspeita de transmissão oral no município de São Roberto, no estado do Maranhão, além de outras 58 amostras provenientes de 18 localidades distribuídas por cinco estados brasileiros e 115 amostras coletadas em 23 localidades distribuídas pelo Equador, Venezuela, Colômbia, Perú, Bolívia, Honduras e Guiana Francesa por colaboradores ou membros do laboratório (Tabela 4.3). As localidades que tiveram pelo menos um representante sequenciado para um determinado marcador são mostradas na Figura 4.3.

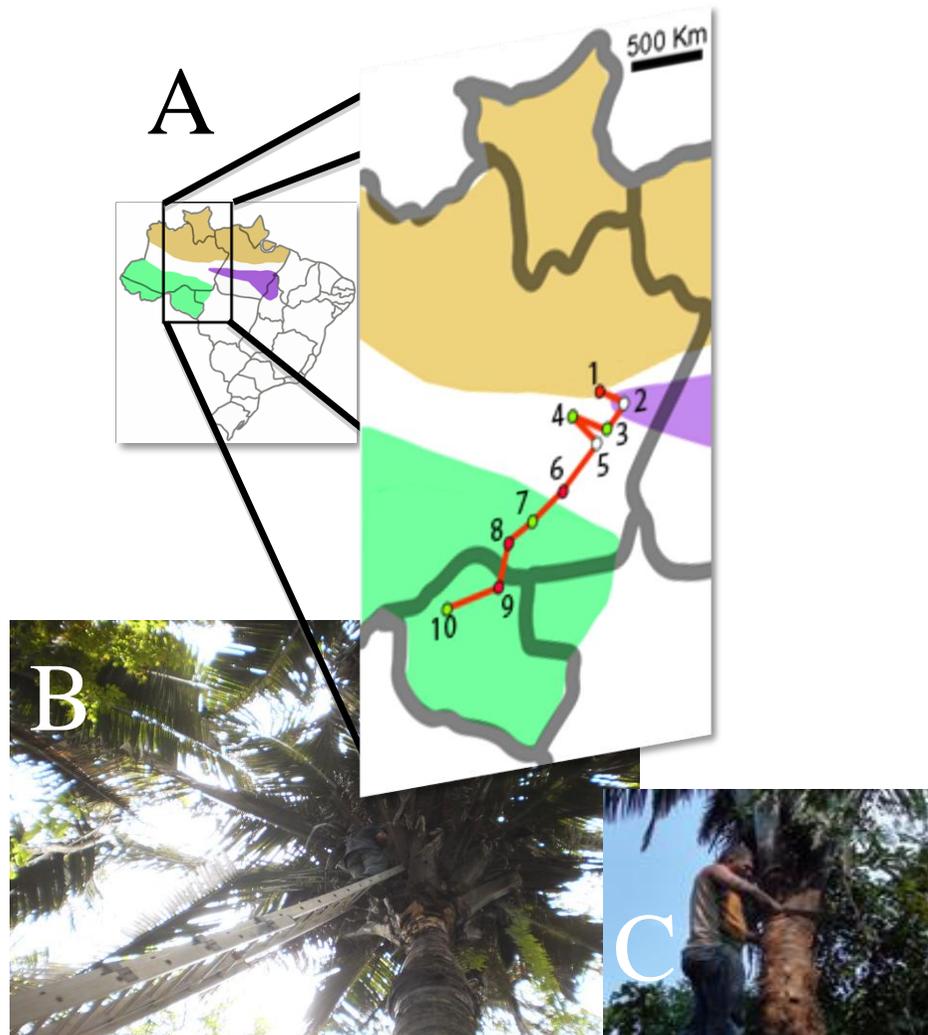


**Figura 4.1.** Transecto percorrido (linha vermelha) por áreas contíguas de distribuição de *R. robustus* III (amarelo) e IV (roxo), com representação dos 18 pontos visitados. Os pontos preenchidos em vermelho representam as localidades onde não se encontrou triatomíneos; em verde, os pontos onde *R. robustus s.l.* foram coletados; e em branco, as localidades onde outras espécies de triatomíneos foram coletados, mas nenhum *R. robustus s.l.* foi encontrado (A). Método de coleta por derrubada da palmeira (B), precedido por coleta manual (C).

**Tabela 4.1.** Localidades visitadas durante o percurso pelo transecto representado na Figura 4.1A.

As localidades estão ordenadas de acordo com os pontos do mapa (sentido norte → sul).  $N_{PA}$  – Número de palmeiras derrubadas;  $N_{PP}$  – Número de palmeiras positivas;  $N_{RR}$  - Número de espécimes de *R. robustus s.l.* coletados em cada palmeira positiva.

| Localidade                            | Espécie de palmeira     | $N_{PA} / N_{PP}$ | $N_{RR}$   | Outras espécies de triatomíneos coletadas  |
|---------------------------------------|-------------------------|-------------------|------------|--|
| 1. Belém, Pará                        | <i>Attalea maripa</i>   | 2 / 0             | 0          | -  |
| 2. Ponta das Pedras, Pará             | <i>Attalea maripa</i>   | 2 / 1             | 5          | -  |
| 3. Santa Maria, Pará                  | <i>Attalea maripa</i>   | 5 / 1             | 1          | -  |
| 4. Aurora do Pará, Pará               | <i>Attalea maripa</i>   | 4 / 0             | 0          | -  |
| 5. Paragominas, Pará                  | <i>Attalea maripa</i>   | 1 / 0             | 0          | -  |
| 6. Ulianópolis, Pará                  | <i>Attalea maripa</i>   | 3 / 1             | 13         | -  |
| 7. Rondon do Pará, Pará               | <i>Attalea maripa</i>   | 4 / 2             | 2 / 32     | -  |
| 8. Marabá, Pará                       | <i>Attalea maripa</i>   | 1 / 1             | 10         | -  |
| 9. Paraupabas, Pará                   | <i>Attalea maripa</i>   | 2 / 2             | 13 / 17    | -  |
| 10. São Domingos do Araguaia, Pará    | <i>Attalea maripa</i>   | 1 / 1             | 15         | -  |
| 11. Estrela, Tocantins                | <i>Attalea speciosa</i> | 1 / 1             | 4          | -  |
| 12. Imperatriz, Maranhão              | <i>Attalea speciosa</i> | 1 / 0             | 0          | -  |
| 13. São Bento do Tocantins, Tocantins | <i>Attalea speciosa</i> | 1 / 1             | 20         | 35 <i>R. neglectus</i> , 4 <i>R. pictipes</i> e 1 <i>Panstrongylus lignarius</i> |
| 14. Tocantinópolis, Tocantins         | <i>Attalea speciosa</i> | 1 / 1             | 0          | 48 <i>R. neglectus</i>   |
| 15. Araguaína, Tocantins              | <i>Attalea speciosa</i> | 1 / 1             | 5          | -  |
| 16. Palmeirante, Tocantins            | <i>Attalea speciosa</i> | 1 / 1             | 0          | 55 <i>R. neglectus</i> / 63 <i>R. neglectus</i>                                  |
| 17. Presidente Kennedy, Tocantins     | <i>Attalea speciosa</i> | 1 / 1             | 0          | 45 <i>R. neglectus</i>   |
| 18. Colméia, Tocantins                | <i>Attalea speciosa</i> | 2 / 2             | 0          | 3 <i>R. neglectus</i> e 1 <i>R. pictipes</i> / 3 <i>R. neglectus</i>             |
| <b>TOTAL:</b>                         |                         |                   | <b>137</b> | <b>252 <i>R. neglectus</i>, 5 <i>R. pictipes</i> e 1 <i>P. lignarius</i>.</b>    |



**Figura 4.2.** Transecto percorrido (linha vermelha) por áreas contíguas de distribuição de *R. robustus* II, III e IV, com representação dos 10 pontos visitados.

Os pontos preenchidos em vermelho representam as localidades cujos triatomíneos não foram encontrados, em verde, os pontos onde *R. robustus s.l.* foram coletados em branco, as localidades onde outras espécies de triatomíneos foram coletados, mas nenhum *R. robustus s.l.* foi encontrado (A). Método de coleta por busca ativa na copa das palmeiras (B), seguido de “limpeza” da palmeira (C) e inspeção visual em um lençol branco do material coletado nas copas (D).

**Tabela 4.2.** Localidades visitadas durante o percurso pelo transecto representado na Figura 4.2A.

As localidades estão ordenadas de acordo com os pontos do mapa (sentido norte → sul).  $N_{PA}$  – Número de palmeiras amostradas;  $N_{PP}$  – Número de palmeiras positivas;  $N_{Rr}$  – Número de espécimes de *R. robustus s.l.* coletados em cada palmeira positiva.

| Localidade                              | Espécie de palmeira      | $N_{PA}$ / $N_{PP}$ | $N_{Rr}$           | Outras espécies de triatomíneos coletadas       |
|---|--------------------------|---------------------|--------------------|---|
| 1. Manaus, Amazonas                     | <i>Oenocarpus bataua</i> | 5 / 0               | 0                  | -   |
| 2. Purupuru, Amazonas                   | <i>Attalea speciosa</i>  | 4 / 1               | 0                  | 3 <i>R. pictipes</i> / - / - / -                |
| 3. Careiro Castanho, Amazonas           | <i>Attalea maripa</i>    | 10 / 3              | 2 / 3 / 6          | 2 <i>R. pictipes</i> / 4 <i>R. pictipes</i> / - |
| 4. Anamá, Amazonas                      | <i>Attalea phalerata</i> | 15 / 5              | 3 / 4 / 8 / 2 / 10 | -   |
| 5. Vilarejo Km 125 BR-319, Amazonas     | <i>Attalea speciosa</i>  | 5 / 2               | 2 / 0              | 2 <i>R. pictipes</i> / 1 <i>R. pictipes</i>     |
| 6. Vilarejo Km 280 BR-319, Amazonas     | <i>Attalea speciosa</i>  | 7 / 0               | 0                  | -   |
| 7. Vilarejo Km 450 BR-319, Amazonas     | <i>Attalea speciosa</i>  | 6 / 1               | 2                  | -   |
| 8. Humaitá, Amazonas                    | <i>Attalea speciosa</i>  | 5 / 0               | 0                  | -   |
| 9. Vilarejo a 50 Km de Calama, Rondônia | <i>Attalea speciosa</i>  | 5 / 0               | 0                  | -   |
| 10. Porto Velho, Rondônia               | <i>Attalea speciosa</i>  | 8 / 4               | 12 / 15 / 10 / 13  | -   |
| <b>TOTAL:</b>                           |                          |                     | <b>91</b>          | <b>12 <i>R. pictipes</i></b>                    |

**Tabela 4.3.** Amostras de *Rhodnius* utilizadas no sequenciamento de sete marcadores. Os números abaixo de cada marcador utilizado (colunas entre “Área biogeográfica” e “Campo/Colônia”) representam o número de indivíduos provenientes de cada localidade / colônia que tiveram o fragmento em questão sequenciados. Localidades que não tiveram indivíduos sequenciados para um determinado marcador são representadas com um hífen (“-“). As informações apresentadas na coluna do ITS-2 são o número de indivíduos sequenciados, seguido do número de clones obtidos para cada indivíduo (em parêntesis). As informações acerca dos outros amplicons (além do AmpG) testados para o desenvolvimento de um método diagnóstico de *R. prolixus* e *R. robustus* s.l. foram omitidas aqui, mas podem ser encontradas no Anexo 1. Maiores detalhes acerca dos marcadores moleculares *period* (*per*), *cryptochrome-2* com iniciadores senso ancorados nos éxons 5 [*cry2* (E5)] e 7 [*cry2* (E7)] do gene ortólogo de *R. prolixus*, e *timeless*, com iniciador senso e anti-senso ancorados nos éxons 15 e 16, respectivamente, do gene ortólogo de *R. prolixus* [*tim* (E15)] podem ser encontrados na seção 5.4 dos Resultados.

| Espécie*                     | Localidade de coleta / Localidade da colônia | Área biogeográfica† | mtCytb | AmpG | ITS-2 | <i>per</i> | <i>cry2</i> (E5) | <i>cry2</i> (E7) | <i>tim</i> (E21) | Campo/Colônia | Ano de coleta |
|------------------------------|--|---------------------|--------|------|-------|------------|------------------|------------------|------------------|---------------|---------------|
| <i>R. prolixus</i>           | "Desconhecida" / IOC, Brasil                 | Orinoco             | 5      | 23   | -     | -          | -                | -                | -                | colônia       | ?             |
|                              | 1. Guatemala / MERTU, Guatemala              | Orinoco             | 5      | 10   | -     | 2          | 1                | 1                | 1                | colônia       | ?             |
|                              | 2. Orico, Francisco Morazan, Honduras        | Orinoco             | 7      | -    | -     | -          | -                | -                | -                | campo         | 1999          |
|                              | 3. Colômbia / CDCI, Atlanta, EUA             | Orinoco             | 10     | 10   | -     | 2          | -                | -                | 1                | colônia       | ?             |
|                              | 3. Tolima, Colômbia                          | Orinoco             | 2      | -    | -     | 5          | -                | -                | 2                | campo         | 2009          |
|                              | 4. Barinas, Venezuela                        | Orinoco             | -      | 3    | -     | -          | -                | -                | -                | campo         | 2004          |
|                              | 5. Pampanito, Trujillo, Venezuela            | Orinoco             | 10     | -    | 1 (1) | -          | -                | -                | 1                | colônia       | 1997          |
|                              | 6. Cojedes, Venezuela                        | Orinoco             | 2      | -    | 1 (1) | -          | -                | -                | -                | colônia       | 2001          |
|                              | 7. San José Tiznados, Guárico, Venezuela     | Orinoco             | 5      | -    | -     | -          | -                | -                | -                | colônia       | 1988          |
| 7. Ortiz, Guárico, Venezuela | Orinoco                                      | 3                   | -      | -    | -     | -          | 2                | -                | campo            | 1995          |               |
| <i>R. robustus</i> I         | 5. Pampanito, Trujillo, Venezuela            | Orinoco             | 4      | -    | 1 (1) | -          | -                | -                | -                | colônia       | 1997          |
|                              | 5. Candelaria, Trujillo, Venezuela           | Orinoco             | 4      | -    | -     | -          | -                | -                | -                | colônia       | 1988          |
|                              | 5. Trujillo, Venezuela                       | Orinoco             | -      | 2    | -     | -          | -                | -                | -                | campo         | 2004          |
| <i>R. robustus</i> II        | 8. Sucumbios, Equador                        | Napo                | 1      | 5    | -     | -          | -                | -                | -                | colônia       | ?             |
|                              | 9. Napo, Equador                             | Napo                | 1      | -    | -     | -          | -                | -                | -                | colônia       | ?             |

| Espécie*                      | Localidade de coleta / Localidade da colônia | Área biogeográfica† | mtCytb | AmpG | ITS-2 | per         | cry2 (E5) | cry2 (E7) | tim (E21) | Campo/ Colônia | Ano de coleta |      |
|-------------------------------|--|---------------------|--------|------|-------|-------------|-----------|-----------|-----------|----------------|---------------|------|
| <i>R. robustus</i> II (cont.) | 10. Loreto, Perú                             | Inambarí            | 10     | -    | -     | 2           | 1         | -         | 1         | colônia        | ?             |      |
|                               | 11. Carauari, Amazonas, Brasil               | Inambarí            | 4      | 4    | 1 (1) | -           | -         | -         | -         | campo          | 2000          |      |
|                               | 12. Porto Velho, Rondônia, Brasil            | Rondônia            | 39     | 45   | -     | 13          | 1         | 2         | 3         | campo          | 2010          |      |
|                               | 13. Santo Antônio, Rondônia, Brasil          | Rondônia            | 1      | 1    | -     | -           | -         | -         | 1         | campo          | 2010          |      |
|                               | 14. Ouro Preto do Oeste, Rondônia, Brasil    | Rondônia            | 9      | 7    | -     | 4           | 1         | 1         | 4         | campo          | 2010          |      |
|                               | 15. Guajará Mirim, Rondônia, Brasil          | Rondônia            | 3      | 2    | -     | -           | -         | -         | -         | campo          | 2010          |      |
|                               | 16. Monte Negro, Rondônia, Brasil            | Rondônia            | 1      | 2    | -     | -           | -         | -         | -         | campo          | 2004          |      |
|                               | 17. Cacoal, Rondônia, Brasil                 | Rondônia            | 1      | 2    | -     | -           | -         | -         | -         | campo          | 2010          |      |
|                               | 18. San Gabriel, Pando, Bolívia              | Inambarí            | **     | 2    | -     | -           | -         | -         | -         | campo          | 2008          |      |
|                               | 19. Guarayos, Santa Cruz, Bolívia            | Inambarí            | **     | 7    | -     | -           | -         | 1         | 1         | campo          | 2008          |      |
|                               | 20. San Ramón, Santa Cruz, Bolívia           | Inambarí            | **     | 1    | -     | -           | 1         | -         | -         | campo          | 2008          |      |
|                               | 21. Montero, Santa Cruz, Bolívia             | Inambarí            | **     | 3    | -     | -           | 1         | -         | 1         | campo          | 2008          |      |
|                               | 22. El Torno, Santa Cruz, Bolívia            | Inambarí            | **     | 2    | -     | -           | 1         | 3         | 2         | campo          | 2008          |      |
|                               | 23. Cochabamba, Chaparé, Bolívia             | Inambarí            | **     | -    | -     | 1 (1 cada)‡ | -         | 1         | -         | -              | campo         | 2008 |
|                               | 24. Apuí, Amazonas, Brasil                   | Rondônia            | 4      | -    | -     | -           | -         | -         | -         | -              | colônia       | 1996 |
|                               | 25. Careiro Castanho, Amazonas, Brasil       | Inambarí            | 2      | -    | -     | 2 (3 e 10)  | 2         | 3         | 2         | 2              | campo         | 2010 |
| 26. Anamá, Amazonas, Brasil   | Inambarí                                     | 2                   | -      | -    | -     | -           | 1         | -         | 1         | campo          | 2010          |      |
| <i>R. robustus</i> III        | 27. Carajás, Pará, Brasil                    | Pará                | 2      | -    | -     | 1           | 1         | -         | -         | campo          | 2008          |      |
|                               | 28. Novo Repartimento, Pará, Brasil          | Pará                | 5      | 2    | 1 (1) | -           | -         | -         | -         | colônia        | 1988          |      |
|                               | 29. Itupiranga, Pará, Brasil                 | Pará                | 5      | -    | -     | -           | -         | -         | -         | colônia        | 1984          |      |
|                               | 30. Marabá, Pará, Brasil                     | Pará                | 1      | -    | -     | 8           | 1         | -         | 2         | campo          | 2009          |      |
|                               | 31. Paraupibas, Pará, Brasil                 | Pará                | 22     | -    | -     | -           | 5         | 1         | 1         | 7              | campo         | 2009 |

| Espécie*                       | Localidade de coleta / Localidade da colônia  | Área biogeográfica† | mtCytb | AmpG | ITS-2           | per | cry2 (E5) | cry2 (E7) | tim (E21) | Campo/ Colônia | Ano de coleta |
|--------------------------------|---|---------------------|--------|------|-----------------|-----|-----------|-----------|-----------|----------------|---------------|
| <i>R. robustus</i> III (cont.) | 32. São Domingos do Araguaia, Pará, Brasil    | Pará                | 4      | 3    | -               | -   | -         | -         | -         | campo          | 2009          |
|                                | 33. São Bento do Tocantins, Tocantins, Brasil | Pará                | 10     | 3    | -               | 3   | 1         | -         | 6         | campo          | 2009          |
|                                | 34. Araguaína, Tocantins, Brasil              | Pará                | 2      | 2    | -               | -   | -         | -         | -         | campo          | 2009          |
|                                | 35. Tocantinópolis, Tocantins, Brasil         | Pará                | 4      | -    | -               | -   | -         | -         | -         | colônia        | ?             |
|                                | 36. Estrela, Tocantins, Brasil                | Pará                | 2      | -    | 1 (10)          | -   | -         | -         | -         | campo          | 2009          |
|                                | 37. Rondon do Pará, Pará, Brasil              | Pará                | 3      | 2    | -               | -   | -         | -         | 1         | campo          | 2009          |
|                                | 38. Ulianópolis, Pará, Brasil                 | Pará                | 6      | 4    | -               | -   | -         | -         | -         | campo          | 2009          |
| <i>R. robustus</i> n.sp.       | 38. Ulianópolis, Pará, Brasil                 | Pará                | 6      | 3    | 3 (10, 12 e 14) | 1   | 3         | 3         | -         | campo          | 2009          |
|                                | 39. São Roberto, Maranhão, Brasil             | Pará                | 6      | -    | 3 (10, 12 e 12) | -   | -         | -         | 2         | campo          | 2011          |
| <i>R. robustus</i> IV          | 40. Santa Maria, Pará, Brasil                 | Belém               | 1      | 1    | -               | -   | -         | -         | -         | campo          | 2009          |
|                                | 41. Ponta das Pedras, Pará, Brasil            | Belém               | 1      | -    | -               | 1   | 1         | -         | 1         | campo          | 2009          |
|                                | 42. Barcarena, Pará, Brasil                   | Belém               | 3      | -    | -               | -   | -         | -         | -         | colônia        | 1996          |
|                                | 43. Abaetetuba, Pará, Brasil                  | Belém               | 1      | -    | 1 (11)          | 1   | -         | -         | -         | campo          | 2011          |
|                                | 44. São José da Povoação, Pará, Brasil        | Belém               | 3      | -    | -               | 1   | 1         | 1         | 2         | campo          | 2008          |
|                                | 45. Curalinho, Pará, Brasil                   | Belém               | 2      | -    | -               | -   | -         | -         | -         | campo          | 2008          |
|                                | 46. Bagre, Pará, Brasil                       | Belém               | 2      | -    | -               | -   | 4         | 2         | 1         | campo          | 2008          |
|                                | 47. Oriximiná, Pará, Brasil                   | Rondônia            | 4      | -    | -               | -   | -         | -         | -         | colônia        | 1986          |
|                                | 47. Oriximiná, Pará, Brasil                   | Rondônia            | 1      | -    | -               | -   | -         | -         | 1         | campo          | 2008          |
|                                | 48. Manaus, Amazonas, Brasil                  | Imerí               | 3      | -    | -               | -   | -         | -         | -         | campo          | 2004          |
|                                | 49. Mucajaí, Roraima, Brasil                  | Guiana              | 2      | -    | -               | -   | -         | -         | -         | campo          | 2005          |
|                                | 50. UHE Paredão, Roraima, Brasil              | Guiana              | 5      | -    | -               | -   | -         | -         | -         | colônia        | 1987          |

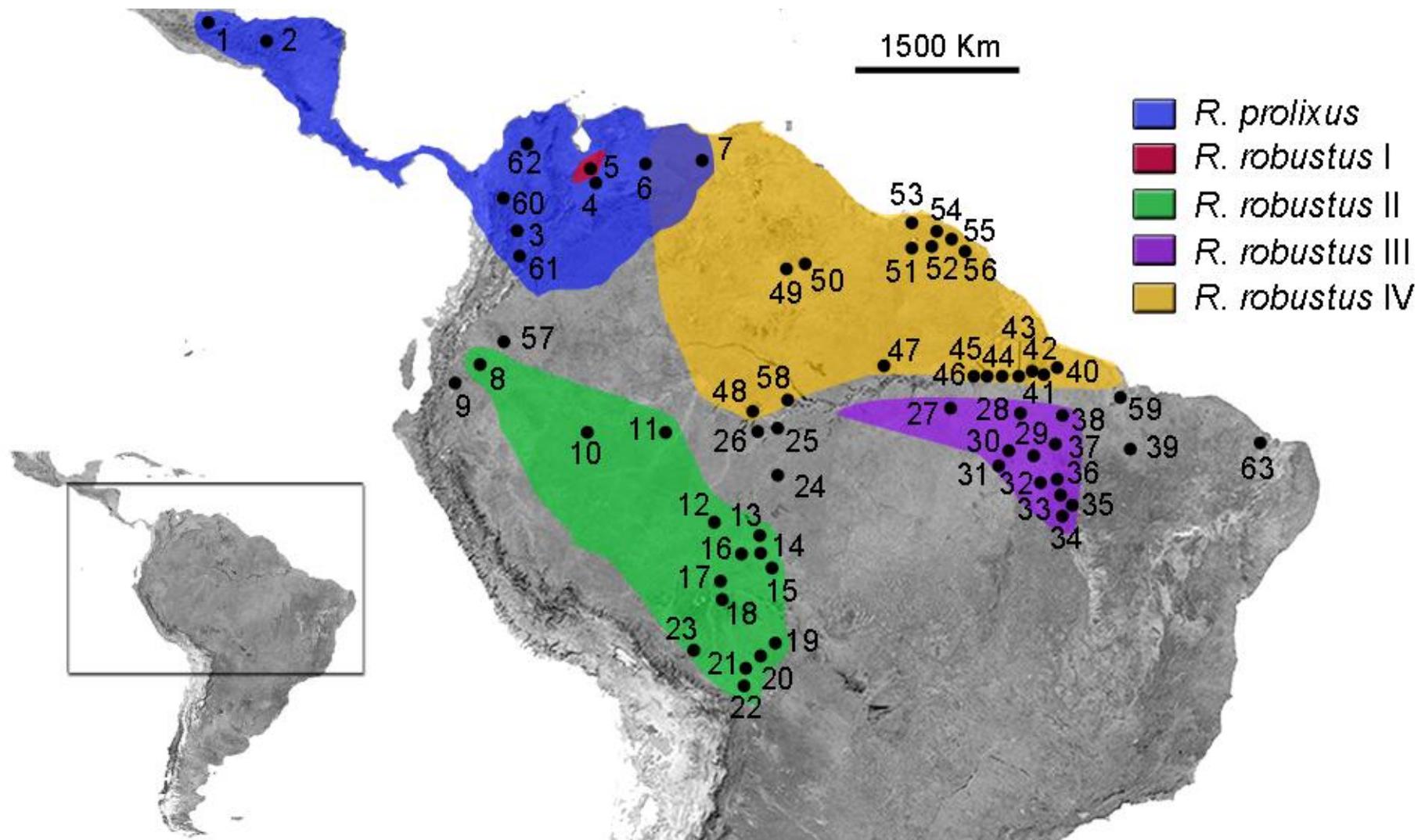
| Espécie*                      | Localidade de coleta / Localidade da colônia     | Área biogeográfica† | mtCytb     | AmpG       | ITS-2           | per       | cry2 (E5) | cry2 (E7) | tim (E21) | Campo/ Colônia | Ano de coleta |
|-------------------------------|--|---------------------|------------|------------|-----------------|-----------|-----------|-----------|-----------|----------------|---------------|
| <i>R. robustus</i> IV (cont.) | 51. Belizón, Guiana Francesa                     | Guiana              | 2          | -          | -               | -         | -         | -         | -         | campo          | 2003          |
|                               | 52. Camopi, Guiana Francesa                      | Guiana              | 1          | -          | -               | -         | -         | -         | -         | campo          | 2002          |
|                               | 53. Macouria, Guiana Francesa                    | Guiana              | 1          | -          | -               | -         | -         | -         | -         | campo          | 2003          |
|                               | 54. Cayenne, Guiana Francesa                     | Guiana              | 3          | 2          | 1 (1)           | -         | -         | -         | 1         | campo          | 2003          |
|                               | 55. Remiré, Guiana Francesa                      | Guiana              | 3          | 1          | -               | -         | -         | -         | -         | campo          | 2003          |
|                               | 56. Kaw, Guiana Francesa                         | Guiana              | 1          | -          | -               | -         | -         | -         | -         | campo          | 2003          |
| <i>R. robustus</i> V          | 8. Localidade próxima à Manaus, Amazonas, Brasil | Imerí               | 1          | -          | -               | -         | -         | -         | -         | campo          | 2001          |
| <i>R. barretti</i>            | 23. Sucumbios, Equador                           | Napo                | 2          | -          | -               | -         | -         | -         | -         | campo          | 1999          |
|                               | 57. Putumayo, Colombia                           | Napo                | 2          | -          | -               | -         | -         | -         | -         | campo          | 2001          |
| <i>R. pictipes</i>            | 58. Itacoatiara, Amazonas, Brasil                | Inambarí            | 1          | -          | -               | -         | -         | -         | -         | campo          | 2005          |
|                               | 49. Mucajaí, Roraima, Brasil                     | Guiana              | 1          | -          | -               | -         | -         | -         | -         | campo          | 2005          |
|                               | 59. Rosário, Maranhão, Brasil                    | Belém               | 1          | -          | -               | -         | -         | -         | -         | campo          | 2008          |
| <i>R. pallescens</i> I        | 60. Vegachi, Antioquia, Colômbia                 | Orinoco             | 1          | -          | -               | -         | -         | -         | -         | campo          | 2007          |
|                               | 61. Santa Fé, Veraguas, Colômbia                 | Orinoco             | 1          | -          | -               | -         | -         | -         | -         | campo          | 2007          |
| <i>R. pallescens</i> II       | 62. San Zenón, Magdalena, Colômbia               | Trans-Andino        | 1          | -          | -               | -         | -         | -         | -         | colônia        | 2007          |
| <i>R. nasutus</i>             | 63. Ceará, Brasil                                | Corredor Seco       | -          | -          | 1 (1)‡          | -         | -         | -         | -         | campo          | 2008          |
| <b>TOTAL:</b>                 |  |                     | <b>260</b> | <b>154</b> | <b>16 (112)</b> | <b>51</b> | <b>25</b> | <b>19</b> | <b>45</b> |                |               |

\* Identificação molecular da espécie, a partir do sequenciamento do mtCytb.

† *Sensu* Morrone (2006).

\*\* Amostras sequenciadas e molecularmente identificadas como sendo *R. robustus* II em outro estudo (Justi, 2009).

‡ Amostras geradas em outro estudo (Justi, 2009).



**Figura 4.3.** Distribuição geográfica dos espécimes amostrados e sequenciados de *R. prolixus*, *R. robustus* s.l., *R. barretti*, *R. pictipes*, *R. pallescens* I e II e *R. nasutus*. Cada localidade está numerada de acordo com a Tabela 4.3. As cores indicam a distribuição geográfica estimada de *R. prolixus* e *R. robustus* I-IV, segundo Monteiro et al. (2003), Pavan e Monteiro (2007) e Pavan et al. (2013).

#### 4.2. Extração de DNA e identificação dos espécimes

O tecido de cada espécime foi colocado em um tubo de 1,5mL, mergulhado em nitrogênio líquido até o congelamento, e macerado com bastões Kimble/Kontes Pellet Pestles<sup>®</sup>. No caso de adultos e ninfas de 5º estágio, foram utilizados somente um par de patas; para os estádios ninfais restantes foram utilizados os insetos inteiros.

O DNA genômico de cada espécime foi extraído utilizando o kit *Genomic DNA Extraction* (Real Genomics<sup>™</sup>), seguindo as recomendações do fabricante para o isolamento do material genético contido no tecido animal macerado. O DNA extraído foi eluído em 30 µL de uma solução disponibilizada pelo fabricante.

A identificação molecular dos espécimes foi feita a partir do sequenciamento de um fragmento de 663-pb do mtCytb sequenciado. Adicionalmente, as amostras de campo provenientes de regiões geográficas de possível encontro de espécies do complexo *R. robustus s.l.* foram seqüenciadas para o marcador ITS-2 (Marcilla et al., 2001). O mesmo foi feito para amostras provenientes de colônias mantidas em laboratório. Estes espécimes tiveram este marcador nuclear sequenciado para a identificação de possíveis eventos de hibridação e introgressão de DNA mitocondrial.

#### 4.3. Amplificação pela PCR e sequenciamento do mtCytb

A identificação dos espécimes provenientes do campo foi feita a partir da amplificação e sequenciamento de um fragmento de 663-pb do mtCytb. Para a amplificação deste fragmento, foram utilizados os iniciadores “CYTB7432F”, 5’ – GGACG(AT)GG(AT)ATTTATTATGGATC, e “CYTB7433R”, 5’ – GC(AT)CCAAT TCA(AG)GTTA(AG)TAA (Monteiro et al., 2003). As amostras foram colocadas em um termociclador Veriti<sup>®</sup> 96 Well (*Applied Biosystems*), ajustado para as seguintes condições: 94 °C por 5 minutos, seguida de 35 ciclos a 94 °C, por 30 segundos, 48 °C por 30 segundos, e 72°C por 45 segundos, e ainda uma extensão final de 10 minutos a 72°C. A purificação dos produtos amplificados pela PCR foi feita utilizando os kits *GFX<sup>™</sup> PCR DNA and gel band purification* (*GE Healthcare*) ou *HiYield gel/PCR DNA extraction* (*Real Genomics<sup>™</sup>*), de acordo com as recomendações dos fabricantes.

As reações de sequenciamento foram realizadas para ambas as fitas de DNA, utilizando o kit *ABI Prism<sup>®</sup> BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing* (*Applied*

*Biosystems*), com os mesmos iniciadores usados na amplificação, sendo submetidas no termociclador Veriti® 96 Well (*Applied Biosystems*) à 96 °C durante 10 segundos, 50 °C por 5 segundos e 60 °C a 4 minutos, sendo estas condições repetidas em 30 ciclos. Os produtos da reação de sequenciamento foram purificados, seguindo o protocolo de precipitação com isopropanol para remoção dos dideoxi-nucleotídeos excedentes. Os produtos purificados foram ressuspensos em formamida *Hi-Di™* (*Applied Biosystems*) e analisados em um sequenciador automático de 48 capilares *ABI 3730* (*Applied Biosystems*), pertencente à Plataforma Genômica - Sequenciamento de DNA/ PDTIS - FIOCRUZ.

#### 4.4. Amplificação pela PCR, clonagem e sequenciamento do ITS-2

A amplificação do ITS-2 (de aproximadamente 720-pb) foi feita com o uso dos iniciadores 5.8T, 5' – CTGAGCGGCGGATCACTCGG e 28T, 5' – GCACTATC AAGCAACACGACTC (Marcilla et al., 2001). As reações foram feitas em um termociclador Veriti® 96 Well (*Applied Biosystems*), programado para a seguinte ciclagem: desnaturação a 94 °C (5 minutos), seguido de 40 ciclos de 94 °C (45 segundos), 60 °C (1 minuto) e 72 °C (1 minuto e meio), e uma extensão final de 15 minutos a 72 °C.

A purificação dos produtos amplificados para a etapa de clonagem foi feita utilizando o kit *GFX™ PCR DNA and gel band purification* (*GE Healthcare*), de acordo com as recomendações dos fabricantes para a purificação direta das reações da PCR. Os DNAs purificados foram submetidos à reação de ligação do vetor à seleção das colônias, foram feitas de acordo com o sugerido pelo fabricante, exceto pelo fato de utilizar bactérias *Escherichia coli* quimio-competentes (cepa DH5α) e meio LB.

Após o crescimento das bactérias (indicado pela turvação do meio), seus DNAs foram extraídos (*mini-prep*) por lise alcalina. Para avaliar a extração do DNA plasmidial, as amostras foram submetidas à eletroforese em gel de agarose a 0,8%. A enzima *EcoRI* (*Promega*) foi utilizada para a digestão do DNA extraído, para confirmar que o inserto (ITS-2) estava presente. O DNA foi incubado com a enzima por duas horas, a 37 °C e posteriormente submetidas à eletroforese em gel de agarose a 0,8%.

Os plasmídeos que continham o inserto foram purificados com polietilenoglicol 8000 (PEG 8000) e submetidas à reação de sequenciamento (nas mesmas condições apresentadas na seção 4.3 para mtCytb). As amostras foram analisadas no sequenciador automático de 48 capilares *ABI 3730 (Applied Biosystems)*, pertencente à Plataforma Genômica - Sequenciamento de DNA/ PDTIS – FIOCRUZ.

#### 4.5. Análise de microssatélites

As amostras de *R. robustus s.l.* coletadas em simpatria e molecularmente identificadas com base no mtCytb como pertencentes a linhagens distintas de *R. robustus s.l.* foram analisadas para 10 *loci* de microssatélites. Estes *loci* foram previamente caracterizados em *R. prolixus*, sendo cinco (R4, R8, R13, R26 e R31) selecionados dos 10 *loci* descritos por Harry e cols. (2008b) e outros cinco (List14-010, List14-013, List14-021, List14-025 e List14-064) dos 10 descritos por Fitzpatrick e cols. (2009). Os alelos de microssatélites foram amplificados pela PCR no termociclador Veriti<sup>®</sup> 96 Well (*Applied Biosystems*) nas seguintes condições: 95 °C por 1 minuto, seguida de 40 ciclos a 94 °C, por 30 segundos, temperatura de anelamento *T<sub>a</sub>* por 30 segundos (Tabela 4.4), e 72°C por 45 segundos, e ainda uma extensão final de 10 minutos a 72°C. Os fluoróforos de 2<sup>a</sup> geração 6-FAM e HEX (*Applied Biosystems*) foram utilizados para marcação dos iniciadores senso (Tabela 4.4).

Todos os produtos de PCR foram analisados a partir da comparação com os tamanhos padrão de GS500 LIZ (*Applied Biosystems*) de 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490 e 500-pb, em sequenciador ABI 3730xl Genetic Analyser (*Applied Biosystems*) da Plataforma de Análise de Fragmentos / PDTIS – FIOCRUZ. As análises dos fragmentos e identificação dos alelos foram efetuadas no programa PEAK SCANNER<sup>™</sup> v.2.0 (*Applied Biosystems*), após a remoção dos picos gerados pelo anelamento dos iniciadores.

**Tabela 4.4.** Características gerais dos 10 *loci* analisados.

Ta – temperatura de anelamento; F – fita senso; R – fita anti-senso; pb – pares de base; A – adenina; C – citosina; G – guanina; T – timina; N – qualquer um dos quatro nucleotídeos.

| <i>Locus</i> | Iniciadores (5' - 3')                                   | Tamanhos Esperados* (pb) | Tamanhos obtidos‡ (pb) | Repetição   | Ta    | Fluoróforo | Autores                   |
|--------------|---|--------------------------|------------------------|---|-------|------------|---------------------------|
| R4           | F: AAGTGGTTAAAATGAAAATATTCC<br>R: CCGGTAAGACGCAGAGTAC   | 252-258                  | 240-266                | (GT) <sub>2</sub> T <sub>2</sub> (GT) <sub>17</sub>                   | 50 °C | FAM        | Harry et al. (2008)       |
| R8           | F: ATGGCAACTTTAATTTCAAGTATTCC<br>R: TCTGACGAAACGCCACTG  | 203-205                  | 200-202                | (GT) <sub>8</sub>   | 50 °C | HEX        | Harry et al. (2008)       |
| R13          | F: TACATCTTCAATAATCATCACACAC<br>R: AAGTAAATTGAATGAATGCC | 190-192                  | 188-190                | (AC) <sub>5</sub> GC(AC) <sub>1</sub>                                 | 48 °C | FAM        | Harry et al. (2008)       |
| R26          | F: AGAAGGAATCTATCCACTTTTCGC<br>R: CCTCGCTATCAGCTGCTACG  | 112-120                  | 112-118                | (CA) <sub>3</sub> CT(CA) <sub>2</sub>                                 | 54 °C | HEX        | Harry et al. (2008)       |
| R31          | F: TGTGGTAAGTCCTGTGTAGAAGG<br>R: TCTGTTGGTCCAGACACGG    | 129-133                  | 120-126                | (GT) <sub>11</sub>  | 52 °C | FAM        | Harry et al. (2008)       |
| List14-010   | F: AATGATGACTGTATTGATGGGC<br>R: TTCGACCAACAACAATTCCC    | 311-339                  | 307-317                | (CA) <sub>9</sub>   | 52 °C | FAM        | Fitzpatrick et al. (2009) |
| List14-013   | F: CATACTACACGCACACAAGACC<br>R: ATACTCGCATCAAAGCCATTGG  | 335-345                  | 333-337                | (AC) <sub>10</sub>  | 55 °C | FAM        | Fitzpatrick et al. (2009) |
| List14-021   | F: AACCTCTGAACACATCAAATGG<br>R: AGCTACCTCTTGCCTCTACG    | 291-299                  | 280-307                | (TG) <sub>10</sub>  | 55 °C | HEX        | Fitzpatrick et al. (2009) |
| List14-025   | F: CCGCTCTATCAACTACTCC<br>R: GATCCCTTATGTTTCTCAGC       | 163-181                  | 151-173                | (TC) <sub>9</sub> (AC) <sub>7</sub> N <sub>13</sub> (AC) <sub>7</sub> | 50 °C | HEX        | Fitzpatrick et al. (2009) |
| List14-064   | F: AGAAAATGAGCAAAACGGCC<br>R: ACAGGCAAACTATGACG         | 237-247                  | 243-260                | (GT) <sub>10</sub>  | 57 °C | HEX        | Fitzpatrick et al. (2009) |

\* Variação do tamanhos dos alelos encontrada nos trabalhos publicados (Fitzpatrick et al. 2009; Harry et al. 2008).

‡ Variação do tamanhos dos alelos encontrada neste estudo.

O programa MICRO-CHECKER foi utilizado para a possível detecção em cada população analisada dos *loci* que apresentaram abandono de alelos (do inglês *allele drop out*) e alelos nulos (Van Oosterhout et al., 2004). Os equilíbrios de ligação (LE) foram testados entre todos os pares de *loci* de cada ponto de coleta utilizando os procedimentos implementados no ARLEQUIN v. 3.5 (Excoffier e Lischer, 2010).

Desvios do equilíbrio de Hardy–Weinberg (H-W) foram calculados em cada população no programa GENALEX v. 6.5 (Peakall e Smouse, 2012), para cada *locus* analisado. Deficiências significativas de heterozigotos, corrigidas para testes múltiplos pelo algoritmo de FDR (Benjamini e Yekutieli, 2001), indicaram possíveis misturas populacionais (efeito Wahlund), alta taxa de endocruzamentos, ou a presença de híbridos naturais. Estes resultados serviram de sinalizadores para análises subsequentes.

A frequência alélica dos *loci* e o número de alelos privados das amostras de cada local de coleta foram estimados tal como implementado no programa GENALEX v. 6.5 (Peakall e Smouse, 2012). O teste do Chi-Quadrado aplicado às análises de Shannon foi feito no mesmo programa. Com base neste teste, é possível saber se as diferenças entre as frequências alélicas observadas em cada *locus* das populações analisadas são estatisticamente significantes.

O coeficiente de endogamia ( $F_{IS}$ ) e a diferenciação genética entre populações (usando  $F_{ST}$  e  $R_{ST}$ ) foram examinados usando o programa ARLEQUIN v. 3.5 (Excoffier e Lischer, 2010). A análise de componentes principais (PCA) foi realizada para a definição preliminar dos agrupamentos populacionais mais prováveis foi feito no programa GENALEX v. 6.5 (Peakall e Smouse, 2012).

A estruturação das populações pertencentes a linhagens distintas de *R. robustus s.l.* coletadas em simpatria foi estimada no programa STRUCTURE v. 2.3 (Hubisz et al., 2009; Pritchard et al., 2000), que utiliza um método bayesiano de agrupamento. Nesta análise, define-se o número mais provável de populações presentes na amostragem, baseando-se na premissa de equilíbrio de H-W e LE dentro das populações. Desta forma, os indivíduos genotipados são atribuídos a uma determinada população de maneira a minimizar o desequilíbrio de ligação e de H-W em cada agrupamento. A vantagem de se utilizar esta análise é a de não definir *a priori* as populações, levando em consideração apenas a localidade de coleta da amostra.

O modelo de ancestralidade de mistura (*admixture model*), no qual não é considerada nenhuma informação geográfica prévia à respeito das populações, foi assumido como sendo verdadeiro. O modelo de frequência dos alelos utilizado foi o correlacionado. Neste modelo é assumido que nenhuma população contém genótipos muito divergentes.

Foram estabelecidas  $1,8 \times 10^6$  cadeias Markov Monte Carlo (MCMC), com  $1,0 \times 10^5$  iterações e o descarte das primeiras 300.000 cadeias. As cadeias MCMC foram checadas visualmente quanto à convergência e estabilização de todos os parâmetros calculados. Vinte rodadas independentes foram executadas para cada número de populações *a posteriori* (ou K). Ao todo, foram testados de dois a seis agrupamentos.

Os valores de verossimilhança e a probabilidade dos dados serem verdadeiros para cada número de K foram calculados e posteriormente utilizados em uma análise estatística *ad hoc*, chamada de delta K ( $\Delta K$ ), que leva em consideração o erro-padrão nos dados entre sucessivos valores de K (neste caso, de 2 a 6), para definir o número de populações K mais provável (Evanno et al., 2005).

#### 4.6. Isolamento de novos marcadores moleculares no genoma de *R. prolixus*

##### 4.6.1. Desenvolvimento de um método diagnóstico para *R. prolixus* e *R. robustus* s.l.

Durante a primeira etapa do projeto genoma, chamada de *genome survey sequencing* (GSS), foram desenvolvidos 96 pares de iniciadores, desenhados com base em 10.000 sequências aleatórias do genoma de *R. prolixus*, que amplificassem 400-500-pb de regiões não-codificantes. Parâmetros como a reprodutibilidade, variabilidade dos amplicons e a presença de regiões espécie-específicas entre os táxons analisados foram levados em consideração para a seleção preliminar de sete marcadores com o potencial diagnóstico para *R. prolixus*.

Maiores detalhes acerca da geração das sequências durante o GSS e amplificação dos marcadores, bem como o método de busca utilizado para localizar o marcador diagnóstico no genoma, são apresentados na seção de metodologia do artigo em anexo a esta tese (Anexo 1).

##### 4.6.2. Identificação dos genes ortólogos do relógio circadiano de *R. prolixus*

A busca pelos genes ortólogos relacionados ao ritmo circadiano *per*, *tim*, *cry1* e *cry2* de insetos foi realizada no banco de dados UNIPROT (Apweiler et al., 2004), utilizando como palavras-chave o “*nome do gene*” e “*Insecta*”. A inclusão na análise preliminar das sequências de aminoácidos contidas nesse banco de dados foi feita com base em três critérios previamente estabelecidos: (1) utilização de apenas sequências completas da proteína; (2) obtenção de sequências manualmente anotadas e revisadas, contidas no banco de dados de conhecimento da estrutura de proteínas Swiss-Prot (European Bioinformatics Institute<sup>®</sup>); (3) escolha de sequências provenientes de Ordens de insetos distintas, de maneira a representar da melhor forma a variabilidade da sequência de cada proteína em *Insecta*. Na ausência de sequências de organismos proximamente relacionados à *Rhodnius prolixus* que satisfizessem tais critérios, proteínas não revisadas e automaticamente anotadas foram obtidas no banco de dados TrEMBL (European Bioinformatics Institute<sup>®</sup>) e comparadas por similaridade de aminoácidos com as demais proteínas curadas. Neste caso, foram comparadas as

regiões de domínio das proteínas, por serem as regiões mais conservadas evolutivamente.

As sequências obtidas foram alinhadas utilizando o programa MUSCLE (Edgar, 2004), implementado no MEGA 5 (Tamura et al., 2011). As penalidades de abertura de *gaps* foram estabelecidas em -4,5 (a fim de facilitar a abertura de *gaps*), de extensão de *gaps* em zero e o multiplicador de hidrofobicidade em 2,5 (maior agrupamento de proteínas hidrofóbicas e hidrofílicas com as suas semelhantes). O método de agrupamento utilizado foi o UPGMB, com tamanho mínimo da diagonal ( $\lambda$ ) igual a 30. Cada alinhamento foi revisado manualmente em uma etapa posterior ao alinhamento automático.

As regiões conservadas entre as sequências de espécies de insetos foram utilizadas como base para a busca subsequente dos genes homólogos de interesse nos *contigs* do genoma de *R. prolixus*. As proteínas completas dos genes relacionados ao controle dos ritmos circadianos de *D. melanogaster* foram usadas de maneira alternativa como base para as buscas, uma vez que este foi o primeiro inseto a ter os genes de interesse caracterizados. O mesmo foi feito com as sequências dos genes homólogos de organismos mais proximamente relacionados a *R. prolixus*. Estas abordagens foram utilizadas com o objetivo de confirmar a localização das regiões-alvo no genoma de *R. prolixus*.

Os genes de interesse foram procurados na base de dados de transcritos preditos no genoma de *R. prolixus* (RproC1, linhagem CDC) com o uso do TBLASTN (Altschul et al., 1997), implementado na página *web VectorBase* (<http://rprolixus.vectorbase.org/Tools/BLAST/>). Este método compara peptídeos inseridos pelo usuário com os nucleotídeos traduzidos do genoma de *R. prolixus*. Os critérios estabelecidos na busca foram de tamanho de palavra = 3 e *e-value* = 1.

A proteína codificante predita que apresentou os maiores escores obtidos no alinhamento (Altschul et al., 2009) e os menores *e-values* foi escolhida como sendo a representante do gene homólogo correspondente. Foi considerada como homóloga a proteína que obtivesse *e-value* menor que  $10^{-30}$  e escores maiores que 100.

Os genes, as regiões transcritas e os peptídeos pertencentes à predição C1 (predição gênica *VectorBase pipeline* usando o agrupamento v3.0), denominada

“predição *VectorBase*”, foram obtidas na página *web VectorBase* (<http://vectorbase-cap.ensemblgenomes.org/>). A predição alternativa, chamada de “predição *Lagerblad 3.0*” foi feita no programa GENEID v1.3 (Blanco et al., 2007), treinado com uma base de dados de ESTs de *R. prolixus* no agrupamento v3.0 do genoma de *R. prolixus* (disponível na página *web VectorBase*). As estruturas de éxons e íntrons preditas segundo os dois modelos foram comparadas visualmente. As predições dos genes homólogos de *R. prolixus* foram posteriormente curadas pelo colaborador Dr. Jose Manuel Latorre Estivalis, do Centro de Pesquisas René Rachou, Fiocruz.

#### 4.6.3. Análise filogenética dos genes homólogos de *per*, *tim* e *cry*

O relacionamento filogenético entre as sequências de cada um dos genes homólogos analisados (de *R. prolixus* e dos outros insetos) foi inferido no programa MEGA 5 (Tamura et al., 2011), utilizando o método de agrupamentos vizinhos (*Neighbor-Joining*, ou NJ) (Saitou e Nei, 1987). O método de substituição empregado foi o Modelo de distribuição de Poisson (Nei, 1987). A confiabilidade dos clados foi inferida a partir de 1000 replicações de *bootstrap*.

#### 4.6.4. Escolha das regiões gênicas dentro de cada gene do relógio circadiano a serem utilizadas como novos marcadores moleculares aplicados à sistemática de triatomíneos

A escolha das regiões gênicas com aplicabilidade potencial em estudos de sistemática molecular foi estabelecida com base em três critérios: (1) seleção de íntrons ao invés de éxons, por serem as regiões de maior variabilidade nos genes, que sejam flanqueados por éxons igualmente preditos em ambos os modelos “*VectorBase*” e “*Lagerblad*” (i.e. início e fim dos éxons não conflitantes nos dois modelos), uma vez que a técnica escolhida para a amplificação dos fragmentos do gene foi a de anelamento dos iniciadores senso e anti-senso nos éxons para a cobertura de um íntron, ou EPIC (*Exon-Primed Intron-Crossing*); (2) escolha de íntrons com mais de 200 e menos de 1300-pb, uma vez que estes são os limites respectivos inferior e superior de nucleotídeos analisados com acurácia em um sequenciador automático de 48 capilares *ABI 3730* (*Applied Biosystems*). A utilização do primeiro critério é relevante para a posterior aplicação destes marcadores em estudos de sistemática

molecular envolvendo outras espécies da subfamília Triatominae, uma vez que os iniciadores estão localizados em regiões (potencialmente) mais conservadas do gene.

O desenvolvimento dos iniciadores foi feito no programa PRIMER-BLAST, disponível na página *web* do NCBI (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). A sequência utilizada como molde para o desenvolvimento do par de iniciadores de cada marcador continha os 100-250 nucleotídeos finais e iniciais dos éxons localizados nas regiões 5' e 3' do íntron, respectivamente, além da sequência do íntron flanqueado. Foram estabelecidos como parâmetros para o desenvolvimento dos iniciadores: (1) a região de anelamento dos iniciadores deveria amplificar, além do íntron completo, 5 a 150-pb de cada éxon; (2) o tamanho do produto da PCR deveria ser maior que o tamanho do íntron e menor que 1300-pb; (3) a temperatura de anelamento (*T<sub>a</sub>*) mínima deveria ser 52 °C, ótima de 60 °C e máxima de 65 °C, com diferença entre as *T<sub>a</sub>* dos iniciadores senso e anti-senso de até 2 °C; e (4) o tamanho dos iniciadores não poderia ser inferior a 18 e superior a 25 nucleotídeos.

#### 4.6.5. Amplificação dos marcadores pela PCR e sequenciamento

Amostras de *R. prolixus* (provenientes da Guatemala) e *R. robustus* II (de Loreto, Perú) foram submetidas a diferentes condições nas reações de PCR para a amplificação das regiões de interesse. A quantidade de reagentes para cada reação foi ajustada após os seguintes testes: (1) variação na quantidade de DNA (de 10 a 80ng); (2) variação na quantidade de cloreto de magnésio [ $MgCl_2$  é o co-fator para o bom funcionamento da enzima *Taq Platinum*<sup>®</sup> *DNA polimerase* (Invitrogen) – 0,5; 0,75; 1,0; 1,25; e 1,5mM]; (3) variação na quantidade dos iniciadores senso e anti-senso (5; 10; 12,5; 15; e 20mM); (4) variação na quantidade da enzima polimerase (de 1 a 6 unidades). As amostras foram submetidas a gradientes de temperatura de anelamento (*T<sub>a</sub>*), de 47 °C a 67 °C (com variações a cada 2 °C), a diferentes tempos nas fases de anelamento dos iniciadores (de 30 segundos a 2 minutos, com variações a cada 30 segundos) e de extensão (de 1 a 3 minutos, com variação de 30 segundos). Todas as reações de PCR foram feitas em um termociclador Veriti<sup>®</sup> 96 Well (*Applied Biosystems*), programado para 40 ciclagens.

A avaliação da correta amplificação dos fragmentos foi feita após a submissão das amostras à eletroforese em gel de agarose 2% e verificação visual quanto ao

aparecimento de uma banda com o tamanho esperado para o fragmento gênico isolado e a presença/ausência de bandas inespecíficas (i.e. amplificação de fragmentos com tamanhos diferentes do esperado). Como os íntrons podem apresentar polimorfismo de tamanho, foi considerado um desvio-padrão de 300-pb no tamanho do produto amplificado. As amostras que tiveram apenas o fragmento de interesse amplificado foram purificadas e diretamente sequenciadas, assim como mencionado nas seções 4.3 e 4.5.

Como um dos objetivos desta tese era a de utilizar estes marcadores em um estudo estritamente filogenético, as informações obtidas pelo sequenciamento direto são válidas, desde que os sítios variáveis entre as fitas de DNA sejam poucos (inferiores a 10% do total de pares de base). Este nível de variação foi checado através da análise de qualidade dos eletroferogramas gerados após o sequenciamento.

Após o ajuste das condições da PCR, foram selecionadas para amplificação e sequenciamento direto dos novos marcadores amostras coletadas de *R. robustus* amazônicos, além de amostras de *R. prolixus* cedidas pelos nossos colaboradores (Tabela 4.3). Foram descartadas das análises filogenéticas as sequências que apresentaram grande quantidade de picos duplos (em mais de 10% da sequência) nos eletroferogramas. De maneira semelhante, os marcadores que apresentaram um baixo sucesso na amplificação pela PCR ou baixa resolução no sequenciamento direto foram descartados das análises subsequentes.

#### 4.7. Edição e alinhamento das sequências

A edição das sequências das fitas senso e anti-senso de cada amostra e a geração da sequência consenso de cada indivíduo para cada marcador molecular foi feita no programa SEQMAN LASERGENE v.7 (DNASTar, Inc.). Todos os sítios ambíguos foram verificados, para que fossem corrigidos os possíveis erros de leitura durante o sequenciamento. As sequências dos iniciadores de mtCytb e AmpG foram removidas, assim como os fragmentos dos éxons que flanqueiam o ITS-2 e os demais marcadores nucleares sequenciados.

As sequências do mtCytb foram alinhadas manualmente, por se tratar de uma região codificante, e portanto, de fácil avaliação das relações de homologia entre os

sítios. Os alinhamentos preliminares das sequências de ITS-2, *per*, *tim* e *cry* foram feitos utilizando o programa MUSCLE (Edgar, 2004), implementado no MEGA 5 (Tamura et al., 2011). As condições utilizadas nesta análise foram idênticas às apresentadas na seção 4.7.2. Cada alinhamento foi revisado manualmente em uma etapa posterior ao alinhamento automático.

#### 4.8. Análises filogenéticas e de distância genética intra e interespecíficas

##### 4.8.1. *mtCytb*

As distâncias genéticas intra e interespecíficas foram calculadas de acordo com a distância de 2-parâmetros de Kimura (K2-p) (Kimura, 1980). De maneira complementar, as sequências foram analisadas utilizando o método de co-variância na Análise de Componentes Principais (PCA), presente no programa GENALEX v. 6.5 (Peakall e Smouse, 2012), para verificar o relacionamento entre as espécies. A técnica multivariada da PCA permite encontrar e ilustrar graficamente os padrões principais de variação nucleotídica nos dados analisados. Este é um processo matemático complexo no qual os eixos principais de variação são ortogonais e localizados dentro dos dados de maneira multidimensional. Cada eixo sucessivo do gráfico explica proporcionalmente menos do total da variação, isto é, o primeiro eixo contém a maior informação acerca da variabilidade entre os dados, seguido pelo segundo eixo e assim sucessivamente. Na maioria das vezes, dois ou três eixos são suficientes para revelar grande parte da variação entre os indivíduos. Apenas os sítios variáveis são incluídos na análise e a identificação taxonômica *a priori* dos espécimes foi feita com base na topologia da árvore de *mtCytb* obtida.

O modelo de substituição nucleotídica adotado nas reconstruções filogenéticas das sequências de *mtCytb* foi selecionado de acordo com os Critérios de Informação Akaike e Bayesiano (AIC e BIC, respectivamente), implementados no programa jMODELTEST v. 2 (Darriba et al., 2012). Os cálculos de verossimilhança incluíram sete esquemas de substituição com frequências de base iguais ou desiguais (+F), proporção de sítios invariáveis (+I) e presença/ausência de distribuição gama (+ $\Gamma$ ), totalizando 56 modelos testados.

As reconstruções filogenéticas das sequências de mtCytb de *R. prolixus* e *R. robustus s.l.* foram feitas utilizando os métodos de máxima-verossimilhança (ML), implementado no programa MEGA5 (Tamura et al., 2011), e Inferência Bayesiana (BI), no programa BEAST v.1.7 (Drummond et al., 2012). A confiabilidade dos clados obtidos pelo primeiro método foi estimada a partir de 1000 replicações de *bootstrap*. Maiores detalhes acerca do segundo método são apresentados nas seções 4.8.3 e 4.8.5 e 4.9.1.

O conceito filogenético de espécies (Mishler e Brandon, 1987) foi aplicado à identificação dos espécimes sequenciados. Indivíduos agrupados em clados monofiléticos foram considerados com sendo pertencentes a uma mesma espécie.

#### 4.8.2. ITS-2

Os cálculos de distância genética entre as sequências de ITS-2 foram feitos de acordo com Mas-Coma e Bargues (2009), que levaram em consideração os valores obtidos a partir da análise de (1) porcentagem de substituições nucleotídicas e inserções e deleções (*indels*) não resultantes de mini e microsatélites; (2) porcentagem de substituições nucleotídicas; e (3) porcentagem de substituições nucleotídicas e todos os *indels* obtidos.

As reconstruções filogenéticas de ITS-2 de *R. prolixus* e *R. robustus s.l.* foram realizadas segundo os métodos de ML e BI. Para a análise do primeiro modelo, as regiões de sequências que continham *indels* foram excluídas (opção de deleção completa). Para fins de comparação, foram adicionadas sequências de *R. prolixus*, *R. robustus* I-IV e *R. robustus* 6.4 (*sensu* Justi 2009), além de *R. nasutus* como grupo externo (estas sequências serão chamadas de sequências-padrão). A confiabilidade dos clados obtidos pelos dois métodos foi estimada a partir de 1000 replicações de *bootstrap* (ML) e probabilidade posterior (BI).

O modelo de substituição nucleotídica adotado nas reconstruções filogenéticas das sequências de ITS-2 foi selecionado de acordo com os mesmos critérios adotados na seção anterior (4.9.1).

Levando em consideração as múltiplas cópias de ITS-2 presentes em cada indivíduo, foi realizada uma terceira reconstrução filogenética utilizando o método de

ML. Nesta reconstrução, foi levada em consideração a proporção da divergência entre sequências intraespecíficas de ITS-2 e mtCytb esperada em espécies de triatomíneos pertencentes a um complexo de espécies. Ao comparar as sequências de ITS-2 de indivíduos pertencentes ao grupo de espécies “1A” do complexo *Triatoma dimidiata* (Bargues et al., 2008), provenientes do sul do México, América Central e Colômbia, foi encontrada uma divergência intraespecífica de 1,7%. Monteiro e cols. (dados não publicados), ao comparar sequências do mtCytb de indivíduos provenientes destas mesmas localidades, encontraram uma porcentagem média de divergência intraespecífica de 4,5%, ou seja, este gene parece evoluir 2,65 vezes mais rápido que ITS-2. Portanto, é de se esperar que as sequências de ITS-2 de uma mesma linhagem de *R. robustus s.l.* não ultrapasassem 0,57% de divergência, quando comparadas com as sequências-padrão [levando em consideração o valor limite de 1,5% de divergência entre sequências do mtCytb de uma mesma espécie de *Rhodnius* (Monteiro et al. 2003)]. As sequências que apresentaram porcentagens de divergência intraespecíficas superiores a este valor em comparações com as sequências-modelo foram excluídas da análise filogenética.

Esta abordagem de estabelecer um valor máximo de divergência intraespecífica para as sequências de ITS-2 de *Rhodnius* teve os seguintes objetivos:

- (1) minimizar a possibilidade de incluir na análise genótipos que podem refletir mais o fluxo gênico entre as linhagens ancestrais de *R. robustus s.l.* amazônicas do que o fluxo gênico recente (i.e. retenção de polimorfismo ancestral). Este fenômeno pode ser visível em marcadores de cópia múltipla quando são analisadas espécies proximamente relacionadas que divergiram recentemente. Como as cópias de ITS-2 de *Rhodnius* não são completamente homogeneizadas pela evolução em concerto (Lazoski, Pavan e Monteiro, dados não publicados), apenas uma parte destas cópias deve representar a homogeneização das características genéticas da população local (Mas-Coma e Bargues, 2009).
- (2) evitar a seleção de cópias paralogas de ITS-2 ou pseudogenes. Quando presentes, parálogos e pseudogenes podem gerar interpretações filogenéticas errôneas (Buckler et al., 1997; Harris e Crandall, 2000). Como estes eventos são pouco comuns, é de se esperar que estas cópias estejam em menor frequência no indivíduo.

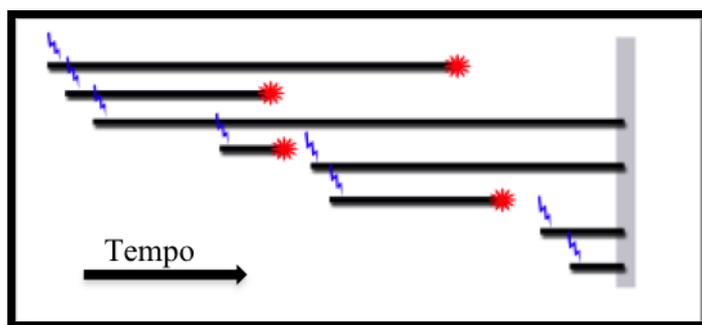
Posteriormente, as sequências intraespecíficas que divergiram em mais 0,6% das sequências-modelo foram inseridas uma a uma nas análises filogenéticas, com o objetivo de compreender o efeito de sua inclusão na topologia resultante.

Um problema encontrado nas análises filogenéticas de sequências de ITS-2 se refere ao tratamento analítico a ser dado para os *indels*. A técnica mais utilizada para a análise dos *indels* é a de substituir os *gaps* de alinhamento por nucleotídeos fictícios (Bargues et al., 2006). Contudo, estas regiões evoluem de maneira independente dos demais sítios nucleotídicos do alinhamento, já que no processo evolutivo relacionado aos *indels* mais de um nucleotídeo pode ser inserido ou deletado em um único evento (Warnow, 2012). Mesmo que um *indel* contendo mais de um nucleotídeo (e os *gaps* de alinhamento correspondentes) seja substituído por um nucleotídeo fictício, os modelos de substituição nucleotídica não refletem o modo de evolução das inserções e deleções no genoma.

Os modelos evolutivos de *indels* atualmente mais aceitos envolvem princípios de parcimônia e processos estocásticos de nascimento, imigração e morte (Nicholls e Gray, 2008). O princípio de parcimônia de Wagner (Kluge e Farris, 1969) assume que os estados de um caráter são reversíveis e com mesma probabilidade de surgimento e perda. Já o modelo Dollo de parcimônia (Farris, 1977) assume que a transição de presença para ausência tem uma probabilidade associada de ocorrência negligível. Neste contexto, é atribuída ao ancestral a ausência de um caráter (estado zero) – este estado pode sofrer apenas uma única transição, de ausência para presença (zero para um). Uma vez que o caráter ancestral é modificado, não é permitida a transição de presença para ausência (Farris, 1977).

A aplicação dos princípios de Wagner e Dollo aos processos estocásticos depende da natureza do caráter evolutivo do *indel*, i.e., se deve ser tratado como sendo um evento raro (Dollo) ou extremamente volátil (Wagner). A mudança dos estados dos *indels* (inserção para deleção, ou vice-versa) pode ser modelada por transições das cadeias de Markov de tempo-contínuo (Suchard et al., 2001), para encontrar o estado ancestral de cada caráter em uma matriz binária (0 e 1). Se, por exemplo, quatro sequências em um total de quatro analisadas apresentam uma inserção (caráter 1), o ancestral terá esta inserção (1), mas se duas apresentarem este caráter (1) e outras duas não (0), o caráter ancestral pode ser 0 ou 1 (Figura 4.4). De maneira

simplificada, podemos dizer que o modelo de Dollo seria empregado no primeiro exemplo e o de Wagner no segundo. Esta chance das mudanças de caráter ocorrer com frequência (Wagner) ou raramente (Dollo) são medidas segundo o modelo estocástico Dollo de multi-estado (Alekseyenko et al., 2008). A partir desta análise, é possível criar inferências filogenéticas com base nas informações genéticas contidas nos *indels*.



**Figura 4.4.** Ganho e perda de caracteres desde o nascimento.

O surgimento de um novo caráter é representado pelos raios azuis. Esses caracteres sobrevivem (linhas pretas) até um evento de morte (estruturas vermelhas). As observações são feitas em um dado ponto no tempo presente (região cinza). Modificado de Alekseyenko et al. (2008).

As reconstruções filogenéticas das sequências de ITS-2 foram feitas utilizando o método de BI, implementado no programa BEAST v.1.7 (Drummond et al., 2012). Todas as cópias de ITS-2 de cada indivíduo sequenciado (1 a 14 cópias por indivíduo) foram adicionadas nesta análise. Para fins de comparação, foram adicionadas sequências de *R. prolixus*, *R. robustus* I-IV e *R. robustus* 6.4 (sensu Justi, 2009) (Tabela 4.3).

Como no programa BEAST v.1.7 todos os modelos *a priori* relacionados ao tratamento dos nós das árvores se baseiam em coalescência (aplicáveis a populações que representem uma única espécie) ou especiação (cujos nós das árvores representam eventos de especiação), em cada uma das sequências foi atribuído um caráter discreto *a priori* – a identificação das sequências pelo nome das espécies. Para esta designação foi utilizada a identificação molecular obtida a partir do sequenciamento do mtCytb. A utilização deste caráter é relevante para a diferenciação durante a análise dos nós das árvores que representam coalescência daqueles que representam especiação. Esta

abordagem é feita no pacote *StarBeast*, ou \*BEAST (Heled e Drummond, 2010), implementada no programa BEAST v.1.7 (Drummond et al., 2012).

O modelo de especiação assumido *a priori* para os nós das árvores foi o Yule (Hardin, 1971) e a árvore filogenética *a priori* foi gerada de maneira aleatória. As duas partições de sequências de ITS-2 foram concatenadas em \*BEAST, tanto as sequências sem os *indels* quanto a matriz binária que representa os *indels*. Nesta matriz binária foram atribuídos os valores “1” para cada nucleotídeo presente em uma inserção e “0” para cada lacuna (*gap*) de alinhamento.

Dez análises preliminares contendo  $10^7$  gerações de cadeias MCMC foram realizadas para observação da convergência dos dados e ajuste fino dos operadores. Posteriormente, duas análises independentes foram efetuadas nas MCMC por  $1,5 \times 10^8$  gerações, com um descarte das primeiras  $1,5 \times 10^7$  gerações. A convergência dos parâmetros e a mistura apropriada das cadeias de MCMC foram confirmadas pelo cálculo do tamanho amostral efetivo (ESS) no programa TRACER v. 1.5 (Drummond e Rambaut, 2007), excluindo 10% das cadeias iniciais de cada análise. Todos os parâmetros acima de 200 foram considerados como calculados apropriadamente (sensu Drummond et al., 2006). As análises foram combinadas no programa LOGCOMBINER, e as árvores de máxima credibilidade determinada com base nas 300.000 árvores geradas (descarte das 60.000 árvores iniciais), com limite de probabilidade posterior de 0.6, no programa TREE ANNOTATOR (ambos os programas são parte do pacote BEAST). A confiabilidade dos clados obtidos pelo foi estimada a partir do método de probabilidade posterior.

#### 4.8.3. Árvore filogenética de espécies com base nas sequências de *mtCytb* e *ITS-2*

Existem muitas fontes potenciais de discrepância entre árvores de genes e árvores de espécies, em função de fenômenos biológicos como transferência horizontal, separação incompleta das linhagens (do inglês *incomplete lineage sorting* ou *ILS*) e duplicação/extinção gênica (Heled e Drummond, 2010). Estas fontes de inconsistência foram alvos de pesquisas envolvendo a concepção de inferências estatísticas para a geração da correta reconstrução filogenética de espécies (revisado em Huang e Knowles, 2009). De maneira simplificada, no pacote \*BEAST as árvores filogenéticas de cada gene são concatenadas em uma única árvore de espécies a partir

da contabilização dos valores de verossimilhança de cada alinhamento de sequências, dada a árvore de um gene, e dos tamanhos populacionais efetivos atuais e ancestrais de cada uma das linhagens (Heled e Drummond, 2010). O número de espécies presentes na árvore filogenética de espécies é definido *a priori* e as árvores de genes funcionam como parâmetros de “perturbação” da *priori*, a serem contabilizados na distribuição posterior.

A árvore filogenética das espécies de *R. robustus s.l.* e *R. prolixus* foi reconstruída com base nas sequências de mtCytb e ITS-2 (contendo suas duas partições concatenadas) utilizando o \*BEAST (Heled e Drummond, 2010). O grupo externo utilizado foi composto por *R. pictipes* e *R. pallescens* I e II. O caráter discreto *a priori* de identificação da espécie que cada sequência representa foi feito com base nos resultados obtidos a partir do sequenciamento do mtCytb.

Dez análises preliminares contendo  $10^7$  gerações de cadeias MCMC foram realizadas para a observação da convergência dos dados e ajuste fino dos operadores. Posteriormente, foram feitas duas análises independentes contendo  $2 \times 10^8$  gerações de cadeias MCMC em cada, com um descarte das primeiras  $2 \times 10^7$  gerações. A convergência dos parâmetros e a mistura apropriada das cadeias MCMC foram confirmadas da mesma maneira apresentada na seção anterior, assim como a combinação das análises. A árvore de máxima credibilidade foi determinada com base nas 400.000 árvores geradas (descarte das 40.000 árvores iniciais), com limite de probabilidade posterior de 0.6, no programa TREE ANNOTATOR (ambos os programas são parte do pacote BEAST). A confiabilidade dos cladogramas obtidos foi estimada a partir do método de probabilidade posterior.

#### 4.8.4. Reconstrução filogenética das sequências dos marcadores relacionados ao ritmo circadiano

As reconstruções filogenéticas de cada um dos novos marcadores foram feitas utilizando os métodos de ML e BI. O primeiro método foi acessado no programa MEGA5 (Tamura et al., 2011) e o segundo no programa BEAST 1.7 (Drummond et al., 2012). Na reconstrução filogenética segundo o primeiro método, as regiões das sequências que continham *indels* foram excluídas (opção de deleção completa), assim

como realizado nas análises do ITS-2. A confiabilidade dos clados obtidos pelo primeiro método foi estimada a partir de 1000 replicações de *bootstrap*.

Os modelos de substituição nucleotídica adotados nas reconstruções filogenéticas das sequências dos ortólogos de *per* e *tim* foram selecionados de acordo com o Critérios de Informação Bayesiano (BIC) e na reconstrução filogenética das sequências do homólogo de *cry* foi selecionado de acordo com o Critérios de Informação Akaike, corrigido para amostragens menores que 30 sequências contendo menos de 1000-pb em cada ( $AIC_C$ ).

As inferências bayesianas foram feitas exatamente iguais aquelas apresentadas na seção 4.8.2, inclusive quanto ao tratamento dos *indels*. Contudo, o número de cadeias de MCMC geradas em cada uma das duas análises foi de  $4 \times 10^8$ , com descarte de  $4 \times 10^7$  e o total de 800.000 árvores analisadas foi analisado (descarte das primeiras 160.000). Os marcadores pertencentes a um mesmo gene foram concatenados, por evoluírem numa mesma taxa.

#### 4.8.5. *Árvore filogenética de espécies com base nas sequências de mtCytb, ITS-2 e dos marcadores relacionados ao ritmo circadiano*

A inferência bayesiana para a busca da árvore de espécies foi feita de maneira idêntica aquela apresentada na seção 4.9.3. Contudo, o número de cadeias de MCMC geradas em cada uma das duas análises foi de  $4 \times 10^8$ , com descarte de  $4 \times 10^7$  e o total 800.000 árvores foi analisado (descarte das primeiras 160.000).

O objetivo desta análise é o de avaliar o quanto cada um destes marcadores analisados influenciam e contribuem para a determinação dos clados filogenéticos das espécies de *R. prolixus* e *R. robustus s.l.*, comparando estes resultados com aqueles obtidos a partir da análise das árvores filogenéticas de mtCytb e mtCytb + ITS-2.

#### 4.9. Análises filogeográficas e biogeográficas

##### 4.9.1. Calibrações das datações dos eventos cladogenéticos

Para as análises filogeográficas e o teste de hipóteses biogeográficas foram utilizadas as sequências obtidas do mtCytb de *R. prolixus* e *R. robustus s.l.* O grupo externo foi composto por três sequências de *R. pictipes* e *R. pallescens* I e II (Tabela 4.3).

O método do relógio molecular aleatório (do inglês *Random Local Clock*, ou *RLC*) foi empregado com o objetivo de avaliar se a taxa de mutação entre as linhagens estudadas era a mesma ou não, i.e., se os dados se encaixavam nas premissas da hipótese do relógio molecular ‘fixo’ ou do relógio molecular ‘relaxado’. Este método foi realizado no \*BEAST (Heled e Drummond, 2010). Foram geradas  $2 \times 10^8$  cadeias de MCMC, com descarte das  $2 \times 10^7$  cadeias iniciais. A hipótese aceita para o conjunto de dados foi a de que as linhagens apresentam taxas evolutivas distintas (cf. seção 5.6.1 dos Resultados para maiores detalhes).

Para os cálculos de divergência entre as linhagens de *R. prolixus* e *R. robustus s.l.* e entre as espécies de *R. robustus s.l.* foram empregadas diferentes calibrações dos eventos cladogenéticos resultantes da especiação das linhagens representadas na topologia da árvore. A hipótese de separação de *R. pictipes* e *R. prolixus* entre 51,5 e 44,2 Maa, calculada por Gaunt e Miles (2002), foi levada em consideração nas análises. O período geológico de maior elevação dos Andes entre 12 e 8 Maa (cf. Hoorn et al., 2010) foi assumido como sendo responsável pela cladogênese de *R. pictipes* e *R. pallescens* I e II. Essa premissa teve como base estudos anteriores acerca da ecologia e filogeografia das espécies (Abad-Franch et al., 2009; Pavan, 2009). Foi assumido ainda que a última grande elevação dos Andes teria sido responsável pela separação de *R. pallescens* I e II, entre 2,5 e 1,5 Maa (Abad-Franch et al., 2009). Todos estes valores das “calibrações suaves” foram delineados de forma a assumirem uma distribuição normal, por se tratarem de eventos geológicos ou de valores obtidos empiricamente. As influências das três “calibrações suaves” impostas *a priori* nas datações posteriores foram analisadas a partir de quatro abordagens distintas (Tabela 4.5):

- (1) Calibração ‘frouxa’ de um nó da árvore – datação do evento de separação entre *R. pictipes* e *R. pallescens* I e II.

- (2) Calibração ‘estrita’ de um nó da árvore – datação do evento de separação entre *R. pictipes* e *R. pallescens* I e II.
- (3) Calibração ‘estrita’ de dois nós da árvore – (a) datação dos eventos de separação entre *R. pictipes* e *R. pallescens* I e II, e entre *R. pallescens* I e II; e (b) datação das separações entre *R. pictipes* e *R. pallescens* I e II e ainda entre *R. prolixus* e *R. pictipes*.
- (4) Calibração ‘estrita’ de três nós da árvore – datação dos eventos de separação entre *R. pictipes* e *R. pallescens* I e II, entre *R. prolixus* e *R. pictipes* e entre *R. pallescens* I e II.

As calibrações ‘frouxa’ e ‘estrita’ se referem ao tamanho do intervalo de confiança assumido para as datações dos eventos geológicos (maior ou menor, respectivamente). Com base na informação prévia a respeito das taxas de mutação encontradas em insetos do gênero *Triatoma* [ $5,5 \times 10^{-3}$  a  $9,0 \times 10^{-3}$  (Pfeiler et al., 2006)] e da Classe Insecta [ $1,15 \times 10^{-2}$  (Brower, 1994)], em todas as análises foi imposta *a priori* uma distribuição normal logarítmica com média -4,4 e desvio-padrão de 0,9 da taxa de mutação (variação da taxa de mutação entre  $2,1 \times 10^{-3}$  e  $7,16 \times 10^{-2}$ , com média de  $1,2 \times 10^{-2}$ ). Para cada uma das análises de calibração das árvores filogenéticas, foram realizadas três corridas independentes, das quais  $2 \times 10^8$  cadeias de MCMC foram analisadas (descarte das primeiras  $2 \times 10^7$  cadeias). A convergência dos parâmetros e a mistura apropriada foram confirmadas da mesma maneira apresentada na seção 4.8.2, assim como a combinação das análises. A árvore de máxima credibilidade foi determinada com base nas 120.000 árvores geradas (descarte das 24.000 árvores iniciais), com limite de probabilidade posterior de 0.6, no programa TREE ANNOTATOR (ambos os programas são parte do pacote BEAST).

**Tabela 4.5.** Tipos de calibração utilizados nos nós das árvores filogenéticas.

Estas calibrações foram utilizadas no modelo de ‘relógio molecular relaxado’, e os respectivos valores *a priori* das médias, medianas, desvios-padrão e alcances obtidos segundo distribuições normais dos valores. **Maa** – Milhões de anos atrás.

| Calibração  | Evidência da calibração*                     | Média e Mediana (Maa)** | Desvio-padrão (Maa) | Alcance (Maa)† |
|---|--|-------------------------|---------------------|----------------|
| Separação entre <i>R. pictipes</i> de <i>R. pallescens s.l.</i> (‘frouxa’)  | Geológica - orogênese dos Andes              | 13,0                    | 2,40                | 17,7 - 8,3     |
| Separação entre <i>R. pictipes</i> de <i>R. pallescens s.l.</i> (‘estrita’) | Geológica - orogênese dos Andes              | 10,0                    | 1,05                | 12,1 - 7,9     |
| Separação entre <i>R. pallescens</i> I e II (‘estrita’)                     | Geológica - última grande elevação dos Andes | 2,5                     | 0,55                | 3,6 - 1,4      |
| Separação entre <i>R. prolixus</i> e <i>R. pictipes</i> (‘estrita’)         | Empírica - Gaunt e Miles (2002)              | 48,0                    | 2,00                | 51,9 - 44,1    |

\* A evidência da separação das espécies por conta do evento vicariante de orogênese dos Andes teve como base os trabalhos publicados de Abad-Franch et al. (2009), Hoorn et al. (2010) e Pavan (2009). A evidência da separação das espécies por conta do evento vicariante da última grande elevação dos Andes teve como base os resultados obtidos por Abad-Franch et al. (2003, 2009) e Ribas et al. (2003).

\*\* A média e mediana são iguais, por se tratar de distribuições normais.

† Levando em consideração os valores contidos no intervalo de confiança de 2,5 a 97,5% da distribuição normal.

#### 4.9.2. Testes das hipóteses biogeográficas de “incurções marinhas” e “refúgios pleistocênicos”

Os tempos estimados a partir do relógio molecular ‘relaxado’ serviram de base para a determinação do tempo geológico de separação das espécies. Adicionados a estes resultados, análises acerca da genealogia e estrutura populacional de cada uma das espécies foram consideradas para a comparação com as predições das hipóteses biogeográficas de ‘incurções marinhas’ e ‘refúgios pleistocênicos’. Nas análises descritas nesta e na seção seguinte, *R. prolixus* e cada uma das espécies de *R. robustus s.l.* foram consideradas como sendo representantes de populações de uma única linhagem ancestral. A idéia central da análise biogeográfica é a de analisar retrospectivamente a informação genética contida nas “folhas” das árvores (indivíduos sequenciados) para compreender como os “ramos” (clados) se bifurcaram. As sequências dos indivíduos de uma mesma espécie contêm uma série de sítios conservados que representa a sequência do ancestral comum mais recente. Portanto, em uma análise retrospectiva, é possível tratar as sequências atuais de uma mesma espécie (mesmo aquelas provenientes de localidades distintas) como pertencentes a uma única população ancestral.

O polimorfismo intraespecífico das sequências foi analisado pelo cálculo dos índices  $\pi$  (diversidade nucleotídica) (Nei e Li, 1979) e Hd (diversidade haplotípica) (Nei, 1987), utilizando o programa DNASP v.5 (Librado e Rozas, 2009).

Redes genealógicas foram construídas para *R. prolixus* e cada uma das linhagens de *R. robustus* utilizando o programa Network v. 4.6 (Fluxus Technology<sup>©</sup> 2012). A partir dessa abordagem é possível observar com mais clareza o relacionamento entre os haplótipos de uma mesma linhagem. O método de *median-joining* (Bandelt et al., 1999) com um processo posterior de máxima-parcimônia foi utilizado para a geração das genealogias. Neste método é gerada uma árvore contendo o menor somatório do comprimento de todos os braços necessários para a conexão de todos os haplótipos (*minimum-spanning*). Posteriormente vetores intermediários (haplótipos não amostrados ou extintos) são adicionados segundo os critérios de máxima-parcimônia.

A consistência das redes foi testada a partir da modificação das taxas de transição e transversão (1:1 ou 1:3), do peso dos caracteres de acordo com suas posições no códon (peso de 1:1:1 ou 2:3:1 para primeira, segunda e terceira posições do códon, respectivamente) e dos parâmetros épsilon (foram testados os valores 0, 10, 30 e 50). Os valores de épsilon conferem um peso as medidas de distância genética. Quanto maior o valor deste parâmetro, maior o peso de cada substituição nucleotídica.

Para inferir na rede a localização do haplótipo mais semelhante ao do ancestral, uma sequência representante de uma linhagem proximamente relacionada foi adicionada como grupo externo em cada uma das redes genealógicas construídas. A partir da inclusão desta sequência, espera-se que o haplótipo que apresentar menor distância em relação ao grupo externo represente o haplótipo ancestral.

O teste de neutralidade  $F_S$  de (Fu, 1997) foi realizado utilizando o programa Arlequin v. 3.5 (Excoffier e Lischer, 2010) com o objetivo de saber se alguma população ancestral (representadas por cada uma das espécies analisadas) sofreu uma expansão populacional recente. Neste teste, resultados de  $F_S$  menores que -2,0 ( $p < 0,02$ ) indicam com um nível de significância de 5% que existe um excesso de alelos raros na população, provavelmente por conta de uma expansão populacional súbita.

Para testar se cada uma das espécies amazônicas de *R. robustus s.l.* sofreu rápida e recente expansão populacional, foram feitas análises de *Bayesian Skyride* utilizando o programa BEAST 1.7 (Drummond et al., 2012). Nesta análise, admitiu-se que a taxa de mutação não varia entre as sequências (modelo de ‘relógio molecular fixo’), uma vez que os indivíduos analisados pertencem a uma mesma espécie e, portanto, devem apresentar o tempo de geração, tamanhos populacionais, taxa metabólica e eficácia na reparação do DNA semelhantes. Com base na informação prévia publicada na literatura (Brower, 1994; Pfeiler et al., 2006), a taxa de substituição fixa foi delineada entre  $5,5 \times 10^{-3}$  e  $1,15 \times 10^{-2}$  por milhão de anos. O tamanho populacional inicial definido *a priori* para todas as espécies foi de 10 a 1000 indivíduos (média = 100). Esse intervalo dilatado reflete a falta de informação prévia acerca desta variável. Dois modelos de crescimento populacional foram testados: constante (a população cresce em uma taxa constante) e variável (a população pode apresentar, em função do tempo, variações na tendência da curva de crescimento). Duas análises independentes para cada um dos dois modelos foram realizadas a partir de  $1,5 \times 10^8$  gerações de cadeias de MCMC (descarte das  $1,5 \times 10^7$  cadeias iniciais). A escolha do modelo que melhor reflete o crescimento populacional de cada uma das espécies foi feita a partir do Fator de Bayes (do inglês *Bayes Factor*, ou BF), que é calculado pela equação:

$$\mathbf{BF}_{1,2} = \mathbf{mL}(\mathbf{modelo1}) - \mathbf{mL}(\mathbf{modelo2}), \text{ onde}$$

**mL** = valor de verossimilhança marginal

Segundo Kass e Raftery (1995), os valores de BF entre 0 e 1 são considerados como evidência pouco apoiada; se  $1 < \text{BF} < 3$ , há uma evidência positiva a favor do modelo 1; se  $\text{BF} > 3$ , há uma forte evidência a favor do modelo 1; e se  $\text{BF} > 5$ , uma fortíssima evidência a favor do modelo 1. Os valores de verossimilhança marginais foram obtidos por amostragem direcional (do inglês *path sampling*, ou PS) e amostragem em *stepping-stone* (SS) (Baele et al., 2012). Para estes cálculos foram amostradas  $1 \times 10^5$  cadeias de MCMC e os valores armazenados a cada  $1 \times 10^3$ . Estes métodos de amostragem garantem a obtenção de valores de verossimilhança mais

acurados do que aqueles calculados nas análises de *Bayesian Skyride* (Baele et al., 2012).

#### 4.9.3. Teste de hipóteses quanto à origem e dispersão da linhagem ancestral de *R. prolixus* e *R. robustus* s.l.

As análises de coalescência entre *R. prolixus* e as linhagens de *R. robustus* s.l. foram realizadas utilizando inferências bayesianas de genealogias de fluxo gênico para a geração de hipóteses acerca da origem e dispersão das espécies na região pan-amazônica, utilizando o programa MIGRATE-N v. 3.3 (Beerli e Palczewski, 2010). Nesta análise, *R. prolixus* e *R. robustus* s.l. foram consideradas como populações distintas de uma única espécie (i.e., populações da linhagem ancestral). Os testes de hipóteses biogeográficas foram feitos de acordo com duas escalas geográficas distintas. Em uma macroescala, foram consideradas nas análises as sequências provenientes da região do Orinoco e da Amazônia, para a determinação da origem das espécies na região pan-amazônica. Em uma mesoescala, foram utilizadas as sequências das espécies amazônicas de *R. robustus* s.l. para determinar como as populações ancestrais teriam se dispersado nesta região e originado o padrão de espécies observado hoje em dia. Cada uma das sequências utilizadas foi agrupada de acordo com a área biogeográfica (sensu Morrone, 2006) de ocorrência da população ancestral de origem (Tabela 4.3). No caso de uma espécie ocorrer em mais de uma área biogeográfica, os resultados das genealogias das espécies atuais foram levados em consideração para determinar sua possível área biogeográfica de origem (ao observar a localidade geográfica de ocorrência do haplótipo apontado como mais semelhante ao haplótipo ancestral). Vinte sequências provenientes de cada área biogeográfica foram utilizadas nas análises em macroescala e 10 sequências de cada linhagem amazônica de *R. robustus* s.l. para as análises em mesoescala, em função da limitação de tempo computacional. Esta quantidade de sequências parece ser suficiente para atingir a convergência das cadeias de MCMC e calcular corretamente os parâmetros de interesse (Beerli, comunicação pessoal). As sequências representantes de cada espécie foram selecionadas aleatoriamente. Quando nas análises em macroescala duas ou mais espécies atuais eram provenientes de uma mesma área biogeográfica, foram selecionadas quantidades iguais de sequências para cada espécie, de maneira a somar 20 indivíduos.

As hipóteses foram construídas de acordo com quatro modelos de dispersão: migração completa, migração por *stepping-stones*, migração por *source-sink* e isolamento completo. A premissa do modelo de migração completa é a de que todas os agrupamentos apresentam migração bidirecional, mas de alguma forma a deriva genética permanece atuante em cada grupo. No modelo de *stepping-stones* é estabelecida migração bidirecional entre agrupamentos geograficamente adjacentes. O modelo de *source-sink* representa a hipótese de fluxo gênico direcional, ocorrendo de uma população fonte para uma região adjacente, cuja sustentabilidade é dependente de um constante provimento de novos imigrantes. Finalmente no modelo de isolamento completo é estabelecida a hipótese de que a linhagem ancestral ocupava todas as regiões analisadas e, com o surgimento de barreiras geográficas naturais, as populações se isolaram e eventualmente especiam em alopatria.

Todos os eventos cladogenéticos estimados segundo o modelo de “relógio molecular relaxado” para as sequências de mtCytb que resultaram em especiação foram datados em um período mais antigo do que duas vezes o tamanho efetivo da população. Portanto, os resultados obtidos refletem o compartilhamento de polimorfismo ancestral entre as espécies e não o fluxo gênico atual (Beerli, 2006). O programa MIGRATE-N retorna estimativas da distribuição posterior dos valores de migração entre as populações, escalonado pela taxa de mutação ( $M$ ), e tamanho populacional efetivo ( $\theta$ ), assim como os valores de verossimilhança marginais para cada hipótese gerada (Beerli, 2006). Como as análises envolvem eventos cladogenéticos antigos,  $M$  e  $\theta$  foram apenas utilizados para a conferência da convergência dos valores estimados nas MCMC.

O modelo de substituição assumido no programa MIGRATE-N é o F84 (Felsenstein, 1981), com ou sem distribuição gama para a variação das taxas de substituição nucleotídica entre os sítios. A estimativa do modelo que melhor se emprega nos dados utilizados, assim como as estimativas de máxima-verossimilhança (MLEs) para as taxas de transição/transversão, frequência nucleotídica e o parâmetro alfa ( $\alpha$ ), foram calculadas no programa PAUP\*, de acordo com o proposto por Gonçalves da Silva e cols. (2012). De maneira resumida, o teste de correlação de verossimilhança foi utilizado para a seleção do modelo de substituição mais adequado para os dados. Posteriormente, uma árvore inicial de NJ foi estimada utilizando o modelo de distância logDet/paralinear (Massingham e Goldman, 2007). Todas as

MLEs foram obtidas e adicionadas ao modelo de substituição selecionado. Estes valores foram utilizados para a busca heurística de árvores filogenéticas com o critério de ML, utilizando a adição aleatória passo-a-passo de amostras e bifurcação-reconexão na árvore. Novas MLEs foram obtidas com base na árvore filogenética gerada. Esse processo foi repetido até as MLEs se estabilizarem entre as gerações de novas árvores. Os valores finais foram então utilizados como parâmetros iniciais no programa MIGRATE-N.

As hipóteses de origem e dispersão da linhagem ancestral de *R. prolixus* e *R. robustus s.l.* foram delineadas em uma macroescala após a seleção de sequências provenientes de três agrupamentos de áreas biogeográficas: (1) região do Orinoco; (2) região do Napo; e (3) Guiana-Inambarí-Imerí-Rondônia-Pará-Belém (Figura 4.5). As áreas foram agrupadas desta forma em função de análises filogeográficas anteriores de PAE e análises cladogenéticas de áreas biogeográficas (análises de UPGMA), utilizando informações acerca da riqueza de unidades taxonômicas operacionais (OTUs) da Sub-Família Triatominae na região pan-amazônica. Abad-Franch e cols. (2009) constataram que a região do Orinoco possivelmente é a região de origem da tribo Rhodniini. Contudo, a descoberta de uma nova linhagem de *R. robustus* filogeneticamente mais basal em relação aos clados *R. prolixus* + *R. robustus* I-IV + *R. neglectus* + *R. nasutus* (cf. seção 5.2.1 dos Resultados) levantou a possibilidade de que a origem da linhagem ancestral do grupo poderia ser em outra área biogeográfica – a de sua ocorrência, no Napo.



**Figura 4.5.** Áreas biogeográficas da América do Sul e América Central.

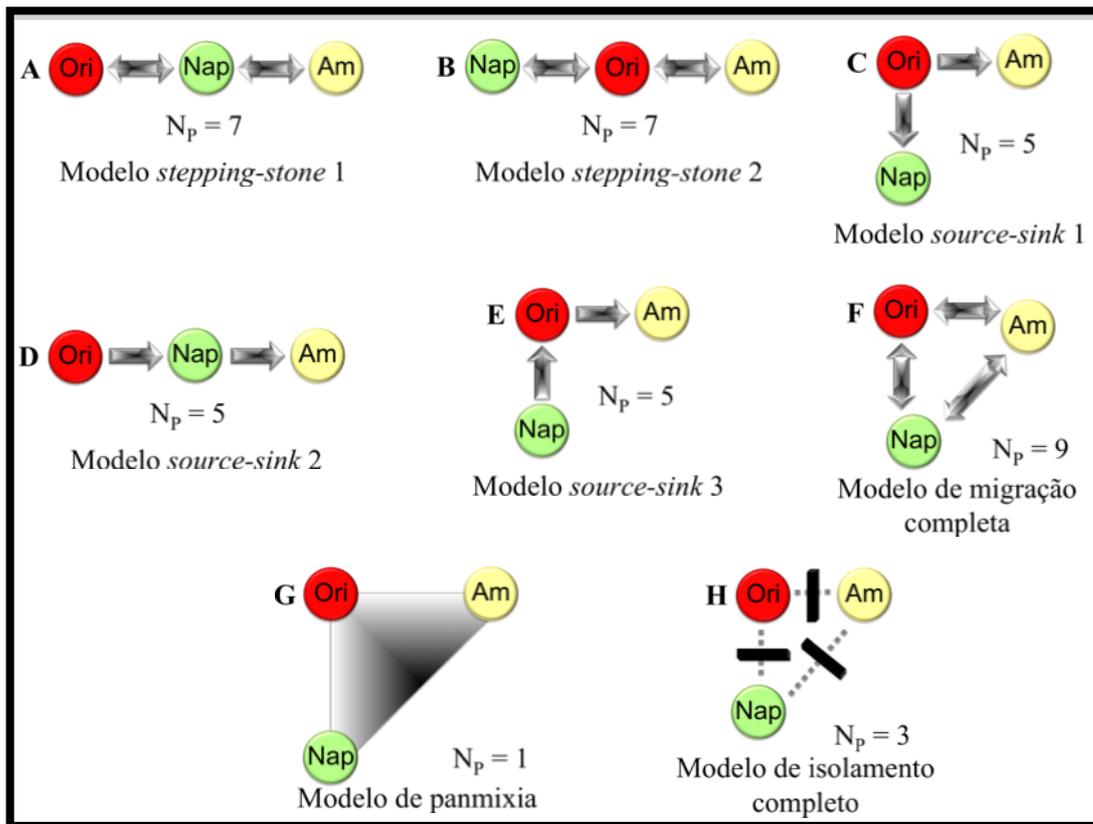
Em destaque de cores, os agrupamentos criados para o teste de hipóteses em macroescala. Cada cor (vermelho, amarelo e verde) representa um agrupamento distinto analisado em uma macroescala. Modificado de Abad-Franch et al. (2009).

Duas hipóteses foram delineadas segundo o modelo *stepping-stones* (Figura 4.6): **(A)** população ancestral do Orinoco teria migrado (ou se expandido) primeiramente para a área biogeográfica do Napo. Uma vez estabelecida esta nova população, um novo evento de migração (ou expansão populacional) teria ocupado as áreas biogeográficas da Guiana, Inambari, Imeri, Rondônia, Pará e Belém; **(B)** a população presente no Napo teria migrado (ou se expandido) para a área do Orinoco. Uma vez estabelecida esta nova população, um novo evento de migração (ou expansão populacional) teria ocupado as áreas biogeográficas de Guiana, Inambari, Imeri, Rondônia, Pará e Belém. Outras três hipóteses foram delineadas segundo o modelo *source-sink* (Figura 4.6): **(C)** parte da população ancestral do Orinoco teria servido de fonte de migrantes para o estabelecimento da população ancestral proveniente do Napo e de Guiana, Inambari, Imeri, Rondônia, Pará e Belém; **(D)** parte da população ancestral do Orinoco teria servido de fonte de migrantes para o estabelecimento da população ancestral proveniente do Napo. Uma vez estabelecida, esta teria sido a fonte de migrantes para as áreas biogeográficas adjacentes de Guiana,

Inambarí, Imerí, Rondônia, Pará e Belém; e **(E)** a população ancestral do Napo teria servido de fonte de migrantes para o estabelecimento da população ancestral do Orinoco. Uma vez estabelecida, esta teria sido a fonte de migrantes para as áreas biogeográficas adjacentes da Guiana, Inambarí, Imerí, Rondônia, Pará e Belém. Além dessas hipóteses, foram testadas ainda outras três que tiveram como base os modelos de **(F)** migração completa (hipótese nula ou  $H_0$ ); **(G)** panmixia; e **(H)** isolamento completo entre as populações. As amostras de *R. robustus* II, inclusive as coletadas da área biogeográfica do Napo, foram consideradas como pertencentes ao terceiro agrupamento (Figura 4.5, em amarelo), uma vez que a ocorrência atual de populações desta espécie nesta área biogeográfica parece ser resultante de uma ocupação secundária (cf. seção 5.6.2 dos Resultados para maiores detalhes).

Apenas as sequências de *R. robustus* II e *R. robustus* IV foram testadas como possíveis representantes da população ancestral que teria originado as espécies amazônicas atuais de *R. robustus* II, III, IV e *n.sp.* Esta imposição se deve ao fato de que estas duas espécies apresentam distribuições geográficas mais adjacentes ao Orinoco e Napo (possíveis áreas biogeográficas de origem de *R. prolixus* e *R. robustus s.l.*) do que as demais espécies (Tabela 4.3). Foram testadas duas hipóteses segundo o modelo *stepping-stone* (Figura 4.7): **(A)** as populações ancestrais da região de ocorrência de *R. robustus* II teriam dispersado (ou expandido) para as regiões de ocorrência de *R. robustus* IV e *R. robustus* III. Posteriormente, as populações ancestrais da região de ocorrência de *R. robustus* III teriam dispersado (ou expandido) para as regiões de ocorrência de *R. robustus n.sp.*; e **(B)** as populações ancestrais da região de ocorrência de *R. robustus* IV teriam dispersado (ou expandido) para as regiões de ocorrência de *R. robustus* II e *R. robustus* III. Posteriormente, as populações ancestrais da região de ocorrência de *R. robustus* III teriam dispersado (ou expandido) para as regiões de ocorrência de *R. robustus n.sp.* Posteriormente, as populações ancestrais da região de ocorrência de *R. robustus* III teriam dispersado (ou expandido) para as regiões de ocorrência de *R. robustus n.sp.* Outras duas hipóteses foram delineadas segundo o modelo *source-sink* (Figura 4.7): **(C)** idêntico a **(A)**, mas com populações-fonte ancestrais na região de ocorrência de *R. robustus* II e posteriormente de *R. robustus* III; e **(D)** idêntico a **(B)**, mas com populações-fonte ancestrais da região de ocorrência de *R. robustus* IV e posteriormente de *R. robustus* III. Foram analisadas ainda duas hipóteses de isolamento completo: **(E)** a linhagem

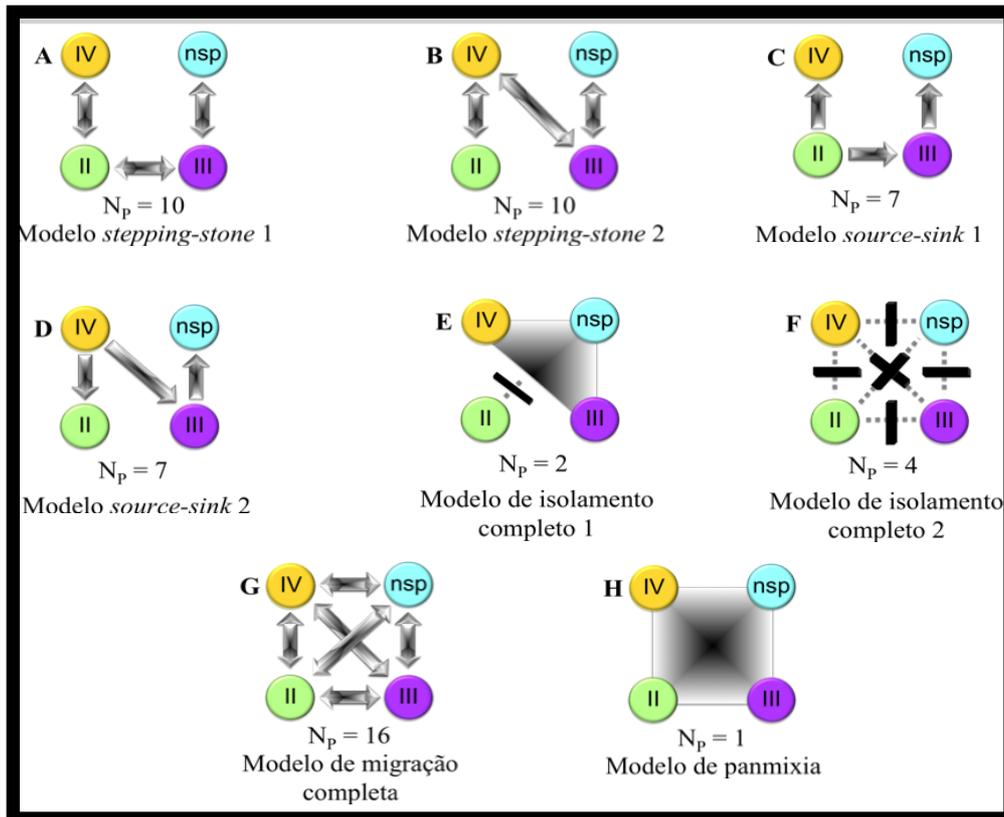
ancestral teria ocupado toda a região amazônica e uma barreira natural ao fluxo gênico teria surgido de maneira a separar *R. robustus* II de *R. robustus* III, IV e *n.sp.*; e (F) a linhagem ancestral teria ocupado toda a região amazônica e barreiras naturais ao fluxo gênico teriam surgido de maneira a separar *R. robustus* II, *R. robustus* III, *R. robustus* IV e *R. robustus n.sp.* Além destas hipóteses, foi testada ainda outra que teve como base o modelo de (G) migração completa (hipótese nula ou  $H_0$ ); e (H) o modelo panmítico.



**Figura 4.6.** Hipóteses biogeográficas de dispersão testadas em uma macroescala. Cada círculo representa um agrupamento testado. As cores dos círculos seguem o mesmo padrão apresentado na Figura 4.5. As setas indicam a direção de fluxo gênico. As linhas pontilhadas interrompidas por uma barra preta significam isolamento entre os agrupamentos. O triângulo representa uma população panmítica. Maiores detalhes acerca das hipóteses são apresentados no texto. **Ori** – Orinoco; **Nap** – Napo; **Am** – agrupamento amazônico contendo representantes das áreas biogeográficas de Guiana, Inambarí, Imerí, Rondônia, Pará e Belém;  $N_p$  = Número de parâmetros calculados em cada hipótese.

As hipóteses testadas em macro e mesoescalas diferiram no número de parâmetros estimados, de 1 a 9 e de 1 a 16, respectivamente (Figuras 4.6 e 4.7), de maneira a permitir análises de panmixia, isolamento, migração assimétrica e migração

completa. Quanto maior o número de parâmetros a serem estimados, maior o número de cadeias de MCMC necessárias para a convergência dos parâmetros estimados. A maneira encontrada para acomodar esses valores é a de utilizar um número de cadeias de MCMC que sejam suficientes para convergir todos os parâmetros calculados no modelo com o maior número de parâmetros (neste caso o de migração completa).



**Figura 4.7.** Hipóteses biogeográficas de dispersão testadas em uma mesoescala.

Cada círculo representa um agrupamento testado. Os círculos seguem o mesmo padrão de cores das distribuições geográficas das espécies, observadas na Figura 4.1A (com a exceção de “nsp”, cuja distribuição não é observada no mapa). As setas indicam a direção de fluxo gênico. As linhas pontilhadas interrompidas por uma barra preta significam isolamento entre os agrupamentos. O triângulo e o quadrado representam populações panmíticas testadas. Maiores detalhes acerca das hipóteses são apresentados no texto. **II** – *R. robustus* II; **III** – *R. robustus* III; **IV** – *R. robustus* IV; **nsp** – *R. robustus* n.sp.;  $N_p$  = Número de parâmetros calculados em cada hipótese.

Para cada um dos modelos de hipóteses, foram analisadas 30 réplicas, nas quais foram realizadas oito cadeias paralelas de temperatura (1,00; 1,17; 1,40; 1,75; 2,33; 3,50; 7,00; e  $1,00 \times 10^6$ ; *Metropolis-coupled* MCMC, ou MCMCMC) (Suchard et al., 2001), com possível troca entre as cadeias a cada 10 passos. Os valores estabelecidos *a priori*, assim como as janelas de amostragem para a determinação dos valores de  $M$  e  $\theta$  foram mantidos conforme o padrão do programa. Em cada réplica

foram gravados  $10^4$  passos das MCMC, em  $10^3$  intervalos, totalizando  $10^7$  passos, com descarte dos  $10^5$  iniciais. A convergência dos parâmetros e a mistura apropriada das cadeias de MCMC foram inspecionadas visualmente (Kuhner, 2009) e confirmadas pelos valores de ESS maiores que  $10^5$  e probabilidade de genealogias aceitas maior do que 10%.

As genealogias de fluxo gênico foram ranqueadas de acordo com os fatores logarítmicos de Bayes (LBF), calculados da seguinte forma:

**LBF =  $\ln(\text{mL}(\text{modelo1})) - \ln(\text{mL}(\text{modelo2}))$** , onde o modelo 2 é o maior valor de máxima-verossimilhança.

Para ranquear os modelos, os valores de LBF devem ser exponenciados. Os modelos que apresentarem os maiores valores serão os mais prováveis de estarem corretos. O cálculo da probabilidade de cada modelo ser o mais correto é feito a partir do valor exponencial de LBF do melhor modelo dividido pela soma dos valores exponenciais de LBF de todos os pares de modelos testados.

#### 4.10. Análise da atividade locomotora de *R. prolixus* e *R. robustus* s.l.

##### 4.10.1. Escolha e preparo das amostras e delineamento experimental

O experimento de atividade locomotora foi feito com 123 *R. prolixus* provenientes da Guatemala e mantidos na colônia do Laboratório de Doenças Parasitárias (IOC), e 67 *R. robustus* II, provenientes de Loreto, Perú, e mantidas na colônia do Laboratório de Referência Nacional e Internacional de Triatomíneos (IOC). Vinte representantes de cada colônia tiveram o mtCytb e o ITS-2 sequenciados, para identificação das espécies e a confirmação da pureza das colônias (i.e., ausência de híbridos). Os espécimes de *R. robustus* s.l. advindos do campo não foram utilizados, pois nenhum sobreviveu em condições de laboratório.

A análise locomotora foi feita no sistema desenvolvido para insetos do gênero *Drosophila*, chamado monitor de atividade (*Drosophila Activity Monitor*, Trikinetics; Waltham, MA, USA). Os monitores são estruturas que contém 32 canais, para o encaixe de tubos de vidro contendo os insetos de estudo. Como as ninfas de *R. prolixus* e *R. robustus* II são maiores do que os adultos de *Drosophila*, foram

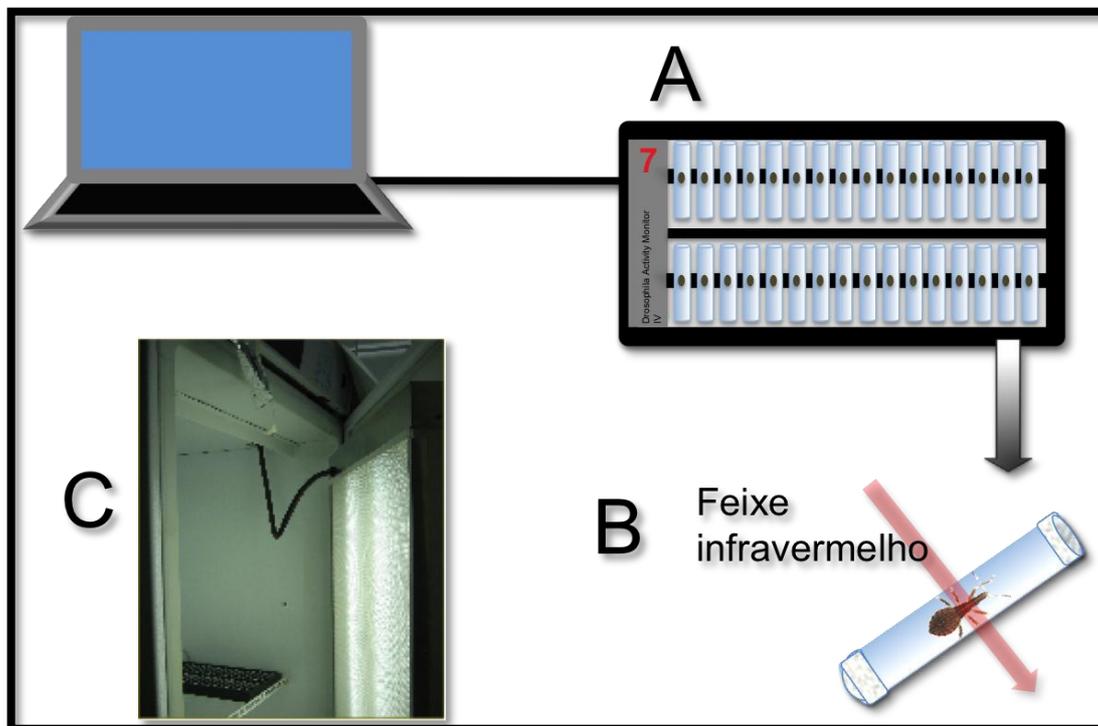
utilizados monitores adaptados para o encaixe de tubos de vidro maiores. Uma ninfa foi colocada em cada tubo medindo sete centímetros de comprimento por um centímetro de diâmetro, cujas extremidades são seladas com *parafilm*.

Os monitores são ligados em um computador para a contabilização dos movimentos de cada inseto. Em cada canal do monitor, um feixe de infravermelho atravessa o tubo de vidro transversalmente. Portanto, a leitura dos movimentos das ninfas é feita de maneira individual. A cada interrupção do feixe infravermelho, decorrente da passagem de uma ninfa pelo meio do tubo, o programa *DAMSystem data collection software* v. 2.0 (Trikinetics Waltham, MA, USA) contabiliza um movimento (Figura 4.8). Estes movimentos foram armazenados em intervalos de cinco e trinta minutos ao longo dos dias de experimento.

Todos os monitores contendo as ninfas foram mantidos em uma incubadora *Precision Scientific* modelo 818 durante o experimento, sob a temperatura constante de 25 °C (Figura 4.8). Uma parte destes monitores recebeu uma cobertura construída com isopor envolvido. O objetivo de utilizar este aparato foi o de criar uma proteção à incidência de luz direta, para mimetizar um abrigo. Esta cobertura forma uma sombra de 2,5 cm em uma extremidade do tubo, deixando livre para incidência de luz na fotofase os 4,5 cm de comprimento restantes do tubo (Figura 4.9). Os insetos foram mantidos em um regime de fotoperíodo de 6 dias com 12 horas de luz e 12 horas de escuridão (LD 12:12), seguido de 20 dias de escuridão completa (DD) e 5 dias em LD 12:12. Foram escolhidos 6 dias de LD para garantir que os insetos sincronizassem o ritmo de atividade de acordo com os indicadores de luminosidade (*entrainment*). A escolha por 20 dias em DD foi feita após alguns testes para avaliar quantos dias seriam necessários para que os picos de atividade voltassem pelo menos uma vez para o ZT determinado em LD. Com isso, é possível calcular o período do relógio circadiano de cada uma das espécies.

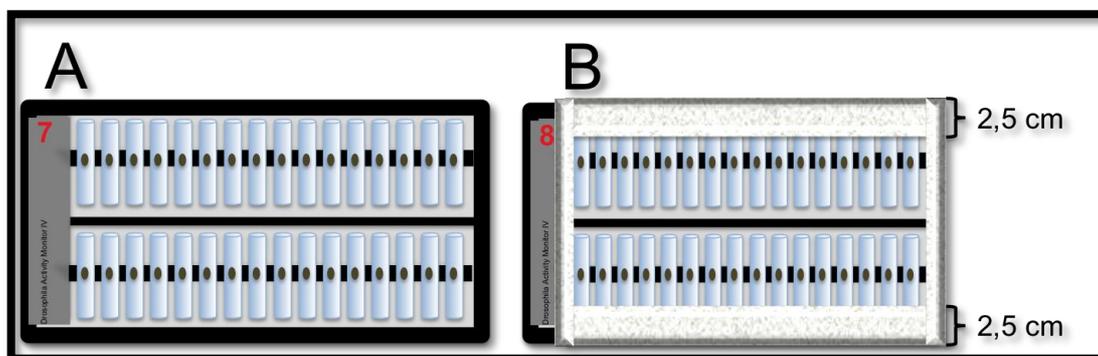
Os experimentos foram delineados de maneira a possibilitar a observação das possíveis diferenças entre as atividades locomotoras de *R. prolixus* e *R. robustus* II quando submetidos à diferentes condições: privação alimentar de dez e cinco dias anteriores ao início do experimento (durante os 31 dias subsequentes todos os espécimes se mantiveram em jejum); influência da presença de um abrigo no

acréscimo ou decréscimo de atividade locomotora; análise de espécimes dos segundo e terceiro estádios (N2 e N3).



**Figura 4.8.** Experimento de atividade locomotora.

(A) Os insetos são colocados individualmente em tubos de vidro. Cada monitor contém 32 canais individuais de monitoramento da atividade locomotora (faixa preta na porção mediana dos tubos) e é ligado a um computador, para contabilidade dos movimentos. (B) Em cada canal é emitido um feixe infravermelho, que atravessa transversalmente a porção mediana do tubo de vidro. A cada interrupção do feixe, contabiliza-se um movimento. (C) Os monitores são colocados em prateleiras de uma incubadora, que produz regimes artificiais diários de luz (localizadas na porta do equipamento) e temperatura (Foto: Carla Gentile).



**Figura 4.9.** Monitores para os experimentos de atividade locomotora.

(A) Monitor sem proteção para a luz. (B) Monitor com o aparato de isopor encaixado para criar uma proteção à incidência de luz direta, mimetizando um abrigo.

#### 4.10.2. Análise dos dados de atividade locomotora

Os dados de atividade locomotora dos espécimes foram analisados de maneira preliminar no programa EXCEL<sup>®</sup> (Microsoft<sup>™</sup>). Os insetos mortos foram contabilizados e suprimidos da análise. O primeiro dia de LD foi tratado como um período de “adaptação” ao regime claro e escuro e igualmente foi excluído da análise.

Os valores brutos das atividades, coletados a cada 30 minutos, foram transformados utilizando a média móvel simples (do inglês *Simple Moving Average*, ou SMA), uma média não ponderada que é calculada da seguinte maneira:

$$\text{SMA}_1 = \frac{Va_1 + Va_2 + Va_3}{3}; \text{SMA}_2 = \frac{Va_2 + Va_3 + Va_4}{3}; \dots \text{SMA}_{N-2} = \frac{Va_{N-2} + Va_{N-1} + Va_N}{3}$$

**Va** = valor absoluto

**N** = número de amostras representantes de um mesmo período de tempo.

Esses valores foram utilizados na confecção dos gráficos de barras de atividade das espécies para visualização do período de maior atividade (diurno ou noturno) e os tempos *zeitgebers* (ZT – tempo do ciclo ambiental medido em horas) com os maiores picos de atividade de *R. prolixus* e *R. robustus* II.

As proporções das atividades diurnas e noturnas de cada indivíduo analisado foram calculadas para uma melhor definição de qual período as espécies apresentaram maior atividade e comparadas a partir do teste t de Student, ao nível de significância de 5%. Como os valores de proporções não são variáveis aleatórias contínuas e, portanto, não apresentam uma distribuição normal, estes devem ser transformados pelo cálculo do arco-seno da raiz quadrada do valor da proporção. Após a transformação, a normalidade dos dados foi avaliada utilizando os testes de Komolgorov-Smirnov (Massey, 1951) e Shapiro-Wilk (Shapiro e Wilk, 1965), disponíveis no programa SPSS v. 17 (IBM<sup>®</sup> 2008, Chicago, EUA). Os dados foram ainda visualmente checados quanto a escala, ao enviesamento e a localização em um histograma e um gráfico de quantis (ou gráfico Q-Q), para comparação dos valores transformados com uma população normal padrão. Os gráficos foram desenhados no programa SPSS v. 17 (IBM<sup>®</sup> 2008, Chicago, EUA).

A comparação dos valores de arco-seno da raiz quadrada da atividade locomotora de ninfas N2 e N3 de *R. prolixus* e *R. robustus* II obtidos sob as diferentes condições de privação alimentar e presença/ausência de abrigo foi feita utilizando a análise de variância de quatro sentidos (*four-way* ANOVA), no programa SPSS v. 17 (IBM<sup>®</sup> 2008, Chicago, EUA), ao nível de significância de 5%.

Os actogramas foram feitos no *plug-in* ACTOGRAMJ (Schmid et al., 2011), disponível no programa FIJI (Schindelin et al., 2012). A função de suavização gaussiana das atividades em função da média com valor sigma igual a 100 foi aplicada para uma melhor visualização dos picos de atividade locomotora. Foram selecionados de 10-22 actogramas para cada espécie e condição testada nas análises de *four-way* ANOVA.

Os períodos dos relógios circadianos de *R. prolixus* e *R. robustus* II foram determinados a partir do atraso ou adiantamento diário dos picos de maior atividade observados nos actogramas durante o regime de DD. Os cálculos foram feitos com base nos 20 dias ou nos 15 dias iniciais de DD. Para esta análise foram selecionados 22 actogramas de *R. prolixus* e 10 de *R. robustus* II que apresentaram as melhores definições dos picos de atividade. O método utilizado foi o de Lomb-Scargle (Lomb, 1976; Scargle, 1982; Townsend, 2010), ao nível de 5% de significância. O período dos relógios circadianos de *R. prolixus* e *R. robustus* II, indiretamente definidos pela atividade locomotora, foram comparados visualmente em gráficos de caixa (*boxplots*), gerados no programa SPSS v. 17 (IBM<sup>®</sup> 2008, Chicago, EUA).

## 5. RESULTADOS

### 5.1. Coleta e identificação morfológica dos triatomíneos

A coleta ao longo do primeiro transecto, definido entre as áreas de distribuição geográfica de *R. robustus* III e IV [de Belém (Pará) até Palmas (Tocantins)], foi realizada no fim do período seco de 2009. Os cinco primeiros pontos de coleta, de um total de 18, foram visitados em uma época na qual não chovia há um mês. Entretanto, do 10° ao 21° (e último) dia de coleta [do ponto de Ulianópolis (Pará) em diante], o ambiente ficou bastante úmido, em decorrência de pancadas de chuva nos fins de tarde [entre às 16 e 17h; pluviosidade mensal = 450 mm (Veronez, 2011)]. Ao todo,

foram obtidos 395 espécimes de triatomíneos, dentre os quais 137 foram identificados por critérios morfológicos (Lent e Wygodzinsky, 1979) como pertencentes ao complexo de linhagens crípticas *R. robustus s.l.* (Tabela 4.1).

A vegetação dos três primeiros pontos localizados no Pará (Belém, Ponta das Pedras e Santa Maria) é característica de floresta pluvial tropical. As palmeiras *Attalea maripa* amostradas estavam localizadas em áreas sem sinais visíveis de desmatamento recente e apresentavam escores de matéria orgânica em decomposição (MOD) de “+” (sensu Abad-Franch et al., 2005). A infestação por triatomíneos das palmeiras foi baixa (com apenas duas palmeiras positivas de um total de 14 amostradas – 14,3%), assim como a densidade de triatomíneos por palmeira (0,43). Em Ponta das Pedras foram coletados cinco *R. robustus s.l.* e em Santa Maria, apenas um (Tabela 4.1). Nesta última localidade, havia uma *Iguana iguana* co-habitando a palmeira de onde foi coletado o espécime de *R. robustus s.l.*

Em Aurora do Pará e Paragominas (Pará) o ambiente silvestre estava completamente modificado, em função de queimadas e desmatamento para criação de monoculturas e cultivos de gado. As poucas palmeiras encontradas apresentavam sinais de queimada (truncos com fuligem) e escore de MOD “1/2+”. Nenhum triatomíneo foi encontrado nas cinco palmeiras *A. maripa* amostradas (Tabela 4.1).

A vegetação dos sete pontos seguintes [Ulianópolis, Marabá, Paraupébas, São Domingos do Araguaia (Pará), Imperatriz (Maranhão), Estrela e São Bento do Tocantins (Tocantins)] pode ser caracterizada por uma transição gradual de floresta pluvial tropical para uma região de cerradão (Cerrado), chamada de floresta pluvial aberta. Nesta floresta, a fitomassa e o fitovolume vão diminuindo gradativamente de densidade, desde o fim do período chuvoso (dezembro a maio) até o fim do período seco (junho a novembro).

Em quase todas as cidades visitadas que apresentavam esse tipo de vegetação foi possível amostrar palmeiras com escore de MOD entre “+” e “+++”, excetuando-se em Imperatriz, onde o padrão era semelhante àquele observado em Aurora do Pará e Paragominas. Em Ulianópolis, São Domingos do Araguaia, Rondon do Pará (Pará) e São Bento do Tocantins (Tocantins) foram amostradas palmeiras do peridomicílio (a menos de 50 m das casas). Nestes pontos, espécimes de *R. robustus s.l.* foram coletados em maior quantidade [além de Marabá (Pará); Tabela 4.1]. Nas demais

localidades foram amostradas palmeiras em áreas de mata secundária ou em bordas de mata primária. A infestação e a densidade de triatomíneos em palmeiras de florestas pluviais abertas foram altas, já que nove das 14 palmeiras analisadas foram positivas (64,3%). Nove triatomíneos foram encontrados, em média, por palmeira amostrada. Ao todo, foram coletados 126 *R. robustus s.l.*, 35 *R. neglectus*, 4 *R. pictipes* e 1 *P. lignarius* (Tabela 4.1).

Um fato interessante foi observado em Rondon do Pará enquanto os ramos da palmeira derrubada estavam sendo retirados – diferentes espécies de triatomíneos notavelmente ocupavam diferentes “estratos” (ou microambientes) da copa. Em um estrato inferior (na base da copa), foram encontrados 4 *R. pictipes*. Cerca de 20 cm acima, 1 *Panstrongylus lignarius* foi coletado. Em seguida (aproximadamente 40 cm acima), 35 *R. neglectus* foram encontrados juntos à um ninho de pássaro aparentemente abandonado. Na porção mais alta da copa da palmeira (cerca de 50 cm acima do ponto anterior) foram encontrados próximos à um porco-espinho (*Coendou prehensilis*) 20 *R. robustus s.l.*

A vegetação dos últimos cinco pontos de coleta [Tocantinópolis, Araguaína, Palmeirante, Presidente Kennedy e Colméia (Tocantins)] é característica de uma transição gradual de floresta pluvial aberta (Tocantinópolis e Araguaína) para região de cerrado (Cerrado). Todas as palmeiras *Attalea speciosa* amostradas apresentaram um escore de MOD entre “+” e “+++”. A maior porcentagem de infestação das palmeiras por triatomíneos e a maior densidade média de triatomíneos por palmeira dos pontos amostrados neste transecto foram observados nessa região (100% e 37,2%, respectivamente). Ao todo, foram coletados em Tocantinópolis, Araguaína, Palmeirante, Presidente Kennedy e Colméia (Tocantins) 217 indivíduos morfológicamente identificados como *R. neglectus* e 1 *R. pictipes*. *R. robustus s.l.* foi encontrada apenas em Araguaína, Tocantins (cinco indivíduos). Em Palmas (Tocantins), não foi possível encontrar uma região adequada ao trabalho de campo (i.e. propriedade privada cujo dono autorizasse a derrubada de palmeiras).

A coleta ao longo do segundo transecto que atravessa áreas de distribuição geográfica de *R. robustus* II, III e IV (de Manaus, Amazonas, até Porto Velho, Rondônia) foi feita no meio/fim do período seco em 2010. Nesta época o ambiente estava muito seco, devido à falta de chuva por aproximadamente três meses. Em

apenas um dos 15 dias de coleta houve precipitação. Cento e três espécimes do gênero *Rhodnius* foram coletados, dos quais 91 foram identificados preliminarmente por critérios morfológicos (Lent e Wygodzinsky, 1979) como sendo *R. robustus s.l.* (Tabela 4.2).

Apesar dos seis primeiros pontos visitados estarem inseridos em uma região de floresta pluvial tropical, a vegetação estava seca e com muitos focos de queimada. As palmeiras amostradas apresentaram escores de MOD entre “+” e “++”. Onze das 46 palmeiras estavam infestadas por triatomíneos, com densidade média de 1,7 triatomíneos por palmeira. Foram coletados nestas localidades 40 *R. robustus s.l.* e 12 *R. pictipes*.

A comunidade localizada entre o Km 450 da BR-319 e a cidade de Humaitá (Amazonas) estão em uma área de transição entre floresta pluvial tropical e floresta pluvial aberta. O escore de MOD das palmeiras amostradas foi de “+”. Nesta região, apenas uma das 11 palmeiras estava infestada, na qual dois *R. robustus s.l.* foram coletados (densidade média de 0,18 triatomíneo por palmeira).

Os últimos dois pontos de coleta se inserem em uma região de floresta pluvial aberta. Foram amostradas nestas localidades apenas palmeiras do peridomicílio (a menos de 50 m das casas), com escore de MOD de “+”. Quatro das 13 palmeiras amostradas estavam infestadas por triatomíneos (30,8%), nas quais foram coletados 50 *R. robustus s.l.* (densidade média de 3,9 triatomíneos por palmeira).

De todos os triatomíneos coletados em ambos os transectos, somente 10 *R. robustus s.l.* provenientes de Paraupébas (Pará) sobreviveram em condições de laboratório. Consequentemente, não foi possível obter amostras suficientes para os experimentos de comportamento, em função do número de indivíduos fundadores da colônia de laboratório ter sido baixo e do tempo de geração da espécie ser relativamente longo (aproximadamente quatro meses).

## 5.2. Filogenia de *R. prolixus* e *R. robustus* s.l. e divergência molecular entre as sequências

### 5.2.1. MtCytb

Duzentos e cinquenta e quatro espécimes de *R. prolixus* e *R. robustus* s.l. tiveram o fragmento de mtCytb sequenciado. Apesar do fragmento alvo conter 682-pb, algumas sequências obtidas ficaram menores em função da baixa qualidade dos eletroferogramas na extremidade inicial (próxima à região de anelamento do iniciador senso). Os 19 sítios iniciais do fragmento foram excluídos do alinhamento final das sequências com o objetivo de padronizar o tamanho do fragmento a ser utilizado nas análises (663-pb).

A análise do fragmento de 663-pb do mtCytb revelou a existência de 47 haplótipos distintos, definidos por 141 sítios variáveis. A partir do padrão das substituições nucleotídicas dos dados (tais como distribuição das substituições ao longo do fragmento, proporção de transições e transversões e de regiões menos variáveis), o programa jModeltest (Darriba et al., 2012) selecionou como mais adequado o modelo de Hasegawa, Kishino e Yano (Hasegawa et al., 1985) com uma proporção de sítios invariáveis e quatro parâmetros da distribuição gama (HKY +I + $\Gamma$ ).

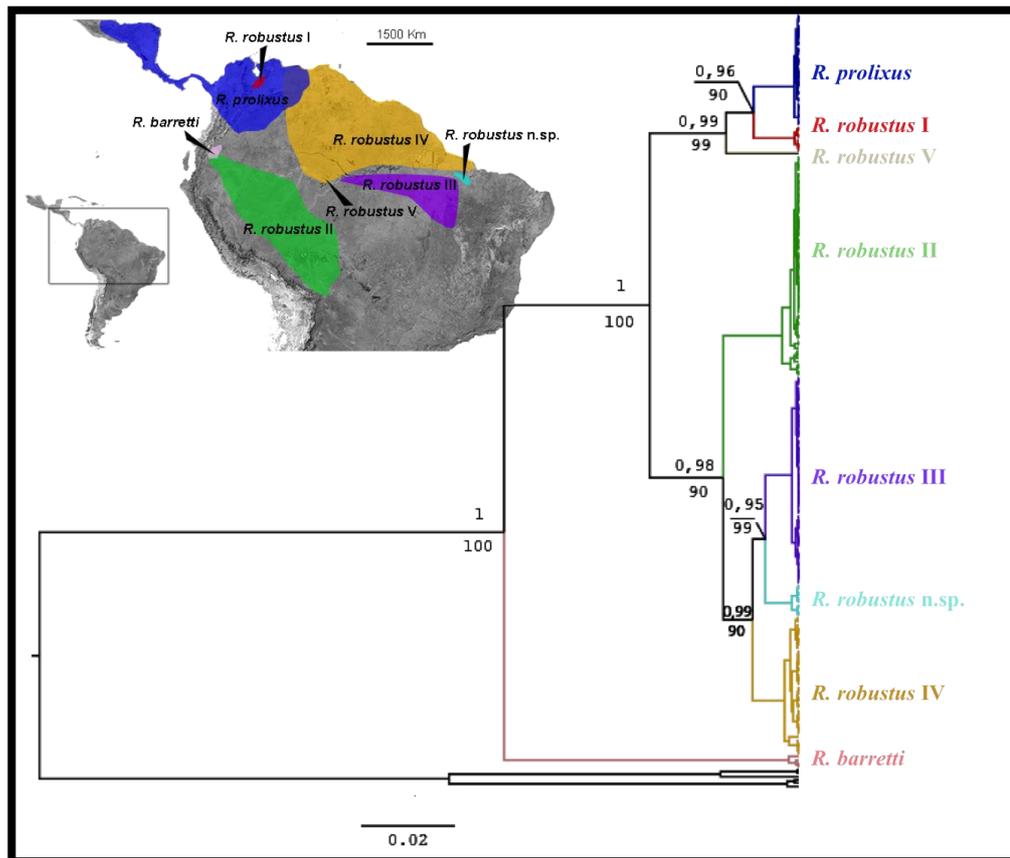
A topologia das árvores geradas a partir das análises de BI e ML foram semelhantes. Praticamente todos os indivíduos amostrados dentro das áreas estimadas de ocorrência de *R. prolixus* e do complexo parafético de espécies crípticas de *R. robustus* (I, II, III e IV) foram genotipados como sendo representantes da espécie correspondente à região. Ao todo, foram obtidas nestas áreas sequências do mtCytb de 49 *R. prolixus*, oito *R. robustus* I, 74 *R. robustus* II, 60 *R. robustus* III e 39 *R. robustus* IV (Tabela 4.3). No nordeste do Equador e no sudoeste da Colômbia, onde se esperava encontrar espécimes de *R. robustus* II, foram descobertas duas novas linhagens.

A primeira linhagem foi evidenciada a partir da análise molecular de um indivíduo proveniente de uma localidade próxima à Manaus, Amazonas. Este indivíduo é filogeneticamente mais próximo de *R. prolixus* e *R. robustus* I do que das linhagens amazônicas de *R. robustus* [*bootstrap* (BS) = 99, probabilidade posterior (PP) = 0,99; Figura 5.1]. Deste ponto em diante esta linhagem será chamada de *R.*

*robustus* V. A segunda linhagem foi evidenciada a partir da análise das sequências de mtCytb de outros dois indivíduos de Sucumbios, Equador, e de dois insetos coletados em Putumayo, Colômbia. Esta linhagem parece formar um grupo basal do complexo de espécies *R. prolixus* e *R. robustus* s.l. (BS = 100; PP = 1,0; Figura 5.1). Adicionalmente, foram observadas diferenças morfológicas entre *R. robustus* e os espécimes desta segunda linhagem (Abad-Franch F, Monteiro FA, Jaramillo N, Pavan MG, Palomeque FS et al. Manuscrito em preparação; Anexo 2). Daqui em diante nesta tese, essa linhagem será chamada de *Rhodnius barretti*.

A investigação das áreas de possível contato entre *R. robustus* III e IV revelou a surpreendente presença de uma terceira linhagem nova, filogeneticamente mais próxima de *R. robustus* III, chamada aqui de *R. robustus* n.sp. (BS = 99; PP = 0,95). Seis indivíduos desta linhagem foram encontrados em uma palmeira amostrada em Ulianópolis (Pará) co-habitada por indivíduos molecularmente identificados como *R. robustus* III. Adicionalmente, as seis amostras coletadas após um surto de infecção oral por *T. cruzi* em uma fazenda no município de São Roberto, Maranhão (ponto fora das áreas de distribuição conhecidas das espécies amazônicas de *R. robustus*) também foram agrupadas neste clado e, portanto, identificadas como *R. robustus* n.sp.

Em relação às amostras coletadas nas áreas de possível contato entre *R. robustus* II, III e IV, no segundo transecto (Anamã e Careiro Castanho, Amazonas), os quatro indivíduos analisados foram agrupados no clado de *R. robustus* II (BS = 90; PP = 0,98). Não foi observado qualquer indício de gradiente genético entre *R. robustus* II, III e IV. As amostras de Careiro Castanho e Anamã (*R. robustus* II) divergem das amostras de *R. robustus* IV encontradas no ponto de coleta adjacente de Manaus, Amazonas [distância geográfica média (dm) = 100 Km] em 2,8-3,4% e de *R. robustus* III de Carajás, Pará, em 3,8-4,1% (dm = 800 Km). A porcentagem de divergência entre essas amostras e aquelas provenientes de Porto Velho, Rondônia, identificadas como *R. robustus* II é de 0,3-0,8% (dm = 700 Km).



**Figura 5.1.** Árvore filogenética do mtCytb.

Árvore consenso de 100.000 árvores construídas pelo método bayesiano com base em uma matriz de distâncias HKY +I + $\Gamma$  para as 254 sequências do mtCytb de *R. prolixus* e *R. robustus s.l.* A confiança dos agrupamentos foi testada a partir do cálculo da probabilidade posterior (mostrados na parte de cima de cada ramo) e 1000 replicações de *bootstrap* (mostrados na parte de baixo de cada ramo), calculados na inferência filogenética feita pelo método de máxima verossimilhança. A barra de escala está representada em substituições. O grupo externo da árvore é formado por duas sequências de *R. pallescens* I, uma sequência de *R. pallescens* II e três sequências de *R. pictipes*. A linhagem descoberta na possível região de encontro entre *R. robustus* III e IV, chamada de *R. robustus n.sp.*, é filogeneticamente mais próxima de *R. robustus* III. A segunda linhagem descoberta, *R. robustus* V, apesar de ter sido coletada na Amazônia brasileira (região biogeográfica do Imeri), está mais próxima das espécies do Orinoco *R. prolixus* e *R. robustus* I do que das demais linhagens de *R. robustus* amazônicas. A terceira nova linhagem apresentada aqui, *R. barretti*, parece formar um clado mais basal em relação ao complexo *R. prolixus* e *R. robustus s.l.*

A porcentagem da divergência genética entre os oito clados da árvore filogenética do mtCytb calculada com base na distância K2-p (em comparações par-a-par) atingiu valores médios superiores a 2%. Portanto, cada um dos oito clados formados por *R. prolixus*, *R. robustus* I, *R. robustus* V, *R. robustus* II, *R. robustus* III, *R. robustus n.sp.*, *R. robustus* IV e *R. barretti* representa uma linhagem distinta (Tabela 5.1). O caráter parafilético do complexo de espécies crípticas de *R. robustus s.l.* é evidenciado pelo fato de *R. robustus* I e *R. robustus* V serem filogeneticamente

mais próximas à *R. prolixus* (porcentagens médias de *K-2p* entre 3,3 e 3,5%), do que às linhagens amazônicas II, III e IV de *R. robustus* (7,1-7,8%).

**Tabela 5.1.** Porcentagem de divergências genéticas (*K-2p*) interespecíficas (abaixo da diagonal fundo branco) e intraespecíficas (na diagonal, fundo cinza) das sequências do mtCyb de *R. prolixus*, *R. robustus s.l.* e *R. barretti* calculadas em comparações par-a-par. Os valores de desvio padrão foram calculados a partir de 1000 iterações de *bootstrap* e são mostrados entre parêntesis.

|                             | 1               | 2               | 3               | 4               | 5               | 6               | 7               | 8              |
|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|----------------|
| 1. <i>R. prolixus</i>       | 0,2<br>(± 0,1)  |                 |                 |                 |                 |                 |                 |                |
| 2. <i>R. robustus</i> I     | 3,3<br>(± 0,7)  | 1,5<br>(± 0,3)  |                 |                 |                 |                 |                 |                |
| 3. <i>R. robustus</i> V     | 3,5<br>(± 0,7)  | 4,5<br>(± 0,8)  | -               |                 |                 |                 |                 |                |
| 4. <i>R. robustus</i> II    | 6,2<br>(± 1,0)  | 7,7<br>(± 1,1)  | 7,5<br>(± 1,1)  | 0,4<br>(± 0,1)  |                 |                 |                 |                |
| 5. <i>R. robustus</i> III   | 6,1<br>(± 0,9)  | 7,5<br>(± 1,1)  | 7,1<br>(± 1,0)  | 3,9<br>(± 0,8)  | 0,2<br>(± 0,1)  |                 |                 |                |
| 6. <i>R. robustus</i> n.sp. | 7,6<br>(± 1,1)  | 8,6<br>(± 1,1)  | 8,8<br>(± 1,2)  | 4,6<br>(± 0,8)  | 2,6<br>(± 0,6)  | 0,3<br>(± 0,2)  |                 |                |
| 7. <i>R. robustus</i> IV    | 6,6<br>(± 1,0)  | 7,7<br>(± 1,4)  | 7,8<br>(± 1,1)  | 3,4<br>(± 0,7)  | 2,3<br>(± 0,5)  | 3,0<br>(± 0,6)  | 0,8<br>(± 0,2)  |                |
| 8. <i>R. barretti</i>       | 13,9<br>(± 1,3) | 15,6<br>(± 0,8) | 14,9<br>(± 1,4) | 13,4<br>(± 1,3) | 14,2<br>(± 1,3) | 14,8<br>(± 1,4) | 14,0<br>(± 1,3) | 0,5<br>(± 0,2) |

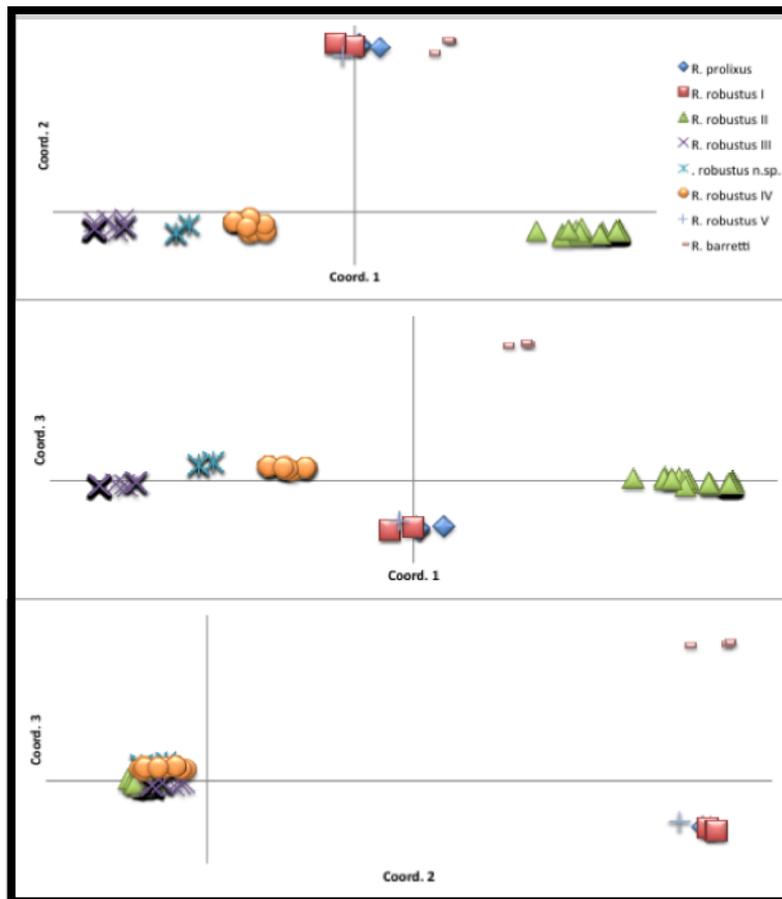
*R. barretti* é a linhagem mais divergente do complexo *R. prolixus* e *R. robustus s.l.*, apresentando porcentagens médias de *K-2p* entre 13,9 e 15,6%. *R. robustus* III e IV são as linhagens filogeneticamente mais próximas, quando comparados os oito clados identificados, atingindo em média 2,3% de divergência. *R. robustus* n.sp. diverge de *R. robustus* III em uma porcentagem um pouco maior, cerca de 2,6% (Tabela 5.1). Já *R. robustus* II diverge de *R. robustus* III, IV e n.sp. em 3,9, 4,6 e 3,4%, respectivamente.

O resultado dos agrupamentos obtidos na análise de componentes principais (PCA) foram distintos daqueles encontrados nas inferências filogenéticas bayesiana e

de máxima verossimilhança. Nas análises filogenéticas foi possível separar as sequências de mtCytb em oito agrupamentos, diferentemente da PCA, que a partir da análise de três eixos as separou em apenas sete. Os indivíduos de *R. robustus* I e de *R. prolixus* não foram identificados como pertencentes à grupamentos distintos (Figura 5.2).

Aproximadamente 80% da variação nucleotídica pôde ser explicada em três coordenadas, sendo 45% desta variação informativa para separação das linhagens na primeira coordenada, 22% na segunda e 13% na terceira coordenada. A primeira coordenada foi suficiente para separar claramente as linhagens amazônicas II, III, n.sp. e IV entre si (sendo *R. robustus* II considerada a linhagem mais distante deste grupo) e ainda separá-las de *R. prolixus*, *R. robustus* I, *R. robustus* V e *R. barretti*. Com base na segunda coordenada, foi possível observar que *R. robustus* V forma um agrupamento distinto de *R. prolixus* e *R. robustus* I, e a partir da observação da terceira coordenada, foi possível adicionar a informação de que *R. barretti* é o agrupamento mais distante do complexo *R. prolixus* e *R. robustus s.l.*

As porcentagens médias de divergência intraespecíficas foram menores do que 2%, variando entre 0,2 e 1,5% (Tabela 5.1). Nas 49 amostras de *R. prolixus* foram observados cinco haplótipos. Apesar de haver uma alta diversidade haplotípica ( $Hd = 0,76$ ), a diversidade nucleotídica observada nas sequências desta espécie foi a mais baixa ( $\pi = 1,60E-04$ ), quando comparada com aquelas encontradas nas sequências de mtCytb dos representantes do complexo *R. robustus s.l.* (Tabela 5.2). A outra espécie presente na região do Orinoco, *R. robustus* I, foi representada por oito indivíduos que compartilham três haplótipos distintos. Apesar de haver uma alta diversidade haplotípica ( $Hd = 0,71$ ), assim como observado em *R. prolixus*, a diversidade nucleotídica foi baixa ( $\pi = 1,51E-03$ ).



**Figura 5.2.** Gráfico das três coordenadas utilizadas na análise de PCA.

De cima para baixo (eixo X e eixo Y) – Coordenada (Coord.) 1 x Coord. 2; Coord. 1 x Coord. 3; e Coord. 2 x Coord. 3.

**Tabela 5.2.** Índices de variabilidade e testes de neutralidade.

$N_S$  = numero de amostras analisadas;  $N_H$  = numero de haplótipos observados;  $Hd$  – diversidade haplotípica;  $\pi$  = diversidade nucleotídica;  $D$  – teste de neutralidade de Tajima;  $F_S$  = teste de neutralidade de Fu.

|                          | $N_S$ | $N_H$ | $Hd$ | Pi ( $\pi$ ) | $D$    | $F_S$ |
|--------------------------|-------|-------|------|--------------|--------|-------|
| <i>R. prolixus</i>       | 49    | 5     | 0,76 | 1,60E-04     | 0,17   | -0,38 |
| <i>R. robustus</i> I     | 8     | 3     | 0,71 | 1,51E-03     | 1,10   | 0,20  |
| <i>R. robustus</i> V     | 1     | 1     | -    | -            | -      | -     |
| <i>R. robustus</i> II    | 78    | 16    | 0,68 | 3,88E-03     | -1,88* | -5,03 |
| <i>R. robustus</i> III   | 65    | 7     | 0,58 | 1,81E-03     | -0,13  | -1,07 |
| <i>R. robustus</i> n.sp. | 11    | 2     | 0,53 | 3,20E-03     | 1,95   | 4,19  |
| <i>R. robustus</i> IV    | 39    | 10    | 0,86 | 8,10E-03     | 0,31   | 1,67  |

\*  $P < 0,05$

Nos 193 indivíduos representantes das linhagens amazônicas de *R. robustus* foram observados 16 haplótipos de *R. robustus* II, sete de *R. robustus* III, dois de *R. robustus* n.sp. e 10 de *R. robustus* IV. A diversidade haplotípica observada foi de moderada a alta, variando de 0,53 (*R. robustus* II) a 0,86 (*R. robustus* IV). Já a diversidade nucleotídica de cada uma dessas quatro linhagens amazônicas foi maior do que aquelas observadas em *R. prolixus* e *R. robustus* I, variando entre 1,81E-03 (em *R. robustus* III) a 8,10E-03 (em *R. robustus* IV).

### 5.2.2. ITS-2

As amostras coletadas nas regiões geográficas de possível encontro de *R. robustus* II, III e IV (Careiro Castanho, Amazonas, e Ulianópolis, Pará) tiveram o ITS-2 sequenciado para a identificação de possíveis eventos de hibridação e introgressão de DNA mitocondrial. Não foi possível obter sequências de ITS-2 das amostras provenientes de Sucumbios (Equador) e Putumayo (Colômbia) identificadas a partir da análise do mtCytb como *R. robustus* V e *R. barretti*. As amostras provenientes de Ulianópolis (Pará), identificadas como *R. robustus* III, e de Anamá (Amazonas), identificadas como *R. robustus* II, tampouco tiveram o ITS-2 sequenciado com sucesso.

Ao todo, foram obtidos 13 clones de ITS-2 a partir de dois espécimes provenientes de Careiro Castanho identificados com base nas sequências do mtCytb como *R. robustus* II e 36 clones de três espécimes de Ulianópolis identificados como *R. robustus* n.sp. Foram incluídas nas análises outras 34 sequências de *R. robustus* n.sp. (provenientes de três espécimes coletados em São Roberto, Maranhão), além de 10 sequências de *R. robustus* III (Estrela, Tocantins) e 11 de *R. robustus* IV (Abaetetuba, Pará). Para fins de comparação, foram analisadas ainda seis sequências (uma de cada linhagem) previamente genotipadas a partir de análises do ITS-2 de *R. prolixus*, *R. robustus* I, II, III, IV e 6.4 (sensu Justi 2009).

Os tamanhos das sequências de ITS-2 dos clones analisados variaram de 702 a 742-pb em função de inserções e deleções observadas no alinhamento das sequências de *R. prolixus*, *R. robustus* I-IV, *R. robustus* n.sp. e *R. robustus* 6.4. Ao todo, foram identificadas 25 regiões de microssatélites (MSAT), sete regiões de *indel* não resultantes de mini ou microssatélites (nMSAT) e 91 sítios polimórficos resultantes de substituições nucleotídicas (MUT). Mais de 70% dessas regiões variáveis e das substituições nucleotídicas foram observadas quando apenas as linhagens amazônicas de *R. robustus* II, III, n.sp. e IV foram analisadas (MUT=71, MSAT=21, nMSAT=5; Tabela 5.3).

**Tabela 5.3.** Sítios variáveis encontrados nas comparações feitas entre as sequências de ITS-2. **MUT** - sítios polimórficos resultantes de substituições nucleotídicas; **MSAT** - regiões de microssatélites; **nMSAT** - regiões de *indel* não resultantes de mini ou microssatélites.

|  | MUT | MSAT | nMSAT |
|--|-----|------|-------|
| <i>R. prolixus</i> + <i>R. robustus</i> I + <i>R. robustus</i> II-IV + <i>R. robustus</i> n.sp. + <i>R. robustus</i> 6.4 | 91  | 25   | 7     |
| <i>R. robustus</i> II-IV + <i>R. robustus</i> n.sp. + <i>R. robustus</i> 6.4   | 75  | 21   | 7     |
| <i>R. robustus</i> II-IV + <i>R. robustus</i> n.sp.  | 71  | 21   | 5     |

Não foi possível confirmar a identidade taxonômica das amostras coletadas nas regiões de possível encontro entre *R. robustus* II, III e IV (Careiro Castanho, Amazonas, Ulianópolis, Pará, e São Roberto, Maranhão) quando analisados todos os clones de ITS-2 obtidos. Ao compará-los com as sequências-referência de *R. robustus* II, III, IV e 6.4, os clones das amostras de Careiro Castanho apresentaram

porcentagens de divergências genéticas de 0-0,8%, 0,7-1,2%, 2,1-2,4% e 1,3-1,7%, respectivamente, ao passo que os clones das amostras de Ulianópolis e São Roberto apresentaram divergências de 0,6-1,7%, 0,6-1,8%, 1,9-2,4% e 1,2-2,1%, respectivamente.

Os intervalos dos resultados das comparações interespecíficas extremamente dilatados levaram aos seguintes questionamentos: (1) a reconstrução filogenética a partir das sequências de ITS-2 representa uma classificação incompleta das linhagens (*ILS*)? (2) Os clones sequenciados representam mais de um *locus* que evoluem de maneira independente? Neste caso, (3) as linhagens *R. robustus* II, III e n.sp. podem ser separadas de alguma maneira, com base nas sequências de ITS-2?

Para responder a primeira questão, foram realizadas reconstruções filogenéticas bayesianas e de ML com todas as sequências obtidas, para inferir se as linhagens são agrupadas em clados distintos.

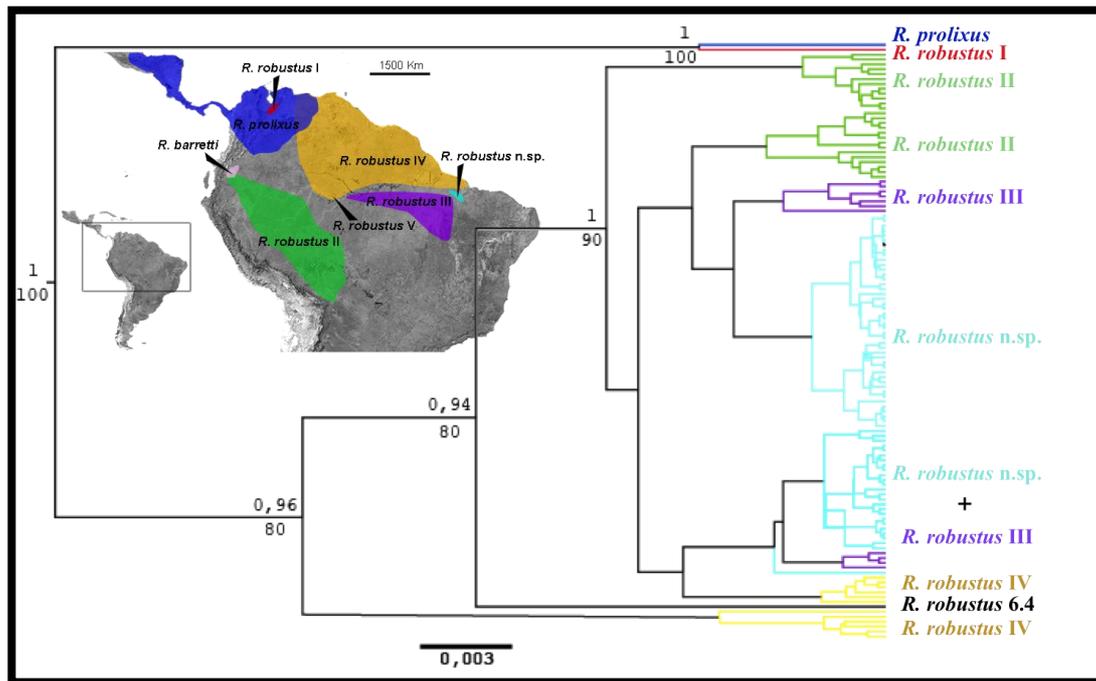
A segunda questão pôde ser respondida indiretamente com base na observação das porcentagens de variabilidade genética intraindividual e de divergências interespecíficas. Caso houvesse mais de um *locus*, era de se esperar que a variabilidade intraindividual nas sequências de ITS-2 fossem altas. Adicionalmente, esperava-se que as amostras provenientes de áreas geográficas com mínimas (senão nenhuma) chances de encontro entre as linhagens apresentassem em comparações interespecíficas intervalos de divergência genética igualmente dilatados aos observados nas comparações feitas com amostras coletadas em regiões geográficas de possível encontro.

Por fim, a terceira questão pôde ser respondida com base na comparação entre as sequências-padrão de *R. robustus s.l.* e os clones obtidos neste estudo. Caso fosse possível obter uma filogenia de ITS-2 coerente com a de mtCytb, era de se esperar que alguns clones sequenciados fossem divergentes das sequências-padrão em menos de 0,6% e que a divergência interespecífica fosse maior ou igual a 0,8%. Com base nas respostas para estas três questões, foi possível compreender melhor se *R. robustus* II, III e n.sp. são de fato espécies distintas (como observado nas árvores filogenéticas de mtCytb) ou se ainda existe fluxo gênico entre estas linhagens.

A partir do padrão das substituições nucleotídicas observado em regiões variáveis não resultantes de *indels*, o modelo de substituição nucleotídica adotado nas reconstruções filogenéticas bayesianas e de ML das sequências de ITS-2 foi o de Tamura e Nei (Tamura e Nei, 1993), com quatro parâmetros da distribuição gama para a variação das taxas de substituição nucleotídica entre os sítios (TrN + $\Gamma$ ). Nas inferências filogenéticas bayesianas, foi utilizado o modelo estocástico Dollo de multi-estado (Alekseyenko et al., 2008) para a análise dos *indels*, substituindo os nucleotídeos e os *gaps* de alinhamento por “1” e “0” em uma matriz binária.

As reconstruções filogenéticas bayesianas e de ML (Figura 5.3) realizadas a partir da análise das sequências de todos os clones de ITS-2 (excluindo as regiões de *indels*) confirmaram o relacionamento parafilético das linhagens de *R. robustus s.l.* ( $BS = 100$ ;  $PP = 1.0$ ). *R. robustus* I e *R. prolixus* formam um clado separado das demais linhagens de *R. robustus* (II, III, n.sp., IV e 6.4), assim como observado nas análises do mtCytb. Com base nestas reconstruções, é possível perceber ainda que *R. robustus* 6.4 é um táxon separado das demais linhagens amazônicas analisadas ( $BS = 80$ ;  $PP = 0.94$ ). Contudo, as sequências de *R. robustus* II, III, n.sp. e IV não foram agrupadas em clados separados e bem apoiados estatisticamente.

A análise dos *indels* observados nas sequências de *R. prolixus* e *R. robustus s.l.* não adicionaram informações filogenéticas relevantes. Sessenta e cinco clones de *R. robustus* II, III, n.sp. e IV foram agrupados com *R. prolixus* e *R. robustus* I, ao passo que os 42 clones demais de *R. robustus* III, n.sp. e IV foram agrupados em outro clado (dados não mostrados).



**Figura 5.3.** Árvore filogenética consenso de ITS-2.

Árvore construída pelo método bayesiano com base em matriz de distâncias TrN + $\Gamma$  para as 112 sequências do ITS-2 de *R. prolixus* e *R. robustus* s.l. A confiança dos agrupamentos foi testada a partir do cálculo da probabilidade posterior (mostrados na parte de cima de cada ramo) e 1000 replicações de *bootstrap* (mostrados na parte de baixo de cada ramo). O enraizamento da árvore foi feito no ponto médio.

A comparação entre clones obtidos de um mesmo indivíduo revelaram uma alta variabilidade intraindividual das sequências de ITS-2, comprovando que as cópias deste marcador não foram completamente homogeneizadas por evolução em concerto. Os 13 clones obtidos a partir das amostras coletadas em Careiro Castanho revelaram 10 sequências distintas, ao passo que os 36 clones de Ulianópolis, 34 de São Roberto, Maranhão (identificados com o mtCytb como *R. robustus* n.sp.), 10 de Estrela, Tocantins (identificados como *R. robustus* III) e 12 de Abaetetuba, Pará (identificados como *R. robustus* IV) apresentaram, respectivamente, 20, 20, 8 e 11 sequências diferentes. Apenas os indivíduos provenientes das localidades de Ulianópolis e São Roberto compartilharam sequências idênticas (cinco, ao todo).

As sequências de ITS-2 obtidas a partir das amostras coletadas nas áreas de possível encontro das linhagens *R. robustus* II-IV revelaram porcentagens de divergência intraindividual de 0 a 1%, ao passo que as sequências das amostras provenientes de regiões com mínimas chances de encontro entre as linhagens (Estrela, Tocantins, e Abaetetuba, Pará) apresentaram divergências entre 0 e 3% (Tabelas 5.4).

**Tabela 5.4.** Porcentagem de divergência K2-p intraindividual das sequências de ITS-2 de *R. robustus* II, III e IV.

N<sub>C</sub> – número de clones de ITS-2 analisados a partir de um mesmo indivíduo.

| Espécie*                 | Amostra                   | N <sub>C</sub> | K-2p (%) |
|--------------------------|---------------------------|----------------|----------|
| <i>R. robustus</i> II    | Careiro Castanho (AM) 1   | 9              | 0-0,6    |
|                          | Careiro Castanho 2 (AM) 2 | 4              | 0-0,8    |
| <i>R. robustus</i> n.sp. | São Roberto (MA) 1        | 10             | 0-0,8    |
|                          | São Roberto (MA) 2        | 12             | 0-1,0    |
|                          | São Roberto (MA) 3        | 12             | 0-1,0    |
| <i>R. robustus</i> n.sp. | Ulianópolis (PA) 1        | 10             | 0-0,4    |
|                          | Ulianópolis (PA) 2        | 12             | 0-0,9    |
|                          | Ulianópolis (PA) 3        | 14             | 0-0,4    |
| <i>R. robustus</i> III   | Estrela (TO)              | 10             | 0-1,6    |
| <i>R. robustus</i> IV    | Abaetetuba (PA)           | 12             | 0-3,0    |

\* Identidade taxonômica obtida a partir da análise do mtCytb.

A variabilidade intraindividual dos clones sequenciados foi maior do que a divergência média observada nas comparações interespecíficas entre *R. robustus* II, III e IV e ainda entre *R. prolixus* e *R. robustus* I (Tabela 5.5).

Este resultado foi observado mesmo quando as amostras provenientes das possíveis áreas de encontro entre as linhagens de *R. robustus* II-IV foram retiradas da análise (para descartar a possibilidade de estar analisando possíveis híbridos). Os clones de ITS-2 do indivíduo proveniente de Estrela (*R. robustus* III), quando comparados com *R. robustus* II e *R. robustus* III apresentam limites máximos de divergência genética muito próximos (1,4% e 1,5%, respectivamente). Um padrão semelhante foi observado quando os clones do indivíduo proveniente de Abaetetuba (*R. robustus* IV) foram analisados – o intervalo observado para as porcentagens de divergência entre estas sequências e a de *R. robustus* IV foi tão alto (0,7-2,6%), que se sobrepôs aos intervalos de divergência interespecíficos (1,3-3,0% quando comparadas com *R. robustus* II e 1,5-3,1%, quando comparadas com *R. robustus* III).

**Tabela 5.5.** Porcentagem de divergências genéticas (K-2p) interespecíficas (abaixo da diagonal fundo branco) e intraespecíficas (na diagonal, fundo cinza) das sequências de ITS-2. Os valores de erro padrão foram calculados a partir de 1000 iterações de *bootstrap* e são mostrados entre parêntesis. Em vermelho são destacados os valores de divergência genética calculados em comparações interespecíficas que atingiram valores iguais ou menores aqueles obtidos em comparações intraespecífica. **A.** Distância genética 1: %[mutações + todos os indels – indels resultantes de mini e microssatélites]; **B.** Distância genética 2: %[mutações]; **C.** Distância genética 3: %[mutações + todos os *indels*].

| <b>A. Distância 1</b>       | 1           | 2           | 3           | 4           | 5           | 6           | 7 |
|-----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|---|
| 1. <i>R. prolixus</i>       | -           |             |             |             |             |             |   |
| 2. <i>R. robustus</i> I     | 0,8 (± 0,3) | -           |             |             |             |             |   |
| 3. <i>R. robustus</i> II    | 3,1 (± 0,6) | 3,0 (± 0,6) | 0,5 (± 0,2) |             |             |             |   |
| 4. <i>R. robustus</i> III   | 3,5 (± 0,7) | 3,5 (± 0,7) | 0,9 (± 0,2) | 0,7 (± 0,2) |             |             |   |
| 5. <i>R. robustus</i> n.sp. | 3,3 (± 0,6) | 3,2 (± 0,6) | 0,7 (± 0,2) | 0,7 (± 0,2) | 0,5 (± 0,2) |             |   |
| 6. <i>R. robustus</i> IV    | 4,0 (± 0,7) | 3,8 (± 0,7) | 1,7 (± 0,3) | 2,1 (± 0,4) | 1,9 (± 0,3) | 1,7 (± 0,4) |   |
| 7. <i>R. robustus</i> 6.4   | 3,4 (± 0,7) | 3,3 (± 0,7) | 1,2 (± 0,4) | 1,6 (± 0,4) | 1,4 (± 0,4) | 2,2 (± 0,4) | - |
| <b>B. Distância 2</b>       | 1           | 2           | 3           | 4           | 5           | 6           | 7 |
| 1. <i>R. prolixus</i>       | -           |             |             |             |             |             |   |
| 2. <i>R. robustus</i> I     | 0,8 (± 0,3) | -           |             |             |             |             |   |
| 3. <i>R. robustus</i> II    | 2,7 (± 0,6) | 2,5 (± 0,6) | 0,4 (± 0,2) |             |             |             |   |
| 4. <i>R. robustus</i> III   | 3,1 (± 0,6) | 3,0 (± 0,6) | 0,8 (± 0,2) | 0,7 (± 0,2) |             |             |   |
| 5. <i>R. robustus</i> n.sp. | 2,8 (± 0,6) | 2,8 (± 0,6) | 0,6 (± 0,2) | 0,7 (± 0,2) | 0,5 (± 0,2) |             |   |
| 6. <i>R. robustus</i> IV    | 3,6 (± 0,6) | 3,5 (± 0,6) | 1,6 (± 0,2) | 2,0 (± 0,4) | 1,8 (± 0,3) | 1,7 (± 0,4) |   |
| 7. <i>R. robustus</i> 6.4   | 3,0 (± 0,6) | 2,8 (± 0,6) | 1,2 (± 0,4) | 1,7 (± 0,5) | 1,4 (± 0,4) | 2,2 (± 0,5) | - |
| <b>C. Distância 3</b>       | 1           | 2           | 3           | 4           | 5           | 6           | 7 |
| 1. <i>R. prolixus</i>       | -           |             |             |             |             |             |   |
| 2. <i>R. robustus</i> I     | 0,7 (± 0,3) | -           |             |             |             |             |   |
| 3. <i>R. robustus</i> II    | 3,8 (± 0,7) | 3,6 (± 0,7) | 0,5 (± 0,2) |             |             |             |   |
| 4. <i>R. robustus</i> III   | 3,7 (± 0,7) | 3,7 (± 0,7) | 0,9 (± 0,2) | 0,7 (± 0,2) |             |             |   |
| 5. <i>R. robustus</i> n.sp. | 3,5 (± 0,7) | 3,5 (± 0,7) | 0,7 (± 0,2) | 0,7 (± 0,2) | 0,5 (± 0,1) |             |   |
| 6. <i>R. robustus</i> IV    | 4,2 (± 0,7) | 4,1 (± 0,7) | 1,8 (± 0,3) | 2,2 (± 0,4) | 1,9 (± 0,3) | 1,8 (± 0,4) |   |
| 7. <i>R. robustus</i> 6.4   | 3,8 (± 0,7) | 3,6 (± 0,7) | 1,5 (± 0,4) | 1,7 (± 0,4) | 1,5 (± 0,4) | 2,6 (± 0,5) | - |

Alguns clones de um mesmo indivíduo e de indivíduos coletados em uma mesma localidade apresentaram porcentagens acima do esperado em comparações intraespecíficas (> 0,6%). Como estes clones também divergiam muito das sequências-padrão (i.e. previamente identificadas com base na análise do ITS-2), eles foram excluídos das análises por existir a suspeita de se tratarem de sequências provenientes de *loci* diferentes daquele amostrado nas sequências-padrão. Ao retirar essas sequências das análises, as divergências entre as linhagens foram recalculadas (Tabela 5.6).

**Tabela 5.6.** Porcentagem de divergências genéticas (K-2p) interespecíficas (abaixo da diagonal fundo branco) e intraespecíficas (na diagonal, fundo cinza) das sequências de ITS-2 que apresentaram porcentagens de divergência inferiores a 0,6% em comparações intraespecíficas.

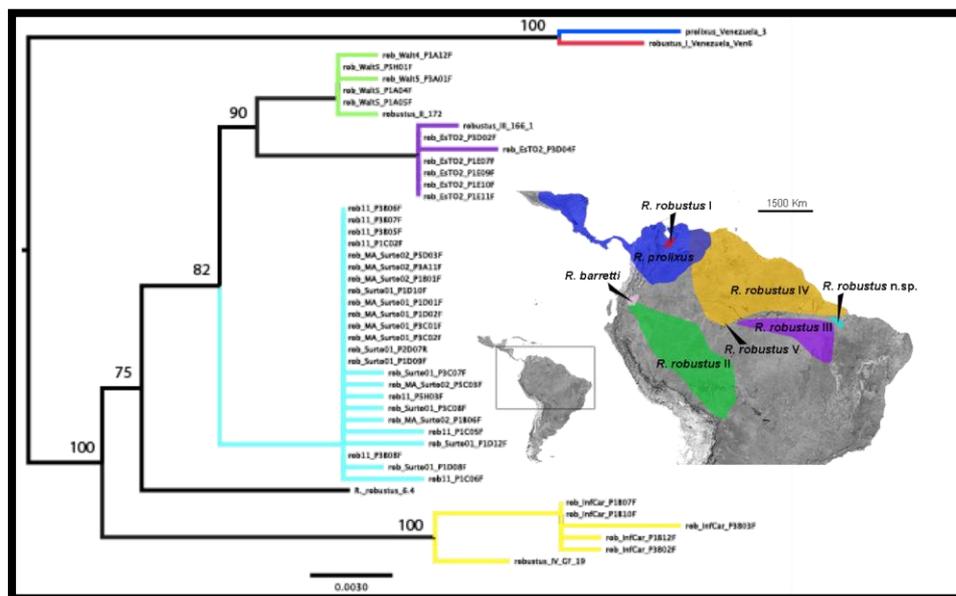
Em vermelho são destacados os valores de divergência genética calculados em comparações interespecíficas que atingiram valores iguais ou menores aqueles obtidos em comparações intraespecífica. **A.** Distância genética 1: %[mutações + todos os indels – indels resultantes de mini e microssatélites]; **B.** Distância genética 2: %[mutações]; **C.** Distância genética 3: %[mutações + todos os indels].

| <b>A. Distância 1</b>       | <b>1</b>    | <b>2</b>    | <b>3</b>    | <b>4</b>    | <b>5</b>    | <b>6</b>    | <b>7</b> |
|-----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|----------|
| 1. <i>R. prolixus</i>       | -           |             |             |             |             |             |          |
| 2. <i>R. robustus</i> I     | 0,8 (± 0,3) | -           |             |             |             |             |          |
| 3. <i>R. robustus</i> II    | 3,5 (± 0,7) | 3,3 (± 0,7) | 0,2 (± 0,1) |             |             |             |          |
| 4. <i>R. robustus</i> III   | 3,8 (± 0,7) | 3,6 (± 0,7) | 1,1 (± 0,3) | 0,1 (± 0,1) |             |             |          |
| 5. <i>R. robustus</i> n.sp. | 3,2 (± 0,6) | 3,3 (± 0,6) | 1,0 (± 0,3) | 1,2 (± 0,3) | 0,4 (± 0,1) |             |          |
| 6. <i>R. robustus</i> IV    | 4,3 (± 0,8) | 4,2 (± 0,8) | 2,7 (± 0,6) | 3,1 (± 0,6) | 2,8 (± 0,6) | 0,5 (± 0,2) |          |
| 7. <i>R. robustus</i> 6.4   | 3,4 (± 0,7) | 3,3 (± 0,7) | 1,5 (± 0,3) | 1,8 (± 0,4) | 1,5 (± 0,4) | 2,7 (± 0,6) | -        |

| <b>B. Distância 2</b>       | <b>1</b>    | <b>2</b>    | <b>3</b>    | <b>4</b>    | <b>5</b>    | <b>6</b>    | <b>7</b> |
|-----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|----------|
| 1. <i>R. prolixus</i>       | -           |             |             |             |             |             |          |
| 2. <i>R. robustus</i> I     | 0,8 (± 0,3) | -           |             |             |             |             |          |
| 3. <i>R. robustus</i> II    | 3,3 (± 0,7) | 3,1 (± 0,7) | 0,2 (± 0,1) |             |             |             |          |
| 4. <i>R. robustus</i> III   | 3,6 (± 0,7) | 3,4 (± 0,7) | 1,0 (± 0,3) | 0,1 (± 0,1) |             |             |          |
| 5. <i>R. robustus</i> n.sp. | 3,0 (± 0,7) | 3,1 (± 0,7) | 1,1 (± 0,3) | 1,4 (± 0,3) | 0,2 (± 0,1) |             |          |
| 6. <i>R. robustus</i> IV    | 4,3 (± 0,8) | 4,1 (± 0,8) | 2,6 (± 0,4) | 2,9 (± 0,5) | 2,5 (± 0,5) | 0,5 (± 0,2) |          |
| 7. <i>R. robustus</i> 6.4   | 3,2 (± 0,7) | 3,1 (± 0,7) | 1,4 (± 0,3) | 1,7 (± 0,4) | 1,6 (± 0,4) | 2,5 (± 0,5) | -        |

| <b>C. Distância 3</b>       | <b>1</b>    | <b>2</b>    | <b>3</b>    | <b>4</b>    | <b>5</b>    | <b>6</b>    | <b>7</b> |
|-----------------------------|-------------|-------------|-------------|-------------|-------------|-------------|----------|
| 1. <i>R. prolixus</i>       | -           |             |             |             |             |             |          |
| 2. <i>R. robustus</i> I     | 0,8 (± 0,3) | -           |             |             |             |             |          |
| 3. <i>R. robustus</i> II    | 4,0 (± 0,7) | 3,8 (± 0,6) | 0,1 (± 0,1) |             |             |             |          |
| 4. <i>R. robustus</i> III   | 3,8 (± 0,7) | 3,6 (± 0,7) | 1,0 (± 0,3) | 0,1 (± 0,1) |             |             |          |
| 5. <i>R. robustus</i> n.sp. | 3,7 (± 0,7) | 4,0 (± 0,7) | 1,4 (± 0,3) | 1,5 (± 0,3) | 0,1 (± 0,1) |             |          |
| 6. <i>R. robustus</i> IV    | 4,3 (± 0,7) | 4,2 (± 0,7) | 2,9 (± 0,5) | 2,9 (± 0,5) | 3,1 (± 0,6) | 0,5 (± 0,1) |          |
| 7. <i>R. robustus</i> 6.4   | 3,8 (± 0,7) | 3,6 (± 0,7) | 1,6 (± 0,4) | 1,7 (± 0,4) | 1,9 (± 0,4) | 2,8 (± 0,6) | -        |

Nessa reanálise as médias das porcentagens de divergência entre as sequências em comparações intraespecíficas foram todas menores do que aquelas obtidas em comparações interespecíficas (0,1-0,5% e 0,8-4,3%, respectivamente). Como resultado, a topologia da árvore de ITS-2 mostrou claramente a separação entre as espécies amazônicas de *R. robustus* s.l. (II, III, IV e 6.4), com apoio estatístico de *bootstrap* acima de 75 (Figura 5.4). Os clones das amostras provenientes de Careiro Castanho foram agrupadas dentro do clado de *R. robustus* II. Já os clones analisados de Ulianópolis e de São Roberto, assim como na topologia obtida a partir do mtCytb, foram agrupados em um clado separado das demais espécies amazônicas (*R. robustus* n.sp.; *BS* = 82). As diferenças entre as topologias obtidas com base nas sequências de ITS-2 e mtCytb foram: (1) *R. robustus* III está filogeneticamente mais relacionada com *R. robustus* II do que com *R. robustus* IV; (2) *R. robustus* III é filogeneticamente mais próxima de *R. robustus* II do que de *R. robustus* n.sp.



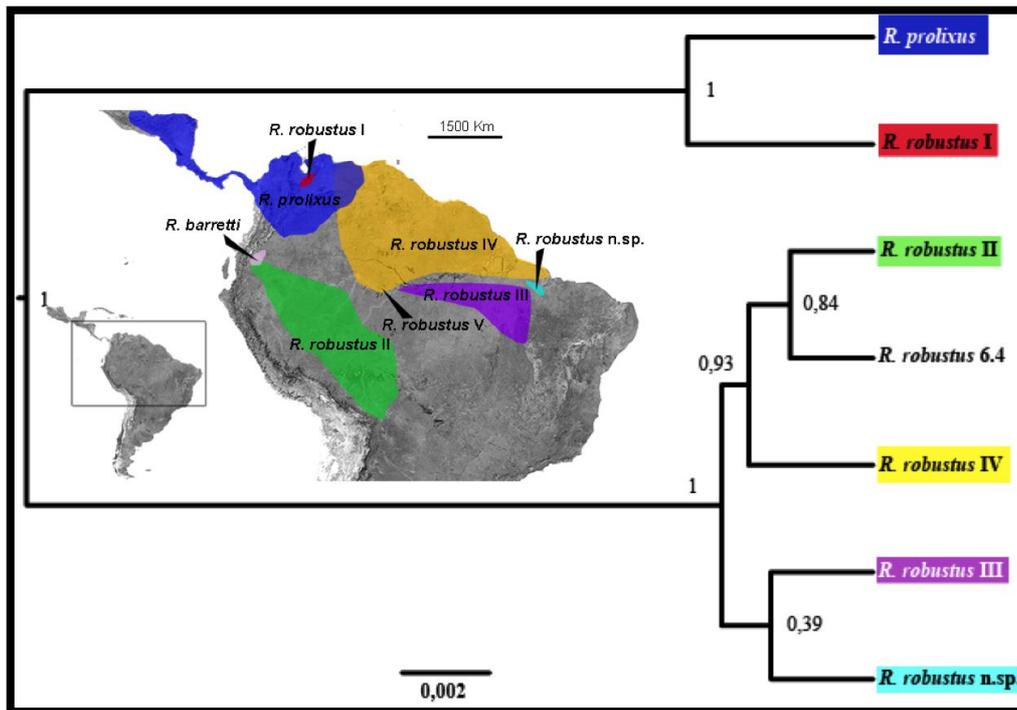
**Figura 5.4.** Árvore filogenética de ITS-2 após análise de homologia das sequências. Árvore consenso construída por ML com base em matriz de distâncias TrN + $\Gamma$  das 46 sequências de ITS-2 de *R. prolixus* e *R. robustus* s.l. selecionadas. A confiança dos agrupamentos foi testada a partir de 1000 replicações de *bootstrap*. O enraizamento da árvore foi feito no ponto médio. Diferente do observado na reconstrução filogenética a partir do mtCytb, *R. robustus* III está filogeneticamente mais próximo de *R. robustus* II do que de *R. robustus* IV.

### 5.2.3. Árvore filogenética de espécies com base nas sequências de mtCytb e ITS-2

A inferência filogenética bayesiana concatenando as informações genéticas contidas nas sequências de mtCytb e nos clones de ITS-2 foi realizada com o objetivo de saber qual o consenso obtido em relação ao relacionamento entre *R. robustus* II, III, IV e n.sp. e se os clones de ITS-2 que apresentaram porcentagens de divergência nucleotídica acima do esperado em comparações intraespecíficas representam uma classificação incompleta da linhagem (*ILS*). Para esta análise, foram incluídos todos os clones de ITS-2 de cada um dos espécimes sequenciados, além das sequências do mtCytb obtidas neste estudo.

A partir da topologia resultante da árvore concatenada de mtCytb e ITS-2 (Figura 5.5), foi possível confirmar a parafilia de *R. robustus s.l.* – *R. robustus* I é filogeneticamente mais próxima de *R. prolixus* do que das demais espécies de *R. robustus* ( $PP = 1.0$ ). Em relação ao grau de parentesco das espécies de *R. robustus* amazônicas, diferente do observado na topologia do mtCytb, *R. robustus* IV foi agrupada mais proximamente a *R. robustus* II do que a *R. robustus* III. *R. robustus* 6.4, diferente do observado na topologia da árvore de ITS-2, está proximamente mais relacionada a *R. robustus* II ( $PP = 0,93$ ).

O clado formado por *R. robustus* III e *R. robustus* n.sp. foi o único não resolvido na árvore filogenética. Apesar destas duas espécies terem sido agrupadas em um clado distinto das demais espécies amazônicas de *R. robustus s.l.* ( $PP = 1.0$ ), a separação entre elas teve baixo apoio estatístico ( $PP = 0,39$ ).



**Figura 5.5.** Árvore filogenética consenso de espécies obtida a partir da concatenação das sequências de mtCytb e ITS-2.

A confiança dos agrupamentos foi testada a partir do cálculo da probabilidade posterior. O enraizamento da árvore foi feito no ponto médio. A topologia final se assemelhou mais aquela obtida a partir das sequências de ITS-2 do que a obtida com base em mtCytb.

### 5.3. Análise de microssatélites

A partir da descoberta inesperada de seis indivíduos molecularmente identificados como *R. robustus* III e seis identificados como *R. robustus* n.sp. habitando uma mesma palmeira em Ulianópolis (Pará), foi possível testar se há fluxo gênico entre estas linhagens. Este teste foi realizado com base na análise de 10 *loci* de microssatélites previamente descritos em *R. prolixus* (Fitzpatrick et al., 2009; Harry et al., 2008b). Para fins de comparação, foram adicionadas cinco amostras provenientes de São Roberto, Maranhão (*R. robustus* n.sp.), e cinco de Rondon do Pará, localidade adjacente de Ulianópolis, onde foram encontrados somente espécimes de *R. robustus* III.

Todos os 10 *loci* foram amplificados com sucesso nas 17 amostras provenientes de Ulianópolis e Rondon do Pará. Apenas o *locus* R8 não foi amplificado nas cinco amostras de São Roberto. Nove dos 10 *loci* isolados de *R.*

*prolixus* apresentaram em *R. robustus* III e n.sp. tamanhos diferentes do esperado (Tabela 4.4).

Dois a sete alelos foram observados em cada um dos *loci* analisados [média de 3,5 alelos por *locus* ( $A_M$ )]. *List014-21* e *List014-64* foram os mais polimórficos, apresentando seis e sete alelos, respectivamente. Já *R8* e *List014-13* foram considerados monomórficos, por apresentarem frequências do alelo mais comum em proporções maiores que 95% e, portanto, foram excluídos das análises subsequentes.

Para testar a hipótese de que há fluxo gênico entre *R. robustus* III e n.sp. em áreas simpátricas ( $H_1$ ), as amostras foram agrupadas em quatro populações, sendo: (1) seis espécimes de *R. robustus* n.sp. provenientes de Ulianópolis; (2) seis de *R. robustus* III de Ulianópolis; (3) cinco de *R. robustus* n.sp. de São Roberto; e (4) cinco *R. robustus* III de Rondon do Pará. A hipótese nula ( $H_0$ ) neste caso é que os indivíduos agrupados em 1 e 2 pertencem a mesma população.

Os *loci* *R13* e *List014-21* da população de São Roberto apresentaram evidências para alelos nulos. Contudo, a inclusão destes alelos nas análises não influenciou os resultados finais e, portanto, eles foram mantidos (dados não mostrados).

Em nenhum dos agrupamentos da  $H_0$  ou  $H_1$  foi detectado indícios de abandono de alelos (*allele drop out*). Ao agrupar as amostras de acordo com a  $H_1$ , o número de alelos observados em cada *locus* variou de um a quatro em cada população. As populações de *R. robustus* n.sp. de Ulianópolis e *R. robustus* III de Rondon do Pará foram as que apresentaram a maior média de alelos por *locus* ( $A_{M1} = 2,22$ ;  $A_{M2} = 1,78$ ;  $A_{M3} = 1,33$ ;  $A_{M4} = 2,22$ ) e a população 3 foi a que mais apresentou *loci* monomórficos ( $L_{M1} = 1$ ;  $L_{M2} = 3$ ;  $L_{M3} = 4$ ;  $L_{M4} = 1$ ; Tabela 5.7).

**Tabela 5.7.** Resumo da variabilidade dos 10 *loci* de microssatélites encontrada nas quatro populações de *R. robustus* III e n.sp. analisadas.

**ULRr n.sp.** = amostras de *R. robustus* n.sp. coletadas em Ulianópolis (Pará); **UL Rr III** = amostras de *R. robustus* III coletadas em Ulianópolis (Pará); **MA Rr n.sp.** = amostras de *R. robustus* n.sp. coletadas em São Roberto (MA); **RP Rr III** = amostras de *R. robustus* III coletadas em Rondon do Pará (Pará); **N** = número de amostras analisadas; **L<sub>A</sub>** = número de *loci* analisados; **L<sub>M</sub>** = número de *loci* monomórficos; **A<sub>P</sub>** = número de alelos privados.

| População      | N | L <sub>A</sub> | L <sub>M</sub> | A <sub>M</sub> | A <sub>P</sub> |
|----------------|---|----------------|----------------|----------------|----------------|
| 1. UL Rr n.sp. | 6 | 10             | 1              | 2,22           | 1              |
| 2. UL Rr III   | 6 | 10             | 3              | 1,78           | 1              |
| 3. MA Rr n.sp. | 5 | 9              | 4              | 1,33           | 3              |
| 4. RP Rr III   | 5 | 10             | 1              | 2,22           | 8              |

Nos oito *loci* analisados, foram encontrados ao todo 13 alelos privados (Tabela 5.7), sendo um na população 1 (120 em *R31*), um na população 2 (308 em *List014-21*), três na população 3 (280 em *List014-21*, 250 e 254 em *List014-64*) e oito na população 4 (257 em *R4*, 153 em *List014-25*, 130 em *R31*, 309 em *List014-10*, 112 e 118 em *List014-10*, 302 em *List014-21* e 256 em *List014-64*).

Após a correção dos níveis de significância de todos os testes com múltiplas comparações ( $\alpha = 0,015$ ), foi possível observar que as diferenças nas frequências alélicas de todos os *loci* (não monomórficos) foram informativos para a separação das populações (Tabela 5.8). Os resultados obtidos no teste do Chi-Quadrado dos valores G aplicados nas análises de Shannon (Tabela 5.9) mostraram entre as populações 1 e 3 diferenças significativas nas frequências alélicas de dois *loci* (*List014-10* e *List014-21*). Adicionalmente, as frequências alélicas de 4 *loci* (*List014-25*, *R26*, *List014-21* e *List014-64*) foram significativamente distintas nas comparações entre as populações 1 e 4. Já as comparações entre as populações 2 e 3, 2 e 4, e ainda 3 e 4 resultaram na identificação de diferenças significativas nas frequências alélicas de cinco (*List014-25*, *List014-10*, *List014-21* e *List-014-64*), quatro (*List014-25*, *R26*, *List014-21* e *List014-64*) e cinco (*List014-25*, *List014-10*, *R26*, *List014-21* e *List014-64*) *loci*, respectivamente. Apesar de não ter sido encontrado nenhum *locus* diagnóstico para as populações simpátricas de *R. robustus* n.sp. e *R. robustus* III, estas apresentaram no *locus R31* frequências alélicas significativamente distintas ( $P = 0,01$ ).

Com base nas análises de PCA, foi possível observar três agrupamentos, compostos pela população 3, população 4 e populações 1 e 2 (Figura 5.6). Portanto,

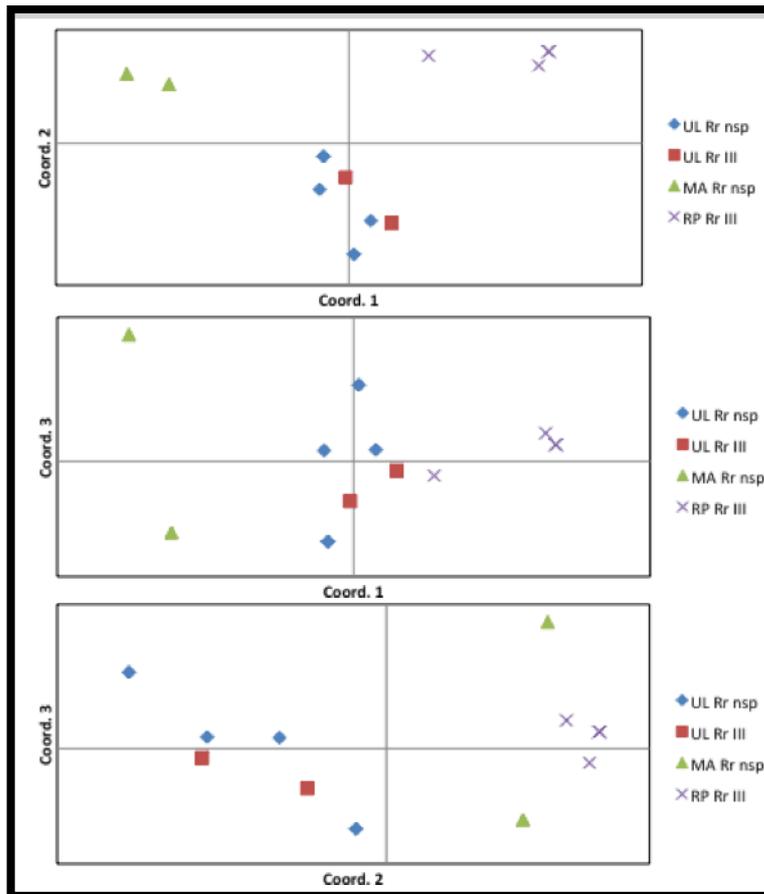
não foi possível separar as populações de *R. robustus* III e n.sp. coletadas em uma mesma palmeira.

**Tabela 5.8.** Frequência alélica encontrada em cada uma das quatro populações dos 10 *loci* de microssatélites analisados.

| <i>Locus</i>       | <b>Alelos</b> | <b>UL Rr nsp</b> | <b>UL Rr III</b> | <b>MA Rr nsp</b> | <b>RP Rr III</b> |
|--------------------|---------------|------------------|------------------|------------------|------------------|
| <b>R13</b>         | <b>188</b>    | 0,167            | 0,000            | 0,400            | 0,000            |
|                    | <b>190</b>    | 0,833            | 1,000            | 0,600            | 1,000            |
| <b>R4</b>          | <b>253</b>    | 0,750            | 0,667            | 1,000            | 0,900            |
|                    | <b>257</b>    | 0,000            | 0,000            | 0,000            | 0,100            |
|                    | <b>266</b>    | 0,250            | 0,333            | 0,000            | 0,000            |
| <b>List014-25</b>  | <b>151</b>    | 0,833            | 0,500            | 1,000            | 0,000            |
|                    | <b>153</b>    | 0,000            | 0,000            | 0,000            | 0,100            |
|                    | <b>170</b>    | 0,167            | 0,500            | 0,000            | 0,900            |
| <b>R31</b>         | <b>120</b>    | 0,333            | 0,000            | 0,000            | 0,000            |
|                    | <b>126</b>    | 0,667            | 1,000            | 1,000            | 0,800            |
|                    | <b>130</b>    | 0,000            | 0,000            | 0,000            | 0,200            |
| <b>List014-010</b> | <b>307</b>    | 0,500            | 0,500            | 0,000            | 0,800            |
|                    | <b>309</b>    | 0,000            | 0,000            | 0,000            | 0,100            |
|                    | <b>311</b>    | 0,250            | 0,333            | 0,000            | 0,000            |
|                    | <b>317</b>    | 0,250            | 0,167            | 1,000            | 0,100            |
| <b>R26</b>         | <b>112</b>    | 0,000            | 0,000            | 0,000            | 0,800            |
|                    | <b>116</b>    | 1,000            | 1,000            | 1,000            | 0,000            |
|                    | <b>118</b>    | 0,000            | 0,000            | 0,000            | 0,200            |
| <b>List014-21</b>  | <b>280</b>    | 0,000            | 0,000            | 0,400            | 0,000            |
|                    | <b>298</b>    | 0,250            | 0,000            | 0,600            | 0,500            |
|                    | <b>300</b>    | 0,500            | 0,500            | 0,000            | 0,000            |
|                    | <b>302</b>    | 0,000            | 0,000            | 0,000            | 0,500            |
|                    | <b>304</b>    | 0,250            | 0,333            | 0,000            | 0,000            |
|                    | <b>308</b>    | 0,000            | 0,167            | 0,000            | 0,000            |
| <b>List014-64</b>  | <b>244</b>    | 0,167            | 0,333            | 0,000            | 0,100            |
|                    | <b>250</b>    | 0,000            | 0,000            | 0,200            | 0,000            |
|                    | <b>252</b>    | 0,583            | 0,667            | 0,600            | 0,000            |
|                    | <b>254</b>    | 0,000            | 0,000            | 0,200            | 0,000            |
|                    | <b>256</b>    | 0,000            | 0,000            | 0,000            | 0,400            |
|                    | <b>258</b>    | 0,083            | 0,000            | 0,000            | 0,400            |
|                    | <b>260</b>    | 0,167            | 0,000            | 0,000            | 0,100            |

**Tabela 5.9.** Valores do índice G calculados a partir de comparações par-a-par das frequências alélicas interpopulacionais de cada um dos 8 *loci* polimórficos analisados. O índice de significância alfa foi corrigido para 0,014 pelo método de FDR. Em vermelho, os P-valores obtidos pelo método do Chi-Quadrado das comparações estatisticamente significativas. Note que a população 1 (*R. robustus* n.sp. proveniente de Ulianópolis, Pará) apresenta frequências alélicas do *locus* R31 significativamente distintas das observadas na população simpátrica de *R. robustus* III (Pop 2). A numeração das populações está de acordo com o apresentado no texto. **Pop** = População; **gl** = graus de liberdade; **Chi** = Chi-Quadrado.

| Comparações   | R13  |    |         | R4   |    |         | List014-25 |    |           | R31  |    |         | List014-10 |    |           | R26   |    |           | List014-21 |    |           | List014-64 |    |           |
|---------------|------|----|---------|------|----|---------|------------|----|-----------|------|----|---------|------------|----|-----------|-------|----|-----------|------------|----|-----------|------------|----|-----------|
|               | G    | gl | P (Chi) | G    | gl | P (Chi) | G          | gl | P (Chi)   | G    | gl | P (Chi) | G          | gl | P (Chi)   | G     | gl | P (Chi)   | G          | gl | P (Chi)   | G          | gl | P (Chi)   |
| Pop 1 x Pop 2 | 2,96 | 1  | 0,086   | 0,20 | 1  | 0,653   | 3,10       | 1  | 0,078     | 6,35 | 1  | 0,012*  | 0,35       | 2  | 0,842     | 0     | 0  | 1         | 7,08       | 3  | 0,07      | 4,91       | 3  | 0,179     |
| Pop 1 x Pop 3 | 1,51 | 1  | 0,219   | 4,03 | 1  | 0,045   | 2,59       | 1  | 0,108     | 5,59 | 1  | 0,018   | 16,3       | 2  | P < E-04* | 0     | 0  | 1         | 18,86      | 3  | P < E-04* | 12,37      | 5  | 0,03      |
| Pop 2 x Pop 3 | 7,40 | 1  | 0,007*  | 5,59 | 1  | 0,018   | 9,15       | 1  | 0,002*    | 0    | 0  | 1       | 19,5       | 2  | P < E-04* | 0     | 0  | 1         | 30,32      | 4  | P < E-04* | 11,19      | 3  | 0,011*    |
| Pop 1 x Pop 4 | 2,59 | 1  | 0,108   | 5,36 | 2  | 0,068   | 19,8       | 2  | P < E-04* | 8,14 | 2  | 0,017   | 6,67       | 3  | 0,082     | 30,32 | 2  | P < E-04* | 19,73      | 3  | P < E-04* | 17,67      | 4  | 0,001*    |
| Pop 2 x Pop 4 | 0    | 0  | 1       | 6,81 | 2  | 0,033   | 10,1       | 2  | 0,006*    | 3,40 | 1  | 0,065   | 7,38       | 3  | 0,061     | 30,32 | 2  | P < E-04* | 30,32      | 4  | P < E-04* | 25,31      | 4  | P < E-04* |
| Pop 3 x Pop 4 | 6,56 | 1  | 0,01*   | 1,44 | 1  | 0,23    | 27,7       | 2  | P < E-04* | 2,99 | 1  | 0,084   | 21,0       | 2  | P < E-04* | 27,73 | 2  | P < E-04* | 12,57      | 2  | 0,002*    | 27,73      | 6  | P < E-04* |



**Figura 5.6.** Gráfico das três coordenadas utilizadas na análise de PCA dos *loci* de microssatélites analisados.

De cima para baixo (eixos X e eixos Y) – Coordenada (Coord.) 1 x Coord. 2; Coord. 1 x Coord. 3; e Coord. 2 x Coord. 3. **Pop** = População; **gl** = graus de liberdade.

Não foi observado qualquer indício de desequilíbrio de ligação nos *loci* de cada população ( $P \geq 0,015$ ). Em referência ao equilíbrio das populações da  $H_1$  segundo o modelo de Hardy-Weinberg (H-W), apenas *List014-10* e *List014-21* na população 1 apresentaram desvios significativos do equilíbrio (Tabela 5.10). Quando 1 e 2 foram agrupadas em uma única população ( $H_0$ ), não foram observados desvios significativos (dados não mostrados).

**Tabela 5.10.** Teste Chi-Quadrado das frequências alélicas dos marcadores microssatélites das populações de *R. robustus* III, IV e sp. nov.

Análise feita para avaliar se as frequências alélicas encontradas em cada população dos oito loci polimórficos estão de acordo com o esperado caso estivesse em equilíbrio de Hardy-Weinberg. Somente a população 1 apresentou dois *loci* em desequilíbrio, *List014-10* e *List014-21*.

| <i>Locus</i>      | Pop 1 – UL Rr nsp |    |               | Pop 2 – UL Rr III |    |       | Pop 3 – MA Rr nsp |    |       | Pop 4 – RP Rr III |    |       |
|-------------------|-------------------|----|---------------|-------------------|----|-------|-------------------|----|-------|-------------------|----|-------|
|                   | Chi-Quadrado      | gl | P             | Chi-Quadrado      | gl | P     | Chi-Quadrado      | gl | P     | Chi-Quadrado      | gl | P     |
| <i>R13</i>        | 0,240             | 1  | 0,624         | 0,240             | 1  | 0,624 | 5,000             | 1  | 0,025 | 5,000             | 1  | 0,025 |
| <i>R4</i>         | 0,667             | 1  | 0,414         | 1,500             | 1  | 0,221 | 1,500             | 1  | 0,221 | 0,062             | 1  | 0,804 |
| <i>List014-25</i> | 0,240             | 1  | 0,624         | 6,000             | 1  | 0,016 | 6,000             | 1  | 0,016 | 0,062             | 1  | 0,804 |
| <i>R31</i>        | 0,375             | 1  | 0,540         | 0,375             | 1  | 0,540 | 0,375             | 1  | 0,540 | 5,000             | 1  | 0,025 |
| <i>List014-10</i> | 12,000            | 3  | <b>0,007*</b> | 6,000             | 3  | 0,112 | 6,000             | 3  | 0,112 | 10,000            | 3  | 0,019 |
| <i>R26</i>        | -                 | -  | -             | -                 | -  | -     | -                 | -  | -     | 5,000             | 1  | 0,025 |
| <i>List014-21</i> | 12,000            | 3  | <b>0,007*</b> | 6,000             | 3  | 0,112 | 5,000             | 1  | 0,025 | 5,000             | 1  | 0,025 |
| <i>List014-64</i> | 6,245             | 6  | 0,396         | 1,500             | 1  | 0,221 | 10,000            | 3  | 0,019 | 15,000            | 6  | 0,020 |

\* valor significativo após a correção do valor-P pelo algoritmo de FDR (P = 0,012) para múltiplas comparações.

Nenhum dos valores de  $F_{IS}$  calculados para cada população apresentou valores significativos, possivelmente em função do tamanho amostral ( $P > 0,08$ ) (Tabela 5.11). Ao contrário do observado em comparações intrapopulacionais, as comparações interpopulacionais feitas com base nos índices de  $F_{ST}$  e  $R_{ST}$  (Tabela 5.11) revelaram que praticamente todas as populações estão estruturadas. A exceção foi observada justamente quando as populações simpátricas de *R. robustus* III e n.sp. foram comparadas ( $F_{ST} = 0,03$ ,  $P = 0,18$ ;  $R_{ST} = 0,12$ ,  $P = 0,05$ ).

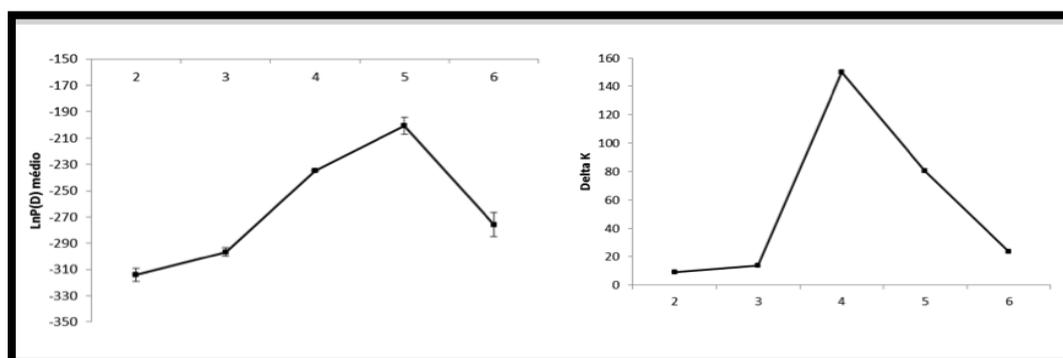
**Tabela 5.11.** Valores dos índices de  $R_{ST}$  (abaixo da diagonal) e  $F_{ST}$  (acima da diagonal) calculados a partir de comparações par-a-par das frequências alélicas interpopulacionais dos oito *loci* polimórficos analisados.

O índice de significância alfa foi corrigido para 0,014 pelo método de FDR. Note que os valores de  $R_{ST}$  e  $F_{ST}$  não foram significativos apenas na comparação entre as populações 1 e 2. Na diagonal (em cinza), os valores de  $F_{IS}$  calculados a partir de comparações par-a-par das frequências alélicas intrapopulacionais dos oito *loci* polimórficos analisados. Nenhum dos valores foi significativo, provavelmente em função do baixo número amostral de cada população.

| População    | 1     | 2     | 3     | 4     |
|--------------|-------|-------|-------|-------|
| 1. UL Rr nsp | 0,11  | 0,03  | 0,25* | 0,43* |
| 2. UL Rr III | 0,12  | -0,61 | 0,36* | 0,41* |
| 3. MA Rr nsp | 0,99* | 0,99* | 0,78  | 0,66* |
| 4. RP Rr III | 0,34* | 0,34* | 0,99* | 0,07  |

\*  $P < 0,0001$

Nas análises bayesianas para a definição do número de agrupamentos (K) em que as populações amostradas estariam inseridas, as cadeias de MCMC permaneceram visualmente estáveis e não apresentaram nenhuma tendência. Com base na observação dos valores de  $\Delta K$  (Figura 5.7), as populações da  $H_1$  foram separadas em agrupamentos distintos (Figura 5.8).

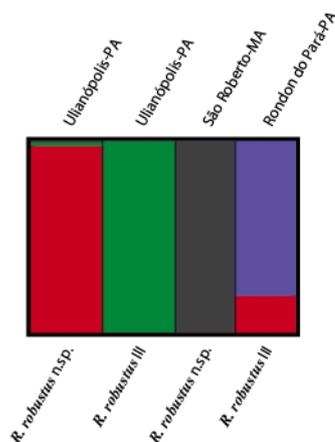


**Figura 5.7.** Gráfico dos valores de verossimilhança para definição do número de agrupamentos (K).

À esquerda, o gráfico dos valores de verossimilhança calculados a partir do log natural da probabilidade dos dados em agrupamentos (K) de dois a seis populações. Note que o valor de verossimilhança do agrupamento contendo quatro populações apresentou o menor desvio-padrão em relação a média e antecede imediatamente o platô do gráfico. À direita, os valores de Delta K ( $\Delta K$ ) plotados em um gráfico. Note que o maior valor obtido foi para o agrupamento contendo quatro populações.

Apesar da falta de um *locus* diagnóstico, os indivíduos de *R. robustus* III e n.sp. coletados em uma mesma palmeira foram identificados como representantes de

populações distintas. O único indício de fluxo gênico significativo entre as quatro populações foi evidenciado no agrupamento de *R. robustus* III proveniente de Rondon do Pará. Um indivíduo desta localidade apresentou 93,4% de chances de pertencer ao agrupamento de *R. robustus* s.l. de Ulianópolis. Este resultado indica um possível evento de hibridação entre *R. robustus* III e n.sp. em Rondon do Pará.



**Figura 5.8.** Gráfico da estruturação das quatro populações analisadas.

A partir do resultado, foi possível confirmar a separação das populações simpátricas de *R. robustus* n.sp. e *R. robustus* III provenientes de Ulianópolis (PA). Um indivíduo da população de *R. robustus* III de Rondon do Pará (PA) foi identificado como geneticamente mais semelhante aos indivíduos pertencentes à população de *R. robustus* n.sp. de Ulianópolis.

#### 5.4. Isolamento de novos marcadores moleculares no genoma de *R. prolixus* e análises filogenéticas

##### 5.4.1. Marcador molecular diagnóstico para *R. prolixus* - AmpG

Em uma etapa inicial do estudo, com o objetivo de identificar marcadores diagnósticos para *R. prolixus*, sete amplicons foram testados em 27 *R. prolixus*, dois *R. robustus* I, quatro *R. robustus* II, quatro de *R. robustus* III e seis de *R. robustus* IV (Tabela 1 do Anexo 1). Nesta primeira análise, as amostras amplificadas por PCR e sequenciadas para esses sete marcadores revelaram 13 polimorfismos, nove dos quais são polimorfismos de um único nucleotídeo (SNPs) [Tabela A.2 do Anexo 1]. Um destes SNPs, localizado no sítio 280 da região de 364-pb amplificada do *AmpG* com os primers *AmpGF*, 5' - GAG AGC TGA AGA TAG GCA AGC G, e *AmpGR*, 5' - TGA TAA CTG GAT TAG GCG CAG C foi diagnóstico para *R. prolixus* (i.e. todos os *R. prolixus* têm uma adenina, ao invés de uma guanina, neste sítio). Para confirmar a consistência e utilidade deste SNP como marcador diagnóstico para identificação de *R. prolixus* foram sequenciados outros 154 espécimes, dos quais 46 *R. prolixus*, sendo

23 provenientes da Guatemala, Colômbia, Venezuela (além de 23 de uma localidade desconhecida), dois *R. robustus* I (Trujillo, Venezuela), 83 *R. robustus* II, provenientes do Equador (Sucumbios), Brasil (uma localidade do Amazonas e sete do Pará) e Bolívia (uma localidade de Pando e quarto de Santa Cruz); 19 *R. robustus* III, provenientes do Brasil (quatro localidades do Pará e duas de Tocantins); e quatro *R. robustus* IV, sendo um do Brasil (Pará) e três da Guiana Francesa (Cayenne e Remiré) (Tabela 4.3; Tabela 1 do Anexo 1). O sítio 280 permaneceu sendo diagnóstico para *R. prolixus*.

O resultado filogenético da árvore bayesiana consenso e da rede de haplótipos construída a partir das sequências do *AmpG* confirmaram a natureza parafilética de *R. robustus s.l.* com respeito à *R. prolixus* (Figura 2 do Anexo 1). *R. prolixus* e *R. robustus* I compartilham oito sítios polimórficos, três dos quais informativos para as análises de parcimônia (A, C, and A, nas posições 174, 270 e 360, respectivamente; Tabela 2 do Anexo 1). A porcentagem de divergência *K-2p* entre as sequências destas espécies é de 0,6%, em contraste com 1,1-1,7% observados nas comparações entre sequências de *R. robustus* I e *R. robustus* II-IV (Tabela 3 do Anexo 1). A região do *AmpG* (amplicon contendo o SNP diagnóstico para *R. prolixus*) está localizada no quarto íntron do gene hipotético de cópia única *Proteína Transmembrana 165* (*TP165*), que parece estar localizado no cromossomo X de *R. prolixus* (Figura 3 do Anexo 1). Maiores detalhes acerca dos resultados desta seção podem ser encontrados no Anexo 1.

#### 5.4.1. Genes do relógio circadiano

Nestas análises, não foi possível obter sequências de *R. robustus* I e V e de *R. barretti*, devido ao mal estado de conservação dos espécimes obtidos.

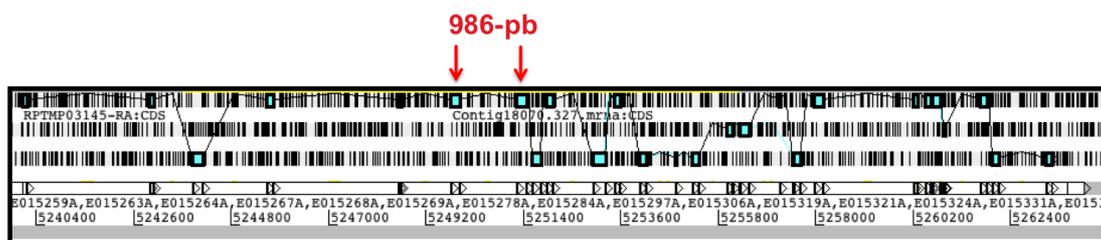
#### 5.4.2. period

Foram utilizadas as sequências das proteínas homólogas de PER de *Antheraea pernyi*, *Drosophila melanogaster*, *Periplaneta americana*, *Apis mellifera*, *Culex quinquefasciatus*, *Gryllus bimaculatus*, *Pediculus humanus corporis*, *Riptortus*

*pedestris* e *Tribolium castaneum* (das quais as três primeiras foram anotadas manualmente e revisadas) para a definição das regiões conservadas em insetos.

A partir do alinhamento da sequência de *D. melanogaster* com as de outros oito insetos, foi possível identificar em todos os ortólogos de PER as regiões de domínio protéico PAS-1 e PAS-2. Os domínios de *R. pedestris* (Hemiptera: Coreoidea) foram escolhidos como base para buscar a região homóloga correspondente no genoma de *R. prolixus*, por conta da maior proximidade filogenética desta espécie com os triatomíneos. Oito *hits* foram encontrados em regiões no sentido senso do Supercontig 18070 (GL563191). A região com maior similaridade (escore = 157 bits; *e-value* = 5e-39) está inserida na proteína codificante RPTMP03145-RA (na posição 5234661-5264037), criada a partir das predições “C1” e “Lagerblad”.

Quando a sequência completa de PER de *D. melanogaster* foi utilizada na busca, foram encontrados 13 *hits*, cujas regiões mais similares continuaram inseridas na proteína predita RPTMP03145-RA (melhor *hit* com escore = 396 e *e-value* = 1e-110). O resultado foi semelhante ao de quando a sequência completa do ortólogo de PER de *R. pedestris* foi submetida como base para a busca (20 *hits*, todos na região de RPTMP03145-RA; melhor *hit* com *e-value* = 0 e escore = 706). De acordo com a predição “C1”, a proteína candidata a ser homóloga de PER apresenta 1.136 AA (entre as posições 5240104 e 5263201 do Supercontig 18070), 88 AA a menos que a maior das proteínas de insetos selecionadas para as análises (*D. melanogaster*). O gene predito é composto por 23.198 pares de base, que compõem em 26 éxons e 25 íntrons. Na predição “Lagerblad”, a proteína RPTMP03145-RA foi predita como sendo um pouco menor, com 884 AA, localizados entre as posições 5249785 e 5263307 do mesmo Supercontig. Nesta predição, a região do genoma de *R. prolixus* que abrange os cinco éxons e íntrons iniciais da predição “C1” não foi considerada como pertencente ao gene (Figura 5.9).



**Figura 5.9.** Éxons (quadrados em ciano) e íntrons (linhas em cinza e ciano) do gene *per* predito.

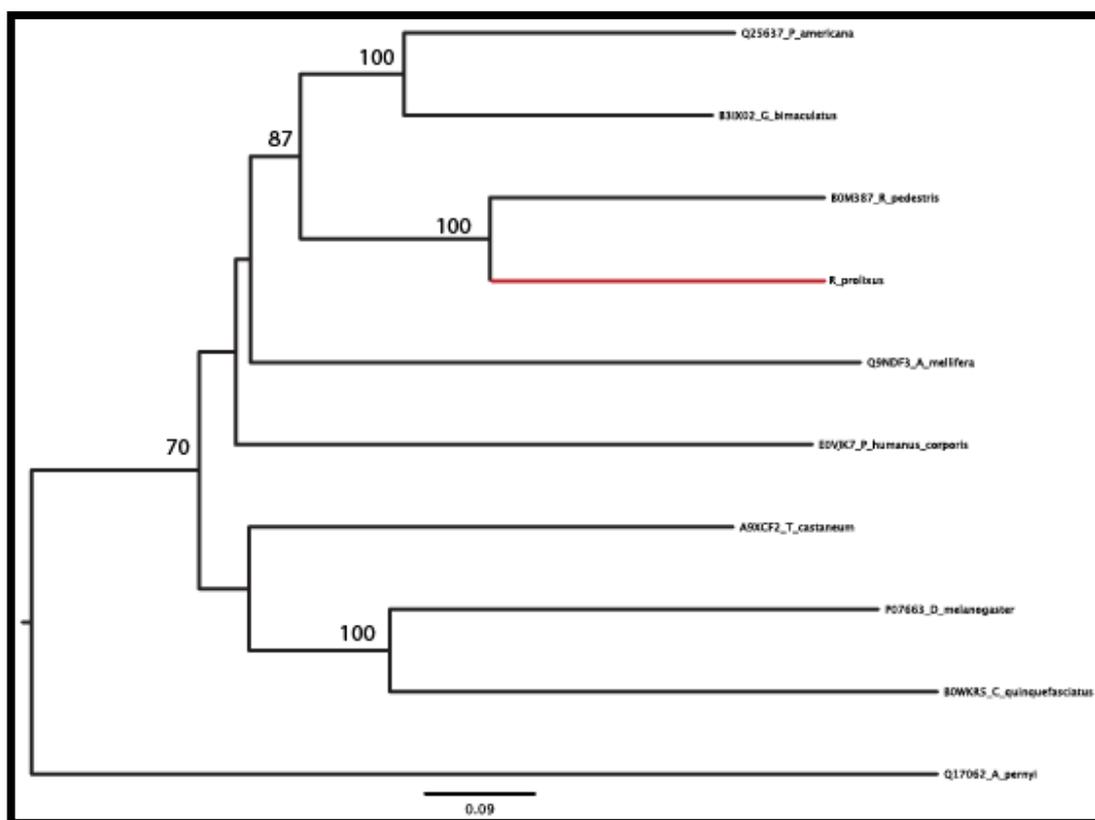
Cada uma das três linhas formadas por barras verticais pretas representa um quadro de leitura do éxon (o mais alto representa leitura na primeira base). As setas voltadas para a direita no retângulo abaixo dos quadros de leitura representam a leitura do gene no sentido senso. As setas em vermelho indicam a posição dos éxons 6 e 7 em que os iniciadores foram desenhados, de forma a amplificarem pela PCR um fragmento de 986-pb contendo o íntron 6 por completo (908-pb).

Quando as sequências de AA de ambas as predições foram inseridas no alinhamento das proteínas de insetos homólogas de PER, observou-se que a predição “C1” é mais correta do que a “Lagerblad”. Os aminoácidos da predição “C1” foram alinhados no início da sequência de *R. pedestris*, sendo os primeiros 50 idênticos. Já a sequência da predição “Lagerblad” foi alinhada apenas no 175º aminoácido dessa espécie. As análises posteriores de curadoria confirmaram que a predição “C1” estava correta (resultados não mostrados).

Com base no alinhamento das sequências de proteínas homólogas de PER com RPTMP03145-RA, foi possível identificar os domínios PAS-1 e PAS-2 na sequência de *R. prolixus*. Ademais, a topologia da árvore filogenética dos ortólogos de PER de insetos confirmaram que RPTMP03145-RA é a proteína ortóloga de PER em *R. prolixus* (Figura 5.10). Esta sequência foi agrupada com a de *R. pedestris* em um mesmo clado. *R. prolixus* e *R. pedestris* são filogeneticamente mais relacionadas às espécies das Ordens Dicyoptera e Orthoptera (*P. americana* e *G. bimaculatus*) do que às espécies da Ordem Diptera (*D. melanogaster* e *Cx. quinquefasciatus*).

A região escolhida para o anelamento dos iniciadores é bastante conservada em insetos da Ordem Hemiptera. O iniciador senso (*PER\_F* 5' – GTGGTTATGCCCTCGTGGAA) está ancorado na porção final do domínio PAS-2 e o anti-senso (*PER\_R* 5' - CGTTACATCAGGATTGGTGGG) no éxon subsequente. Além do íntron de 908-pb, os iniciadores amplificam os 25 aminoácidos finais de PAS-2 e dois aminoácidos do éxon seguinte, totalizando 989-pb.

Após os diferentes testes nas reações de PCR para amplificação de amostras de *R. prolixus* e *R. robustus* II, os melhores resultados foram obtidos a partir da utilização, em cada volume final de reação de 25µL, de 10ng de DNA, 2,5 µL de tampão, 1,25 µL de cloreto de magnésio (MgCl<sub>2</sub>), 2,5 µM de cada primer e 1,25 unidades de *Taq Platinum*<sup>®</sup> *DNA polimerase* (Invitrogen). As amostras foram submetidas à 96 °C por 5 min, seguido de 94 °C por 30 segundos, 63 °C por 3 min e 72 °C por 3min, e ainda à uma extensão final de 72 °C por 30 min.



**Figura 5.10.** Árvore filogenética de NJ construída com base em seqüências de aminoácidos dos ortólogos de PER em Insecta.

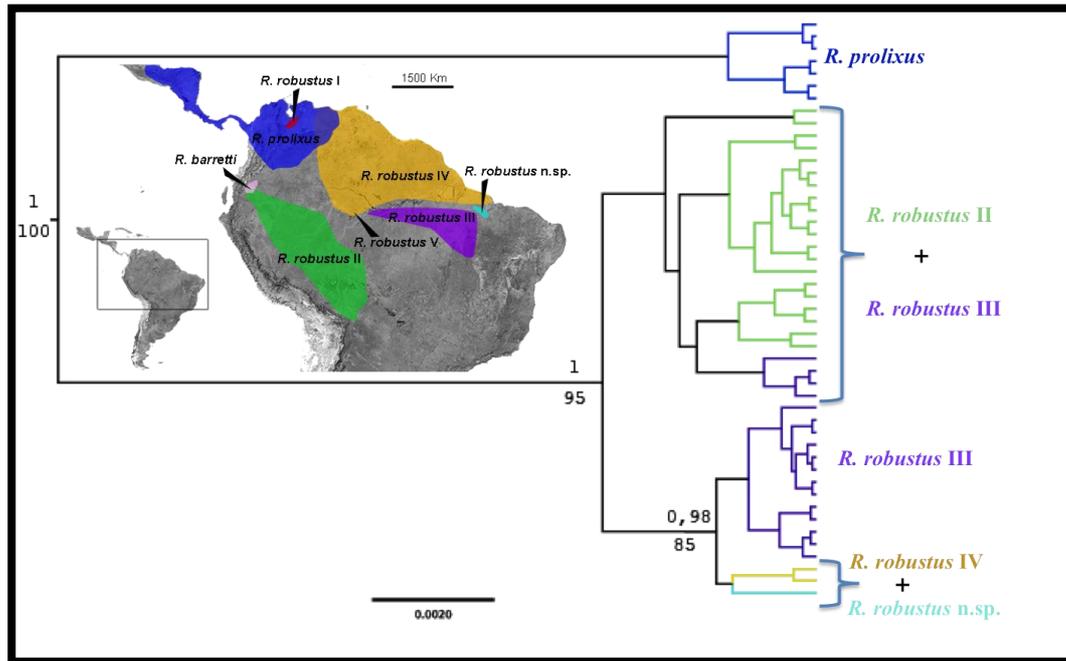
O modelo de substituição utilizado foi o Dayhoff. Apenas as regiões das proteínas que não continham *gaps* de alinhamento foram analisadas. A confiança dos agrupamentos foi testada a partir de 1000 replicações de *bootstrap*. Somente os valores maiores ou iguais a 70 são mostrados. Note que a seqüência de AA de *R. prolixus* se insere em um agrupamento interno da árvore dos ortólogos de PER (está proximamente relacionada a um outro hemíptero, *R. pedestris*), provando que o gene isolado do genoma de *R. prolixus* é ortólogo de *per*. Em vermelho, o ramo da árvore representado por *R. prolixus*.

Cinquenta e uma amostras tiveram o fragmento ortólogo de *per* sequenciado (nove *R. prolixus*, 21 *R. robustus* II, 17 *R. robustus* III, um *R. robustus* n.sp. e três *R. robustus* IV). Trinta e nove sítios variáveis e nove *indels* (que variam de 1 a 38 nucleotídeos) foram observados.

O modelo de substituição nucleotídica adotado nas reconstruções filogenéticas bayesianas e de ML das sequências do ortólogo de *per* foi o Hasegawa, Kishino e Yano (Hasegawa et al., 1985), com quatro parâmetros da distribuição gama para a variação das taxas de substituição nucleotídica entre os sítios (HKY + $\Gamma$ ). Nas inferências filogenéticas bayesianas, foi utilizado o modelo estocástico Dollo de multi-estado (Alekseyenko et al., 2008) para a análise dos *indels*, substituindo os nucleotídeos e os *gaps* de alinhamento por “1” e “0” em uma matriz binária.

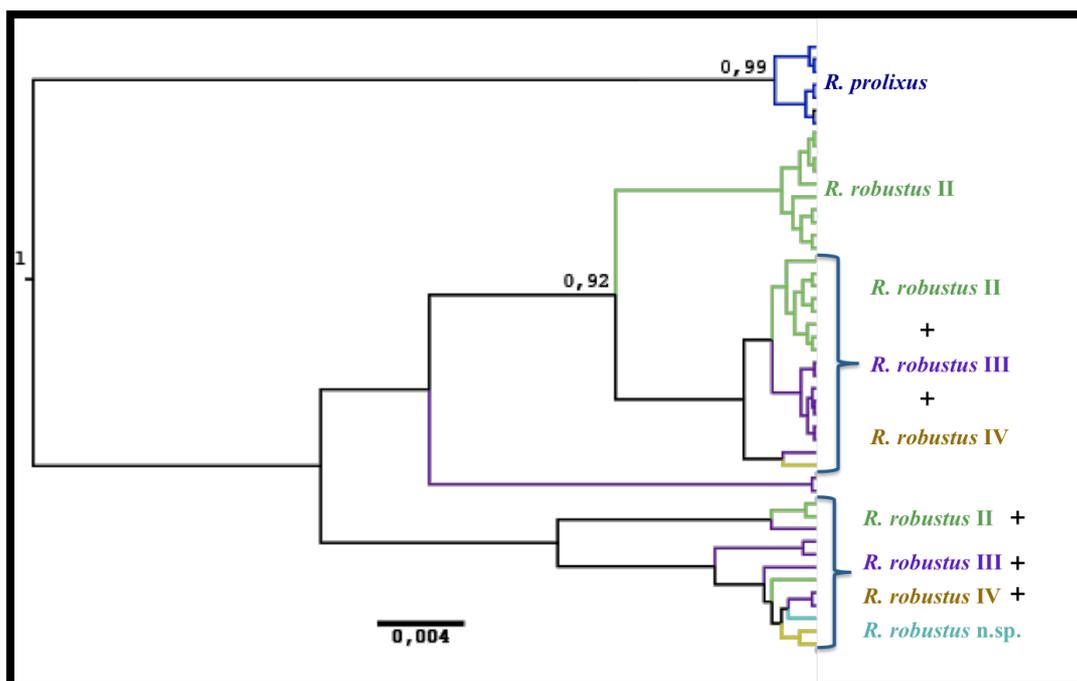
Com base na topologia resultante da análise das sequências de *per* sem os *indels* (Figura 5.11), é possível identificar a separação entre *R. prolixus* e *R. robustus s.l.* Contudo, o mesmo não pode ser dito em relação às linhagens de *R. robustus s.l.* Quatro clados foram observados na árvore filogenética resultante. No primeiro clado, as amostras de *R. prolixus* foram agrupadas separadamente das demais (ML = 100; PP = 1,0). No segundo clado foram agrupados todos os espécimes de *R. robustus* II com quatro espécimes de *R. robustus* III provenientes de Paraupebas (Pará). Este clado é filogeneticamente mais distante dos outros dois clados de *R. robustus s.l.* (BS = 95; PP = 1,0). No terceiro, estão agrupados os demais indivíduos de *R. robustus* III. Finalmente, o quarto clado foi composto por sequências de *R. robustus* IV e *R. robustus* n.sp. A separação entre as sequências destas duas espécies não foi bem apoiada estatisticamente (BS < 70 e PP < 0,8).

A topologia da árvore contendo as informações dos *indels* (Figura 5.12) confirmou a separação entre *R. prolixus* e *R. robustus s.l.* (PP = 1,0). Entretanto, não foi observada nenhuma separação confiável entre os demais clados (PP < 0,8).



**Figura 5.11.** Árvore filogenética de *per*.

Árvore consenso construída pelo método bayesiano com base em matriz de distâncias HKY +I + $\Gamma$  para as sequências de *per* de *R. prolixus* e *R. robustus* s.l. A confiança dos agrupamentos foi testada a partir do cálculo da probabilidade posterior (mostrados na parte de cima de cada ramo) e 1000 replicações de *bootstrap* (mostrados na parte de baixo de cada ramo), calculados na inferência filogenética feita pelo método de máxima verossimilhança. Apenas os valores de probabilidade posterior acima de 0,8 e de bootstrap acima de 80 são mostrados. O enraizamento da árvore foi feito no ponto médio. Note que o segundo e terceiro clados (de cima para baixo) agrupam sequências de *R. robustus* II com algumas de *R. robustus* III, e de *R. robustus* IV com as de *R. robustus* n.sp., respectivamente.



**Figura 5.12.** Árvore filogenética dos indels de *per*.

Árvore consenso construída pelo método bayesiano com base no algoritmo Dollo de multiestado para os indels do ortólogo de *per* de *R. prolixus* e *R. robustus s.l.* codificados em uma matriz binária. A confiança dos agrupamentos foi testada a partir do cálculo da probabilidade posterior. Apenas os valores de probabilidade posterior acima de 0,8 são mostrados. O enraizamento da árvore foi feito no ponto médio. Note que a partir da análise dos indels, é possível diferenciar apenas *R. prolixus* de *R. robustus s.l.*

#### 5.4.3. timeless

A definição das regiões mais conservadas nas proteínas homólogas de TIM em insetos foi realizada com base no alinhamento das sequências de aminoácidos (AA) de *D. melanogaster* (isoforma D), *T. castaneum* (isoforma A), *P. humanus corporis* e *Cx. quinquefasciatus*. Apenas a proteína de *D. melanogaster* foi anotada manualmente e revisada.

Foi feita uma busca no alinhamento das sequências homólogas de TIM por uma região de AA já conhecida em *D. melanogaster*, que está relacionada ao bom funcionamento do relógio circadiano (5' – VSTLQKLLSLWFEASLSESSSEDNESNT SPPKQ). Devido à sua importância funcional, esperava-se que esta região fosse conservada nas demais proteínas homólogas e, portanto, muito útil para uma busca inicial no genoma de *R. prolixus* pela proteína homóloga de TIM.

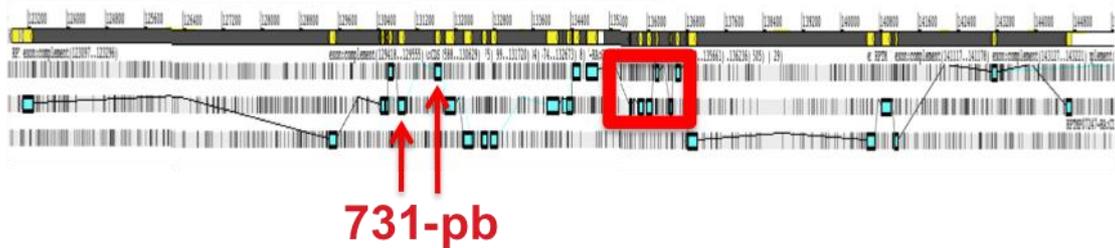
Apenas no penúltimo aminoácido dessa região funcional foi encontrada uma substituição (de K para E) em *T. castaneum* e *P. humanus corporis*. Adicionalmente,

os 17 aminoácidos anteriores à região-alvo eram idênticos em todos os insetos (5' – WGVTMVQLIALIYKDQH). Portanto, além da região de importância funcional já conhecida em *D. melanogaster*, esses 17 aminoácidos foram adicionados à sequência-base para a comparação com *R. prolixus*.

Quando esta sequência de AA foi buscada no genoma de *R. prolixus*, o *e-value* e o escore resultantes foram insatisfatórios ( $5e-07$  e 49 *bits*, respectivamente), provavelmente em função do tamanho curto da sequência-base. Contudo, esta busca inicial revelou uma região no Contig de *R. prolixus* GL562808 (Supercontig 17687) de grande identidade com a sequência-base (25 aminoácidos semelhantes de 29 comparados), levantando a suspeita de que esta deveria ser a provável localização da proteína homóloga de TIM em *R. prolixus*. Esta sequência de 29 AA está inserida em uma proteína predita por “C1” e “Lagerblad”, chamada RPTMP07247-RA.

Uma nova busca foi realizada, desta vez utilizando a sequência completa de AA de TIM de *D. melanogaster*, por ser esta a única anotada manualmente e revisada. A proteína codificante predita RPTMP07247-RA no sentido anti-senso foi confirmada como a principal candidata a ser homóloga de TIM em *R. prolixus* (melhor hit com escore de 543 *bits* e *e-value* igual a  $1e-154$ ).

A proteína RPTMP07247-RA predita por “C1” está inserida entre as posições 123097 e 144756 do Supercontig 17687 do genoma de *R. prolixus*, e contém 1086 AA, sendo, portanto, a terceira maior proteína homóloga de TIM em relação às aquelas selecionadas de outros insetos (21 AA a menos que a isoforma A de *T. castaneum* e 10 a mais que a proteína de *P. humanus corporis*). O gene de RPTMP07247-RA apresenta 21.660 nucleotídeos que compõem 27 éxons e 26 íntrons. A predição “Lagerblad” está localizada no meio da predição “C1” (posições 130468-135093; Figura 5.13), contendo apenas 513 AA, quase a metade do tamanho da menor proteína selecionada de insetos (996 AA, em *Cx. quinquefasciatus*). O gene predito por “Lagerblad” contém 12 éxons e 11 íntrons e negligencia os éxons 1-12, 24, 26 e 27 do gene de “C1”. Adicionalmente, as predições são discordantes em relação à posição de início dos éxons 13 e 22 e do fim dos éxons 15, 19 e 25 de “C1” (1 e 11; 3, 7 e 13 em “Lagerblad”, respectivamente).



**Figura 5.13.** Éxons (quadrados em ciano e amarelo) e íntrons (linhas em cinza e ciano) do gene ortólogo *tim* predito.

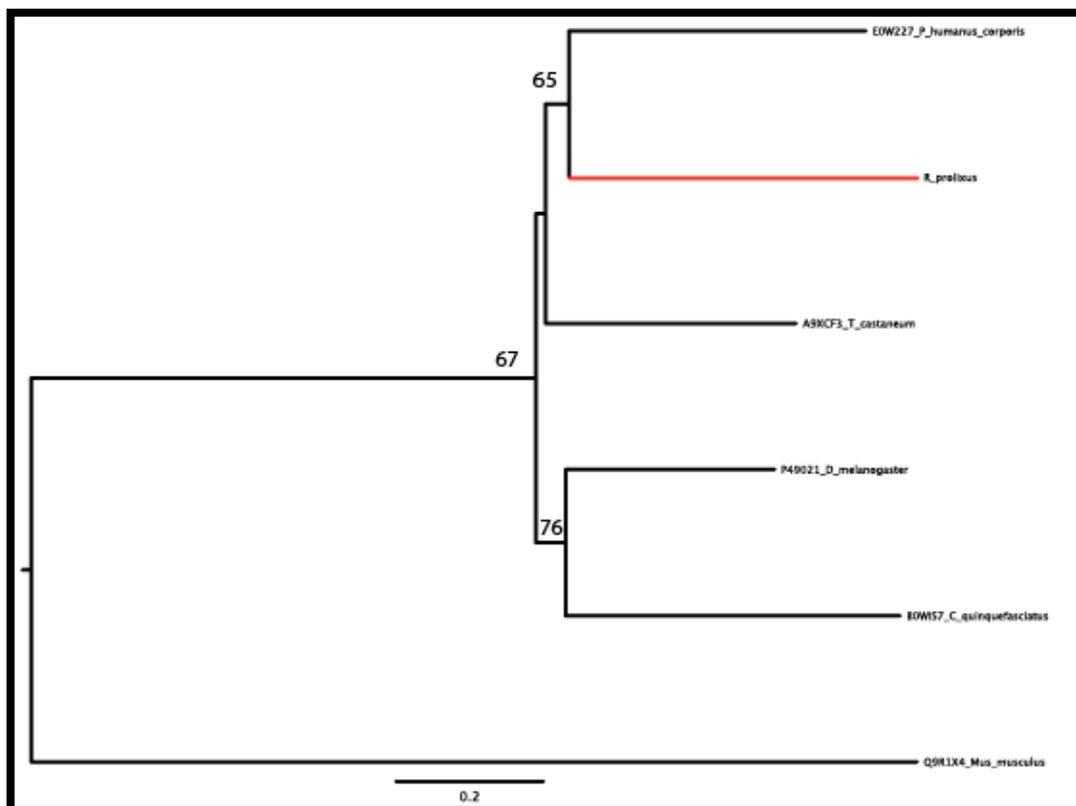
Cada uma das três linhas formadas por barras verticais pretas representa um quadro de leitura do éxon (o mais alto representa leitura na primeira base). O retângulo em vermelho representam os éxons preditos erroneamente. As setas voltadas para a direita no retângulo cinza acima dos quadros de leitura representam a leitura do gene no sentido antissenso. As setas em vermelho indicam a posição dos éxons 15 e 16 em que os iniciadores foram desenhados, de forma a amplificarem pela PCR um fragmento de 731-pb contendo o íntron 15 por completo (623-pb).

Ao comparar a sequência de aminoácidos da predição “C1” com as sequências homólogas de TIM, foi possível encontrar a região conservada em insetos que está relacionada ao bom funcionamento do relógio circadiano (37 de 44 AA idênticos). Contudo, os três últimos aminoácidos desta região foram separados por 123 AA, produtos dos éxons 7 a 12. A sequência aminoácidos da predição “Lagerblad” se inicia nestes três últimos aminoácidos da região conservada e, portanto, parece não abranger a proteína homóloga de TIM por completo. Com base nas análises posteriores de curadoria, foi possível identificar que a região entre os éxons 7 e 12 eram uma predição errônea e, conseqüentemente, foram retirados da análise. A proteína TIM de *R. prolixus* contém 963 AA e o gene 21 éxons e 20 íntrons.

Com base na topologia da árvore, foi possível confirmar que RPTMP07247-RA é a proteína homóloga de TIM de *R. prolixus* (Figura 5.14). Esta proteína parece ser filogeneticamente mais próxima do clado formado pelas proteínas de espécies da Ordem Coleoptera e Phthiraptera (*T. castaneum* e *P. humanus corporis*, respectivamente) do que do clado das espécies da Ordem Diptera (*D. melanogaster* e *Cx. quinquefasciatus*). Contudo, devido aos baixos valores de *bootstrap* na árvore, os resultados acerca do parentesco não são conclusivos. Para fins de comparação, a proteína homóloga de TIM em mamíferos (ou m-TIM) de *Mus musculus* foi adicionada na análise filogenética como grupo externo.

Como o íntron subsequente da região mais conservada entre insetos era muito curto (29-pb), a região terminal do éxon 15 e inicial do éxon 16 foram escolhidas para

para o desenho dos iniciadores senso e anti-senso (ambos os éxons foram igualmente preditos por “C1” e “Lagerblad”). Os iniciadores (TIM\_E1516F 5’ – AATCGCGGTC CAGTAACACA e TIM\_E1516R 5’ – AACAAATCGCCACAGGAT) amplificam em *R. prolixus* um fragmento de 731-pb, dos quais 623-pb são correspondentes ao íntron 15, 9-pb do éxon 15 e 93-pb do éxon 16.



**Figura 5.14.** Árvore filogenética de NJ dos genes ortólogos de TIM em Insecta. O modelo de substituição utilizado foi o Dayhoff. Apenas as regiões das proteínas que não continham *gaps* de alinhamento foram analisadas. A confiança dos agrupamentos foi testada a partir de 1000 replicações de *bootstrap*. Somente os valores maiores ou iguais a 65 são mostrados. Note que a sequência de AA de *R. prolixus* se insere em um agrupamento interno da árvore dos ortólogos de TIM (está proximamente relacionada a *P. humanus corporis*), provando que o gene isolado do genoma de *R. prolixus* é ortólogo de *tim*. Em vermelho, o ramo da árvore representado por *R. prolixus*.

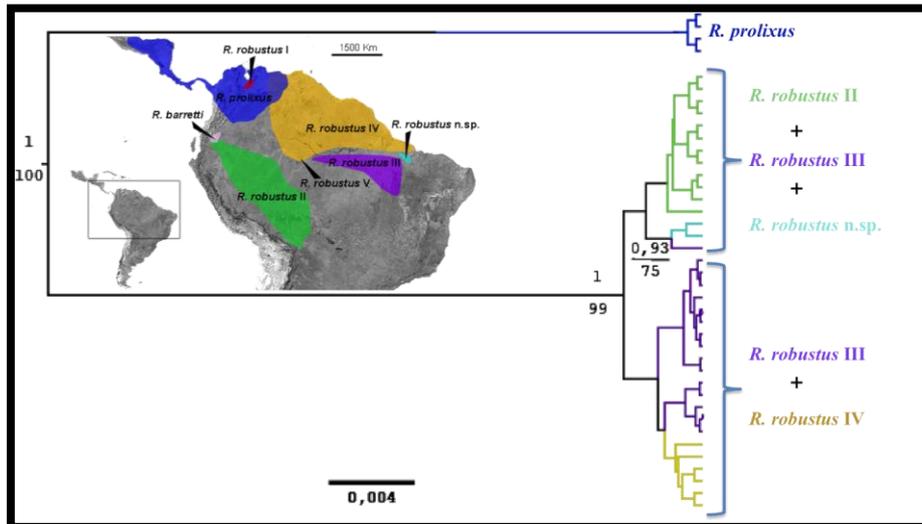
Após os diferentes testes nas reações de PCR para amplificação de amostras de *R. prolixus* e *R. robustus* II, os melhores resultados foram obtidos a partir da utilização, em cada um final de reação de 25µL, das mesmas concentrações de reagentes utilizadas para amplificação do fragmento homólogo de *per*. As amostras foram submetidas à 96 °C por 5 min, seguido de 94 °C por 30 segundos, 65 °C por 2 min e 72 °C por 3min, e ainda à uma extensão final de 72 °C por 20 min.

Quarenta e cinco amostras tiveram o fragmento ortólogo de *tim* sequenciado, sendo cinco *R. prolixus*, 16 *R. robustus* II, 16 *R. robustus* III, dois *R. robustus* n.sp. e seis *R. robustus* IV. Não foi possível obter sequências de *R. robustus* I e V e de *R. barretti*. Trinta e um sítios variáveis e oito *indels* (que variam de 1 a 156-pb) foram observados. O íntron 15 apresenta um polimorfismo de tamanho. As sequências de *R. prolixus* apresentam 210-pb a mais que as de *R. robustus s.l.*

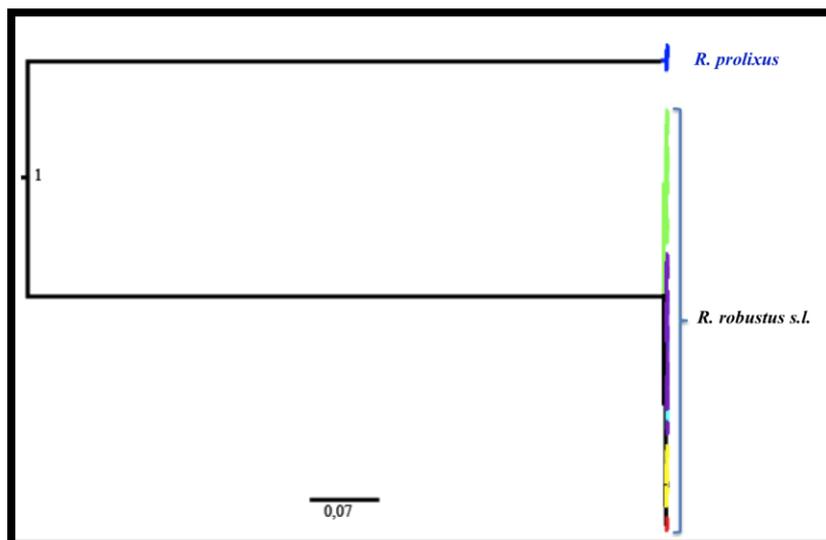
O modelo de substituição nucleotídica adotado nas reconstruções filogenéticas bayesianas e de ML das sequências do ortólogo de *tim* foi o HKY (Hasegawa et al., 1985), selecionado segundo BIC. Nas inferências filogenéticas bayesianas dos *indels*, foi utilizado o modelo estocástico Dollo de multi-estado (Alekseyenko et al., 2008), substituindo os nucleotídeos e os *gaps* de alinhamento por “1” e “0”, respectivamente, em uma matriz binária.

A topologia das árvores bayesiana e de ML (Figura 5.15) obtidas a partir das substituições nucleotídicas mostraram claramente a separação entre *R. prolixus* e *R. robustus s.l.* ( $BS = 100$ ;  $PP = 1$ ). As sequências de *R. robustus s.l.* foram separadas em dois clados ( $BS = 99$ ;  $PP = 1,0$ ). O primeiro clado foi formado por todas as sequências de *R. robustus* II, além de uma sequência de *R. robustus* III, proveniente de Marabá (Pará), e as duas sequências de *R. robustus* n.sp. O segundo clado foi formado pelas amostras restantes de *R. robustus* III e todas as amostras de *R. robustus* IV.

A árvore filogenética bayesiana construída a partir dos *indels* (Figura 5.16) mostrou apenas dois clados bem definidos ( $PP = 1$ ): o primeiro formado por *R. prolixus* e o segundo por *R. robustus s.l.* Desta forma, foi possível concluir que os *indels* do íntron de *tim* analisado são filogeneticamente informativos.



**Figura 5.15.** Árvore filogenética consenso de *tim*.  
 Árvore construída pelo método bayesiano e ML (HKY; *PP*; 1000 replicações de *BP*; raiz da árvore no ponto médio) para as sequências de *tim* de *R. prolixus* e *R. robustus s.l.* Apenas os valores de probabilidade posterior acima de 0,8 e de bootstrap acima de 70 são mostrados. Note que a partir das sequências do ortólogo de *tim* é possível diferenciar apenas *R. prolixus* de *R. robustus s.l.*



**Figura 5.16.** Árvore filogenética dos *indels* de *tim*.  
 Árvore consenso construída a partir do método bayesiano, utilizando o modelo Dollo multiestado (*PP*; raiz da árvore no ponto médio) dos *indels* de *tim* de *R. prolixus* e *R. robustus s.l.* Assim como na reconstrução feita com base nas sequências de *tim* sem os *indels*, é possível diferenciar apenas *R. prolixus* de *R. robustus s.l.*

#### 5.4.4. *cryptochrome1* e *cryptochrome2*

De acordo com as análises anteriores, foram selecionadas sequências de AA de CRY1 de *D. melanogaster* e das proteínas homólogas de *Ae. aegypti* e *Cx. quinquefasciatus* (todas anotadas manualmente e revisadas), além de CRY2 de *Apis*

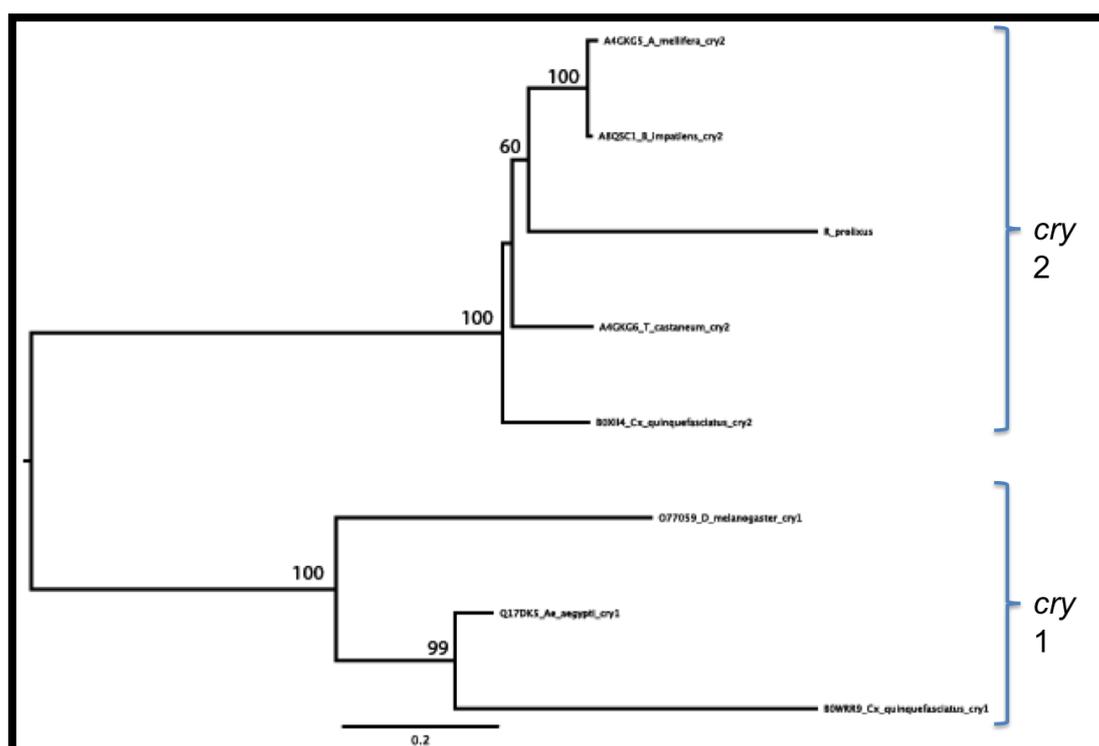
*mellifera* e as proteínas homólogas de *Bombus impatiens*, *Cx. quinquefasciatus* e *T. castaneum* para definir as regiões mais conservadas e exclusivas em cada uma das duas proteínas em insetos. Esta abordagem inicial serviu para explorar primariamente a variabilidade das proteínas homólogas dos diferentes insetos.

As sequências de AA da proteína CRY1 de insetos são muito divergentes entre si (entre 46 e 70% de divergência) e não apresentam uma região claramente conservada. Portanto, a sequência de *D. melanogaster* foi escolhida como sequência-base de CRY1 para a busca no genoma de *R. prolixus* pela proteína homóloga correspondente (por ter sido a primeira espécie que teve a proteína isolada). Dois *hits* foram obtidos na comparação, ambos em uma proteína codificante predita no sentido anti-senso por “C1” e “Lagerblad”, chamada RPTMP15799-RA. O escore e *e-value* do melhor *hit* foi de 227 bits e 1e-59, respectivamente. A proporção de semelhanças encontradas nos aminoácidos analisados foi de 53% (194 de 360 aminoácidos), indicando que RPTMP15799-RA poderia ser homóloga à CRY1.

Ao contrário do observado nas sequências de CRY1, todas as sequências de CRY2 analisadas são muito semelhantes, tanto no tamanho, como na composição de aminoácidos. A sequência mais divergente em termos de tamanho é a de *Cx. quinquefasciatus*, que é composta por 828 aminoácidos (em comparação com aproximadamente 550 AA nas demais proteínas) e diverge em aproximadamente 30% das outras proteínas selecionadas (as outras sequências divergem entre si de 5 e 23%). Devido a semelhança entre as sequências, foi escolhida arbitrariamente a proteína de *A. mellifera* como sequência base de CRY2. O critério estabelecido para esta escolha foi arbitrário, levando-se em consideração simplesmente o fato de *A. mellifera* ter sido a primeira espécie de inseto que teve esta proteína isolada.

Seis *hits* foram encontrados exatamente na mesma proteína predita no genoma de *R. prolixus* apontada como candidata a ser homóloga de CRY1, a RPTMP15799-RA. Contudo, o melhor *hit* nesta comparação apresentou escore e *e-value* muito maiores do que na primeira observação (563 bits e 1e-164). A proporção de semelhanças entre a sequência-base e a proteína predita de *R. prolixus* também foi mais elevada do que na análise anterior, onde 285 dos 348 aminoácidos (84%) comparados foram semelhantes.

O resultado da árvore filogenética confirmou que a proteína predita RPTMP15799-RA no genoma de *R. prolixus* é de fato ortóloga a CRY2. A topologia mostra dois grandes grupos, representados por sequências de CRY1 e CRY2 de insetos. A sequência de *R. prolixus* foi agrupada no segundo grupo (CRY2) e está filogeneticamente mais próxima das sequências de insetos da Ordem Hymenoptera (*A. mellifera* e *B. impatiens*) do que dos insetos das Ordens Coleoptera e Diptera (*T. castaneum* e *Cx. quinquefasciatus*, respectivamente) ( $BS = 84$  e  $100$ ; Figura 5.17).

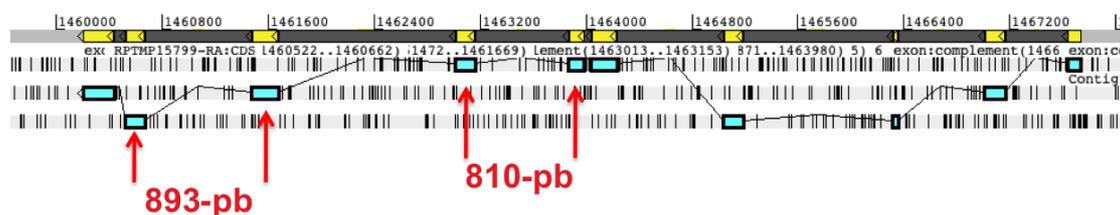


**Figura 5.17.** Árvore filogenética (NJ; Dayhoff; 1000 replicações *bootstrap*) construída com base em sequências de aminoácidos dos homólogos de CRY em Insecta. Apenas as regiões das proteínas que não continham *gaps* de alinhamento foram analisadas. Somente os valores maiores ou iguais a 60 são mostrados. Note que a sequência de AA de *R. prolixus* se insere no agrupamento da árvore composto por outros dos ortólogos de CRY2, provando que o gene isolado do genoma de *R. prolixus* é ortólogo de *cry2*. Em vermelho, o ramo da árvore representado por *R. prolixus*.

A semelhança da proteína predita de *R. prolixus* com CRY1 e CRY2 se deve ao fato de estas proteínas fotorreceptoras pertencerem as mesmas famílias de proteínas, as famílias DNA-fotoliase e ligante-7 de FAD (base de dados Pfam - <http://pfam.sanger.ac.uk/>). Estas famílias protéicas apresentam regiões de domínio semelhantes e, portanto, parte de suas sequências de aminoácidos são similares. Uma nova busca por uma proteína ortóloga de CRY1 na montagem dos supercontigs,

transcritos e ESTs do genoma de *R. prolixus* foi feita utilizando o método TBLASTN, mas nenhum resultado significativo foi encontrado ( $e\text{-value} < 1e\text{-}25$ ).

A proteína RPTMP15799-RA está localizada no Supercontig 17792 do genoma de *R. prolixus* no sentido anti-senso, entre as posições 1460203 e 1467731. As predições feitas por “C1” e “Lagerblad” foram idênticas. A proteína contém 477 AA e é a de menor tamanho em comparação com as outras proteínas dos insetos analisados. O gene homólogo de *cry2* têm 7.529-pb, distribuídos em 10 éxons e 9 íntrons (Figura 5.18).



**Figura 5.18.** Éxons (quadrados em ciano e amarelo) e íntrons (linhas em cinza) do gene ortólogo *cry2* predito no genoma de *R. prolixus*.

Cada uma das três linhas formadas por barras verticais pretas representa um quadro de leitura do éxon (o mais alto representa leitura na primeira base). As setas voltadas para a direita nos retângulos em amarelo representam a leitura do gene no sentido antissenso. As setas em vermelho indicam a posição dos éxons 6 e 7 e ainda 8 e 9 em que os iniciadores foram desenhados, de forma a amplificarem pela PCR fragmentos de 810-pb e 893-pb, contendo os íntrons 6 (717-pb) e 7 (819-pb) por completo, respectivamente.

Como as sequências de aminoácidos do gene *cry2* em insetos são relativamente bem conservadas e as predições da proteína-alvo de *R. prolixus* foram idênticas, levou-se em consideração o tamanho dos íntrons para a escolha das regiões a serem amplificadas (íntrons com mais de 200-pb e menos de 1300-pb). Dois pares de iniciadores foram desenhados. O primeiro par, *cry2\_E6\_F* 5' – GACCTGGAGTT TGGCAAGGT e *cry2\_E7\_R* 5' – ATGGAAGCAGTGATTGCGGA, amplifica o íntron 5 (717-pb) mais 57-pb da região terminal do éxon 5 e 36-pb da região inicial do éxon 6, totalizando 810-pb. O segundo par, *cry2\_E8\_F* 5' – TGGGACAGAAATCCT GAGGC e *cry2\_E9\_R* 5' – ACCTTCTTCGCGTAACTGGG, amplifica o íntron 8 de 809-pb, 48-pb finais do éxon 8 e 36-pb iniciais do éxon 9, totalizando 893-pb.

Após os diferentes testes nas reações de PCR para amplificação de amostras de *R. prolixus* e *R. robustus* II, os melhores resultados foram obtidos a partir da utilização, em cada um final de reação de 25µL, das mesmas concentrações de reagentes utilizadas para amplificação dos fragmentos homólogos de *per* e *tim*. As amostras foram submetidas à 96 °C por 5 min, seguido de 94 °C por 30 segundos, 63 °C por 2 min e 72 °C por 3min, e ainda à uma extensão final de 72 °C por 20 min.

Vinte e cinco amostras tiveram o íntron 6 do ortólogo de *cry2* sequenciado, sendo um *R. prolixus*, 11 *R. robustus* II, quatro *R. robustus* III, três *R. robustus* n.sp. e quatro *R. robustus* IV. Quinze sítios variáveis e três *indels* foram observados. Dois destes *indels* (cada um de 1-pb) são autapomorfias de *R. prolixus*; o outro *indel* é uma autapomorfia de *R. robustus* IV.

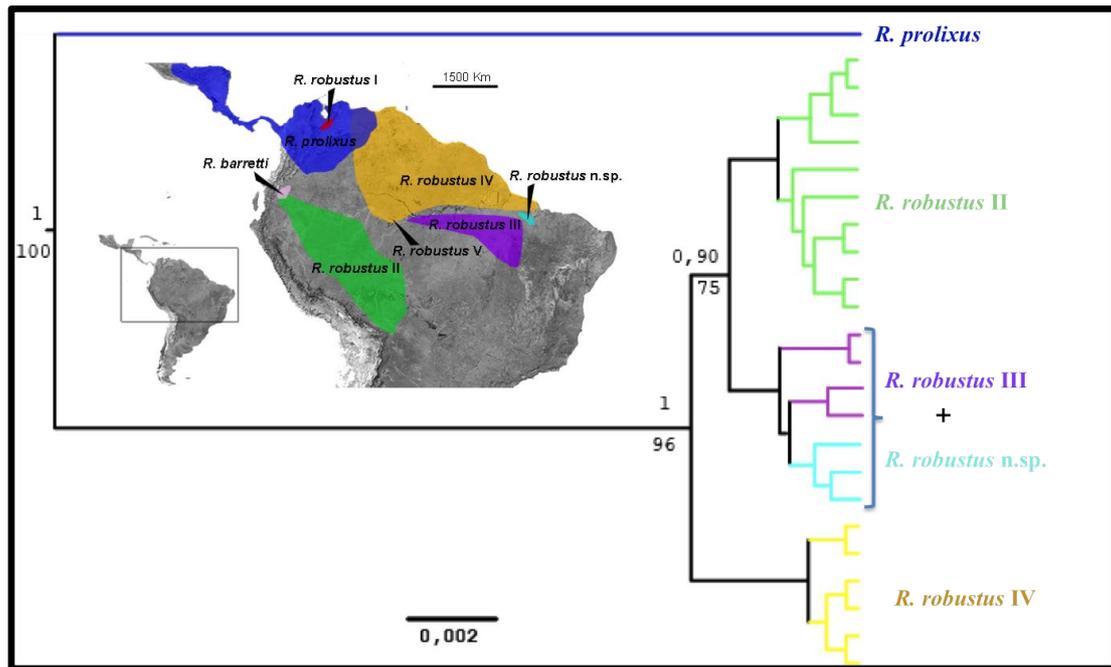
O resultado obtido para o íntron 8 do ortólogo de *cry2* de *R. prolixus* foi semelhante. Quatorze sítios variáveis e quatro *indels* foram observados com base na análise de três sequências de *R. prolixus*, nove de *R. robustus* II, uma de *R. robustus* III, três de *R. robustus* n.sp. e três de *R. robustus* IV. Assim como no íntron 6, os *indels* observados também separam *R. prolixus* de *R. robustus s.l.* (um *indel* de 1-pb) e ainda *R. robustus* IV das demais linhagens de *R. robustus* (um *indel* de 1-pb e outro de 14-pb).

O modelo de substituição nucleotídica adotado nas reconstruções filogenéticas bayesianas e de ML das sequências dos íntrons 6 e 8 do ortólogo de *cry2* foi o HKY+I+Γ e HKY+Γ, respectivamente (Hasegawa et al., 1985). Nas inferências filogenéticas bayesianas, foi utilizado ainda o modelo estocástico Dollo de multi-estado (Alekseyenko et al., 2008) para a análise dos *indels*, substituindo os nucleotídeos e os *gaps* de alinhamento por “1” e “0” em uma matriz binária.

A topologia das árvores bayesiana e de ML obtida a partir das substituições nucleotídicas dos íntrons 6 e 8 de *cry2* (Figuras 5.19 e 5.20) mostraram, assim como para *per* e *tim*, a separação entre *R. prolixus* e *R. robustus s.l.* (*BS* = 100; *PP* = 1). A maior diferença entre as topologias das árvores dos íntrons foi que na primeira (íntron 6) as linhagens de *R. robustus* II e III foram agrupadas separadamente em dois clados (*BS* = 75; *PP* = 0,9), ao passo que na segunda (íntron 8), três espécimes de *R. robustus* II [provenientes de Ouro Preto do Oeste e Porto Velho (Rondônia) e Careiro Castanho (Amazonas)] foram agrupadas no clado de *R. robustus* III. Em ambas as topologias, as

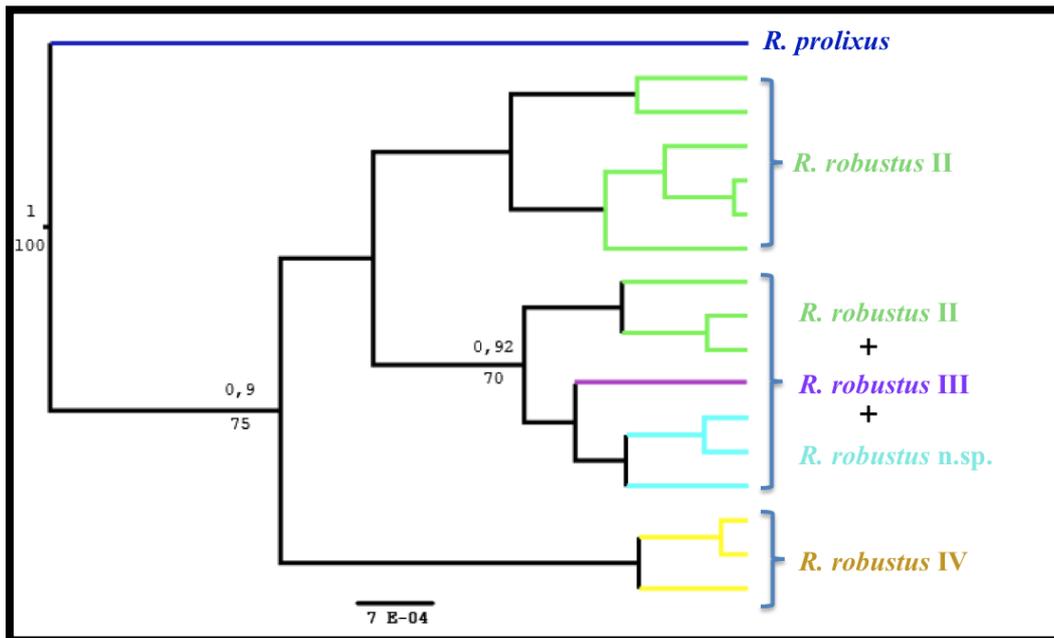
sequências de *R. robustus* IV foram agrupadas em um clado filogeneticamente mais distante ( $BS = 96$  e  $PP = 1$ , e  $BS = 90$  e  $PP = 0,95$ ) e as sequências de *R. robustus* n.sp. agrupadas com *R. robustus* III. A partir destes resultados foi possível concluir que, apesar de estarem localizados em um mesmo gene, os íntrons 6 e 8 apresentam diferenças quanto à informação filogenética gerada. As substituições nucleotídicas do íntron 6 de *cry2* são filogeneticamente mais informativas que as presentes no íntron 8.

A árvore filogenética bayesiana construída a partir dos *indels* do íntron 6 de *cry2* mostrou quatro clados bem definidos (Figura 5.21). Os primeiros três clados foram compostos por (1) *R. prolixus* ( $PP = 1,0$ ), (2) *R. robustus* II ( $PP = 0,79$ ), e (3) *R. robustus* III e *R. robustus* n.sp. ( $PP = 0,11$ ). O quarto clado agrupou as sequências de *R. robustus* IV. Esta apareceu como a linhagem mais divergente do complexo *R. prolixus* e *R. robustus* s.l. Já a topologia da árvore filogenética bayesiana construída a partir dos *indels* do íntron 8 de *cry2* mostrou apenas uma clara separação entre *R. robustus* IV e as demais linhagens de *R. robustus* s.l. e *R. prolixus* (Figura 5.23). É possível concluir com base nestes resultados que os *indels* não contém informações filogenéticas relevantes.



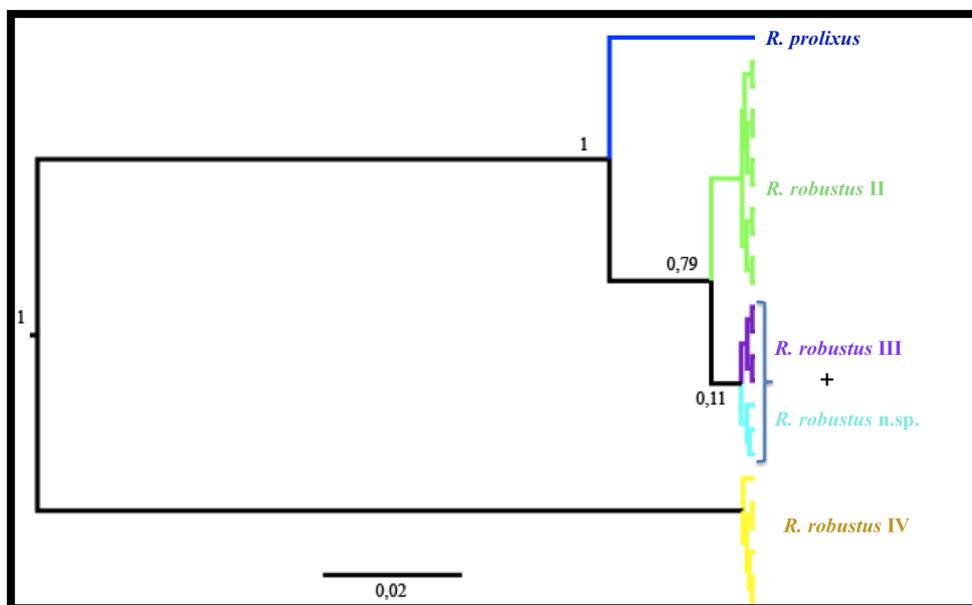
**Figura 5.19.** Árvore filogenética do íntron 6 de *cry2*.

Árvore consenso (método bayesiano; HKY+I+ $\Gamma$ ; PP; 1000 replicações de BP a partir da reconstrução por ML; raiz da árvore no ponto médio) para as sequências do íntron 6 de *cry2* de *R. prolixus* e *R. robustus s.l.* Apenas os valores de probabilidade posterior acima de 0,8 e de bootstrap acima de 70 são mostrados. Note que a partir das sequências do ortólogo de *cry2* é possível *R. prolixus* de *R. robustus s.l.* e ainda as linhagens II, III, IV de *R. robustus* entre si. O único clado não resolvido foi aquele que contém *R. robustus III* e *R. robustus n.sp.*



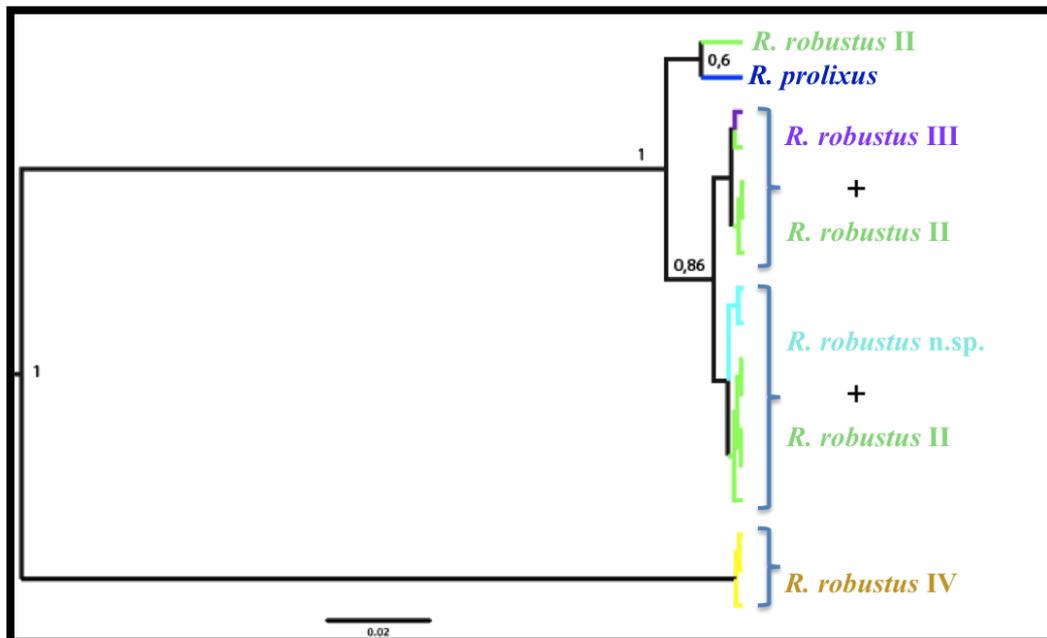
**Figura 5.20.** Árvore filogenética do íntron 8 de *cry2*.

Árvore consenso (método bayesiano; HKY+ $\Gamma$ ; *PP*; 1000 replicações de *BP* a partir da reconstrução por *ML*; raiz da árvore no ponto médio) para as sequências do íntron 8 de *cry2* de *R. prolixus* e *R. robustus s.l.* Apenas os valores de probabilidade posterior acima de 0,8 e de *bootstrap* iguais ou acima de 70 são mostrados. Note que nesta topologia, *R. robustus* II é parafilético. Três espécimes de *R. robustus* II foram agrupados mais próximos de *R. robustus* III e n.sp. do que aos demais espécimes de *R. robustus* II. A partir das sequências do íntron 8 de *cry2* é possível diferenciar não apenas *R. prolixus* de *R. robustus s.l.*, como também a linhagens IV das demais linhagens de *R. robustus*.



**Figura 5.21.** Árvore filogenética dos *indels* do íntron 6 de *cry2*.

Árvore consenso (método bayesiano; Dollo multiestado; *PP*; raiz da árvore no ponto médio) dos *indels* do íntron 6 de *cry2* de *R. prolixus* e *R. robustus s.l.* Ao contrário do observado na reconstrução feita com base nas sequências do íntron 6 de *cry2* não resultantes de *indels*, *R. robustus* II, III e n.sp. estão filogeneticamente mais próximas de *R. prolixus* do que de *R. robustus* IV.



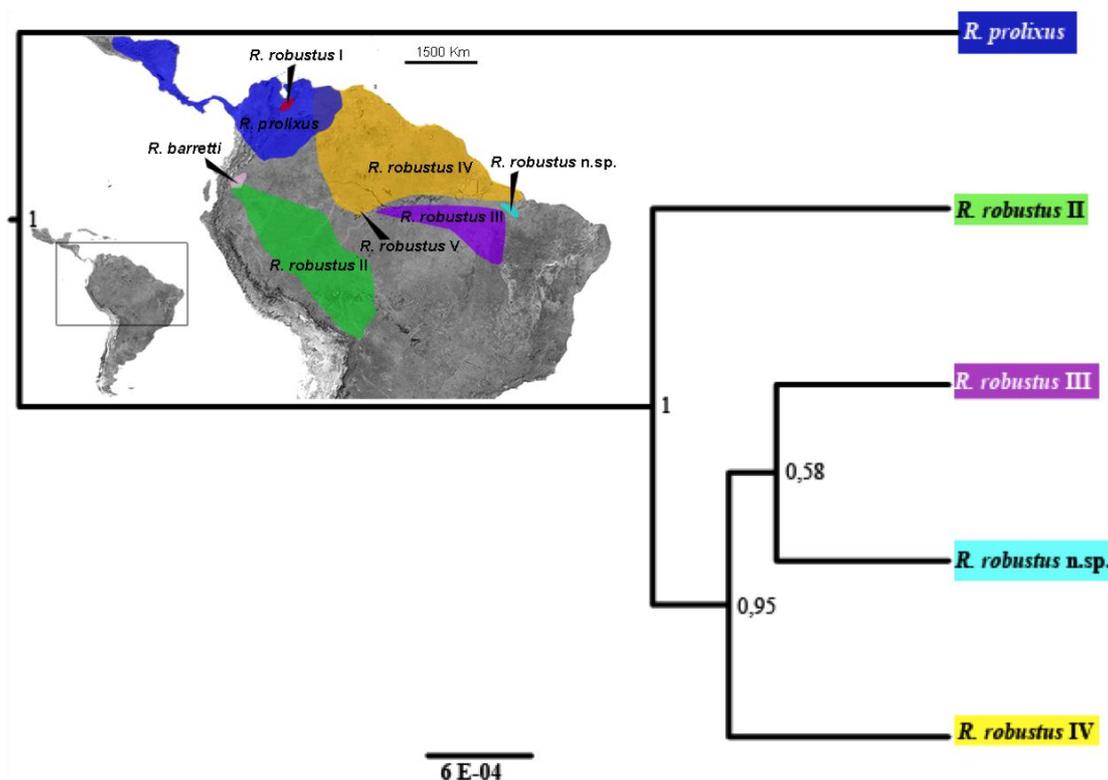
**Figura 5.22.** Árvore filogenética dos *indels* do íntron 8 de *cry2*. Árvore consenso (método bayesiano; Dollo multiestado; *PP*; raiz da árvore no ponto médio dos *indels* do íntron 8 de *cry2* de *R. prolixus* e *R. robustus s.l.* Assim como na reconstrução feita com base nos *indels* do íntron 6 de *cry2*, os *indels* não contêm informações filogenéticas relevantes.

### 5.5. Filogenia de *R. prolixus* e *R. robustus s.l.* com base na análise concatenada dos marcadores nucleares e mitocondrial

Os eventos de cladogênese que resultaram na especiação das linhagens de *R. prolixus* e *R. robustus s.l.* foram inferidos a partir da concatenação das sequências de mtCytb, ITS-2, *per*, *tim* e *cry2*. A partir desta abordagem foi possível inferir se as fontes de discrepância nas relações entre as linhagens de *R. robustus s.l.* observadas nas topologias das árvores filogenéticas de genes individuais são resultado da classificação incompleta da linhagem (*ILS*) ou se estas linhagens representam na realidade populações de uma mesma espécie. Nesta abordagem não foi possível analisar *R. robustus I*, *R. robustus V* e *R. barretti*, pois não havia sequências de todos os marcadores para estas espécies.

O relacionamento entre as espécies na topologia resultante foi idêntico ao observado nas análises mitocondriais (Figura 5.22), por conta do baixo sinal filogenético observado nos demais marcadores. *R. prolixus* é uma espécie distinta de *R. robustus* (*PP* = 1.0) e *R. robustus* representa um complexo de, pelo menos, três espécies distintas (além de *R. robustus I*, *R. robustus V* e *R. barretti*) – *R. robustus II*, III e IV, sendo que *R. robustus II* é a espécie filogeneticamente mais distante (*PP* =

1.0). Apesar de *R. robustus* n.sp. aparecer como uma espécie distinta de *R. robustus* III, o apoio estatístico para esta separação foi baixo ( $PP = 0,58$ ).



**Figura 5.23.** Árvore filogenética consenso de espécies obtida a partir da concatenação das seqüências de mtCytb, ITS-2, *per*, *tim* e *cry2*. A confiança dos agrupamentos foi testada a partir do cálculo da probabilidade posterior. O enraizamento da árvore foi feito no ponto médio. A topologia final foi semelhante aquela obtida a partir das seqüências do mtCytb, com a exceção de que a separação entre *R. robustus* III e n.sp. não foi bem apoiada estatisticamente.

### 5.6. Análises filogeográficas de *R. prolixus* e *R. robustus* s.l.

Como as topologias das árvores filogenéticas obtidas a partir dos marcadores nucleares apresentaram indícios de *ILS*, todas as análises filogeográficas foram realizadas com base apenas nas seqüências de mtCytb.

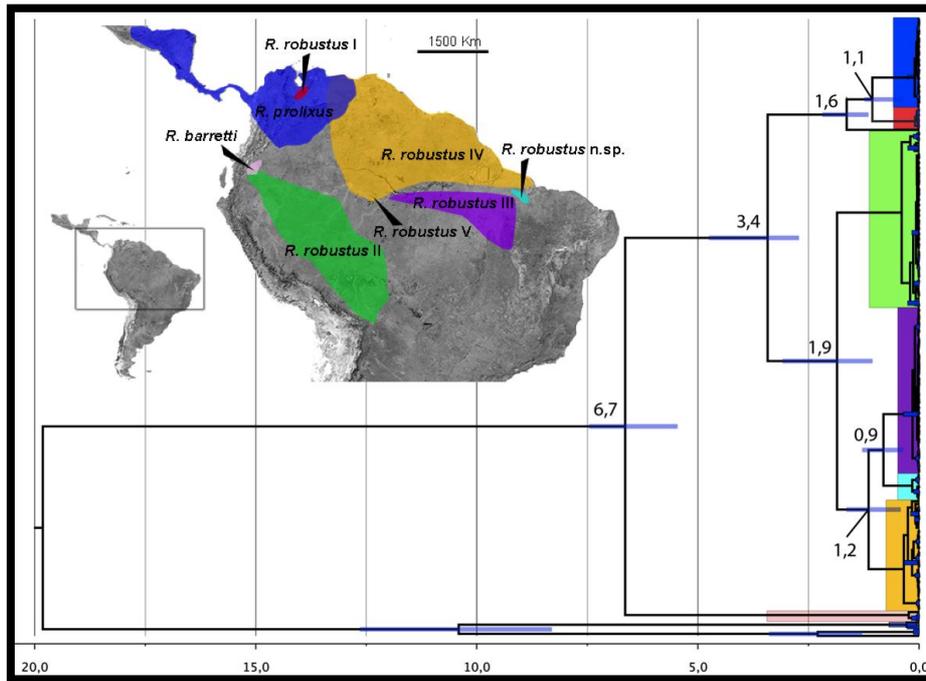
#### 5.6.1. Hipótese do relógio molecular e datação das cladogêneses

A hipótese do relógio molecular fixo foi testada utilizando o método do relógio molecular aleatório (*RLC*), para definir se as taxas evolutivas das linhagens de *R. prolixus* e *R. robustus* s.l. variam na topologia da árvore. A média do número de taxas evolutivas distintas na topologia foi de 2,57 ( $ESS = 3000$ ). Em função deste resultado, foram assumidas taxas independentes nos diferentes ramos das árvores. O

relógio molecular relaxado utilizado foi o “não-correlacionado” (do inglês *uncorrelated*), no qual nenhuma correlação *a priori* é assumida entre as taxas evolutivas da linhagem e de seus ancestrais.

A partir dos quatro tipos de calibração utilizados nos nós das árvores filogenéticas (Tabela 4.5), foi possível observar disparidades nas estimativas de tempo para os eventos de cladogênese que originaram as espécies do complexo *R. prolixus/R. robustus s.l.* Nos dois tipos cujo nó da raiz da árvore foi calibrado (separação entre *R. pictipes* e *R. prolixus/R. robustus s.l.*), as estimativas de tempo para os demais nós da topologia foram aproximadamente duas vezes mais antigas do que as observadas ao utilizar somente calibrações baseadas em eventos geológicos vicariantes (Figura 5.24; Tabela 5.12). As taxas evolutivas estimadas, por consequência, apresentaram diferenças nesta mesma proporção. Considerando as datações mais antigas obtidas nas árvores contendo a calibração no nó da raiz, as linhagens teriam divergido em 0,64-1,38% a cada milhão de anos, em comparações par-a-par. Com base na utilização das árvores filogenéticas que não possuem a calibração no nó da raiz, estas taxas de substituição nucleotídica entre as linhagens analisadas aumentaram para 1,0-2,1%.

Com base na análise de LBF, foi possível observar uma forte evidência de que as topologias que utilizaram as duas calibrações biogeográficas ‘estritas’ para a separação entre *R. prolixus/R. robustus s.l.* e *R. pictipes/R. pallescens s.l.* e ainda entre *R. pallescens* I e II apresentaram as melhores datações para os eventos vicariantes ( $3,8 < \text{LBF} < 4,2$ ).



**Figura 5.24.** Reconstrução filogenética bayesiana contendo as datações dos eventos vicariantes do complexo *R. robustus s.l.* e *R. prolixus*.

A média do tempo de separação em milhão de anos atrás (Maa) é apresentada nos nós da árvore. Os retângulos em azul mostram o intervalo de confiança para as datações dos eventos vicariantes. As cores dos clados correspondem às cores de cada espécie ilustrada no mapa. O intervalo de tempo representado na árvore é mostrado na régua abaixo da topologia, em ordem decrescente (de 20 Maa aos dias de hoje).

A separação da população ancestral em *R. barretti* e *R. prolixus/R. robustus s.l.* teria ocorrido durante o Mioceno Médio/Superior, há aproximadamente 7,5 Ma (12 – 6 Maa). Durante o Plioceno Inferior/Médio (entre 5 e 2,7 Maa), a população ancestral de *R. prolixus*, *R. robustus I* e *R. robustus V* teria se separado da população ancestral de *R. robustus II*, III, IV e nsp.

A Idade do Plioceno Superior/Pleistoceno Inferior foi marcada pela divergência de *R. robustus V* das populações da linhagem ancestral do Orinoco (entre 2,5 e 1,0 Maa). Ainda durante o Pleistoceno Inferior (entre 1,8 e 0,8 Maa), estas populações da linhagem ancestral teriam originado *R. prolixus* e *R. robustus I*.

**Tabela 5.12.** Datação dos eventos vicariantes, em milhão de anos atrás (Maa) calculadas com base nas cinco diferentes estratégias de calibração dos nós das árvores.

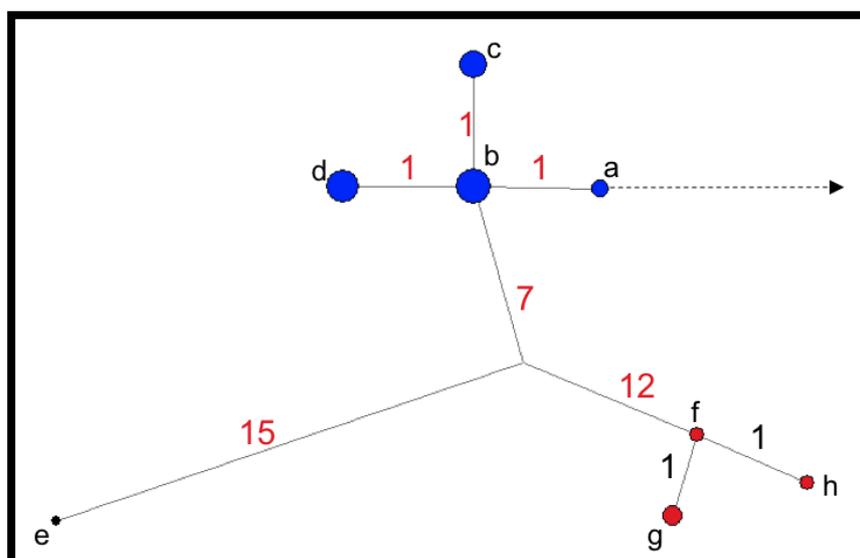
Em parêntesis, o intervalo de confiança (em milhões de anos atrás) de cada evento vicariante.

|   | <i>R. pro</i> e<br><i>R. rob I</i> | <i>R. pro</i> +<br><i>R. rob I</i> e<br><i>R. rob V</i> | Orinoco e<br>Amazônia* | <i>R. rob II</i> e<br><i>R. rob III</i><br>+ IV + nsp | <i>R. rob III</i> e <i>R.</i><br><i>rob IV</i> | <i>R. rob III</i> e <i>R.</i><br><i>rob n.sp.</i> | Complexo <i>R.</i><br><i>pro</i> + <i>R. rob</i><br><i>s.l.</i> e <i>R. bar</i> | <i>R. pro</i> + <i>R.</i><br><i>rob s.l.</i> + <i>R.</i><br><i>barr</i> e <i>R.</i><br><i>pict</i> |
|---|------------------------------------|---|------------------------|---|--|---|---|--|
| <b>1 calibração frouxa</b>  | 1,3<br>(3,0-0,5)                   | 2,0<br>(5,1-0,8)  | 4,3<br>(9,6-1,5)       | 2,3<br>(5,3-0,8)                                      | 1,4<br>(3,3-0,5)                               | 1,0<br>(2,3-0,4)                                  | 8,3<br>(18,0-3,5)   | 24,5<br>(44,0-11,3)  |
| <b>1 calibração estrita</b>   | 1,4<br>(1,8-1,0)                   | 2,2<br>(2,5-1,0)  | 4,5<br>(5,3-3,5)       | 2,3<br>(3,5-1,8)                                      | 1,4<br>(2,0-0,9)                               | 1,0<br>(1,5-0,6)                                  | 8,8<br>(12,0-6,0)   | 22,7<br>(28,5-16,6)  |
| <b>2 calibrações estritas (Andes + raiz)</b>                                      | 2,0<br>(2,7-1,5)                   | 3,2<br>(4,2-2,3)  | 7,1<br>(9,2-5,9)       | 3,7<br>(4,5-2,8)                                      | 2,2<br>(3,0-1,6)                               | 1,5<br>(2,2-1,0)                                  | 13,4<br>(15-10,6)   | 47,7<br>(56,0-39,5)  |
| <b>2 calibrações estritas (Andes + <i>R. pallescens</i>)</b>                      | 1,1<br>(1,5-0,8)                   | 1,6<br>(2,1-1,2)  | 3,4<br>(4,6-2,7)       | 1,9<br>(3,0 - 1,6)                                    | 1,2<br>(1,6-0,9)                               | 0,8<br>(1,2-0,5)                                  | 6,7<br>(7,2-6,0)  | 19,8<br>(23,1-15,2)  |
| <b>3 calibrações estritas (Andes + <i>R. pallescens</i> + <i>R. pictipes</i>)</b> | 1,8<br>(2,3-1,3)                   | 3,0<br>(3,5-2,2)  | 6,5<br>(7,5-5,8)       | 3,5<br>(4,0-2,9)                                      | 2,1<br>(2,8-1,8)                               | 1,4<br>(1,8-0,9)                                  | 13,0<br>(14,5-11,3)   | 47,5<br>(54-40,0)  |

A separação da população ancestral amazônica em dois clados (*R. robustus* II e *R. robustus* III + IV + nsp.) ocorreu possivelmente entre 3,5 e 1,6 Maa, durante o Plioceno Médio/Superior e Pleistoceno Inferior. Durante o Plioceno Superior/Pleistoceno Inferior (entre 2,0 e 0,9 Maa), parte da população ancestral do segundo clado se divergiu, originando *R. robustus* IV. A outra parte da população ancestral teria divergido entre 1,5 e 0,5 Maa, no Pleistoceno Inferior/Médio.

### 5.6.2. Genealogia molecular e testes de expansão populacional súbita

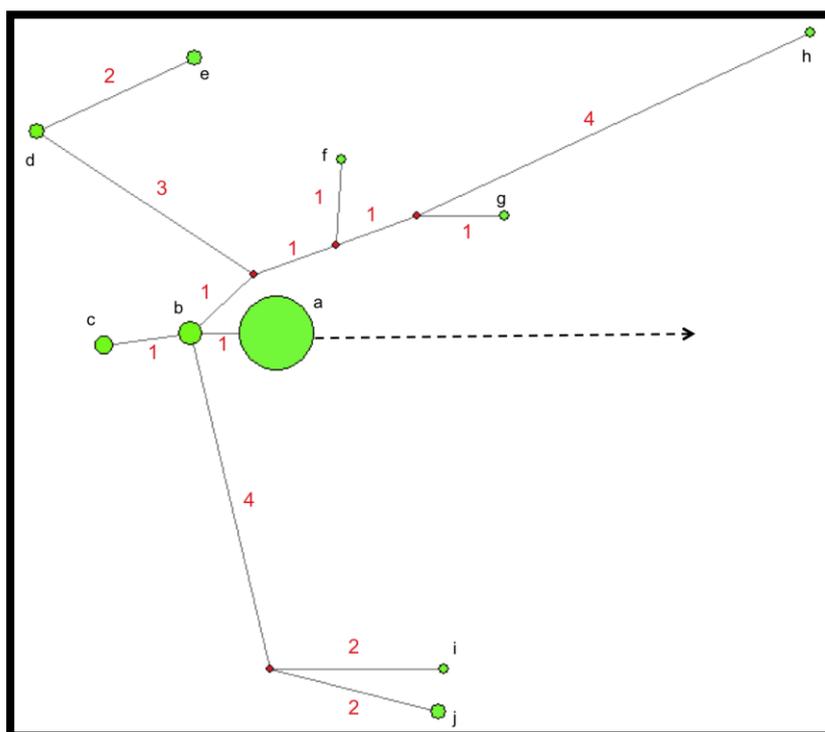
Na rede genealógica contruída a partir das sequências de mtCytb de *R. prolixus*, *R. robustus* I e *R. robustus* V, foi possível inferir que o Orinoco (extremidade nordeste) foi a região biogeográfica de origem deste grupo, uma vez que o grupo externo (*R. barretti*) está mais próximo do haplótipo encontrado em Guárico, Venezuela (Figura 5.25). A região biogeográfica do Imerí (onde *R. robustus* V foi coletada) parece ser um resultado de uma migração secundária da população ancestral. A partir do formato da rede genealógica, é possível observar um padrão de fluxo gênico entre regiões adjacentes, condizente com o modelo de migração por *stepping-stones*. Adicionalmente, as genealogias de *R. prolixus* e *R. robustus* I apresentam “formas de estrela”, uma característica das genealogias de populações que sofreram uma expansão repentina, ou um processo evolutivo de varredura seletiva (do inglês *selective sweep*).



**Figura 5.25.** Rede de haplótipos construída pelo algoritmo de *Median-joining* com base nas sequências de mtCytb de *R. prolixus* (azul) e *R. robustus* I (vermelho).

Cada círculo representa um haplótipo. O tamanho do círculo é proporcional à quantidade de indivíduos que compartilham o mesmo haplótipo. Os números em vermelho representam a quantidade de passos mutacionais entre um haplótipo e outro. As cores dos haplótipos estão de acordo com as utilizadas na Figura 4.3. Note que o haplótipo “a” é o mais próximo do grupo externo (seta com linha pontilhada). O haplótipo “b” está numa posição central da rede e seus ramos parecem formar uma “estrela”. Localidades geográficas que contêm representantes dos haplótipos amostrados: (a) Guárico (Venezuela); (b) Guárico e Cojedes (Venezuela) e Tolima (Colômbia); (c) Francisco Morazan (Honduras) e Guatemala; (d), (f), (g) e (h) Trujillo (Venezuela).

Com base na genealogia de *R. robustus* II (Figura 5.26), foi possível inferir que Guajará-Mirim, Ouro Preto do Oeste ou Porto Velho (Rondônia) representam a área de origem da linhagem. Praticamente todos os haplótipos encontrados nas localidades próximas à possível região de encontro entre *R. robustus* II, III e IV estão representados nas porções terminais da rede genealógica (distantes do centro de origem) e, portanto, representam uma ocupação secundária. A única exceção foi o haplótipo encontrado em Careiro Castanho (Amazonas), que é compartilhado com a amostra proveniente de Santo Antônio (Rondônia) e está no centro da rede genealógica. Este haplótipo, além de outro encontrado em Porto Velho, Ouro Preto do Oeste e Guajará-Mirim apresentam sinais de expansão súbita ou um processo evolutivo de varredura seletiva, uma vez que esta parte da rede genealógica tem forma de “estrela”. Com base nestas observações, é possível inferir que o centro-sul de Rondônia é a região biogeográfica de origem de *R. robustus* II e que a ocupação de Inambarí foi secundária.

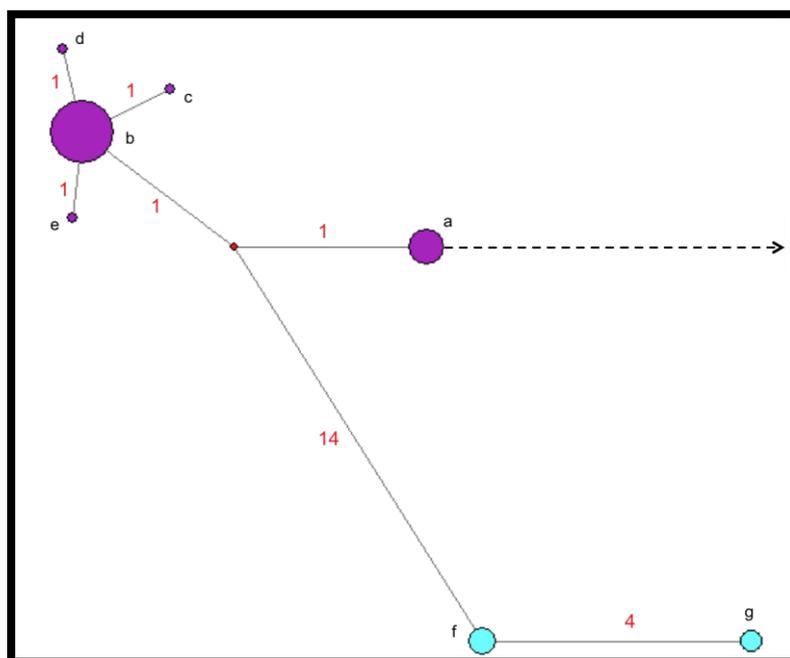


**Figura 5.26.** Rede de haplótipos construída pelo algoritmo de *Median-joining* com base nas sequências de mtCytb de *R. robustus* II.

As cores dos haplótipos estão de acordo com as utilizadas na Figura 4.3. Os pontos em vermelho representam haplótipos ancestrais ou não amostrados. Note que o haplótipo “a” é o mais próximo do grupo externo (seta com linha pontilhada). O haplótipo “b” está numa posição central da rede e seus ramos parecem formar uma “estrela”. Localidades geográficas que contêm representantes dos haplótipos amostrados: (a) Porto Velho, Ouro Preto do Oeste e Guajará-Mirim (Rondônia); (b) Careiro Castanho (Amazonas) e Santo Antonio, Guajará-Mirim e Monte Negro (Rondônia); (c) Apuí e Carauari (Amazonas); (d) e (e) Apuí

(Amazonas); **(f)** Anamã (Amazonas); **(g)** Sucumbios e Napo (Equador) e Loreto (Perú); **(h)** Anamã (Amazonas); **(i)** Porto Velho (Rondônia); **(j)** Guajará-Mirim e Cacoal (Rondônia).

A partir da genealogia de *R. robustus* III (Figura 5.27), foi possível observar que o haplótipo compartilhado por indivíduos coletados em Rondon do Pará e São Domingos do Araguaia (Pará) e ainda em Araguaína e Estrela (Tocantins) foi o mais próximo do grupo externo (*R. robustus* II). Portanto, a extremidade sudoeste do Pará aparece como região biogeográfica mais provável de ocorrência da população ancestral desta linhagem. Adicionalmente, a topologia desta parte da rede genealógica em forma de “estrela” indica rápida e recente expansão populacional ou um processo evolutivo de varredura seletiva. Já a localidade de São Roberto (Maranhão), também pertencente a região biogeográfica do Pará, foi considerada a área de origem de *R. robustus* n.sp. A localidade onde *R. robustus* III e *R. robustus* n.sp. foram encontradas em simpatria (Ulianópolis, Pará) conseqüentemente parece ter sido ocupada após eventos de migração de ambas as espécies.

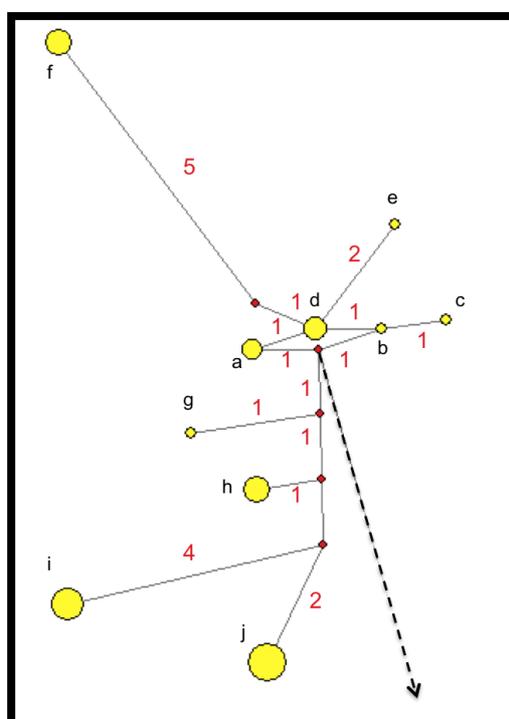


**Figura 5.27.** Rede de haplótipos construída pelo algoritmo de *Median-joining* com base nas seqüências de mtCytb de *R. robustus* III e *R. robustus* n.sp.

As cores dos haplótipos estão de acordo com as utilizadas na Figura 4.3. Os pontos em vermelho representam haplótipos ancestrais ou não amostrados. Note que o haplótipo “a” é o mais próximo do grupo externo (seta com linha pontilhada). Os ramos do haplótipo “b” parecem ter um formato de “estrela”. A localidade de origem estimada de *R. robustus* n.sp. provavelmente foi na região representada pelo haplótipo “f”. Localidades geográficas que contêm representantes dos (cont.) haplótipos amostrados: **(a)** Rondon Pará e São Domingos

do Araguaia (Pará) e Araguaína e Estrela (Tocantins); **(b)** Parauebas, Ulianópolis, Carajás, Novo Repartimento, Itupiranga e Marabá (Pará), São Bento do Tocantins, Estrela e Tocantinópolis (Tocantins); **(c)** Parauebas (Pará); **(d)** São Bento do Tocantins (Tocantins); **(e)** Parauebas (Pará); **(f)** São Roberto (Maranhão); **(g)** Ulianópolis (Pará).

Por fim, uma grande quantidade de haplótipos não amostrados e/ou ancestrais foi observada na genealogia de *R. robustus* IV (Figura 5.28) e por esta razão não foi possível determinar a sua origem com exatidão. Populações contendo um haplótipo ancestral (ou não amostrado) teriam migrado inicialmente para as regiões biogeográficas do Imerí (Manaus, Amazonas) e da Guiana (Mucajaí, Roraima) e posteriormente ocupado as demais regiões da distribuição atual de *R. robustus* IV.



**Figura 5.28.** Rede de haplótipos construída pelo algoritmo de *Median-joining* com base nas sequências de mtCytb de *R. robustus* IV.

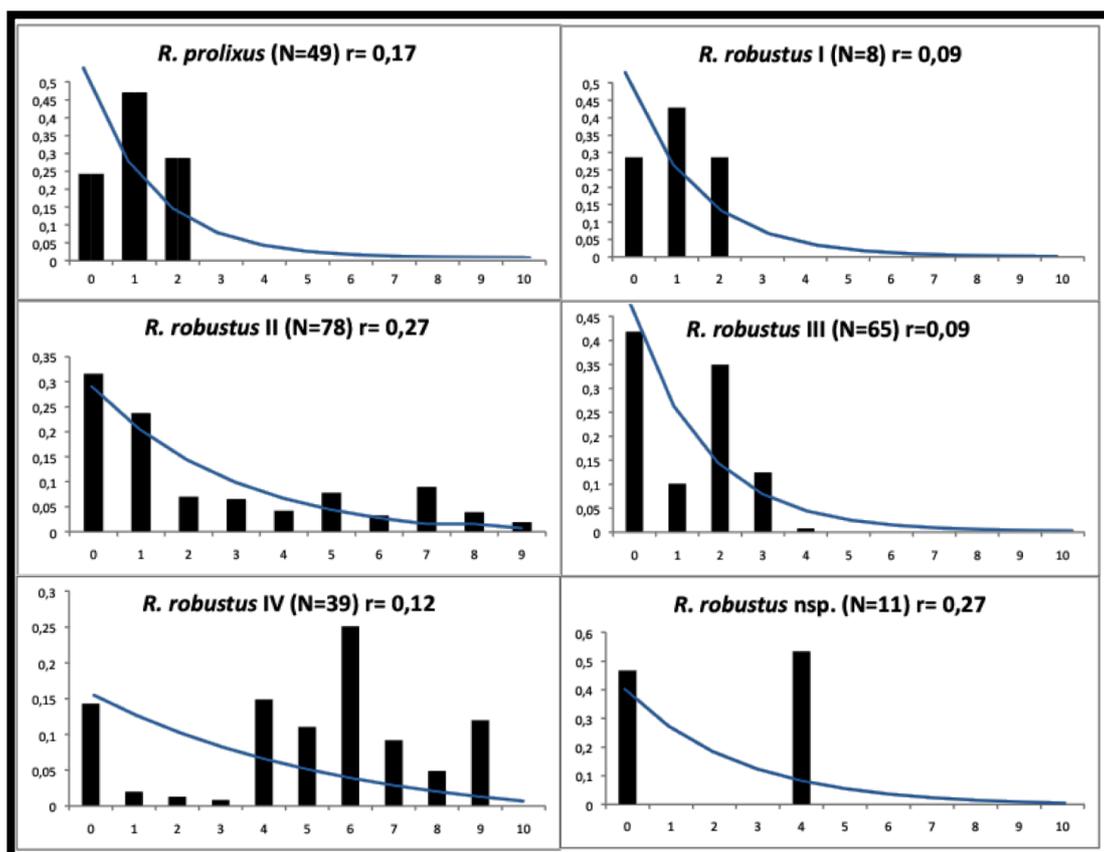
As cores dos haplótipos estão de acordo com as utilizadas na Figura 4.3. Note que o grupo externo (seta com linha pontilhada) está ligado no vetor, impossibilitando hipotetizar com clareza a origem da linhagem. Os pontos mais próximos deste vetor são representados pelos haplótipos “a” e “b”. Localidades geográficas que contêm representantes dos haplótipos amostrados: **(a)** Manaus (Amazonas); **(b)** e **(c)** Mucajaí (Roraima); **(d)** Barcarena (Pará); **(e)** Oriximiná (Pará); **(f)** UHE Paredão (Roraima); (cont.) **(g)** Oriximiná (Pará); **(h)** Santa Maria e Barcarena (Pará); **(i)** Currálinho, Bagre, São José Povoação e Abaetetuba (Pará); **(j)** Guiana Francesa.

Apesar dos indícios de expansão populacional súbita em *R. prolixus* e nas linhagens II e III de *R. robustus*, nenhum dos valores obtidos nos testes de Tajima (D) e Fu foi significativo (Tabela 5.2). Contudo, nos testes de *mismatch distribution*

(Figura 5.29), a frequência de substituições nucleotídicas das sequências de *R. robustus* II seguiu um padrão unimodal ( $r = 0,27$ ) e significativamente semelhante à frequência esperada em populações que sofreram expansão populacional súbita.

Os resultados obtidos nas análises bayesianas de *skyride* indicaram uma fortíssima evidência a favor do modelo de crescimento populacional variável das linhagens de *R. robustus* II, III e IV ao longo do tempo (Fator de Bayes  $> 5$ ; Tabela 5.13). Apenas o gráfico bayesiano de *Skyride* de *R. robustus* II mostrou um súbito crescimento populacional. Sua população aumentou de tamanho 3,25 vezes nos últimos 125 mil anos (Figura 5.30). Estima-se que a linhagem teria aproximadamente 430.000 indivíduos há 125 mil anos e que nos dias de hoje 1.400.000 espécimes habitem a região amazônica. Já as populações efetivas de *R. robustus* III e IV parecem ter permanecido estáveis; nos últimos 125.000 anos, estas populações cresceram de 300.000 para 400.000 e de 600.000 para 700.000 indivíduos, respectivamente.

A partir das análises bayesianas da genealogia de *R. robustus* II, foi possível observar que as áreas de possível contato entre *R. robustus* II, III e IV foram ocupadas há apenas 80.000 anos (dados não mostrados).



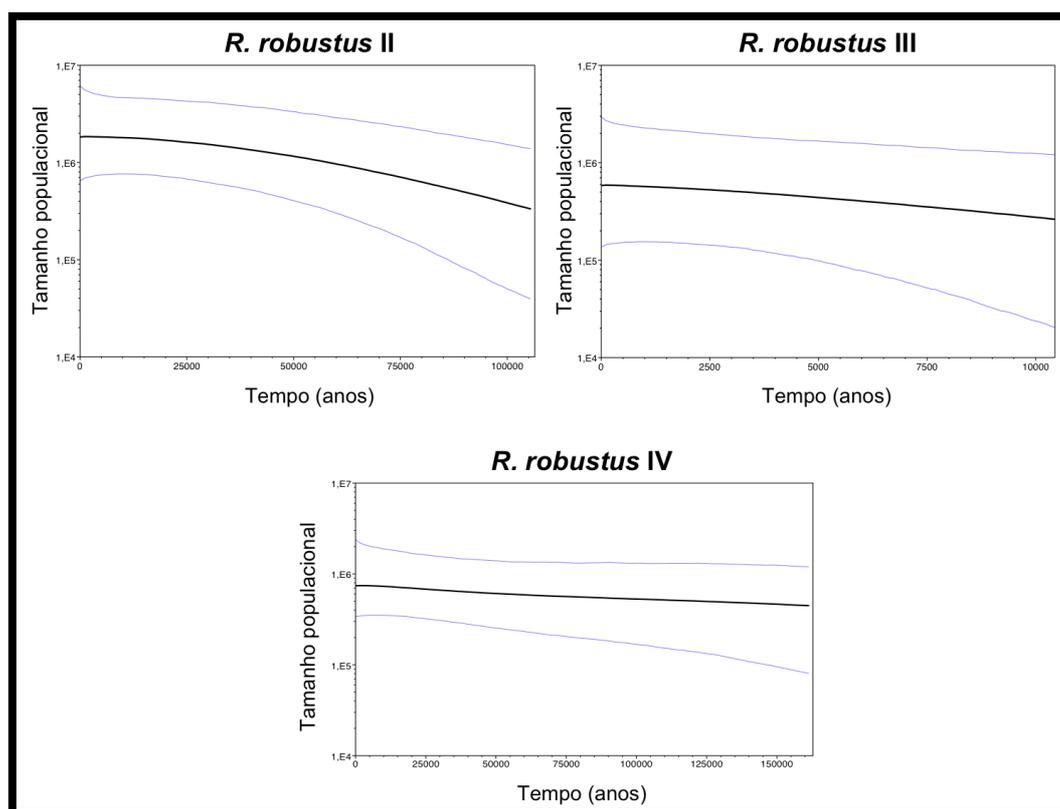
**Figura 5.29.** Mismatch-distribution das amostras de *R. prolixus* e *R. robustus* I-IV.

O eixo X mostra o número de diferenças encontradas entre as seqüências de cada espécie quando comparadas entre si, e o eixo Y mostra os valores das frequências destas diferenças na população. As barras (em preto) representam os valores das frequências observadas, enquanto as linhas em azul representam as frequências esperadas.

**Tabela 5.13.** Valores do Fator de Bayes (BF) obtidos a partir da comparação entre os valores de verossimilhança calculados pelos algoritmos de *pathlikelihood* e *stepping-stone* para os modelos de crescimento populacional constante e variável.

Para todas as linhagens de *R. robustus* testadas, foram indicados com uma fortíssima evidência o modelo de tamanho populacional variável.

| Espécie                | Crescimento populacional variável |                       |             | Crescimento populacional uniforme |                       |             | BF   |
|------------------------|-----------------------------------|-----------------------|-------------|-----------------------------------|-----------------------|-------------|------|
|                        | <i>pathlikelihood</i>             | <i>stepping-stone</i> | mL (média)* | <i>pathlikelihood</i>             | <i>stepping-stone</i> | mL (média)* |      |
| <i>R. robustus</i> II  | -1285,90                          | -1285,92              | -1286,24    | -1253,87                          | -1253,92              | -1252,26    | 34,0 |
|                        | -1286,54                          | -1286,61              |             | -1250,60                          | -1250,66              |             |      |
| <i>R. robustus</i> III | -1028,64                          | -1028,72              | -1031,85    | -1026,74                          | -1026,80              | -1026,79    | 5,1  |
|                        | -1034,98                          | -1035,04              |             | -1026,77                          | -1026,84              |             |      |
| <i>R. robustus</i> IV  | -1170,08                          | -1170,14              | -1169,98    | -1150,86                          | -1150,96              | -1145,64    | 24,3 |
|                        | -1169,83                          | -1169,88              |             | -1140,33                          | -1140,39              |             |      |



**Figura 5.30.** Gráfico bayesiano *Skyride* mostrando que apenas *R. robustus* II apresenta indícios de expansão demográfica súbita.

Nos últimos 125 mil anos, sua população efetiva cresceu de 430.000 para 1.400.000 de espécimes.

### 5.6.3. Hipóteses biogeográficas de dispersão de *R. prolixus* e *R. robustus* s.l. nas regiões Amazônica e do Orinoco

De acordo com os resultados obtidos nas genealogias, as populações ancestrais que teriam originado *R. prolixus*, *R. robustus* I e *R. robustus* V foram consideradas como provenientes da região do Orinoco. A população ancestral que deu origem à *R. barretti* foi considerada como natural da região biogeográfica do Napo e as demais populações ancestrais que originaram as outras linhagens de *R. robustus* foram consideradas como pertencentes à Amazônia.

Com base na estimativa de ML (MLE) dos parâmetros do modelo F84 de substituição nucleotídica, foi possível observar que existem diferenças quanto a distribuição dos sítios variáveis ao longo das sequências (LRT = 127,3, P = 1,0E-06) e, portanto, o modelo gama deveria ser utilizado. O valor do parâmetro de formato da distribuição gama foi alto ( $\alpha = 0,17$ ), o que significa que há uma grande proporção de sítios variáveis. Este resultado reflete a alta divergência entre as sequências analisadas.

As cadeias de MCMC visualmente permaneceram estáveis e não apresentaram nenhuma tendência. Todos os valores de ESS obtidos foram maiores que  $1,0 \times 10^5$  e o número de genealogias amostradas foi maior que 20%, indicando boa mistura das cadeias e convergência dos parâmetros estimados.

Nas análises em macroescala, a hipótese biogeográfica de dispersão com maior probabilidade de ser verdadeira foi a *source-sink* 1 (Probabilidade = 99%; Tabela 5.14), na qual a população ancestral que originou *R. prolixus* e todas as linhagens de *R. robustus* s.l. teria surgido na região do Orinoco. Posteriormente, esta população teria contribuído como fonte constante de migrantes para as regiões biogeográficas do Napo e da Amazônia (Guiana, Inambarí, Imerí, Rondônia, Pará e Belém). Este modelo apresentou mais do que 5 unidades de log ML maiores do que a hipótese biogeográfica ranqueada em segundo lugar, a *source-sink* 3, na qual admitiu-se que a população ancestral teria sido originada na região do Napo e servido como fonte de migrantes para o Orinoco e que, uma vez estabelecida, esta última teria sido a fonte de migrantes para as áreas biogeográficas adjacentes à Amazônia.

**Tabela 5.14.** Valores marginais dos logaritmos de verossimilhança dos modelos biogeográficos em macroescala, calculados segundo o modelo harmônico e corrigidos pelo método de Bezier.

A partir do cálculo do fator bayesiano logarítmico (LBF), os modelos foram ranqueados de acordo com a probabilidade de cada um estar correto. O modelo escolhido com probabilidade de 99% é destacado em negrito.

| Modelo                      | lmL Bezier      | lmL Harmônico   | LBF (Bezier) | Probabilidade modelo* | Escolha (Bezier)** |
|-----------------------------|-----------------|-----------------|--------------|-----------------------|--------------------|
| Migração completa           | -3830,18        | -3534,08        | -379,4       | 1,69E-165             | 6                  |
| <i>Stepping-stone</i> 1     | -3466,59        | -3057,79        | -15,81       | 1,36E-07              | 3                  |
| <b><i>Source-sink</i> 1</b> | <b>-3450,78</b> | <b>-3054,91</b> | <b>0</b>     | <b>0,99</b>           | <b>1</b>           |
| <i>Stepping-stone</i> 2     | -3602,69        | -3374,7         | -151,91      | 1,06E-66              | 4                  |
| <i>Source-sink</i> 2        | -3458,12        | -3053,2         | -7,34        | 6,49E-04              | 2                  |
| <i>Source-sink</i> 3        | -3638,28        | -3414,2         | -187,5       | 3,71E-82              | 5                  |
| Panmixia                    | -5407,45        | -4203,36        | -1956,67     | < 1,00E-250           | 8                  |
| Isolamento completo         | -4015,25        | -3623,24        | -564,47      | 7,14E-246             | 7                  |

\* Probabilidade do modelo estar correto.

\*\* Escolha do modelo, com base na comparação dos valores logarítmicos marginais de verossimilhança corrigidos pelo método de Bezier.

As análises em mesoescala revelaram que os modelos que melhor explicam a dispersão das populações ancestrais de *R. robustus s.l.* para as regiões biogeográficas da Guiana, Inambarí, Imerí, Rondônia, Pará e Belém são *stepping-stone* 2 e *source-sink* 2 (Probabilidades de 58,7% e 39,3%, respectivamente; (Tabela 5.15). A origem das linhagens amazônicas de *R. robustus* II, III, IV e n.sp. teria acontecido em uma região de ocorrência de *R. robustus* IV. Estas populações ancestrais teriam sido uma fonte de migrantes para as demais regiões biogeográficas ou teriam se expandido para as regiões de ocorrência de *R. robustus* II e *R. robustus* III. Finalmente, as populações ancestrais da região de ocorrência de *R. robustus* III teriam migrado ou expandido para as regiões de ocorrência de *R. robustus* n.sp.

**Tabela 5.15.** Valores marginais dos logaritmos de verossimilhança dos modelos biogeográficos em mesoescala, calculados segundo o modelo harmônico e corrigidos pelo método de Bezier.

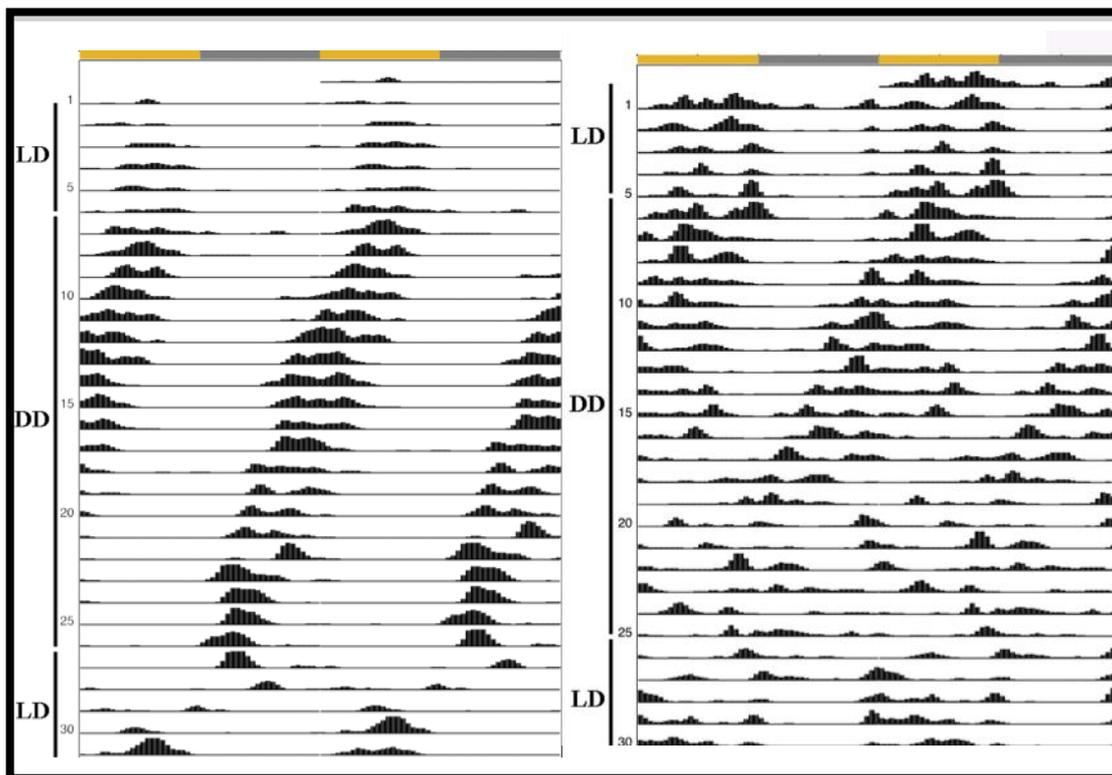
A partir do cálculo do fator bayesiano logarítmico (LBF), os modelos foram ranqueados de acordo com a probabilidade de cada um estar correto. O modelo escolhido com probabilidade de 99% é destacado em negrito.

| Hipótese              | lmL Bezier      | lmL Harmônico   | LBF (Bezier) | Probabilidade modelo* | Choice (Bezier)** |
|-----------------------|-----------------|-----------------|--------------|-----------------------|-------------------|
| <b>Source-sink 2</b>  | <b>-1571,49</b> | <b>-1529,42</b> | <b>0</b>     | <b>0,59</b>           | <b>1</b>          |
| Stepping-stone 2      | -1571,89        | -1529,08        | -0,4         | 0,39                  | 2                 |
| Stepping-stone 1      | -1575,25        | -1527,23        | -3,76        | 0,01                  | 3                 |
| Source-sink 1         | -1576,4011      | -1532,11        | -4,9111      | 4,31E-03              | 4                 |
| Migração completa     | -1577,01        | -1527,67        | -5,52        | 2,35E-03              | 5                 |
| Panmixia              | -1582,98        | -1528,54        | -11,49       | 6,00E-06              | 6                 |
| Isolamento completo 1 | -1596,83        | -1540,83        | -25,34       | 5,80E-12              | 7                 |
| Isolamento completo 2 | -1668,38        | -1635,94        | -96,89       | 4,89E-43              | 8                 |

### 5.7. Atividade locomotora de *R. prolixus* e *R. robustus* II

*R. prolixus* e *R. robustus* II apresentaram um relógio circadiano funcional, uma vez que mantiveram um ritmo endógeno de atividade diário, mesmo na ausência de sinais externos (luz) e foram capazes de ajustar o padrão do ritmo mesmo após 20 dias sem sinais externos de dia e noite (exemplo na Figura 5.31).

A distribuição dos valores das proporções das atividades diurnas e noturnas foram “normalizadas”, utilizando-se os valores de arco seno da raiz quadrada da proporção de atividade diurna e noturna. A normalidade foi avaliada através do emprego dos testes de Komolgorov-Smirnov ( $d=0,064$ ; Lilliefors  $p=0,06$ ) e Shapiro-Wilk ( $W=0,988$ ;  $p=0,112$ ).



**Figura 5.31.** Atividade locomotora de ninfas de triatomíneos.

Neste caso, foram utilizadas como exemplo uma ninfa N3 de *R. robustus* II em 31 dias de experimento (esquerda) e uma ninfa N3 de *R. prolixus* em 30 dias de experimento (direita). Cada linha representa um dia. As primeiras linhas (1 a 6) mostram os 6 dias de LD 12:12, as 20 seguintes representam os dias de DD e as últimas cinco, os dias LD 12:12. Em DD, os dias são subjetivos, uma vez que não há regime de luz. Note que a atividade das ninfas é maior durante o dia. Como esperado, *Rhodnius* tem um relógio funcional, pois apresenta um padrão rítmico de atividade, mesmo na ausência de sinais externos (luz). Quando os espécimes são submetidos a 24h de escuridão, note que o pico de atividade (colunas mais altas) é cada dia mais cedo, o que mostra que seu relógio endógeno é inferior a 24h. Ao serem submetidos a regimes de LD novamente (linhas 26 a 31), note que nos dias 30 e 31 a atividade passa a ser sincronizada novamente ao ciclo ambiental após um período de transição. Ambas as espécies possuem dois picos de atividade, um no meio de LD (ZT 7) e outro próximo ao fim de LD (ZT 12).

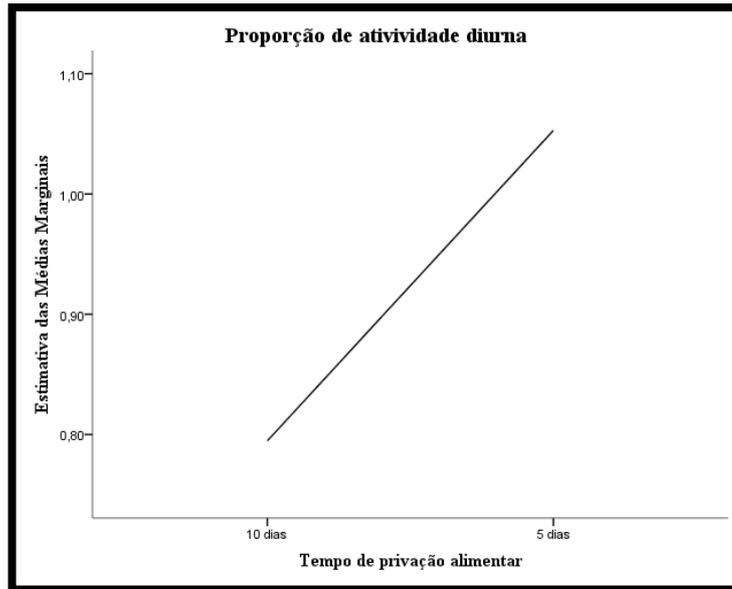
Os resultados da atividade locomotora mostraram que as ninfas de *R. prolixus* e *R. robustus* II apresentaram maior atividade durante o dia ( $t$ -Student = -4,64,  $P < 0,0001$ ,  $gl = 188$ ). Os picos de atividade de ambas as espécies ocorrem no meio da fotofase (fase clara) e na transição claro/escuro, pouco antes das luzes serem apagadas. Estes picos estão localizados aproximadamente em ZT 7 e ZT12 (o que corresponderia as 13h e 18h, respectivamente; Figura 5.32).

A modelagem da análise de variância de quatro sentidos (do inglês *four-way* ANOVA) foi significativa ( $F = 10,31$ ,  $P < 0,0001$ ,  $gl = 14$ ) (Tabela 5.16). Esta análise revelou que as atividades locomotoras de ambas as espécies foi menor quando as

espécies foram submetidas a mais tempo de privação alimentar ( $F = 52,94$ ,  $P < 0,0001$ ,  $gl = 1$ ; Figura 5.32).

**Tabela 5.16.** Análise Four-way ANOVA, comparando a presença/ausência de abrigo, tempo de privação de alimento, espécie e estágio ninfal. Note que as únicas fontes de comparação significativas são “alimentação”, “abrigo+alimentação” e “abrigo+estádio+espécie”.

| Fonte   | Soma dos<br>Quadrados Tipo III | Graus de<br>liberdade | Média dos<br>quadrados | F      | P         |
|---|--------------------------------|-----------------------|------------------------|--------|-----------|
| Modelo 4-Way ANOVA                              | 3,881                          | 14                    | ,277                   | 10,305 | < 0,0001* |
| Univariado                                      |                                |                       |                        |        |           |
| Presença de abrigo                              | 0,019                          | 1                     | 0,019                  | 0,72   | 0,40      |
| Estádio ninfal                                  | 0,014                          | 1                     | 0,014                  | 0,52   | 0,47      |
| Espécie   | 0,000                          | 1                     | 0,000                  | 0,02   | 0,89      |
| Privação alimentar                              | 1,424                          | 1                     | 1,424                  | 52,94  | < 0,0001* |
| Abrigo * Estádio                                | 0,025                          | 1                     | 0,025                  | 0,93   | 0,34      |
| Abrigo * Espécie                                | 0,017                          | 1                     | 0,017                  | 0,63   | 0,43      |
| Abrigo * Priv. alimentar                        | 0,226                          | 1                     | 0,226                  | 8,39   | 0,004*    |
| Estádio * Espécie                               | 0,018                          | 1                     | 0,018                  | 0,68   | 0,41      |
| Estádio * Priv. alimentar                       | 0,022                          | 1                     | 0,022                  | 0,82   | 0,37      |
| Espécie * Experimento                           | 0,016                          | 1                     | 0,016                  | 0,59   | 0,44      |
| Abrigo * Estádio *<br>Espécie                   | 0,114                          | 1                     | 0,114                  | 4,24   | 0,04*     |
| Abrigo * Estádio * Priv.<br>alimentar           | 0,004                          | 1                     | 0,004                  | 0,15   | 0,70      |
| Abrigo * Espécie * Priv.<br>alimentar           | 0,029                          | 1                     | 0,029                  | 1,07   | 0,30      |
| Estádio * Espécie * Priv.<br>alimentar          | 0,006                          | 1                     | 0,006                  | 0,23   | 0,63      |
| Abrigo * Estádio *<br>Espécie * Priv. alimentar | 0,000                          | 0                     | .                      | .      | .         |

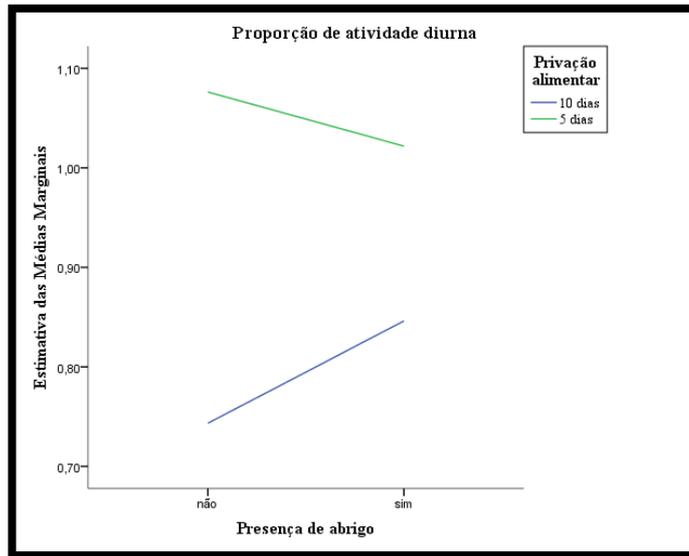


**Figura 5.32.** Médias marginais estimadas do arco seno da raiz quadrada das proporções de atividade diurna nas condições LD 12:12.

Nota-se que quanto mais tempo privados de alimentação, menor a atividade locomotora.

Quando privadas de alimentação sanguínea por cinco dias, *R. prolixus* e *R. robustus* II se movimentaram mais nos tubos que não continham abrigos à incidência direta de luz. O contrário foi observado quando esta privação perdurou por 10 dias, onde estas espécies foram mais ativas nos tubos contendo abrigos (Figura 5.33).

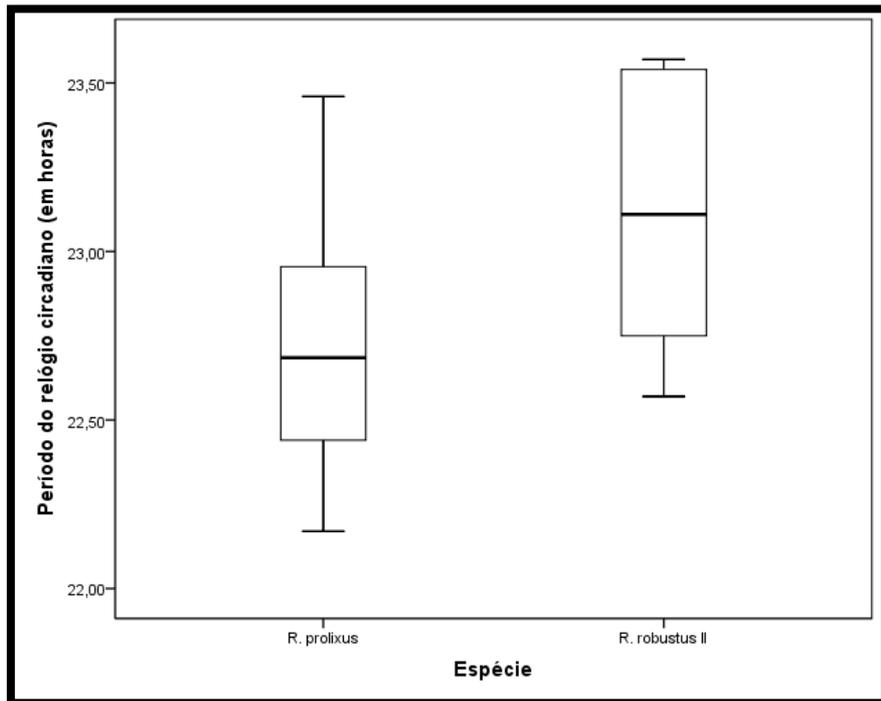
As ninfas N3 de *R. prolixus* permaneceram mais ativas nos tubos que continham abrigos, ao passo que as ninfas N2 se movimentaram mais quando não havia qualquer proteção à incidência direta de luz. O oposto foi observado em ninfas de *R. robustus* II. As ninfas N2 permaneceram mais ativas nos tubos que continham abrigos, ao passo que as ninfas N3 se movimentaram mais quando não havia qualquer proteção à incidência direta de luz.



**Figura 5.33.** Médias marginais estimadas do arco seno da raiz quadrada das proporções de atividade diurna nas condições LD 12:12 em comparação com a presença e ausência de abrigo.

Quando privados de alimentação por mais tempo (10 dias), *R. prolixus* e *R. robustus* II se movimentaram mais nos tubos contendo abrigos do que nos tubos com incidência direta de luz. O contrário foi observado quando privados de alimentação por menos tempo (5 dias).

Os períodos ( $\tau$ ) dos relógios circadianos de *R. prolixus* e *R. robustus* II são visivelmente menores que 24 horas, por conta da gradativa antecipação diária dos picos de atividade na ausência de indicadores ambientais (fase escura ou DD; Figura 5.31). A partir do cálculo do período do relógio circadiano de cada um dos 23 indivíduos de *R. prolixus* e dos 15 *R. robustus* II, foi possível observar uma alta variação em torno da mediana dos valores encontrados, sobretudo em *R. robustus* II (Figura 5.34). Apesar desta variação, a diferença entre os períodos dos relógios circadianos de *R. robustus* II ( $\tau = 23,2$  horas) e *R. prolixus* ( $\tau = 22,7$  horas) é significativa (*One Way ANOVA* -  $F = 4,31$ ,  $P = 0,04$ ,  $gl = 1$ ).



**Figura 5.34.** Diagrama de caixas dos períodos do relógio circadiano de 23 *R. prolixus* e 15 *R. robustus II*.

Note que o segundo quartil do diagrama de *R. prolixus* se sobrepõe ao primeiro quartil do diagrama de *R. robustus II*. Contudo, a mediana dos períodos dos relógios não se sobrepõem ( $\tau_{R. prolixus} = 22,7$  horas e  $\tau_{R. robustus} = 23,2$  horas).

## 6. DISCUSSÃO

### 6.1. Análises filogenéticas e identidades taxonômicas de *R. prolixus* e *R. robustus* s.l.

Desde o início da década de 80, marcadores moleculares mitocondriais vêm sendo utilizados com exclusividade para resolver questões taxonômicas envolvendo espécies filogeneticamente próximas (Ferris et al., 1983). Por ser uma molécula de cópia única, de herança uniparental (matrilinear) e presente em abundância nas células somáticas e germinativas dos animais, o mtDNA não sofre recombinação [salvo raras exceções; (Rokas et al., 2003)] e pode ser sequenciado diretamente, sem a necessidade de empregar métodos laboriosos como a clonagem artificial em vetores biológicos. Ademais, a ineficiência dos mecanismos de reparo do DNA na organela mitocondrial contribui para a ocorrência de altas taxas de substituição nucleotídica (i.e. mutação) nesta molécula.

Ao analisar a topologia da árvore de um gene mitocondrial de linhagens proximamente relacionadas, deve ser levado em consideração que o caráter uniparental do marcador impede a identificação de possíveis eventos de introgressão do material genético de uma espécie em outra (Chan e Levin, 2005). Nestes casos, a utilização de marcadores nucleares de cópia única (i.e. de origem biparental) é imprescindível.

Os íntrons dos genes codificantes de proteínas são bons candidatos para complementar as informações genéticas contidas no mtDNA acerca do relacionamento filogenético entre espécies próximas, por estarem em regiões do genoma nuclear que evoluem rapidamente e de maneira neutra (Irimia e Roy, 2008). Contudo, é importante salientar que cada *locus* evolui de maneira independente (quando não estão ligados) e processos de recombinação podem afetar o sinal filogenético contido no DNA nuclear. Diferentemente do genoma mitocondrial, alelos distintos de um único gene ou mesmo de vários genes que evoluem de maneira independente podem ser agrupados por recombinação.

Infelizmente não havia um marcador já descrito em triatomíneos que fosse de origem biparental, de cópia única e de evolução rápida (como é o caso de muitos íntrons). Praticamente todas as publicações envolvendo análises filogenéticas de complexos de espécies crípticas em triatomíneos utilizaram o segundo espaçador

intergênico ribossomal (ITS-2), que constitui uma família de genes com múltiplas cópias (e.g. Bargues et al., 2006; Marcilla et al., 2001). Apesar de as cópias de ITS-2 nas espécies analisadas dos gêneros *Triatoma* e *Panstrongylus* aparentemente estarem completamente homogeneizadas pelo mecanismo evolução em concerto (Bargues et al., 2006; Marcilla et al., 2002; Martinez et al., 2006), em *Rhodnius* isso parece não ocorrer com a mesma eficiência (Lazoski, Pavan e Monteiro, dados não publicados).

A eficiência da homogeneização do rDNA por evolução em concerto é normalmente alta nos organismos de maneira geral, desde bactérias e fungos até mamíferos (Ganley e Kobayashi, 2007; Liao 1999). Esta homogeneização garante a ortologia das cópias presentes em um indivíduo e, portanto, sua utilidade em estudos filogenéticos. Contudo, a utilização destas regiões do rDNA em estudos populacionais deve ser feita com cautela. Como fruto da homogeneização das sequências e fluxo gênico entre as populações de uma espécie (ou à sua falta), a divergência intraespecífica pode ser aumentada. Fenômenos como *crossing-over* desigual e conversões gênicas afetam o número de cópias variáveis dos minissatélites contidos no genoma, gerando um padrão de segregação não Mendeliano (Bower et al., 2008). Estas trocas de material genético podem ocasionar a inserção ou deleção de regiões de repetições. Este ciclo ocorre muito mais rápido do que a taxa de mutação pontual (i.e. não evolui de forma neutra), o que restringe o uso desses marcadores em estudos populacionais (Dover, 2000).

Marcadores nucleares provenientes de famílias de multigenes tampouco são úteis em estudos filogenéticos quando as taxas de mutação dos *loci* são maiores do que as taxas de homogeneização por evolução em concerto. Desta forma, a variação das sequências intraindividuais podem ser maiores do que as observadas em comparações interpopulacionais e interespecíficas e o sinal filogenético resultante é pouco claro (Wilkerson et al. 2004; Onyabe e Conn 1999; Li e Wilkerson 2007). Nestes casos, não há como garantir a ortologia das cópias de rDNA sequenciadas de cada indivíduo estudado.

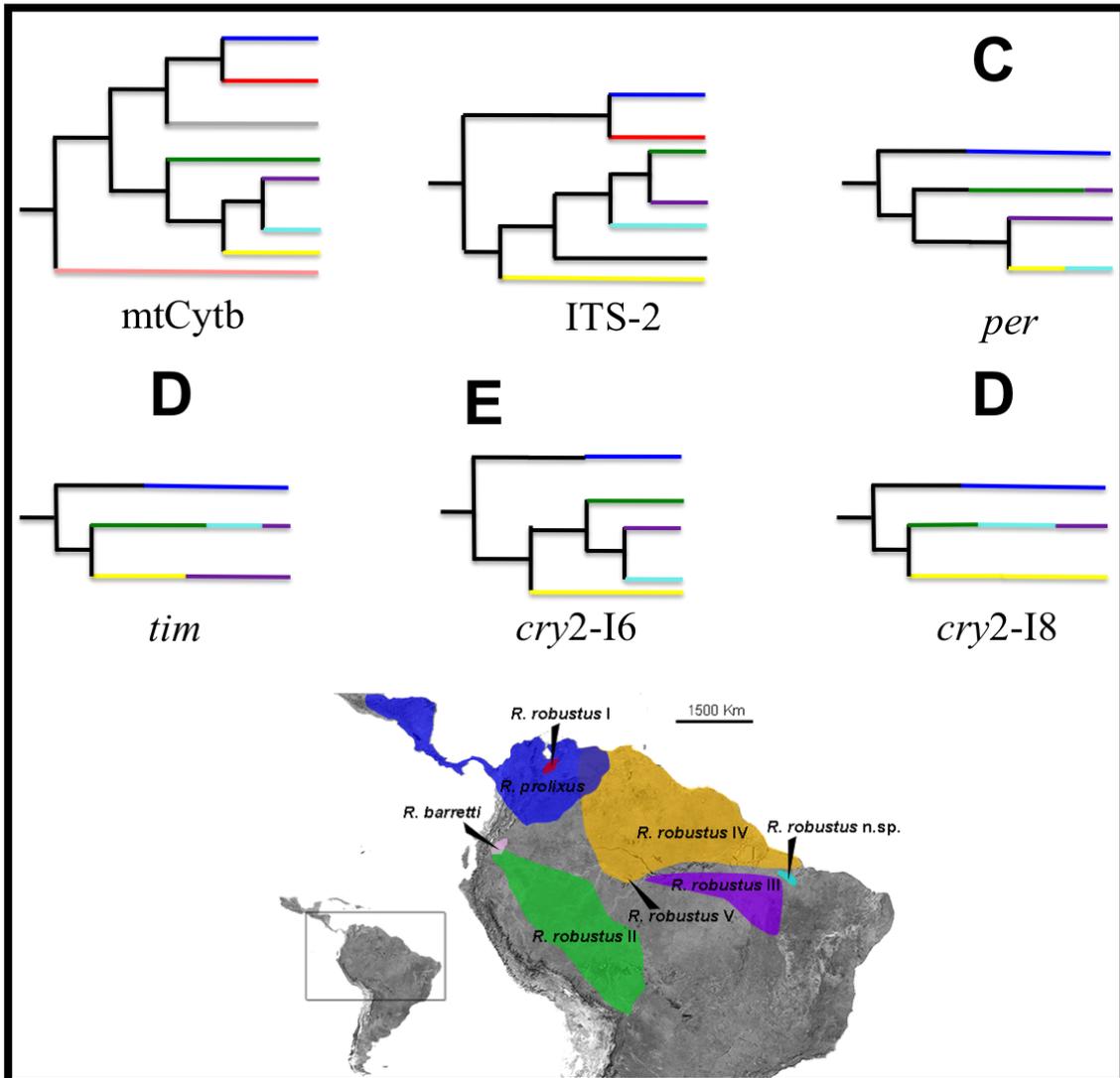
O isolamento de novos marcadores moleculares foi imprescindível para o desenvolvimento desta tese, uma vez que a determinação de possíveis zonas híbridas em regiões limítrofes das distribuições geográficas conhecidas das linhagens II, III e IV de *R. robustus* (sensu Monteiro et al., 2003; Pavan e Monteiro, 2007) dependia da

comparação entre as topologias das árvores resultantes do mtCytb e de genes nucleares. Além das análises filogenéticas com base na região ITS-2, que se mostrou pouco confiável pela falta de homogeneização das cópias múltiplas (e, portanto, não há garantia de ortologia das sequências analisadas), foram isolados do genoma de *R. prolixus* quatro íntrons completos de três genes relacionados ao comportamento de insetos – *per*, *tim* e *cry*. Estes genes estão associados com o comportamento sexual e isolamento reprodutivo e são, portanto, potencialmente úteis em estudos relacionados à especiação (e.g. Araki et al., 2009; Rona et al., 2010).

#### 6.1.1. Análise filogenética do mtCytb

Os resultados obtidos a partir da análise filogenética de 663-pb do mtCytb de *R. prolixus* e *R. robustus s.l.* corroboraram o que havia sido observado anteriormente por Monteiro e cols. (2003) – a distinção de *R. prolixus* e *R. robustus* como unidades taxonômicas distintas e ainda a identificação de um complexo parafilético de *R. robustus*. Dentro deste complexo parafilético foram descobertas durante esta tese três novas linhagens matrilineares, além das quatro já conhecidas (*R. robustus* I, II, III e IV). Estas linhagens foram denominadas aqui de *R. robustus* V, *R. robustus* n.sp. e *R. barretti* (Figura 5.1 e Tabela 5.1).

*R. robustus* V foi encontrada na ecorregião de florestas tropicais Japurá/Solimões-Negro da região biogeográfica do Imerí, próximo a Manaus (Amazonas). Esta linhagem, juntamente com aquela proveniente da região do Orinoco, *R. robustus* I, são filogeneticamente mais próximas a *R. prolixus* (porcentagens médias de  $K-2p$  entre 3,3 e 3,5%), do que às linhagens amazônicas II, III e IV de *R. robustus* (7,1-7,8%) (Figura 6.1A). O único espécime de *R. robustus* V coletado em uma região próxima de Manaus, Amazonas, não foi bem conservado e, conseqüentemente, não foi possível obter sequências de qualquer marcador molecular diferente do mtCytb. Como seu relacionamento parafilético em relação às demais linhagens de *R. robustus s.l.* não foi confirmado com base em análises adicionais de um marcador de origem biparental, não há como saber se este indivíduo é representante de uma nova linhagem, ou ainda se esta sequência representa uma introgressão de material genético interespecífico.



**Figura 6.1.** Esquema das topologias obtidas a partir do sequenciamento dos genes mtCytb (A), ITS-2 (B), *period* (C); *timeless* (D); *cryptochrome2-I6* (E), e *cryptochrome2-I8*. São mostrados apenas os agrupamentos “consenso” ( $PP > 0,8$ ). Os ramos com mais de uma cor representam os clados formados por mais de uma linhagem. O tamanho das cores nos ramos é proporcional ao número de indivíduos da linhagem que foram agrupadas.

Na porção oeste da distribuição geográfica de *R. robustus* II, localizada na região biogeográfica do Napo, foi encontrada uma linhagem que forma um grupo aparentemente basal em relação ao complexo *R. prolixus* e *R. robustus* s.l., que pretendemos descrever como *R. barretti*, em homenagem ao pesquisador Toby Vincent Barrett, apresentando porcentagens médias de  $K-2p$  entre 13,9 e 15,6%. *R. barretti* foi encontrada ainda em Putumayo, Colômbia, por Márquez e cols. (2011). Apesar de não ter sido possível sequenciar nenhum marcador nuclear desta linhagem em função da má preservação do material, análises morfométricas de cabeça e a razão

dos comprimentos da cabeça e do abdômen são diagnósticas com respeito à linhagem simpátrica *R. robustus* II (Anexo 2).

A partir da análise dos espécimes coletados nas duas regiões de possível encontro entre *R. robustus* II, III e IV, foram reveladas duas situações completamente distintas. Na região de Anamá e Careiro Castanho (Amazonas), região peripátrica dessas três linhagens, todos os espécimes coletados foram identificados com base no mtCytb como *R. robustus* II. Estas amostras foram analisadas para outros marcadores, que serão discutidos mais adiante. Contudo, em Ulianópolis (Pará), possível área de encontro entre *R. robustus* III e IV, surpreendentemente foi encontrada uma nova linhagem, que diverge de *R. robustus* III e IV em 2,6%, chamada aqui de *R. robustus* n.sp. Esta linhagem foi encontrada ocasionalmente em uma única palmeira, coabitada por indivíduos molecularmente identificados como *R. robustus* III (Figura 6.1A).

Posteriormente, recebemos outras seis amostras de *R. robustus* s.l. provenientes de São Roberto (Maranhão), relacionadas a um surto de infecção oral. Em uma localidade deste município foi realizada uma festa onde 13 pessoas consumiram caldo de cana e posteriormente apresentaram os sintomas característicos de infecção por *T. cruzi* transmitida via oral. A equipe da Secretaria Estadual de Vigilância do Pará foi até o local e capturou sete *R. robustus* s.l. em uma palmeira localizada ao lado do moedor de cana, confirmando a origem da contaminação. Destes sete triatomíneos coletados, seis puderam ser identificados por mtCytb como pertencentes à linhagem *R. robustus* n.sp.

As amostras de *R. robustus* n.sp. foram analisadas filogeneticamente para outros marcadores moleculares e sua validade taxonômica separada de *R. robustus* III foi avaliada a partir da análise de 10 *loci* de microssatélites. Estes resultados são comentados mais adiante.

Diferentemente do observado na topologia da árvore filogenética bayesiana do mtCytb, não foi possível observar a partir dos três primeiros eixos nas análises de PCA *R. robustus* I em um agrupamento distinto de *R. prolixus*. Esta incongruência provavelmente ocorreu em função da utilização de um modelo estatístico no PCA que não leva em consideração diferentes taxas de substituição entre transições e transversões, uma vez que os nucleotídeos são substituídos por códigos numéricos. A

outra possibilidade é a de que a divisão de *R. robustus* I de *R. prolixus* só pudesse ser observada em outros eixos cartesianos não analisados.

#### 6.1.2. Identificação taxonômica utilizando o mtCytb

De acordo com o ‘conceito filogenético de espécie’ (Mishler e Brandon, 1987), as espécies são definidas como “o menor agrupamento diagnosticável de organismos individuais, dentro do qual exista um padrão de ancestralidade e descendência”. A aplicação deste conceito de espécie aos resultados obtidos nesta tese a partir da análise do mtCytb, leva a conclusão inexorável que *R. robustus* é um complexo parafilético de seis espécies distintas. Particularmente em *Rhodnius barretti*, análises de morfologia e morfometria corroboraram os resultados moleculares (Anexo 2).

Não há consenso na literatura com respeito ao limite mínimo da porcentagem de divergência entre sequências que caracterizaria comparações entre espécies distintas. Hebert e colaboradores (2003), proponentes do *barcode of life*, propuseram para sequências de citocromo oxidase I (COI) que divergências entre 2 e 3% (2,7% originalmente) poderiam ser consideradas como interespecíficas. Contudo, valores mínimos de divergência interespecífica não podem ser aplicados “universalmente”, uma vez que diferentes grupos taxonômicos podem apresentar mecanismos de isolamento reprodutivo, tempos de geração e tamanhos efetivos de população distintos.

Monteiro e colaboradores (2004) sugeriram o valor de 8% como divergência interespecífica limítrofe entre sequências do mtCytb para espécies do gênero *Triatoma*. Contudo, vale ressaltar que existe uma lacuna entre as porcentagens de divergência intra e interespecíficas, o que garante a acurácia da separação de espécies. Esta lacuna é chamada de *barcoding gap* (Hebert et al. 2003). Em *Triatoma*, divergências intraespecíficas parecem não ultrapassar os 2%.

De maneira concordante, a maior porcentagem de divergência intraespecífica observada em *R. robustus s.l.*, considerando as linhagens como unidades taxonômicas distintas, foi de 1,5%. Em *Rhodnius*, as porcentagens de divergência genética observadas entre espécies próximas parecem ser menores do que em *Triatoma*. Por

exemplo, na ‘linhagem *pictipes*’, a divergência pode chegar a valores tão baixos quanto 4%, como visto entre *R. brethesi* e *R. stali*, que são incontestavelmente espécies boas.

### 6.1.3. Análises filogenéticas segundo o modelo coalescente de especiação

Um problema inerente à interpretação de topologias de árvores filogenéticas consiste em separar os nós que representam eventos macroevolutivos (especiação e extinção) daqueles que refletem processos microevolutivos (coalescência no nível populacional). Na tentativa de solucionar este problema, modelos baseados na distribuição de probabilidade discreta, como o Yule (Hardin, 1971), vêm sendo empregados nas análises estatísticas de topologias de árvores filogenéticas (Pons et al. 2006). Neste modelo é assumido que cada bifurcação dos ramos em uma topologia representa um evento de especiação.

Um outro modelo baseado em coalescência é o de ‘árvores de espécies’, no qual é possível estimar com precisão o tamanho efetivo populacional ancestral de cada espécie com base em sequências *multilocus* e comparar padrões heterogêneos em topologias de árvores de genes (Heled e Drummond, 2008). Contudo, existem dois grandes vieses metodológicos que comprometem a aplicabilidade desta análise em problemas relacionados a espécies filogeneticamente próximas. O primeiro é que no modelo de ‘árvore de espécies’ a premissa é de que qualquer incongruência entre árvores filogenéticas de genes é resultado de *ILS*. Portanto, nas árvores de espécies não existe a possibilidade de distinguir *ILS* de especiação incipiente (i.e. linhagens que são distintas, mas apresentam zonas híbridas em áreas de contato secundárias). O segundo viés se refere a informação *a priori* que deve ser preenchida em relação à taxonomia dos espécimes analisados nas análises bayesianas de geração da ‘árvore de espécies’. Todas as sequências obrigatoriamente devem ser assinaladas a uma determinada unidade taxonômica, mesmo quando na realidade o número de espécies é o questionamento principal do estudo.

De fato, processos como *ILS* e retenção de polimorfismo ancestral normalmente são a fonte principal de incongruência entre árvores gênicas contendo linhagens que divergiram recentemente ou cuja história evolutiva é caracterizada por intervalos curtos entre eventos de diversificação (Avice, 1989; Philipe et al. 2011). Na

ausência de um método capaz de diferenciar *ILS* de introgressão, a melhor maneira de interpretar os dados sob um ponto de vista histórico é o de analisar a maior quantidade de *loci* separadamente.

#### 6.1.4. Análise filogenética do ITS-2

Com base nas análises filogenéticas preliminares contendo todas as 110 sequências de ITS-2 de *R. prolixus* e *R. robustus* I, II, III, IV, 6.4 e n.sp., foi possível observar altas porcentagens de divergência genética intraindividual, inclusive maiores do que as porcentagens de divergência intra e interespecíficas (considerando as linhagens como unidades taxonômicas distintas). Foram observados pelo menos dois agrupamentos distintos dos clones de ITS-2 divergentes em até aproximadamente 2% quando foram analisadas sequências intraindividuais. A inclusão de todas estas sequências em uma inferência filogenética bayesiana resultou na formação de sete clados distintos estatisticamente bem apoiados ( $BS \geq 80$ ;  $PP > 0,9$ ). A parafilia de *R. robustus s.l.* foi confirmada com o uso deste marcador nuclear. *R. robustus* I forma um clado filogeneticamente mais próximo de *R. prolixus* do que das demais linhagens de *R. robustus s.l.* As sequências de *R. robustus* II foram separadas em dois clados distintos, o primeiro contendo apenas sequências de *R. robustus* II e o segundo contendo ainda sequências de *R. robustus* III, IV e n.sp. Parte das sequências de *R. robustus* IV foram ainda agrupadas separadamente em um outro clado. Finalmente, a sequência de *R. robustus* 6.4 formou um clado separado das demais linhagens de *R. robustus s.l.* (Figura 6.1B).

Este resultado é contraditório àquele obtido a partir do mtCytb e pode ser explicado de duas maneiras: (1) existem *loci* ortólogos de rDNA que evoluíram de maneira independente e ambos foram amostrados; ou (2) existem pelo menos dois *loci* de rDNA não ortólogos em cada linhagem de *R. robustus s.l.*

Em referência à primeira hipótese, isto pode acontecer por consequência de conversão gênica ineficiente e recombinação gênica (Li e Wilkerson, 2006). A segunda hipótese pode ser explicada por localizações múltiplas do gene de rDNA - já foram observadas em vertebrados e invertebrados, como por exemplo, cinco localizações distintas em humanos (Gonzalez e Sylvester 2001) e duas em fungos *Gibberella fujikuroi*, moscas-de-fruta *Drosophila hydei* (Hennig et al. 1975),

gafanhotos *Warramaba virgo* (White et al. 1982) e mosquitos *Anopheles albitarsis* (Li e Wilkerson, 2006).

Como não houve tempo suficiente durante esta tese para se fazer a predição das estruturas secundárias de ITS-2, as sequências foram separadas com base na proposta do *barcoding gap*. Ao contrário de *Triatoma*, em *Rhodnius* não existem dados anteriores referentes a lacuna entre divergência máxima observada entre sequências de ITS-2 intraespecíficas e de divergência mínima esperada em comparações interespecíficas.

Portanto, foi feita inicialmente uma comparação entre as sequências de mtCytb e ITS-2 das espécies crípticas de *T. dimidiata* disponíveis na literatura (Bargues et al. 2006; Monteiro et al. no prelo). A escolha de *T. dimidiata* foi baseada no fato de este táxon, assim como *R. robustus s.l.*, representar um complexo de linhagens filogeneticamente próximas. Os valores mínimos de divergência interespecífica calculados em mtCytb e ITS-2 foram utilizados para estimar nestas espécies em que proporção o primeiro marcador evolui mais rapidamente que o segundo. Assumindo esta proporção como verdadeira em *Rhodnius*, esta foi utilizada nas sequências de mtCytb de *R. robustus s.l.* para calcular o valor mínimo de divergência interespecífica esperado em comparações par-a-par entre sequências de ITS-2 das diferentes linhagens dessa espécie. As sequências de *R. prolixus* e *R. robustus s.l.* sequenciadas anteriormente foram utilizadas como referência. Como resultado, todos os clones que apresentaram porcentagens de divergência intraespecífica menores ou iguais a 0,6% em comparações com as sequências-base foram considerados como sendo de um mesmo *locus*.

Com base na nova reconstrução filogenética de ML, foi possível observar a clara separação entre as linhagens amazônicas de *R. robustus s.l.* Contudo, o relacionamento filogenético inferido para estas linhagens a partir das sequências de ITS-2 foi diferente daquele obtido com sequências de mtCytb. A topologia da árvore do mtCytb mostrou *R. robustus* III filogeneticamente mais próxima de *R. robustus* IV e n.sp. e *R. robustus* n.sp. mais próxima de *R. robustus* III, ao passo que a de ITS-2 apresentou *R. robustus* III mais relacionada a *R. robustus* II e *R. robustus* n.sp. a *R. robustus* II e III (Figura 6.1B).

Após a inclusão na topologia da árvore de ITS-2 dos clones de *R. robustus s.l.* que apresentaram porcentagens de divergência nucleotídica acima do esperado em comparações intraespecíficas, o resultado mais inusitado foi a formação de um novo clado bem apoiado estatisticamente contendo sequências de *R. robustus* IV provenientes de Abaetetuba (Pará). Esta sequência está filogeneticamente mais próxima à *R. robustus* n.sp. do que à *R. robustus* IV.

A presença desta sequência na análise levantou suspeitas em relação à identidade taxonômica de *R. robustus* 6.4. Esta suposta linhagem de *R. robustus s.l.* foi encontrada no Chaparé, na Bolívia, por Justi (2009), a partir de um único clone de um indivíduo identificado com base no mtCytb como *R. robustus* II. É possível que este clone, assim como aquele de *R. robustus* IV, se trate de uma cópia rara, proveniente de um *locus* que tenha evoluído de maneira independente dos demais, ou ainda de um *locus* não ortólogo em relação ao *locus* analisado. O sequenciamento de mais clones deste indivíduo, além de outros espécimes provenientes desta mesma localidade estão sendo analisados para a confirmação deste novo táxon.

Com o objetivo de testar se as regiões de microssatélites presentes em ITS-2 apresentam algum sinal filogenético relevante em *Rhodnius prolixus* e *R. robustus s.l.*, estas regiões foram analisadas separadamente, utilizando o modelo estocástico Dollo de multiestado (Alekseyenko et al., 2008). Nenhum resultado filogeneticamente relevante foi encontrado – a separação entre as linhagens da região do Orinoco e da Amazônia não foi observada. Estes resultados mostram que a melhor maneira de se analisar filogeneticamente as sequências de ITS-2 é excluindo todos os *indels*. Microssatélites altamente variáveis podem gerar ‘ruídos’ desnecessários nas análises filogenéticas, por evoluírem de maneira distinta das demais regiões do ITS-2. Consequentemente, sua utilização pode resultar na geração de topologias de árvores equivocadas (Li e Wilkerson, 2006).

Ao utilizar as informações genéticas contidas nas sequências de mtCytb e ITS-2 para a geração de uma ‘árvore filogenética de espécies’, a parafilia de *R. robustus s.l.* foi confirmada, o que já era esperado, uma vez que em ambas as árvores de mtCytb e ITS-2 os espécimes de *R. robustus* I estão agrupados filogeneticamente mais próximos de *R. prolixus* do que das demais linhagens de *R. robustus*. Contudo, o relacionamento entre as linhagens amazônicas de *R. robustus* não pôde ser resolvido.

Surpreendentemente, a topologia da ‘árvore de espécies’ diferiu daquelas observadas nas árvores de mtCytb e ITS-2. *R. robustus* IV e *R. robustus* 6.4 foram agrupadas mais proximamente a *R. robustus* II. Já a cladogênese de *R. robustus* III e n.sp. não foi apoiada estatisticamente.

A discordância do relacionamento filogenético das linhagens amazônicas de *R. robustus* s.l. nas três árvores é um indicativo de inexistência de um sinal filogenético comum nos dados analisados (Sota e Vogler, 2001). Este resultado abre precedentes para uma série de interpretações acerca das limitações dos marcadores e da metodologia utilizada para a solução deste problema filogenético, e ainda dos padrões e processos responsáveis pela geração da diversidade genética encontrada nestas linhagens. Nesse caso, a decisão mais parcimoniosa foi a de tentar responder primeiro as questões mais simples, que são as de origem metodológica.

No pacote \*BEAST (Heled e Drummond, 2008) a ‘árvore filogenética de espécies’ é estimada em conjunto com as árvores filogenéticas de cada marcador molecular utilizado, em uma única análise bayesiana de cadeias de Markov Monte-Carlo. Nesta análise, o tamanho populacional efetivo de cada linhagem para cada marcador utilizado deve ser calculado com acurácia para que a topologia da ‘árvore de espécies’ seja bem resolvida e, conseqüentemente, o relacionamento filogenético entre as linhagens seja o mais verossímil possível. Estes parâmetros são de suma importância para a separação dos eventos coalescentes resultantes de eventos de especiação daqueles que refletem apenas relacionamentos entre populações de uma mesma espécie ou uma separação incompleta da linhagem (*ILS*). Quanto maior o tamanho populacional, maior o tempo em número de gerações para atingir a coalescência.

A acurácia dos cálculos do tamanho populacional efetivo (ancestral e presente) de cada linhagem é dependente do número de *loci* filogeneticamente informativos e independentes que estão sendo analisados. A ‘árvore de espécies’ gerada a partir de apenas dois *loci* pode conter uma taxa de erro de até 30% (Heled e Drummond, 2008).

#### 6.1.5. Análise filogenética dos genes relacionados ao relógio circadiano

Para a obtenção das sequências dos quatro íntrons presentes nos genes ortólogos de *per*, *tim* e *cry* de *R. prolixus* e *R. robustus s.l.*, optou-se pelo sequenciamento direto sem a etapa prévia de clonagem por se tratarem de genes nucleares supostamente de cópia única. O sequenciamento direto de uma região nuclear não é indicado quando se analisa regiões que apresentam no genoma múltiplas cópias contendo alta variabilidade nucleotídica entre elas (e.g. ITS-2 de *Rhodnius*), quando a heterozigosidade das fitas de DNA é alta (evidenciada pela grande quantidade de picos duplos nos eletroferogramas em diferentes sítios da sequência), ou quando o objetivo de estudo é a análise populacional de uma espécie (para avaliação de homozigotos e heterozigotos). As sequências obtidas pelo sequenciamento direto representam, na realidade, a cópia do fragmento que ocorre em maior frequência, no caso de genes de cópia múltipla, ou uma sequência “consenso” do organismo estudado, sem informações relevantes acerca da heterozigosidade.

Com base na inspeção visual dos eletroferogramas, foi possível concluir que a heterozigosidade das fitas de DNA dos íntrons analisados é baixa. Ademais, como o intuito nesta tese era o de se avaliar o relacionamento filogenético (e não populacional) de *R. prolixus* e *R. robustus s.l.*, a opção pelo sequenciamento direto é válida.

A partir da análise filogenética das sequências do íntron de *per* de *R. prolixus* e *R. robustus s.l.*, foi possível observar uma topologia parecida com aquela obtida em mtCytb. *R. prolixus* é uma espécie distinta de *R. robustus s.l.* e grande parte das sequências de *R. robustus* III são filogeneticamente mais próximas de *R. robustus* IV e n.sp. do que de *R. robustus* II. Contudo, algumas discrepâncias foram observadas nesta comparação - quatro indivíduos de *R. robustus* III provenientes de Paraupébas (Pará) foram agrupados no mesmo clado composto por *R. robustus* II. Já *R. robustus* n.sp. foi agrupado com *R. robustus* IV em um mesmo clado (Figura 6.1C).

Estes resultados podem ser interpretados de três maneiras distintas: (1) existe fluxo gênico atual entre fêmeas de *R. robustus* III e machos de *R. robustus* II, e ainda entre fêmeas de *R. robustus* n.sp. e machos de *R. robustus* IV; (2) separação incompleta da linhagem; ou (3) *R. robustus* II, III, IV e n.sp. são populações de uma mesma espécie. Em relação a primeira hipótese, dificilmente existe fluxo gênico atual

entre espécimes de *R. robustus* II e fêmeas de *R. robustus* III provenientes de Paraupébas (Pará), em função da distância geográfica (distância mínima = 1192 km) e da baixa capacidade de dispersão dos triatomíneos (de dezenas de metros até 1 km; Barbu et al. 2010). Ademais, é pouco provável que a terceira hipótese seja verdadeira, uma vez que a partir de um marcador mitocondrial (mtCytb) e outro nuclear (ITS-2) foi possível separar as linhagens de *R. prolixus* e *R. robustus* s.l. Portanto, a melhor explicação seria a de que em *R. robustus* II, III, IV e n.sp. são observados sinais de *ILS* relacionados à retenção de polimorfismo ancestral.

Na filogenia reconstruída a partir de um íntron do gene ortógo de *tim* (Figura 6.1D) foi possível observar que *R. prolixus* é uma espécie distinta de *R. robustus* s.l. Apesar do número de sítios variáveis entre as sequências de *R. prolixus* e *R. robustus* s.l. de *tim* ser praticamente o mesmo que o observado em *per* (31 e 39, respectivamente), a proporção de autapomorfias de *R. prolixus* nas sequências de *tim* é muito maior do que em *per*, representando 80% da variação nucleotídica total. Em *per*, apenas 46% dos sítios polimórficos não resultantes de *indels* correspondem a autapomorfias de *R. prolixus*. Conseqüentemente, a informação genética contida em *tim* para a diferenciação das linhagens de *R. robustus* s.l. é menor do que em *per*.

Na topologia da árvore filogenética resultante as sequências de *R. robustus* II, III, IV e n.sp. não foram agrupadas em clados exclusivos. As sequências de *R. robustus* II estão contidas em um clado que é composto ainda por representantes de *R. robustus* n.sp. e ainda um espécime de *R. robustus* III proveniente de Marabá (Pará). As demais sequências de *R. robustus* III estão agrupadas em um outro clado, que contém ainda sequências de *R. robustus* IV (Figura 6.1D).

É possível que haja fluxo gênico presente ou recente entre *R. robustus* III e IV, por conta da proximidade geográfica destas linhagens. Porém o mesmo não pode ser dito em relação às sequências de *R. robustus* II, III e n.sp. incluídas em um único clado – estas duas últimas linhagens distam geograficamente de *R. robustus* II mais de 1500 km. Por conta da pouca quantidade de sítios polimórficos entre estas linhagens, os resultados sugerem a possibilidade de estar sendo observado *ILS* neste gene.

Ao observar a informação contida nos *indels* de *tim*, foi constatado um polimorfismo de tamanho que diferencia os íntrons de *R. prolixus* e de *R. robustus* s.l.

em 210-pb. Portanto, a exclusão dos *indels* de *tim* nas análises filogenéticas pode resultar uma perda importante de informação filogenética.

Os resultados mais interessantes sob o ponto de vista evolutivo das linhagens amazônicas de *R. robustus s.l.* foram obtidos a partir das análises filogenéticas dos dois íntrons do gene ortólogo de *cry2*. Assim como em todos os marcadores analisados nesta tese, as análises filogenéticas dos dois íntrons de *cry2* mostraram claramente a separação entre *R. prolixus* e *R. robustus s.l.* Contudo, estes íntrons de *cry2* analisados apresentam ‘assinaturas’ filogenéticas distintas nas linhagens amazônicas de *R. robustus*, apesar de estarem presentes no mesmo gene (Figura 6.1E e F). Em ambas as topologias, as sequências de *R. robustus* IV foram agrupadas em um clado filogeneticamente mais distante das demais linhagens de *R. robustus s.l.*, mas apenas no íntron 6 de *cry2* *R. robustus* III e n.sp. foram agrupadas separadamente de *R. robustus* II.

Uma das possíveis explicações para tal fato é que o íntron 8 contenha regiões mais conservadas do que o íntron 6, como sítios de *splicing* alternativo. A outra possibilidade é que tenham ocorrido processos de recombinação nesta região do gene que afetaram o sinal filogenético. Estas hipóteses ainda precisam ser testadas.

A grande quantidade de dados gerados a partir das análises filogenéticas de *R. prolixus* e *R. robustus s.l.* utilizando seis marcadores nucleares presentes em 5 *loci* independentes e um marcador mitocondrial foram muito úteis para o entendimento do fenômeno de especiação deste complexo. Nas topologias das árvores filogenéticas de mtCytb, ITS-2 e o íntron 6 de *cry2* (*cry2*-I6), todos os clados foram formados exclusivamente por uma única linhagem em cada, ao passo que as de *per*, *tim* e o íntron 8 de *cry2* (*cry2*-I8) mostraram as linhagens amazônicas em clados não exclusivos.

Nas topologias das árvores obtidas a partir das análises de *per* e mtCytb, a maioria das sequências de *per* (e todas as de mtCytb) de *R. robustus* II foram agrupadas em um clado mais distante de *R. robustus* III, IV e n.sp. Já nas árvores filogenéticas de ITS-2 e *cry2*, a maioria das sequências de *cry2*-I8 (senão todas de ITS-2 e *cry2*-I6) de *R. robustus* IV forma um clado exclusivo mais distante de *R. robustus* II e III. Portanto, a posição de *R. robustus* III na topologia da árvore parece ser a fonte de discrepância entre as topologias analisadas. O pior sinal filogenético foi

observado nas topologias da árvore do íntron de *tim* em *R. prolixus* e *R. robustus s.l.*, pois não foi possível separar com clareza nenhum clado que seja representado uma linhagem de *R. robustus* separado.

A ‘árvore de espécies’ calculada no \*BEAST a partir da análise dos seis marcadores confirmou a posição de *R. robustus* III filogeneticamente mais relacionada a *R. robustus* IV do que a *R. robustus* II. Portanto, a mistura de duas ou mais linhagens de *R. robustus s.l.* em um mesmo clado nas topologias de *per*, *tim* e do íntron 8 de *cry2*, e ainda a posição de *R. robustus* III filogeneticamente mais próxima à *R. robustus* II, são interpretadas como sendo decorrentes de *ILS*. Com a utilização de seis *loci* independentes na geração da ‘árvore de espécies’, as taxas de erro na determinação da topologia final são reduzidas para 10-20% (Heled e Drummond, 2008). Contudo, vale ressaltar que em todos os marcadores nucleares foram observados poucos sítios polimórficos nas comparações entre as linhagens de *R. robustus s.l.* e, portanto, o sinal filogenético do mtCytb prevaleceu.

Como dito anteriormente, o pacote \*BEAST utiliza como premissa a ausência de fluxo gênico primário ou secundário entre as unidades taxonômicas. Árvores filogenéticas de linhagens próximas comumente apresentam discordâncias na topologia quando diferentes genes são analisados (Bessu e Near, 2009; Bottecchia et al. 2004; Mallet et al. 2007; Near e Keck 2013; Karsten e Knowles, 2007; Philippe et al. 2013; Mullen et al. 2008). Este fato pode ser observado em função de uma ou mais árvores gênicas conter paralogias, *ILS*, ou introgressão (Wiens et al., 2010).

A topologia resultante da ‘árvore de espécies’ feita a partir da análise de todos os marcadores nucleares em conjunto com o marcador mitocondrial foi exatamente igual àquela obtida a partir da análise do mtCytb. Assim como nos resultados obtidos nesta tese, Philippe e cols. (2011) observaram que a inclusão de marcadores mitocondriais em topologias de ‘árvores filogenéticas de espécies’ melhora substancialmente a resolução dos clados que aparentemente estão sujeitos a *ILS* em alguns *loci* nucleares.

O grande desafio encontrado no entendimento das topologias das árvores gênicas foi o evento de homoplasia observado em *R. robustus* II, III e IV. Desta forma, fica difícil compreender se *R. robustus* III é filogeneticamente mais próxima de *R. robustus* II ou de *R. robustus* IV. Possíveis explicações para este tipo de

incongruência envolvem a idealização de cenários demográficos mais complexos que levem em consideração migração entre populações ancestrais e especiação com fluxo gênico histórico ou recente. Modelos genealógicos coalescentes podem fornecer uma base estatística sólida sem a necessidade de estabelecer premissas como *ILS* e ausência de migração (Knowles e Carstens, 2007).

Marcadores mitocondriais apresentam maior probabilidade de resolverem os relacionamentos entre complexos de espécies próximas utilizando apenas um único *locus* ou mesmo quando agregado a quantidades substanciais de *loci* nucleares (Bossu e Near, 2009; Hudson e Coyne, 2002). Compreender os processos biogeográficos e populacionais históricos envolvidos na origem e dispersão do complexo de espécies utilizando estes marcadores é fundamental para localizar as possíveis áreas de encontro (e possível fluxo gênico) primárias e secundárias entre linhagens filogeneticamente próximas. Desta forma, é possível estimar a possibilidade de introgressão do mtDNA histórica e recente e, conseqüentemente, resolver as discrepâncias observadas nas topologias de árvores gênicas (Bossu e Near, 2009).

Uma outra incongruência observada entre as árvores dos marcadores utilizados e a ‘árvore de espécies’ foi em relação ao status taxonômico de uma das linhagens descobertas nesta tese. *R. robustus* n.sp. aparece em quase todas as topologias das árvores dos marcadores analisados (com a exceção de *per*) proximamente relacionada a *R. robustus* III. Contudo, a identidade taxonômica desta linhagem como sendo diferente de *R. robustus* III só foi confirmada (i.e. bem apoiada estatisticamente) nas topologias das árvores de ITS-2 e mtCytb. Na ‘árvore de espécies’ estimada em \*BEAST, a separação entre *R. robustus* n.sp. e *R. robustus* III não foi apoiada estatisticamente ( $PP = 0,58$ ). Ramos curtos internos na árvore podem fornecer uma baixa resolução filogenética em função da estocasticidade de poucas substituições nucleotídicas ao longo dos ramos e sinais conflitantes atribuídos a *ILS* entre os genes (Philippe et al. 2013). Neste caso, a melhor maneira encontrada para definir a existência de fluxo gênico entre estas linhagens foi a de analisar marcadores moleculares que evoluem mais rapidamente que os íntrons nucleares e o gene mitocondrial mtCytb – os microssatélites.

## 6.2. Análise de microssatélites

As causas e conseqüências da hibridação de linhagens filogeneticamente próximas vem sendo extensivamente estudadas no contexto de zonas híbridas, áreas de interação entre grupos de indivíduos geneticamente diferentes que resulta na produção de pelo menos uma geração de ancestralidade misturada (Harrison 1990). As zonas híbridas estão frequentemente localizadas em áreas geográficas tangenciais em relação às distribuições das populações parentais e tipicamente associadas a variações fenotípicas e genotípicas clinais (Barton e Hewitt 1989). Estas zonas podem surgir tanto em uma situação de intergradação primária, como em um contato secundário.

As zonas de hibridação primárias são resultado da diferenciação de populações interconectadas em variações clinais de múltiplos caracteres fenotípicos e genotípicos. Em contraste, as zonas híbridas de origem secundária são formadas quando populações que divergiram em alopatria entram ocasionalmente em contato, após eventos de expansão demográfica e/ou geográfica.

Sinais de ‘reforço’ (do inglês *reinforcement*), homogeneização de caracteres morfológicos e/ou genéticos e divergência adaptativa são alguns dos eventos evolutivos que podem ocorrer em áreas de encontro de linhagens distintas (e.g. Araki et al., 2009; Johnson e Lachance, 2012; Ortiz-Barrientos et al., 2004). A possibilidade de ocorrência destes eventos faz com que o estudo de zonas híbridas seja rico em informações sobre o processo de especiação de diferentes organismos. A hibridação das linhagens, em última análise, depende do balanço entre seleção natural e dispersão.

Ao contrário do postulado no conceito biológico de espécies, no qual a espécie é definida como “agrupamento de populações naturalmente intercruzantes e reprodutivamente isolado de outros grupos com as mesmas características” (Mayr, 1963), a hibridação entre espécies boas parece ser um evento regular. Pelo menos 10% das espécies de animais são conhecidas por hibridarem na natureza (Mallet, 2005). Esta porcentagem pode ser ainda maior em grupos que radiaram subitamente ou recentemente – de 29 a 42% das espécies de dois gêneros de borboleta, *Heliconius* e *Eueides* (Mallet, 2007).

Treze híbridos naturais de *R. prolixus* e *R. robustus* IV foram encontrados recentemente no estado de Portuguesa, Venezuela (Fitzpatrick et al. 2008). Todas estas ninfas foram identificadas com o marcador 28S D2 nrDNA como sendo *R. prolixus* e com o mtCytb como *R. robustus* IV. Neste caso, o mais provável é que tenha ocorrido introgressão em *R. prolixus* do material genético mitocondrial de *R. robustus* IV.

A amostragem de *R. robustus* II, III e IV foi feita durante o período da tese a partir da seleção criteriosa de regiões que representam duas áreas contíguas das linhagens, ao longo das quais foram realizados transectos de amostragem. Conforme comentado anteriormente, nos pontos de coleta compreendidos na possível zona de encontro entre *R. robustus* II, III e IV no estado do Amazonas, Brasil, não foram observados indícios de qualquer evento de introgressão. Contudo, em Ulianópolis (Pará), região de possível encontro entre *R. robustus* III e IV, surpreendentemente foram encontrados em uma palmeira seis espécimes identificados como *R. robustus* n.sp. co-ocorrendo com seis indivíduos identificados como *R. robustus* III. Esta descoberta inesperada veio a requerer a utilização de marcadores moleculares que evoluem mais rapidamente que os genes mitocondriais e nucleares, os microssatélites, para que fosse testado o isolamento reprodutivo entre os dois táxons. Por serem marcadores que fornecem informações de nível populacional, microssatélites são particularmente úteis em estudos que investigam a existência de espécies incipientes que ocorram em simpatria.

Doze espécimes simpátricos de *R. robustus* III e n.sp. foram genotipados para 10 *loci* de microssatélites. Aparentemente, com base nas frequências gênicas obtidas, não há barreiras ao fluxo gênico entre estes dois táxons nesta área. Para fins de comparação, foram adicionadas à análise cinco amostras de *R. robustus* III provenientes de Rondon do Pará (Pará), uma área adjacente à região de simpatria, e cinco amostras de *R. robustus* n.sp. provenientes de São Roberto (Maranhão), a outra localidade onde esta linhagem foi encontrada.

É importante ressaltar que o fator limitante nesta análise foi o número de amostras em cada população. Os alelos nulos encontrados e os desvios significativos do equilíbrio segundo o modelo de H-W podem ser resultado do baixo tamanho amostral das quatro populações.

Dois *loci* das amostras de *R. robustus* n.sp. de São Roberto apresentaram indícios de alelos nulos. Apesar da frequência de heterozigotos ter sido numericamente desproporcional em quase todas as populações para a maioria dos *loci*, apenas dois destes *loci* da população de Ulianópolis de *R. robustus* n.sp. apresentaram após a correção do nível de significância para múltiplas comparações desvios significativos do equilíbrio segundo o modelo de H-W. Ainda possivelmente em função da baixa amostragem, nenhum valor calculado de  $F_{IS}$  foi significativo.

Em relação ao número de alelos observados nas quatro populações analisadas, oito dos 10 *loci* são polimórficos. Nenhum *locus* foi diagnóstico para as populações simpátricas de *R. robustus* III e n.sp. Contudo, as frequências dos alelos observados em *R31* foram significativamente diferentes para *R. robustus* III e n.sp. na localidade de Ulianópolis ( $G = 6,35$ ; g.l. = 1;  $P = 0,01$ ).

A partir da análise dos componentes principais (PCA) e dos índices de estruturação  $F_{ST}$  e  $G_{ST}$ , não foi possível separar as (supostas) populações simpátricas de *R. robustus* III e n.sp. Em contraste com o algoritmo bayesiano implementado no programa STRUCTURE, na análise de PCA não é feita nenhuma premissa em relação ao equilíbrio de H-W ou LD e, portanto, é útil apenas para uma avaliação visual preliminar do grau de diferenciação das populações amostradas. Quanto aos índices de estruturação  $F_{ST}$  e  $G_{ST}$ , a premissa básica é que as populações são panmíticas e que estão em equilíbrio de H-W. O maior problema em relação a estas premissas é que dificilmente populações naturais as satisfazem e muito menos populações pertencentes a complexos de espécies incipientes, onde normalmente a heterozigosidade observada é baixa e a zona de possível fluxo gênico é restrita (Leng e Zhang 2011).

O índice mais utilizado em análises de microssatélites, o  $G_{ST}$ , não quantifica satisfatoriamente a diferenciação populacional em condições de desequilíbrio (Leng e Zhang, 2011). Quando a heterozigosidade de uma população analisada é baixa, os valores de  $G_{ST}$  se dilatam de maneira artificial e o oposto é observado quando a heterozigosidade é alta. Esta última condição (alta heterozigosidade) pode ser observada em populações naturais quando a taxa de mutação do *locus* é alta ou quando a população inicial ancestral foi formada por indivíduos majoritariamente heterozigotos.

Ao contrário das estimativas obtidas pelos índices de  $F_{ST}$  e  $G_{ST}$ , na inferência bayesiana implementada no programa STRUCTURE não é assumido o equilíbrio de ligação ou equilíbrio de H-W. Com base na heterozigosidade observada, o programa agrupa as populações estabelecidas *a priori* de forma que o equilíbrio de H-W seja atingido em cada um dos K agrupamentos testados (Pritchard et al. 2000). Portanto, a vantagem em se utilizar esta metodologia reside no fato de não haver necessidade de se admitir determinados pressupostos nem de se estimar o número de populações naturais estudadas (chamadas aqui de agrupamentos).

A partir da análise populacional feita no STRUCTURE, as (supostas) populações simpátricas de *R. robustus* III e n.sp. foram agrupadas separadamente. Contudo, um indivíduo de Rondon do Pará identificado como *R. robustus* III apresentou 93,4% de chance de pertencer ao agrupamento de *R. robustus* n.sp. de Ulianópolis. Este resultado pode ser interpretado de três maneiras distintas: (1) as populações simpátricas de *R. robustus* III e n.sp. não hibridizam e o que se observa provavelmente é uma retenção de polimorfismo ancestral; (2) as populações simpátricas não hibridizam e a semelhança entre os alelos observados em *R. robustus* III e n.sp. é consequente de uma alta variabilidade dos *loci* de microssatélites; e (3) as populações simpátricas podem eventualmente hibridizar.

Em referência à primeira hipótese, pode-se dizer que a diversidade genética observada nas populações atuais reflete o tempo de coalescência dos *loci* de microssatélites que, por sua vez, depende da história demográfica e populacional da linhagem. Segundo a teoria da coalescência, populações maiores acumulam mais mutações e perdem menos alelos por ‘extinção’ em comparação com populações menores. Em vertebrados, a variação genética dos microssatélites diminui significativamente (aproximadamente 20% da heterozigosidade observada) em populações que sofreram diminuições drásticas no tamanho total (Chapuis et al., 2012). Portanto, é de se esperar no caso de as populações de *R. robustus* III ou *R. robustus* n.sp. terem sofrido eventos históricos de *bottleneck* (ou ‘efeito gargalo de garrafa’) severos, que a variabilidade dos microssatélites também tenha diminuído. Além disso, em função do longo tempo de geração, é de se esperar que *R. robustus* s.l. acumule menos mutações do que um inseto com o tempo de geração curto (como mosquitos) e, portanto, que apresente *loci* de microssatélites menos variáveis do que é comumente observado em mosquitos.

Ao considerar a segunda hipótese como verdadeira, é de se esperar que os microssatélites de *R. robustus* III e n.sp. tenham uma alta taxa de mutação, assim como observado em outros insetos, como em mosquitos (Lovin et al. 2009). Além de os microssatélites estarem sujeitos à recombinação, novos alelos são gerados por estocasticidade e oscilam por deriva. Isso faz com que ‘assinaturas’ evolutivas mais antigas sejam apagadas do sinal filogenético, de modo que o que é observado atualmente está mais relacionado à processos populacionais contemporâneos. A igualdade dos alelos em cada *locus*, portanto, é consequente de convergência e paralelismo de eventos independentes. Para testar essa hipótese, seria necessário fazer um estudo genealógico para medir a taxa de mutação destes *loci* a cada geração e, com base no cálculo a partir do relógio molecular relaxado (seção a seguir), estimar o acúmulo de mutações retrospectivamente até o tempo geológico de vicariância estimado. Para isso, uma simulação poderia ser realizada, utilizando o modelo estocástico de Dollo multiestado (Alekseyenko et al. 2008).

Por fim, se a terceira hipótese for a mais correta, o programa STRUCTURE estaria superestimando o número de agrupamentos em quatro. Na realidade, casos como este são comuns quando uma das premissas da análise não é respeitada: a de que as amostras analisadas não sejam pertencentes a uma mesma família (i.e. pai, mãe, filhos, etc.) (Pritchard et al. 2000). O problema é que triatomíneos silvestres do gênero *Rhodnius* são filopátricos, i.e. passam normalmente toda sua vida intimamente associados à ninhos de aves, mamíferos e répteis nas copas de palmeiras e só saem em busca de um novo abrigo e fonte alimentar no caso do vertebrado abandonar seu ninho ou esconderijo. Desta forma, possivelmente em Ulianópolis cada uma das populações de *R. robustus* III e n.sp. deve conter representantes de uma única família (ou de poucas famílias) de cada uma das linhagens de *R. robustus*. Como nas análises de PCA,  $F_{ST}$  e  $G_{ST}$  somente as (supostas) populações simpátricas de *R. robustus* III e n.sp. não estão estruturadas, estas amostras podem estar representando um encontro secundário das linhagens. Os resultados referentes à origem e dispersão das linhagens de *R. robustus* III e n.sp. na região amazônica corroboram este encontro secundário (seção a seguir).

### 6.3. Filogeografia de *R. prolixus* e *R. robustus* s.l. na região amazônica

As análises biogeográficas do complexo de espécies crípticas *R. prolixus* e *R. robustus* s.l. foram feitas com base nas sequências do mtCytb. O motivo principal desta escolha foi que a topologia da árvore obtida para este marcador foi idêntica àquela observada na ‘árvore de espécies’, quando todos os marcadores foram analisados ao mesmo tempo em diferentes partições, o que demonstra a falta de informação genética nos marcadores nucleares para a diferenciação de *R. robustus* s.l. Para qualquer *locus*, a informação que pode ser extraída sobre o tamanho populacional (parâmetro de extrema importância em análises coalescentes, conforme dito anteriormente) depende da acurácia com a qual as árvores filogenéticas e genealógicas são resolvidas (Atkinson et al. 2008). Desta forma, os outros marcadores moleculares não foram incluídos nestas análises, pois algumas discrepâncias em relação às topologias das árvores são observadas nas comparações com a topologia resultante do mtCytb e da ‘árvore de espécies’, o que poderia comprometer a acurácia da datação dos eventos cladogenéticos. Adicionalmente, a não inclusão destes dados foi motivada ainda pela falta de informação prévia em triatomíneos ou organismos filogeneticamente próximos à respeito das taxas de mutação médias para cada marcador.

Os marcadores mitocondriais, apesar de serem de origem estritamente materna, são considerados como bastante apropriados para estudos filogeográficos (Avice, 2000, 2009; Lemey et al., 2009; Ohshima e Yoshizawa, 2011). A presença no organismo de um número grande de mitocôndrias, sua rápida taxa de substituição (Howell et al., 1996) e o fato de o ancestral comum mais recente ser estimado em um número de gerações quatro vezes menor do que o estimado para marcadores nucleares (Avice, 2000), faz com que as informações contidas nos marcadores mitocondriais resolvam com acurácia árvores filogenéticas e os processos de demografia histórica segundo um modelo coalescente. Para suprir a limitação da utilização de apenas um *locus*, foram acrescentados nas análises o maior número possível de indivíduos de cada linhagem. Esta estratégia parece contribuir para o aumento da precisão dos resultados obtidos (Heled e Drummond 2008).

### 6.3.1. Relógio molecular relaxado, calibração dos nós da árvore filogenética e datação dos eventos cladogenéticos

Com base nas inferências bayesianas do relógio molecular aleatório (*RLC*), a hipótese do relógio molecular ‘fixo’ foi rejeitada. Ao todo, foram encontradas de duas a três taxas evolutivas distintas na topologia analisada. Dentro do complexo de espécies crípticas *R. prolixus* e *R. robustus s.l.*, as sequências provenientes de Roraima de *R. robustus* IV, além de *R. robustus* V e n.sp. evoluem a uma taxa distinta das demais (dados não mostrados). Portanto, o modelo de relógio molecular ‘relaxado’ implementado no programa BEAST (Drummond et al. 2012) foi empregado nas análises. Este modelo não-uniforme descreve a diferença de taxas evolutivas entre as linhagens e modela a correlação entre estas taxas. Como este é um método ‘linhagem-específico’ (i.e. cada sequência é analisada de maneira independente), uma média de mudança das taxas é calculada para cada ramo da árvore. Os tempos estimados *a posteriori* para os eventos vicariantes dependem de uma série de variáveis iniciais (*prioris*), como a taxa de substituição nucleotídica estimada e sua variância, e as calibrações dos nós da árvore utilizadas e suas variâncias (Battistuzzi et al., 2010).

A variação da taxa mutacional foi determinada *a priori* de acordo com as informações prévias publicadas na literatura para espécies do gênero *Triatoma* e para a Classe Insecta (Brower, 1994; Pfeiler et al., 2006), o que resultou em um intervalo de confiança compreendido entre  $2,1 \times 10^{-3}$  e  $7,16 \times 10^{-2}$  substituições por milhões de anos (com média de  $1,2 \times 10^{-2}$ ).

A aplicação de parâmetros contendo certo grau de incerteza na topologia da árvore tende a aumentar a variância nas estimativas das datações *a posteriori*, podendo gerar densidades posteriores bimodais ou multimodais (Ho et al. 2009; Lee et al. 2012). Apesar deste possível revés, a opção por um intervalo de confiança dilatado nesta estimativa foi uma maneira de evitar vieses potenciais na análise, mesmo ao custo de reduzir a precisão de inferir definitivamente as proposições (Ho et al. 2005).

Foram testadas cinco abordagens distintas para a datação dos nós da árvore com base em quatro tipos de calibrações, sendo um destes de base empírica e três estimados a partir de evidências geológicas. O valor de verossimilhança marginal foi

calculado para cada abordagem e a escolha do melhor filograma foi feita a partir da análise do Fator Logarítmico de Bayes (*LBF*). Esta estratégia teve como objetivo identificar quais calibrações eram as mais indicadas para os dados analisados.

Utilizando a mesma motivação para a escolha dos valores *a priori* das taxas de substituição nucleotídica (i.e. o de garantir a acurácia dos dados, mesmo ao custo de uma menor precisão), nas datações das calibrações dos nós da árvore foram assumidos intervalos de confiança dilatados que pudessem abrigar as estimativas de tempo geológico da orogênese dos Andes em sua porção central e norte (cf. Abad-Franch et al. 2009; Hoorn et al. 2010).

Em todas as calibrações foram assumidas distribuições paramétricas normais. A simetria da distribuição normal em torno da média é muito útil para calibrações biogeográficas, pois neste caso é praticamente impossível especificar um limite severo ao tempo da cladogênese (Ho et al. 2009). Este tipo de distribuição é igualmente válido no caso de se importar datações de eventos cladogenéticos estimadas em outros estudos, pois permite a inferência de um erro associado ao valor outrora calculado (Ho et al., 2008).

Com base nos resultados de *LBF*, foi possível concluir que a datação utilizada na calibração do nó referente à separação entre *R. pictipes* e *R. prolixus/R. robustus s.l.* (*sensu* Gaunt e Miles 2002) está superestimada. A adição desta calibração na árvore filogenética modificou substancialmente a posição do evento de cladogênese de *R. barretti* no tempo geológico. A separação entre *R. barretti* e *R. prolixus/R. robustus s.l.* foi situada em um período mais antigo do que a separação de *R. pictipes* e *R. pallescens s.l.*

Este resultado parece pouco provável, uma vez que as sequências de mtCytb de *R. pictipes* e *R. pallescens* divergem em aproximadamente 19% e as sequências de *R. barretti* e *R. prolixus/R. robustus s.l.* em 10%. Se as datações dos nós das árvores que contêm esta calibração proposta por Gaunt e Miles (2002) para a separação de *R. prolixus* e *R. pictipes* estivessem corretas, a diferença entre as taxas evolutivas da ‘linhagem *pictipes*’ e ‘linhagem *robustus*’ seria muito grande – 1,9% de divergência, em média, entre as sequências de *R. pictipes* e *R. pallescens s.l.* a cada milhão de anos (Ma) e 0,72% de substituições por Ma entre *R. barretti* e *R. prolixus/R. robustus s.l.*

Ao considerar os filogramas contendo apenas as calibrações biogeográficas, as estimativas das taxas evolutivas foram muito mais próximas entre as duas linhagens e aparentemente mais realistas – 1,9% de divergência por Ma, em média, entre *R. pictipes* e *R. pallescens*, e 1,6% por Ma entre *R. barretti* e *R. prolixus/R. robustus s.l.* De fato, a topologia da árvore que contém as duas calibrações biogeográficas ‘estritas’ para a separação entre *R. prolixus/R. robustus s.l.* e *R. pictipes/R. pallescens s.l.* e ainda entre *R. pallescens* I e II apresentou fortes evidências de ser o melhor filograma ( $3,8 < LBF < 4,2$ ).

A utilização dos valores de verossimilhança marginais calculados em *LBF* para determinação do melhor filograma é válida, pois tanto as calibrações dos nós quanto as taxas evolutivas calculadas para cada ramo com base nas estimativas de tempo estabelecidas nas calibrações influenciam na topologia final da árvore (Drummond, comunicação pessoal).

As datações obtidas nos nós mais profundos (i.e. mais próximos da raiz) calculadas a partir de calibrações dos nós mais superficiais (mais distantes da raiz) costumam apresentar desvios significativos dos tempos reais e o oposto também é verdadeiro (Battistuzzi et al. 2010; Linder et al. 2005). Portanto, é de se esperar que os nós que contém as melhores datações sejam aqueles localizados em tempos geológicos próximos aos dos nós calibrados. No caso de se utilizar apenas um único nó calibrado, é provável que todos os nós distantes dele apresentem datações equivocadas. Provavelmente esta é a explicação para a topologia com duas calibrações biogeográficas ter sido selecionada ao invés das topologias contendo apenas uma calibração. Apesar da média obtida para os eventos cladogenéticos serem semelhantes nas árvores que tiveram um ou dois nós calibrados, os intervalos de confiança para os eventos cladogenéticos não o são (mais dilatados nas inferências de tempo a partir de um nó calibrado).

A influência nos resultados da utilização de uma calibração inadequada em um nó da árvore foi tão grande para a estimativa dos demais eventos cladogenéticos que mesmo adicionando na topologia outras duas calibrações biogeográficas consideradas como ‘boas’, o valor de verossimilhança marginal resultante foi menor do que aquele obtido nas árvores contendo apenas um ponto de calibração (orogênese da porção Central dos Andes). Este resultado está de acordo com as simulações realizadas por

Battistuzzi e cols. (2010), nas quais foi observado que a adição de novas calibrações são inúteis para a adequação das estimativas temporais das cladogêneses quando existe uma calibração que não se encaixa com a taxa evolutiva das demais.

As estimativas *a posteriori* dos tempos geológicos dos nós calibrados foram fortemente influenciados pelos valores indicados *a priori*, assim como observado anteriormente por Ho e cols. (2005, 2009). Esta limitação metodológica pode ser superada com a utilização marcadores moleculares contendo sequências nucleotídicas mais extensas ou ainda com a análise de mais *loci* polimórficos.

A partir da árvore filogenética contendo duas calibrações biogeográficas, foi possível situar os eventos cladogenéticos no tempo geológico. A cladogênese de *R. barretti* teria ocorrido durante o Mioceno Médio/Superior, há aproximadamente 7,5 Ma (12 – 6 Maa). Durante o Plioceno Inferior/Médio (entre 5 e 2,7 Maa), a população ancestral de *R. prolixus*, *R. robustus* I e *R. robustus* V teria se separado da população ancestral de *R. robustus* II, III, IV e n.sp. Por fim, as cladogêneses das linhagens inseridas nestes clados teriam ocorrido durante o Plioceno Médio/Superior e Pleistoceno Inferior/Médio, entre 3,5 e 0,5 Maa. Portanto, os períodos cladogenéticos calculados coincidem com o tempo em que as ‘incurções marinhas’ e os ‘refúgios pleistocênicos’ teriam supostamente promovido modificações no ambiente pan-amazônico. Para determinar se estas hipóteses biogeográficas explicariam o padrão de diversificação revelado em *R. prolixus* e *R. robustus s.l.*, foram conduzidas análises genealógicas adicionais e testes de hipóteses referentes à origem e dispersão das linhagens na região pan-amazônica.

### 6.3.2. Padrões populacionais e genealogia molecular

Com base no estudo genealógico e populacional de *R. prolixus* e de cada uma das linhagens de *R. robustus s.l.*, foi possível determinar se as modificações no ambiente amazônico durante o Mioceno, Plioceno e Pleistoceno deixaram ‘assinaturas’ referentes aos processos de diversificação genética nas populações atuais. Apesar de *R. robustus* V ter sido coletado na ecorregião de floresta tropical Japurá/Solimões-Negro (região biogeográfica do Imerí), sua sequência foi arbitrariamente considerada como pertencente à região biogeográfica do Orinoco, com base nos resultados filogenéticos de mtCytb. Esta decisão conservadora foi

tomada, pois *R. robustus* V é representada por um único espécime genotipado apenas com o mtCytb. Como seu status taxonômico não foi confirmado utilizando um marcador de origem biparental, não há como saber se este indivíduo é de fato representante de uma nova linhagem, ou ainda se esta sequência representa uma retenção de polimorfismo ancestral.

Assim como observado em outros triatomíneos (cf. Monteiro et al. 2004; Pavan 2009), apesar de *R. prolixus* e as linhagens de *R. robustus* apresentarem uma grande diversidade haplotípica ( $Hd > 0,5$ ), as diversidades nucleotídicas em comparações par-a-par são baixas ( $\pi = 1,60E-04 - 8,10E-03$ ). As maiores diversidades haplotípica e nucleotídica foram encontradas em *R. robustus* IV, provavelmente a linhagem de *R. robustus s.l.* com maior distribuição geográfica. A menor diversidade haplotípica foi encontrada em *R. robustus* III e a menor diversidade nucleotídica em *R. prolixus*. O único teste de neutralidade significativamente negativo foi o D de Tajima em *R. robustus* II, o que indicaria sinais de expansão populacional súbita ou seleção purificadora. Apesar de o teste de  $F_s$  de Fu não ter sido significativo, a análise de distribuição ‘desencontrada’ (do inglês *mismatch distribution*) apresentou um padrão unimodal ( $r = 0,27$ ) característico de populações que sofreram súbita expansão populacional e/ou demográfica.

A partir da construção de redes genealógicas para os haplótipos encontrados em *R. prolixus* e *R. robustus s.l.*, foi possível definir (empiricamente) as possíveis regiões de origem e dispersão das linhagens. Esta abordagem foi bastante útil para esclarecer se as regiões de encontro entre as linhagens II, III e IV são consequentes de ocupações primárias ou secundárias.

Apesar de as sequências de *R. prolixus* não terem apresentado valores significativos nos testes de neutralidade, a região do Guárico (nordeste da região biogeográfica do Orinoco) parece ter sido o local de origem e expansão súbita para as regiões adjacentes do Orinoco e ainda para a região biogeográfica do Imerí, se considerarmos *R. robustus* V uma linhagem válida e distinta das demais. Contudo, a grande distância geográfica entre o Orinoco e o Imerí sugere que esta rede contém ainda muitas peças faltantes. De maneira alternativa, esta sequência mitocondrial de *R. robustus* V pode representar apenas uma retenção de polimorfismo ancestral. Neste caso, teríamos fortes indicativos para uma possível origem de *R. prolixus* e *R.*

*robustus s.l.* na região do Orinoco, conforme postulado empiricamente por Abad-Franch e cols. (2009).

As informações interessantes extraídas da rede genealógica de *R. robustus* II foram que a origem da linhagem deve ter sido na região biogeográfica de Rondônia, e que o Inambari foi ocupado secundariamente. Portanto, a região de possível encontro entre as linhagens II, III e IV (Careiro Castanho e Anamã, Amazonas), onde não foi encontrado nenhum indício de fluxo gênico interespecífico, foi ocupada de maneira secundária. Um haplótipo compartilhado entre sequências provenientes de ambas as regiões biogeográficas (haplótipo b da Figura 5.26) parece ter sido o centro de uma expansão geográfica repentina.

A expansão populacional súbita de *R. robustus* II foi confirmada após as análises de *Bayesian Skyride*. Sua população teria aumentado de tamanho 3,25 vezes nos últimos 125 mil anos, atingindo aproximadamente 1,4 milhão de insetos. Cerca de 80 mil anos atrás, a distribuição de *R. robustus* II teria atingido as regiões de possível encontro com *R. robustus* III e IV. Cabe ressaltar que a sensibilidade do modelo *Skyride* para as estimativas feitas com apenas um *locus* permite apenas uma visualização das tendências gerais dos dados e não os detalhes finos, e somente é possível observar (quando presentes) fenômenos de expansão súbita que tenham ocorrido mais recentemente (Atkinson et al. 2008; Heled e Drummond 2009).

A análise genealógica de *R. robustus* III mostrou Rondon do Pará e São Domingos do Araguaia, Pará (porção leste da região biogeográfica do Pará) como regiões prováveis de origem do grupo. São Roberto (Maranhão) foi assinalada como possível região de origem de *R. robustus* n.sp. Portanto, Ulianópolis (Pará), a região de possível fluxo gênico entre as linhagens, é resultado de uma ocupação secundária. Estes resultados são condizentes com a hipótese de que *R. robustus* III e n.sp. são linhagens distintas e que o indivíduo de Rondon do Pará assinalado como possível migrante de *R. robustus* n.sp. possa ser uma retenção de polimorfismo ancestral. De maneira concordante, a região de possível encontro entre *R. robustus* III e *R. robustus* II e IV também parece ser uma ocupação secundária. De acordo com as análises de *Bayesian Skyride*, não há indícios de expansão populacional súbita em *R. robustus* II.

A genealogia de *R. robustus* IV foi sem dúvida a menos conclusiva. Provavelmente em função da amostragem ter sido falha na porção oeste e noroeste de

sua distribuição geográfica, ficou difícil definir a possível região de origem da linhagem. As regiões biogeográficas do Imerí e Guiana são as possíveis candidatas. Caso *R. robustus* V seja uma linhagem válida e diferente das demais, é possível que o Imerí tenha sido uma importante região de diferenciação de *R. robustus* s.l.

### 6.3.3. Hipóteses de origem e dispersão das linhagens

Uma das grandes limitações em estudos biogeográficos é a ausência de análises que possibilitem o teste de hipóteses biogeográficas para a diversificação dos organismos. Normalmente as afirmações são baseadas em estudos comparativos de diversidade de organismos pertencentes a diferentes níveis hierárquicos que coabitam uma mesma região biogeográfica (Avise 2009). Contudo, muitas vezes os critérios de separação são pouco intuitivos (cf. premissas mutuamente não-exclusivas em Morrone e Crisci, 1995) e minimalistas, com o emprego do método de máxima-parcimônia. Neste método, o número de passos mutacionais é minimizado nas topologias de árvore e, conseqüentemente, pode gerar erros quando as probabilidades de passos mutacionais entre os táxons são desiguais (Cunningham et al., 1998).

A partir da utilização do programa MIGRATE-N, foi possível extrapolar a teoria da coalescência aplicada ao fluxo gênico atual (coalescência recente) para testar hipóteses biogeográficas de origem e dispersão de linhagens que se separaram há milhões de anos (coalescência antiga). Como se trata de estimativas de fluxo gênico antigo, as estimativas dos parâmetros  $\Theta$  e  $M$  foram utilizadas apenas para examinar a convergência dos dados quando submetidos à diferentes modelos de redes de fluxo gênico (Beerli e Palczewski 2010).

Esta abordagem viola algumas premissas do modelo coalescente, comentadas a seguir, mas que dificilmente invalidam os resultados aqui obtidos (cf. Gonçalves da Silva 2012 para aplicação análoga com uma única espécie). Conforme observado nas análises populacionais e de *Bayesian Skyride*, *R. robustus* II viola a premissa de estabilidade demográfica. No caso de uma expansão recente súbita, ainda não se sabe qual a influência no ranqueamento dos dados por *LBF* (Beerli, 2009). Contudo, é esperado que este viés seja minimizado pelo fato das linhagens compartilharem variação genética em função de eventos migratórios antigos e ancestralidade comum (Goncalves da Silva et al., 2012). No caso de separações antigas, o programa

MIGRATE-N irá substituir pequenas taxas de migração (decorrentes de retenção de polimorfismo ancestral) por cladogêneses. Portanto, os modelos testados não são destinados a diferenciar como as linhagens se separaram (i.e. não diferenciam o tipo de migração). Com base na retenção de polimorfismo ancestral, os modelos designarão quais sequências se “assemelhariam” mais com a população estoque.

A finalidade desta análise foi a de testar estatisticamente, através do emprego de *LBF*, a hipótese de Abad-Franch e cols. (2009) referente à origem de *Rhodnius* na região do Orinoco e tentar compreender como a região amazônica foi colonizada pela população ancestral de *R. robustus s.l.*

Em uma macroescala, o modelo melhor ranqueado corroborou a hipótese de Abad-Franch e cols. de que as espécies de *Rhodnius* teriam sido originadas na região biogeográfica do Orinoco. Como o modelo mais aceito foi o de *source-sink 2*, não foi possível diferenciar para qual região biogeográfica a população estoque migrou/expandiu primeiro – se Napo ou Amazônia.

Em uma mesoescala, dois modelos foram aceitos com probabilidades semelhantes de estarem corretos, os modelos *stepping-stone 2* e *source-sink 2*. Em ambos os modelos, a origem de *R. robustus s.l.* teria sido em uma das seguintes regiões biogeográficas: Guiana, Inambarí, Imerí, Rondônia, Pará e Belém. Como os resultados da genealogia de *R. robustus IV* indicam a origem das espécies em Manaus ou Roraima, pode-se especular que as possíveis regiões biogeográficas de origem foram a Guiana e o Imerí.

O fato das cladogêneses serem muito antigas e as análises terem sido realizadas com um único *locus* pesaram na hora de diferenciar um modelo *source-sink* de um *stepping-stone* nas análises em macroescala. Neste caso, estas análises serviram para responder questões biogeográficas mais simples, embora muito relevantes para o conhecimento acerca da especiação em *Rhodnius* – a origem do complexo *R. prolixus* e *R. robustus s.l.* e a diversificação das linhagens amazônicas de *R. robustus s.l.* É provável que análises contendo marcadores nucleares possam ser ainda mais úteis para corroborar (ou refutar) as hipóteses levantadas aqui, uma vez que estes marcadores podem conter “assinaturas” referentes aos eventos coalescentes mais antigos, o que poderia recuperar com maior acurácia o polimorfismo ancestral.

Os resultados aqui obtidos, em conjunto com as deduções feitas a partir das genealogias de *R. prolixus* e *R. robustus* s.l. apóiam a hipótese proposta por Abad-Franch e cols. (2009) de relacionamento entre as ecorregiões da pan-Amazônia, obtidas com base nas análises de parcimônia de endemismo dos triatomíneos. As regiões biogeográficas da Guiana e Imerí estão proximamente relacionadas e formam um clado juntamente com o Orinoco (1), evidenciado pela separação de *R. prolixus*/*R. robustus* I/*R. robustus* V de *R. robustus* II, III, IV e n.sp.; e as regiões do Inambarí e Rondônia formam um clado separado de (1), evidenciado pela origem e dispersão de *R. robustus* II. Mesmo de caráter especulativo, propõe-se aqui que a partir da origem na região do Orinoco, populações da linhagem ancestral tenham ocupado a região do Imerí secundariamente, de onde teriam sido originadas *R. robustus* IV e V.

#### 6.4. As cladogêneses da linhagem *R. robustus* s.l. no contexto biogeográfico da região pan-amazônica

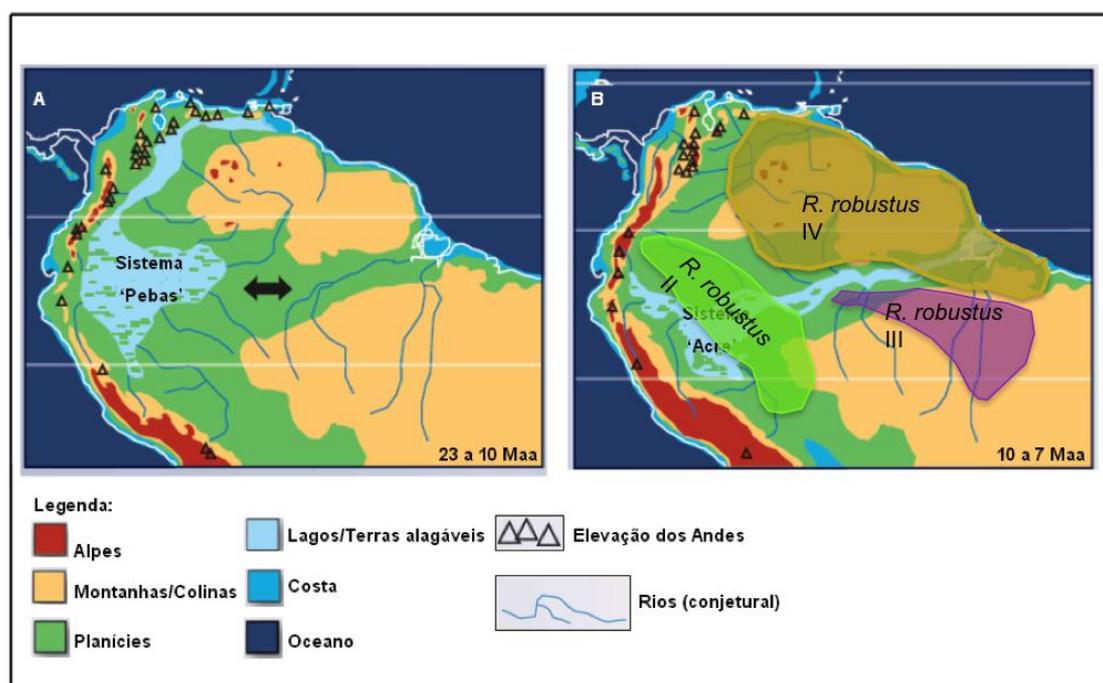
De acordo com as calibrações do relógio molecular relaxado, a ‘linhagem *robustus*’ e a ‘linhagem *pictipes*’ teriam se separado no período do Mioceno Inferior/Médio (23 - 15 Maa), durante as ‘incurções marinhas’ e a consequente formação do Sistema Pebas (Hoorn et al. 2010). A ‘linhagem *robustus*’ teria permanecido no Escudo das Guianas, na região biogeográfica atual do Orinoco.

O período calculado da cladogênese destas linhagens está de acordo com o proposto por Hwang e Weirauch (2012). Contudo, é importante ressaltar que esta datação pode conter erros em função do nó da raiz estar muito distante das calibrações dos nós mais superficiais utilizadas na topologia do filograma. Uma outra estimativa desta separação foi publicada por Gaunt e Miles (2002) e testada nesta tese (separação entre 56 e 46 Maa), mas esta calibração empírica (utilizando uma calibração ‘fixa’ do relógio molecular de Reduviidae) não parece estar de acordo com as calibrações sugeridas com base nos eventos biogeográficos vicariantes de orogênese dos Andes.

A Idade do Serravalliano, durante o Mioceno Médio (13 – 11 Maa), foi caracterizada por um período de rápido soerguimento da porção central e norte do oriente dos Andes (Hoorn et al. 2010). Estes eventos teriam gerado uma profunda incisão e erosão nas porções central e norte dos Andes, principalmente na vertente oriental (Andes Venezuelanos e leste da Colômbia), onde foram desprendidas grandes

porções de terras aluvionais, que se acumularam na base oriental dos Andes (Bermúdez et al. 2010).

Durante o Mioceno Médio/Superior (10 - 7 Maa), o sistema Pebas (Figura 6.2A), localizado nas terras baixas da porção oeste da Amazônia, sofreu modificações em função da diminuição do nível do mar e do aumento de chuvas na vertente oriental dos Andes (Figueiredo et al., 2009). O sistema antes lacustrino (ou marinho) foi modificado para fluvial (chamado de Sistema do Acre; Figura 6.2B), em função da diminuição do nível do mar e da grande ocorrência de chuvas. Os rios localizados ao noroeste da região modificaram seus cursos, iniciando o sistema fluvial como é conhecido atualmente (com a drenagem transcontinental em direção ao Oceano Atlântico). A porção de terras baixas dos Andes Venezuelanos e leste da Colômbia secaram por completo (Latrubesse et al. 2010).



**Figura 6.2.** (A) Sistema Pebas, que supostamente teria ocorrido entre 23 e 10 Maa. (B) Sistema Acre, que supostamente teria sido formado entre 10 e 7 Maa.

Note que o Rio Amazonas (antigo sistema Acre) divide as distribuições de *R. robustus* III e IV e, portanto, parece ter sido uma importante barreira para o fluxo gênico entre as linhagens. Apenas uma pequena parte da distribuição de *R. robustus* IV está localizada na margem oriental do Rio Amazonas, no norte do Pará. Possivelmente essa região deve ter sido ocupada secundariamente, durante o período de glaciação. Já *R. robustus* II ocupa toda a região amazônica de terras baixas, antes alagada no Sistema Acre. É possível que *R. robustus* II tenha ocupado a porção oriental do Rio Amazonas, na região biogeográfica de Rondônia e, de maneira súbita, expandido para as demais regiões há 80 mil anos. Modificado de Hoorn et al. (2010).

Os padrões filogeográficos obtidos apóiam as ‘incurções marinhas’ como a hipótese que melhor explica a diversificação de *R. prolixus* e *R. robustus s.l.*, pois os resultados obtidos satisfazem duas das três predições listadas a seguir (*sensu* Conn e Mirabello 2007): (a) as populações ancestrais e basais devem ter ocupado um dos blocos [população ancestral – região do Orinoco (estimada no modelo coalescente); população mais basal – Andes (*R. barretti*)]; e (b) populações mais derivadas devem se encontrar nas terras baixas da porção oeste amazônica (região central e oeste amazônico – *R. robustus* II, III, IV e n.sp.). A predição (c) de que as linhagens provenientes de cada um dos três blocos de maior altitude devem ser monofiléticas entre si não foi satisfeita, uma vez que *R. robustus* V, presente na região biogeográfica do Imerí, filogeneticamente está agrupada em um clado com *R. prolixus* e *R. robustus* I, que são naturais da região do Orinoco. Contudo, as datações dos eventos cladogenéticos obtidas não conferem com este período. Somente a cladogênese de *R. barretti* foi datada no período do Mioceno Médio/Superior (7,2 – 6,0 Maa). As demais cladogêneses foram todas datadas em um período posterior às ‘incurções marinhas’ do Mioceno, durante o Plioceno e Pleistoceno (entre 3,4 e 0,8 Maa).

O período do Mioceno Superior/Plioceno (7 – 2,5 Maa) foi o de maior radiação de plantas de altitude, roedores, pássaros e insetos, como abelhas (cf. Hoorn et al. 2010). Com o fechamento do istmo do Panamá, este período também ficou conhecido pelas altas taxas de migração de mamíferos provenientes da América do Norte que colonizaram regiões de altitude e regiões baixas da Amazônia (Marshall e Cifelli, 1990). Particularmente nas regiões baixas do oeste da Amazônia ocorreu a maior diversificação das florestas, promovendo a criação de novos nichos (Hoorn et al. 2010).

O período Quaternário foi marcado pelos ciclos glaciais e interglaciais ocasionados pela influência astronômica dos ciclos de Milankovitch (Bradley, 1999). Haffer (1969) e Vanzolini & Williams (1970) propuseram que neste contexto as florestas úmidas foram contraídas em pequenas porções separadas por vegetações de savana durante as glaciações e que estes ‘refúgios’ teriam mantido isoladas populações de diversas linhagens ancestrais, levando-as eventualmente à especiação.

Apesar das contradições em torno do fato das florestas úmidas tropicais terem se contraído e expandido durante os períodos glaciais e interglaciais (cf. Haffer 2001 e Hoorn 2010 para opiniões conflitantes), esta é a hipótese de especiação vicariante do Pleistoceno mais citada e utilizada (e.g. Aleixo, 2004; Cabanne et al., 2008; Dick et al., 2004; Haffer e Prance, 2002; Ribas et al., 2012).

Os resultados filogeográficos, populacionais e genealógicos obtidos em *R. prolixus* e *R. robustus s.l.* apóiam parcialmente as primeiras três das quatro previsões existentes para esta hipótese biogeográfica: (a) expansão populacional súbita (observada apenas em *R. robustus* II); (b) variabilidade genética baixa (todos os táxons analisados, com a exceção de *R. robustus* IV que apresentou porcentagens de divergência K-2p de até 1,5%); (c) baixa estruturação filogeográfica (evidenciada a partir das contradições observadas nas diferentes topologias das árvores filogenéticas dos marcadores mtCytb, ITS-2, *per*, *tim* e *cry2*); e (d) sinais de *bottleneck* (ou efeito ‘gargalo de garrafa’) na população, associados às contrações e isolamento, causados pelos ‘refúgios’ (nenhuma evidência foi encontrada).

A previsão mais importante para definir se a hipótese de ‘refúgios pleistocênicos’ é a hipótese biogeográfica que melhor explica o padrão filogeográfico obtido para os táxons estudados não foi satisfeita por completo, a de que “táxons amplamente distribuídos na região pan-amazônica refletem um processo súbito de crescimento populacional e de migração, ocorridos em períodos interglaciais, quando as florestas tropicais úmidas teriam se expandido” (Haffer e Prance 2002). Existe a possibilidade de a utilização de um único marcador molecular (mtCytb) não ter sido suficiente para detectar eventos de súbita expansão populacional mais antigos. Esta justificativa se adequa ao observado nos resultados das análises de *Bayesian Skyride*.

Apesar de não ter sido possível determinar qual hipótese biogeográfica, se ‘incursões marinhas’ ou ‘refúgios pleistocênicos’, melhor explicaria o padrão de diversificação revelado para *R. prolixus* e *R. robustus s.l.* na região pan-amazônica, os resultados indicam dicotomias evolutivas de larga amplitude no tempo geológico (desde o Mioceno Médio até o Pleistoceno Médio) e, portanto, é provável que mais de um evento biogeográfico tenha contribuído para as cladogêneses observadas. Apesar da amostragem utilizada não ser ideal para se testar a hipótese de que os rios foram barreiras para o fluxo gênico das populações ancestrais da linhagem, a partir da



Durante o início do Mioceno Superior, parte da população ancestral deve ter migrado para região trans-andina, ocupando o norte dos Andes. Nesta época, a cordilheira ainda não havia soerguido. Esta subpopulação ancestral possivelmente permaneceu isolada e especiou em alopatria, originando *R. neivai*. Durante o fim do Mioceno Superior, é provável que parte da população ancestral de *R. nasutus* tenha se isolado em regiões de altitude (e, portanto, mais secas), no Escudo da Guianas, entre 7,1 e 5,8 Maa (assumindo a taxa média de 1,6% de divergência para cada 1 Ma). Esta população teria permanecido nesta região até o Pleistoceno, período em que corredores áridos ligaram a região amazônica com a Caatinga (Silva e Bates, 2002; Webb, 1991). Esta população teria migrado e permanecido isolada neste bioma, originando *R. nasutus* (Pavan e Monteiro, dados não publicados).

Durante o Mioceno Superior/Plioceno, após a transição do sistema Pebas para o sistema Acre, a porção de terras baixas dos Andes Venezuelanos e leste da Colômbia secaram por completo e provavelmente serviram de rota para a dispersão da população estoque. De 10 a 7 Maa, o Sistema do Acre deve ter representado uma barreira importante para o fluxo gênico e, desta forma, *R. barretti* teria ocupado a região do Imerí e especiada em alopatria. De maneira alternativa, *R. barretti* pode ter migrado e se adaptado ao ambiente de florestas úmidas. Durante este período, as florestas aumentaram de densidade, graças ao fornecimento de nutrientes pela erosão e formação de depósitos dos Andes (Hoorn et al. 2010).

A alta diversificação das florestas nas regiões baixas do oeste da Amazônia possivelmente proporcionou a dispersão do estoque ancestral de *R. robustus* II, III e IV no Plioceno (4,6-2,7 Maa), ocupando inicialmente a região biogeográfica da Guiana e do Imerí. Neste mesmo período, *R. neglectus*, espécie endêmica do Cerrado, teria se separado da linhagem ancestral de *R. prolixus*/*R. robustus* I e especiada. É possível que sua linhagem ancestral tenha se dispersado pela região oeste cis-Andina e se adaptado as florestas secas formadas no fim do Plioceno/Pleistoceno durante os períodos glaciais (aproximadamente há 3,5 Ma, assumindo a taxa média de 1,6% de divergência para cada 1 Ma).

O Pleistoceno foi o período mais importante de diferenciação genética de *R. robustus* s.l. e *R. prolixus*. Entre 3,0 e 1,6 Maa, a população ancestral de *R. robustus* II teria sido originada na região biogeográfica de Rondônia. Recentemente, há 125 – 80

mil anos, *R. robustus* II possivelmente se dispersou para as regiões biogeográficas do Inambarí, Napo e Imerí. No período compreendido entre 2,1 e 1,2 Maa, uma nova dispersão da região do Orinoco para Amazônia teria ocorrido e a população ancestral alcançado a região biogeográfica do Imerí. De acordo com esta hipótese, é possível que *R. robustus* V seja encontrado ainda na região biogeográfica da Guiana. Posteriormente, entre 1,6 e 0,5 Maa, a população ancestral de *R. robustus* III/n.sp e IV teria dispersado da região da Guiana até o Pará e possivelmente se separado por um evento vicariante, originando *R. robustus* IV. A origem de *R. robustus* IV é controversa, podendo ter sido na região biogeográfica da Guiana (em Roraima) ou no Imerí (em Manaus, Amazonas). Por fim, parte desta população ancestral de *R. robustus* III/n.sp teria ocupado a região oeste do Maranhão e sofrido uma radiação adaptativa, dando origem à *R. robustus* n.sp. Posteriormente, a população de *R. robustus* n.sp. teria crescido e ocupado a região de possível encontro entre *R. robustus* III e IV. Já a origem de *R. robustus* III provavelmente foi no território que hoje faz parte do sudeste do estado do Pará e norte do Tocantins (entre Rondon do Pará e Estrela, respectivamente). Todos estes eventos de especiação podem ser igualmente explicados pela hipótese de ‘refúgios pleistocênicos’. Neste caso, as populações ancestrais teriam se isolado em fragmentos de floresta úmida e especiada em alopatria. Cabe ressaltar que nos períodos glaciais os rios amazônicos não devem ter representado barreiras difíceis de transpor, uma vez que estavam supostamente congelados, o que facilitaria a migração passiva em pelos de animais.

Cabe ressaltar aqui que os resultados originados nesta tese são apenas uma pequena “peça” de um “quebra-cabeça” muito mais complexo, uma vez que envolve diretamente a explicação dos eventos de especiação ocorridos apenas entre *R. prolixus* e *R. robustus* s.l.. Contudo esta peça representa um importante passo nos estudos biogeográficos em triatomíneos, por se tratar da primeira vez em que são realizados testes explícitos de hipóteses.

#### 6.5. Análise locomotora de *R. prolixus* e *R. robustus* II

Apesar de as ninfas de triatomíneos apresentarem um papel importante na transmissão da doença para humanos por serem capazes de transmitir o protozoário *T.*

*cruzi* assim como os adultos, ainda carecemos de informações acerca do comportamento de triatomíneos durante os estádios juvenis.

Adultos de *R. prolixus* apresentam atividade noturna bimodal, com picos de atividade durante as primeiras horas da noite, quando saem de seus refúgios em busca de alimento, e no fim da noite, quando regressam aos abrigos (Lorenzo e Lazzari, 1998). Com base nos resultados obtidos nesta tese, foi possível descobrir que, surpreendentemente, os estádios ninfais de *R. prolixus* e *R. robustus* II apresentam maior atividade durante o dia. Esta atividade também é bimodal, com picos no meio e no fim do dia, em ZT7 e ZT12 (o que corresponderia as 13h e 18h, respectivamente).

Em insetos, um padrão semelhante foi observado em *Gryllus bimaculatus*, cujas ninfas têm atividade diurna e os adultos atividade noturna (Uemura e Tomioka, 2006), assim como *R. prolixus* e *R. robustus s.l.* Esta diferença de atividade em diferentes estágios de um mesmo organismo parece estar relacionada à competição intraespecífica por alimento, decorrente da ocupação do mesmo nicho ecológico. Em vertebrados, como roedores, anfíbios e répteis (Haim e Rozenfeld, 1993; Oishi et al. 2004), e em invertebrados como crustáceos de água doce (Miranda-Anaya 2002), é possível observar esta mudança de fase da atividade quando organismos conspecíficos ocupam um mesmo nicho ecológico. Esta mudança de fase parece ocorrer, pelo menos em vertebrados, em resposta aos sinais químicos contidos nas fezes da outra espécie (Haim e Rozenfeld, 1993). Existem indícios de que esta resposta deve ser promovida de maneira circadiana por osciladores periféricos, como moléculas de *output* humorais, melatonina, ou ainda por outros substratos neurais (Vivanco et al., 2007, Schwartz et al., 2004).

Uma possível explicação para esta diferença observada nas atividades locomotoras entre ninfas e adultos de triatomíneos pode vir da própria biologia (e ecologia) de *R. prolixus* e *R. robustus* II. Os estágios juvenis dessas espécies, por serem ápteros, permanecem próximos aos ninhos de roedores e marsupiais nas copas das palmeiras. Como estes mamíferos são noturnos, o encontro entre o vetor e o hospedeiro juvenil tem maior chance de ocorrer durante o dia. Provavelmente, assim como nos adultos, o padrão bimodal de atividade explica a saída de refúgios em busca de alimento no meio da fase clara do dia (ZT7) e a procura por um refúgio no fim desta fase (ZT12). Para provar esta hipótese como sendo verdadeira, seria interessante

comparar a atividade locomotora de *R. prolixus* e *R. robustus s.l.* com uma espécie de *Rhodnius* estritamente associada a aves diurnas, como *R. neglectus*. Neste caso, em função da biologia do hospedeiro, espera-se que a atividade de ninfas de *R. neglectus* seja oposta (maior atividade durante a noite, período em que a ave se encontra no ninho).

*R. prolixus* e *R. robustus* II não apresentaram diferenças significativas quando os períodos de atividade foram comparados. Contudo, este resultado não descarta a hipótese de existirem barreiras naturais pré-zigóticas de isolamento temporal entre espécies simpátricas, uma vez que as linhagens testadas não se encontram na natureza (i.e. possuem distribuição disjunta). Infelizmente não foi possível manter em laboratório espécimes de linhagens distintas provenientes de áreas de simpatria. De qualquer forma, seria interessante testar futuramente a existência desta barreira natural ao fluxo gênico entre *R. prolixus* e *R. robustus* I provenientes de Trujillo, Venezuela, e entre *R. robustus* III e n.sp. provenientes de Ulianópolis, Pará.

Apesar de não haver uma diferença significativa entre *R. prolixus* e *R. robustus* II em relação aos picos de atividade diurna, a diferença entre a média dos períodos dos relógios circadianos foi significativa, ainda que não conclusiva em função da sobreposição de parte destas atividades (Figura 5.34). Essa diferença entre os períodos do relógio circadiano deve estar mais relacionada a uma questão evolutiva do funcionamento do relógio das diferentes espécies (do que ecológica), por se tratar de uma medição do tempo do ciclo das alças regulatórias principal e secundária na ausência de estímulos externos e, portanto, estritamente endógena.

Resultados acerca dos padrões de atividade e dispersão dos triatomíneos podem ser úteis para o desenvolvimento de novas estratégias de controle vetorial. Lazzari e Lorenzo (2009), com base nos padrões de atividade noturna dos triatomíneos adultos, propuseram a aplicação de inseticidas no interior das casas no horário no período inicial da noite, momento em que estão mais ativos e que permanecem mais tempo fora de seus abrigos. Após a descoberta inesperada de que as ninfas de *R. prolixus* são mais ativas durante à tarde (momento em que estão buscando alimento), a aplicação de inseticidas nas casas infestadas seria mais eficaz se realizada em dois períodos do dia (à tarde, entre 13 e 15h e entre 20 e 21h). Essa estratégia garantiria o sucesso de encontrar ninfas e adultos fora de seus abrigos (e

portanto mais vulneráveis) e ingurgitados, já que a distensão da cutícula abdominal após o repasto sanguíneo facilitaria a absorção dos componentes químicos (Lazzari e Lorenzo, 2009).

#### 6.6. Taxonomia molecular de *R. prolixus* e *R. robustus* s.l.

O status taxonômico de *R. robustus* s.l. foi questionado durante muitos anos em função de sua semelhança morfológica e isoenzimática com *R. prolixus* (Harry, 1993, 1994). A descrição de *R. robustus* foi feita por Larousse (1927) com base em dois espécimes, um proveniente da Guiana Francesa e outro da desembocadura do Rio Tefé (Amazonas). Naquele momento, Larousse desconfiou que *R. robustus* era uma espécie nova, pois apresentava um tamanho maior e uma coloração mais escura do que o holótipo de *R. prolixus*. Posteriormente, Lent e Jurberg (1969) e Lent e Wygodzinsky (1979) adicionaram novos caracteres diagnósticos à chave taxonômica de *Rhodnius* para identificação desta espécie, duas estruturas da genitália dos machos e diferenças na coloração da tíbia posterior de ninfas de quarto e quinto estádios.

Provavelmente tanto Larousse quanto Lent e Jurberg devem ter encontrado muitas dificuldades em encontrar caracteres morfológicos homogêneos em *R. robustus* e diagnóstico em relação à espécie críptica *R. prolixus*, pois sabemos hoje com base na região geográfica de ocorrência, que os espécimes utilizados na descrição original de *R. robustus* são provenientes de duas linhagens distintas – *R. robustus* II e *R. robustus* IV (Pavan e Monteiro, 2007). Como as características da genitália dos machos são muito variáveis e, portanto, não indicadas como caráter diagnóstico (Harry 1993), e as diferenças cromáticas restritas a dois estádio ninfais, a identificação morfológica precisa desta espécie é impossível.

Após a constatação de que *R. robustus* não só é apenas uma espécie válida e distinta de *R. prolixus*, como também um complexo de quatro linhagens crípticas (Lyman et al. 1999; Monteiro et al. 2003), houve a necessidade de se desenvolver um método de identificação simples, confiável e de baixo custo para distinguir *R. prolixus* de *R. robustus* s.l. e, portanto, suprir as limitações da chave taxonômica de identificação morfológica.

Pavan e Monteiro (2007) desenvolveram um PCR-Multiplex diagnóstico para *R. prolixus* e *R. robustus s.l.* que utiliza iniciadores espécie-específicos para a amplificação de fragmentos de tamanhos distintos do gene mitocondrial Cytb. Embora eficiente quando utilizado para identificar molecularmente indivíduos destas espécies, a identificação pode ser problemática em regiões de simpatria e possível introgressão de DNA mitocondrial de uma espécie em outra, conforme observado por Fitzpatrick e cols. (2008) na Venezuela. Para superar essa limitação, era necessária a criação de um método molecular complementar ao PCR-multiplex que utilizasse um marcador nuclear de cópia única.

#### 6.6.1. AmpG

Nesta tese foi possível isolar e testar um marcador nuclear de cópia única, o quarto íntron da proteína transmembrana 165 (TPS165), de possível localização no cromossomo X de *R. prolixus* e *R. robustus s.l.*, de caráter diagnóstico para *R. prolixus*. No marcador molecular isolado, esta espécie contém um SNP na posição 280 em relação a *R. robustus s.l.* (uma adenina, ao invés de uma guanina). O caráter diagnóstico do marcador foi comprovado após a genotipagem de 46 *R. prolixus*, dois *R. robustus* I, 83 *R. robustus* II, 19 *R. robustus* III e quatro *R. robustus* IV, coletados em uma ampla área geográfica.

A análise filogenética de *R. prolixus* e *R. robustus* I com base no sequenciamento deste marcador provou a natureza parafilética de *R. robustus s.l.*, evidenciada pelo relacionamento de *R. robustus* I mais próximo de *R. prolixus* do que de *R. robustus* II-IV.

Adicionalmente, *R. prolixus* é considerada uma espécie modelo em estudos bioquímicos e fisiológicos de insetos [e.g. Ribeiro et al. (1993) e Wigglesworth (1957)] e muitos laboratórios do mundo mantêm colônias destas espécies. Existem evidências de que a maioria destas colônias estão contaminadas com ou mal identificadas como *R. robustus s.l.* (Pavan, Lazoski, Dotson e Monteiro, dados não publicados). O SNP descrito (Anexo 1) poderá ser bastante útil para certificação molecular dessas colônias e de outras de *R. prolixus* mantidas em laboratórios mundo afora.

#### 6.6.2. Utilização dos indels de *timeless* como marcador diagnóstico

Após a identificação do gene ortólogo *timeless* (*tim*) de *R. prolixus*, o íntron 15 foi analisado em cinco *R. prolixus*, 16 *R. robustus* II, 16 *R. robustus* III, dois *R. robustus* n.sp. e seis *R. robustus* IV. Apesar de *tim* não separar filogeneticamente as linhagens de *R. robustus s.l.*, este marcador contém um polimorfismo de tamanho potencialmente diagnóstico para *R. prolixus*. O íntron 15 de *R. prolixus* contém oito inserções nucleotídicas (que variam de 1 a 156-pb) em comparação com as sequências de *R. robustus s.l.*, o que resulta em uma diferença de 210-pb entre as sequências de *tim* destas espécies. O caráter diagnóstico do íntron 15 de *tim* ainda precisa ser confirmado, com a adição de mais indivíduos de cada uma das linhagens de *R. robustus s.l.* e *R. prolixus*. Em caso de sucesso, será possível criar um método ainda mais simples que o AmpG para identificação deste complexo, uma vez que será necessário apenas amplificar as amostras de DNA de *R. prolixus* e *R. robustus s.l.* utilizando um par de iniciadores, submeter as amostras amplificadas à eletroforese em um gel de agarose 1% e comparar o tamanho dos fragmentos amplificados (731-pb para *R. prolixus* e 521-pb para *R. robustus s.l.*), sem a necessidade de sequenciamento.

#### 6.7. Utilidade dos novos marcadores nucleares na sistemática molecular de triatomíneos

Atualmente existe apenas um único marcador nuclear isolado em triatomíneos e útil em estudos de sistemática molecular envolvendo complexos de espécies crípticas – o ITS-2 (Marcilla et al., 2001). Apesar de largamente utilizado, existem diversos problemas associados ao seu uso em estudos populacionais ou filogenéticos (c.f. seção 6.1.4).

A técnica escolhida para a amplificação dos genes *per*, *tim* e *cry2* de *R. prolixus* foi a de anelamento dos iniciadores senso e anti-senso nos éxons para a cobertura de um íntron, ou EPIC (*Exon-Primed Intron-Crossing*). As regiões de anelamento destes iniciadores são bastante conservadas em insetos de maneira geral. Portanto, os iniciadores desenvolvidos nesta tese poderão ser úteis ainda para outras espécies de triatomíneos, mesmo aquelas não pertencentes ao gênero *Rhodnius*.

As sequências de *per* e *cry2* de *R. prolixus* e *R. robustus s.l.* apresentaram uma variabilidade genética moderada e alta, respectivamente. Portanto, é de se esperar que estes marcadores nucleares sejam bastante úteis em trabalhos de sistemática molecular de complexos de espécies filogeneticamente próximas, tanto em espécies do gênero *Rhodnius*, como em espécies de outros gêneros da subfamília Triatominae. Estes íntrons podem ter ainda potencial para serem utilizados em estudos de genética de populações.

As sequências de *R. prolixus* e *R. robustus s.l.* do íntron isolado de *tim* apresentaram menos sítios polimórficos do que os íntrons isolados de *per* e *cry2*. Logo, é provável que este marcador seja útil em estudos filogenéticos envolvendo espécies filogeneticamente mais distantes.

Até o momento já foi possível amplificar os fragmentos de *per*, *tim* e *cry2* de 11 espécies de triatomíneos (além de *R. robustus s.l.* e *R. prolixus*) pertencentes a três gêneros distintos: *Meccus longipennis*, *Meccus pallidipennis*, *Triatoma dimidiata*, *Triatoma sordida*, *Triatoma matogrosensis*, *Triatoma sherlocki*, *Triatoma brasiliensis*, *R. stali*, *R. pictipes*, *R. neglectus* e *R. nasutus* (resultados não mostrados).

## 7. CONCLUSÕES

Após a determinação do padrão filogenético e filogeográfico de *R. prolixus* e *R. robustus s.l.* foi possível concluir que:

(1) *R. prolixus* é uma espécie distinta de *R. robustus s.l.*, de acordo com os resultados dos sete marcadores moleculares utilizados, além das diferenças observadas entre os períodos do relógio circadiano.

(2) *R. robustus s.l.* representa um complexo parafilético de sete linhagens distintas, das quais três foram identificadas neste trabalho, *R. barretti*, *R. robustus V* e *R. robustus n.sp.*

(3) A distribuição geográfica das linhagens II, III e IV de *R. robustus s.l.* é disjunta e, surpreendentemente na região de peripatria de *R. robustus* III e IV ocorre uma linhagem até então desconhecida, denominada nesta tese de *R. robustus n.sp.*

(4) *R. robustus* III e n.sp. parecem hibridizar em Ulianópolis (Pará), região de encontro secundário entre essas linhagens. Essas linhagens representam um complexo de espécies incipientes, em *status nascendi*.

(5) Durante o Mioceno Superior/Plioceno Inferior (7,2-6,0 Maa), a população ancestral presente no Orinoco teria se expandido desde a porção de terras baixas dos Andes Venezuelanos até o oeste da Colômbia. Neste período, em função do Sistema do Acre, esse estoque teria sido isolado em florestas úmidas e divergido em alopatria, se diferenciando em *R. barretti*. Durante o Plioceno (4,6-2,7 Maa), a alta diversificação das florestas nas regiões baixas do oeste da Amazônia possivelmente proporcionou a dispersão do estoque ancestral de *R. robustus* II, III e IV, ocupando inicialmente a região biogeográfica da Guiana e posteriormente do Imerí. O Pleistoceno foi o período mais importante de diferenciação genética de *R. robustus s.l.* e *R. prolixus*. Eventos de migração e adaptação a diferentes microhabitats ou ainda a contração de florestas úmidas teriam sido responsáveis pela separação dos estoques ancestrais e divergência em alopatria, ocasionando a especiação de *R. prolixus* e *R. robustus* I, II, III, IV, V e n.sp.

(6) Cinco novos marcadores nucleares foram isolados do genoma de *R. prolixus* e analisados nesta tese. AmpG (TPS165) contém um SNP diagnóstico para *R. prolixus* e, portanto, pode ser muito útil identificar possíveis híbridos naturais. *per*, *tim*, *cry2-E6*, *cry2-E8* apresentaram baixo polimorfismo nas comparações entre as linhagens de *R. robustus s.l.* e, portanto, devem ser úteis apenas em estudos evolutivos contendo espécies filogeneticamente mais distantes.

(7) Apesar do marcador ITS-2 ser bastante útil em análises filogenéticas de espécies do gênero *Triatoma*, será muito difícil utilizá-lo em *Rhodnius*, uma vez que neste gênero este marcador de cópia múltipla não parece ter sido homogeneizado por completo por evolução em concerto, dificultando garantir a ortologia das sequências analisadas.

(8) *R. prolixus* e *R. robustus* II apresentam um ritmo endógeno do relógio circadiano significativamente diferentes, provando uma divergência evolutiva entre as espécies quanto ao funcionamento das alças regulatórias principal e secundária.

(9) Diferente dos adultos, as ninfas de *R. prolixus* e *R. robustus* II possuem atividade locomotora diurna, com picos de atividade em ZT 7 e ZT 12, que corresponderiam a saída em busca de alimento e regresso para os refúgios para passar a noite, respectivamente. Essa diferença no período de atividade entre ninfas e adultos parece envolver questões evolutivas relacionadas à competição intraespecífica por alimento, decorrente da ocupação do mesmo nicho ecológico, e uma associação histórica entre estes vetores e mamíferos noturnos.



Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: [www.elsevier.com/locate/meegid](http://www.elsevier.com/locate/meegid)

## 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)

Márcio G. Pavan<sup>a</sup>, Rafael D. Mesquita<sup>b</sup>, Gena G. Lawrence<sup>c</sup>, Cristiano Lazoski<sup>d</sup>, Ellen M. Dotson<sup>c</sup>, Sahar Abubucker<sup>e</sup>, Makedonka Mitreva<sup>e</sup>, Jennifer Randall-Maher<sup>e</sup>, Fernando A. Monteiro<sup>a,\*</sup>

<sup>a</sup> Laboratório de Sistemáticas Bioquímica, Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, Brazil

<sup>b</sup> Laboratório de Bioinformática, Departamento de Bioquímica, Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>c</sup> Centers for Disease Control and Prevention, Division of Parasitic Diseases and Malaria, Entomology Branch, 1600 Clifton Road NE, Atlanta, GA 30329, United States

<sup>d</sup> Laboratório de Biodiversidade Molecular, Instituto de Biologia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil

<sup>e</sup> The Genome Institute, Washington University School of Medicine, Campus Box 8501, St. Louis, MO 63108, United States

### ARTICLE INFO

#### Article history:

Received 19 June 2012

Received in revised form 16 October 2012

Accepted 23 October 2012

Available online 6 December 2012

#### Keywords:

*Rhodnius*  
Cryptic species  
SNP  
Nuclear DNA

### ABSTRACT

The design and application of rational strategies that rely on accurate species identification are pivotal for effective vector control. When morphological identification of the target vector species is impractical, the use of molecular markers is required. Here we describe a non-coding, single-copy nuclear DNA fragment that contains a single-nucleotide polymorphism (SNP) with the potential to distinguish the important domestic Chagas disease vector, *Rhodnius prolixus*, from members of the four sylvatic *Rhodnius robustus* cryptic species complex. A total of 96 primer pairs obtained from whole genome shotgun sequencing of the *R. prolixus* genome (12,626 random reads) were tested on 43 *R. prolixus* and *R. robustus* s.l. samples. One of the seven amplicons selected (*AmpG*) presented a SNP, potentially diagnostic for *R. prolixus*, on the 280th site. The diagnostic nature of this SNP was then confirmed based on the analysis of 154 *R. prolixus* and *R. robustus* s.l. samples representing the widest possible geographic coverage. The results of a 60% majority-rule Bayesian consensus tree and a median-joining network constructed based on the genetic variability observed reveal the paraphyletic nature of the *R. robustus* species complex, with respect to *R. prolixus*. The *AmpG* region is located in the fourth intron of the Transmembrane protein 165 gene, which seems to be in the *R. prolixus* X chromosome. Other possible chromosomal locations of the *AmpG* region in the *R. prolixus* genome are also presented and discussed.

© 2013 Elsevier B.V. All rights reserved.

### 1. Introduction

Chagas disease is transmitted through *Trypanosoma cruzi* infected feces of triatomine bugs (Hemiptera: Reduviidae) and is endemic in Latin America and in the Caribbean (Coura, 2007), ranking fourth in epidemiologic importance among all neglected diseases in those regions (Hotez et al., 2008). Although case incidence estimates show a 90% decline (from 500,000 to 50,000 new cases per year), 8 million people are still infected and 14,000 die every year (Senior, 2007). Domestic insect vectors are responsible for most of the disease transmission to humans. Because no vaccine is available, the elimination of household-infesting triatomines is the

major goal of many Chagas disease control programs (Abad-Franch et al., 2010).

The design and implementation of any vector control intervention must rely on accurate species identification. However, the identification of Chagas disease vectors based on morphological characters alone can be technically challenging, since some groups of species are cryptic (i.e., isomorphic), and therefore not amenable to be distinguished in such a manner. This taxonomical issue is even more relevant when applied to the transmission of *T. cruzi* by *Rhodnius prolixus* and *Rhodnius robustus* s.l. to humans. Although morphologically indistinguishable, they play very different epidemiological roles: the former are efficient domestic vectors, and the latter are not, as all species of the complex are entirely sylvatic (Monteiro et al., 2003).

*Rhodnius prolixus* is one of the most important Chagas disease vectors in Latin America, particularly in Venezuela, Colombia, and Central America, due to its ability to efficiently colonize human dwellings. Although certain countries such as Guatemala have successfully controlled the transmission mediated by this species

\* Corresponding author. Address: Laboratório de Sistemáticas Bioquímica, Instituto Oswaldo Cruz, FIOCRUZ, Avenida Brasil, 4365, Pavilhão Arthur Neiva, sala 21A/22, Mangueiras, Rio de Janeiro, RJ 21045-900, Brazil. Tel.: +55 21 25621255; fax: +55 21 22803740.

E-mail address: [fam@ioc.fiocruz.br](mailto:fam@ioc.fiocruz.br) (F.A. Monteiro).

(Guhl et al., 2009), Venezuela and Colombia still face serious problems of re-infestation of insecticide-treated houses by sylvatic populations (Fitzpatrick et al., 2008; Lopez et al., 2007). These events are commonplace in endemic regions where vectors are autochthonous, and will demand the development of innovative surveillance/control strategies that contemplate the complexities of sylvatic populations dynamics (Abad-Franch and Monteiro, 2005).

To allow for the accurate taxonomic identification of *R. prolixus* and the members of the *R. robustus* cryptic species complex, Pavan and Monteiro (2007) have developed an objective and cost-effective mitochondrial DNA (mtDNA) multiplex-PCR assay. However, due to its mitochondrial-based nature, this method could generate misleading results in areas where *R. prolixus* and members of the *R. robustus* complex might come into contact and hybridize, causing mitochondrial DNA introgression (Fitzpatrick et al., 2008).

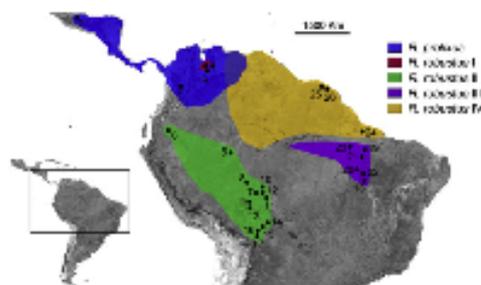
Thus, in order to overcome such limitation, we describe a new non-coding, single-copy nuclear DNA fragment containing a single-nucleotide polymorphism (SNP) with the potential to separate *R. prolixus* from members of the *R. robustus* cryptic species complex.

## 2. Materials and methods

### 2.1. Molecular identification of specimens and sampling

Since accurate taxonomic identification of *R. prolixus* and the four cryptic species of *R. robustus* [sensu Monteiro et al. (2003)] based on morphology alone is problematic, all specimens used in this study were identified through molecular taxonomy based on a fragment of the mitochondrial cytochrome *b* gene (Pavan and Monteiro, 2007). Moreover, we sequenced the ribosomal second internal transcribed spacer (ITS-2) of all specimens analyzed (data not shown), according to Marcilla et al. (2001), to identify possible introgression of mitochondrial DNA specimens.

For an exploratory screening aimed at finding unique, non-coding regions that separate these species, we tested 25 *R. prolixus* from two well established laboratory colonies kept at the Centers for Disease Control and Prevention (CDC) in Atlanta (19 specimens) and at the Medical Entomology Research and Training Unit in Guatemala (six specimens), plus two field-collected *R. prolixus* from Venezuela, two *R. robustus* I, four of each *R. robustus* II and III, and six *R. robustus* IV [sensu Monteiro et al. (2003)]. This approach was done during the initial phase of *Rhodnius prolixus* genome project, named survey sequencing, aimed to select a non-introgressed *R. prolixus* colony with low genetic variability. Fifty-two colony specimens, 43 *R. prolixus* and nine *R. robustus* II, and 102 field-collected insects, three *R. prolixus*, two *R. robustus* I, 74 *R. robustus* II, 19 *R. robustus* III, and four *R. robustus* IV (Fig. 1 and Table 1), were



**Fig. 1.** Geographical distribution of sampled specimens of *R. prolixus* and *R. robustus* s.l. Each location is numbered according to Table 1. The colors indicate the range of *R. prolixus* and the four members of the *R. robustus* cryptic species complex (*R. robustus* I–IV).

assayed to confirm the diagnostic applicability of possible species-specific SNPs.

### 2.2. DNA purification and library construction

DNA from the CDC and Guatemala Colony samples were extracted from ovaries or testis using the Promega genomic extraction kit. For all other *Rhodnius* specimens, one or two legs of a single *Rhodnius* specimen was placed in a 1.5 mL microtube, dipped into liquid nitrogen until frozen, and ground to a powder with Kimble/Kontes Pellet Pestles<sup>®</sup> prior to standard Genomic DNA Extraction with a Real Genomics<sup>™</sup> kit.

From a small WGS plasmid library, 12,626 random Genome Survey Sequences (GSS) were generated as described by Abubucker et al. (2008), using genomic DNA extracted from a population of *R. prolixus* from the Atlanta colony, and running on an ABI 3730 automated sequencer. Ten thousand GSS reads were screened for non-coding regions, searched against a transposon database, then against RFAM (Gardner et al., 2011), and finally against a non-redundant protein database (with an *e*-value cutoff of  $10^{-5}$ ). Regions with hits were masked. Vector sequences were removed and sequences with unmasked regions of 400–500 base pairs (bp) were extracted and sent through an in-house primer-calling pipeline. Primers with the following criteria were picked: (1) all bases of each primer in the pair must have a quality score of at least Q25; (2) minimum and maximum acceptable amplicon size was 400 and 500, respectively; and (3) excluded regions were 150-bp long (i.e., region with repetitive elements in the middle of the amplicon that would compromise primer annealing).

### 2.3. PCR amplification and DNA sequencing

From the putative unique, non-coding regions, 96 primer pairs were designed for the PCR amplification of regions of approximately 400–500-bp in length (Table A.1), during the survey sequencing approach.

PCR amplifications of the nuclear gene fragments were performed in a Veriti<sup>®</sup> 96 Well Thermal Cycler (Applied Biosystems), programmed for one denaturation step at 96 °C for 5 min, followed by 35 cycles at 94 °C for 30 s, 60 °C for 30 s, and 72 °C for 45 s, and a final 10-min extension step at 72 °C. Each 25 µL final volume reaction, contained 5–10 ng of DNA template, 2.5 µL buffer 10×, 2.5 mM of MgCl<sub>2</sub>, 1.25 units of Taq polymerase (Biotools), 0.5 mM of dNTPs, and 10 pmol of each primer. Purification of PCR products was performed with the Hi Yield<sup>™</sup> Gel/PCR DNA Extraction Kit (Real Genomics<sup>™</sup>), and both fragment strands were subjected to fluorescent dye-terminator cycle sequencing reactions (ABI Prism<sup>®</sup> BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems), using the same primers above, and run on an ABI 3730 automated sequencer.

### 2.4. Analyses of generated DNA sequence data

The removal of primer sequences, editing of both forward and reverse strands, and the generation of a consensus sequence for each sample were done using the SEQMAN LASERGENE 7.0 program (DNASTAR, Inc.). Sequences were aligned using the default parameters of CLUSTALW2 program (Larkin et al., 2007). Polymorphic sites were identified using MEGA 5 (Tamura et al., 2011). RECON (Bao and Eddy, 2002), and BLASTX (Altschul et al., 1997) were used to identify repeated regions and coding sequences, respectively.

A Bayesian phylogenetic tree constructed based on the diagnostic amplicons for 154 *R. prolixus* and *R. robustus* s.l. sequences was inferred in BEAST 1.7 (Drummond et al., 2012). Bayesian Information Criterion (BIC) in jMODELTEST (Posada, 2008) was used to elect Jukes and Cantor (1969) as the best-fit evolutionary model (delta

**Table 1**

Information on the samples used in this study. Field-collected samples are numbered according to Fig. 1.  $N_A$  – number of individuals tested for seven amplicons (AmpA to AmpG);  $N_B$  – number of individuals sequenced for AmpG. <sup>1</sup>Centers for Disease Control and Prevention; <sup>2</sup>Medical Entomology Research and Training Unit; <sup>3</sup>Instituto Oswaldo Cruz; <sup>4</sup>Instituto Nacional de Pesquisa da Amazonia.

| Species                                       | Collection site/colony location                         | Geographical coordinates                   | $N_A$           | $N_B$ | Field/colony | Date of collection | GenBank accession numbers    |                    |
|---|---|--|-----------------|-------|--------------|--------------------|------------------------------|--------------------|
| <i>R. prolixus</i>                            | 1. Colombia/CDC <sup>1</sup> , Atlanta, USA             | –  | 19              | 10    | Colony       | –                  | JQ432871 to 432880           |                    |
|   | 2. Guatemala/MERTU <sup>2</sup> , Guatemala             | –  | 6               | 10    | Colony       | –                  | JQ432881 to 432890           |                    |
|   | “Unknown”/IOC <sup>3</sup> , Brazil                     | –  | –               | 23    | Colony       | –                  | JQ432844 to 432866           |                    |
|   | 3. Barinas, Venezuela                                   | 08°37'N 70°12'W                            | 2               | 3     | Field        | 2004               | JQ432891 to 432893           |                    |
| <i>R. robustus</i> I                          | 4. Trujillo, Venezuela                                  | 09°22'N 70°25'W                            | 2               | 2     | Field        | 2004               | JQ432894, JQ432895           |                    |
| <i>R. robustus</i> II                         | 5. Caruaru, Amazonas, Brazil/INPA <sup>4</sup> , Brazil | –  | –               | 4     | Colony       | –                  | JQ432867 to 432870           |                    |
|   | 6. Sucumbios, Ecuador                                   | –  | –               | 5     | Colony       | –                  | JQ432896 to 432900           |                    |
|   | 7. Monte Negro, Rondônia, Brazil                        | 10°15'N 63°17'W                            | 4               | 2     | Field        | 2004               | JQ432961, JQ432962           |                    |
|   | 8. Porto Velho, Rondônia, Brazil                        | 08°44'S 63°25'W                            | –               | 45    | Field        | Apr/2009           | JQ432916 to 432960           |                    |
|   | 9. Cacoal, Rondônia, Brazil                             | 11°26'S 61°27'W                            | –               | 2     | Field        | Apr/2009           | JQ432973, JQ432974           |                    |
|   | 10. Santo Antônio, Rondônia, Brazil                     | 12°31'S 63°33'W                            | –               | 1     | Field        | Feb/2010           | JQ432971                     |                    |
|   | 11. Guajará Mirim, Rondônia, Brazil                     | 10°47'S 65°20'W                            | –               | 2     | Field        | Jul/2010           | JQ432969, JQ432970           |                    |
|   | 12. Ouro Preto do Oeste, Rondônia, Brazil               | 10°43'S 62°15'W                            | –               | 7     | Field        | Oct/2009           | JQ432963 to 432968, JQ432972 |                    |
|   | 13. San Gabriel, Pando, Bolivia                         | 10°10'S 67°03'W                            | –               | 2     | Field        | Sep/2007           | JQ432912, JQ432913           |                    |
|   | 14. Guarayos, Santa Cruz, Bolivia                       | 15°52'S 63°20'W                            | –               | 7     | Field        | Mar/2008           | JQ432901 to 432907           |                    |
|   | 15. San Ramón, Santa Cruz, Bolivia                      | 16°29'S 62°29'W                            | –               | 1     | Field        | Mar/2008           | JQ432908                     |                    |
|   | 16. Montero, Santa Cruz, Bolivia                        | 17°17'S 63°08'W                            | –               | 3     | Field        | Mar/2008           | JQ432909 to 432911           |                    |
|   | 17. El Torno, Santa Cruz, Bolivia                       | 18°04'S 63°18'W                            | –               | 2     | Field        | Sep/2007           | JQ432914, JQ432915           |                    |
|   | <i>R. robustus</i> III                                  | 18. Araguaína, Pará, Brazil                | 07°11'S 48°12'W | –     | 2            | Field              | Nov/2009                     | JQ432975, JQ432976 |
|   |   | 19. Ulianópolis, Pará, Brazil              | 03°43'S 47°29'W | –     | 7            | Field              | Nov/2009                     | JQ432977 to 432983 |
|   |   | 20. São Domingos do Araguaia, Pará, Brazil | 05°30'S 48°44'W | –     | 3            | Field              | Nov/2009                     | JQ432984 to 432986 |
|   |   | 21. Rondon do Pará, Pará, Brazil           | 04°54'S 48°20'W | –     | 2            | Field              | Nov/2009                     | JQ432987, JQ432988 |
| 22. São Bento do Tocantins, Tocantins, Brazil |   | 06°02'S 47°55'W                            | –               | 3     | Field        | Nov/2009           | JQ432989 to 432991           |                    |
| 23. Novo Repartimento, Pará, Brazil           |   | 04°20'S 49°50'W                            | 4               | 2     | Field        | Aug/1998           | JQ432992, JQ432993           |                    |
| <i>R. robustus</i> IV                         | 24. Santa Maria, Pará, Brazil                           | 01°20'S 47°32'W                            | –               | 1     | Field        | Oct/2009           | JQ432994                     |                    |
|   | 25. Cayenne, French Guiana                              | 04°56'N 52°20'W                            | 5               | 2     | Field        | Jan/2003           | JQ432995, JQ432996           |                    |
|   | 26. Rémire, French Guiana                               | 04°53'N 52°16'W                            | 1               | 1     | Field        | Aug/2001           | JQ432997                     |                    |
| Total   |   |  | 43              | 154   |              |                    |                              |                    |

AIC = 0) for the data set. Tree prior was randomly generated and Yule process of speciation was imposed for all tree reconstructions. Two independent runs were performed for  $10^7$  generations, with a burn-in of  $10^6$  generations. Convergence of parameters and a proper mixing were confirmed by calculating the Effective Sample Size (ESS) in TRACER 1.5 (Drummond and Rambaut, 2007), excluding the initial 10% (burn-in) of each run. All considered parameters showed ESS values above  $10^4$ . Runs were combined using LOGCOMBINER, and a maximum credibility tree based on the 20,000 trees generated (burn-in = 2000) and a posterior probability limit of 0.6 was produced using TREE ANNOTATOR (both programs part of the BEAST package). Statistical support for clades was assessed by the posterior probability method, and trees were visualized in FIGTREE v.1.0 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Genealogy of the DNA sequences generated for the selected diagnostic amplicon was also inferred with NETWORK 4.6 (Fluxus-Engineering), using the Median-Joining network method (Bandelt et al., 1999), with a Maximum-Parsimony (MP) calculation in the post-processing. Character weights and epsilon values were given a value of 10 (transition/transversion weighting was 1:1, since we analyzed a non-coding region). Also the Maximum-Composite Likelihood (M-CL) algorithm (Tamura et al., 2004) with 1000 bootstrap replicates of variance, available in MEGA 5 (Tamura et al., 2011), was used to construct intra and interspecific distance matrices.

## 2.5. Genomic search of selected amplicon

Using the MEGABLAST web service ([http://www.ncbi.nlm.nih.gov/sutils/genom\\_table.cgi](http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi)), the diagnostic amplicon was searched for in the following 40 complete or partial arthropod genomes:

*Acromyrmex echinator*, *Acyrtosiphon pisum*, *Aedes aegypti*, *Anopheles gambiae*, *Anopheles darlingi*, *Apis mellifera*, *Apis florea*, *Atta cephalotes*, *Bombus impatiens*, *Bombus terrestris*, *Bombyx mori*, *Camponotus floridanus*, *Culex quinquefasciatus*, *Daphnia pulex*, *Drosophila melanogaster*, *Drosophila pseudoobscura pseudoobscura*, *Drosophila simulans*, *Drosophila yakuba*, *Drosophila ananassae*, *Drosophila erecta*, *Drosophila grimshawi*, *Drosophila mojavensis*, *Drosophila persimilis*, *Drosophila sechellia*, *Drosophila virilis*, *Drosophila willistoni*, *Harpegnathos saltator*, *Ixodes scapularis*, *Lepeophtheirus salmonis*, *Mayetiola destructor*, *Nasonia giraulti*, *Nasonia longicornis*, *Pediculus humanus corporis*, *Pogonomyrmex barbatus*, *Rhipicephalus microplus*, *R. prolixus*, *Solenopsis invicta*, *Varroa destructor*, *Nasonia vitripennis*, and *Tribolium castaneum*. A second search was done in the BLASTN web service ([http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST\\_PROGRAMS=megaBlast&PAGE\\_TYPE=BlastSearch&SHOW\\_DEFAULTS=on&LINK\\_LOC=blasthome](http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome)), with standard parameters of word size = 7 and e-value = 10, and the NCBI nucleotide (nt) database arthropod (taxid 6656) subset.

## 2.6. Selected amplicon genomic contig localization and chromosome synteny

### 2.6.1. Databases

The *R. prolixus* genome assembly v.3.0 and the hard-masked contig GL563069 (“N” masked transposons, repetitive and low complexity regions) were downloaded from VectorBase (<http://www.vectorbase.org>). Transposable elements and *Rhodnius* EST databases were kindly provided by Dr. J.M. Ribeiro. *A. mellifera*, *B. terrestris* and *T. castaneum* protein sequences were extracted from the NCBI non-redundant (nr) database using a Perl script that selected and extracted all sequences related to any NCBI taxonomic

index point (taxids used were respectively 7460, 30195 and 7070), using taxdump and gi-taxid-prot NCBI relational datasets (<ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz>, and [ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi\\_taxid\\_prot.dmp.gz](ftp://ftp.ncbi.nih.gov/pub/taxonomy/gi_taxid_prot.dmp.gz)). The *T. castaneum* genome assembly version 3.0 was downloaded from BeetleBase ([ftp://bioinformatics.ksu.edu/pub/BeetleBase/3.0/Tribolium\\_genome\\_sequence.fasta](ftp://bioinformatics.ksu.edu/pub/BeetleBase/3.0/Tribolium_genome_sequence.fasta)). *R. prolixus* genes, transcripts and peptides belonging to prediction version C1 (VectorBase gene prediction pipeline using assembly v3.0) were also downloaded from VectorBase (named VectorBase prediction). Alternative gene prediction version Lagerblad 3.0 was done using GENEID v1.3 (Blanco et al., 2007) trained with a large *R. prolixus* EST database in genome assembly v3.0 (named Lagerblad and available in *R. prolixus* genomic browser at VectorBase). Transposable elements were identified in contig GL563069 using local RPSBLAST, BLASTN and TBLASTX (Tatusova, 2010), and transposable element databases.

#### 2.6.2. Selected amplicon localization in genomic contigs

The selected amplicon region was searched on *R. prolixus* genomic contigs using BLASTN (Tatusova, 2010).

#### 2.6.3. Genes and transposons curation in amplicon flanking regions

Genes from both predictions flanking the diagnostic amplicon region were searched for chimeras using BLASTP, NCBI nr protein, UNIPROT databases, and TBLASTN of *Rhodnius* EST database. Chimera genes were split manually to allow gene synteny analysis. Gene prediction revision was done manually on ARTEMIS (Rutherford et al., 2000) based on EST database hits and GENEWISE (Birney et al., 2004) analysis using contig GL563069 and coded proteins best-hit in nr and UNIPROT databases. Transposable elements revision was also carried out in GENEWISE with BLASTN and TBLASTX hits, using TE database.

#### 2.6.4. Gene synteny

Reciprocal BLASTP best-hit strategy (Tatusov et al., 1997) was used to identify homologs at the protein level for 14 genes in the diagnostic amplicon flanking region. We used in the comparison species with chromosome-mapped genomes (*A. mellifera*, *B. terrestris* and *T. castaneum*). The *R. prolixus* database included predicted proteins from both VectorBase and Lagerblad predictions. Bee and beetle homolog proteins GIs from the NCBI protein database (<http://www.ncbi.nlm.nih.gov/protein>) were used to identify coding genes and their chromosome location. Transposable elements were not considered in this analysis.

#### 2.6.5. Whole contig synteny

*Rhodnius prolixus* hard-masked contig GL563069 was compared with all *T. castaneum* chromosomes separately using TBLASTX with a  $10^{-10}$  e-value cutoff. The results were loaded in ACT (Rutherford et al., 2000) and limited by 50 or 70% identity cutoff when confirming selected gene homolog location, and 50% identity in GL563069 open reading frames (ORFs) larger than 100 bp when analyzing entire contig synteny.

### 3. Results

#### 3.1. Amplicon selection

Seven amplicons were selected based on reproducibility, variability, and presence of species-specific sites across all taxa tested. In this first screening, 27 *R. prolixus*, two *R. robustus* I, four of each *R. robustus* II and III, and six *R. robustus* IV samples PCR-amplified and sequenced for these seven nuclear loci yielded 13 polymorphic sites, nine of which are SNPs (Table A.2). One such SNP located on the 280th site of the 364-bp region amplified with primers

*AmpGF*, 5'-GAG AGC TGA AGA TAG GCA AGC G, and *AmpGR*, 5'-TGA TAA CTG GAT TAG GCG CAG C was diagnostic for *R. prolixus* (i.e., all *R. prolixus* have an adenine, instead of a guanine, on that particular site). To further evaluate the reliability and usefulness of this particular SNP to serve as an important diagnostic marker for the identification of *R. prolixus*, an extensive effort was made to obtain the greatest amount of *R. prolixus* and *R. robustus* s.l. samples aimed at covering the widest geographical distribution for the species. A total of 154 specimens, 46 *R. prolixus*, two *R. robustus* I, 83 *R. robustus* II, 19 *R. robustus* III, and four *R. robustus* IV, were sequenced for the 364-bp *AmpG* region. The 280th site still proved to be diagnostic. This autapomorphic character of *R. prolixus* is preceded by 5'-GTCATAAGA and followed by 5'-TCGTTGGTAG conserved sequences for all species analyzed.

Ten polymorphic sites were observed, eight of which are shared between *R. prolixus* and *R. robustus* I, and three of those are parsimony-informative sites (positions 174, 270, and 360; Table 2). The paraphyletic assemblage of *R. robustus* s.l., previously observed with mitochondrial cytochrome b and ribosomal DNA sequencing analyses (Monteiro et al., 2003), is corroborated here, based on a Bayesian phylogenetic tree reconstruction and a Median-Joining network with Maximum-Parsimony post-processing (Fig. 2). The *R. robustus* I clade is more closely related to the *R. prolixus* clade than to the other *R. robustus* clades. The *R. robustus* I and *R. prolixus* sister-species status is also evidenced by M-CL distance comparisons, as sequence divergence between these species is lower than between *R. robustus* I and *R. robustus* II-IV (0.6% in comparison with 1.1–1.7%; Table 3).

#### 3.2. Amplicon G genomic contig and flanking region

The *AmpG* region is located in the *R. prolixus* genomic contig GL563069 from positions 1116977 to 1117340. It was not found in any other genome searched, not even when high e-values and smaller word sizes were used (allowing for more dissimilar matches to occur), which suggests that this region is unique.

Genes in the *AmpG* flanking region (contig GL563069 from 925995 to 1319022 positions) were curated (Fig. 3B) to remove chimera and wrong predictions. Lagerblad genes 17948\_72 and 17948\_73 were joined (*Src tyrosine kinase*) and 17948\_78 was split into two genes (*X11Lbeta* and *Transmembrane protein 165*). Genes 17948\_79 and 17948\_84 were wrong predictions and then were removed (not shown). No chimera or false positive gene was found in the VectorBase prediction, but four genes were absent (17948\_70 – *Ephrin*, 17948\_75 – *Sarcoplasmic calcium-binding protein*, 17948\_81 – *Interleukin 16* and 17948\_88 – *KN motif and Ankyrin repeat domain-containing protein*). Final genes in the *AmpG* flanking regions were: *Ephrin*, *Ribonuclease*, *Src tyrosine kinase*, *Olfactory receptor*, *Sarcoplasmic calcium-binding protein*, *Small glutamine-rich tetratricopeptide repeat-containing protein alpha*, *X11Lbeta*, *Transmembrane protein 165*, *Interleukin 16*, *WNT4 protein*, *UNC-112 related protein*,

**Table 2**

Polymorphic sites of *AmpG* sequences (364 bp) observed in 154 *Rhodnius prolixus* and *R. robustus* I-IV. The column in bold highlights the single-nucleotide polymorphism (SNP) that separates *R. prolixus* from members of the *R. robustus* cryptic species complex (site 280). A dot (.) indicates identity with the consensus nucleotide (in this case, the consensus sequence is from *R. prolixus*). Codes: A – adenine; C – cytosine; G – guanine; T – thymine; S – cytosine or guanine; R – guanine or adenine; K – guanine or thymine; Y – cytosine or thymine.

| Species                | 74 | 80 | 167 | 174 | 270 | <b>280</b> | 321 | 337 | 341 | 360 |
|------------------------|----|----|-----|-----|-----|------------|-----|-----|-----|-----|
| <i>R. prolixus</i>     | C  | G  | G   | A   | C   | <b>A</b>   | G   | C   | C   | A   |
| <i>R. robustus</i> I   | S  | -  | -   | -   | -   | <b>G</b>   | -   | -   | -   | -   |
| <i>R. robustus</i> II  | -  | R  | R   | T   | A   | <b>G</b>   | K   | Y   | S   | T   |
| <i>R. robustus</i> III | -  | -  | -   | T   | A   | <b>G</b>   | T   | T   | G   | T   |
| <i>R. robustus</i> IV  | -  | -  | -   | T   | A   | <b>G</b>   | T   | -   | -   | T   |



NADH dehydrogenase, KN motif, and ankyrin repeat domain-containing protein and Band 4.1-like protein 4 (Fig. 3B). The detailed information about this region was included in Table A.3.

The *AmpG* region was found in the fourth intron of the *Transmembrane protein 165* gene. This gene's original predictions (RPRC07420-RA and 17948\_78) were manually curated (Fig. 3A), including complete EST support. The last intron of this gene also has an *Outcast* transposable element (class I) with a truncated reverse transcriptase (Fig. 3A). Other transposable elements identified were in intergenic regions (Fig. 3B).

### 3.3. Amplicon G chromosome localization

Arthropod choice for synteny analysis was based on the following criteria: (1) to be the phylogenetically closest organisms with chromosome-mapped genomes available; and (2) to have the closest number of chromosomes to *R. prolixus*, which was known to be 20 autosomes, plus XY (Panzeria et al., 2007). *R. prolixus* (taxid 13249) taxonomic classification in the NCBI taxonomic web database (<http://www.ncbi.nlm.nih.gov/taxonomy>) was followed back by tracing the branches of the Insecta phylogenetic tree toward its trunk to search for organisms with chromosome-mapped genomes. Endopterygota Infraclass (taxid 33392), inside Neoptera Subclass (taxid 33340) was the closest taxonomic point that satisfied those two mentioned criteria, where the following 13 species genomes (and their respective number of chromosomes in parenthesis) were found: *A. gambiae* (4), *D. pseudoobscura* (7), *A. mellifera* (16), *Drosophila miranda* (6), *B. terrestris* (18), *M. destructor* (4), *N. vitripennis* (5), *D. virilis* (6), *T. castaneum* (10), *D. simulans* (6), *D. yakuba* (7), *D. melanogaster* (7) and *A. aegypti* (3). Based on the second criterion, we chose for the analysis two bees, *A. mellifera* and *B. terrestris*, and the beetle *T. castaneum*.

Almost all homolog genes of the *AmpG* flanking region were found in the three selected genomes, when we used reciprocal best-hit strategy in gene synteny analysis (Tatusov et al., 1997). The only exception was in the *T. castaneum* analyzed region, where we did not confirm a homolog for the *WNT4 protein* (GI 91086553; Fig. 3C), as the best-hit was not reciprocal to *R. prolixus*. However, this protein has the same chromosomal localization as UNC-112 related protein and NADH dehydrogenase (Fig. 3C), and thus was not discarded from analysis. X11Lbeta homolog searches in bees did not retrieve any similar gene, probably due to a lack of prediction, as the TBLASTN web search showed similar genomic regions in unmapped genomic scaffolds (data not shown). *Transmembrane protein 165* gene contained the *AmpG* region inside its fourth intron, but it was the only protein that exhibited homologs located in different chromosomes in all three compared insects (Fig. 3C). Many homolog *R. prolixus* genes were not found, or their chromosome localization did not suggest a pattern in bee genomes, especially *B. terrestris*. Considering *A. mellifera*, four of the first six homolog genes were localized in chromosome 1 (CHR1), but this pattern did not seem to overlap the *AmpG* region. Moreover, no other pattern was observed until the end of the gene block. *T. castaneum* homologs showed a mixed localization pattern in CHRX and CHR7 (six and five occurrences, respectively), but the two homologs flanking *Transmembrane protein 165* gene are located at the X chromosome (Fig. 3C).

Whole contig synteny of contig GL563069 was done with all *T. castaneum* chromosomes separately. Synteny results from the whole contig have confirmed the localization previously identified in reciprocal best-hit approach for the following genes: *Ephrin*, *Src tyrosin kinase*, *Small glutamine-rich tetratricopeptide repeat-containing protein alpha*, *X11Lbeta*, *Transmembrane protein 165*, *Interleukin 16*, *WNT4 protein*, *UNC-112 related protein*, *NADH dehydrogenase*, *KN motif and ankyrin repeat domain-containing protein* and *Band 4.1-like protein 4* (Fig. A.1). Although the *Ribonuclease* gene did not

show similarity regions neither in CHR7 or CHRX (using an  $e$ -value  $< 10^{-10}$ ), its homolog was found in reciprocal best-hit approach (Fig. A.1B). Genes *Odorant receptor* and *Sarcoplasmic calcium-binding protein* were confirmed to be located in CHR6 (data not shown). The overall synteny of this contig in *T. castaneum* CHRX, CHR7 and CHR4 showed 161, 158 and 52 hits, respectively (Fig. A.2) and rates of 14.6, 7.7, and 3.7 hits/Mbase, respectively. The synteny of all other chromosomes presented between 36 and 6 hits and 3.1–0.3 hits/Mbase.

## 4. Discussion

The *R. prolixus* diagnostic SNP was identified during the initial phase of the *Rhodnius prolixus* genome project, known as *genome survey sequencing* (GSS). GSS sequences will be used, amongst other things, as a framework for the mapping and sequencing of genome size pieces, and this step is required for the assessment of genetic variability values, as high heterozygosity levels could lead to difficulties during genome assembly. This task is accomplished through the identification of variable non-coding single-copy nuclear loci that are then used as markers to estimate colony heterozygosity. In the particular case of the *Rhodnius prolixus* genome project, it was also used with the purpose of identifying a colony that was indeed *R. prolixus*, and not *R. robustus*, or a mixture of the two species. The *amplicon G* (*AmpG*) was the only locus to display a derived character unique to (and thus distinctive of) the *R. prolixus* lineage: an adenine, instead of a guanine, on the 280th position. Due to its usefulness in the colony-screening phase of the project, we decided to evaluate its applicability over a larger sample and geographic area. Our results are very encouraging as they show proven reliability after being tested on an extensive collection of specimens spanning a significant geographical area. Our sampling is evidently not ideal as there are still large areas from where we were unable to obtain specimens. Needless to say, some of these regions are very remote and difficult to access. With respect to the sampling of *R. robustus* I, it is worth mentioning that this species is rare and of restricted geographic distribution. To date, it has only been found in the state of Trujillo, Venezuela.

The ribosomal second internal transcribed spacer (ITS-2) is frequently used for solving taxonomical problems in triatomines. This region seems to be in concerted evolution in some organisms, a molecular process of homogenization between different loci in gene families, as a consequence of gene conversion and unequal recombination (Liao, 1999). This homogenization can be complete, as seems to be the case of *Triatoma* species (Bargues et al., 2002, 2006), or reveal intragenomic and intraspecific variation, as showed in other insects as *Culex* and *Lutzia* species (Diptera: Culicidae) (Vesgueiro et al., 2011). As *Rhodnius* species seem to fit the latter case (CL, MGP, and FAM, unpublished), ITS-2 copies must be cloned and sequenced, which makes its use more expensive and time-consuming.

Although this methodology entails the sequencing of PCR products, as opposed to multiplex-PCR methods (Pavan and Monteiro, 2007), no cloning is required (such as with most rDNA targets). It is also simple and straightforward, because it only involves the identification of a single nucleotide site. We hope that this SNP will be of use for a number of professionals with a broad variety of interests regarding this particular group of insects, however, evaluation of the applicability of this marker over a larger sampling of *R. robustus* I and *R. robustus* IV is desirable. The validation of its potential awaits further usage by the research community that will put it to the test.

In order to determine its genome localization, the *AmpG* region was searched for in all available partial or complete arthropod genomes (40 in total), but it was only located in the genome of *R.*

*prolixus* itself, in contig GL563069, inside the fourth intron of *Transmembrane protein 165* gene (TP165). Fourteen genes flanking the *AmpG* region were annotated and their homologs searched at the protein level, using reciprocal best-hit strategy (Tatusov et al., 1997) in *A. mellifera*, *B. terrestris* and *T. castaneum*. These were the three arthropods that have shown the closest chromosome number and phylogenetic relationship to *R. prolixus*. Despite the detection of the fourth TP165 intron as the location of the *AmpG* region, this protein was the only one of the *R. prolixus* contig whose homologs were located on different chromosomes in all three species searched, and thus could not be used to infer *AmpG* localization. Considering all 14 gene homologs' localization in all three species analyzed, *T. castaneum* showed the most homogenous pattern concerning genes location, with six and five occurrences in X and 7 chromosomes (CHRX and CHR7), respectively; notwithstanding, both *AmpG* flanking genes homologs were located in CHRX.

The whole contig synteny approach represents a more reliable method to look at genomic similarities, as DNA sequences are compared along the complete extension of each contig. *R. prolixus* contig GL 563069 synteny in comparison with *T. castaneum* contig also pointed CHRX and CHR7 with almost equal number of hits (161 and 158) as the possible chromosome location, even though their lengths are very different. The frequency of hits in CHRX was two times higher than in CHR7 (14,6 to 7,7 hits/Mbase, respectively) and about ten times higher than all other chromosomes, showing that contig GL563069 shares more potential gene content with *T. castaneum* CHRX. Thus, our results suggest that this contig could be part of *R. prolixus* X chromosome.

DNA sequencing analysis of the *AmpG* region of *R. prolixus* and *R. robustus* *s.l.* samples shows that *R. prolixus* and *R. robustus* I share eight polymorphic sites, three of these are derived nucleotides (A, C, and A, at positions 174, 270, and 360, respectively; Table 2), and hence Maximum-Composite Likelihood divergence between these species reached 0.6% in contrast to 1.1–1.7% between *R. robustus* I and *R. robustus* II–IV. A Bayesian maximum credibility tree and a Median-Joining network with Maximum-Parsimony post-processing further support the paraphyletic nature of *R. robustus* (Fig. 2), previously observed with *cyt b* and rDNA sequences (Monteiro et al., 2003).

In addition to its utility in taxonomy, this marker could be used in combination with the *cyt b* multiplex-PCR (Pavan and Monteiro, 2007) for the detection of natural crossbreeding and introgression of mitochondrial DNA along putative hybrid zones, as reported by Fitzpatrick et al. (2008) in Venezuela. The presence of an adenine/guanine double-peak in electropherograms could be indicative of a recent hybridization event. In addition, as *R. prolixus* has long been used as a model for insect physiology and biochemistry (e.g. Ribeiro et al. (1993) and Wigglesworth (1957)), several laboratories worldwide currently maintain colonies of this species. There is evidence, though, that some of these colonies have been contaminated with, or misidentified as, *R. robustus s.l.* (MGP, CL, EMD, FAM, unpublished). The SNP here described could also be used to evaluate the identity and purity of *R. prolixus* colonies. However, although the possibility of using this SNP as a diagnostic marker for this important vector species seems very promising, its "universal" applicability remains to be evaluated by the research community.

#### Disclosure statement

The authors declare no conflict of interest.

#### Role of the funding source

This work was supported by the Brazilian Research Council (CNPq), the Brazilian Ministry of Science, Technology and

Innovation (MCTI), and NIH Grant No. U54 HG003079. The sponsors did not take part nor influenced the any aspect of this study.

#### Acknowledgements

To John Spieth at the Genome Institute at Washington University School of Medicine in St. Louis for the assistance with library construction and identification of nuclear loci. To Sebastião Aldo Valente, Luis Herman S. Gil, Toby Barrett, José R. Coura, Christine Aznar, and Celia Cordon-Rosales, for kindly providing specimens. We thank José Marcos Ribeiro for providing the Transposon database. The authors are also grateful to the comments of two anonymous referees that helped improve the original manuscript. We also thank the PDTIS-FIOCRUZ DNA sequencing core for running sequencing reactions.

#### Appendix A. Supplementary data

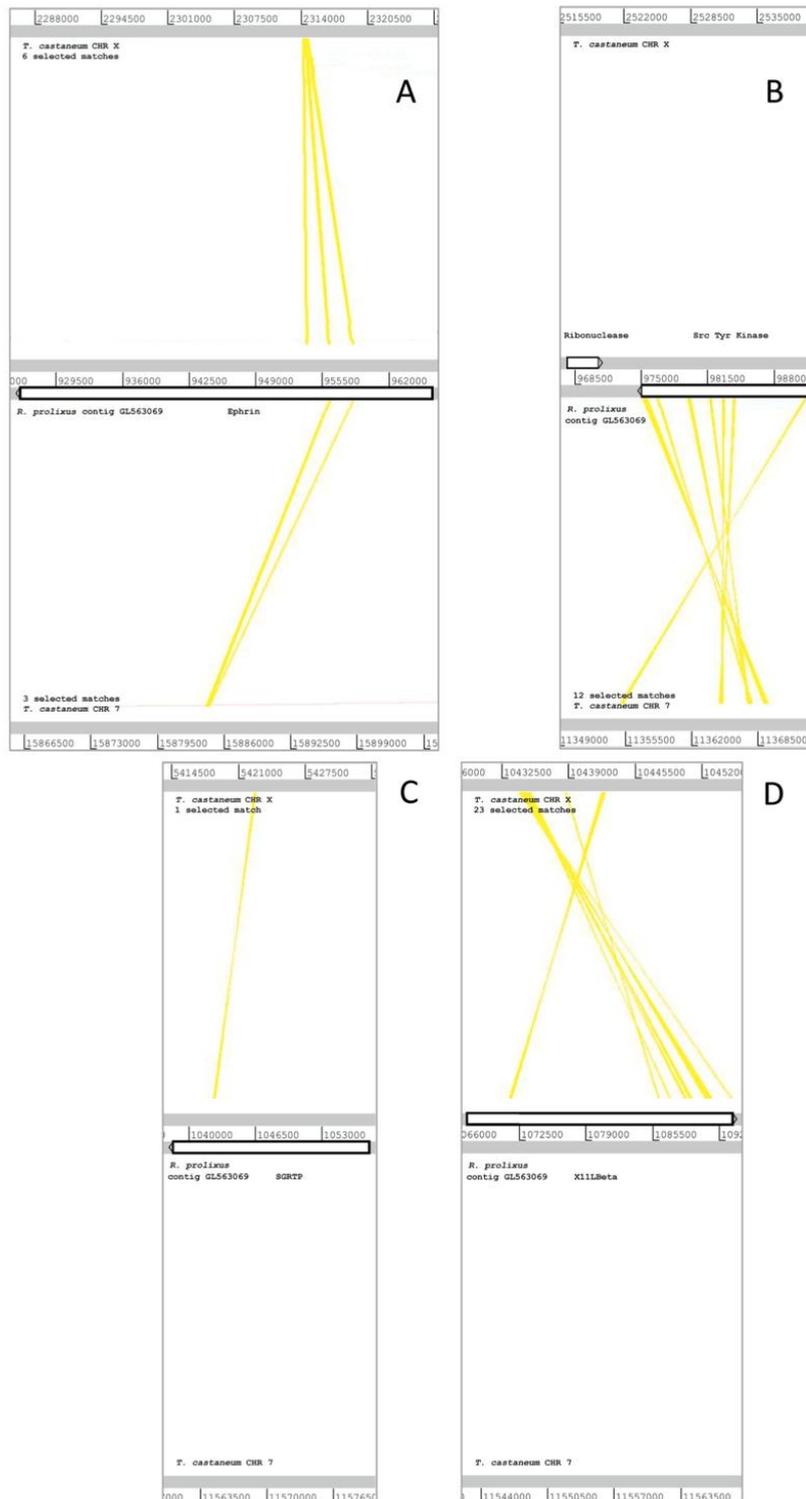
Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.meegid.2012.10.018>.

#### References

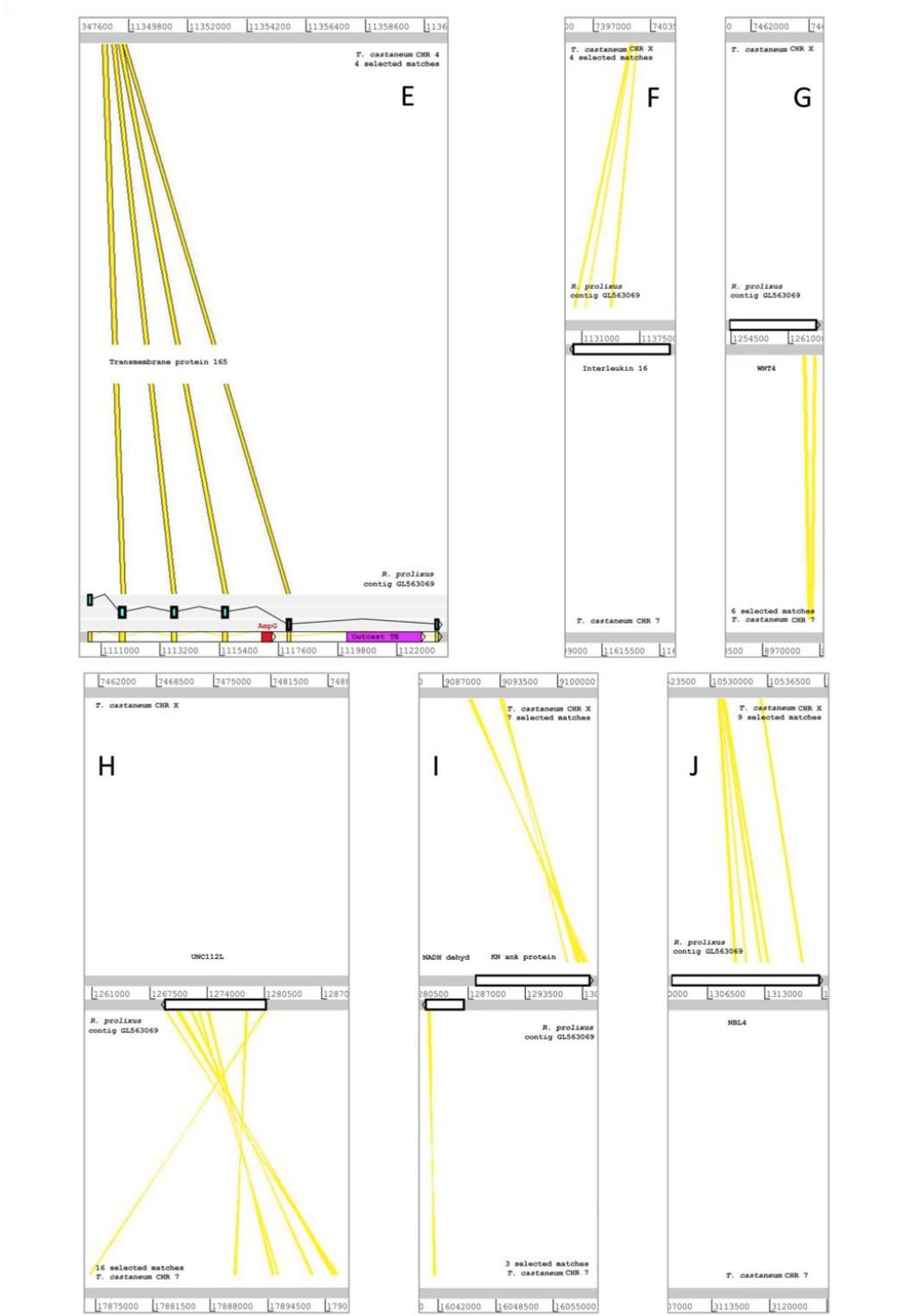
- Abad-Franch, F., Monteiro, F., 2005. Molecular research and the control of Chagas disease vectors. *An. Acad. Bras. Cienc.* 77, 437–454.
- Abad-Franch, F., Ferraz, G., Campos, C., Palomeque, F.S., Grijalva, M.J., Aguilar, H.M., Miles, M.A., 2010. Modeling disease vector occurrence when detection is imperfect: infestation of Amazonian palm trees by triatomine bugs at three spatial scales. *PLoS Negl. Trop. Dis.* 4, e620.
- Abubucker, S., Martin, J., Yin, Y., Fulton, L., Yang, S.P., Hallsworth-Pepin, K., Johnston, J.S., Hawdon, J., McCarter, J.P., Wilson, R.K., Mitreva, M., 2008. The canine hookworm genome: analysis and classification of *Ancylostoma caninum* survey sequences. *Mol. Biochem. Parasitol.* 157, 187–192.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402.
- Bandelt, H.J., Forster, P., Rohlf, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16, 37–48.
- Bao, Z., Eddy, S.R., 2002. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* 12, 1269–1276.
- Bargues, M.D., Marcilla, A., Dujardin, J.P., Mas-Coma, S., 2002. Triatomine vectors of *Trypanosoma cruzi*: a molecular perspective based on nuclear ribosomal DNA markers. *Trans. R. Soc. Trop. Med. Hyg.* 96 (Suppl. 1), S159–S164.
- Bargues, M.D., Kłisiwicz, D.R., Panzera, F., Noireau, F., Marcilla, A., Perez, R., Rojas, M.G., O'Connor, J.E., Gonzalez-Candelas, F., Galvao, C., Jurberg, J., Carcavallo, R.U., Dujardin, J.P., Mas-Coma, S., 2006. Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infect. Genet. Evol.* 6, 46–52.
- Birney, E., Clamp, M., Durbin, R., 2004. GeneWise and Genomewise. *Genome Res.* 14, 988–995.
- Blanco, E., Paiva, G., Guigó, R., 2007. Using geneid to identify genes. *Curr. Protoc. Bioinform.* 4, Unit 4.3.
- Coura, J.R., 2007. Chagas disease: what is known and what is needed—a background article. *Mem. Inst. Oswaldo Cruz* 102 (Suppl. 1), 113–122.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973.
- Fitzpatrick, S., Feliciangeli, M.D., Sanchez-Martin, M.J., Monteiro, F.A., Miles, M.A., 2008. Molecular genetics reveal that sylvatic *Rhodnius prolixus* do colonize rural houses. *PLoS Negl. Trop. Dis.* 2, e210.
- Gardner, P.P., Daub, J., Tate, J., Moore, B.L., Osuch, I.H., Griffiths-Jones, S., Finn, R.D., Nawrocki, E.P., Kolbe, D.L., Eddy, S.R., Bateman, A., 2011. Rfam: Wikipedia, clans and the "decimal" release. *Nucleic Acids Res.* 39, D141–145.
- Gubli, F., Pinto, N., Aguilera, G., 2009. Sylvatic triatominae: a new challenge in vector control transmission. *Mem. Inst. Oswaldo Cruz* 104 (Suppl. 1), 71–75.
- Hotez, P.J., Bottazzi, M.E., Franco-Paredes, C., Ault, S.K., Perago, M.R., 2008. The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Negl. Trop. Dis.* 2, e300.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules. In: Munro, H.N. (Ed.), *Mammalian Protein Metabolism*. Academic Press, New York, pp. 21–123.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948.

- Liao, D., 1999. Concerted evolution: molecular mechanism and biological implications. *Am. J. Hum. Genet.* 64, 24–30.
- Lopez, D.C., Jaramillo, C., Gubí, F., 2007. Population structure and genetic variability of *Rhodnius prolixus* (Hemiptera: reduviidae) from different geographic areas of Colombia. *Biomedica* 27 (Suppl. 1), 28–39.
- Marcilla, A., Bargas, M.D., Ramsey, J.M., Magallon-Gastelum, E., Salazar-Schettino, P.M., Abad-Franch, F., Dujardin, J.P., Schofield, C.J., Mas-Coma, S., 2001. The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and phylogenetic relationships in Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. *Mol. Phylogenet. Evol.* 18, 136–142.
- Monteiro, F.A., Barrett, T.V., Fitzpatrick, S., Cordon-Rosales, C., Feliciangeli, D., Beard, C.B., 2003. Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. *Mol. Ecol.* 12, 997–1006.
- Panzer, F., Ferrandis, I., Ramsey, J., Salazar-Schettino, P.M., Cabrera, M., Monroy, C., Bargas, M.D., Mas-Coma, S., O'Connor, J.E., Angulo, V.M., Jaramillo, N., Perez, R., 2007. Genome size determination in chagas disease transmitting bugs (hemiptera-triatominae) by flow cytometry. *Am. J. Trop. Med. Hyg.* 76, 516–521.
- Pavan, M.G., Monteiro, F.A., 2007. A multiplex PCR assay that separates *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). *Trop. Med. Int. Health* 12, 751–758.
- Posada, D., 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256.
- Ribeiro, J.M., Hazzard, J.M., Nussenzweig, R.H., Champagne, D.E., Walker, F.A., 1993. Reversible binding of nitric oxide by a salivary heme protein from a bloodsucking insect. *Science* 260, 539–541.
- Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. *Bioinformatics* 16, 944–945.
- Senior, K., 2007. Chagas disease: moving towards global elimination. *Lancet Infect. Dis.* 7, 572.
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. USA* 101, 11030–11035.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739.
- Tatusov, R.L., Koonin, E.V., Lipman, D.J., 1997. A genomic perspective on protein families. *Science* 278, 631–637.
- Tatusova, T., 2010. Genomic databases and resources at the National Center for Biotechnology Information. *Methods Mol. Biol.* 609, 17–44.
- Vesgueiro, F.T., Demari-Silva, B., Malafrente Rdos, S., Sallum, M.A., Marrelli, M.T., 2011. Intragenomic variation in the second internal transcribed spacer of the ribosomal DNA of species of the genera *Culex* and *Lotzou* (Diptera: Culicidae). *Mem. Inst. Oswaldo Cruz* 106, 1–8.
- Wigglesworth, V., 1957. The physiology of insect cuticle. *Annu. Rev. Entomol.* 2, 37–54.

Supplementary Fig. A.1. AmpG neighbor genes synteny. AmpG neighbor genes Ephrin (A), Src tyrosin kinase (B), Small glutamine-rich tetratricopeptide repeat-containing protein alpha – SGRTP (C), X11Lbeta (D), Transmembrane 165 protein (E), Interleukin 16 (F), WNT4 protein (G), UNC-112 related-protein (H), NADH dehydrogenase and KN motif and ankyrin repeat domain-containing protein (I), and Band 4.1-like protein 4 – NBL4 (J) were compared with *T. castaneum* chromosomes X and 7 and 4 (CHRX, CHR7 and CHR4) as indicated in panels. Synteny used ACT to select matches overlapping genes area with at least 50% identity (70% to B and D), based on TBLASTX hits from the hard-masked contig GL563069 versus the whole *T. castaneum* chromosomes, with  $10^{-10}$ -value cutoff. Open rectangles in *R. prolixus* contig represent its genes (their names are outside for better visualization). Red rectangle represents AmpG region. Purple rectangle represents Outcast transposable element. Blue rectangles represent coding sequences, and specifically in the transmembrane 165 protein gene, exons are represented in yellow. Yellow lines represent one or more selected matches.



Supplementary Fig. A.2. *prolixus* contig GL563069 whole synteny. Contig GL563069 (in black) was compared with *T. castaneum* chromosomes X and 7 (A) and 4 (B). Synteny used ACT to select matches overlapping open reading frames (ORFs) from contig GL563069 bigger than 100bp, with at least 50% identity based on TBLASTX hits from the hard-masked contig GL563069 versus the whole *T. castaneum* chromosomes with  $10^{-10}$  e-value cutoff. Yellow lines can represent one or more selected matches. Red and blue lines represent unselected matches above cutoffs (originated from ORFs smaller than 100 bp).



## Supplementary Material

**Table A.1.** The 96 primer-pair sequences used for an exploratory screening in the *R. prolixus* genome. All primer-pairs were designed for a PCR-amplification of regions of approximately 500 base pairs (bp) in length. We sequenced 27 *R. prolixus*, two *R. robustus* I, four of each *R. robustus* II and III, and six *R. robustus* IV for each PCR-amplified region (results not shown). Primer sequences for the diagnostic amplicon are shown in bold.

| Primer number | Forward primer Sequences (5' - 3') | Reverse primer sequences (5' - 3') |
|---------------|------------------------------------|------------------------------------|
| 1             | TTCAATGCAACCTGAAATGTGG             | TCCCTTCCGTAACAGAGGAACC             |
| 2             | TCTGAGAGTCCACAGGAGAGCC             | ATGCATCTAGCGATGAAGTGGC             |
| 3             | GCAACAGGTTCTTGTTCCTGC              | AAGTGAAATTAGCATAGCCGTTTCG          |
| 4             | TTCAGTGCTTGAGCATGATTCC             | TCTGAACCTCTATTACCAATTCCAGC         |
| 5             | AAATCGTTTACGCGTTTAGGC              | AAATTAATCATCATGGCGTATCTAAAGC       |
| 6             | CAAAGAAGCTCCTGATGAAGGG             | CGTTGTCAACATCTATCGCAGC             |
| 7             | GGAAGGGCTCATCAGAGAATGC             | GATAGATGGCGTCGTTGAAAGC             |
| 8             | GCACAACCTTCTTTCTGCAATG             | AAGACGAAAGCTACACGCTGATG            |
| 9             | AAAGATGCTAATGTTGGCGAGC             | AAACGGCTTCATGGTAATTTGG             |
| 10            | TCTAGCCTATTAACCGCTCGGC             | GAGGGTTCCTCAGTTGGTAGC              |
| 11            | ATACCGCCAACACACCGACTAC             | TGCCAGTAAGTTTAATTGCTTCAAATG        |
| 12            | TTGGGTAACAGAAGCAAATTGATG           | AGCTTAACCGATTTCCGGTCTCC            |
| 13            | GACAGCCATCTCTGCAACTGG              | ATATGCATTAAGCACGCTGTGG             |
| 14            | ACGCAATCTCTTGATCTCCAGC             | TCAGTCTCTGGACAGTGTTCGC             |
| 15            | GAAGATGGAGGATATGGCGAGG             | CTAGTCTGGCTATCGTCCCACG             |
| 16            | GAGCGGTTAGAAGCTAGACCCG             | GTCAATCGTGTATTACGCGG               |
| 17            | AATCATGCTTGAGCAGGAAACC             | GCAGGCAATACTTCACCTCACC             |
| 18            | GTGGGAGATATAGAACGGCCATAA           | AACTCCCCTGTGAACCACGTATT            |
| 19            | ACCATTTGTTTGTTCACGCGGG             | CGCATGCGCACAAGAGATAGG              |
| 20            | TGTGAACCACGTATTGGAATGG             | CACAGTGGGAGATATAGAACGGC            |
| 21            | TTTACTGCACATCACGTGACCC             | AAAGCCGGCCGATATTCC                 |
| 22            | CACTAGCTACCATCCACCAACTCC           | TGCCTAACAGTTAATGAGGCCG             |
| 23            | CCATCTGTACAGCAAGAGGTCG             | CCATAACGATTTGGAAGGTTGC             |
| 24            | TGGATGTTACAGAGCTGAACGC             | TAAGACTTTGGATCGCTTTGGC             |
| 25            | ACTGGGTTGTTGAAGTTCTGGC             | TTTGCCTTGCCTCGCC                   |
| 26            | GAAGCATAATGCACCATCTCCC             | TAAATGACCACGGATCTGGACC             |
| 27            | ATTCCGGCGATTGTGTCTATTCC            | ACAGATGGACAGACGGAGTTGG             |
| 28            | CCTTGTTACACTGGCACAGAC              | CTCGTTGCACAAGCTTTGGTC              |

29 **GAGAGCTGAAGATAGGCAAGCG**  
30 AGCCAGAAGCCTTCAAAGTGG  
31 CTTCTGCAAGGTCGTTATCTTGG  
32 CCAGCTTCCTGATCTTGTGAGC  
33 ATGGTGACCACTCTGAGTTCCG  
34 GGTACCTCCTTTCTTTTCGCTGG  
35 TCCTTACTAGGCGAAGCACTCC  
36 TTTGAGAAATGGCAGTAAATCTTGC  
37 CACAGAAACGTTTGGCAAAGAG  
38 GCCAATAAAACAACCAGACCAGC  
39 TGCTGCTTAATTGGACGAAACA  
40 CCCTTATACAACGTCTACCGGA  
41 TGCAGCAAAGTCAACAAGATGG  
42 GGGTAATGGGATGGATAAGGGA  
43 CGTCTCTATGCCCAGGTCAACT  
44 GGGCGAATTTACACAGAGAGTGG  
45 ATCAGCTTGATTCGTCGGACC  
46 CGTGATTGTTTGATGAAGCTTGG  
47 TCTGTTGTGTTGCACGGATACC  
48 CTGTTACTGGATTCCCAGACGG  
49 TCTTGGCTTTCTAAACCGGACA  
50 TCCGGGAGGACAGTTGTATAAGG  
51 TCGGGTACCAATTTCAATTCATC  
52 AACAAAGGAAAGAAGCTTCCGC  
53 TAATCATCGAATTGGACGAGGG  
54 TACTACACATTCTTGGCGAGCG  
55 CAGCTTGATTCGTTGCATTCTC  
56 CAAAGAACGCTCTTGGTTATTTCG  
57 AGAGATTTCGTCTTTCTAACGGC  
58 ACCTACTGAAAGAATGGAAGCTAAATTG  
59 CGCAGTTAATTCACTTGATGCAAA  
60 TGGTCTGAGAAGTGTTGTTGGC  
61 TCCAAATAAATTTAGGACTGGAAATCTG  
62 ACCATGCACCTGTCTATCTGCC  
63 CACCACATTGTTTACCCGAAGC  
64 TGGACTGGAAGTTCCGAAGAAG  
65 GGGATCCAGAAATCACTGTCAGG

**TGATAACTGGATTAGGCGCAGC**  
CAGAGCATTCTCCAATCGTTCC  
AGGAGCCAAGAATTTGGAAACC  
GGATTGCAATTAGAAACTGGCTTG  
TTAAATAAAGCGAAGTGCCTCTCC  
GTTGGTTCATGTCGGGAGTAGG  
CCTAAAAGAATTGTTGTGCCGAGG  
CTCAGATATTGGTCGGGCTTTG  
GCCTATATTTGTGGTTTGCATGAAG  
TCGTAGAAAATCAGTCGTGGTTCG  
GTGCGCATTATGAGCTTCAAGA  
ATATGTTTCGCTAGGTCGCCTCC  
GACGTTAGGTTGAAGGGAAGGC  
AAACGTTATCAATTGTGCTGGCT  
AAACTCCATTGCAAAGTTTTATTCC  
TGGATGAAGAGAGAGATTCCGC  
AGTGATCATCAGTTTCAGTTCAATCAG  
TTCCTGAGTCACTGCATCATCC  
GAATGTTGGGCGTAAGATCCC  
GCGTATGTTGCAGAAATCTCCC  
CCCTTATACAACGTCTACCGGA  
CGGCACATAAAGGGACTTAAAGC  
TTCGGGAAACATAATGCTGTAGG  
TACACATATCCCAGGCGATGC  
CGCCATCTTGTGTACTACAGCG  
TTTAGTGACGCCCTTTGTGAGG  
GGTTTAAATGCGTCCAGAACGTC  
TCGACTAGTTTAAAGCTGTCTTGTTC  
TATTGCCGGTGTATGTCTCTGG  
ATTACGCTAGCGAGGTCACACC  
TAGCCTCTCGATCACAGCATCA  
AAGTGGTGGTTCAGTAATGGGC  
CTAAAGTCGGTCGACAAGAATCATC  
ACGATGGAGAGTAGATCGTCGG  
AATGGAAATATCCGTCTGCCG  
CAAGAAGAAGGATGACACGAACAC  
GTTCAGTACAGGTCAGGGCAGG

|    |                              |                              |
|----|------------------------------|------------------------------|
| 66 | CAATGTTTCCTGCTGAAGTGAGG      | AGAGTGAATGCGTTGACTTGGG       |
| 67 | TCGACCTAGGGTCATCTTCAGG       | GCAGCAATTGGTATCCTGATCG       |
| 68 | CAAGATTCCAAGTCCTGTATTGGG     | AATGGCACTCTGTTCCCTTCAGC      |
| 69 | GAATAACTGAGCACGACCTGCC       | GATTCTTTGGGCATGTTTGAGG       |
| 70 | ATCAGCTTGAATTCGTTACCC        | TCCCACCCAACATACATGTGC        |
| 71 | TTGTGCACAGATCTACAACAGCC      | AGTGGGAGATATAGAACGGCCATA     |
| 72 | TTCGGAACAATGGTAAATTTCGC      | AATATGCCGGTTATGCTCATGG       |
| 73 | CCTTGCACGAAGCACATTACAG       | TTGTGGTCCAATAAATACCGTCAG     |
| 74 | TTCAAATAAATCCGTTAGGCAAATTC   | GACATCCCAAAGCAAAGTGTCC       |
| 75 | CAATCGCAATGAATTCGAGACC       | CCGTATGTGTGTGTGCCTTCAG       |
| 76 | TCTAACAAATACCAAACGGGCG       | GAATTCGATCATCGGAACTTGG       |
| 77 | TCGTTCTCGATCAACCTTCTCC       | CCAATACAGGTCGAGATGTGGG       |
| 78 | GCAGCGGGTGAATAGGAGAATC       | TGATGCCACAATTAATCTAGTACAGAGC |
| 79 | TGTTTCAGACTGTTGCTAAATCTTC    | GGCCTTAGGCGAGGAATACTTG       |
| 80 | TTATAGTGGACTGTCGGCCTCC       | TCACGCAAAGAAGTTGAAGACG       |
| 81 | CTAGTTTACCCGATTGTCTTGCG      | ATTGGTACCTCATGCAAGGACG       |
| 82 | AAGGTAAAAGTCAACCCGAGGC       | CCTAACATCCAAAAGGTAGCTTGC     |
| 83 | TGCAACTCTTCAGTCATCTTCAGG     | CAAAGCTGAATGGAACCTTGGG       |
| 84 | AGCTAGAATGTCGGCCGTAGTG       | TGAGCTGATGAATTATTAGCGGAAG    |
| 85 | GGCATGTGAGCAGTTACAATCG       | TCAAAGTCGAGAGGAGCAGTGG       |
| 86 | CGGAGCGTTCAGTGGAGTTAG        | GATTTGACTTCACTGCTCTCCC       |
| 87 | CACGGCTTAATTGCAACATAATTT     | ACGTCGAACACGCTTAAATGGT       |
| 88 | ACATAATGGGCCCAACTGAAGC       | TGGTATGGTCAGTTAGGTCGGC       |
| 89 | CCTCCAACAATCAATAGCAAAGC      | ACAGACGTAACCTCCACCAGGC       |
| 90 | CAGGCTTCGGGTGTAGTTCTCC       | CGTCAGGATGTCCAAGAAGTCC       |
| 91 | TTGCATTGGAAGTAATTGACGG       | CATGAGATGGTCGACTGTCAGC       |
| 92 | TTAAACTTCTTGTTGTTGCGCCG      | AGTAGGCACTGAAACACAATGGC      |
| 93 | ACGGAACATGTCTGAACCGTCT       | AACTTGCACAGTGTAGCAGCCA       |
| 94 | GACTTCTGGGTTGATGACAGC        | GCTGGGATTTCTCTCAGTGTGG       |
| 95 | TTTTATGATTTACATGATCTGATCCAGC | CAAGAACAGAAAATGCTAGGCCC      |
| 96 | AGTGATTTGGTCCGTATTGTTGG      | TCAACAGCTGAGGTCATTAGCG       |

---

**Table A.2.** Polymorphic sites in all seven amplicons sequenced. Note that *AmpG* contains a polymorphic site (**SNP<sub>G3</sub>**) that seems to be diagnostic for *R. prolixus*. Codes: bp – base pairs.

| Species                | <i>AmpA</i>      |                    | <i>AmpB</i>      |                    | <i>AmpC</i>       |                   | <i>AmpD</i>        |                    | <i>AmpE</i>        | <i>AmpF</i>      | <i>AmpG</i>       |                   |                   |
|------------------------|------------------|--------------------|------------------|--------------------|-------------------|-------------------|--------------------|--------------------|--------------------|------------------|-------------------|-------------------|-------------------|
|                        | SNP <sub>A</sub> | Indel <sub>B</sub> | SNP <sub>B</sub> | Indel <sub>C</sub> | SNP <sub>C1</sub> | SNP <sub>C2</sub> | Indel <sub>D</sub> | SNP <sub>D</sub>   | Indel <sub>E</sub> | SNP <sub>F</sub> | SNP <sub>G1</sub> | SNP <sub>G2</sub> | SNP <sub>G3</sub> |
| <i>R. prolixus</i>     | T                | no                 | C                | no                 | T                 | G                 | no                 | A                  | no                 | T                | G                 | C                 | A                 |
| <i>R. robustus</i> I   | C                | 1-bp               | T                | yes*               | A/T               | A/G               | no                 | G                  | 16-bp              | A/T              | G/T               | C/G               | G                 |
| <i>R. robustus</i> II  | T or C           | no                 | C                | yes*<br>or no      | A or T            | G                 | no                 | A                  | 16-bp<br>or no     | T                | G                 | C                 | G                 |
| <i>R. robustus</i> III | T                | no                 | C                | no                 | A or T            | G                 | no                 | A                  | 16-bp<br>or no     | T                | G                 | C                 | G                 |
| <i>R. robustus</i> IV  | T or C           | no                 | C                | no                 | A or T            | G                 | 10bp<br>or no      | A or a<br>deletion | 16-bp<br>or no     | T                | G                 | C                 | G                 |

\* Variable indel, from 10 to 15bp.

**Table A.3. *AmpG* region neighborhood genes prediction details and homolog results.** This table shows in the first ten columns detailed information about prediction, annotation and curation of genes and transposable elements in the *AmpG* region neighborhood. The following columns show Gis and chromosome location to reciprocal best-hit strategy homologs found in *A. mellifera*, *T. castaneum* and *B. terrestris*. “Code” column numbers is according to Figure 3. Black and green lines show TP165 gene and its exons details, as this protein was completely curated. Orange line shows *AmpG* region details. Vector base prediction gene codes have "RPTMP prefix. Lagerblad prediction has a "contig" prefix. Other codes: \* = not homolog, only similar (no reciprocal best-hit); *na* = not applicable; *ns* = not searched; *nf* = not found; *unk* = unknown; *pred error* = prediction error; *transposon* and *TE* = transposable elements.

| Start position | End position | Frame | Abbreviation     | Annotation   | Homologous proteins (GI) |                        |                         | Homologous proteins Chromosome |                        |                         |
|----------------|--------------|-------|------------------|--|--------------------------|------------------------|-------------------------|--------------------------------|------------------------|-------------------------|
|                |              |       |                  |  | A.<br><i>mellifera</i>   | T.<br><i>castaneum</i> | B.<br><i>terrestris</i> | A.<br><i>mellifera</i>         | T.<br><i>castaneum</i> | B.<br><i>terrestris</i> |
| 925995         | 966186       | -     | Ephrin           | Ephrin   | 110766111                | 270015107              | 340710982               | 2                              | X                      | 2                       |
| 967657         | 970654       | +     | Ribonuclease     | Ribonuclease   | 110766219                | 91087325               | nf                      | 1                              | 7                      | na                      |
| 975073         | 991396       | -     | Src_Tyr_K        | Src Tyrosin Kinase   | 328788805                | 91086687               | 340716501               | 1                              | 7                      | 7                       |
| 1014493        | 1023165      | +     | Odorant_receptor | Odorant Receptor   | 201023349                | 226334904              | 340728952               | 1                              | 6                      | unk                     |
| 1026686        | 1030171      | -     | SCBP             | Sarcoplasmic calcium-binding protein                                   | 66513995                 | 91085141               | 340729116               | 3                              | 6                      | unk                     |
| 1038468        | 1057541      | -     | SGRTP            | Small glutamine-rich tetratricopeptide repeat-containing protein alpha | 328791828                | 91076010               | 340708935               | 1                              | X                      | 1                       |
| 1058630        | 1066570      | -     | hAT_TE           | hAT transposon (class II)  | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1067427        | 1093204      | +     | X11LBeta         | X11Lbeta 5'truncated   | nf                       | 189234056              | nf                      | na                             | X                      | na                      |
| 1110523        | 1123517      | +     | TP165            | Transmembrane protein 165  | 328790251                | 91081711               | 340711903               | 5                              | 4                      | 3                       |
| 1110523        | 1110645      | +     | na               | EXON Transmembrane protein 165   | na                       | na                     | na                      | na                             | na                     | na                      |
| 1111691        | 1111886      | +     | na               | EXON Transmembrane protein 165   | na                       | na                     | na                      | na                             | na                     | na                      |
| 1113618        | 1113793      | +     | na               | EXON Transmembrane protein 165   | na                       | na                     | na                      | na                             | na                     | na                      |
| 1115516        | 1115686      | +     | na               | EXON Transmembrane protein 165   | na                       | na                     | na                      | na                             | na                     | na                      |
| 1116977        | 1117340      | .     | <i>AmpG</i>      | <i>AmpG</i>  | na                       | na                     | na                      | na                             | na                     | na                      |
| 1117914        | 1118019      | +     | na               | EXON Transmembrane protein 165   | na                       | na                     | na                      | na                             | na                     | na                      |
| 1118579        | 1119562      | +     | na               | pred error   | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1120144        | 1122900      | +     | Outcast_TE       | Outcast transposon (Class I)   | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1120322        | 1121146      | +     | RT_TE            | Reverse Transcriptase fragment (outcast transposon)                    | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1123441        | 1123517      | +     | na               | EXON Transmembrane protein 165   | na                       | na                     | na                      | na                             | na                     | na                      |
| 1129966        | 1140839      | -     | Interleukin      | interleukin 16   | 328789728                | 270014802              | 340710900               | 2                              | X                      | 2                       |
| 1141206        | 1142245      | +     | Mariner_TE       | Mariner transposon fragment (Class II)                                 | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1162063        | 1164863      | -     | L1_TE            | L1 transposon (Class I)  | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1163028        | 1163570      | -     | RT_TE            | Reverse Transcriptase (L1 transposon - Class I)                        | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1230613        | 1230891      | -     | na               | pred error   | ns                       | ns                     | ns                      | na                             | na                     | na                      |
| 1254443        | 1264046      | +     | WNT4             | protein wnt-4 fragment   | 328790423<br>*           | 91086553*              | 340715611<br>*          | 1                              | 7                      | 7                       |
| 1269195        | 1280540      | -     | UNC112L          | unc-112-related protein  | 328785955                | 270010684              | 340714139               | 5                              | 7                      | 5                       |
| 1282266        | 1286411      | -     | NADH_dehyd       | NADH dehydrogenase   | 328793182                | 91087395               | 340727519               | unk                            | 7                      | 17                      |
| 1287896        | 1300699      | +     | KN_ank_ptn       | KN motif and ankyrin repeat domain-containing protein                  | 328786345                | 91076206               | 340710743               | 2                              | X                      | 2                       |
| 1302514        | 1319022      | +     | NBL4             | Band 4.1-like protein 4 (NBL4)   | 328778071                | 189234071              | 340709952               | 2                              | X                      | 2                       |

## 9. ANEXO 2

### ***Rhodnius barretti*, a new species of Triatominae (Heteroptera, Reduviidae) from western Amazonia**

Fernando Abad-Franch<sup>1,2/+</sup>, Márcio G Pavan<sup>3,4</sup>, Nicolás Jaramillo-O<sup>5</sup>, Francisco S Palomeque<sup>2/\*</sup>, Carolina Dale<sup>6</sup>, Duverney Chaverra<sup>5</sup>, Fernando A Monteiro<sup>3</sup>

<sup>1</sup> Instituto Leônidas e Maria Deane – Fiocruz Amazônia, Manaus, AM, Brazil

<sup>2</sup> Centro de Investigación en Enfermedades Infecciosas, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

<sup>3</sup> Laboratório de Epidemiologia e Sistemática Molecular, Instituto Oswaldo Cruz – Fiocruz, Rio de Janeiro, RJ, Brazil

<sup>4</sup> Laboratório de Biologia Molecular de Insetos, Instituto Oswaldo Cruz – Fiocruz, Rio de Janeiro, RJ, Brazil

<sup>5</sup> Grupo BCEI, Instituto de Biología, Universidad de Antioquia, Medellín, Colombia

<sup>6</sup> Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos, Instituto Oswaldo Cruz – Fiocruz, Rio de Janeiro, RJ, Brazil

+ Corresponding author: fernando@amazonia.fiocruz.br

\* Current address: Influenza Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

**Financial support:** UNDP/UNICEF/World Bank/WHO TDR Special Program (grants A20441 and 970195), CNPq (Brazil), Secretaria de Vigilância em Saúde (Ministry of Health, Brazil), Fiocruz-Fapeam agreement (Brazil). This work benefited from international collaboration through the ECLAT Network. MGP received scholarship from CNPq, and CD from CAPES.

## **Abstract**

*Rhodnius barretti*, a new triatomine species, is described based on adult specimens collected in rainforest environments within the Napo ecoregion of western Amazonia (Colombia and Ecuador). *R. barretti* resembles *R. robustus s.l.*, but mitochondrial cytochrome *b* gene sequences reveal that it is a strongly divergent member of the ‘robustus lineage’, i.e., basal to the clade encompassing *R. nasutus*, *R. neglectus*, *R. prolixus*, and five members of the *R. robustus* species complex. Morphometric analyses also reveal consistent divergence from *R. robustus s.l.*, including head shape and the length ratios of some anatomical structures. *R. barretti* occurs, often at high densities, in *Attalea butyracea* and *Oenocarpus bataua* palms. It is strikingly aggressive, and adults may invade houses flying from peridomestic palms. *R. barretti* must therefore be regarded as a potential *Trypanosoma cruzi* vector in the Napo ecoregion, where Chagas disease is endemic.

**Key words:** *Rhodnius barretti* sp. nov., Triatominae, Rhodniini, Amazonia

The Triatominae are blood-sucking Reduviidae (Hemiptera, Heteroptera) notorious as the vectors of *Trypanosoma cruzi*, causative agent of Chagas disease. While all of the over 140 currently recognized triatomine species can potentially transmit the parasite among mammals, just a few are epidemiologically relevant vectors of human infection (Lent & Wygodzinsky 1979, Barrett 1991, Schofield & Galvão 2009).

The tribe Rhodniini comprises two paraphyletic genera, *Rhodnius* and *Psammolestes* (Monteiro et al. 2000, 2002). Most *Rhodnius* species preferentially breed in palm trees across their native ranges (Barrett 1991, Abad-Franch et al. 2009). Flying from palms, adult specimens frequently invade, and sometimes colonize, houses and peridomestic structures (Fitzpatrick et al. 2008, Gurgel-Gonçalves et al. 2008, Justi et al. 2010).

Throughout Amazonia and several further Neotropical ecoregions, invasion of houses and other premises by adult *Rhodnius* spp. is probably the main mechanism underlying Chagas disease transmission, either as an endemic infection or in the form of outbreaks linked to food contamination (Grijalva et al. 2003, Aguilar et al. 2007, Calzada et al. 2010, Amunárriz et al. 2011, Valente et al. 2009).

While the morphological traits of *Rhodnius* allow for easy genus-level identification, many species lack clear-cut diagnostic characters and remain difficult to identify: a few species are near-sibling, and others are in fact species complexes encompassing several cryptic taxa (Lent & Wygodzinsky 1979, Monteiro et al. 2003, Abad-Franch et al. 2009). Currently, the genus comprises 18 named species occurring in Central and South America. Biogeographic and genetic data suggest a northern Paleo-Amazonian origin of the tribe Rhodniini, which comprises two main lineages: the ‘pictipes lineage’ and the ‘robustus lineage’ (Abad-Franch & Monteiro 2007, Abad-Franch et al. 2009). Cryptic speciation seems to have occurred along both lineages, including the clades comprising *R. pallescens*, *R. robustus-prolixus*, and *R. pictipes*

(Monteiro et al. 2003, Pavan & Monteiro 2007, Abad-Franch et al. 2009, Gómez-Palacio et al. 2012). On the other hand, phenotypic variation without significant genetic divergence has been recorded in several triatomine species (Schofield et al. 1999, Abad-Franch et al. 2003). When cryptic speciation and phenotypic plasticity both occur within a single genus, it comes as no surprise that taxonomic errors arise: while cryptic taxa are overlooked, some species are formally described based on minor morphological variation. Thus, *R. milesi* (Valente et al. 2001) is probably a *R. neglectus* variant from south-eastern Amazonia, *R. zeledoni* (Jurberg et al. 2009) closely resembles *R. domesticus*, and *R. montenegrensis* (Da Rosa et al. 2012) likely represents one of the *R. robustus* lineages of Monteiro et al. (2003). Furthermore, some *Rhodnius* species, such as *R. dalessandroi*, *R. paraensis*, or *R. amazonicus*, have seldom been collected; their ecological-evolutionary characteristics and affiliations remain therefore unclear (Barrett 1991, Abad-Franch et al. 2009).

Integrative taxonomic approaches are particularly useful when the organisms of interest are prone to cryptic speciation, phenotypic plasticity, or both (Dayrat 2005, Bickford et al. 2006). By combining morphological and molecular evidence, such approaches help solve systematic problems that would otherwise remain elusive (Monteiro et al. 2001, Abad-Franch & Monteiro 2005). In addition, they yield information on characters that can be measured under a variety of circumstances, including morphological traits allowing for rapid identification in the field; doubtful assignments can be resolved later by checking further traits in the laboratory (e.g., Monteiro et al. 2001, Gurgel-Gonçalves et al. 2008).

Here, we use an integrative approach, including qualitative and quantitative phenotypic assessments and DNA sequencing, to elucidate the specific status of a *Rhodnius* population from the Napo ecoregion of western Amazonia (Ecuador and

Colombia). These bugs were first identified as *R. robustus*; yet, detailed analyses showed them to belong to a new species (Abad-Franch 2003, Abad-Franch & Monteiro 2005, Abad-Franch et al. 2009, Márquez et al. 2011), which we name *Rhodnius barretti* and formally describe in the present paper.

## **Materials and methods**

*Material examined and specimen deposition.* *Rhodnius barretti* sp. nov. male holotype: municipality of Puerto Asís (0.50N, 76.50W), department of Putumayo, Colombia, #3374; female allotype: municipality of Puerto Asís, department of Putumayo, Colombia, #3375; paratypes: 4 males, municipality of Puerto Asís, department of Putumayo, Colombia, #3376-3379; 4 females, municipality of Puerto Asís, department of Putumayo, Colombia, #3380-3383; 3 males, municipality of Francisco de Orellana (0.45S, 77.00W), Orellana province, and municipality of Lago Agrio (0.08S, 76.83W), Sucumbíos province, Ecuador, #3384-3386. These specimens are deposited in the Triatominae Collection of the Instituto Oswaldo Cruz (CT-IOC), Fiocruz, Rio de Janeiro, Brazil, maintained by the Laboratório Nacional e Internacional de Referência em Taxonomia de Triatomíneos (LNIRTT). *Rhodnius barretti* sp. nov., five males, municipality of Puerto Asís, department of Putumayo, Colombia, #CRUA-298, 300, 303, 304, 487; ten females, municipality of Puerto Asís, department of Putumayo, Colombia, #CRUA-297, 299, 301, 302, 342, 447-450, 484. These specimens are deposited in the Reduviidae Collection maintained by the Grupo de Biología y Control de Enfermedades Infecciosas (BCEI), Universidad de Antioquia, Medellín, Colombia.

*Morphology and morphometrics.* A Zeiss Stemi SV11 stereomicroscope was used for the measurement and description of morphological characters. Linear measurements

and head capsule landmark coordinates (Fig. 1 and Table) were taken from 37 adult *R. barretti* specimens (34 from Colombia and 3 from Ecuador) and 12 *R. robustus s.l.* (from Colombia, Ecuador, Bolivia, Brazil, French Guiana, Peru, Suriname, and Venezuela). Landmark coordinates were recorded using *tpsDig* 1.18 (Rohlf 1999a). Shape variables were computed using the Generalized Procrustes superimposition algorithm (Rohlf 1996), which derives a least-squares consensus configuration from raw coordinate data. “Partial warps” were then computed as deformations of each structure in relation to the consensus configuration using *tpsRelw* 1.18 (Rohlf 1999b). Shape variables derived from partial warps were used as input for canonical variate (discriminant) analysis, and a plot showing the position of each specimen on the ‘shape discriminant space’ was built with the first two canonical vectors (Fig. 2). Multivariate analyses were conducted using JMP 3.2.6 (SAS Institute, Cary, NC, USA).

*Molecular analyses.* Samples of four putative *R. barretti* specimens (two from Ecuador and two from Colombia), consisting of one or two legs, were placed apiece in a 1.5mL microtube, dipped into liquid nitrogen, and ground to powder for DNA extraction using a Real Genomics™ kit. A 663-basepair (bp) fragment of the mitochondrial (mt) cytochrome *b* gene (*cyt b*) was PCR-amplified as described in Monteiro et al. (2003). Amplicons were purified (Hi Yield™ Gel/PCR DNA Extraction Kit, Real Genomics™) and both strands were subjected to Sanger sequencing reactions (ABI Prism® BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems) and run on an ABI 3730 sequencer. Consensus sequences were computed and manually aligned as in Monteiro et al. (2003). Homologous sequences from other members of the ‘robustus lineage’ (*R. robustus* I-V, *R. prolixus*, *R. neglectus*, and *R. nasutus*) were added to the alignment; one *R. pictipes* sequence was

used to root the phylograms. The Hasegawa, Kishino and Yano (HKY) model with a proportion of invariable sites was the best-fit model of nucleotide substitution (lowest Akaike Information Criterion value) for phylogenetic reconstruction; it was selected using jModeltest 0.1.1 (Posada 2008).

A Bayesian phylogenetic tree for the complete data set of 13 *Rhodnius* sequences was inferred in BEAST 1.6.1 (Drummond and Rambaut 2007) using a ‘two-partitions’ ((1+2), 3) scheme of codon positions. A prior tree was randomly generated and a Yule process of speciation was imposed for all tree reconstructions. Two independent runs were performed for  $10^7$  generations, with a burn-in of  $10^6$  generations. Convergence of parameters and proper mixing were confirmed by calculating the Effective Sample Size (ESS) in TRACER 1.5 (Rambaut and Drummond 2007), excluding the initial 10% (burn-in) of each run. All considered parameters had  $ESS > 10^4$ . Runs were combined using LogCombiner, and a maximum credibility tree based on the 10,000 trees (burn-in = 2,000) was generated with a posterior probability limit of 0.6 using Tree Annotator (both part of the BEAST package). Statistical support for clades was assessed by the posterior probability method. The maximum-composite likelihood (M-CL) algorithm available in MEGA 5 (Tamura et al. 2011) was used to construct a species distance matrix, and a heatmap was constructed in the R 2.13 program (Hornik 2011). Distance variance was estimated using 1,000 bootstrap pseudo-replicates.

## **Results**

*Rhodnius barretti* Abad-Franch and Monteiro, sp. nov. (Fig. 3).

*Systematics.* Family Reduviidae Latreille, 1807; subfamily Triatominae Jeannel, 1919; tribe Rhodniini Pinto, 1926; genus *Rhodnius* Stål, 1859; type species, *Rhodnius prolixus* Stål, 1859; *Rhodnius barretti* sp. nov.

*Description.* *Rhodnius barretti* sp. nov. Total length: male 19.10-22.66 mm, female 20.30-24.66 mm. Maximum width of pronotum: male 4.0-4.7 mm, female 4.1-5.0 mm; maximum width of abdomen: male 5.6-6.2 mm, female 5.6-7.2 mm. Overall color dark brown, slightly paler in some atypical specimens, with yellowish and/or brown-reddish markings and stripes on various parts of the body and appendages (Fig. 3). Head elongated, almost three times as long as wide across the eyes (1:0.36-0.41) and distinctly longer than the pronotum (1:0.72-0.80); granulose both laterally and ventrally; dark brown-blackish, with a dorsal, narrow, light-colored medial longitudinal stripe extending from clypeus to neck base. Anteocular region approximately four times as long as the postocular region (1:0.23-0.26); eyes large, in lateral view surpassing the lower but not attaining the upper surface of the head. Ratio eye width-synthlipsis 1:0.9-1.0. Antennae with second article uniformly blackish brown, third article light yellow except dark brown basal fourth; fourth article light-colored. Ratio of antennal segments 1:8.90-10.20:4.25-5.30:3.80-4.70. Rostrum dark brown-blackish, with the second segment extending beyond the level of ocelli and almost attaining the neck; ratio of rostral segments 1:3.93-4.00:0.88-0.90.

Pronotum dark brown-blackish, with submedian carinae and lateral margins light yellowish-brown, usually as very narrow stripes but wider in paler specimens (Fig. 3). Posterior lobe with two dark brown-blackish longitudinal stripes between submedian carinae and on each side between submedian carinae and lateral margins; these stripes are separated by slender light bands in typical dark specimens, and by lighter, more conspicuous bands in paler specimens. Anterolateral angles of pronotum not

prominent. Scutellum dark brown, with 1 + 1 light yellowish brown carinae, somewhat widened anteriorly and fused at the base of the light-colored, narrowly-pointed scutellar process.

Hemelytra dark brown, with veins bordered by conspicuous but narrow light-colored stripes; space between veins uniformly blackish-brown on the corium, sometimes mixed with lighter diffuse spots on the membrane. Legs with coxae almost entirely blackish- to dark-reddish-brown and lighter brown trochantera; femora dark reddish brown, tibiae of similar or slightly lighter color, darker at apex in this latter case. Legs slender, with fore femora about 7.5-8 times as long as wide.

Ventral surface of the abdomen dark brown, speckled with lighter brown in most but not all specimens; longitudinal medial blackish line, which narrows towards the inter-segment lines and is flanked by 1 + 1 submedian, light, wider bands; in each segment, these bands are in turn flanked by 1 + 1 sublateral oblique longitudinal marks.

Urosternites with a marginal, longitudinal darker band on each side, also bordered by lighter bands. Ventrally, the connexivum is barely visible or completely covered by the urosternites; dorsally light yellowish brown, with each segment almost completely covered by a large brownish-black spot that narrows somewhat, but only slightly in typical dark specimens, towards the posterior end.

*Morphometrics.* Linear measurements and ratios are presented in the Table.

Discriminant analysis of shape shows how specimens designated as *R. barretti* are distinct from *R. robustus s.l.* specimens from a wide geographical range (Fig. 2). The three Ecuadorian *R. barretti* paratypes are clearly nested within the Colombian type material; in addition, this type group overlaps with further specimens from Colombia.

*Molecular analysis.* The analysis of 13 sequences of *Rhodnius* for the *cyt b* gene fragments (663 bp) revealed 201 variable sites, 31 (15.4%) of which were

autapomorphies of *R. barretti*. *Cyt b* sequences of *R. barretti* (GenBank accession numbers JX273159 to JX273162) are unmistakably different from those of other ‘robustus lineage’ members. Such deep divergence is shown in the phylogram presented in Fig. 4; all *R. barretti* specimens cluster together in a 100%-supported clade that is basal to the rest of ‘robustus lineage’ species, including *R. nasutus* and *R. neglectus*. Pairwise *cyt b* sequence divergence between *R. barretti* and the other *Rhodnius* species, computed via M-CL, was very large, ranging from 7.4% to 10.7% (darker colors in the heatmap of Fig. 4). These values are larger than those between the four members of the *R. robustus* cryptic species complex and *R. nasutus*, or between *R. robustus* I-V and *R. neglectus*. The pairwise sequence divergence between Ecuadorian *R. barretti* and sympatric, field-collected *R. robustus* is 7.5%. Overall, the genetic distance between *R. robustus* and *R. barretti* is 8.7% (variance = 2.0%; range, 7.4-9.7%). This is strong evidence that *R. barretti* represents a distinct entity within the ‘robustus lineage’. Two *R. barretti* specimens collected in Colombia share the same *cyt b* haplotype; while this haplotype closely resembles the other two, present in Ecuador, the finding of three haplotypes in a sample of four bugs (mean distance = 0.3%; variance = 0.1%) is suggestive of high levels of genetic diversity.

*Biology and bionomics.* *R. barretti* has been collected in *Attalea butyracea* and *Oenocarpus bataua* palm trees in the Ecuadorian Amazon, but may also occur in other palm species (Abad-Franch et al. 2001, Noireau et al. 2002). These palm tree-living colonies are common and can reach high densities (Abad-Franch et al. 2010); starved bugs are strikingly aggressive, readily attacking collectors in open daylight (FA-F & FSP, pers. obs.). In contrast, sympatric *R. robustus* are typically found as small colonies (FA-F & FSP, pers. obs.). Field observations in Ecuador indicate that *R. barretti* adult specimens can invade houses (FSP, pers. obs.), probably by flying

from heavily infested peridomestic palms. Palomeque et al. (2003) studied the biology and bionomics of Ecuadorian *R. barretti* specimens (at the time identified as *R. robustus*) under laboratory conditions; the results show that, when fed on mammal (mouse) blood, *R. barretti* (i) can produce two generations per year, with a mean egg-to-adult time of  $128 \pm 7$  days; (ii) require about 290mg of blood to complete development; and (iii) often defecate (especially older nymphs and adults) during or a few minutes after feeding. These bionomic traits, comparable to those of *R. prolixus*, suggest that *R. barretti* may be an efficient vector of *T. cruzi* (Palomeque et al. 2003).

*Etymology.* This species is dedicated to Dr Toby Vincent Barrett (Instituto Nacional de Pesquisas da Amazônia – INPA, Brazil), in all likelihood the world's top authority on Amazonian triatomine ecology.

*Taxonomic discussion.* Phenotype-based differentiation of *R. barretti* from *R. robustus*, the most closely related species found in sympatry, is difficult. However, sympatric Ecuadorian *bona fide* *R. robustus* (genotype II) are conspicuously larger and lighter-colored (Fig. 4), which should provide a means for field identification by well-trained local entomologists. In addition, the length ratios of several anatomical structures appear to be consistently different from those typically reported for *R. robustus s.l.* (Table). In particular, *R. barretti* heads seem to be longer relative to total body length (with the head accounting for ~20-22% of total length) than those of *R. robustus s.l.* (~17-20%). The third antennal segment appears to be relatively shorter in *R. barretti* (up to 5.3 times as long as the first segment, vs. up to 5.6 in *R. robustus s.l.*); similarly, the second rostral segment is up to 4.0 times as long as the first one in *R. barretti*, vs. up to 4.4 times in *R. robustus s.l.*. Finally, the scutellar process is narrowly pointed in *R. barretti* but stout/rounded in sympatric *R. robustus* (Fig. 4). Although these differences will need to be confirmed with further samples, dubious

assignments can be resolved, as shown above and in Márquez et al. (2011), using morphometrics and/or DNA analysis.

## Discussion

We described *R. barretti* sp. nov. on the basis of an integrative assessment including mtDNA sequencing, morphometrics, and qualitative phenotype traits (Abad-Franch & Monteiro 2005, Schofield & Galvão 2009). As for the rest of species within the ‘robustus lineage’, *R. barretti* is hard to distinguish from its closest relatives based only on phenotype (Lent & Wygodzinsky 1979). Still, *R. robustus* genotype II, the only member of the lineage recorded to date in the Napo ecoregion (Abad-Franch et al. 2009), can be distinguished from *R. barretti* by its overall larger size and lighter color (see Fig. 4); it was this difference, in fact, that led one of us to suspect that a distinct taxon might co-occur with *R. robustus* in north-eastern Ecuador (Abad-Franch 2003). In addition, the length ratios of some anatomic structures and the apex of the scutellar process appear to be consistently different from those of *R. robustus s.l.*. Although we have so far systematically compared only a small sample, *R. barretti* nymphs also seem to differ from those of sympatric *R. robustus* populations. The former are overall dark brown-reddish with strikingly black markings on the thorax and wing pads; the head is very dark except for a slender, median dorsal stripe, and the legs are brown-reddish with little difference between tibiae and femora. *R. robustus* nymphs are of much lighter overall color, with brown markings on a yellowish background; their tibiae are usually visibly lighter than the femora (see also Lent & Wygodzinsky 1979).

Morphometric analyses can help delineate species boundaries among nearly-cryptic triatomine taxa, and are useful for solving taxonomic doubts when adult specimens are involved (Gurgel-Gonçalves et al. 2008). However, and despite consistent

phenotype-based evidence, it was the results of mtDNA sequence analysis that gave us confidence to conclude that the atypical Napo population belongs to a new species (see also Márquez et al. 2011). MtDNA sequencing is now widely available at a reasonable cost; furthermore, any developmental stage, from eggs to adult bugs, can be used for DNA sequencing. The results not only provide insight on the taxonomic status and phylogenetic relationships of problem specimens (Abad-Franch & Monteiro 2005, Justi et al. 2010), but also allow for the development of PCR-based species identification assays (Pavan & Monteiro 2007, Pavan et al. 2013). While mtDNA sequencing results might be difficult to interpret when reciprocally monophyletic sister clades are separated by low levels (<4%) of sequence divergence (Wiemers et al. 2007), this was not the case with *R. barretti*, whose divergence from its closest relatives is >8.7%. According to the *cyt b* strict molecular clock proposed by Pfeiler et al. (2006) for triatomines, both lineages have probably been evolving independently for about 4-9 million years. Márquez et al. (2011) also reported *cyt b* distances (Kimura 2-parameter)  $\geq 10\%$ ; inter-species comparisons show that the smallest *cyt b* distance within the genus *Rhodnius* (a 4.3% distance) is that between the phenotypically and ecologically well-differentiated *R. brethesi* and *R. stali* (FAM & MGP, unpublished). Márquez et al. (2011) showed, in addition, that *R. barretti* nuclear 28S rRNA D2 variable region sequences are unique when compared with those of other *Rhodnius* species and genotypes available in GenBank.

We finally note that there could remain some doubts as to whether the Napo population could in fact be *R. dalessandroi* Carcavallo & Barreto 1976. This little-known species, collected from *Oenocarpus* palms in the Amazon-Orinoco transition, was excluded from the classic monograph by Lent & Wygodzinsky (1979), and there is no DNA sequence available. Yet, comparison with *R. dalessandroi* descriptions

(Carcavallo & Barreto 1976, Martínez 1984) clearly suggests that *R. barretti* and *R. dalessandroi* are distinct entities. Morphological differences include head/neck and urotergite coloration and hemelytron:abdomen, antenna article and rostrum segment length ratios (Carcavallo & Barreto 1976, Martínez 1984). These species, in addition, occupy sharply different ecoregions (Abad-Franch et al. 2009).

In conclusion, the results we report here show how an integrative systematic appraisal can help uncover apparently cryptic diversity even within relatively well-studied taxonomic groups. A hitherto undescribed species of *Rhodnius* was discovered in the rainforest of the Napo ecoregion – one of the major biodiversity hotspots on Earth (Olson & Dinerstein 1998, Gaston 2010). This sub-region of Amazonia has also been the focus of detailed surveys on Chagas disease epidemiology over the last two decades (Amunárriz et al. 1991, 2011, Chico et al. 1997, Aguilar et al. 1999, Grijalva et al. 2003, Guevara et al. 2013). All of them converge towards a coherent set of conclusions: (i) that the infection is endemic in the region, with about 2-4% seropositivity among local residents; (ii) that *T. cruzi* transmission to humans has been continuous over the last decades and is currently active; and (iii) that such transmission is maintained by adventitious adult triatomines that invade but do not colonize human dwellings. Within the genus *Rhodnius*, *R. pictipes* and *R. robustus* were up to now blamed for this scenario (Abad-Franch et al. 2001, Noireau et al. 2002); *R. barretti* is a new candidate vector species, and its biological and behavioral characteristics suggest that it may play a significant role in Chagas disease epidemiology in the Napo ecoregion.

## Acknowledgements

We thank MA Miles, M Aguilar and MJ Grijalva for continuous support, SA Valente and MM Lima for kindly providing *Rhodnius neglectus* and *R. nasutus* specimens, and A Gonçalves da Silva for useful remarks. We are particularly grateful to Cléber Galvão for his insightful comments and suggestions. This is contribution number XX of the Research Program on Infectious Disease Ecology in the Amazon (RP-IDEA) of the Instituto Leônidas e Maria Deane. This paper is dedicated to the memory of Alexandre Afrânio Peixoto, whose inspiring research on speciation in insect disease vectors deeply influences much of our own work.

## References

- Abad-Franch F 2003. *The ecology and genetics of Chagas disease vectors in Ecuador, with emphasis on Rhodnius ecuadoriensis (Triatominae)*. PhD Thesis, London School of Hygiene and Tropical Medicine, University of London, UK.
- Abad-Franch F, Ferraz G, Campos C, Palomeque FS, Grijalva MJ, Aguilar HM, Miles MA 2010. Modeling disease vector occurrence when detection is imperfect: infestation of Amazonian palm trees by triatomine bugs at three spatial scales *PLoS Negl Trop Dis* 4: e620.
- Abad-Franch F, Monteiro FA 2005. Molecular research and the control of Chagas disease vectors. *An Acad Bras Ciências* 77: 437-454.
- Abad-Franch F, Monteiro FA 2007. Biogeography and evolution of Amazonian triatomines (Heteroptera: Reduviidae): implications for Chagas disease

- surveillance in humid forest ecoregions. *Mem Inst Oswaldo Cruz* 102 (Suppl. 1): 57-69.
- Abad-Franch F, Monteiro FA, Jaramillo N, Gurgel-Gonçalves R, Dias FBS, Diotauiti L 2009. Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: a multi-scale appraisal of the tribe Rhodniini (Triatominae). *Acta Trop* 110: 159-177.
- Abad-Franch F, Paucar CA, Carpio CC, Cuba Cuba CA, Aguilar VHM, Miles MA 2001. Biogeography of Triatominae (Hemiptera: Reduviidae) in Ecuador: implications for the design of control strategies. *Mem Inst Oswaldo Cruz* 96: 611-620.
- Aguilar HM, Abad-Franch F, Racines J, Paucar CA 1999. Epidemiology of Chagas disease in Ecuador. A brief review. *Mem Inst Oswaldo Cruz* 94 (Suppl. 1): 387-393.
- Aguilar HM, Abad-Franch F, Dias JCP, Junqueira ACV, Coura JR 2007. Chagas disease in the Amazon region. *Mem Inst Oswaldo Cruz* 102 (Suppl. 1): 47-55.
- Amunárriz MU, Chico ME, Guderian RH 1991. Chagas disease in Ecuador: a sylvatic focus in the Amazon region. *J Trop Med Hyg* 94: 145-149.
- Amunárriz M, Quito S, Tandazo V, López M 2010. Seroprevalencia de la enfermedad de Chagas en el cantón Aguarico, Amazonía Ecuatoriana. *Rev Panam Salud Pública* 28: 25-29.
- Barrett TV 1991. Advances in triatomine bug ecology in relation to Chagas disease. In KH Harris (org.) *Advances in Disease Vector Research*, Vol. 8. Springer-Verlag, New York, p. 143-176.

- Bickford D, Lohman DJ, Sodhi NS, Ng PKL, Meier R, Winker K, Ingram KK, Das I 2006. Cryptic species as a window on diversity and conservation. *Trends Ecol Evol* 22: 148-155.
- Calzada JE, Pineda V, Garisto JD, Samudio F, Santamaría AM, Saldaña A 2010. Human trypanosomiasis in the eastern region of the Panama Province: new endemic areas for Chagas disease. *Am J Trop Med Hyg* 82: 580-582.
- Carcavallo R, Barreto P 1976. Una nueva especie de *Rhodnius* Stål (Hemiptera, Reduviidae, Triatominae) de Colombia. *Bol Dir Malariol Saneamiento Ambiental* 16: 176-183.
- Chico HM, Sandoval C, Guevara EA, Calvopiña HM, Cooper PJ, Reed SG, Guderian RH 1997. Chagas disease in Ecuador: evidence for disease transmission in an indigenous population in the Amazon region. *Mem Inst Oswaldo Cruz* 92: 317-320.
- Da Rosa JA, Rocha CS, Gardim S, Pinto MC, Mendonça VJ, Ferreira Filho JCR, de Carvalho EOC, Camargo LMA, de Oliveira J, Nascimento JD, Cilense M, Almeida CE 2012. Description of *Rhodnius montenegrensis* sp. nov. (Hemiptera: Reduviidae: Triatominae) from the state of Rondônia, Brazil. *Zootaxa* 3478: 62-76.
- Dayrat B 2005. Towards integrative taxonomy. *Biol J Linn Soc* 85: 407-415.
- Drummond AJ, Rambaut A 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
- Fitzpatrick S, Feliciangeli MD, Sánchez-Martín M, Monteiro FA, Miles MA 2008. Molecular genetics reveal that silvatic *Rhodnius prolixus* do colonise rural houses. *PLoS Negl Trop Dis* 2: e210.

- Galvão C, Carcavallo RU, Rocha DS, Jurberg J 2003. A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa* 202: 1-36.
- Gaston KJ 2010. Biodiversity. In: NS Sodhi, PR Ehrlich (eds.) *Conservation Biology for All*. Oxford University Press, Oxford, p. 27-42.
- Gómez-Palacio A, Jaramillo-O N, Caro-Riaño H, Diaz S, Monteiro FA, Pérez R, Panzera F, Triana O 2012. Morphometric and molecular evidence of intraspecific biogeographical differentiation of *Rhodnius pallescens* (HEMIPTERA: REDUVIIDAE: RHODNIINI) from Colombia and Panama. *Infect Genet Evol* 12: 1975-1983.
- Grijalva MJ, Escalante L, Paredes RA, Costales JA, Padilla A, Rowland EC, Aguilar HM, Racines J 2003. Seroprevalence and risk factors for *Trypanosoma cruzi* infection in the Amazon region of Ecuador. *Am J Trop Med Hyg* 69: 380-385.
- Guevara AG, Atherton RD, Wauters MA, Vicuña Y, Nelson M, Prado J, Kato H, Calvopiña MH, Hashiguchi Y 2013. Seroepidemiological study of Chagas disease in the southern Amazon region of Ecuador. *Trop Med Health* 41: 21-25.
- Gurgel-Gonçalves R, Abad-Franch F, Ferreira JBC, Santana DB, Cuba Cuba CA 2008. Is *Rhodnius prolixus* (Triatominae) invading houses in central Brazil? *Acta Trop* 107: 90-98.
- Harry M 1994. Morphometric variability in the Chagas disease vector *Rhodnius prolixus*. *Jpn J Gen* 69: 233-250.

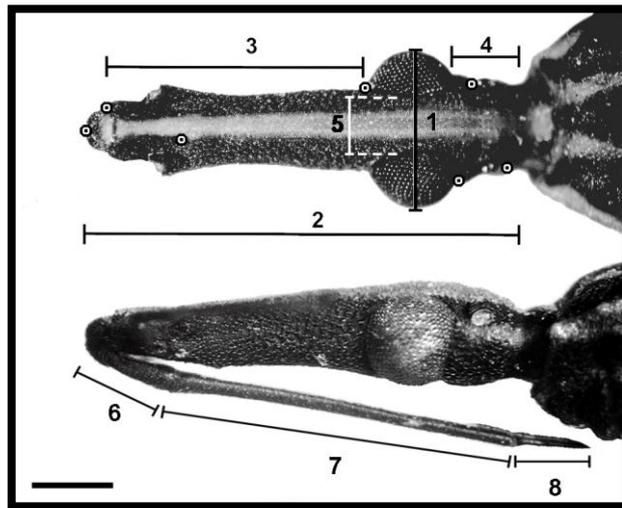
- Hornik K 2011. The R FAQ. Available online at <http://CRAN.R-project.org/doc/FAQ/>.
- Jurberg J, Rocha DS, Galvão C 2009. *Rhodnius zeledoni* sp. nov. afim de *Rhodnius paraensis* Sherlock, Guitton & Miles, 1977 (Hemiptera, Reduviidae, Triatominae). *Biota Neotrop* 9: 123-1218.
- Justi SA, Noireau F, Cortez MR, Monteiro FA 2010. Infestation of peridomestic *Attalea phalerata* palms by *Rhodnius stali*, a vector of *Trypanosoma cruzi* in the Alto Beni, Bolivia. *Trop Med Int Health* 15: 727-732.
- Lent H, Wygodzinsky P 1979. Revision of the Triatominae (Hemiptera, Reduviidae), and their significance as vectors of Chagas' disease. *Bull Am Mus Nat Hist* 163: 125-520.
- Márquez E, Jaramillo-O N, Gómez-Palacio A, Dujardin JP 2011. Morphometric and molecular differentiation of a *Rhodnius robustus*-like form from *R. robustus* Larousse, 1927 and *R. prolixus* Stål, 1859 (Hemiptera, Reduviidae). *Acta Trop* 120: 103-109.
- Martínez A 1984. Caracterización taxionómica de *Rhodnius dalessandroi* Carcavallo & Barreto, 1976 (Hemiptera, Reduviidae, Triatominae). *Chagas I*: 29-31.
- Monteiro FA, Barrett TV, Fitzpatrick S, Cordon-Rosales C, Feliciangeli D, Beard CB 2003. Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. *Mol Ecol* 12: 997-1006.
- Monteiro FA, Escalante AA, Beard CB 2001. Molecular tools and triatomine systematics: a public health perspective. *Trends Parasitol* 17: 344-347.

- Monteiro FA, Lazoski C, Noireau F, Solé-Cava AM 2002. Allozyme relationships among ten species of Rhodniini, showing paraphyly of *Rhodnius* including *Psammolestes*. *Med Vet Entomol* 16: 83-90.
- Monteiro FA, Wesson DM, Dotson EM, Schofield CJ, Beard CB 2000. Phylogeny and molecular taxonomy of the Rhodniini derived from mitochondrial and nuclear DNA sequences. *Am J Trop Med Hyg* 62: 460-465.
- Noireau F, Abad-Franch F, Valente SAS, Dias-Lima A, Lopes CM, Cunha V, Valente VC, Palomeque FS, de Carvalho-Pinto CJ, Sherlock I, Aguilar M, Steindel M, Grisard EC, Jurberg J 2002. Trapping Triatominae (Hemiptera: Reduviidae) in silvatic habitats. *Mem Inst Oswaldo Cruz* 97: 61-63.
- Olson D, Dinerstein E 1998. The Global 200: a representation approach to conserving the Earth's most biologically valuable ecoregions. *Conserv Biol* 12: 502-515.
- Palomeque FS, Abad-Franch F, Grijalva MJ 2003. Notas sobre la biología y bionomía de *Rhodnius robustus* Larrouse, 1927 (Hemiptera: Reduviidae: Triatominae). *Rev Pontif Univ Católica Ecuador* 71: 49-60.
- Pavan MG, Mesquita RD, Lawrence GG, Lazoski C, Dotson EM, Abubucker S, Mitreva M, Randall-Maher J, Monteiro FA 2013. A nuclear single-nucleotide polymorphism (SNP) potentially useful for the separation of *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). *Infect Genet Evol* 14: 426-433.
- Pavan MG, Monteiro FA 2007. A multiplex PCR assay that separates *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). *Trop Med Int Health* 12: 751-758.

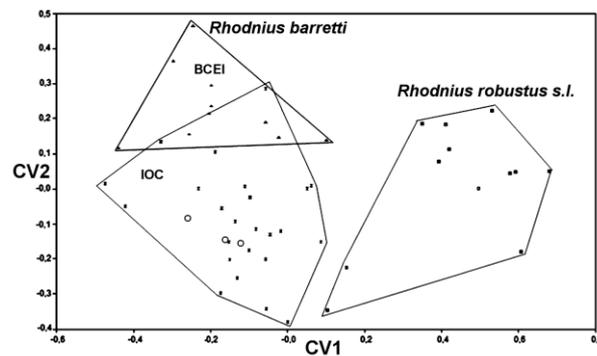
- Pfeiler E, Bitler BG, Ramsey JM, Palacios-Cardiel C, Markow TA 2006. Genetic variation, population structure, and phylogenetic relationships of *Triatoma rubida* and *T. recurva* (Hemiptera: Reduviidae: Triatominae) from the Sonoran desert, insect vectors of the Chagas' disease parasite *Trypanosoma cruzi*. *Mol Phylogenet Evol* 41: 209-221.
- Posada D 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol* 25: 1253-1256.
- Rohlf FJ 1996. Morphometric spaces, shape components and the effects of linear transformations. In: LF Marcus, M Corti, A Loy, GJP Naylor, D Slice (eds.) *Advances in morphometrics*. Plenum, New York, p. 117-129.
- Rohlf FJ 1999a. *tpsDig*, version 1.18. Department of Ecology and Evolution, State University of New York, Stony Brook, NY (<http://life.bio.sunysb.edu/morph/>).
- Rohlf FJ 1999b. *tpsRelw*, version 1.18. Department of Ecology and Evolution, State University of New York, Stony Brook, NY (<http://life.bio.sunysb.edu/morph/>).
- Schofield CJ, Diotaiuti L, Dujardin JP 1999. The process of domestication in Triatominae. *Mem Inst Oswaldo Cruz* 94 (Suppl. 1): 375-378.
- Schofield CJ, Galvão C 2009. Classification, evolution, and species groups within the Triatominae. *Acta Trop* 110: 88-100.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739.

- Valente SAS, Valente VC, Pinto AYN, Barbosa CMJ, dos Santos MP, Miranda CO, Cuervo P, Fernandes O 2009. Analysis of an acute Chagas disease outbreak in the Brazilian Amazon: human cases, triatomines, reservoir mammals and parasites. *Trans R Soc Trop Med Hyg* 103: 291-297.
- Valente VC, Valente SAS, Carcavallo RU, Rocha DS, Galvão C, Jurberg J 2001. Considerações sobre uma nova espécie do gênero *Rhodnius* Stål, do estado Pará, Brasil (Hemiptera, Reduviidae, Triatominae). *Entomol Vectores* 8: 65-80.
- Wiemers M, Fiedler K 2007. Does the DNA barcoding gap exist? – A case study in blue butterflies (Lepidoptera: Lycaenidae). *Front Zool* 4: 8.

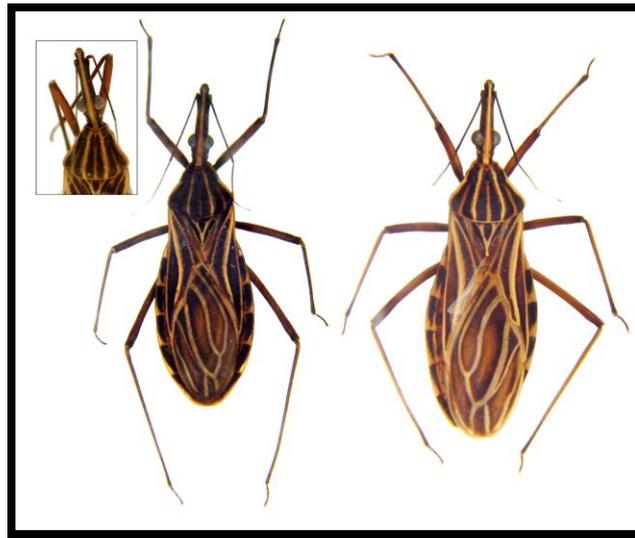
## Figures and Tables



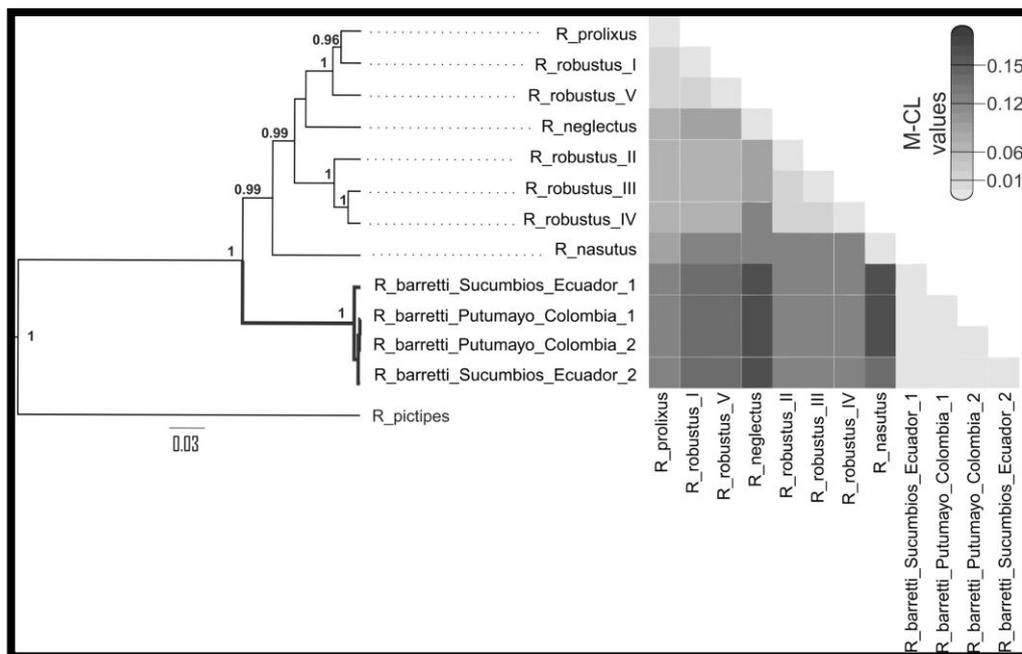
**Figure 1.** Dorsal (top) and lateral (bottom) views of a *Rhodnius barretti* head, showing linear measurements (1-8; see Table) and landmarks (target spots) used in morphometric analyses (scale bar = 1mm).



**Figure 2.** Geometric morphometrics results: factorial map built with the first two canonical vectors (CV) showing 100% head shape discrimination between *Rhodnius barretti* sp. nov. (Colombia and Ecuador) and *R. robustus s.l.* (Bolivia, Brazil, Colombia, Ecuador, and French Guiana). *R. barretti* are deposited at the BCEI (Medellín, Colombia) and IOC (Rio de Janeiro, Brazil) collections, as indicated. Ecuadorian specimens are marked as open circles.



**Figure 3.** *Rhodnius barretti* sp. nov. (left) and sympatric *R. robustus* s.l. (right) from north-eastern Ecuador: general aspect of adult specimens in dorsal view. The inset shows the head and pronotum of an atypical, paler form of *R. barretti*.



**Figure 4.** Bayesian consensus tree of 13 663bp cyt *b* *Rhodnius* sequences and a heatmap illustrating genetic distance values. Posterior probabilities >0.95 are shown for key nodes. Squares in the heatmap illustrate pairwise genetic distance values between sequences, with darker colors representing larger distances (see scale bar). Collection sites and GenBank accession numbers are: *R. prolixus* – Portuguesa, Venezuela, EF011723; *R. robustus* I – Trujillo, Venezuela, AF421340; *R. robustus* II – Napo, Ecuador, AF421341; *R. robustus* III – Pará, Brazil, AF421342; *R. robustus* IV – Pará, Brazil, AF421342; *R. robustus* V – Novo Airão, Amazonas, Brazil, JX273158; *R. nasutus* – Ceará, Brazil, JX273155; *R. neglectus* – Tocantins, Brazil, JX273156; *R. pictipes* – Amazonas, Brazil; *R. barretti* – Lago Agrio, Sucumbíos, Ecuador (1), JX273159; *R. barretti* – Lago Agrio, Sucumbíos, Ecuador (2), JX273160; *R. barretti* – Puerto Asís, Putumayo, Colombia (1 and 2), JX273161 and JX273162.

**Table.** Anatomical structure metrics in adult *Rhodnius barretti* sp. nov. and *Rhodnius robustus* s.l. Note that some measurements are given as ranges and some as the mean  $\pm$  2 standard errors (SE) and are therefore not directly comparable; see the Table footnote.

| Metric            | Description   | <i>Rhodnius barretti</i> sp. nov. |             | <i>Rhodnius robustus</i> s.l. |                          |
|-------------------|---|-----------------------------------|-------------|-------------------------------|--------------------------|
|                   |   | Male                              | Female      | Male                          | Female                   |
| Measurements (mm) | Total length  | 19.1-22.7                         | 20.3-24.7   | 20.0-23.5 <sup>a</sup>        | 23.0-26.0 <sup>a</sup>   |
|                   | Maximum width of pronotum                           | 4.00-4.70                         | 4.10-5.00   | 4.50-5.00 <sup>a</sup>        | 5.00-6.00 <sup>a</sup>   |
|                   | Maximum width of abdomen                            | 5.70-6.00                         | 5.60-7.20   | 6.00-7.00 <sup>a</sup>        | 6.00-7.50 <sup>a</sup>   |
|                   | Outer (maximum) distance across eyes (#1 in Fig. 1) | 1.60-1.87                         | 1.60-1.93   | Nr                            | Nr                       |
|                   | Synthlipsis (#5 in Fig. 1)                          | 0.47-0.61                         | 0.56-0.70   | 0.46-0.52 <sup>b</sup>        | 0.52-0.57 <sup>b</sup>   |
|                   | Anteocular distance (#3 in Fig. 1)                  | 2.50-3.05                         | 2.50-3.30   | 2.52-2.67 <sup>b</sup>        | 2.66-2.83 <sup>b</sup>   |
|                   | Postocular distance (#4 in Fig. 1)                  | 0.65-0.80                         | 0.55-0.85   | 0.65-0.72 <sup>b</sup>        | 0.68-0.72 <sup>b</sup>   |
|                   | Total length of head (#2 in Fig. 1)                 | 4.10-5.00                         | 4.00-5.25   | 3.97-4.24 <sup>b</sup>        | 4.17-4.44 <sup>b</sup>   |
|                   | Length of first rostral segment (#6 in Fig. 1)      | 0.80-1.00                         | 0.96-1.08   | 0.70-0.75 <sup>b</sup>        | 0.75-0.81 <sup>b</sup>   |
|                   | Length of second rostral segment (#7 in Fig. 1)     | 3.50-4.00                         | 3.30-4.20   | 2.99-3.14 <sup>b</sup>        | 3.08-3.36 <sup>b</sup>   |
|                   | Length of third rostral segment (#8 in Fig. 1)      | 0.78-1.00                         | 0.80-0.91   | 0.87-0.93 <sup>b</sup>        | 0.87-0.96 <sup>b</sup>   |
| Ratios            | Total length:head length                            | 1:0.21-0.22                       | 1:0.20-0.21 | 1:0.18-0.20 <sup>c</sup>      | 1:0.17-0.18 <sup>c</sup> |
|                   | Total length:pronotum width                         | 1:0.21                            | 1:0.20      | 1:0.21-0.23 <sup>a</sup>      | 1:0.22-0.23 <sup>a</sup> |
|                   | Total length:abdomen width                          | 1:0.27-0.30                       | 1:0.28-0.29 | 1:0.30 <sup>a</sup>           | 1:0.26-0.29 <sup>a</sup> |
|                   | Abdomen width:pronotum width                        | 1:0.70-0.78                       | 1:0.69-0.73 | 1:0.71-0.75 <sup>a</sup>      | 1:0.80-0.83 <sup>a</sup> |
|                   | Head length:head width                              |                                   | 1:0.36-0.41 |                               | 1:0.40 <sup>a</sup>      |
|                   | Head length:pronotum length                         |                                   | 1:0.72-0.80 |                               | 1:0.70-0.80 <sup>c</sup> |
|                   | Eye width:synthlipsis                               |                                   | 1:0.90-1.00 |                               | 1:0.75-1.00 <sup>a</sup> |
|                   | Anteocular region:postocular region                 |                                   | 1:0.23-0.26 |                               | 1:0.25 <sup>a</sup>      |
|                   | 1st:2nd antennal segments                           |                                   | 1:8.90-10.2 |                               | 1:7.60-10.2 <sup>a</sup> |
|                   | 1st:3rd antennal segments                           |                                   | 1:4.25-5.30 |                               | 1:5.40-6.30 <sup>a</sup> |
|                   | 1st:4th antennal segments                           |                                   | 1:3.80-4.70 |                               | 1:3.40-4.40 <sup>a</sup> |
|                   | 1st:2nd rostral segments                            |                                   | 1:3.93-4.00 |                               | 1:4.10-4.40 <sup>a</sup> |
|                   | 1st:3rd rostral segments                            |                                   | 1:0.88-0.90 |                               | 1:0.90-1.00 <sup>a</sup> |

*a*: after Lent & Wygodzinsky (1979); *b*: mean $\pm$ 2SE, with values taken from Harry (1994); *c*: with head length taken from Harry (1994); Nr: not reported.

## 10. REFERÊNCIAS BIBLIOGRÁFICAS

- Abad-Franch, F., Diotaiuti, L., Gurgel-Goncalves, R., Gurtler, R.E., 2013. Certifying the interruption of Chagas disease transmission by native vectors: cui bono? *Mem Inst Oswaldo Cruz* 108, 251-254.
- Abad-Franch, F., Ferraz, G., Campos, C., Palomeque, F.S., Grijalva, M.J., Aguilar, H.M., Miles, M.A., 2010a. Modeling disease vector occurrence when detection is imperfect: infestation of Amazonian palm trees by triatomine bugs at three spatial scales. *PLoS Negl Trop Dis* 4, e620.
- Abad-Franch, F., Monteiro, F., 2005. Molecular research and the control of Chagas disease vectors. *Anais da Academia Brasileira de Ciências* 77, 437-454.
- Abad-Franch, F., Monteiro, F.A., 2007. Biogeography and evolution of Amazonian triatomines (Heteroptera: Reduviidae): implications for Chagas disease surveillance in humid forest ecoregions. *Mem Inst Oswaldo Cruz* 102 Suppl 1, 57-70.
- Abad-Franch, F., Monteiro, F.A., Jaramillo, O.N., Gurgel-Goncalves, R., Dias, F.B., Diotaiuti, L., 2009. Ecology, evolution, and the long-term surveillance of vector-borne Chagas disease: a multi-scale appraisal of the tribe Rhodniini (Triatominae). *Acta Trop* 110, 159-177.
- Abad-Franch, F., Palomeque, F.S., Aguilar, H.M., Miles, M.A., 2005. Field ecology of sylvatic *Rhodnius* populations (Heteroptera, Triatominae): risk factors for palm tree infestation in western Ecuador. *Trop Med Int Health* 10, 1258-1266.
- Abad-Franch, F., Santos, W.S., Schofield, C.J., 2010b. Research needs for Chagas disease prevention. *Acta Trop* 115, 44-54.
- Abruzzi, K.C., Rodriguez, J., Menet, J.S., Desrochers, J., Zadina, A., Luo, W., Tkachev, S., Rosbash, M., 2011. *Drosophila* CLOCK target gene characterization: implications for circadian tissue-specific gene expression. *Genes Dev* 25, 2374-2386.
- Aguilar, H.M., Abad-Franch, F., Dias, J.C., Junqueira, A.C., Coura, J.R., 2007. Chagas disease in the Amazon region. *Mem Inst Oswaldo Cruz* 102 Suppl 1, 47-56.
- Aguilar, V.H., Abad-Franch, F., Racines, V.J., Paucar, C.A., 1999. Epidemiology of Chagas disease in Ecuador. A brief review. *Mem Inst Oswaldo Cruz* 94 Suppl 1, 387-393.
- Aleixo, A., 2004. Historical diversification of a terra-firme forest bird superspecies: a phylogeographic perspective on the role of different hypotheses of Amazonian diversification. *Evolution* 58, 1303-1317.
- Alekseyenko, A.V., Lee, C.J., Suchard, M.A., 2008. Wagner and Dollo: a stochastic duet by composing two parsimonious solos. *Syst Biol* 57, 772-784.
- Allada, R., Chung, B.Y., 2010. Circadian organization of behavior and physiology in *Drosophila*. *Annu Rev Physiol* 72, 605-624.
- Allada, R., White, N.E., So, W.V., Hall, J.C., Rosbash, M., 1998. A mutant *Drosophila* homolog of mammalian Clock disrupts circadian rhythms and transcription of period and timeless. *Cell* 93, 791-804.
- Altschul, S.F., Gertz, E.M., Agarwala, R., Schaffer, A.A., Yu, Y.K., 2009. PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res* 37, 815-824.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389-3402.
- Anderson, J.M., Lai, J.E., Dotson, E.M., Cordon-Rosales, C., Ponce, C., Norris, D.E., Beard, C.B., 2002. Identification and characterization of microsatellite markers in the Chagas disease vector *Triatoma dimidiata*. *Infect Genet Evol* 1, 243-248.
- Anhuf, D., Ledru, M.P., Behling, H., Da Cruz Jr, F.W., Cordeiro, R.C., Van der Hammen, T., Karmann, I., Marengo, J.A., De Oliveira, P.E., Pessenda, L., Siffedine, A., Albuquerque, A.L., Da Silva Dias, E.L., 2006. Paleo-environmental change in Amazonian and African rainforest during the LGM. *Palaeogeography, Palaeoclimatology, Palaeoecology* 239, 510-527.
- Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N., Yeh, L.S., 2004. UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res* 32, D115-119.
- Araki, A.S., Vigoder, F.M., Bauzer, L.G., Ferreira, G.E., Souza, N.A., Araujo, I.B., Hamilton, J.G., Brazil, R.P., Peixoto, A.A., 2009. Molecular and behavioral differentiation among Brazilian populations of *Lutzomyia longipalpis* (Diptera: Psychodidae: Phlebotominae). *PLoS Negl Trop Dis* 3, e365.
- Aschoff, J., 1960. Exogenous and endogenous components in circadian rhythms. *Cold Spring Harb Symp Quant Biol* 25, 11-28.
- Atkinson, Q.D., Gray, R.D., Drummond, A.J., 2008. mtDNA variation predicts population size in humans and reveals a major Southern Asian chapter in human prehistory. *Mol Biol Evol* 25, 468-474.
- Avice, J.C., 2000. *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge, MA.

- Avice, J.C., 2009. Phylogeography: retrospect and prospect. *J Biogeogr* 36, 3-15.
- Avice, J.C., Ball, R.M., Arnold, J., 1988. Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Mol Biol Evol* 5, 331-344.
- Baele, G., Li, W.L., Drummond, A.J., Suchard, M.A., Lemey, P., 2012. Accurate Model Selection of Relaxed Molecular Clocks in Bayesian Phylogenetics. *Mol Biol Evol*.
- Ball, R.M., Freeman, S., James, F.C., Bermingham, E., Avice, J.C., 1988. Phylogeographic population structure of Red-winged Blackbirds assessed by mitochondrial DNA. *Proc Natl Acad Sci U S A* 85, 1558-1562.
- Bandelt, H.J., Forster, P., Rohl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16, 37-48.
- Bargues, M.D., Klisiowicz, D.R., Gonzalez-Candelas, F., Ramsey, J.M., Monroy, C., Ponce, C., Salazar-Schettino, P.M., Panzera, F., Abad-Franch, F., Sousa, O.E., Schofield, C.J., Dujardin, J.P., Guhl, F., Mas-Coma, S., 2008. Phylogeography and genetic variation of *Triatoma dimidiata*, the main Chagas disease vector in Central America, and its position within the genus *Triatoma*. *PLoS Negl Trop Dis* 2, e233.
- Bargues, M.D., Klisiowicz, D.R., Panzera, F., Noireau, F., Marcilla, A., Perez, R., Rojas, M.G., O'Connor, J.E., Gonzalez-Candelas, F., Galvao, C., Jurberg, J., Carcavallo, R.U., Dujardin, J.P., Mas-Coma, S., 2006. Origin and phylogeography of the Chagas disease main vector *Triatoma infestans* based on nuclear rDNA sequences and genome size. *Infect Genet Evol* 6, 46-62.
- Barrozo, R.B., Minoli, S.A., Lazzari, C.R., 2004. Circadian rhythm of behavioural responsiveness to carbon dioxide in the blood-sucking bug *Triatoma infestans* (Heteroptera: Reduviidae). *J Insect Physiol* 50, 249-254.
- Battistuzzi, F.U., Filipowski, A., Hedges, S.B., Kumar, S., 2010. Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Mol Biol Evol* 27, 1289-1300.
- Bauzer, L.G., Gesto, J.S., Souza, N.A., Ward, R.D., Hamilton, J.G., Kyriacou, C.P., Peixoto, A.A., 2002a. Molecular divergence in the period gene between two putative sympatric species of the *Lutzomyia longipalpis* complex. *Mol Biol Evol* 19, 1624-1627.
- Bauzer, L.G., Souza, N.A., Ward, R.D., Kyriacou, C.P., Peixoto, A.A., 2002b. The period gene and genetic differentiation between three Brazilian populations of *Lutzomyia longipalpis*. *Insect Mol Biol* 11, 315-323.
- Beerli, P., 2006. Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics* 22, 341-345.
- Beerli, P., Palczewski, M., 2010. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics* 185, 313-326.
- Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. *Ann Statist* 29, 1165-1188.
- Benton, M.J., 1999. Early origins of modern birds and mammals: molecules vs. morphology. *Bioessays* 21, 1043-1051.
- Blanco, E., Parra, G., Guigó, R., 2007. Using geneid to identify genes. *Curr Protocols in Bioinformatics* 4, Unit 4.3.
- Bodin, A., Barrozo, R.B., Couton, L., Lazzari, C.R., 2008. Temporal modulation and adaptive control of the behavioural response to odours in *Rhodnius prolixus*. *J Insect Physiol* 54, 1343-1348.
- Boothroyd, C.E., Young, M.W., 2008. The in(put)s and out(put)s of the *Drosophila* circadian clock. *Ann N Y Acad Sci* 1129, 350-357.
- Bossu, C.M., Near, T.J., 2009. Gene trees reveal repeated instances of mitochondrial DNA introgression in orangethroat darters (percidae: *etheostoma*). *Syst Biol* 58, 114-129.
- Bower, J.E., Downton, M., Cooper, R.D., Beebe, N.W., 2008. Intraspecific concerted evolution of the rDNA ITS1 in *Anopheles farauti sensu stricto* (Diptera: Culicidae) reveals recent patterns of population structure. *J Mol Evol* 67, 397-411.
- Bradley, R.S., 1999. *Paleoclimatology: Reconstructing Climates of the Quaternary*. Academic Press, San Diego.
- Brenière, S.F., Pietrokovsky, S., Gastelum, E.M., Bosseno, M.F., Soto, M.M., Ouaiissi, A., Kasten, F.L., Wisnivesky-Colli, C., 2004. Feeding patterns of *Triatoma longipennis* Usinger (Hemiptera, Reduviidae) in peridomestic habitats of a rural community in Jalisco State, Mexico. *J Med Entomol* 41, 1015-1020.
- Brower, A.V.Z., 1994. Rapid morphological radiation and convergence in the butterfly, *Heliconius erato*, inferred from patterns of mitochondrial DNA evolution. *Proc Natl Acad Sci U S A* 91, 6491-6495.
- Buckler, E.S.t., Ippolito, A., Holtsford, T.P., 1997. The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* 145, 821-832.

- Bush, M.B., De Oliveira, P.E., Colinvaux, P.A., Millere, M.C., Moreno, J.E., 2004. Amazonian paleoecological histories: one hill, three watersheds. *Palaeogeography, Palaeoclimatology, Palaeoecology* 214, 359-393.
- Cabanne, G.S., d'Horta, F.M., Sari, E.H.R., Santos, F.R., Miyaki, C.Y., 2008. Nuclear and mitochondrial phylogeography of the Atlantic forest endemic *Xiphorhynchus fuscus* (Aves: Dendrocolaptidae): Biogeography and systematics implications. *Mol Phyl Evol* 49, 760-773.
- Calleros, L., Panzera, F., BARGUES, M.D., Monteiro, F.A., Klisiowicz, D.R., Zuriaga, M.A., Mas-Coma, S., Perez, R., 2010. Systematics of *Mepraia* (Hemiptera-Reduviidae): cytogenetic and molecular variation. *Infect Genet Evol* 10, 221-228.
- Ceballos, L.A., Piccinalli, R.V., Marcet, P.L., Vazquez-Prokopec, G.M., Cardinal, M.V., Schachter-Broide, J., Dujardin, J.P., Dotson, E.M., Kitron, U., Gurtler, R.E., 2011. Hidden sylvatic foci of the main vector of Chagas disease *Triatoma infestans*: threats to the vector elimination campaign? *PLoS Negl Trop Dis* 5, e1365.
- Cecere, M.C., Vasquez-Prokopec, G.M., Gurtler, R.E., Kitron, U., 2006. Reinfestation sources for Chagas disease vector, *Triatoma infestans*, Argentina. *Emerg Infect Dis* 12, 1096-1102.
- Ceriani, M.F., Hogenesch, J.B., Yanovsky, M., Panda, S., Straume, M., Kay, S.A., 2002. Genome-wide expression analysis in *Drosophila* reveals genes controlling circadian behavior. *J Neurosci* 22, 9305-9319.
- Chan, K.M., Levin, S.A., 2005. Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution* 59, 720-729.
- Chapuis, M.P., Streiff, R., Sword, G.A., 2012. Long microsatellites and unusually high levels of genetic diversity in the Orthoptera. *Insect Mol Biol* 21, 181-186.
- Christensen, H.A., de Vasquez, A.M., 1981. Host feeding profiles of *Rhodnius pallescens* (Hemiptera: Reduviidae) in rural villages of Central Panama. *The American journal of tropical medicine and hygiene* 30, 278-283.
- Clements, A.N., 1999. *The Biology of Mosquitoes: Sensory Reception and Behaviour*. CABI Publishing, New York.
- Colinvaux, P.A., De Oliveira, P.E., 2001. Amazon plant diversity and climate through the Cenozoic. *Palaeogeography, Palaeoclimatology, Palaeoecology* 166, 51-63.
- Conn, J.E., Mirabello, L., 2007. The biogeography and population genetics of neotropical vector species. *Heredity (Edinb)* 99, 245-256.
- Constantinou, C., 1979. *Circadian rhythms in insects*. University of London, London, p. 270.
- Constantinou, C., 1984. Circadian rhythm of oviposition in the blood sucking bugs, *Triatoma phyllosoma*, *T. infestans* and *Panstrongylus megistus* (Hemiptera: Reduviidae). *J Interdisc Cycle Res* 15, 203-212.
- Coura, J.R., Junqueira, A.C., 2012. Risks of endemicity, morbidity and perspectives regarding the control of Chagas disease in the Amazon Region. *Mem Inst Oswaldo Cruz* 107, 145-154.
- Coura, J.R., Junqueira, A.C., Fernandes, O., Valente, S.A., Miles, M.A., 2002. Emerging Chagas disease in Amazonian Brazil. *Trends Parasitol* 18, 171-176.
- Craw, R.C., 1978. Two biogeographical frameworks: implications for the biogeography of New Zealand. A review. *Tuatara: Journal of the Biological Society* 23, 81-108.
- Croizat, L., 1958. *Panbiogeography*. Publicado pelo próprio autor, Caracas.
- Croizat, L., Nelson, G., Rosen, D.E., 1974. Centres of origin and related concepts. *Systematic Zoology* 23, 265-287.
- Cunningham, C.W., Omland, K.E., Oakley, T.H., 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol Evol* 13, 361-366.
- Cyran, S.A., Buchsbaum, A.M., Reddy, K.L., Lin, M.C., Glossop, N.R., Hardin, P.E., Young, M.W., Storti, R.V., Blau, J., 2003. *vriille*, *Pdp1*, and *dClock* form a second feedback loop in the *Drosophila* circadian clock. *Cell* 112, 329-341.
- Darlington, T.K., Wager-Smith, K., Ceriani, M.F., Staknis, D., Gekakis, N., Steeves, T.D., Weitz, C.J., Takahashi, J.S., Kay, S.A., 1998. Closing the circadian loop: CLOCK-induced transcription of its own inhibitors per and tim. *Science* 280, 1599-1603.
- Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9, 772.
- Darwin, C., 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. John Murray, London.
- de la Rua, N., Stevens, L., Dorn, P.L., 2011. High genetic diversity in a single population of *Triatoma sanguisuga* (LeConte, 1855) inferred from two mitochondrial markers: Cytochrome b and 16S ribosomal DNA. *Infect Genet Evol* 11, 671-677.

- de Paula, A.S., Diotaiuti, L., Galvão, C., 2007. Systematics and biogeography of *Rhodniini* (Heteroptera: Reduviidae: Triatominae) based on 16S mitochondrial rDNA sequences. *J Biogeogr* 34, 699-712.
- De Queiroz, A.N., De Carvalho, O.A., 2008. Problems in the interpretation of Brazilian archaeofaunas: different contexts and the important role of taphonomy. *Quaternary International* 180, 75-89.
- Dias, J.C., 2009. Elimination of Chagas disease transmission: perspectives. *Mem Inst Oswaldo Cruz* 104 Suppl 1, 41-45.
- Dick, C.W., Roubik, D.W., Gruber, K.F., Bermingham, E., 2004. Long-distance gene flow and cross-Andean dispersal of lowland rainforest bees (Apidae: Euglossini) revealed by comparative mitochondrial DNA phylogeography. *Mol Ecol* 13, 3775-3785.
- Diotaiuti, L., Silveira, A.C., Elias, M., 1984. [Finding of *Rhodnius prolixus* Stal, 1859, in babassu palm trees]. *Rev Bras Malariol Doencas Trop* 36, 11-14.
- Donnelly, P., Tavaré, S., 1995. Coalescents and genealogical structure under neutrality. *Annu Rev Genet* 29, 401-421.
- Dorn, P.L., Calderon, C., Melgar, S., Moguel, B., Solorzano, E., Dumonteil, E., Rodas, A., de la Rua, N., Garnica, R., Monroy, C., 2009. Two distinct *Triatoma dimidiata* (Latreille, 1811) taxa are found in sympatry in Guatemala and Mexico. *PLoS Negl Trop Dis* 3, e393.
- Dover, G., 2000. Results may not fit well with current theories. *Nature* 408, 17.
- Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4, e88.
- Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161, 1307-1320.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7, 214.
- Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22, 1185-1192.
- Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29, 1969-1973.
- Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32, 1792-1797.
- Ellegren, H., 2004. Microsatellites: simple sequences with complex evolution. *Nat Rev Genet* 5, 435-445.
- Emery, P., So, W.V., Kaneko, M., Hall, J.C., Rosbash, M., 1998. CRY, a *Drosophila* clock and light-regulated cryptochrome, is a major contributor to circadian rhythm resetting and photosensitivity. *Cell* 95, 669-679.
- Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14, 2611-2620.
- Excoffier, L., Lischer, H.E., 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* 10, 564-567.
- Farris, J.S., 1977. Phylogenetic analysis under Dollo's law. *Syst Zool* 26, 77-88.
- Feliciangeli, M.D., Dujardin, J.P., Bastrenta, B., Mazzarri, M., Villegas, J., Flores, M., Munoz, M., 2002. Is *Rhodnius robustus* (Hemiptera: Reduviidae) responsible for Chagas disease transmission in Western Venezuela? *Trop Med Int Health* 7, 280-287.
- Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17, 368-376.
- Ferris, S.D., Sage, R.D., Prager, E.M., Ritte, U., Wilson, A.C., 1983. Mitochondrial DNA evolution in mice. *Genetics* 105, 681-721.
- Figueiredo, L.T., Moreli, M.L., Borges, A.A., de Figueiredo, G.G., Badra, S.J., Bisordi, I., Suzuki, A., Capria, S., Padula, P., 2009. Evaluation of an enzyme-linked immunosorbent assay based on Araraquara virus recombinant nucleocapsid protein. *The American journal of tropical medicine and hygiene* 81, 273-276.
- Fitzpatrick, S., Feliciangeli, M.D., Sanchez-Martin, M.J., Monteiro, F.A., Miles, M.A., 2008. Molecular genetics reveal that silvatic *Rhodnius prolixus* do colonise rural houses. *PLoS Negl Trop Dis* 2, e210.
- Fitzpatrick, S., Watts, P.C., Feliciangeli, M.D., Miles, M.A., Kemp, S.J., 2009. A panel of ten microsatellite loci for the Chagas disease vector *Rhodnius prolixus* (Hemiptera: Reduviidae). *Infect Genet Evol* 9, 206-209.
- Freitas, S.P., Lorosa, E.S., Rodrigues, D.C., Freitas, A.L., Goncalves, T.C., 2005. [Feeding patterns of *Triatoma pseudomaculata* in the state of Ceara, Brazil]. *Rev Saude Publica* 39, 27-32.
- Fu, Y.X., 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147, 915-925.

- Funk, W.C., Caldwell, J.P., Peden, C.E., Padial, J.M., De la Riva, I., al., e., 2007. Tests of biogeographic hypotheses for diversification in the Amazonian forest frog, *Physalaemus petersi*. *Mol Phyl Evol* 44, 825-837.
- Galvão, C., Carcavallo, R., Rocha, D.S., Jurberg, J., 2003. A checklist of the current valid species of the subfamily Triatominae Jeannel, 1919 (Hemiptera, Reduviidae) and their geographical distribution, with nomenclatural and taxonomic notes. *Zootaxa* 202, 1-36.
- Gamble, T., Simons, A.M., Colli, G.R., Vitt, L.J., 2008. Tertiary climate change and the diversification of the Amazonian gecko genus *Gonatodes* (Sphaerodactylidae, Squamata). *Mol Phyl Evol* 46, 269-277.
- Garcia, A.L., Carrasco, H.J., Schofield, C.J., Stothard, J.R., Frame, I.A., Valente, S.A., Miles, M.A., 1998. Random amplification of polymorphic DNA as a tool for taxonomic studies of triatomine bugs (Hemiptera: Reduviidae). *J Med Entomol* 35, 38-45.
- Garcia, B.A., Powell, J.R., 1998. Phylogeny of species of *Triatoma* (Hemiptera: Reduviidae) based on mitochondrial DNA sequences. *J Med Entomol* 35, 232-238.
- Garcia, B.A., Zheng, L.O., Perez de Rosas, A.R., Segura, E.L., 2004. Isolation and characterization of polymorphic microsatellite loci in the Chagas' disease vector *Triatoma infestans* (Hemiptera: Reduviidae). *Mol. Ecol. Notes* 4, 568-571.
- Gascon, C., Malcolm, J.R., Patton, J.L., da Silva, M.N., Bogart, J.P., Lougheed, S.C., Peres, C.A., Neckel, S., Boag, P.T., 2000. Riverine barriers and the geographic distribution of Amazonian species. *Proc Natl Acad Sci U S A* 97, 13672-13677.
- Gaunt, M.W., Miles, M.A., 2002. An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 19, 748-761.
- Gazave, E., Marques-Bonet, T., Fernando, O., Charlesworth, B., Navarro, A., 2007. Patterns and rates of intron divergence between humans and chimpanzees. *Genome Biol* 8, R21.
- Gentile, C., Meireles-Filho, A.C., Britto, C., Lima, J.B., Valle, D., Peixoto, A.A., 2006. Cloning and daily expression of the timeless gene in *Aedes aegypti* (Diptera: Culicidae). *Insect Biochem Mol Biol* 36, 878-884.
- Gentile, C., Rivas, G.B., Meireles-Filho, A.C., Lima, J.B., Peixoto, A.A., 2009. Circadian expression of clock genes in two mosquito disease vectors: *cry2* is different. *J Biol Rhythms* 24, 444-451.
- Geurgas, S.R., Rodrigues, M.T., Moritz, C., 2008. The genus *Coleodactylus* (Sphaerodactylinae, Gekkota) revisited: a molecular phylogenetic perspective. *Mol Phylogenet Evol* 49, 92-101.
- Giordano, R., Cortez, J.C., Paulk, S., Stevens, L., 2005. Genetic diversity of *Triatoma infestans* (Hemiptera: Reduviidae) in Chuquisaca, Bolivia based on the mitochondrial cytochrome b gene. *Mem Inst Oswaldo Cruz* 100, 753-760.
- Glaser, F.T., Stanewsky, R., 2007. Synchronization of the *Drosophila* circadian clock by temperature cycles. *Cold Spring Harb Symp Quant Biol* 72, 233-242.
- Glossop, N.R., Houl, J.H., Zheng, H., Ng, F.S., Dudek, S.M., Hardin, P.E., 2003. VRILLE feeds back to control circadian transcription of Clock in the *Drosophila* circadian oscillator. *Neuron* 37, 249-261.
- Goncalves da Silva, A., Cunha, I.C., Santos, W.S., Luz, S.L., Ribolla, P.E., Abad-Franch, F., 2012. Gene flow networks among American *Aedes aegypti* populations. *Evol Appl* 5, 664-676.
- Gonçalves da Silva, A., Cunha, I.C., Santos, W.S., Luz, S.L., Ribolla, P.E., Abad-Franch, F., 2012. Gene flow networks among American *Aedes aegypti* populations. *Evol Appl* 5, 664-676.
- Gourbière, S., Dorn, P., Tripet, F., Dumonteil, E., 2012. Genetics and evolution of triatomines: from phylogeny to vector control. *Heredity (Edinb)* 108, 190-202.
- Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711-732.
- Griffiths, R.C., Tavare, S., 1994. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci* 344, 403-410.
- Guhl, F., Pinto, N., Aguilera, G., 2009. Sylvatic triatominae: a new challenge in vector control transmission. *Mem Inst Oswaldo Cruz* 104 Suppl 1, 71-75.
- Haffer, J., 1969. Speciation in Amazonian forest birds. *Science* 165, 131-137.
- Haffer, J., Prance, G.T., 2002. Climatic forcing of evolution in Amazonia during the Cenozoic: on the refuge theory of biotic differentiation. *Estudos Avançados* 46, 175-208.
- Hall, J.C., 2003. Genetics and molecular biology of rhythms in *Drosophila* and other insects. *Adv Genet* 48, 1-280.
- Hall, J.P.W., Harvey, D.J., 2002. The phylogeography of Amazonia revisited: new evidence from Riodinid butterflies. *Evolution* 56, 1489-1497.
- Hardin, E., 1971. The probabilities of rooted tree-shapes generated by random bifurcation. *Adv in Appl Probab* 3, 44-77.
- Hardin, P.E., 2005. The circadian timekeeping system of *Drosophila*. *Curr Biol* 15, 714-722.

- Hardin, P.E., 2011. Molecular genetic analysis of circadian timekeeping in *Drosophila*. *Adv Genet* 74, 141-173.
- Harold, A.S., Mooi, R.D., 1994. Areas of endemism: definition and recognition criteria. *Syst Biol* 43, 261-266.
- Harris, D.J., Crandall, K.A., 2000. Intragenomic variation within ITS1 and ITS2 of freshwater crayfishes (Decapoda: Cambaridae): implications for phylogenetic and microsatellite studies. *Mol Biol Evol* 17, 284-291.
- Harry, M., 1993. Use of the median process of the pygophore in the identification of *Rhodnius nasutus*, *R. neglectus*, *R. prolixus* and *R. robustus* (Hemiptera: Reduviidae). *Ann Trop Med Parasitol* 87, 277-282.
- Harry, M., 1994. Morphometric variability in the Chagas' disease vector *Rhodnius prolixus*. *Jpn J Genet* 69, 233-250.
- Harry, M., Dupont, L., Quartier, M., Diotaiuti, L., Walter, A., Romana, C., 2009. New perspectives for population genetics of Chagas'disease vectors in the Northeastern Brazil: isolation of polymorphic microsatellite markers in *Triatoma brasiliensis*. *Infect Genet Evol* 9, 633-637.
- Harry, M., Dupont, L., Romana, C., Demanche, C., Mercier, A., Livet, A., Diotaiuti, L., Noireau, F., Emperaire, L., 2008a. Microsatellite markers in *Triatoma pseudomaculata* (Hemiptera, Reduviidae, Triatominae), Chagas' disease vector in Brazil. *Infect Genet Evol* 8, 672-675.
- Harry, M., Poyet, G., Romana, C.A., Solignac, M., 1998. Isolation and characterization of microsatellite markers in the bloodsucking bug *Rhodnius pallescens* (Heteroptera, Reduviidae). *Mol Ecol* 7, 1784-1786.
- Harry, M., Roose, C.L., Vautrin, D., Noireau, F., Romana, C.A., Solignac, M., 2008b. Microsatellite markers from the Chagas disease vector, *Rhodnius prolixus* (Hemiptera, Reduviidae), and their applicability to *Rhodnius* species. *Infect Genet Evol* 8, 381-385.
- Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22, 160-174.
- Hashimoto, K., Schofield, C.J., 2012. Elimination of *Rhodnius prolixus* in Central America. *Parasit Vectors* 5, 45.
- Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97-109.
- Hayes, F.E., Sewlal, J.A.N., 2004. The Amazon River as a dispersal barrier to passerine birds: effects of river width, habitat and taxonomy. *Journal of Biogeography* 31, 1809-1818.
- Heled, J., Drummond, A.J., 2008. Bayesian inference of population size history from multiple loci. *BMC Evol Biol* 8, 289.
- Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. *Mol Biol Evol* 27, 570-580.
- Hlaing, T., Tun-Lin, W., Somboon, P., Socheat, D., Setha, T., al., e., 2009. Mitochondrial pseudogenes in the nuclear genome of *Aedes aegypti* mosquitoes: implications for past and future population genetic studies. *BMC Genomics* 10, 11.
- Ho, S.Y., Phillips, M.J., Drummond, A.J., Cooper, A., 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Mol Biol Evol* 22, 1355-1363.
- Ho, S.Y., Saarma, U., Barnett, R., Haile, J., Shapiro, B., 2008. The effect of inappropriate calibration: three case studies in molecular ecology. *PLoS One* 3, e1615.
- Hooghiemstra, H., Van der Hammen, T., 1998. Neogene and Quaternary development of the neotropical rain forest: the forest refugia hypothesis, and a literature overview. *Earth- Science Reviews* 44, 147-183.
- Hoorn, C., 1993. Marine incursions and the influence of Andean tectonics on the Miocene depositional history of northwestern Amazonia: results of a palynostratigraphic study. *Palaeogeography, Palaeoclimatology, Palaeoecology* 105, 267-309.
- Hoorn, C., 1994. Fluvial palaeoenvironments in the intracratonic Amazonas Basin (Early Miocene – Early Middle Miocene, Colombia). *Palaeogeography, Palaeoclimatology, Palaeoecology* 109, 1-54.
- Hoorn, C., Wesselingh, F.P., ter Steege, H., Bermudez, M.A., Mora, A., Sevink, J., Sanmartin, I., Sanchez-Meseguer, A., Anderson, C.L., Figueiredo, J.P., Jaramillo, C., Riff, D., Negri, F.R., Hooghiemstra, H., Lundberg, J., Stadler, T., Sarkinen, T., Antonelli, A., 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* 330, 927-931.
- Howell, N., Kubacka, I., Mackey, D.A., 1996. How rapidly does the human mitochondrial genome evolve? *Am J Hum Genet* 59, 501-509.
- Huang, H., Knowles, L.L., 2009. What is the danger of the anomaly zone for empirical phylogenetics? *Syst Biol* 58, 527-536.
- Hubby, J.L., Lewontin, R.C., 1966. A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics* 54, 577-594.

- Hubisz, M.J., Falush, D., Stephens, M., Pritchard, J.K., 2009. Inferring weak population structure with the assistance of sample group information. *Mol Ecol Resour* 9, 1322-1332.
- Hudson, R.R., 1990. Gene genealogies and the coalescent process. *Oxf Surv Evol Biol* 7, 1-14.
- Hudson, R.R., Coyne, J.A., 2002. Mathematical consequences of the genealogical species concept. *Evolution* 56, 1557-1565.
- Hwang, W.S., Weirauch, C., 2012. Evolutionary history of assassin bugs (insecta: hemiptera: Reduviidae): insights from divergence dating and ancestral state reconstruction. *PLoS One* 7, e45523.
- Hypsa, V., Tietz, D.F., Zrzavy, J., Rego, R.O., Galvao, C., Jurberg, J., 2002. Phylogeny and biogeography of Triatominae (Hemiptera: Reduviidae): molecular evidence of a New World origin of the Asiatic clade. *Mol Phylogenet Evol* 23, 447-457.
- Irimia, M., Roy, S.W., 2008. Spliceosomal introns as tools for genomic and evolutionary analysis. *Nucleic Acids Res* 36, 1703-1712.
- Johnson, N.A., Lachance, J., 2012. The genetics of sex chromosomes: evolution and implications for hybrid incompatibility. *Ann N Y Acad Sci* 1256, E1-22.
- Jukes, T.H., Cantor, C.R., 1969. Evolution of protein molecules, in: Munro, H.N. (Ed.), *Mammalian protein metabolism*. Academic Press, New York, pp. 21-123.
- Justi, S.A., 2009. Sistemática molecular e importância epidemiológica de *Rhodnius* spp. (Hemiptera: Reduviidae) na Bolívia. , *Biologia Celular e Molecular*. Instituto Oswaldo Cruz, FIOCRUZ, Rio de Janeiro, p. 126pp.
- Justi, S.A., Noireau, F., Cortez, M.R., Monteiro, F.A., 2010. Infestation of peridomestic *Attalea phalerata* palms by *Rhodnius stali*, a vector of *Trypanosoma cruzi* in the Alto Beni, Bolivia. *Trop Med Int Health* 15, 727-732.
- Kass, R.E., Raftery, A.E., 1995. Bayes Factors. *J Am Stat Assoc* 90, 773-795.
- Keegan, K.P., Pradhan, S., Wang, J.P., Allada, R., 2007. Meta-analysis of *Drosophila* circadian microarray studies identifies a novel set of rhythmically expressed genes. *PLoS Comput Biol* 3, e208.
- Kimura, M., 1968. Evolutionary rate at the molecular level. *Nature* 217, 624-626.
- Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16, 111-120.
- Kimura, M., Crow, J.F., 1964. The number of alleles that can be maintained in a finite population. *Genetics* 49, 725-738.
- Kimura, M., Ohta, T., 1978. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc Natl Acad Sci U S A* 75, 2868-2872.
- Kingman, J.F.C., 1982. The coalescent. *Stoch Proc Appl* 13, 235-248.
- Klammer, G., 1984. The relief of the extra-Andean Amazon Basin, in: Sioli, H. (Ed.), *The Amazon: Limnology and landscape ecology of a mighty tropical river and its basin*. Dr. W. Junk Publishers, Dordrecht, pp. 47-83.
- Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I., Khorana, H.G., 1971. Studies on polynucleotides: XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. *J Mol Bio* 56, 341-361.
- Kline, K., McCarthy, J.S., Pearson, M., Loukas, A., Hotez, P.J., 2013. Neglected tropical diseases of Oceania: review of their prevalence, distribution, and opportunities for control. *PLoS Negl Trop Dis* 7, e1755.
- Klowden, M.J., Zwiebel, L., 2004. Vector Olfaction and Behavior, in: Marquardt, W.C., al., e. (Eds.), *Biology of Disease Vectors*, 2nd edition ed. Elsevier Academic Press, Amsterdam.
- Kluge, A.G., Farris, J.S., 1969. Quantitative phyletics and the evolution of anurans. *Syst Zool* 18, 1-32.
- Knowles, L.L., Carstens, B.C., 2007. Delimiting species without monophyletic gene trees. *Syst Biol* 56, 887-895.
- Krishnan, B., Levine, J.D., Lynch, M.K., Dowse, H.B., Funes, P., Hall, J.C., Hardin, P.E., Dryer, S.E., 2001. A new role for cryptochrome in a *Drosophila* circadian oscillator. *Nature* 411, 313-317.
- Kuhner, M.K., 2009. Coalescent genealogy samplers: windows into population history. *Trends Ecol Evol* 24, 86-93.
- Kumar, S., 2005. Molecular clocks: four decades of evolution. *Nat Rev Genet* 6, 654-662.
- Kyriacou, C.P., Hall, J.C., 1986. Interspecific genetic control of courtship song production and reception in *Drosophila*. *Science* 232, 494-497.
- Latrubesse, E.M., Da Silva, S.A.F., Cozzuol, M., Absy, M.L., 2007. Late Miocene continental sedimentation in southwestern Amazonia and its regional significance: biotic and geological evidence. *Journal of South American Earth Sciences* 23, 61-80.
- Lazzari, C.R., 1991. Circadian rhythm of egg hatching in *Triatoma infestans* (Hemiptera: Reduviidae). *J Med Entomol* 28, 740-741.
- Lazzari, C.R., 1992. Circadian organization of locomotion activity in the haematophagous bug *Triatoma infestans*. *Journal of Insect Physiology* 38, 895-903.

- Lazzari, C.R., Lorenzo, M.G., 2009. Exploiting triatomine behaviour: alternative perspectives for their control. *Mem Inst Oswaldo Cruz* 104 Suppl 1, 65-70.
- Lemey, P., Rambaut, A., Drummond, A.J., Suchard, M.A., 2009. Bayesian phylogeography finds its roots. *PLoS Comput Biol* 5, e1000520.
- Leng, L., Zhang, D.E., 2011. Measuring population differentiation using GST or D? A simulation study with microsatellite DNA markers under a finite island model and nonequilibrium conditions. *Mol Ecol* 20, 2494-2509.
- Lent, H., Wygodzinsky, P., 1979. Revision of the Triatominae (Hemiptera: Reduviidae) and their significance as vectors of Chagas disease. *Bulletin of the American Museum of Natural History* 163, 123-520.
- Li, W.L., Drummond, A.J., 2012. Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics. *Mol Biol Evol* 29, 751-761.
- Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451-1452.
- Liò, P., Goldman, N., 1998. Models of molecular evolution and phylogeny. *Genome Research* 8, 1233-1244.
- Lomb, N.R., 1976. Least-squares frequency analysis of unequally spaced data. *ApS Sci* 39, 447-462.
- Lorenzo, M.G., Lazzari, C.R., 1998. Activity pattern in relation to refuge exploitation and feeding in *Triatoma infestans* (Hemiptera: Reduviidae). *Acta Trop* 70, 163-170.
- Lovejoy, N.R., Albert, J.S., Crampton, W.G.R., 2006. Miocene marine incursions and marine/freshwater transitions: Evidence from Neotropical fishes. *Journal of South American Earth Sciences* 21, 5-13.
- Lyman, D.F., Monteiro, F.A., Escalante, A.A., Cordon-Rosales, C., Wesson, D.M., Dujardin, J.P., Beard, C.B., 1999. Mitochondrial DNA sequence variation among triatomine vectors of Chagas' disease. *The American journal of tropical medicine and hygiene* 60, 377-386.
- Lyons, L.C., Roman, G., 2009. Circadian modulation of short-term memory in *Drosophila*. *Learn Mem* 16, 19-27.
- Mallet, J., 2005. Hybridization as an invasion of the genome. *Trends Ecol Evol* 20, 229-237.
- Mallet, J., 2007. Hybrid speciation. *Nature* 446, 279-283.
- Marcet, P.L., Lehmann, T., Groner, G., Gurtler, R.E., Kitron, U., Dotson, E.M., 2006. Identification and characterization of microsatellite markers in the Chagas disease vector *Triatoma infestans* (Heteroptera: Reduviidae). *Infect Genet Evol* 6, 32-37.
- Marcilla, A., Bargues, M.D., Abad-Franch, F., Panzera, F., Carcavallo, R.U., Noireau, F., Galvao, C., Jurberg, J., Miles, M.A., Dujardin, J.P., Mas-Coma, S., 2002. Nuclear rDNA ITS-2 sequences reveal polyphyly of *Panstrongylus* species (Hemiptera: Reduviidae: Triatominae), vectors of *Trypanosoma cruzi*. *Infect Genet Evol* 1, 225-235.
- Marcilla, A., Bargues, M.D., Ramsey, J.M., Magallon-Gastelum, E., Salazar-Schettino, P.M., Abad-Franch, F., Dujardin, J.P., Schofield, C.J., Mas-Coma, S., 2001. The ITS-2 of the nuclear rDNA as a molecular marker for populations, species, and phylogenetic relationships in Triatominae (Hemiptera: Reduviidae), vectors of Chagas disease. *Mol Phylogenet Evol* 18, 136-142.
- Margoliash, E., 1963. Primary structure and evolution of cytochrome c. *Proc Natl Acad Sci USA* 50, 672-679.
- Marques, N., Menna-Barreto, L., 2003. *Cronobiologia: Princípios e Aplicações.*, 3rd ed. Edusp, São Paulo.
- Márquez, E., Jaramillo, O.N., Gomez-Palacio, A., Dujardin, J.P., 2011. Morphometric and molecular differentiation of a *Rhodnius robustus*-like form from *R. robustus* Larousse, 1927 and *R. prolixus* Stal, 1859 (Hemiptera, Reduviidae). *Acta Trop* 120, 103-109.
- Marroig, G., Cerqueira, R., 1997. Plio-Pleistocene South American history and the Amazon lagoon hypothesis: a piece of the puzzle of Amazonian diversification. *J Comp Biol* 2, 103-119.
- Martinez, F.H., Villalobos, G.C., Cevallos, A.M., Torre Pde, L., Laclette, J.P., Alejandro-Aguilar, R., Espinoza, B., 2006. Taxonomic study of the *Phyllosoma* complex and other triatomine (Insecta: Hemiptera: Reduviidae) species of epidemiological importance in the transmission of Chagas disease: using ITS-2 and mtCytB sequences. *Mol Phylogenet Evol* 41, 279-287.
- Mas-Coma, S., Bargues, M.D., 2009. Populations, hybrids and the systematic concepts of species and subspecies in Chagas disease triatomine vectors inferred from nuclear ribosomal and mitochondrial DNA. *Acta Trop* 110, 112-136.
- Massey, F.J., 1951. The Kolmogorov-Smirnov Test for goodness of fit. *J Am Stat Assoc* 46, 68-78.
- Massingham, T., Goldman, N., 2007. Statistics of the Log-Det estimator. *Mol Biol Evol* 24, 2277-2285.
- Mayle, F.E., Beerling, D.J., Gosling, W.D., Bush, M.B., 2004. Responses of Amazonian ecosystems to climatic and atmospheric carbon dioxide changes since the last glacial maximum. *Philos Trans R Soc Lond B Biol Sci* 359, 499-514.

- Mayr, E., 1969. Principles of Systematic Zoology. McGraw Hill, New York.
- Mazzoni, C.J., Souza, N.A., Andrade-Coelho, C., Kyriacou, C.P., Peixoto, A.A., 2006. Molecular polymorphism, differentiation and introgression in the period gene between *Lutzomyia intermedia* and *Lutzomyia whitmani*. BMC Evol Biol 6, 85.
- McDonald, M.J., Rosbash, M., 2001. Microarray analysis and organization of circadian gene expression in *Drosophila*. Cell 107, 567-578.
- McEwen, P.K., Lehane, M.J., 1993. Factors influencing flight initiation in the triatomine bug *Triatoma infestans* (Hemiptera: Reduviidae). Insect Sci Applications 14, 461-464.
- Meireles-Filho, A.C., Amoretty, P.R., Souza, N.A., Kyriacou, C.P., Peixoto, A.A., 2006a. Rhythmic expression of the cycle gene in a hematophagous insect vector. BMC Mol Biol 7, 38.
- Meireles-Filho, A.C., da S.R.G.B., Gesto, J.S., Machado, R.C., Britto, C., de Souza, N.A., Peixoto, A.A., 2006b. The biological clock of an hematophagous insect: locomotor activity rhythms, circadian expression and downregulation after a blood meal. FEBS Lett 580, 2-8.
- Mendonça, V.J., da Silva, M.T., de Araujo, R.F., Junior, J.M., Junior, M.B., Almeida, C.E., Costa, J., Graminha, M.A., Cicarelli, R.M., da Rosa, J.A., 2009. Phylogeny of *Triatoma sherlocki* (Hemiptera: Reduviidae: Triatominae) inferred from two mitochondrial genes suggests its location within the *Triatoma brasiliensis* complex. The American journal of tropical medicine and hygiene 81, 858-864.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equations of state calculations by fast computing machines. J Chem Phys 21, 1087-1091.
- Mishler, B., Brandon, R., 1987. Individuality, Pluralism, and the Phylogenetic Species Concept. Biol Phil 2, 37-54.
- Monteiro, F.A., Barrett, T.V., Fitzpatrick, S., Cordon-Rosales, C., Feliciangeli, D., Beard, C.B., 2003. Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. Mol Ecol 12, 997-1006.
- Monteiro, F.A., Donnelly, M.J., Beard, C.B., Costa, J., 2004. Nested clade and phylogeographic analyses of the Chagas disease vector *Triatoma brasiliensis* in Northeast Brazil. Mol Phylogenet Evol 32, 46-56.
- Monteiro, F.A., Jurberg, J., Lazoski, C., 2009. Very low levels of genetic variation in natural peridomestic populations of the Chagas disease vector *Triatoma sordida* (Hemiptera: Reduviidae) in southeastern Brazil. The American journal of tropical medicine and hygiene 81, 223-227.
- Monteiro, F.A., Lazoski, C., Noireau, F., Sole-Cava, A.M., 2002. Allozyme relationships among ten species of *Rhodniini*, showing paraphyly of *Rhodnius* including *Psammolestes*. Med Vet Entomol 16, 83-90.
- Morrone, J.J., 2006. Biogeographic areas and transition zones of Latin America and the Caribbean islands based on panbiogeographic and cladistic analyses of the entomofauna. Annu Rev Entomol 51, 467-494.
- Morrone, J.J., Crisci, J.V., 1995. Historical biogeography - Introduction to methods. Annual Review of Ecology and Systematics 26, 373-401.
- Myers, E.M., 2003. The circadian control of eclosion. Chronobiol Int 20, 775-794.
- Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York.
- Nei, M., Li, W.H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci U S A 76, 5269-5273.
- Nicholls, G.K., Gray, R.D., 2008. Dated ancestral trees from binary trait data and their application to the diversification of languages. Journal of the Royal Statistical Society Series B 70, 545-566.
- Nores, M., 1999. An alternative hypothesis for the origin of Amazonian bird diversity. Journal of Biogeography 26, 475-485.
- Ohshima, I., Yoshizawa, K., 2011. The utility of indels in population genetics: the Tpi intron for host race genealogy of *Acrocercops transecta* (Insecta: Lepidoptera). Mol Phylogenet Evol 59, 469-476.
- OMS, 2007. Global plan to combat neglected tropical diseases 2008–2015., Geneva.
- OMS, 2010. First WHO report on neglected tropical diseases: working to overcome the global impact of neglected tropical diseases, Geneva.
- Opazo, J.C., Wildman, D.E., Prychitko, T., Johnson, R.M., Goodman, M., 2006. Phylogenetic relationships and divergence times among New World monkeys (Platyrrhini, Primates). Mol Phylogenet Evol 40, 274-280.
- Ortiz-Barrientos, D., Counterman, B.A., Noor, M.A., 2004. The genetics of speciation by reinforcement. PLoS Biol 2, e416.
- Pacheco, R.S., Almeida, C.E., Klisiowicz, D.R., Costa, J., Pires, M.Q., Panzera, F., Bar, M.E., Mas-Coma, S., Bargues, M.E., 2007. Genetic variability of *Triatoma rubrovaria* (Reduviidae: Triatominae) from Brazil, Argentina and Uruguay as revealed by two different molecular markers. Parasite 14, 231-237.
- Pagel, M., 1999. Inferring the historical patterns of biological evolution. Nature 401, 877-884.

- Patel, S., Weckstein, J.D., Patane, J.S., Bates, J.M., Aleixo, A., 2011. Temporal and spatial diversification of *Pteroglossus aracaris* (AVES: Ramphastidae) in the neotropics: constant rate of diversification does not support an increase in radiation during the Pleistocene. *Mol Phylogenet Evol* 58, 105-115.
- Pates, H., Curtis, C., 2005. Mosquito behavior and vector control. *Annu Rev Entomol* 50, 53-70.
- Patterson, J.S., Gaunt, M.W., 2010. Phylogenetic multi-locus codon models and molecular clocks reveal the monophyly of haematophagous reduviid bugs and their evolution at the formation of South America. *Mol Phylogenet Evol* 56, 608-621.
- Patton, J.L., Da Silva, M.F., Malcolm, J.R., 1994. Gene genealogy and differentiation among arboreal spiny rats (Rodentia: Echimyidae) of the Amazon basin: a test of the riverine barrier hypothesis. *Evolution* 48, 1314-1323.
- Patton, J.L., Da Silva, M.N.F., 2005. The history of Amazonian mammals: mechanisms and timing of diversification, in: Bermingham, E., Dick, C.W., Moritz, C. (Eds.), *Tropical Rainforest: Past, Present, and Future*. The University of Chicago, Chicago, pp. 107-126.
- Patton, J.L., Da Silva, M.N.F., Malcolm, J.R., 2000. Mammals of the Rio Juruá and the evolutionary and ecological diversification of Amazonia. *Bulletin of the American Museum of Natural History* 244, 1-306.
- Pavan, M.G., 2009. Filogeografia de *Rhodnius pictipes* (Hemiptera: Reduviidae) na região amazônica, *Biologia Parasitária*. Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, p. 127pp.
- Pavan, M.G., Mesquita, R.D., Lawrence, G.G., Lazoski, C., Dotson, E.M., Abubucker, S., Mitreva, M., Randall-Maher, J., Monteiro, F.A., 2013. A nuclear single-nucleotide polymorphism (SNP) potentially useful for the separation of *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). *Infect Genet Evol* 14, 426-433.
- Pavan, M.G., Monteiro, F.A., 2007. A multiplex PCR assay that separates *Rhodnius prolixus* from members of the *Rhodnius robustus* cryptic species complex (Hemiptera: Reduviidae). *Trop Med Int Health* 12, 751-758.
- Peakall, R., Smouse, P.E., 2012. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* 28, 2537-2539.
- Peixoto, A.A., 2002. Evolutionary behavioral genetics in *Drosophila*. *Advances in Genetics* 47, 117-150.
- Pfeiler, E., Bitler, B.G., Ramsey, J.M., Palacios-Cardiel, C., Markow, T.A., 2006. Genetic variation, population structure, and phylogenetic relationships of *Triatoma rubida* and *T. recurva* (Hemiptera: Reduviidae: Triatominae) from the Sonoran Desert, insect vectors of the Chagas' disease parasite *Trypanosoma cruzi*. *Mol Phylogenet Evol* 41, 209-221.
- Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., al., e., 2011. Resolving Difficult Phylogenetic Questions: Why More Sequences Are Not Enough. *PLoS Biol* 9, e1000602.
- Piccinali, R.V., Marcet, P.L., Ceballos, L.A., Kitron, U., Gurtler, R.E., Dotson, E.M., 2011. Genetic variability, phylogenetic relationships and gene flow in *Triatoma infestans* dark morphs from the Argentinean Chaco. *Infect Genet Evol* 11, 895-903.
- Pittendrigh, C.S., 1958. On the significance of transients in daily rhythms. *Proc Natl Acad Sci U S A* 44, 965-973.
- Platnick, N.I., Nelson, G., 1988. Spanning-tree biogeography - shortcut, detour, or dead-end. *Systematic Zoology* 37, 410-419.
- Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. *Genetics* 155, 945-959.
- Pybus, O.G., 2006. Model selection and the molecular clock. *PLoS Biol* 4, e151.
- Pybus, O.G., Rambaut, A., 2002. GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics* 18, 1404-1405.
- Pybus, O.G., Rambaut, A., Harvey, P.H., 2000. An integrated framework for the inference of viral population history from reconstructed genealogies. *Genetics* 155, 1429-1437.
- Quisberth, S., Waleckx, E., Monje, M., Chang, B., Noireau, F., Breniere, S.F., 2011. "Andean" and "non-Andean" ITS-2 and mtCytB haplotypes of *Triatoma infestans* are observed in the Gran Chaco (Bolivia): population genetics and the origin of reinfestation. *Infect Genet Evol* 11, 1006-1014.
- Räsänen, M.E., Linna, A.M., Santos, J.C., Negri, F.R., 1995. Late miocene tidal deposits in the amazonian foreland basin. *Science* 269, 386-390.
- Rassi, A., Jr., Rassi, A., Marin-Neto, J.A., 2010. Chagas disease. *Lancet* 375, 1388-1402.
- Ribas, C.C., Aleixo, A., Nogueira, A.C., Miyaki, C.Y., Cracraft, J., 2012. A palaeobiogeographic model for biotic diversification within Amazonia over the past three million years. *Proc Biol Sci* 279, 681-689.
- Ribeiro, J.M., Hazzard, J.M., Nussenzveig, R.H., Champagne, D.E., Walker, F.A., 1993. Reversible binding of nitric oxide by a salivary heme protein from a bloodsucking insect. *Science* 260, 539-541.

- Rokas, A., Ladoukakis, E., Zouros, E., 2003. Animal mitochondrial DNA recombination revisited. *TRENDS Ecol Evol* 18, 411-417.
- Rona, L.D., Carvalho-Pinto, C.J., Gentile, C., Grisard, E.C., Peixoto, A.A., 2009. Assessing the molecular divergence between *Anopheles (Kerteszia) cruzii* populations from Brazil using the timeless gene: further evidence of a species complex. *Malar J* 8, 60.
- Rona, L.D., Carvalho-Pinto, C.J., Mazzoni, C.J., Peixoto, A.A., 2010. Estimation of divergence time between two sibling species of the *Anopheles (Kerteszia) cruzii* complex using a multilocus approach. *BMC Evol Biol* 10, 91.
- Ronquist, F., 2004. Bayesian inference of character evolution. *Trends in Ecology and Evolution* 19, 475-481.
- Rosen, B.R., 1988. Biogeographic patterns: a perceptual review, in: Myers, A.A., Giller, P.S. (Eds.), *Analytical Biogeography: an integrated approach to the study of animal and plant distributions*. Chapman and Hall, London & New York, pp. 23-55.
- Rubin, E.B., Shemesh, Y., Cohen, M., Elgavish, S., Robertson, H.M., Bloch, G., 2006. Molecular and phylogenetic analyses reveal mammalian-like clockwork in the honey bee (*Apis mellifera*) and shed new light on the molecular evolution of the circadian clock. *Genome Res* 16, 1352-1365.
- Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B., Erlich, H.A., 1986. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. *Nature* 324, 163-166.
- Saitou, N., Nei, M., 1987. The Neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425.
- Sakai, T., Ishida, N., 2001. Circadian rhythms of female mating activity governed by clock genes in *Drosophila*. *Proc Natl Acad Sci U S A* 98, 9221-9225.
- Salvatella, R., Calegari, L., Puime, A., Basmadjian, Y., Rosa, R., Guerrero, J., Martinez, M., Mendaro, G., Briano, D., Montero, C., et al., 1994. [Feeding pattern of *Triatoma rubrovaria* (Blanchard, 1843) (Hemiptera, Triatominae) in peridomestic habitats, of a rural area of Uruguay]. *Rev Inst Med Trop Sao Paulo* 36, 311-320.
- Sandoval, C.M., Duarte, R., Gutierrez, R., Rocha Dda, S., Angulo, V.M., Esteban, L., Reyes, M., Jurberg, J., Galvao, C., 2004. Feeding sources and natural infection of *Belminus herreri* (Hemiptera, Reduviidae, Triatominae) from dwellings in Cesar, Colombia. *Mem Inst Oswaldo Cruz* 99, 137-140.
- Sarov-Blat, L., So, W.V., Liu, L., Rosbash, M., 2000. The *Drosophila* takeout gene is a novel molecular link between circadian rhythms and feeding behavior. *Cell* 101, 647-656.
- Saunders, D.S., 2002. *Insect Clocks*, 3rd ed. Elsevier Science, Amsterdam.
- Scargle, J.D., 1982. Studies in astronomical time series analysis. II - Statistical aspects of spectral analysis of unevenly spaced data. *ApJ* 263, 835-853.
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., al., e., 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods* 9, 676-682.
- Schmid, B., Helfrich-Föster, C., Yoshii, T., 2011. A new image J plug-in "ActogramJ" for chronobiological analyses. *J Biol Rhythms* 26, 464-467.
- Schofield, C.J., 1988. *Biosystematics of Haematophagous Insects, Biosystematics of the Triatominae*. Clarendon Press, London, pp. 285-312.
- Sehadova, H., Glaser, F.T., Gentile, C., Simoni, A., Giesecke, A., Albert, J.T., Stanewsky, R., 2009. Temperature entrainment of *Drosophila*'s circadian clock involves the gene *nocte* and signaling from peripheral sensory tissues to the brain. *Neuron* 64, 251-266.
- Senior, K., 2007. Chagas disease: moving towards global elimination. *Lancet Infect Dis* 7, 572.
- Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 591-611.
- Soares, R.P., Gontijo, N.F., Romanha, A.J., Diotaiuti, L., Pereira, M.H., 1998. Salivary heme proteins distinguish *Rhodnius prolixus* from *Rhodnius robustus* (Hemiptera: Reduviidae: Triatominae). *Acta Trop* 71, 285-291.
- Sota, T., Vogler, A.P., 2001. Incongruence of mitochondrial and nuclear gene trees in the Carabid beetles *Ohomopterus*. *Syst Biol* 50, 39-59.
- Stanewsky, R., Kaneko, M., Emery, P., Beretta, B., Wager-Smith, K., Kay, S.A., Rosbash, M., Hall, J.C., 1998. The *cryb* mutation identifies cryptochrome as a circadian photoreceptor in *Drosophila*. *Cell* 95, 681-692.
- Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18, 1001-1013.
- Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10, 512-526.

- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28, 2731-2739.
- Tauber, E., Kyriacou, C.P., 2005. Molecular evolution and population genetics of circadian clock genes. *Methods Enzymol* 393, 797-817.
- Tauber, E., Roe, H., Costa, R., Hennessy, J.M., Kyriacou, C.P., 2003. Temporal mating isolation driven by a behavioral gene in *Drosophila*. *Curr Biol* 13, 140-145.
- Townsend, R.H.D., 2010. Fast calculation of the Lomb-Scargle periodogram using graphics processing units. *ApJ Supp* 191, 247.
- Turcq, B., Cordeiro, R.C., Sifeddine, A., Simoes Filho, F.F., Albuquerque, A.L., Abrao, J.J., 2002. Carbon storage in Amazonia during the last glacial maximum: secondary data and uncertainties. *Chemosphere* 49, 821-835.
- Uemura, H., Tomioka, K., 2006. Postembryonic changes in circadian photo-responsiveness rhythms of optic lobe interneurons in the cricket *Gryllus bimaculatus*. *J Biol Rhythms* 21, 279-289.
- UNEP, 2004. Amazon Basin, GIWA Regional Assessment 40b, in: Kalmar, U.o. (Ed.). Barthem, R. B.Charvet-Almeida, P.Montag, L. F. A., Lanna, A. E., Kalmar, Sweden.
- Vafopoulou, X., Terry, K.L., Steel, C.G., 2010. The circadian timing system in the brain of the fifth larval instar of *Rhodnius prolixus* (hemiptera). *J Comp Neurol* 518, 1264-1282.
- Valente, S.A.S., Valente, V.C., Fraiha-Neto, H., 1999. Considerations on the epidemiology and transmission of Chagas disease in the Brazilian Amazon. *Mem Inst Oswaldo Cruz* 94, 395-398.
- Van der Hammen, T., Hooghiemstra, H., 2000. Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. *Quaternary Science Reviews* 19: 725-742. 19, 725-742.
- Van Oosterhout, C., Hutchinson, W., Wills, D.P.M., Shipley, P., 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* 4, 535-538.
- Vanin, S., Bhutani, S., Montelli, S., Menegazzi, P., Green, E.W., Pegoraro, M., Sandrelli, F., Costa, R., Kyriacou, C.P., 2012. Unexpected features of *Drosophila* circadian behavioural rhythms under natural conditions. *Nature* 484, 371-375.
- Vanzolini, P.E., Williams, E.E., 1970. South American anoles: Geographic differentiation and evolution of the *Anolis cloyolopsis* species group (Suaria, Iguanidae). *Arquivos de Zoologia* 19, 1-298.
- Vazquez-Prokopec, G.M., Ceballos, L.A., Kitron, U., Gurtler, R.E., 2004. Active dispersal of natural populations of *Triatoma infestans* (Hemiptera: Reduviidae) in rural northwestern Argentina. *J Med Entomol* 41, 614-621.
- Vigoder, F.M., Araki, A.S., Bauzer, L.G., Souza, N.A., Brazil, R.P., Peixoto, A.A., 2010. Lovesongs and period gene polymorphisms indicate *Lutzomyia cruzi* (Mangabeira, 1938) as a sibling species of the *Lutzomyia longipalpis* (Lutz and Neiva, 1912) complex. *Infect Genet Evol* 10, 734-739.
- Villalobos, G., Martinez-Hernandez, F., de la Torre, P., Laclette, J.P., Espinoza, B., 2011. Entomological indices, feeding sources, and molecular identification of *Triatoma phyllosoma* (Hemiptera: Reduviidae) one of the main vectors of Chagas disease in the Istmo de Tehuantepec, Oaxaca, Mexico. *The American journal of tropical medicine and hygiene* 85, 490-497.
- Villegas, J., Feliciangeli, M.D., Dujardin, J.P., 2002. Wing shape divergence between *Rhodnius prolixus* from Cojedes (Venezuela) and *Rhodnius robustus* from Merida (Venezuela). *Infect Genet Evol* 2, 121-128.
- Waleckx, E., Salas, R., Huaman, N., Buitrago, R., Bosseno, M.F., Aliaga, C., Barnabe, C., Rodriguez, R., Zoveda, F., Monje, M., Baune, M., Quisberth, S., Villena, E., Kengne, P., Noireau, F., Breniere, S.F., 2011. New insights on the Chagas disease main vector *Triatoma infestans* (Reduviidae, Triatominae) brought by the genetic analysis of Bolivian sylvatic populations. *Infect Genet Evol* 11, 1045-1057.
- Wallace, A.R., 1852. On the monkeys of the Amazon. *Proc Zool Soc London* 20, 107-110.
- Warnow, T., 2012. Standard maximum likelihood analyses of alignments with gaps can be statistically inconsistent. *PLoS Curr* 4, RRN1308.
- Weirauch, C., 2008. Cladistic analysis of Reduviidae (Heteroptera: Cimicomorpha) based on morphological characters. *Systematic Entomology* 33, 229-274.
- Weirauch, C., Munro, J.B., 2009. Molecular phylogeny of the assassin bugs (Hemiptera: Reduviidae), based on mitochondrial and nuclear ribosomal genes. *Mol Phylogenet Evol* 53, 287-299.
- Wesselingh, F.P., Hoorn, C., Guerrero, J., Räsänen, M.E., Romero-Pittmann, L., Salo, J., 2006. The stratigraphy and regional structure of Miocene deposits in western Amazonia (Peru, Colombia and Brazil), with implications for late Neogene landscape evolution. *Scripta Geologica* 133, 291-322.

- Wheeler, D.A., Kyriacou, C.P., Greenacre, M.L., Yu, Q., Rutila, J.E., Rosbash, M., Hall, J.C., 1991. Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. *Science* 251, 1082-1085.
- Wiens, J.J., Kukzyski, C.A., Stephens, P.R., 2010. Discordant mitochondrial and nuclear gene phylogenies in emydid turtles: implications for speciation and conservation. *Biol J Linn Soc* 99, 445-461.
- Wigglesworth, V., 1957. The physiology of insect cuticle. *Annual Reviews in Entomology* 2, 37-54.
- Wiley, E.O., 1988. Parsimony Analysis and vicariance biogeography. *Syst Biol* 37, 271-290.
- Yoder, A.D., Young, Z.H., 2000. Estimation of primate speciation dates using local molecular clocks. *Mol Biol Evol* 17, 1081-1090.
- Yu, W., Zheng, H., Price, J.L., Hardin, P.E., 2009. DOUBLETIME plays a noncatalytic role to mediate CLOCK phosphorylation and repress CLOCK-dependent transcription within the *Drosophila* circadian clock. *Mol Cell Biol* 29, 1452-1458.
- Yuan, Q., Metterville, D., Briscoe, A.D., Reppert, S.M., 2007. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. *Mol Biol Evol* 24, 948-955.
- Zhu, H., Yuan, Q., Briscoe, A.D., Froy, O., Casselman, A., Reppert, S.M., 2005. The two CRYs of the butterfly. *Curr Biol* 15, R953-954.
- Zuckermandl, E., Pauling, L., 1962. in: Kasha, M., Pullman, B. (Eds.), *Horizons in Biochemistry*. Academic Press, New York.