

UNIVERSIDADE ESTADUAL PAULISTA
“JULIO DE MESQUITA FILHO”
INSTITUTO DE BIOCIÊNCIAS
CAMPUS DE BOTUCATU

**Heterogeneidade genética de *Anopheles darlingi* e
suas implicações para epidemiologia da malária**

MELINA AULINO CAMPOS DE LIMA

Tese apresentada ao Curso de Pós-
graduação em Ciências Biológicas,
Área de Concentração Genética do
Instituto de Biociências de Botucatu,
Universidade Estadual Paulista –
UNESP.

BOTUCATU- SP
2016

MELINA AULINO CAMPOS DE LIMA

Bacharel em Ciências Biomédicas

**Heterogeneidade genética de *Anopheles darlingi* e
suas implicações para epidemiologia da malária**

Orientador: Prof. Dr. Paulo Eduardo Martins
Ribolla

Tese apresentada ao Curso de Pós-
graduação em Ciências Biológicas,
Área de Concentração Genética do
Instituto de Biociências de Botucatu,
Universidade Estadual Paulista –
UNESP.

BOTUCATU- SP
2016

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Lima, Melina Aulino Campos de.

Heterogeneidade genética de *Anopheles darlingi* e suas implicações para epidemiologia da malária / Melina Aulino Campos de Lima. - Botucatu, 2016

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu
Orientador: Paulo Eduardo Martins Ribolla
Capes: 20202008

1. *Anopheles*. 2. Polimorfismo de nucleotídeo único. 3. Genética de populações. 4. Malária. 5. Microssatélites (Genética).

Palavras-chave: *Anopheles darlingi*; SNPs; genética de população; malária; microssatélites.

“A ciência nos traz conhecimento;
a vida, sabedoria.”

Will Durant

Agradeço e celebro primeiramente a vida, a oportunidade de existir e ser;
a chance de ter conhecido pessoas incríveis que me ajudaram a crescer
e amadurecer, em todos os sentidos, ao longo dos quatro anos de
doutorado.

Agradeço ao Paulo Ribolla, meu orientador e personagem da
minha primeira inspiração no mundo científico; e a todo o
grupo Pangene, em especial ao Diego Alonso.

Agradeço ao apoio e oportunidades oferecidas por pesquisadores que, para mim, são
grandes exemplos: Maria Anice Sallum, João Pinto, Jan Conn e Kevin Emerson.
Pontuo também as experiências enriquecedoras vivenciadas e, por eles
supervisionadas, em Portugal e nos Estados Unidos.

Agradeço a minha família e marido, Vinícius, por
mesmo de longe, apoiarem o meu sonho que
parecia estar ainda mais distante de
realizar-se.

Agradeço ao financiamento da CAPES: BEX 9230/12-2 pela
bolsa de Doutorado sanduiche em Lisboa, Portugal;
A Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP,
processo: 2012/04881-9, pela bolsa de Doutorado Direto no país e
bolsa estágio de pesquisa no exterior
(BEPE) , processo: 2014/09461-3.

Sumário

1	Introdução	14
1.1	Malária	14
1.1.1	Malária no Brasil	15
1.2	Localidades do estudo	16
1.3	Vetores da Malária	18
1.3.1	<i>Anopheles (Nyssorhynchus) darlingi</i>	19
1.4	Marcadores moleculares	22
1.4.1	Estudos anteriores com <i>An. darlingi</i>	22
1.4.2	Técnicas utilizadas	23
2	Objetivos	25
3	Materiais e Métodos	26
3.1	Local do Estudo	26
3.2	Coleta de Anofelinos	26
3.3	Preparação do DNA	26
3.4	Genotipagem de Microsatélites	26
3.5	Genotipagem de SNP	28
3.5.1	Preparação das bibliotecas	28
3.5.2	Sequenciamento das bibliotecas	29
3.5.3	Processamento dos dados	29
3.6	Análises dos dados	30
3.6.1	Análises com microsatélites	31
3.6.2	Análises com SNPs	32
3.6.3	Análises conduzidas para ambas técnicas	33
4	Resultados	35
4.1	Microsatélites <i>versus</i> SNPs	35

4.1.1	Diversidade e Estrutura Genética dos microsatélites	35
4.1.2	Diversidade e Estrutura Genética dos SNPs	37
4.1.3	Comparação entre as duas técnicas	39
4.2	Geografia e Sazonalidade	40
4.2.1	Análise microgeográfica por mês de coleta	41
4.2.2	Análise sazonal em Granada	43
4.2.3	Análise sazonal em Remansinho	43
4.2.4	Análise sazonal em ambas as localidades	43
4.3	Comportamento Hematófágico	46
4.3.1	Exofagia <i>vs.</i> Endofagia	46
4.3.2	Crepuscular <i>vs.</i> Matinal: Granada	49
5	Discussão	56
6	Conclusão	63
7	Referências	64
8	Anexos	77

Lista de Figuras

- 1 **Taxa global de mortalidade por malária.** Porcentagem da mudança na taxa de mortalidade por malária nos 106 países com transmissão dessa doença no período de 2010-2013 (WHO, 2014). 14
- 2 **Mapa de risco por município de infecção por malária.** Gradiente em verde apresenta o risco de infecção por malária nos municípios do Brasil, em 2014. 16
- 3 **Brasil e região das coletas.** (A) Mapa do Brasil destacando os Estados do Acre e Amazonas e os três locais de coletas. (B) Imagem de satélite mostrando degradação florestal de Granada (vermelho) e Remansinho (verde). Os assentamentos rurais são conectados pela rodovia BR364 (amarelo) e Rio Iquiri (azul). 18
- 4 ***Anopheles (Nyssorhynchus) darlingi*.** Imagem de uma fêmea do mosquito *An. darlingi* durante repasto sanguíneo. 20
- 5 **Distribuição prevista de *An. darlingi* nas Américas.** Mapa das Américas mostrando a distribuição de *An. darlingi* no continente em probabilidade (1: vermelho, 0: azul), os pontos pretos foram 318 ocorrências de registro desse espécie (Sinka et al., 2010). 21
- 6 **RADseq versus Dupla Digestão RADseq.** (A) *Restriction-Site Associated DNA Sequencing* (RADseq) tradicional usa uma única enzima de restrição e fragmentação aleatória para gerar uma biblioteca de representação reduzida de todo o genoma nas regiões adjacentes ao sítio de restrição (segmentos vermelhos). (B) Dupla Digestão RADseq (ddRADseq), por outro lado, usa duas enzimas de restrição para digerir e precisa seleção de tamanho, exclui regiões muito próximas e muito distantes, a biblioteca portanto, consiste somente em fragmentos próximos ao sítio alvo (segmentos vermelhos). A representação da biblioteca nesse caso tende a ser correlacionada entre os indivíduos (barras amarelas e verdes). Adaptado de Peterson et al., (2012). 24
- 7 **Fluxograma do processamento dos dados de SNPs.** (A) Preparação e sequenciamento das bibliotecas, (B) Filtragem, alinhamento e processamento das sequências. (C) Seleção dos SNPs com o programa populations. 30

8	Análise Bayesiana (BA) e Análise de Componente Principal (PCA) de genótipos de <i>An. darlingi</i> das três localidades. (A) Resultados da BA dos microsatélites ($K=2$), cada coluna representa um indivíduo, divididos em três localidades (Granada, Remansinho e Cruzeiro do Sul) e a proporção que pertence a cada <i>cluster</i> (cinza e preto). (B) Resultados da BA dos SNPs ($K=2$). (C) PCA dos 9 marcadores microsatélites, cada ponto representa um indivíduo. (D) PCA de 1865 SNPs. (C-D) as cores correspondem as populações assinaladas: Granada, vermelho; Remansinho, verde e Cruzeiro do Sul, azul.	39
9	Sumário da Análise Discriminante de Componente Principal (DAPC) e clusterização hierárquica em <i>An. darlingi</i>. (A) Valores do Critério de Informação Bayesiana (BIC) para cada $K(1-10)$, o melhor valor de K foi 3. (B) Três <i>clusters</i> que separam as localidades em um eixo, a área representa a densidade de indivíduos, nota-se uma sobreposição entre Granada (vermelho) e Remansinho (verde); Cruzeiro do Sul em azul. (C) Três <i>clusters</i> que separam as localidade em dois eixos.(D) Clusterização hierárquica por Ward (1963), as cores correspondem a origem de coleta dos indivíduos.	40

10	Análise de Componente Principal (PCA), Análise Bayesiana (BA) e Análise Discriminante de Componente Principal (DAPC) de genótipos de <i>An. darlingi</i> coletados em Granada e Remansinho em três meses de 2012. (A) PCA dos espécimes genotipados em Granada (vermelhos) e Remansinho (verdes); cada formato representa um indivíduo, e os diferentes formatos correspondem ao mês de coleta: Março (quadrados); Maio (triângulos) e Novembro (pontos). (B) Valores do Critério de Informação Bayesiana (BIC) para cada K , melhor K foi 2. (C) Dois <i>clusters</i> separaram <i>An. darlingi</i> de Granada (vermelho) e Remansinho (verde) apenas com informação do primeira função discriminante (DAPC), área representa a densidade de indivíduo. (D) Acima: resultado da BA, o melhor K foi 2 para todas as análises; cada coluna representa um indivíduo e sua porcentagem pertencente a determinado <i>cluster</i> (cinza e preto); abaixo: resultado da BA após filtragem pelo valor de F_{ST} entre as populações.	42
11	Análise de Componente Principal (PCA) de genótipos de <i>An. darlingi</i> coletados em Granada e Remansinho em três meses diferentes. (A) PCA dos espécimes de Granada, os formatos correspondem ao mês de coleta: Março (quadrados), Maio (triângulos) e Novembro (pontos). (B) PCA dos espécimes de Remansinho em três meses. (C) PCA dos espécimes coletados nos três meses nas duas localidades.	44
12	Sumário da Análise Discriminante de Componente Principal (DAPC) e clusterização hierárquica em <i>An. darlingi</i> coletados no primeiro e segundo semestre. (A) Valores do Critério de Informação Bayesiana (BIC) para cada K (1-10), o melhor valor de K foi 2. (B) Dois <i>clusters</i> que separam as localidades em um eixo, a área representa a densidade de indivíduos coletados no primeiro semestre (cinza escuro) e segundo (cinza claro). (C) Clusterização hierárquica por Ward (1963), a primeira derivação foi entre mosquitos coletados nos diferentes semestres.	45

13	Diagrama de Venn. (A) <i>Loci</i> compartilhados entre a comparação semestral, ou seja, espécimes coletados no primeiro e no segundo semestre, e geográfica, sendo todos os meses de Granada vs. todos os meses de Remansinho. (B) <i>Loci</i> compartilhados entre a comparação de <i>An. darlingi</i> endofágico vs. exofágico coletados em Granada e Remansinho em Abril de 2011, e geográfica entre espécimes de ambos locais do mesmo mês.	46
14	Análise de Componente Principal (PCA) de genótipos de <i>An. darlingi</i> endo/exofágicos de ambos assentamentos rurais. (A) PCA de <i>An. darlingi</i> endofágicos (pontos vermelhos preenchidos) e exofágicos (pontos vermelhos vazios) em Granada. (B) PCA de <i>An. darlingi</i> endofágicos (pontos verdes preenchidos) e exofágicos (pontos verdes vazios) em Remansinho. (C) PCA dos 4 grupos em (A) e (B).	48
15	Sumário da Análise Discriminante de Componente Principal (DAPC) em <i>An. darlingi</i> endo/exofágicos de ambos assentamentos rurais. (A) Valores de Critério de Informação Bayesiana (BIC) para cada K , o menor valor foi 3. (B) Proporção de indivíduos dos 4 grupos incluídos nos três <i>clusters</i> inferidos (vermelho, verde e roxo). (C) Ordem dos três <i>clusters</i> em dois eixos, cada ponto representa um indivíduo que pertence ao determinado <i>cluster</i> ; as caixas representam o <i>eigenvalues</i> para PCA e DA.	49
16	Densidade <i>An. darlingi</i> coletados em Granada das 6 pm até 6 am. (A) Coleta no peridomicílio. (B) Coleta no intradomicílio. Os números 1 e 2 indicam os dois picos de densidade analisados, crepúsculo e amanhecer.	50
17	Análise de Componente Principal (PCA) com SNPs de <i>An. darlingi</i> exofágicos de Granada. (A) PCA de 175 SNPs genotipados em <i>An. darlingi</i> exofágicos coletados das 18-21 horas (crepúsculo) pontos verdes; e das 3-6 horas (amanhecer) pontos rosas.	50

18	Sumário da Análise Discriminante de Componente Principal (DAPC) em <i>An. darlingi</i> exofágicos em Granada. (A) Valores de Critério de Informação Bayesiana (BIC) para cada K , o menor valor foi 3. (B) Mesa com proporção de indivíduos dos 2 grupos incluídos nos três <i>clusters</i> inferidos (rosa, amarelo e verde). (C) Ordem dos três <i>clusters</i> , cada ponto representa um indivíduo que pertence ao determinado <i>cluster</i> ; as caixas representam o <i>eigenvalues</i> para PCA e DA.	51
19	Análise de Componente Principal (PCA) com SNPs de <i>An. darlingi</i> endofágicos de Granada. (A) PCA de 1.228 SNPs genotipados no intradomicílio das 18-21 horas (crepúsculo) pontos roxos; das 3-6 horas (amanhecer) pontos amarelos.	52
20	Sumário da Análise Discriminante de Componente Principal (DAPC) em <i>An. darlingi</i> endofágicos em Granada. (A) Valores de Critério de Informação Bayesiana (BIC) para cada K , o menor valor foi 2. (B) Mesa com proporção de indivíduos dos 2 grupos incluídos nos dois <i>clusters</i> inferidos (roxo e amarelo). (C) Ordem dos dois <i>clusters</i> em um eixo, a área representa a densidade de indivíduos que pertencem ao determinado <i>cluster</i>	52
21	Análise de Componente Principal (PCA) e clusterização hierárquica de genótipos de <i>An. darlingi</i> endofágicos e exofágicos de Granada. (A) PCA de 557 SNPs genotipados em 96 indivíduos no intradomicílio (roxo e amarelo) e no peridomicílio (verde e rosa), nos dois picos de densidade, 18-21 horas (verde e roxo) e das 3-6 horas (rosa e amarelo). (B) Clusterização hierárquica por Ward (1963), * ramo conduzido a análise de genes ontólogos (GO). (C) Divisão dos grupos para a análise de genes ontólogos (I-IV).	54
22	Esquema geral do ciclo gonotrófico de mosquitos exofágico & exofílico ou endofágico & endofílico. Os mosquitos se alimentam e descansam fora das casas são considerados exofágicos e exofílicos, respectivamente; e os mosquitos que fazem o repasto sanguíneo e descansam dentro das casas ou proximidades, são respectivamente, endofágicos e endofílicos. Figura baseada na figura 1 de Paaijmans & Thomas, 2011.	60

Lista de Tabelas

1	Sequência dos <i>primers</i> de nove <i>loci</i> microsatélites padronizados para <i>An. darlingi</i> . . .	27
2	Sequência dos <i>primers</i> para ddRADseq.	28
3	Estimativas de R_s , H_e e F_{IS} dos microsatélites de <i>An. darlingi</i> em três localidades. .	35
4	Estimativas de alelos de microsatélites privados em <i>An. darlingi</i>	36
5	Análise <i>locus</i> por <i>locus</i> dos microsatélites in <i>An. darlingi</i>	37
6	Valores de F_{ST} par a par para microsatélites e SNPs genotipados em <i>An. darlingi</i> . . .	38
7	Sumário dos dados de 1.865 SNPs (variáveis e fixos) nas populações de <i>An. darlingi</i> . .	38
8	Sumário dos dados SNPs (variáveis e fixos) de <i>An. darlingi</i> de Granada e Remansinho em três meses diferentes.	45
9	Sumário dos dados SNPs (variáveis e fixos) de <i>An. darlingi</i> endo/exofágicos de ambos os assentamentos rurais.	48
10	Sumário dos dados SNPs (variáveis e fixos) de <i>An. darlingi</i> endo/exofágicos crepuscu- lar e matinal.	53

Resumo

Degradação florestal, alterações ambientais antropogênicas e mudanças climáticas são fatores que podem modificar a dinâmica populacional de anofelinos. *Anopheles (Nyssorhynchus) darlingi* é o principal vetor de malária no Brasil e em outros países na América do Sul. Estudos observaram o aumento da abundância do vetor *An. darlingi* e número de casos de malária após desflorestação e demais modificações ambientais antropogênicas. Além de ter ampla distribuição geográfica, essa espécie possui plasticidade comportamental e diversidade morfológica, biológica e genética. Tendo em vista essa heterogeneidade, o presente estudo avaliou a diversidade genética populacional em *An. darlingi* ligada a distribuição geográfica, dinâmica sazonal e comportamento hematofágico, através de marcadores microssatélites e SNPs. Espécimes de *An. darlingi* foram coletados em dois assentamentos rurais próximos e, em uma área urbana à aproximadamente 600 km de distância. Além disso, as coletas foram realizadas no primeiro e segundo semestre e, dentro e fora das casas, durante o período de atividade hematofágica do vetor, ou seja, das 18-6 horas. Os resultados apresentaram subpopulações de *An. darlingi* em aspecto geográfico, em escalas macro e microgeográficas, acessadas com mais profundidade com os dados dos SNPs. Além disso, o estudo corroborou o prévio achado de duas subpopulações de *An. darlingi* relacionadas ao regime de chuvas. Por fim, pela primeira vez, a heterogeneidade genética em populações simpátricas desse vetor foi encontrada e relacionada com o fenótipo de comportamento hematofágico, endo e exofagia. Esses achados demonstram a importância do entendimento e vigilância entomológica, pois possuem potencial impacto na transmissão de malária. Dado que cada localidade possui características ambientais peculiares e portanto, diferentes composições populacionais dos vetores, intervenções específicas em menor escala passam a ser abordagens interessantes no controle da malária.

Abstract

Forest degradation, human environmental alteration and climate changes are all influence anopheline populations. *Anopheles darlingi* is the main vector of malaria parasite in Brazil and other countries of South America. Deforestation and others anthropogenic activities have been accompanied by sharp increases in both abundance of the primary malaria vector *Anopheles darlingi* and numbers of malaria cases. Besides it is widely distributed, this species display great behavioral plasticity and morphological, biological and genetic diversity. The aim of this study was to analyze population genetic diversity of *An. darlingi* related to geographical distribution, seasonal dynamics and hematophagic behavior, using microsatellites and SNPs markers. *An. darlingi* specimens were collected in two close rural settlements and in an urban area about 600 km away. In addition, collections were performed in both semesters of the year and, indoor and outdoor during the biting activity period of the vector, i.e., 6pm-6am. The results showed subpopulations of *An. darlingi* related to geographical aspect, in a macro and micro geographic scales, better accessed with SNPs dataset. Moreover, this study corroborated with a previous finding showing two subpopulations of *An. darlingi* related to rainfall. Finally, for the first time, genetic heterogeneity was found in sympatric populations of this vector, and it was associated with a phenotype of hematophagic behavior, endo and exophagy. These outcomes demonstrate that genetic heterogeneity may represent an important vector phenotypic variation with potentially highly significant consequences for malaria transmission. Since each site has unique environmental characteristics and therefore, different population compositions of vectors, specific interventions on a smaller scale become interesting approaches for optimal targeted malaria transmission interventions.

Símbolos e Abreviaturas

°	graus
°C	graus Celsius
'	minutos
µl	microlitros
µM	micromolar
h	hora
min	minutos
ng	nanogramas
pb	par de base
®	registrado
s	segundos
™	marca comercial
V	voltz
<i>A</i>	diversidade alélica
BIC	critério de informação Bayesiana (<i>Bayesian Information Criterion</i>)
BP	processo biológico (<i>Biological Process</i>)
<i>COI</i>	citocromo c oxidase subunidade 1
DAPC	Análise Discriminante de Componente Principal
ddRADseq	dupla digestão RADseq (<i>Double Digest RADseq</i>)
DDT	diclorodifeniltricloreto

DNA	ácido desoxirribonucleico
F_{IS}	coeficiente de endogamia
F_{ST}	índice de fixação (diversidade alélica)
GO	gene ontólogo
He	heterozigosidade esperada
Ho	heterozigosidade observada
HWE	Equilíbrio de Hardy-Weinberg
IAM	<i>Infinite Allele Model</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IRS	borrifação residual em ambientes internos (<i>Indoor Residual Spray</i>)
ITN	telas tratadas com inseticidas (<i>Insecticide-treated Nets</i>)
<i>ITS2</i>	<i>internal transcribed spacer 2</i>
K	número de clusters genéticos
MDE	Equilíbrio de Deriva Mutacional n número amostral
ODM	Objetivos de Desenvolvimento do Milênio
ONU	Organização das Nações Unidas
PAD	Projeto de Assentamento Dirigido
PCR	reação em cadeia da DNA polimerase (<i>Polymerase Chain Reaction</i>)
PCA	Análise de Componente Principal
Φ_{ST}	índice de fixação (diversidade haplotípica)

RADseq	sequenciamento de DNA associado a sítio de restrição (<i>Restriction-site Associate DNA sequencing</i>)
RAPD	polimorfismo de DNA amplificado ao acaso (<i>Random Amplified Polymorphic DNA</i>)
RFLP	Polimorfismo de tamanho dos fragmentos de restrição (<i>Restriction Fragment Length Polymorphism</i>)
<i>Rs</i>	riqueza alélica
SMM	<i>Stepwise Mutation Model</i>
SNP	Polimorfismo de base única (<i>Single Nucleotide Polymorphism</i>)
SSR	sequência simple repetidas
UNESP	Universidade Estadual Paulista “Júlio de Mesquita Filho”
<i>vs.</i>	<i>versus</i>
WHO	Organização Mundial da Saúde (<i>World Health Organization</i>)

1 Introdução

1.1 Malária

Malária é a mais importante parasitose humana em todo o mundo. Segundo o *World Malaria Reports 2015*, uma estimativa de 214 milhões de casos de malária foram reportados em 2014 e consequentes 438 mil mortes, sendo que 88% ocorreram no continente africano. O combate a malária faz parte dos Objetivos de Desenvolvimento do Milênio (ODM) estabelecidos pela Organização das Nações Unidas (ONU), em 2000, e outros esforços internacionais como *Roll Back Malaria* pela Organização Mundial da Saúde. Em 2015, 33 países alcançaram o marco de zero casos de malária reportados, e um total de 57 dos 106 países com transmissão de malária reduziram os casos em mais 75% (Figura 1) (WHO, 2015).

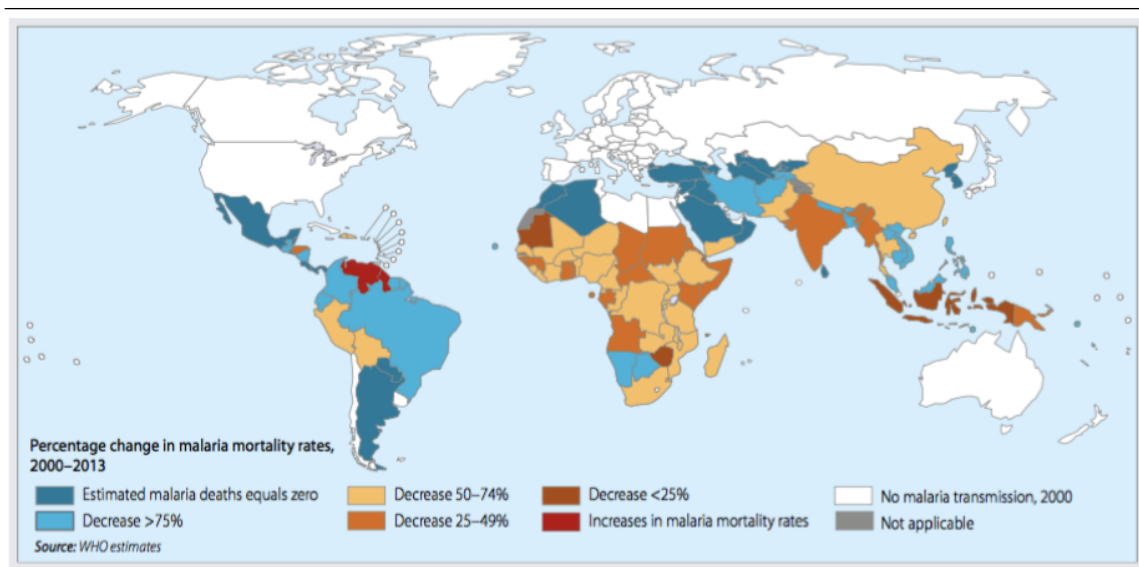


Figura 1: **Taxa global de mortalidade por malária.** Porcentagem da mudança na taxa de mortalidade por malária nos 106 países com transmissão dessa doença no período de 2010-2013 (WHO, 2014).

Os agentes etiológicos da malária humana pertencem ao gênero *Plasmodium*, sendo *Plasmodium vivax*, *P. falciparum*, *P. malariae*, *P. knowlesi* (Cox-Singh et al., 2008; Oliveira-Ferreira et al., 2010) e as sub-espécies *P. ovale curtisi* e *P. ovale wallikeri* (Sutherland et al.,

2010), que são transmitidos por mosquitos do gênero *Anopheles* (Manguin et al., 2008). *P. falciparum* é mais prevalente no continente africano, e é o responsável pela maioria das mortes causadas por malária (WHO, 2015). *P. vivax* tem uma distribuição geográfica maior que *P. falciparum*, explicado pelo seu desenvolvimento em temperaturas mais baixas, e sobrevivência em altas altitudes. Além disso, *P. vivax* tem um estágio dormente no fígado, forma hipnozoíta, que pode ativar meses depois da infecção inicial, causando um reaparecimento dos sintomas, permitindo assim sua manutenção mesmo na ausência do mosquito alado (WHO, 2015).

Portanto, a ecologia natural da malária envolve parasitas plasmódios infectando sucessivamente dois tipos de hospedeiros: humanos e fêmeas de mosquitos anofelinos. Inicialmente, nos humanos, o parasita cresce e multiplica-se nas células hepáticas. Após, ocorre a passagem para o ciclo sanguíneo, desencadeando o aparecimento de sintomas, que pode ser desde febre, sudorese, náusea até anemia grave, malária cerebral e complicações que levam a óbito (CDC).

1.1.1 Malária no Brasil

No histórico de malária no Brasil, a década de 1960 foi registrada com menor número de casos (Tauil & Daniel-Ribeiro, 1998). Esse marco foi conquistado devido ao programa de erradicação do Ministério de Saúde do Brasil, baseado na aplicação de DDT nas paredes internas das casas, bem como no tratamento com cloroquina dos casos febris (Loiola et al., 2002). No entanto, em meados de 1970, a região Amazônica começou um processo acelerado e sem planejamento de ocupação, gerando migração expressiva à assentamentos rurais recém formados (Marques, 1987). Esse movimento migratório levou a um progressivo aumento no número de casos reportados no país, que cresceu de 52 mil em 1970 para 578 mil em 1989 (Tauil & Daniel-Ribeiro, 1998).

Malária ainda é um importante problema de saúde pública no Brasil, concentrou 42% dos 427 mil casos de malária notificados nas Américas em 2013 (WHO, 2014). Embora o principal vetor, o mosquito *An. darlingi*, esteja presente em ampla parte do país, atualmente a incidência da malária no Brasil é quase exclusivamente restrita à região da Bacia Amazônica (Figura 2) (Tauil & Daniel-Ribeiro, 1998; WHO, 2014). O *P. vivax* é responsável pela maioria

dos casos registrados, seguido por casos por *P. falciparum* e *P. malariae* é raramente observado (Oliveira-Ferreira et al., 2010). Embora a malária por *P. vivax* cause pouca mortalidade, quando comparada à malária por *P. falciparum*, é responsável por grande parte da morbidade e por grandes encargos à prosperidade das comunidades endêmicas (Oliveira-Ferreira et al., 2010).

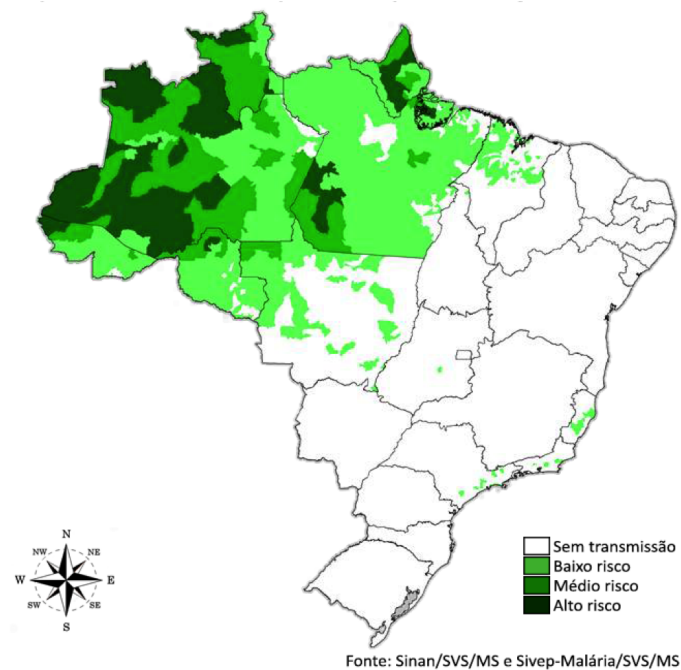


Figura 2: **Mapa de risco por município de infecção por malária.** Gradiente em verde apresenta o risco de infecção por malária nos municípios do Brasil, em 2014.

1.2 Localidades do estudo

O estado do Acre está inserido no grupo que compõe a Amazônia Legal. O seu histórico de urbanização é semelhante a outras regiões do bioma Amazônico, iniciado entre o final do século XIX e o começo do século XX com o ciclo da borracha, um forte extrativismo econômico, que atraiu intensamente mão de obra (Adas, 2006). Em 1977, o Projeto de Assentamento Dirigido de Pedro Peixoto (PAD Peixoto), do Governo Federal, direcionou a migração para a região causando a segunda grande colonização do Acre (Souza, 2002). Uma grande porção

da região oeste deste estado, que engloba os municípios de Acrelândia, Plácido de Castro, Senador Guiomar e a capital Rio Branco foi loteada e distribuída durante o PAD Peixoto. O objetivo era instalar pequenos agricultores em lotes com menos de 100 hectares por família em meio à mata nativa da região. O resultado foi a continuidade da exploração desde os próprios recursos da mata até o extrativismo mineral e a agropecuária, causando gradativas e constantes alterações do espaço físico.

Duas das três localidades do presente estudo, Granada e Remansinho, são assentamentos rurais que surgiram por incentivo deste projeto governamental. Remansinho foi inaugurado em 2007 (Barbosa et al., 2014), enquanto que Granada começou em 1982 (Vital et al., 2014). Essa última, conseqüentemente, tem sofrido a mais tempo o impacto antropogênico, caracterizado pelo maior número de casas e áreas desflorestadas. Ambos os assentamentos nascem na rodovia BR-364 e adentram a floresta em uma estrada não-pavimentada por cerca de 30 km em Granada, e 60 km em Remansinho (Figura 3). Na Amazônia brasileira, assentamentos rurais estão associados com a chamada “malária de fronteira” que explica a maior proporção de números de casos em assentamentos rurais recentes, quando comparados com assentamentos antigos e urbanizados (de Castro et al., 2006; da Silva-Nunes et al., 2012). Malária de fronteira é um fenômeno biológico, ecológico e sócio-demográfico operando em três escalas espaciais (micro/individual, comunidade, e governamental). Em linhas gerais, a primeira escala está relacionada a abundância de *An. darlingi* encontrada nessa área devido às transformações ecológicas favoráveis (Barros & Honório, 2015); conseqüente alta exposição humana à picada do mosquito; e migração de pessoas vindas de áreas livres de malária, e que portanto, possuem baixa imunidade (Castro et al., 2006; Moutinho et al., 2011). A segunda está caracterizada com o baixo nível político organizacional nos assentamentos rurais, impossibilitando o bom gerenciamento clínico de saúde; e por fim, o desenvolvimento de assentamentos rurais, não devidamente planejados pelo governo, caracteriza a terceira escala que sustenta este fenômeno (Castro et al., 2006).

Cruzeiro do Sul, cidade localizada no Alto Rio Juruá, é a segunda mais populosa no Estado do Acre, com 78.444 habitantes (IBGE, 2010). Em 2010, 21.397 casos autóctones de

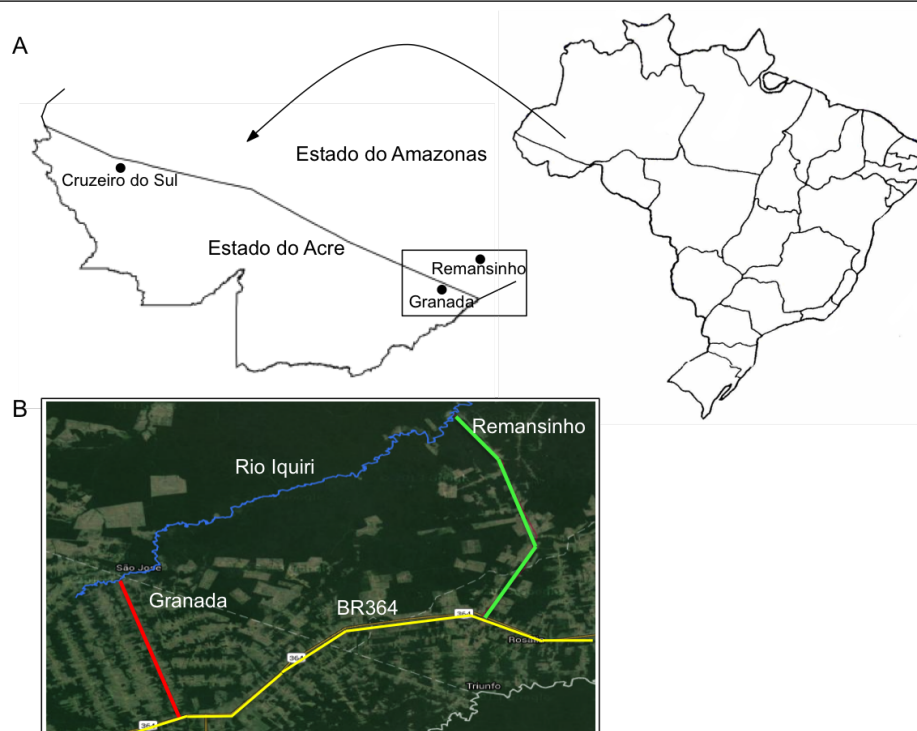


Figura 3: **Brasil e região das coletas.** (A) Mapa do Brasil destacando os Estados do Acre e Amazonas e os três locais de coletas. (B) Imagem de satélite mostrando degradação florestal de Granada (vermelho) e Remansinho (verde). Os assentamentos rurais são conectados pela rodovia BR364 (amarelo) e Rio Iquiri (azul).

malária foram notificados na cidade, totalizando 58,5% de todas as notificações do Acre (Braz et al., 2012). O trabalho de Costa et al. (2010), realizado em Cruzeiro do Sul, observou um expressivo aumento na transmissão de malária após a escavação de 179 novos poços para criação de peixes na região em 2005. Em Mâncio Lima, cidade próxima de Cruzeiro do Sul, Reis et al. (2015) encontraram associação direta entre a abundância de larvas de *An. darlingi* em tanques e prevalência de malária; por conseguinte, foi sugerido que os tanques de piscicultura foram modificações antropogênicas importantes para o aumento da transmissão na região.

1.3 Vetores da Malária

Diferentemente da África, onde as regiões endêmicas para malária abrangem uma variedade de nichos ecológicos, ou seja, desde a savana até floresta úmida, nas Américas Central e do Sul, as

regiões endêmicas estão principalmente concentradas na região Amazônica na América do Sul e florestas tropical e subtropical na América Central (Conn & Ribolla, 2015). Essa diferença pode refletir na diversidade de vetores da malária nos dois continentes, enquanto que na África, *Anopheles coluzzi*, *Anopheles gambiae*, *Anopheles arabiensis*, e *Anopheles funestus s.l.* são os principais vetores, nas Américas, a diversidade é reduzida, com *An. darlingi* como vetor primário na América do Sul, e *Anopheles albimanus* na América Central (Molina-Cruz et al., 2004; Loaiza et al., 2008). No Brasil, *Anopheles aquasalis* e *Anopheles albitarsis s.l.* foram considerados como vetores secundários da malária (Forattini, 2002; Hiwat & Bretas, 2011).

1.3.1 *Anopheles (Nyssorhynchus) darlingi*

Anopheles (Nyssorhynchus) darlingi (Figura 4) é o principal vetor de malária na região Amazônica (Deane et al., 1948; Conn & da Silva-Vasconcelos, 2002; Galardo et al., 2007, da Rocha et al., 2008).

Essa espécie foi inicialmente descrita por Root in 1926 e posteriormente nomeada pelo líder em doenças tropicais, Dr. Samuel Taylor Darling (1872-1925). *An. darlingi* está contida no gênero *Anopheles*, juntamente com mais de 450 espécies, porém menos de 10% são reconhecidas como transmissoras de malária ao homem (Harbach, 2004; Maguin et al., 2008). Esse mosquito tem ampla distribuição geográfica, desde do sul do México até o norte da Argentina, do leste das cordilheiras dos Andes até a costa do oceano Atlântico (Figura 5) (Forattini, 1987; Sinka et al., 2012). Três fatores contribuem para o papel de *An. darlingi* na transmissão de malária: a susceptibilidade ao agente etiológico humano, o comportamento antropofílico ou oportunístico (Zimmerman et al., 2006; Sinka et al., 2012) e a rápida adaptação às modificações ambientais (Rozendaal, 1990; Vittor et al., 2006; Hiwat & Bretas, 2011).

O ciclo de vida dos anofelinos é dividido em quatro fases: ovo, larva, pupa e adulto alado, sendo que as três primeiras são aquáticas. As larvas de *An. darlingi* foram encontradas tanto em criadouros naturais (lagos, margens de rios, córregos e áreas alagadas) quanto artificiais (tanques de piscicultura, valas e drenos), porém normalmente em áreas sombreadas, com vegetação flutuantes, pobres em sais e matéria orgânica (Forattini, 2002; Vittor et al., 2009).



Figura 4: *Anopheles (Nyssorhynchus) darlingi*. Imagem de uma fêmea do mosquito *An. darlingi* durante repasto sanguíneo.

Uma vez que as fêmeas necessitam de alimentação sanguínea para a maturação dos ovos, tornam-se as responsáveis pela transmissão de doenças. O padrão da atividade hematofágica de *An. darlingi* mostrou-se variável quanto ao local e horário, podendo ser dentro (endofágico) ou fora das casas (exofágico) (Charlwood, 1996; da Silva-Vasconcelos et al., 2002; Zimmerman et al., 2013); e a taxa de densidade de picadas uni-bi ou trimodal no intervalo das 18 horas até as 6 horas (Elliot, 1972; Loureno-de-Oliveira et al., 1989; Voorham, 2002; Hiwat & Bretas, 2011). Além disso, houve evidência de variação no repouso desse vetor, podendo ser dentro das casas (endofilia) (Roberts et al., 1987; Quinones & Suarez, 1990) ou fora das casas (exofilia), todavia, o comportamento exofílico predominou em grande parte de sua distribuição (Zimmerman et al., 2004; Hiwat & Bretas, 2011).

O controle do vetor é realizado com borrifação de inseticida no interior das casas, do inglês *indoor residual spray* (IRS), e uso de mosquiteiros impregnados com inseticida para dormir, do inglês *insecticide-treated nets* (ITN) (Lourenço-de-Oliveira et al., 1989), porém nenhum dos dois atingem populações exofílicas/exofágicas dos vetores (Hiwat & Bretas, 2011; da Silva-Nunes et al., 2012; Orjuela et al., 2015). Ademais, a efetividade no uso dos mosquiteiros depende do padrão local de picada dos vetores. Em uma pequena comunidade na Colômbia, onde *An. darlingi* é o principal vetor, sua atividade é principalmente endofágica e da 00:00 à 1:00 hora, horário viável para efetiva intervenção contra o mosquito (Jiménez et al., 2014). Por

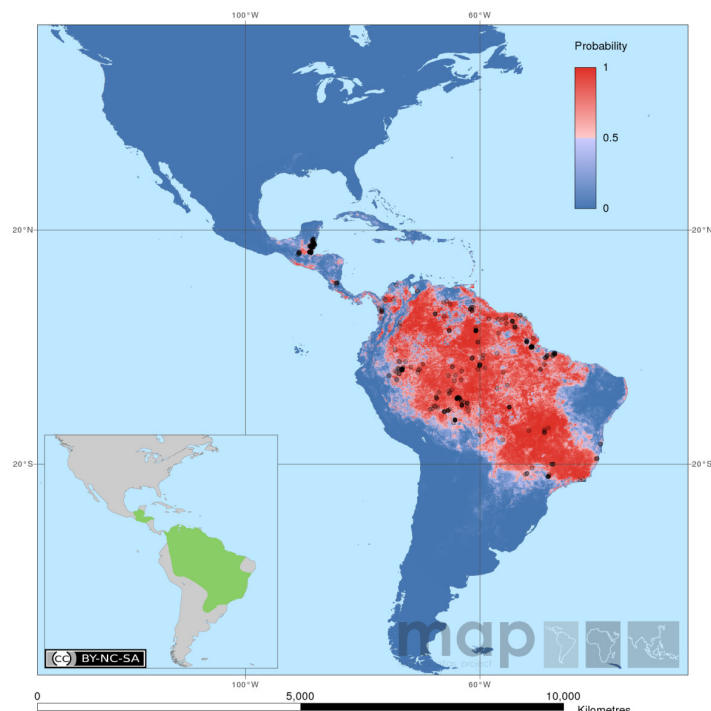


Figura 5: **Distribuição prevista de *An. darlingi* nas Américas.** Mapa das Américas mostrando a distribuição de *An. darlingi* no continente em probabilidade (1: vermelho, 0: azul), os pontos pretos foram 318 ocorrências de registro desse espécie (Sinka et al., 2010).

outro lado, em uma pequena cidade também na Colômbia, espécimes de *An. darlingi*, tanto endofágicos quanto exofágicos, foram coletados sobretudo no começo da noite e no começo da manhã, período no qual as pessoas não estão protegidas por mosquiteiros (Jiménez et al., 2012).

A densidade populacional de *An. darlingi* e outros anofelinos neotropicais varia sazonalmente e, normalmente, apresenta picos no final da estação chuvosa (Sinka, 2010; Herrera et al., 2014). Além das mudanças climáticas, a degradação florestal, bem como outras alterações ambientais antropogênicas são os fatores que mais influenciam a dinâmica populacional de anofelinos. Áreas desmatadas e ecologicamente alteradas tiveram aumento da presença de larvas de *An. darlingi* e decorrente aumento de picadas em humanos pela forma adulta dessa espécie (Vittor et al., 2006; Vittor et al., 2009). Kamdem et al. (2012) mostraram rápida adaptação de linhagens divergentes de *An. gambiae* aos novos nichos criados por intervenções

humanas. Sendo assim, o estudo do comportamento e biologia do vetor, bem como sua dispersão, adaptação e interação com humanos é de grande importância para entomologia médica, a fim de melhor direcionar medidas para o controle vetorial.

1.4 Marcadores moleculares

1.4.1 Estudos anteriores com *An. darlingi*

Os marcadores moleculares contribuem para a entendimento da heterogeneidade genética das espécies, elucidando importantes informações sobre o vetor e sua estrutura populacional (Lounibos & Conn, 2000). Estudos com *An. darlingi* revelaram de moderada a alta heterogeneidade genética em espécimes coletados no Brasil e em outros países da América do Sul. A análise de sequências da região *ITS2* ribossomal apresentou aproximadamente 5% de divergência entre populações do mosquito das regiões norte e sul do Brasil (Malafronte et al., 1999), corroborando o polimorfismo em cromossomos politênicos encontrado por Kreutzer et al. (1972). Baseados em sequências do DNA mitocondrial (gene *COI*) e microssatélites, estudos mostraram uma divisão genética significativa entre os mosquitos dessa espécie da América Central e do Sul (Mirabello & Conn 2006; Mirabello et al., 2008). Conn et al. (2006) demonstraram isolamento por distância de populações de *An. darlingi* do norte e do sul do rio Amazonas através da genotipagem de microssatélites. Nas populações da América do Sul, barreiras geográficas como o citado rio Amazonas e os Andes, por exemplo, dividiram as populações de *An. darlingi* em pelo menos quatro grupos, utilizando sequenciamento do gene *COI* (Pedro & Sallum, 2009). Por outro lado, em um nível microgeográfico, Gutiérrez et al. (2010) mostraram baixa diferenciação genética, analisando o gene *COI* e microssatélites, entre populações desse vetor no noroeste da Colômbia. Em uma observação a um nível climático, com diferente regime de chuvas, marcadores microssatélites mostraram subpopulações de *An. darlingi* moduladas sazonalmente ao longo do rio Madeira, em Rondônia, Brasil (Angella et al., 2014).

A grande distribuição geográfica, a diversidade morfológica e comportamental, junto a heterogeneidade genética são de grande importância epidemiológica, uma vez que podem refletir na capacidade vetorial das populações de *An. darlingi*. Ademais, conduz a hipótese da

existência de um complexo de espécies ao invés de uma espécie monofilética (Charlwood, 1996). Manguin et al. (1999) rejeitaram essa hipótese com base em resultados de um estudo com diversos marcadores (morfologia, isoenzimas, *ITS2* e RAPD), porém Emerson et al. (2015) detectaram três *clusters* genéticos principais em *An. darlingi*, os quais foram sugeridos como três espécies putativas.

1.4.2 Técnicas utilizadas

Microsatélites Os microsatélites, do inglês *simple sequences repeats* (SSRs), são marcadores moleculares utilizados em uma variedade de estudos genéticos populacionais (Goldstein et al., 1999; Ndo et al., 2010; Ma et al., 2011), em sistemática (Feldman et al., 1997; Sun et al., 2009) e no mapeamento genômico (Weissenbach, 1992; Solignac et al., 2007; Jaari et al., 2009). Esses marcadores são regiões do genoma dos organismos eucariontes que possuem repetições de 2 a 6 nucleotídeos em cadeia. As repetições estão distribuídas aleatoriamente no genoma em uma densidade de aproximadamente 1 microsatélite a cada 10 a 15 mil pares de bases, dependendo da espécie (Tautz, 1989). Tem alto grau polimórfico em relação ao número de repetições (taxa de mutação estimada em 10^{-2} a 10^{-3} por *locus* por gameta por geração) podendo influenciar na expressão e funcionalidade do gene (Nelson & Warren, 1993). No campo de entomologia médica, os microsatélites vêm sendo utilizados para o estudo populacional de vários insetos.

Dupla digestão RADseq Algumas abordagens moleculares tem utilizado o sequenciamento de pequenos fragmentos ancorados aos sítios de enzima de restrição. A técnica chamada “Sequenciamento do DNA Associado ao Sítio de Restrição” (RADseq), do inglês *Restriction Site Associated DNA Sequencing*, é uma delas, capaz de identificar milhares de polimorfismos de única base (SNP), aleatoriamente distribuídos no genoma, através do sequenciamento pela plataforma Illumina (Davey & Blaxter, 2010). A técnica utiliza fragmentação específica por uma enzima de restrição, assim RADseq analisa uma sub-amostragem do genoma total da amostra, permitindo não somente a genotipagem e descoberta de SNPs, mas também análises mais complexas como genética quantitativa e estudos filogenéticos (Davey & Blaxter, 2010;

Emerson et al., 2010; Hohenlohe et al., 2010). No presente estudo, foi utilizada a “Dupla Digestão RADseq”, uma adaptação de RADseq descrita por Peterson et al. (2012) (Figura 6).

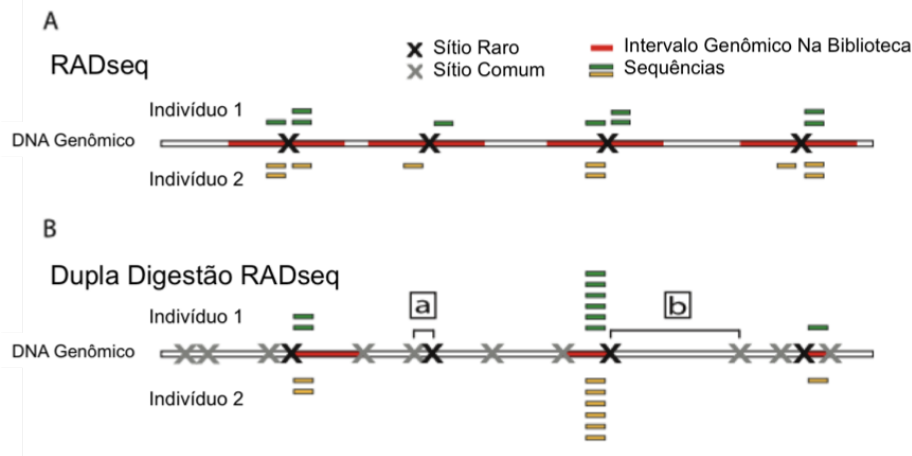


Figura 6: **RADseq versus Dupla Digestão RADseq.** (A) *Restriction-Site Associated DNA Sequencing* (RADseq) tradicional usa uma única enzima de restrição e fragmentação aleatória para gerar uma biblioteca de representação reduzida de todo o genoma nas regiões adjacentes ao sítio de restrição (segmentos vermelhos). (B) Dupla Digestão RADseq (ddRADseq), por outro lado, usa duas enzimas de restrição para digerir e precisa seleção de tamanho, exclui regiões muito próximas e muito distantes, a biblioteca portanto, consiste somente em fragmentos próximos ao sítio alvo (segmentos vermelhos). A representação da biblioteca nesse caso tende a ser correlacionada entre os indivíduos (barras amarelas e verdes). Adaptado de Peterson et al., (2012).

Nessa abordagem a digestão do DNA é feita com duas enzimas de restrição, uma com corte frequente e outra rara, dispensando assim os passos de fragmentação aleatória e *end repair* necessários na técnica original. Essa técnica aumenta a eficiência e robustez enquanto diminui custo, além de necessitar de apenas 100ng ou menos de DNA genômico inicial (Peterson et al., 2012). O passo mais importante é a escolha das enzimas utilizadas para digerir o genoma, pois isso determinará o tamanho dos fragmentos, a porcentagem do genoma a ser estudado, e consequentemente, o número de marcadores SNPs.

2 Objetivos

O estudo visou analisar a heterogeneidade genética ligada a distribuição geográfica de populações de *An. darlingi* contrastando áreas rurais e urbana na Amazônia brasileira, bem como a dinâmica sazonal do vetor e polimorfismos genéticos relacionados aos diferentes comportamentos hematofágicos em populações simpátricas. Para tanto, duas metodologias de genotipagem foram utilizadas: marcadores microsatélites e polimorfismos de base única (SNPs). Além disso, foi testado o grau de discriminação genética a nível microgeográfico dessas duas metodologias.

3 Materiais e Métodos

3.1 Local do Estudo

O estudo foi realizado em três localidades, dois assentamentos rurais, Granada (S9°47'; W67°05') em Acrelândia, Acre e Remansinho (S9°31'; W66°33') em Lábrea, Amazonas, e a cidade de Cruzeiro do Sul (S7°37'; W72°40') também Estado do Acre (Figura 3).

3.2 Coleta de Anofelinos

Foram realizadas coletadas intradomiciliar e extradomiciliar (10 metros de distâncias das casas) por captura sob atração humana (grupo de pesquisa do laboratório Pangene dirigido pelo Prof. Dr. Paulo Ribolla) nos assentamentos rurais em Abril de 2011, Março, Maio e Novembro de 2012. Em Cruzeiro do Sul, coletas extradomiciliares foram realizados com o auxílio de armadilhas Shanon de captura com atração humana protegida, em Março de 2013, pelo grupo da Dr. Marinete Póvoa (Instituto Evandro Chagas, Belém, Pará). Todos os espécimes coletados foram morfológicamente identificados e somente os espécimes de *An. darlingi* foram utilizados nas análises seguintes (Consoli & Oliveira, 1994).

3.3 Preparação do DNA

Duas metodologias foram utilizadas para extração e quantificação do DNA. O protocolo que utiliza a resina Chelex100® *Molecular Biology Grade Resin* (Bio-Rad Laboratories), a 5% e a quantificação pelo espectrofotômetro NanoDrop® (ND-1000), foram aplicados às amostras submetidas posteriormente a genotipagem de microssatélites. Já as amostras submetidas a genotipagem de SNPs foram extraídas seguindo o protocolo do kit ReliaPrep™ *Blood gDNA Miniprep System* (Promega) e quantificadas pelo fluorômetro Qubit® 2.0 (Life Technologies).

3.4 Genotipagem de Microssatélites

O DNA de cada mosquito foi amplificado por reação de polimerização em cadeia (PCR) com a utilização dos oligonucleotídeos específicos para cada microssatélite ligados à marcadores

Tabela 1: Sequência dos *primers* de nove *loci* microsatélites padronizados para *An. darlingi*.

<i>Locus</i>	Unid. de repetição	Sequência dos <i>primers</i> (5' -3')	Temperatura anelamento (°C)	Fluorescência
ADC01	(GA) ₂₁	GCT CTA ATG ATG CTC GTA ACC GCT	55	FAM
AF322185		CAG GTA GTA GCA AGT GCG GTG TCG		
ADC02	(GA) ₇₊₇	CAC ACT GGG GCA TCA TTC ATT TC	52	FAM
AF322186		CAG TGT CTA AGG ACG CAG TGT GA		
ADC28	(GA) ₉	CTC GTC GTC AGC GTC GTG C	56	FAM
AF322187		CCG TTA CGC AGT GGA TGG GCA		
ADC29	(GA) ₂₂	GAT CGT TGG CCG AGA ATG	56	NED
AF322188		GCA TCG TGC ACC GTG ATT		
ADC110	(GT) ₁₆	CCG AAC AAC AGC CAA CAG CTG TG	54	HEX
AF322190		CGT GTG TAA CGA TTG TGT CGA ACG		
ADC137	(GT) ₁₁	TCT TAC GGG AAT GGT GCG ACG CTC	56	HEX
AF322191		TGG TCA ACA GCG TAT GGG TGG CTG		
ADC138	(AC) ₁₄₊₃	CTT TGA GCC GGT GCT GTG CTG C	57	NED
AF322192		GTC CTG GAG GCT GCG AGA ATG G		
ADS P2	(GTA) ₇	GCATATATTCTCGCCGCATT	53	FAM
DV729944		TGTCGTCGGGTAGTAGCTGA		
ADM P9	(AAC) ₁₁	ACA ACG TCC AAT GCA ACA AC	57	HEX
DV729762		CGATACAGAAAGCCCTCGAG		

fluorescentes nos *primers reverse* FAM, NED e VIC (Applied Biosystems) (Conn et al., 2001; Angella et al., 2014) (Tabela 1). Cada reação de PCR foi realizada num volume total de 20µl contendo GoTaq® *Colorless Master Mix*, 2X; 0,5µM de cada oligonucleotídeo e 5ng de DNA genômico. As condições da PCR para ambas etapas foram: desnaturação inicial a 94°C por 5min seguida por 30 ciclos a 94°C por 30s, 52-57°C (dependendo do *locus*) por 30s e 72°C por 35s e um alongamento final a 72°C por 5-10min. Todas as reações tiveram a amplificação confirmadas no gel de agarose (1,5%) corado com GelRed™ (Uniscience), submetidas a corrida de eletroforese a 90V por 50 minutos juntamente com um *ladder* de peso molecular 100 pb (Promega).

Os fragmentos amplificados foram organizados em três grupos multiplex, a fim de otimizar a técnica, e separados por eletroforese capilar em sequenciadores automáticos (ABI 3500) e os seus tamanhos analisados pelo *software* GeneMarker (SoftGenetics, USA). Uma base de dados em Excel® (Microsoft) dos genótipos por indivíduo e *locus* foi organizada, contendo o tamanho dos alelos em pares de bases.

3.5 Genotipagem de SNP

3.5.1 Preparação das bibliotecas

Para a genotipagem de SNPs foi utilizada a técnica de sequenciamento do DNA associado a dupla digestão por enzimas de restrição (ddRADseq). Inicialmente, um total de 200ng de DNA genômico foi individualmente digerido com a combinação das enzimas de restrição EcoRI (Promega) e HpaII (Promega) em um volume final de 40µl, o qual foi purificado com as *beads* AMPure® XP (Agencourt®), segundo o protocolo do fabricante.

Um par de adaptadores (P1 e P2) foi desenhado incluindo a sequência complementar Nextera® Index Primers (Illumina), a fim de realizar a indexação com Nextera® DNA Sample Preparation Kit (Illumina) (Tabela 2). Quantidades apropriadas dos adaptadores P1 (0.3µM) e P2 (4.8µM) foram ligados com o DNA digerido pela enzima *T4 DNA Ligase* (Promega).

Tabela 2: Sequência dos *primers* para ddRADseq.

RADseq dupla digestão		
	Nome da sequência	Sequência
PCR primers Nextera® Index Kit	P5	5' AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC 3'
	P7	3' GGCTCGGGTGCTCTG[i7]TAGAGCATACGGCAGAAGACGAAC 5'
Adaptador P1	1.1	5' TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG 3'
	1.2	3' AGCAGCCGTCGCAGTCTACACATATTCTCTGTCTTAA 5'
Adaptador P2	2.1	5'CGCTGTCTCTTATACGTCCTCTT 3'
	2.2	3' GACAGAGAATATGTGTAGAGGCTCGGGTGCTCTG 5'

Depois de outra purificação com AMPure® XP , os fragmentos de DNA foram selecionados em um intervalo de 350-400 pares de base em gel agarose e purificados novamente. A amplificação por PCR pelo Nextera® Indexing Kit foi feito para gerar as bibliotecas de sequenciamento Illumina, segundo a seguinte condição de ciclagem: denaturação inicial a 72°C por 3min e a 95°C por 30s, seguido por 16 ciclos a 95°C por 10s, anelamento a 55°C por 30s, alongação a 72°C por 30s, e uma extensão final a 72°C por 5 minutos, e então o produto foi

purificado uma última vez.

3.5.2 Sequenciamento das bibliotecas

Todas as amostras foram individualmente quantificadas com KAPA *qPCR Library Quantification Kit* (Biotik®) e combinadas com razões equimolares para compor a biblioteca final. A biblioteca foi quantificada com o mesmo método, depois normalizada e desnaturada, e finalmente carregada em um cartucho de reagentes Illumina. Quatro bibliotecas, 3 com 24 amostras e uma com 15 amostras, foram sequenciadas por ambas extremidades em 150-ciclos em um Miseq *Desktop Sequencer* (Departamento de Genética, Universidade Estadual de São Paulo), e outras 3 bibliotecas, 2 com 96 amostras e uma com 90 amostras, foram sequenciadas nas mesmas especificações, porém em um Hiseq 2500 (*Genomics and Bioinformatics Core*, Universidade Estadual de Nova York, em Buffalo), portanto, um total de 369 amostras foram sequenciadas.

3.5.3 Processamento dos dados

Stacks (versão 1) *pipeline* (Catchen et al. 2013) foi usado para identificação de SNP *loci* (Figura 7). O primeiro programa utilizado do Stacks foi *process_radtags*, responsável por remover sequências errôneas dos dados brutos do Illumina, erros tais na identificação das amostras (*barcode*), nos sítios de restrição das enzimas e/ou baixa qualidade das bases. Depois de filtradas, todas as sequências foram alinhadas pelo programa Bowtie2 (Langmead & Salzberg, 2012) ao genoma de referência de *An. darlingi* (Marinotti et al., 2013), a fim de selecionar somente as sequências oriundas dessa espécie.

Os arquivos em formato SAM foram então executados com o programa *ref_map.pl*, o qual contem outros programas tais como o *pstacks*, *cstacks* e *sstacks*. O primeiro programa combinou as sequências, construiu *loci* considerando uma profundidade mínima de 5 cópias, e definiu os SNPs de cada indivíduo; já o programa *cstacks* agrupou os dados de todos os indivíduos em um catálogo único. O programa *sstacks* comparou cada indivíduo com o catálogo afim de identificar SNPs entre eles. Baseados nos dados obtidos, um passo de correções

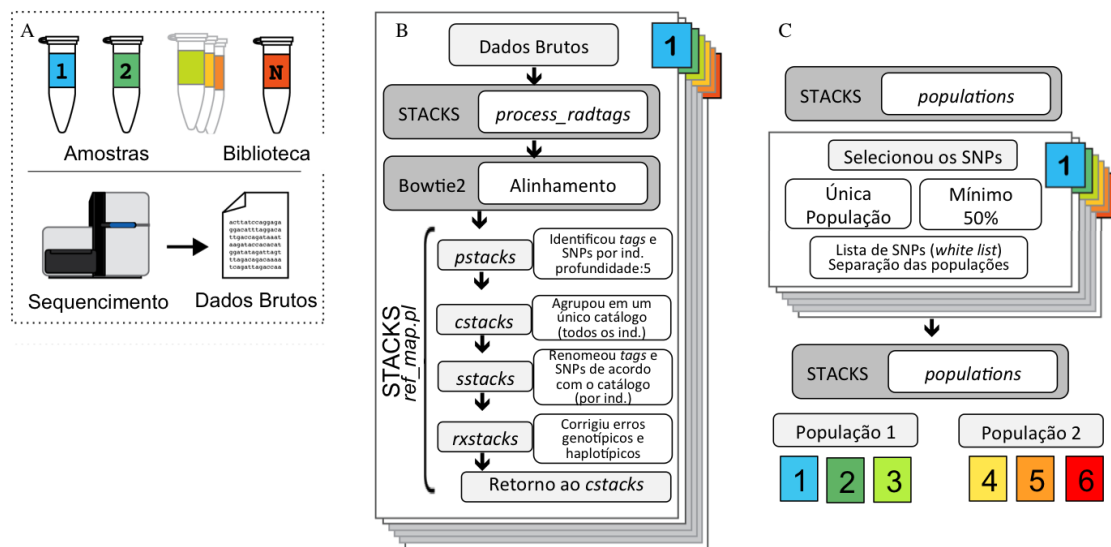


Figura 7: **Fluxograma do processamento dos dados de SNPs.** (A) Preparação e sequenciamento das bibliotecas, (B) Filtragem, alinhamento e processamento das sequências. (C) Seleção dos SNPs com o programa *populations*.

genotípicas e haplotípicas se fez necessário, feito pelo programa *rxstacks*. Os programas *cstacks* e *sstacks* foram executados novamente, e finalmente o programa *populations* foi executado, o qual identificou os SNPs presentes em um dado conjunto de indivíduos. Inicialmente, todos os indivíduos foram considerados como uma única população, a fim de evitar viés populacional na seleção dos SNPs. Além disso, os marcadores deveriam estar presente em pelo menos 50% do total de indivíduos, com a finalidade de evitar viés por falta de dados.

3.6 Análises dos dados

As análises foram conduzidas em três perspectivas diferentes: 1) comparação das metodologias de genotipagem (microsatélites e SNPs); 2) geográfica e sazonal, e 3) relacionada ao comportamento hematofágico do vetor, quanto ao local (intra e peridomicílio) e horário (crepúsculo e amanhecer).

3.6.1 Análises com microsatélites

Frequências alélicas As frequências alélicas foram obtidas por contagem direta através do pacote de programa Arlequin v.3.1 (Excoffier et al., 2005). Este cálculo é resultado da soma do número de indivíduos heterozigotos para um determinado alelo com duas vezes o número de indivíduos homozigotos, e divisão do resultado pelo dobro do número total de indivíduos.

Diversidade alélica e riqueza alélica A diversidade alélica (A) é o número total de alelos presentes numa amostra para um dado *locus*. A riqueza alélica (R_s) é uma medida de diversidade alélica ajustada pelo tamanho da amostra de menor dimensão. Esta medida permite uniformizar as amostras em estudo de modo a torná-las comparáveis. Estes parâmetros foram determinados através do programa Fstat versão 2.9.3.2 (Goudet, 1995).

Testes de heterozigosidade A heterozigosidade observada (H_o) é a proporção de genótipos heterozigóticos encontrada numa amostra de n indivíduos. A heterozigosidade esperada (H_e) é a proporção esperada de genótipos heterozigóticos numa população de n indivíduos. Esse parâmetro baseia-se nas frequências alélicas num dado *locus*, assumindo o princípio de Hardy-Weinberg (Nei, 1978). Ambos parâmetros foram determinados com auxílio do programa Arlequin. Testes de heterozigosidade (Cornuet & Luikart, 1996) foram realizados para detectar desvios do equilíbrio de deriva mutacional (MDE). Estes testes comparam duas estimativas de heterozigosidade esperada, uma baseada nas frequências alélicas (H_e), assumindo as proporções de Hardy-Weinberg, e outra baseada no número de alelos e tamanho amostral (H_{eq}), assumindo o MDE. Se a população experimenta um efeito de gargalo, os alelos raros são rapidamente perdidos e com isso o número de alelos diminuirá mais rápido do que a frequência alélica, ou seja, $H_e > H_{eq}$, enquanto que o contrário ($H_e < H_{eq}$) pode indicar uma expansão populacional. Estimativas de heterozigosidade esperada sob MDE foram calculadas com os modelos SMM (*Stepwise Mutation Model*) e IAM (*Infinite Allele Model*) através do software Bottleneck 1.2.02. (Cornuet & Luikart, 1996).

Equilíbrio de Hardy-Weinberg (HW) A verificação do equilíbrio de HW, que aplica o método de Guo & Thompson (1992), foi feita com auxílio do programa Arlequin. Este teste foi feito com os dados dos genótipos, *locus* por *locus*. Para cada *locus* microssatélite testou-se a hipótese nula (H_0 = união aleatória de gametas) com duas hipóteses alternativas (H_1 = déficit de heterozigóticos ou H_1 = excesso de heterozigóticos).

Análise de alelos nulos Os alelos nulos são alelos não amplificados que, quando em heterozigose, resultam em aparente homozigose. Para os microssatélites, os alelos nulos podem surgir devido a mutações nas regiões flangeadoras, inviabilizando a ligação dos *primers*. A presença de alelos nulos em marcadores microssatélites é frequente, especialmente em grandes populações naturais, e é um fenômeno comum em populações de anofelinos (Lehmann et al. 1997, Conn et al., 2006). A sua ocorrência pode causar um erro importante nas proporções genotípicas, observado por um desvio do equilíbrio de HW, devido ao excesso de homozigotos. Para a detecção de possíveis alelos nulos, o software Micro-Checker (Van Oosterhout et al., 2004) foi utilizado.

Endogamia O coeficiente de endogamia (F_{IS}), que traduz os níveis de heterozigotia e homozigotia numa dada população, foi obtido através do programa Arlequin para cada *locus* microssatélite. Se $F_{IS}=0$ o acasalamento entre os indivíduos de uma população ocorre de modo aleatório e $H_e = H_o$.

Desequilíbrio de ligação O desequilíbrio de ligação, que é a associação não aleatória de alelos em *loci* diferentes, foi analisado com auxílio do programa Genepop versão 1.2 (Raymond & Rousset, 1995; Rousset, 2008).

3.6.2 Análises com SNPs

Resumos estatísticos Os resumos estatísticos, tais como heterozigosidade esperada e observada, F_{IS} das populações e Φ_{ST} e F_{ST} entre populações, foram gerados pelo programa Stacks, assim como arquivos compatíveis para os *softwares* Structure (Pritchard et al., 2000)

e Arlequin. Além disso, uma lista com Φ_{ST} e F_{ST} *locus* por *locus*, de todos os *locus* incluídos em cada análise foi obtido com o mesmo programa.

Diagrama de Venn O pacote gplots (Warnes, 2011) em R (R Developed Core Team, 2016) foi utilizado para detectar os *loci* compartilhados entre dois grupos.

Análise de clusterização hierárquica O pacote ape 3.4 (Paradis et al., 2004) em R foi utilizado para calcular a matriz de distância genética com o método 'pairwise', onde a distancia d entre dois indivíduos é igual ao número de *loci* diferentes, e a variância associada é $d(L - d)/L$, onde L é o número de *loci*. A clusterização hierárquica, baseada nessa matriz de distância, foi computada pelo método Ward's (1963) (Murtagh & Legendre 2014), onde a dissimilaridades são elevadas ao quadrado antes da clusterização.

Análise de Genes Ontólogos Uma vez que um dos passos no processamento dos dados foi o alinhamento ao genoma disponível de *An. darlingi*, foi possível verificar os genes próximos aos SNPs utilizados no estudo. Inicialmente, foram selecionados somente 5% dos SNPs com maiores valores de Φ_{ST} associados na comparação entre dois grupos. O pacote biomaRt (Durinck et al., 2009) em R foi utilizado para identificar genes anotados em *An. darlingi*, bem como genes ortólogos em *Aedes aegypti*, *Culex quinquefasciatus* e *Drosophila melanogaster*, localizados a até 5000 pb de cada SNP para os dois sentidos. Uma lista de genes foi obtida a partir dessa análise e, convertida para uma lista de genes ontólogos (GO) correspondentes, através de um *script* escrito em *python* (Anexo Texto 2). Finalmente, o pacote topGO (Alexa & Rahnenfuhrer, 2010) em R foi utilizado para a análise de enriquecimento de termos GO de processo biológico (BP), que normalizou, correlacionou, e gerou uma tabela com os resultados.

3.6.3 Análises conduzidas para ambas técnicas

Estrutura populacional As análises da estrutura populacional foram calculadas empregando a análise de variância molecular (AMOVA), com o programa Arlequin. Este método estatístico se baseia na geração de estimativas da variância genética em diferentes níveis hierár-

quicos (inter e intrapopulacional). Para as análises Bayesianas de subestrutura populacional recorreu-se ao programa Structure. A análise de *clusters*, segundo as recomendações de Kimberly et al. (2012), foi realizada com o modelo de mistura, anulação das primeiras 100.000 das seguintes 1.000.000 de simulações, com 40 repetições para K de 1 ao número máximo de populações mais 2. O valor ótimo de K foi inferido pelo método de Evanno (Evanno et al., 2005), implementado em *structureHarvester* (Earl & VonHoldt, 2012). Além disso, duas abordagens foram realizadas: uma com o conjunto total de SNPs encontrados; e outra somente com SNPs informativos na diferenciação de duas populações, ou seja, com F_{ST} maior ou igual a 0,05. A finalidade foi testar se em uma análise Bayesiana os marcadores não-informativos mascaram uma real estruturação entre as populações.

Análises multivariadas O pacote adegenet 2.0.0 (Jombart, 2010) em R foi usado para realizar a Análise de Componente Principal (PCA) e a Análise Discriminante de Componente Principal (DAPC). A primeira sintetiza uma visão geral de divergência entre os indivíduos independente do grupo de origem. Já a segunda foca nas diferenças entre os grupos, enquanto negligencia as diferenças dentro de cada grupo. Dessa maneira, o Critério de Informação Bayesiana (BIC) é acessado para cada valor de K , e o número ótimo de *clusters* genéticos é o menor desses valores. Após essa definição, a DAPC permite a designação probabilística de cada indivíduo em relação a um *cluster*, assim como nos métodos de clusterização Bayesiana. Essas análises multivariadas tem sido utilizada como alternativas para a estruturação populacional, as quais exigem um tempo computacional drasticamente menor, principalmente tratando-se de grande bancos de dados, como os gerados por genotipagem de SNPs por sequenciamento de nova geração.

Significância Os parâmetros determinados na análise estatística foram testados a um nível nominal de significância ($\alpha = 0,05$) para a rejeição da hipótese nula (H_0) e, para testes múltiplos, este valor foi corrigido de acordo com o método sequencial de Bonferroni (α') (Holm, 1979) de modo a evitarem-se erros do tipo I (falsa rejeição de H_0).

4 Resultados

4.1 Microssatélites *versus* SNPs

4.1.1 Diversidade e Estrutura Genética dos microssatélites

Um total de 60 espécimes de *An. darlingi* coletados em Remansinho, 59 em Granada e 58 em Cruzeiro do Sul foram genotipados para os 9 microssatélites *loci*. Essas amostras foram conduzidas à análise de diferenciação geográfica entre as três localidades citadas, e também, à comparação com uma técnica de genotipagem de SNPs, ddRADseq. Todos os microssatélites foram polimórficos nos três grupos analisados, sendo que o número de alelos por *locus* variou de 4 (ADSP2) a 43 (ADC01) (Tabela 3).

Tabela 3: Estimativas de R_s , H_e e F_{IS} dos microssatélites de *An. darlingi* em três localidades.

Localidade		Locus									
	N	ADSP2	ADMP9	ADC28	ADC29	ADC138	ADC02	ADC137	ADC110	ADC01	All Loci
Granada	R_s	3	11	9	17	10	16	16	16	31	14
	60 H_e	0,329*	0,808	0,823	0,895*	0,863*	0,830	0,869	0,912	0,953	0,809
	F_{IS}	0,394	0,027	0,069	0,181	0,432	0,056	-0,017	0,032	0,005	0,153
Remansinho	R_s	4	11	11	20	13	18	15	16	29	15
	60 H_e	0,253	0,753	0,841	0,843*	0,899*	0,857	0,859	0,875*	0,951	0,792
	F_{IS}	-0,121	0,022	0,109	0,347	0,416	-0,032	0,069	0,110	0,003	0,123
Cruzeiro do Sul	R_s	3	6	9	10	8	9	12	9	26	10
	58 H_e	0,232	0,718	0,699	0,784*	0,575*	0,812	0,863*	0,810	0,922	0,713
	F_{IS}	0,042	0,006	0,056	0,412	0,433	0,099	0,258	0,053	0,012	0,111

N= tamanho amostral; R_s = riqueza alélica por locus (baseada no mínimo populacional de 58 indivíduos); H_e = heterozigosidade esperada; F_{IS} = índice de fixação; em negrito: p valor significativo depois da correção de Bonferroni;

*presença de alelos nulos detectado pelo Micro-checker.

A riqueza alélica teve o menor e o maior valores endereçados aos mesmos *loci*, sendo 3,3 (ADSP2) e 30,5 (ADC01), respectivamente. O maior total de alelos ($A = 15$) e maior riqueza alélica ($R_s = 14,8$) foram encontrados no grupo de Remansinho, e o menores em Cruzeiro do Sul ($A = 10$, $R_s = 10,2$). A maior quantidade de alelos privados (P) para todos os *loci* foi encontrada também em Remansinho (Tabelas 3, 4). A heterozigosidade esperada (H_E) variou

de 0,232 para o *locus* ADSP2, em Cruzeiro do Sul, a 0,953 para o *locus* ADC01, em Granada (Tabela 3).

Tabela 4: Estimativas de alelos de microsatélites privados em *An. darlingi*.

	Número de alelos privados
Granada	8
Remansinho	15
Cruzeiro do Sul	8
Granada + Remansinho	46

Entre populações ($p=0.575$); Granada + Remansinho, ($p=0.002$)

Os resultados dos testes de desequilíbrio de ligação não apresentaram associação significativa entre nenhum par de *loci* analisado. Os maiores valores de F_{IS} foram encontrados para o *locus* ADC29 e ADC138 em todas os agrupamentos (Tabela 3). Desvios significativos do equilíbrio de HW foram detectados nos mesmos dois *loci* ($p < 0,0001$), bem como a presença de alelos nulos; portanto, ambos foram excluídos de análises posteriores.

Uma vez com 7 *loci*, na análise *locus* por *locus*, o *locus* ADC28 teve o maior F_{ST} (0,04; $p < 0,0001$) e o *locus* ADC01 teve o menor (0,019; $p < 0,0001$) (Tabela 5). As análises de índice de fixação par a par revelaram diferenciação significativa entre ambos os assentamentos rurais e Cruzeiro do Sul, e por outro lado um F_{ST} baixo e não significativo foi encontrado entre Granada e Remansinho ($F_{ST} = 0,003$; $p = 0,1$) (Tabela 6). A análise de componente principal (PCA) não formou agrupamentos delimitados, porém os indivíduos de Cruzeiro do Sul estão mais distantes e não se sobrepuseram tanto quanto os indivíduos de Granada e Remansinho (Figura 8C). Adicionalmente, a análise Bayesiana inferiu dois *clusters*, um que incluiu em sua maioria os indivíduos de Cruzeiro do Sul e outro os de Granada e Remansinho (Figura 8A).

Os resultados obtidos pelo teste de efeito de gargalo apresentaram excesso de homozigotos em Remansinho e Cruzeiro do Sul, o que poderia indicar uma retração populacional recente. Porém, essa alteração demográfica não foi consistente, uma vez que a probabilidade foi significativa somente pelo método S.M.M.

Tabela 5: Análise *locus* por *locus* dos microssatélites in *An. darlingi*.

<i>Locus</i>	Índice de Fixação	
	F_{ST}	<i>P</i> valor
ADSP2	0,02990	0,00684
ADMP9	0,02747	0,00000
ADC28	0,04538	0,00000
ADC02	0,02136	0,00098
ADC137	0,03106	0,00000
ADC110	0,02960	0,00000
ADC01	0,01902	0,00000

4.1.2 Diversidade e Estrutura Genética dos SNPs

Um total de 40 espécimes foram genotipados para SNPs, sendo 12 indivíduos de Granada, 14 de Remansinho e 14 de Cruzeiro do Sul. Essa abordagem foi realizada para mensurar diferenciação genética entre as três localidades baseada em dados de SNPs, bem como para comparação com os resultados obtidos por genotipagem de microssatélites. Uma média de 1.009.015,72 (intervalo: 609.048 – 2.752.364) sequências de 150 pares de base (pb) por indivíduo passaram pelo primeiro processo de qualidade (*process_radtags*) (Anexo- Tabela 8). O alinhamento ao genoma de *An. darlingi* (AdarC3 – 20/10/2015) reteve em média 32,2% (intervalo: 27,1% – 36,7%) das sequências, as quais deram origem à, em média, 17.265,97 (desvio padrão: +/- 6.977,24) genótipos *loci* por indivíduo. A taxa média de *loci* polimórficos foi de 23,47% (desvio padrão: +/- 2,65), ou seja 6.650,5 (intervalo: 2.509 – 11.810) SNPs por indivíduo (Anexo- Tabela 1). Ao final, 1.865 marcadores SNPs foram selecionados utilizando os parâmetros de seleção do programa *populations*, e portanto, cada *locus* esteve presente em pelo menos 50% dos indivíduos.

Na análise das três populações, os valores de F_{ST} par a par foram significativos entre Cruzeiro do Sul e ambos assentamento rurais (Tabela 6). Novamente, Remansinho mostrou-se o grupo com o maior número de alelos privados (478) e taxa de *loci* polimórficos (0,43) (Tabela 7). A análise Bayesiana revelou duas subpopulações distintas ($K=2$), agrupando Remansinho e Granada no mesmo *cluster* e Cruzeiro do Sul em outro (Figura 8B). Por outro lado, a PCA

Tabela 6: Valores de F_{ST} par a par para microssatélites e SNPs genotipados em *An. darlingi*.

	Granada	Remansinho	Cruzeiro do Sul
Granada		0,004	0,234**
Remansinho	0,003		0,198**
Cruzeiro do Sul	0,065*	0,058*	

* p valor significante

Tabela 7: Sumário dos dados de 1.865 SNPs (variáveis e fixos) nas populações de *An. darlingi*.

Localidade	No. indivs.	Privado	Posição	Variante	Polimórfico	% Polimorfismo
Granada	12	386	279843	1854	1067	0.3813
Remansinho	14	478	281349	1865	1211	0.4304
Cruzeiro do Sul	14	223	186883	1236	452	0.2419

baseado na diversidade genética dos SNPs entre os indivíduos separou as três localidades, sendo que 22,5% de todos marcadores explicam a separação entre os assentamentos rurais e Cruzeiro do Sul, e da mesma maneira 4,9% entre Granada e Remansinho (Figura 8D). Tendo em vista a discriminação populacional encontrada na PCA, os dados foram conduzidos a DAPC, a fim de verificar o agrupamento entre Granada e Remansinho (Figura 9). Previamente, o número ótimo de *clusters* foi calculado com o Critério de Informação Bayesiana (BIC), o qual é definido pelo menor valor (o “cotovelo” do gráfico) (Figure 9A). As figura 9B e 9C mostraram a separação entre os três grupos, em uma função discriminante e em duas funções discriminantes, respectivamente.

A clusterização hierárquica apresentou resultados semelhantes: a primeira ramificação ocorreu entre Cruzeiro do Sul e os assentamentos rurais, e logo entre Granada e Remansinho; no entanto, ocorreu ainda que pequena a presença de alguns indivíduos juntos à outros grupos (Figura 9D).

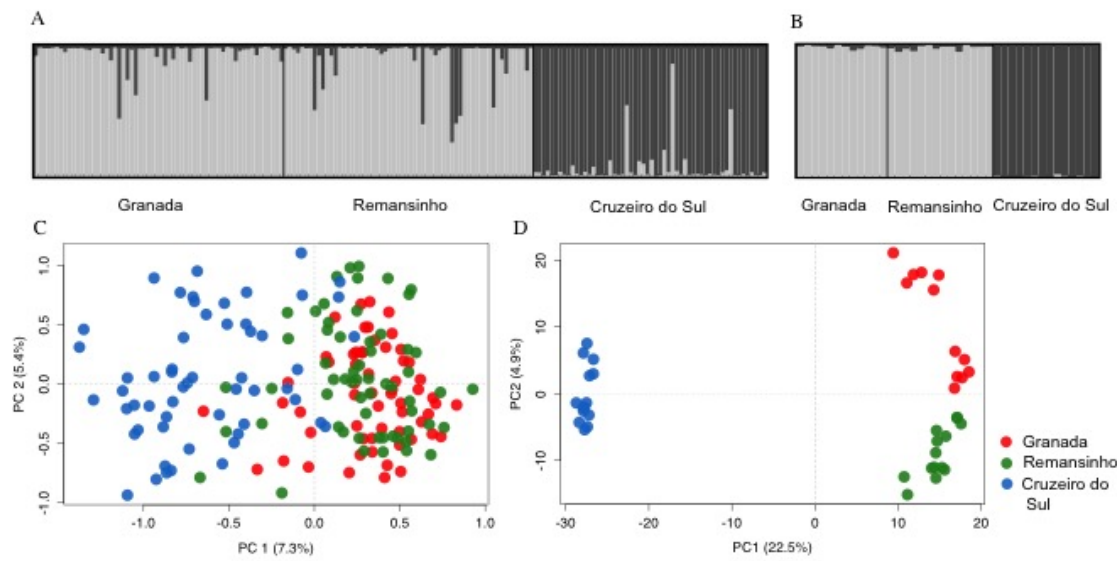


Figura 8: **Análise Bayesiana (BA) e Análise de Componente Principal (PCA) de genótipos de *An. darlingi* das três localidades.** (A) Resultados da BA dos microssatélites ($K=2$), cada coluna representa um indivíduo, divididos em três localidades (Granada, Remansinho e Cruzeiro do Sul) e a proporção que pertence a cada *cluster* (cinza e preto). (B) Resultados da BA dos SNPs ($K=2$). (C) PCA dos 9 marcadores microssatélites, cada ponto representa um indivíduo. (D) PCA de 1865 SNPs. (C-D) as cores correspondem as populações assinaladas: Granada, vermelho; Remansinho, verde e Cruzeiro do Sul, azul.

4.1.3 Comparação entre as duas técnicas

A figura 8 evidenciou as diferenças encontradas entre as duas técnicas utilizadas, no que diz respeito ao potencial genético discriminatório entre as três localidades. Além da separação mais nítida e bem definida entre os indivíduos de cada localidade, a PCA baseada nos dados de SNPs teve uma porcentagem maior indicando a divisão entre elas. Ademais, as estimativas do F_{ST} par a par foram mais altas utilizando o conjunto de SNPs encontrados. Os resultados das análises Bayesianas também demonstraram maior acurácia na corrida com dados de SNPs do que com microssatélites; no primeiro, quase nenhuma mistura entre os dois *clusters* foi encontrada. Por outro lado, nessa mesma análise, não foi possível verificar uma divisão entre Granada e Remansinho, o que pode indicar a panmixia entre elas ou uma questão hierárquica de análise, uma vez que Cruzeiro do Sul mostrou-se consistentemente diferente em relação as

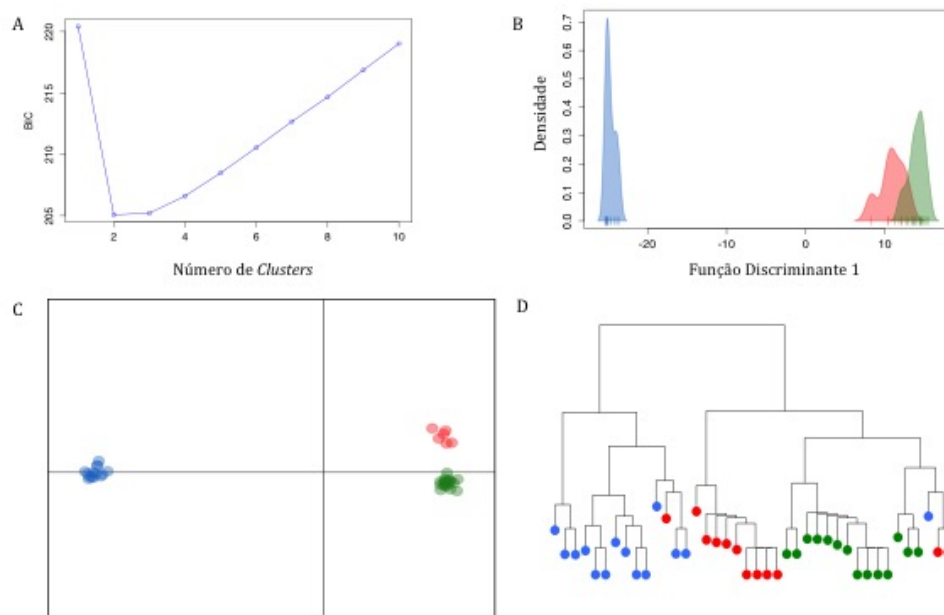


Figura 9: **Sumário da Análise Discriminante de Componente Principal (DAPC) e clusterização hierárquica em *An. darlingi*.** (A) Valores do Critério de Informação Bayesiana (BIC) para cada $K(1-10)$, o melhor valor de K foi 3. (B) Três *clusters* que separam as localidades em um eixo, a área representa a densidade de indivíduos, nota-se uma sobreposição entre Granada (vermelho) e Remansinho (verde); Cruzeiro do Sul em azul. (C) Três *clusters* que separam as localidade em dois eixos.(D) Clusterização hierárquica por Ward (1963), as cores correspondem a origem de coleta dos indivíduos.

outras duas localidades (Figura 9). Análises com indivíduos apenas de Granada e Remansinho foram conduzidas posteriormente, com a finalidade de resolver essa questão.

4.2 Geografia e Sazonalidade

Nesta seção as análises foram conduzidas de quatro maneiras diferentes: 1) comparação geográfica entre Granada e Remansinho em três meses diferentes (Março, Maio e Novembro de 2012); 2) comparação sazonal entre todos os meses de Granada; 3) comparação sazonal entre todos os meses de Remansinho; 4) comparação sazonal entre todos os meses de ambas as localidades.

4.2.1 Análise microgeográfica por mês de coleta

Março de 2012 Foram genotipados 25 espécimes de *An. darlingi* dos assentamentos de Granada (11) e Remansinho (14) coletados em Março de 2012. Uma média de 816.200,12 (intervalo: 555.996-1.262.460) sequências de 150 pb foram alinhadas ao genoma de *An. darlingi* e resultaram em 13.465 (desvio padrão: $\pm 2.222,75$) genótipos *loci* por indivíduo. Em cada indivíduo, 35,77% (desvio padrão: $\pm 8,42$) dos *loci* foram heterozigotos (Anexo-Tabela 9). Para selecionar os SNPs o mesmo processo de filtragem foi utilizado, o qual resultou em 1.941 marcadores que foram incluídos para a PCA (Figura 10A), onde as duas localidades foram separadas. O mesmo conjunto de SNPs foi utilizado para a DAPC, na qual o menor BIC e consequentemente o número ótimo de *clusters* foi 2 (Figura 10B). Cada cluster conteve indivíduos de uma das duas localidades (Figura 10C). A Análise Bayesiana revelou $K = 2$, quando todos os SNPs foram utilizados, porém não foi identificada estruturação (Figura 10D-acima). Em um segundo filtro foram selecionados 545 marcadores, os quais apresentavam F_{ST} maior ou igual a 0,05 entre as duas localidades e, neste caso, foi identificada estruturação populacional (Figura 10D-abixo).

Mai de 2012 Foram genotipados 27 espécimes de *An. darlingi* dos assentamentos de Granada (13) e Remansinho (14) coletados em Maio de 2012. Uma média de 1.714.548,15 (intervalo: 666.662-3.957.660) sequências de 150 pb alinhadas ao genoma de *An. darlingi* e resultaram em 17.127 (desvio padrão: $\pm 6.141,58$) genótipos *loci* por indivíduo. Em cada indivíduo, 34,78% $\pm 7,67$ dos *loci* foram heterozigotos (Anexo- Tabela 10). Um total de 2.048 SNPs foram identificados em pelo menos 50% dos indivíduos e foram utilizados para a PCA e DAPC, onde os indivíduos das duas localidades foram separados (Figura 10A,B,C). A Análise Bayesiana revelou $K = 2$, com o mesmo conjunto de SNPs, e novamente, não foi identificada estruturação (Figura 10D-acima). Após realizar o mesmo processo de seleção dos SNPs segundo o valor de seu F_{ST} associado, 405 SNPs restaram e a estruturação populacional foi identificada entre Granada e Remansinho (Figura 10D-abixo).

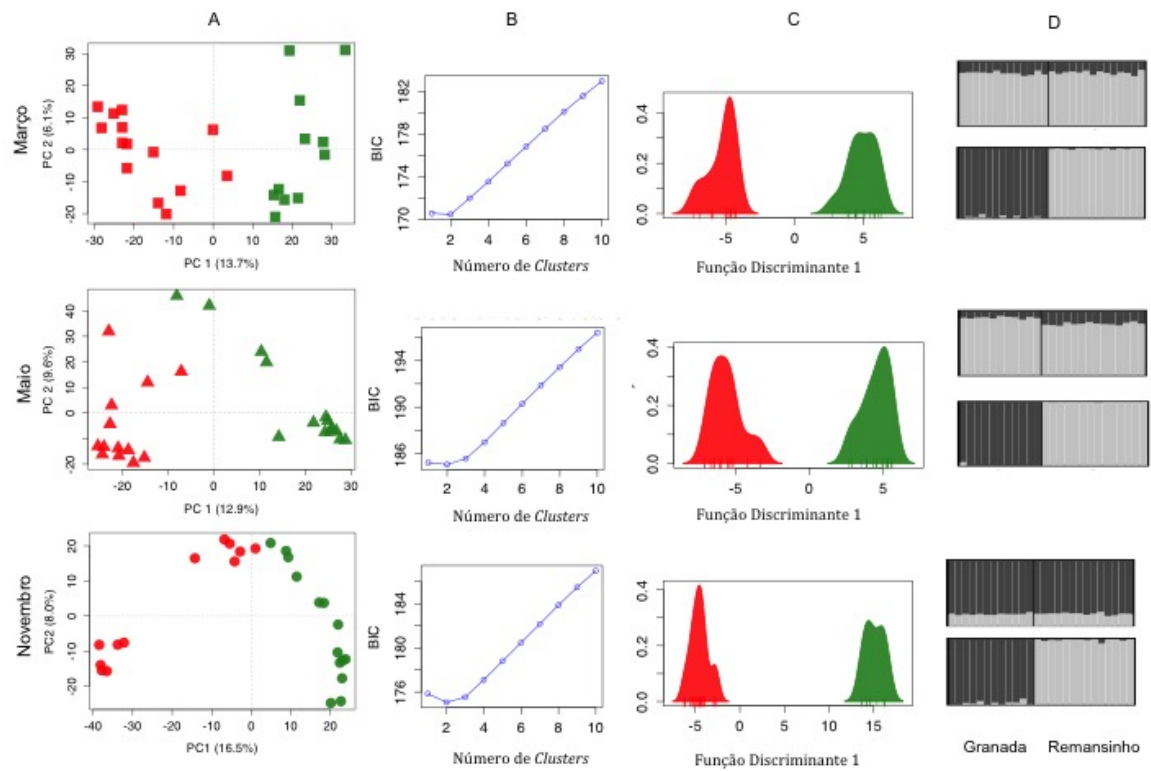


Figura 10: **Análise de Componente Principal (PCA), Análise Bayesiana (BA) e Análise Discriminante de Componente Principal (DAPC) de genótipos de *An. darlingi* coletados em Granada e Remansinho em três meses de 2012.** (A) PCA dos espécimes genotipados em Granada (vermelhos) e Remansinho (verdes); cada formato representa um indivíduo, e os diferentes formatos correspondem ao mês de coleta: Março (quadrados); Maio (triângulos) e Novembro (pontos). (B) Valores do Critério de Informação Bayesiana (BIC) para cada K , melhor K foi 2. (C) Dois *clusters* separaram *An. darlingi* de Granada (vermelho) e Remansinho (verde) apenas com informação do primeira função discriminante (DAPC), área representa a densidade de indivíduo. (D) Acima: resultado da BA, o melhor K foi 2 para todas as análises; cada coluna representa um indivíduo e sua porcentagem pertencente a determinado *cluster* (cinza e preto); abaixo: resultado da BA após filtragem pelo valor de F_{ST} entre as populações.

Novembro de 2012 Os 26 espécimes genotipados de *An. darlingi* dos assentamentos de Granada (12) e Remansinho (14) coletados em Novembro de 2012 foram os mesmos utilizados na primeira análise, com Cruzeiro do Sul (Tabela Anexo 1). Após a filtragem, 1.872 SNPs foram identificados e utilizados para a PCA (Figura 10A), onde as duas localidades foram separadas. A DAPC foi realizada com o mesmo conjunto de marcadores e o número ótimo

de *clusters* foi 2 novamente, e espécimes de Granada e Remansinho foram agrupados separadamente (Figura 10B,C). A análise Bayesiana não apresentou estruturação quando todos os SNPs foram utilizados, porém quando os marcadores de maior F_{ST} associado (326 SNPs), foi identificada estruturação populacional (Figura 10D).

4.2.2 Análise sazonal em Granada

Para a seguinte análise, as 11 espécimes de *An. darlingi* coletadas em Granada em Março de 2012, mais 13 espécimes de Maio e 12 espécimes de Novembro do mesmo ano foram utilizadas. Um total de 2.791 SNPs foram detectados em pelo menos 50% desses 36 indivíduos. Na PCA, os três meses formaram três grupos separados, o primeiro eixo de diferenciação foi entre o primeiro semestre (Março e Maio) e o segundo (Novembro); e no segundo eixo da análise, os meses de Março e Maio também foram distinguidos entre eles (Figura 11A).

4.2.3 Análise sazonal em Remansinho

As 14 espécimes de *An. darlingi* coletadas em Remansinho em Março de 2012, foram adicionadas as mesmas quantidades de espécimes de Maio e Novembro do mesmo ano. Um total de 3.690 SNPs foram detectados em pelo menos 50% dos 42 indivíduos. A PCA revelou, no primeiro eixo, uma divisão semestral, ou seja, meses do primeiro e segundo semestres (Figura 11B).

4.2.4 Análise sazonal em ambas as localidades

Os dados dos três meses de Granada e Remansinho foram agrupados, totalizando 78 genótipos. Neste caso, 3.358 SNPs foram identificados em pelo menos 50% dos indivíduos e foram utilizados para a PCA (Figura 11C). Ainda na figura 11C, o primeiro eixo separou os indivíduos coletados em Novembro dos dois outros meses, Março e Maio. A DAPC e a clusterização hierárquica também confirmaram essa divisão sazonal, entre o primeiro e segundo semestre (Figura 12).

Um conjunto de *loci* com F_{ST} maior ou igual a 0,01 foi utilizado no diagrama de Venn, o

qual mostrou que grande parte dos marcadores informativos que diferenciaram a sazonalidade (primeiro e segundo semestres), também responderam às diferenças geográficas entre Granada e Remansinho (Figura 13A). As porcentagens de *loci* polimórficos foram maiores nos meses do assentamento de Remansinho, bem como o número de alelos privados (Tabela 8).

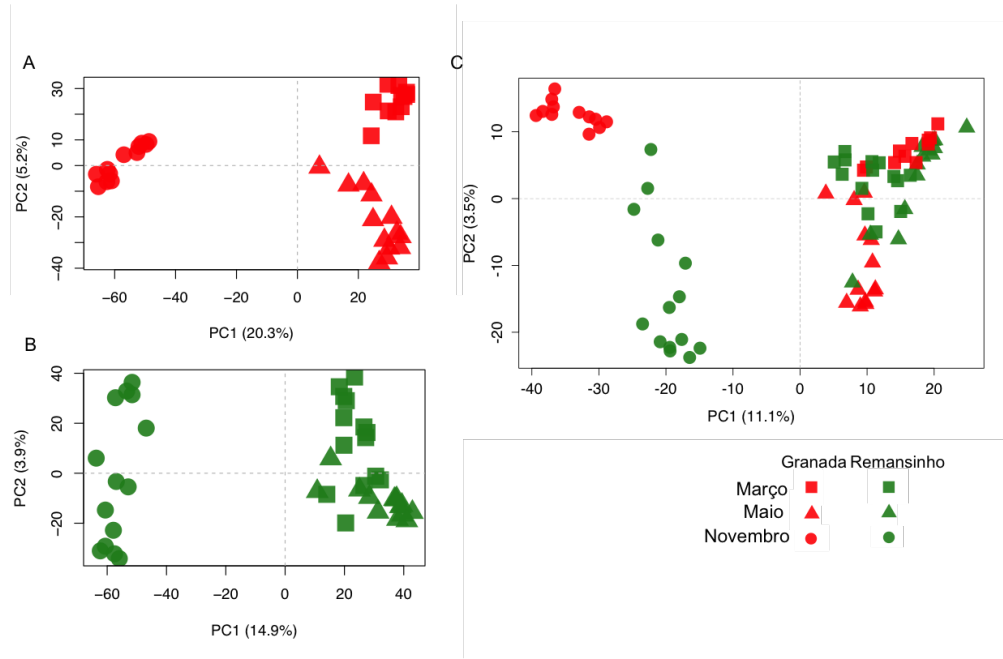


Figura 11: **Análise de Componente Principal (PCA) de genótipos de *An. darlingi* coletados em Granada e Remansinho em três meses diferentes.** (A) PCA dos espécimes de Granada, os formatos correspondem ao mês de coleta: Março (quadrados), Maio (triângulos) e Novembro (pontos). (B) PCA dos espécimes de Remansinho em três meses. (C) PCA dos espécimes coletados nos três meses nas duas localidades.

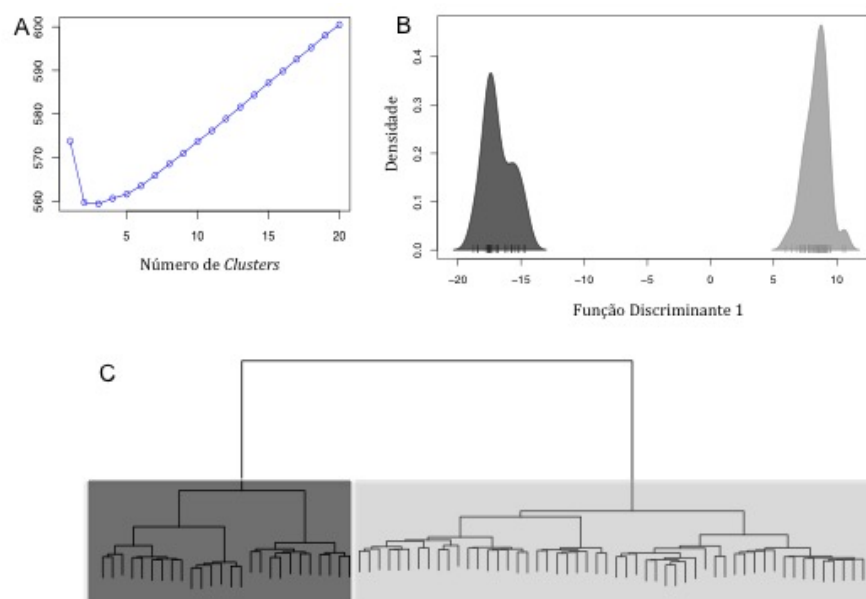


Figura 12: Sumário da Análise Discriminante de Componente Principal (DAPC) e clusterização hierárquica em *An. darlingi* coletados no primeiro e segundo semestre. (A) Valores do Critério de Informação Bayesiana (BIC) para cada $K(1-10)$, o melhor valor de K foi 2. (B) Dois *clusters* que separam as localidades em um eixo, a área representa a densidade de indivíduos coletados no primeiro semestre (cinza escuro) e segundo (cinza claro). (C) Clusterização hierárquica por Ward (1963), a primeira derivação foi entre mosquitos coletados nos diferentes semestres.

Tabela 8: Sumário dos dados SNPs (variáveis e fixos) de *An. darlingi* de Granada e Remansinho em três meses diferentes.

Localidade	Mês	No. indivs.	Privado	Posição	Variante	Polimórfico	% Polimorfismo
Granada	Março	11	255	508102	3352	926	0.1822
	Maio	13	241	508707	3357	1199	0.2357
	Novembro	12	256	344337	2262	776	0.2254
Remansinho	Março	14	325	508557	3357	1390	0.2733
	Maio	14	382	508708	3358	1507	0.2962
	Novembro	14	351	432566	2841	1089	0.2518

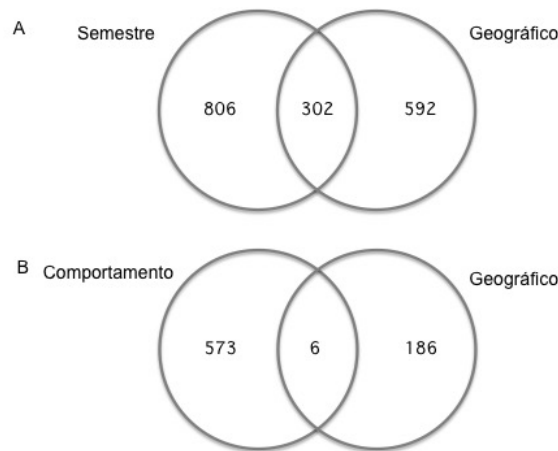


Figura 13: **Diagrama de Venn.** (A) *Loci* compartilhados entre a comparação semestral, ou seja, espécimes coletados no primeiro e no segundo semestre, e geográfica, sendo todos os meses de Granada vs. todos os meses de Remansinho. (B) *Loci* compartilhados entre a comparação de *An. darlingi* endofágico vs. exofágico coletados em Granada e Remansinho em Abril de 2011, e geográfica entre espécimes de ambos locais do mesmo mês.

4.3 Comportamento Hematofágico

Nesta seguinte seção o comportamento hematofágico do vetor *An. darlingi* foi analisado em diferentes perspectivas: 1) espécimes do mosquito coletadas dentro de casa (endofágicas) *versus* fora da casa (exofágicas), considerando até 10 metros de distância da casa, em Remansinho, no intervalo das 18-21 horas; 2) a mesma análise de (1), porém de outra localidade, Granada; 3) grupos de (1) e (2) analisados em conjunto; 4) espécimes exofágicas coletadas em Granada às 18-21 horas (crepuscular) *versus* 3-6 horas (matinal); 5) mesma análise de (4), porém com mosquitos endofágicos; 6) grupos de (4) e (5) analisados em conjunto.

4.3.1 Exofagia vs. Endofagia

Granada Foram genotipados espécimes de *An. darlingi* coletados no intradomicílio (endofágicos) e peridomicílio (exofágicos), em Abril de 2011, em Granada. A Tabela 12, em anexo, apresentou o resultado da genotipagem de SNPs de 54 espécimes *An. darlingi* exofágicos. Em média 612.128,67 (intervalo: 232.796- 2.491.061) sequências alinharam-se ao genoma de *An. darlingi*, e 7.121,78 (desvio padrão: +/- 3141,72) genótipos *loci* por indivíduo foram obtidos.

Em cada indivíduo, uma média de 20,41% $\pm$ 8,03 dos *loci* foram heterozigotos (Tabela 13). A Tabela 13, em anexo, apresentou o resumo do sequenciamento de 42 espécimes *An. darlingi* endofágicos. Nesse grupo uma média de 965.033,62 (intervalo: 485.453- 3.009.945) sequências alinharam-se ao genoma de *An. darlingi* e resultaram em 12.034,4 (desvio padrão: $\pm$ 3.331,72) genótipos *loci* por indivíduo. Em cada indivíduo em média 27,29% $\pm$ 1,97 dos *loci* foram heterozigotos. A PCA foi realizada com 498 SNPs selecionados, e aproximadamente 20% desses marcadores explicaram as diferenças entre os dois grupos.

Remansinho Um total de 32 espécimes de *An. darlingi* coletados em Abril 2011 em Remansinho foram genotipados. Uma média de 974.466,78 (intervalo: 602.930- 1.650.403) sequências de 150 pb foram alinhadas ao genoma de *An. darlingi* e resultaram em 12.617,69 (desvio padrão: $\pm$ 3.324,90) genótipos *loci* por indivíduo (Anexo- Tabela 11). Em cada indivíduo, 29,62% $\pm$ 2,22 dos *loci* foram heterozigotos. A PCA foi realizada com um total de 1.478 SNPs, e cerca de 10% desses marcadores estiveram relacionados com a diferença entre endofágicos e exofágicos (Figura 14B).

Ambas as localidades A PCA dos genótipos dos quatro grupos (Granada- endofágico, Granada-exofágico, Remansinho - endofágico , Remansinho – exofágico) foi realizada com um total de 859 marcadores (Figura 14C). O mesmo conjunto de SNPs foi utilizado para a DAPC, na qual o menor valor de BIC foi 3 (Figura 15A). Os espécimes de *An. darlingi* endofágicos de Granada e Remansinho agruparam-se tanto na PCA quanto na DAPC, neste último, o *cluster* 3 os agrupou quase que em sua totalidade (Figura 14C,15). Por outro lado, os grupos exofágicos encontraram-se separados, tanto na PCA quanto na DAPC (Figura 14C,15). Os dois grupos endofágicos mostraram maior porcentagem de polimorfismo nos *loci* identificados (Tabela 9).

Na figura 13B, o diagrama de Venn, com o conjunto de *loci* com F_{ST} maior ou igual a 0,01, exibiu que somente 6 desses marcadores informativos explicaram as diferenças geográficas entre Granada e Remansinho em Abril de 2011 e as diferenças ligadas ao comportamento hematofágico do vetor no mesmo mês citado.

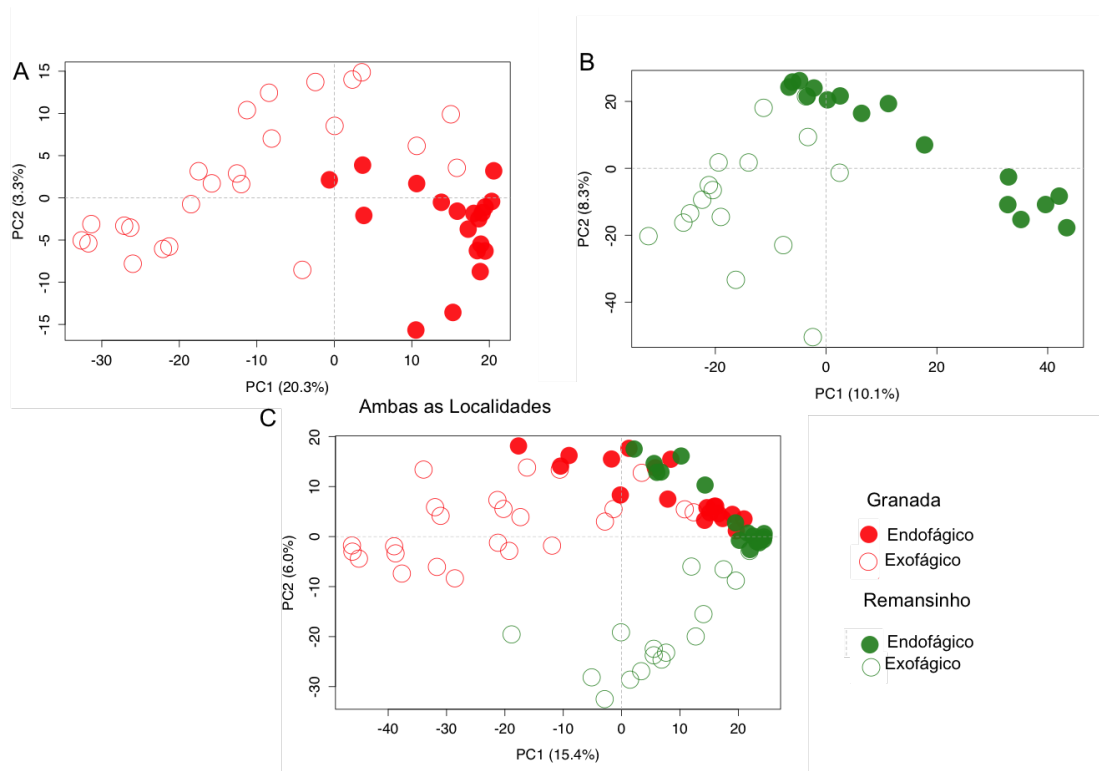


Figura 14: Análise de Componente Principal (PCA) de genótipos de *An. darlingi* endo/exofágicos de ambos assentamentos rurais. (A) PCA de *An. darlingi* endofágicos (pontos vermelhos preenchidos) e exofágicos (pontos vermelhos vazios) em Granada. (B) PCA de *An. darlingi* endofágicos (pontos verdes preenchidos) e exofágicos (pontos verdes vazios) em Remansinho. (C) PCA dos 4 grupos em (A) e (B).

Tabela 9: Sumário dos dados SNPs (variáveis e fixos) de *An. darlingi* endo/exofágicos de ambos os assentamentos rurais.

Localidade	Tipo	Privado	Posições	Variante	Polimórfico	% Polimorfismo
Granada	Exofágico	121	130009	859	348	0.2677
	Endofágico	114	128650	850	459	0.3568
Remansinho	Exofágico	82	129858	858	391	0.3011
	Endofágico	115	130009	859	475	0.3654

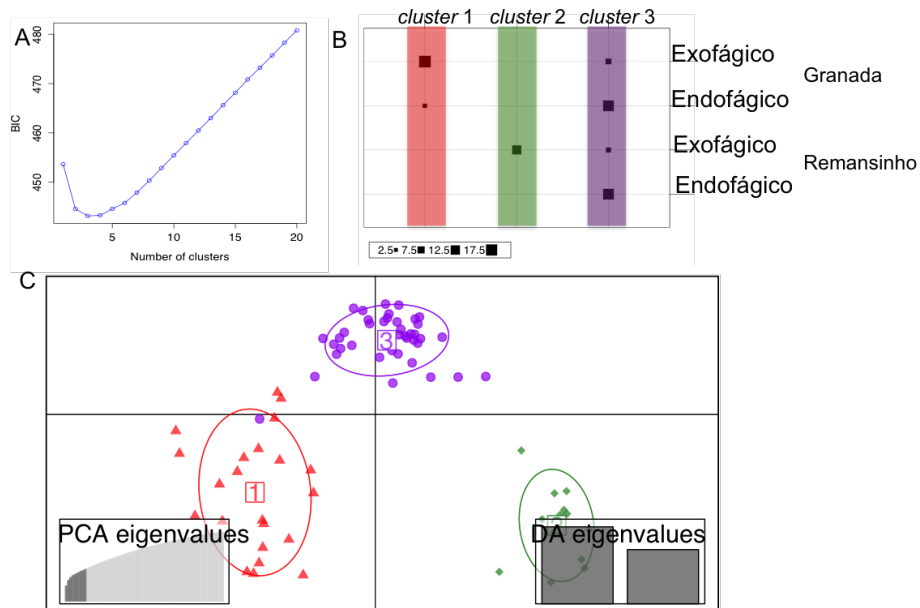


Figura 15: **Sumário da Análise Discriminante de Componente Principal (DAPC) em *An. darlingi* endo/exofágicos de ambos assentamentos rurais.** (A) Valores de Critério de Informação Bayesiana (BIC) para cada K , o menor valor foi 3. (B) Proporção de indivíduos dos 4 grupos incluídos nos três *clusters* inferidos (vermelho, verde e roxo). (C) Ordem dos três *clusters* em dois eixos, cada ponto representa um indivíduo que pertence ao determinado *cluster*; as caixas representam o *eigenvalues* para PCA e DA.

4.3.2 Crepuscular vs. Matinal: Granada

Exofágico Os espécimes de *An. darlingi* coletados durante o primeiro pico de densidade do vetor, 18-21 horas (crepúsculo), foram comparados com espécimes do terceiro pico de densidade, 3-6 horas (amanhecer) (Figura 16A). Para a PCA, 54 genótipos foram utilizados e um conjunto de 175 SNPs, identificados em 50% dos indivíduos (Figura 17). O mesmo conjunto foi utilizado para a DAPC, na qual o menor valor de BIC foi 3 (Figura 18A). Tanto na PCA quanto na DAPC, não foi evidenciado diferenciação entre mosquitos exofágicos coletados no crepúsculo e pela manhã.

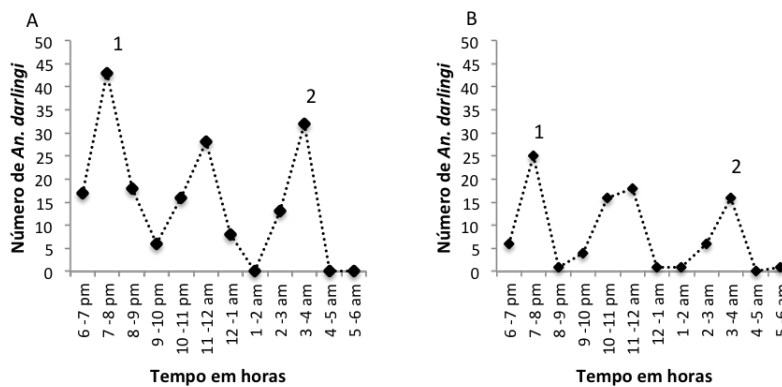


Figura 16: Densidade *An. darlingi* coletados em Granada das 6 pm até 6 am. (A) Coleta no peridomicílio. (B) Coleta no intradomicílio. Os números 1 e 2 indicam os dois picos de densidade analisados, crepúsculo e amanhecer.

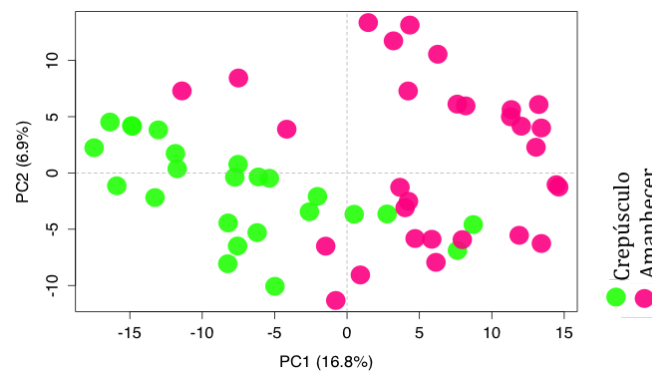


Figura 17: Análise de Componente Principal (PCA) com SNPs de *An. darlingi* exofágicos de Granada. (A) PCA de 175 SNPs genotipados em *An. darlingi* exofágicos coletados das 18-21 horas (crepúsculo) pontos verdes; e das 3-6 horas (amanhecer) pontos rosas.

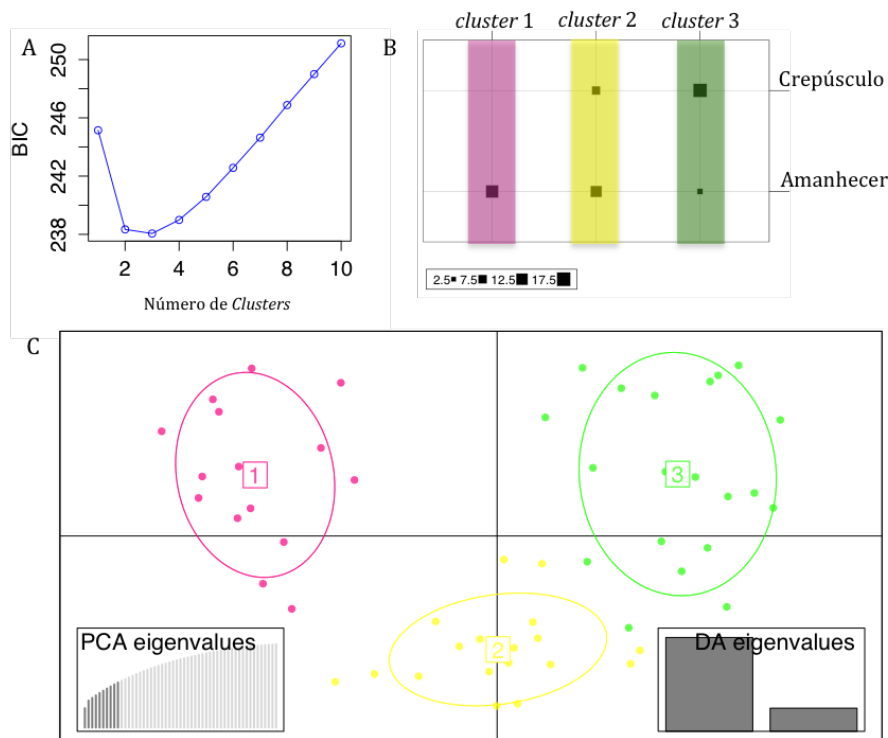


Figura 18: **Sumário da Análise Discriminante de Componente Principal (DAPC) em *An. darlingi* exofágicos em Granada.** (A) Valores de Critério de Informação Bayesiana (BIC) para cada K , o menor valor foi 3. (B) Mesa com proporção de indivíduos dos 2 grupos incluídos nos três *clusters* inferidos (rosa, amarelo e verde). (C) Ordem dos três *clusters*, cada ponto representa um indivíduo que pertence ao determinado *cluster*; as caixas representam o *eigenvalues* para PCA e DA.

Endofágico Os espécimes de *An. darlingi* coletados dentro da casa (endofágicos) durante o primeiro pico de densidade do vetor, 18-21 horas (crepúsculo), foram comparados com espécimes do terceiro pico de densidade, 3-6 horas (amanhecer) (Figura 16B). Para a PCA, 42 genótipos foram utilizados e um conjunto de 1.228 SNPs foram identificados (Figura 19). A DAPC foi feita com o mesmo conjunto de SNPs e o menor valor de BIC foi 2 (Figura 20A). Ao contrário da análise anterior, a PCA e DAPC discriminaram os mosquitos endofágicos coletados no crepúsculo e pela manhã.

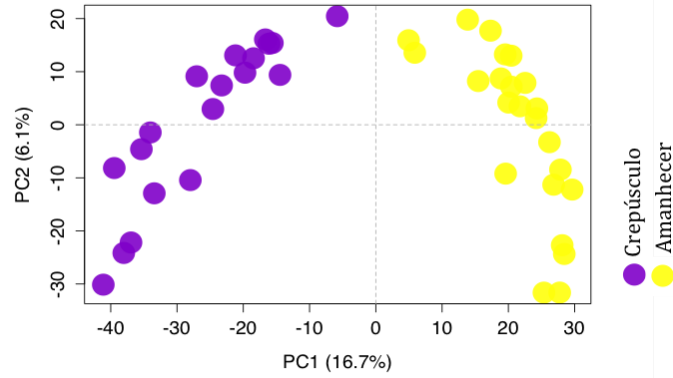


Figura 19: **Análise de Componente Principal (PCA) com SNPs de *An. darlingi* endofágicos de Granada.** (A) PCA de 1.228 SNPs genotipados no intradomicílio das 18-21 horas (crepúsculo) pontos roxos; das 3-6 horas (amanhecer) pontos amarelos.

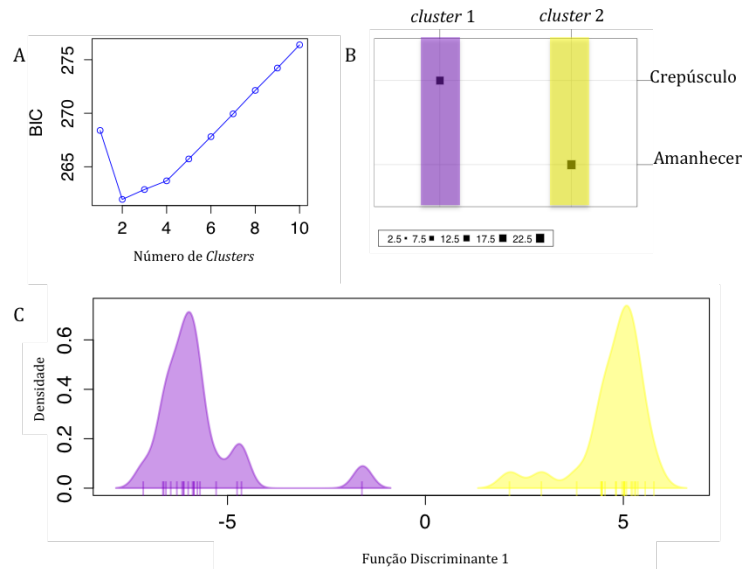


Figura 20: **Sumário da Análise Discriminante de Componente Principal (DAPC) em *An. darlingi* endofágicos em Granada.** (A) Valores de Critério de Informação Bayesiana (BIC) para cada K , o menor valor foi 2. (B) Mesa com proporção de indivíduos dos 2 grupos incluídos nos dois *clusters* inferidos (roxo e amarelo). (C) Ordem dos dois *clusters* em um eixo, a área representa a densidade de indivíduos que pertencem ao determinado *cluster*.

Tabela 10: Sumário dos dados SNPs (variáveis e fixos) de *An. darlingi* endo/exofágicos crepuscular e matinal.

	Tipo	Privado	Posições	Variante	Polimórfico	% Polimorfismo
Exofágico	Crepúsculo	70	83956	556	190	0.2263
	Amanhecer	91	84107	557	291	0.346
Endofágico	Crepúsculo	55	83956	556	243	0.2894
	Amanhecer	85	84107	557	315	0.3745

Exofágico e Endofágico As amostragens anteriores de espécimes de *An. darlingi* endofágicos e exofágicos foram juntamente analisadas. Portanto, para a PCA, 96 genótipos foram utilizados e um conjunto de 557 SNPs foram identificados em 50% desses indivíduos (Figura 21). Nesta análise, o primeiro eixo diferenciou majoritariamente espécimes endofágicos de exofágicos, explicado por 16,5% dos marcadores; no segundo eixo, a diferenciação foi em relação ao horário da hematofagia, explicado por 7,5% do conjunto de SNPs. Da mesma maneira, a clusterização hierárquica apresentou uma primeira divisão entre mosquitos endo/exofágicos, e posteriormente, entre hábito hematofágico crepuscular e matinal (Figura 21B). Os espécimes de *An. darlingi* tanto endofágicos quanto exofágicos com comportamento matinal tiveram porcentagem de polimorfismo maiores, bem como número de alelos privados (Tabela 10).

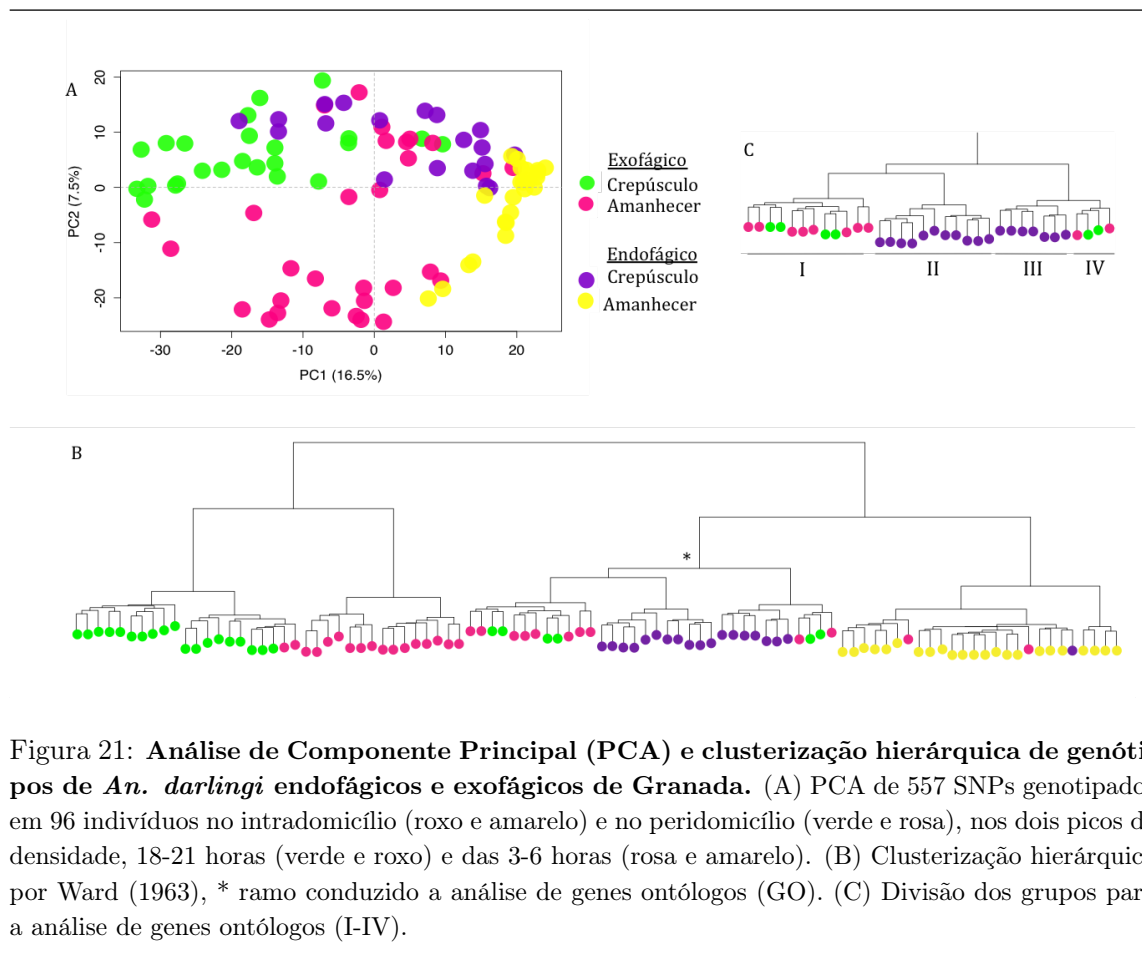


Figura 21: **Análise de Componente Principal (PCA) e clusterização hierárquica de genótipos de *An. darlingi* endofágicos e exofágicos de Granada.** (A) PCA de 557 SNPs genotipados em 96 indivíduos no intradomicílio (roxo e amarelo) e no peridomicílio (verde e rosa), nos dois picos de densidade, 18-21 horas (verde e roxo) e das 3-6 horas (rosa e amarelo). (B) Clusterização hierárquica por Ward (1963), * ramo conduzido a análise de genes ontólogos (GO). (C) Divisão dos grupos para a análise de genes ontólogos (I-IV).

A análise de genes ontólogos (GO) foi conduzida entre espécimes de *An. darlingi* que encontraram-se próximos na clusterização hierárquica, porém com fenótipos diferentes; pois, dessa maneira, analisamos indivíduos similares geneticamente e, portanto, as diferenças encontradas na análise de GO seriam relacionadas com o fenótipo. As duas análises foram entre os mosquitos exofágicos dos dois horários e endofágicos coletados no crepúsculo (Figura 21C). A primeira comparação, entre os grupos I *vs.* II, apresentou enriquecimento significativo de dois grupos de processos biológicos, processo metabólico em geral (GO: 0008152; número esperado (ne) = 22,94; número obtido (no) = 31; $p = 0,018$) e processo metabólico de substâncias orgânicas (GO: 0071704; ne = 17,75; no = 25; $p = 0,026$). Já na segunda comparação, entre os grupos III *vs.* IV, apesar do processo metabólico ter tido 64 genes relacionados, o enriquecimento não foi estatisticamente significativo ($p = 0,06$). No entanto, dois outros grupos de

GO foram significativos, processo metabólico de amino ácido celular (GO:0006520; $ne = 1,51$; $no = 5$; $p = 0,017$) e de ácido graxo (GO:0006631; $ne = 0,7$; $no = 3$; $p = 0,03$). Demais comparações entre indivíduos mais distantes na clusterização hierárquica foram realizadas, porém nenhum enriquecimento para GO foi significativo.

5 Discussão

Estudos que geo-referenciam e identificam focos de casos de malária podem representar um incremento junto aos esforços para a eliminação dessa doença (Bousema et al., 2010; Valle & Lima, 2014). Dessa maneira, correlações locais podem ser detectadas, destacando características peculiares, e conseqüentemente, tornando o controle da transmissão da malária mais efetivo. Correlações positivas foram encontradas entre a proximidade das habitações humanas dos potenciais criadouros dos vetores anofelinos, com o risco de contrair malária (Staedke, 2003; Clark et al., 2008). Áreas de mineração, bem como campos de exploração madeireira e assentamentos rurais também encontraram-se correlacionados com risco a essa doença, pois apresentaram aumento da densidade vetorial (de Castro et al., 2006, 2007; da Silva-Nunes et al., 2008). Portanto, destacamos a importância de um estudo a nível microgeográfico da dinâmica vetorial e transmissão de malária, assim como realizado em outros estudos (Scarpassa & Conn, 2007; Dery et al., 2010; Moutinho et al., 2011).

As análises com os nove marcadores microssatélites discriminaram os espécimes de *An. darlingi* com distâncias geográficas maiores ($\sim 600\text{km}$), ou seja, Cruzeiro do Sul de ambos os assentamentos rurais. Em contrapartida, Granada e Remansinho, os quais são separados por aproximadamente 60 km, apresentaram populações geneticamente similares. Com a finalidade de confirmar se amostras de *An. darlingi* coletadas em Granada e Remansinho formam realmente uma única população ou, se essa observação foi resultante de uma limitação da técnica de microssatélites empregada, SNPs foram genotipados nas amostras dos mesmos locais através da técnica de ddRADseq. Conseqüente, a nítida diferenciação entre os dois assentamentos rurais obtidos com dados dos SNPs na Análise de Componente Principal (PCA) em conjunto com Cruzeiro do Sul (Figura 8), demonstrou que microssatélites são limitantes para acessar subdivisão populacional a nível microgeográfico em *An. darlingi*.

Além disso, os espécimes coletados em Granada e Remansinho, em três meses diferentes, foram consistentemente separados em dois grupos independentes (Figuras 10). A abordagem nas análises Bayesianas que incluiu somente os SNPs informativos na diferenciação das populações, ou seja, $F_{ST} \geq 0,05$, distinguiu novamente, desta vez com um modelo biológico

genético-populacional que assume equilíbrio de Hardy-Weinberg (HWE) e desequilíbrio de ligação (LD), *An. darlingi* de Granada e Remansinho (Figuras 10D). Essa separação pode não ser absoluta, no entanto, o fluxo genético entre elas justificado pela proximidade geográfica não seria relevante. Tendo em vista essa organização subpopulacional de *An. darlingi*, que adapta-se às características típicas de cada ambiente, a necessidade de um controle específico que leve em consideração essas diferenças focais é reforçada.

O regime de chuvas foi considerado o principal fator ecológico que influencia na distribuição e abundância de *An. darlingi*, o qual tem apresentado densidade sazonal de acordo, especialmente, com a disponibilidade de habitats larvais (Sinka et al., 2010; Hiwat & Bretas, 2011). Em áreas ribeirinhas e rurais a densidade desse mosquito coincide com o período de chuvas (Gil et al., 2003), já nas áreas urbanas e semi-urbanas ou a densidade vetorial ocorre na transição do período chuvoso para o período seco e vice-versa (Gil et al., 2007; Souza-Santos, 2002), ou permanece constante, independente da pluviosidade, devido aos criadouros artificiais (Gil et al., 2003; Moutinho et al., 2011).

Em espécimes de *An. darlingi* coletados próximos à Porto Velho (Rondônia, Brasil), ao longo do Rio Madeira, foram encontradas duas subpopulações geneticamente distintas adaptadas a diferentes períodos de chuvas (Angella et al., 2014). Baseado nos dados de SNPs, o presente estudo corroborou a existência dessas subpopulações de *An. darlingi*, uma vez que espécimes coletados em Novembro, durante a estação seca, diferenciaram-se dos espécimes coletados nos outros dois meses, Março e Maio, durante a estação chuvosa. Assim, como descrito por Angella et al. (2014), a subpopulação da estação seca pode estar mais adaptada a criadouros permanentes, comumente criados pelo homem. Estes criadouros, diferentemente dos criadouros naturais, não são afetados dramaticamente pelas flutuações sazonais das chuvas. A influência humana pode, portanto, alterar a estruturação populacional de *An. darlingi* com decorrentes implicações epidemiológicas, uma vez que a subpopulação da estação seca pode promover a constante transmissão de malária ao longo do ano.

Apesar de complexa, a relação entre o grau de desflorestação e a presença de criadouros larvais de *An. darlingi*, bem como aumento da taxa de picada na região Amazônica peru-

ana, foram positivamente associados (Vittor et al., 2006; 2009). Assentamentos rurais estão sujeitos a gradual e contínua mudança da paisagem por intervenção humana, como o desenvolvimento agrícola, exploração mineral, degradação da floresta nativa e aumento da população humana (Rozendall, 1990; de Castro et al., 2006; Sinka et al., 2010; Moutinho et al., 2011). Tais mudanças ambientais combinadas com a variação climática sazonal, como temperatura, pluviosidade e umidade, tem grande influência na sobrevivência, densidade e distribuição dos anofelinos (Yasuoka & Levins, 2007; Angella et al., 2014).

As localidades rurais do presente estudo, Granada e Remansinho, tem presenciado modificações ambientais similares a estas sendo que em Granada, inaugurada há 33 anos, um maior impacto é observado quando comparado com Remansinho, inaugurado há apenas 8 anos (Moutinho et al., 2011; Barbosa et al., 2014). As diferentes idades desses assentamentos e procedente grau de modificação ambiental podem também explicar as diferenças genéticas, a um nível microgeográfico, aqui encontradas entre os dois grupos de mosquitos *An. darlingi* coletados em Granada e Remansinho.

Contundentemente, ambos os marcadores moleculares do presente estudo, tanto micros-satélites quanto os SNPs, detectaram maior diversidade genética no assentamento rural mais recente, Remansinho, seja pelo maior número de alelos privados (Tabela 4) ou pela maior porcentagem de posições polimórficas (Tabela 7). Esse assentamento possui uma proporção maior de floresta intacta quando comparado com Granada, o assentamento mais antigo (Barbosa et al., 2014), corroborando a hipótese de que desflorestação pode ser associada com perda de biodiversidade (Vieira et al., 2008; Gadner et al., 2009).

A discreta subdivisão encontrada nas amostras de Granada, coletadas no mês de Novembro (Figura 10A), indicou possíveis subpopulações simpátricas de *An. darlingi* que adaptaram-se de forma independente aos mesmos estímulos ambientais locais. Plasticidade no comportamento foi detectada nessa espécie ligadas a preferência para criadouro larval (de Castro et al., 2006; Vittor et al., 2009), a preferência de hospedeiro (Zimmerman et al., 2006), ao horário do repasto sanguíneo (Moreno et al., 2007) e a endofagia ou exofagia (Moreno et al. 2015). No presente estudo, foi investigado se sinais genéticos estão ligados ao comportamento hema-

tofágico, no que diz respeito ao local do repasto sanguíneo, dentro das casas (endofagia) ou no peridomicílio (exofagia); e ao horário, no começo da noite (18-21 horas) e no amanhecer (3-6 horas).

Inicialmente os espécimes de *An. darlingi* endofágicos e exofágicos foram analisados em cada assentamento separadamente. A subdivisão entre eles foi melhor suportada na PCA em Granada (20,3%) do que Remansinho (10,1%) (Figura 14). Posteriormente, numa análise com os quatro grupos, os espécimes de *An. darlingi* endofágicos formaram um único agrupamento, independente da origem do assentamento rural, enquanto que os exofágicos formaram dois grupos avulsos (Figura 14). O mesmo resultado foi obtido na DAPC, onde a maioria dos indivíduos endofágicos enquadraram-se em um único *cluster* inferido (Figura 15). Sendo assim, a pressão seletiva de origem antropogênica para o comportamento endofágico pode ter sido a mesma em dois assentamentos independentes.

Uma outra questão a ser considerada é o local de descanso do mosquito durante o ciclo gonotrófico, ou seja, no intervalo entre os repastos sanguíneos quando ocorre a digestão sanguínea, maturação dos ovos e oviposição. Os mosquitos podem ter comportamento exofágico e exofilico (descansa fora das casas) ou endofágico e endofilico (descansa dentro das casas) (Figura 22; Paaijmans & Thomas, 2011); ambos já observados em *An. darlingi* (Hiwat & Bretas, 2011). Sendo assim, nós sugerimos que os espécimes de *An. darlingi* exofágicos são uma subpopulação “selvagem” que se abriga na floresta e se utiliza de criadouros naturais, enquanto que os endofágicos são “domiciliados”, abrigando-se dentro das casas ou nas suas proximidades e se utilizam de criadouros artificiais próximos das habitações.

A adaptação rápida a novos nichos criados pelo homem foi detectada em linhagens divergentes de *An. gambiae* (Kamdem et al., 2012). Ayala & Coluzzi (2005) mostraram que a estreita relação dos vetores com os seres humanos tem sido reconhecida como uma grande força evolutiva para moldar a população de *An. gambiae*. Recentemente, larvas de *An. gambiae* foram surpreendentemente identificadas em ambientes urbanos (Awolola et al., 2007; Omlin et al., 2007). Em *An. darlingi*, estudos tem mostrado adaptação a área urbana / periurbana, principalmente em associação a poços de piscicultura, na Amazônia peruana (Maheu-Giroux

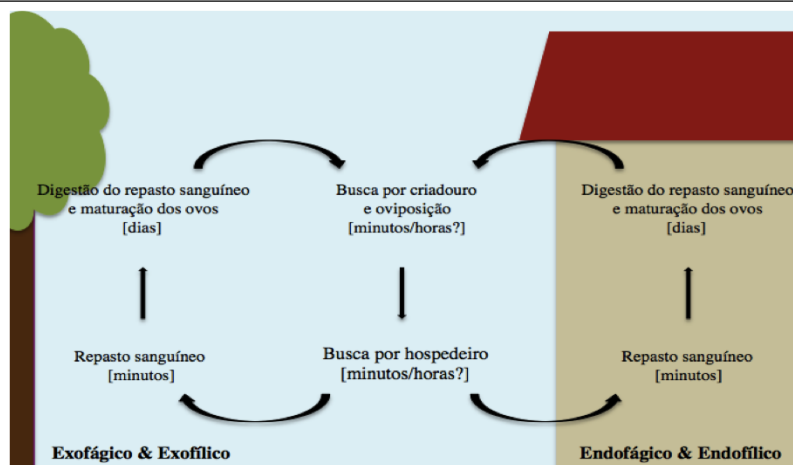


Figura 22: **Esquema geral do ciclo gonotrófico de mosquitos exofágico & exofílico ou endofágico & endofílico.** Os mosquitos se alimentam e descansam fora das casas são considerados exofágicos e exofílicos, respectivamente; e os mosquitos que fazem o repasto sanguíneo e descansam dentro das casas ou proximidades, são respectivamente, endofágicos e endofílicos. Figura baseada na figura 1 de Paaijmans & Thomas, 2011.

et al., 2010) e brasileira (Costa et al., 2010). O trabalho de Costa et al. (2010) foi realizado em um dos locais do presente estudo, Cruzeiro do Sul, onde foi observado um expressivo aumento na transmissão de malária após a escavação de novos poços para criação de peixes na região. Visto isso, as adaptações dos vetores ao ambiente, mediante modificações humanas, podem mudar o paradigma de transmissão e controle da malária: de uma doença que afeta principalmente áreas rurais e pouco povoadas para áreas urbanas e densamente habitadas (Kamdem et al., 2012).

Borrifação de inseticida no interior das casas (IRS) e mosquiteiros impregnados com inseticida (ITNs) reduziram a densidade de anofelinos intra-domiciliares e, causaram variação no horário da alimentação sanguínea por esses vetores (Santos et al., 1999; Pates & Curtis, 2005). Dois estudos reportaram mudanças no comportamento hematofágico em *An. gambiae*, o horário da taxa de picada foi adiantado, comparando vilas tratadas e não-tratadas com mosquiteiros impregnados com inseticida (ITNs) (Njau et al., 1993; MBogo et al., 1996). Em *An. darlingi*, o controle com ITNs foi efetivo apenas em alguns países, ou seja, onde a endofilia e endofagia vetorial eram mais significativas, porém a erradicação foi inviável devido

às populações “selvagens”, exofílica e exofágica (Giglioli, 1956; Rozendaal et al., 1989; Harris et al., 2006; Magris et al., 2007; Hiwat et al., 2010).

Em relação ao horário do repasto sanguíneo, populações de *An. darlingi* tem apresentado diferentes ritmos, uni-bi ou trimodal, no intervalo das 18 horas – 6 horas (Lourenço-de-Oliveira et al., 1989; Hiwat & Bretas, 2011). No presente estudo, a coleta de *An. darlingi* apresentou-se com um ritmo trimodal no assentamento de Granada para mosquitos endofágicos e exofágicos, com o primeiro pico de densidade do vetor às 19 horas, seguido de outros dois às 23 horas e 3 horas da noite (Figura 16). Os SNPs encontrados suportaram diferenças de ~17% entre os espécimes de *An. darlingi*, tanto endofágicos quanto exofágicos, coletados no crepúsculo e no amanhecer (Figura 17, 19). Para os vetores exofágicos, 3 *clusters* genéticos foram inferidos, dois exclusivamente relacionados a cada pico de densidade vetorial e um terceiro compartilhado entre eles (Figura 18B, C). Por outro lado, para os espécimes endofágicos somente 2 *clusters* foram inferidos, os quais compreenderam em sua totalidade um dos dois grupos, ou crepuscular ou matinal (Figura 20B, C). A subdivisão quanto ao horário da alimentação sanguínea foi melhor definida portanto, nos vetores endofágicos. Além disso, na análise em conjunto dos 4 grupos, os espécimes com comportamento endofágico e matinal apresentaram maior homogeneidade entre si (Figura 20A), o que pode refletir uma maior pressão seletiva para este comportamento.

A análise de clusterização hierárquica destes 4 grupos concordou com os dados da análise de componente principal, ou seja, a divisão mais basal foi entre espécimes endofágicos e exofágicos (Figura 21B). Essa primeira distinção pode, de fato, representar subpopulações adaptadas às moradias humanas, consequentemente “domiciliadas”, oriundas das subpopulações “selvagens” que habitam o interior da floresta. Neste caso portanto, as mudanças antropogênicas, através do desenvolvimento de novos nichos ecológicos para o vetor, desempenham uma importante influência no que diz respeito aos aspectos comportamentais de *An. darlingi*.

As derivações posteriores nos dois maiores ramos da clusterização hierárquica, representaram sobretudo a separação entre os horários de hematofagia. Considerando apenas os mosquitos exofágicos, propomos duas hipóteses para explicar a diferença quanto ao horário da

alimentação: a distância do criadouro à fonte de alimentação, no caso a habitação humana, ou seja, os mosquitos da manhã seriam oriundos de criadouros mais longínquos; ou a associação com caráter genético comportamental, independente do criadouro. Para testar as hipóteses, se faz necessária uma investigação genética das larvas provenientes de diversos criadouros, bem como de mosquitos adultos coletados no mesmo espaço temporal.

Alternativamente, a análise de genes fisicamente ligados aos SNPs mais informativos pode indicar grupos gênicos relacionados em uma dada comparação, associando assim, um fator genético ao fenótipo observado. Um importante resultado obtido por essa análise, no presente estudo, foi o enriquecimento de genes para o processo metabólico de substâncias orgânicas na comparações de *An. darlingi* endofágicos vs. exofágicos. Este grupo de genes contem diversos genes associados ao metabolismo de inseticidas e, neste caso, podem estar envolvidos na resistência à inseticida. Mais adiante, o padrão de agrupamento mais terminal pode representar indivíduos muito relacionados como por exemplo, famílias. Hipoteticamente, essas famílias representariam agrupamentos de mosquitos vindos do mesmo criadouro e com hábitos hematofágicos similares. Esse último poderia ser explicado por um grau de treinamento comportamental, a nível de comunidade, existente no mosquito, como encontrado em outros insetos sociais. Portanto, o criadouro de *An. darlingi* parece desempenhar um papel de unidade genética populacional.

6 Conclusão

Em nosso estudo mostramos heterogeneidade genética em *An. darlingi* relacionada desde a aspectos geográficos (escalas macro e microgeográfica), até diferenças entre indivíduos simpátricos ligadas a sazonalidade e ao comportamento. Adicionalmente, acreditamos que o homem desempenha um papel importante ao modificar o ambiente natural, criando novas condições ecológicas para o vetor e influenciando na sua dinâmica populacional. A subdivisão populacional dentro dessa espécie situa a importância do entendimento e vigilância entomológica para epidemiologia da malária. Dado que cada localidade possui características ambientais peculiares e portanto, diferentes composições populacionais dos vetores, intervenções específicas em menor escala passam a ser abordagens interessantes no controle da malária.

7 Referências

- ADAS** M 2006 Construção do Espaço Geográfico Brasileiro. Ed. Moderna, São Paulo. 5ª Edição. Cap10, pag.: 154-67.
- ALEXA** A & RAHNENFUHRER J 2010 topGO: Enrichment analysis for Gene Ontology. R package version 2.22.0.
- ANGÊLLA** AF, SALGUEIRO P, GIL LHS, VICENTE JL, PINTO J, RIBOLLA PEM 2014 Seasonal genetic partitioning in the neotropical malaria vector, *Anopheles darlingi*. Malaria Journal, 13:203.
- AWOLOLA** TS, ODUOLA AO, OBANSA JB, CHUKWURAR NJ, UNYIMADU JP 2007 *Anopheles gambiae* s.s. breeding in polluted water bodies in urban Lagos, southwestern Nigeria. Journal of Vector Borne Diseases 44: 241–244.
- AYALA** FJ & COLUZZI M 2005 Chromosome speciation: humans, Drosophila, and mosquitoes. Proceedings of the National Academy of Sciences U S A 102 Suppl 1: 6535–6542.
- BARBOSA** S, GOZZE AB, LIMA NF, BATISTA CL, BASTOS MDA S, NICOLETE VC, FONTOURA PS, GONÇALVES RM, VIANA SA, MENEZES MJ, SCOPEL KK, CAVASINI CE, MALAFRONTES RS, DA SILVA-NUNES M, VINETZ JM, CASTRO MC, FERREIRA MU 2014 Epidemiology of disappearing *Plasmodium vivax* malaria: a case study in rural Amazonia. PLoS Negl Trop Dis. 8(8):e3109.
- BARROS** FS & HONÓRIO NA 2015 Deforestation and Malaria on the Amazon Frontier: Larval Clustering of *Anopheles darlingi* (Diptera: Culicidae) Determines Focal Distribution of Malaria. Am J Trop Med Hyg. 93(5):939-53.
- BOUSEMA** T, DRAKELEY C, GESASE S, HASHIM R, MAGESA S, MOSHA F, OTIENO S, CARNEIRO I, COX J, MSUYA E, KLEINSCHMIDT I, MAXWELL C, GREENWOOD B, RILEY E, SAUERWEIN R, CHANDRAMOHAN D, GOSLING R 2010 Identification of Hot Spots of Malaria Transmission for Targeted Malaria Control. The Journal of Infectious Diseases; 201(11):1764–1774.
- BRAZ** RM, DUARTE EC, TAUIL PL 2012 Epidemiology of malaria in the municipality of Cruzeiro do Sul, State of Acre, Brazil, in 2010: uses of a control chart at the local level. Revista da Sociedade Brasileira de Medicina Tropical 45(4):526-529.

- BROOKER** S, CLARKE S, NJAGI JK, POLACK S, MUGO B, ESTAMBALE B, MUCHIRI E, MAGNUSSEN P, COX J 2004 Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya. *Trop Med Int Health*, 9:757–766.
- CATCHEN** J, HOHENLOHE PA, BASSHAM S, AMORES A, CRESKO WA 2013 Stacks: an analysis tool set for population genomics. *Molecular Ecology*; 22(11):3124–40.
- CHARLWOOD** JD 1996 Biological variation in *Anopheles darlingi* Root. *Mem Inst Oswaldo Cruz*, 91:391-398.
- CLARK** TD, GREENHOUSE B, NJAMA-MEYA D, NZARUBARA B, MAITEKI-SEBUGUZI C, STAEDKE SG, SETO E, KAMYA MR, ROSENTHAL PJ, DORSEY G 2008 Factors determining the heterogeneity of malaria incidence in children in Kampala, Uganda. *J Infect Dis*, 198(3):393-400.
- CONN** JE, BOLLBACK JP, ONYABE DY, ROBINSON TN, WILKERSON RC, PÓVOA MM 2001 Isolation of polymorphic microsatellite markers from the malaria vector *Anopheles darlingi*. *Molecular Ecology Notes*; 1, 223-225.
- CONN** JE & RIBOLLA PEM 2015 Ecology of *Anopheles darlingi*, the Primary Malaria Vector in the Americas and Current Nongenetic Methods of Vector Control. In: Zach N. Adelman, editors, *Genetic Control of Malaria and Dengue*. Oxford: Academic Press, pp. 81-102.
- CONN** JE, VINEIS JH, BOLLBACK JP, ONYABE DY, WILKERSON C, PÓVOA MM 2006 Populations Structure of the Malaria Vector *Anopheles darlingi* in an Endemic Region of Eastern Amazon Brazil. *Am. J. Trop. Med. Hyg.*, 74(5), pp. 798-806.
- CONSOLI** RAGB & LOURENÇO-DE-OLIVEIRA R 1994 Principais mosquitos de importância sanitária no Brasil. Rio de Janeiro: Fiocruz, p225.
- CORNUET** JM & LUIKART G 1996 Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* 144: 2001–2014.
- COSTA** KM, DE ALMEIDA WA, MAGALHÃES IB, MONTOYA R, MOURA MS, DE LACERDA MV 2010 Malária em Cruzeiro do Sul (Amazônia Ocidental brasileira): análise da série histórica de 1998 a 2008 *Rev Panam Salud Publica*. 28(5):353-60.

- COULIBALY D**, REBAUDET S, TRAVASSOS M, PLOWE CV, DOUMBO OK et al. 2013 Spatio-temporal analysis of malaria within a transmission season in Bandiagara, Mali. *Malaria Journal*, 12:82.
- DA ROCHA JAM**, DE OLIVEIRA SB, PÓVOA MM, MOREIRA LA, KRETTLI AU. 2008 Malaria vectors in areas of *Plasmodium falciparum* epidemic transmission in the Amazon region, Brazil. *American Journal of Tropical Medicine and Hygiene*;78(6) 872-877.
- DA SILVA-NUNES M**, CODECO CT, MALAFRONTES RS, DA SILVA NS, JUNCANSEN C, MUNIZ PT, FERREIRA MU 2008 Malaria on the Amazonian frontier: transmission dynamics, risk factors, spatial distribution, and prospects for control. *Am. J. Trop. Med. Hyg.* 79, 624–635.
- DA SILVA-NUNES M**, MORENO M, CONN JE, GAMBOA D, ABELES S, ET AL. 2012 Amazonian malaria: asymptomatic human reservoirs, diagnostic challenges, environmentally driven changes in mosquito vector populations, and the mandate for sustainable control strategies. *Acta Trop* 121:281–91.
- DA SILVA-VASCONCELOS A**, KATÓ MY, MOURÃO EN, DE SOUZA RT, LACERDA RN, SI-BAJEV A, TSOURIS P, PÓVOA MM, MOMEN H, ROSA-FREITAS MG 2002 Biting indices, host-seeking activity and natural infection rates of anopheline species in Boa Vista, Roraima, Brazil from 1996 to 1998. *Memórias do Instituto Oswaldo Cruz*;97(2) 151-161.
- DAVEY JW**, BLAXTER ML 2010 RADSeq: next-generation population Genetics Briefings in Functional Genomics. vol 9. Nº 5. 416-423.
- DE CASTRO MC**, MONTE-MOR RL, SAWYER DO, SINGER BH 2006 Malaria risk on the Amazon frontier. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2452–2457.
- DE CASTRO MC**, SAWYER DO, SINGER BH 2007 Spatial patterns of malaria in the Amazon: implications for surveillance and targeted interventions. *Health Place* 13, 368–380.
- DEANE LM**, CAUSEY OR, DEANE MP 1948 Notas sobre a distribuição e a biologia dos anofelinos das regiões nordestina e amazônica do Brasil. *Rev Serv Espec Saude Publ*, 1:827-965.
- DERY DB**, BROWN C, ASANTE KP, ADAMS M, DOSOO D, AMENGA-ETEGO S, WILSON M, CHANDRAMOHAN D, GREENWOOD B, OWUSU-AGYEI S 2010 Patterns and seasonality of malaria transmission in the forest-savannah transitional zones. *Malaria Journal*, 9:314.

- DURINCK S, SPELLMAN P, BIRNEY E AND HUBER W** 2009 Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, 4, pp. 1184–1191.
- EARL D & VONHOLDT B** 2012 STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Cons Genet Resour.* 2012; 4(2):359–61.
- ELLIOT R** 1972 The influence of vector behavior on malaria transmission. *Am J Trop Med Hyg*, 21:755–763.
- EMERSON KJ, MERZ CR, CATCHEN JM, et al.** 2010 Resolving postglacial phylogeography using high-throughput sequencing. *Proc Natl Acad Sci USA*;107:16196–200.
- EMERSON KJ, CONN JE, BERGO ES, RANDEL MA, SALLUM MAM** 2015 Brazilian *Anopheles darlingi* Root (Diptera: Culicidae) Clusters by Major Biogeographical Region. *PLoS ONE* 10(7): e0130773.
- EVANNO G, REGNAUT S, GOUDET J** 2005 Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 2005; 14(8):2611–20.
- EXCOFFIER L, LAVAL G, SCHNEIDER S** 2005 Arlequin ver. 3.0: An integrated software Packaged for population genetics data analysis. *Evol. Bioinf. Online*, v.1, p.47–50.
- FELDMAN MW, BERGMAN A, POLLOCK DD & GOLDSTEIN DB** 1997 Microsatellite genetic distances with range constraints: Analytic description and problems of estimation. *Genetics*. 145 (1): 207–16.
- FORATTINI OP** 2002 *Culicidologia Médica* vol. 2: Identificação, Biologia, Epidemiologia. Editora da Universidade de São Paulo, São Paulo, 2002.
- GALARDO AK, ARRUDA M, D’ALMEIDA COUTO AA, WIRTZ R, LOUNIBOS LP, ZIMMERMAN RH.** 2007 Malaria vector incrimination in three rural riverine villages in the Brazilian Amazon. *American Journal of Tropical Medicine and Hygiene*;76(3) 461–469.
- GARDNER TA, BARLOWJ, CHAZDON R, EWERS RM, HARVEY CA, PERES CA, SODHI NS** 2009 Prospects for tropical forest biodiversity in a human-modified world. *Ecology Letter*. 561–582.

- GIGLIOLI** G 1956 Biological variations in *Anopheles darlingi* and *Anopheles gambiae*. Their effect on practical malaria control in the Neotropical Region. Bull. World. Health Org. 15:461–71.
- GIL** LHS, ALVES FP, ZIELER H, SALCEDO JMV, DURLACHER RR, ALMEIDA CUNHA RP, TADA MS, CAMARGO LMA, CAMARGO EP, PEREIRA DA SILVA LH 2003 Seasonal malaria transmission and variation of anopheline density in two distinct endemic areas in Brazilian Amazonia. J Med Entomol, 40:636–641.
- GIL** LHS, TADA MS, KATSURAGAWA TH, RIBOLLA PEM, PEREIRA-DA-SILVA LH 2007 Urban and suburban malaria in Rondônia (Brazilian Western Amazon) II: Perennial transmissions with high anopheline densities are associated with human environmental changes. Mem Inst Oswaldo Cruz, 102:271–276.
- GOLDSTEIN** DB, ROEMER GW, SMITH DA, REICH DE, BERGMAN A & WAYNE RK 1999 The use of microsatellite variation to infer population structure and demographic history in a natural model system. Genetics. 151(2): 797-801.
- GOUDET** J 1995 FSTAT (Version 1.2): a computer program to calculate F-Statistics. The Journal of Heredity 86 (6), pp.485-86.
- GUO**, SW & THOMPSON, EA 1992 Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics, v.48, n.2, p.361-372.
- GUTIÉRREZ** LA, GÓMEZ GF, GONZÁLEZ JJ, CASTRO MI, LUCKHART S, CONN JE, CORREA MM 2010 Microgeographic Genetic Variation of the Malaria Vector *Anopheles darlingi* Root (Diptera: Culicidae) from Córdoba and Antioquia, Colombia. Am. J. Trop. Med. Hyg., 83(1), pp. 38-47.
- HARBACH** RE 2004 The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. Bulletin of Entomological Research, 94, 537-553.
- HARRIS** AF, MATIAS-ARNEZ A, HILL N 2006 Biting time of *Anopheles darlingi* in the Bolivian Amazon and implications for control of malaria. Trans R Soc Trop Med Hyg, 100:45-47.
- HIWAT** H, ISSALY J, GABORIT P, SOMAI A, SAMJHAWAN A, SARDJOE P, SOEKHOE T, GIROD R 2010 Behavioral heterogeneity of *Anopheles darlingi* (Diptera: Culicidae) and malaria transmission dynamics along the Maroni River, Suriname, French Guiana. Trans R Soc Trop Med Hyg, 104(3):207-13.

- HIWAT H & BRETAS G** 2011 Ecology of *Anopheles darling* Root with respect to vector importance: a review. *Parasit. Vectors*, 4, 177.
- HOLM S** 1979 A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6 (2), pp.65-70.
- HOHENLOHE PA, BASSHAM S, ETTER PD, et al.** 2010 Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *PLoS Genet*;6:e1000862.
- JAARI S, MENG-HUA LI, MERILÄ J** 2009 A first-generation microsatellite-based genetic linkage map of the Siberian jay (*Perisoreus infaustus*): insights into avian genome evolution. *BMC Genomics*, 10:1.
- JIMÉNEZ IP, CONN JE, WIRTZ R, BROCHERO H** 2012 *Anopheles* (Diptera: Culicidae) vectors of malaria in Puerto Carreno municipality, Vichada, Colombia. *Biomedica*;32 (Suppl. 1):13 21.
- JIMÉNEZ IP, CONN JE, BROCHERO H** 2014 Malaria vectors in San José del Guaviare, Orinoquia, Colombia. *J Am Mosq Control Assoc*;30(2):918.
- JOMBART T & AHMED I** 2011 adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* 27, 3070-3071.
- KAMDEM C, TENE FOSSOG B, SIMARD F, ETOUNA J, NDO C, KENGNE P, BOUSSÈS P, ETOA FX, AWONO-AMBENE P, FONTENILLE D, ANTONIO-NKONDJIO C, BESANSKY NJ, COSTANTINI C** 2012 Anthropogenic habitat disturbance and ecological divergence between incipient species of the malaria mosquito *Anopheles gambiae*. *PLoS One*. 7(6):e39453.
- KIMBERLY JG, ROSE LA, DAN GB, MICHELLE TF, NOLAN CK, JEAN-SEBASTIEN M, BROOK TM, SEBASTIEN R, DIANA JR, THOR V, TIMOTHY HV** 2012 Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program STRUCTURE *Molecular Ecology* 21, 4925–4930.
- KREUTZER RD, KITZMILLER JB, FERREIRA E** 1972 Inversion polymorphism in the salivary gland chromosomes of *Anopheles darlingi* Root. *Mosq News*, 32: 555- 565.
- LANGMEAD B & SALZBERG SL** 2012 Fast gapped-read alignment with Bowtie 2. *Nat Methods*, 9:357-359.

- LEHMANN** T, BESANKY NJ, HAWLEY WA, FAHEY TG, KAMAU L, COLLINS FH 1997 Micro-geographic structure of *Anopheles gambiae* in western Kenya based on mtDNA and microsatellite loci. Mol Ecol. 6:243–253.
- LOAIZA** JR, BERMINGHAM E, SCOTT ME, ROVIRA JR, CONN JE 2008 Species composition and distribution of adult *Anopheles* (Diptera: Culicidae) in Panama. J Med Entomol;45 (5):841–51.
- LOIOLA** CC, DA SILVA CJ, TAUIL PL 2002 Malaria control in Brazil: 1965 to 2001. Rev Panam Salud Publica, 11(4):235-244.
- LOUNIBOS** LP & CONN JE 2000 Malaria vector heterogeneity in South America. Am Entomol, 46:238-49.
- LOURENÇO-DE-OLIVEIRA** R 1989 Hemoparasitos encontrados em alguns mamíferos de Balbina, estado do Amazonas. Mem. Inst Oswaldo Cruz 83:233-44.
- MA** Y, YANG M, FAN Y, WU J, MA Y, et al. 2011 Population Structure of the Malaria Vector *Anopheles sinensis* (Diptera: Culicidae) in China: Two Gene Pools Inferred by Microsatellites. PLoS ONE 6(7): e22219.
- MAGRIS** M, RUBIO-PALIS Y, ALEXANDER N, RUIZ B, GALVÁN N, FRIAS D, BLANCO M, LINES J 2007 Community-randomized trial of lambdacyhalothrin-treated hammock nets for malaria control in Yanomami communities in the Amazon region of Venezuela. Trop Med Int Health, 12:392-403.
- MAHEU-GIROUX** M, CASAPIA M, SOTO-CALLE VE, FORD LB, BUCKERIDGE DL, COOMES OT, et al. 2010 Risk of malaria transmission from fish ponds in the Peruvian Amazon. Acta Trop. 115: 112–118.
- MALAFRONTÉ** RS, MARRELLI MT, MARINOTTI O 1999 Analysis of ITS2 DNA sequences from Brazilian *Anopheles darlingi* (Diptera: Culicidae). J Med Entomol 36: 631–634.
- MANGUIN** S, CARNIVALE P, MOUCHET J 2008 Biodiversity of Malaria in the World. John Libbey Eurotext, Paris.
- MARINOTTI** O, ET. AL 2013 The Genome of *Anopheles darlingi*, the main neotropical malaria vector. Nucleic Acids Research, 1 – 14.

- MARQUES AC** 1987 Human migration and the spread of malaria in Brazil. *Parasitol Today*, 3:166-170.
- MBOGO CNM, BAYA NM, OFULLA AVO, GITHURE JI, SNOW RW** 1996 The impact of permethrin-impregnated bednets on malaria vectors of the Kenyan coast. *Med. Vet. Entomol.* 10:251-59.
- MIRABELLO L & CONN JE** 2006 Molecular population genetics of the malaria vector *Anopheles darlingi* in Central and South America. *Heredity*, 96:311-321.
- MIRABELLO L, VINEIS JH, YANOVIK SP, SCARPASSA VM, PÓVOA MM, PADILLA N, ACHEE NL, CONN JE** 2008 Microsatellite data suggest significant population structure and differentiation within the malaria vector *Anopheles darlingi* in Central and South America. *BMC Ecology*, 8:3.
- MOLINA-CRUZ A, DE MERIDA AM, MILLS K, et al.** 2004 Gene flow among *Anopheles albimanus* populations in Central America, South America, and the Caribbean assessed by microsatellites and mitochondrial DNA. *Am J Trop Med Hyg*;71(3):350-9.
- MORENO JE, RUBIO-PALIS Y, PAEZ E, PEREZ E, SANCHEZ V** 2007 Abundance, biting behaviour and parous rate of anopheline mosquito species in relation to malaria incidence in gold-mining areas of southern Venezuela. *Med Vet Entomol*, 21:339-349.
- MORENO M, SAAVEDRA MP, BICKERSMITH SA, LAINHART W, TONG C, ALAVA F, VINETZ JM, CONN JE** 2015 Implications for changes in *Anopheles darlingi* biting behaviour in three communities in the peri-Iquitos region of Amazonian Peru. *Malaria Journal*, 14:290.
- MOUTINHO PR, GIL LHS, CRUZ RB, RIBOLLA PEM** 2011 Population dynamics, structure and behavior of *Anopheles darlingi* in a rural settlement in the Amazon rainforest of Acre, Brazil. *Malaria Journal*, 10:174.
- MURTAGH F & LEGENDRE P** 2014 Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, Volume 31, Issue 3, pp 274-295.
- NDO C, ANTONIO-NKONDJIO C, COHUET A, AYALA D, KENGNE P, MORLAIS I, AWONO-AMBENE PH, COURET D, NGASSAM P, FONTENILLE D, SIMARD F** 2010 Population

- genetic structure of the malaria vector *Anopheles nili* in sub-Saharan Africa. *Malaria Journal* , 9:161.
- NEI**, M 1978 Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, v.89, n.3, p.583-590.
- NELSON** DL & **WARREN** ST 1993 Trinucleotide repeat instability: when and where? *Nat Genet.*4 (2): 107-8.
- NJAU** RJA, **MOSHA** FW, **NGUMA** JFM 1993 Field trials of pyrethroid impregnated bed- nets in northern Tanzania. 1. Effects on malaria transmission *Insect Sci. Appl.* 5: 575–84.
- OLIVEIRA-FERREIRA** J, **LACERDA** MV, **BRASIL** P, **LADISLAU** JL, **TAUIL** PL, **DANIELRI-BEIRO** CT 2010 Malaria in Brazil: an overview. *Malaria J*, 9:11.
- OMLIN** FX, **CARLSON** JC, **OGBUNUGAFOR** CB, **HASSANALI** A 2007 *Anopheles gambiae* exploits the treehole ecosystem in western Kenya: a new urban malaria risk? *American Journal of Tropical Medicine and Hygiene* 77: 264–269.
- ORJUELA** LI, **AHUMADA** ML, **AVILA** I, **HERRERA** S, **BEIER** JC, **QUINONES** ML 2015 Human biting activity, spatial-temporal distribution and malaria vector role of *Anopheles calderoni* in the southwest of Colombia. *Malaria Journal*. 14: 256.
- PAAIJMANS** KP & **THOMAS** MB 2011 The influence of mosquito resting behaviour and associated microclimate for malaria risk. *Malaria Journal*, 10:183.
- PARADIS** E, **CLAUDE** J, **STRIMMER** K 2004 APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289-290.
- PATES** H & **CURTIS** C 2005 Mosquito behavior and vector control. *Annual Review of Entomology*, 50:53-70.
- PEDRO** PM & **SALLUM** MAM 2009 Spatial expansion and population structure of the neotropical malaria vector, *Anopheles darlingi* (Diptera: Culicidae). *Biol J Linn Soc*; 97(4):854 66.
- PETERSON** BK, **WEBER** JN, **KAY** EH, **FISHER** HS, **HOEKSTRA** HE 2012 Double Digest RAD-seq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE* 7(5): e37135. doi:10.1371/journal.pone.0037135.

- PRITCHARD** JK, STEPHENS M, DONNELLY P 2000 Inference of population structure using multilocus genotype data. *Genetics*, v.155, n.2, p.945-959.
- QUIÑONES** ML & SUAREZ MF 1990 Indoor resting heights of some anophelines in Colombia. *J Am Mosq Control Assoc*;6(4):602-4.
- R** Core Team 2016 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RAYMOND** M & ROUSSET F 1995 GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. *Journal of Heredity* 86, pp.248-49.
- REIS** IC, HONÓRIO NA, BARROS FSM, BARCELLOS C, KITRON U, CAMARA DCP, et al. 2015 Epidemic and Endemic Malaria Transmission Related to Fish Farming Ponds in the Amazon Frontier. *PLoS ONE* 10(9): e0137521.
- ROBERTS** DR, ALECRIM WD, TAVARES AM, RADKE MG 1987 The house-frequenting, host-seeking and resting behavior of *Anopheles darlingi* in southeastern Amazonas, Brazil. *J Am Mosq Control Assoc*;3(3):433-41.
- ROUSSET** F 2008 Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux.
- ROZENDAAL** JA, VAN HOOFF JPM, VOORHAM J, OOSTBURG BFJ 1989 Behavioral responses of *Anopheles darlingi* in Suriname to DDT residues on house walls. *J. Am. Mosq. Control Assoc.* 5:339-50.
- ROZENDAAL** JA 1990 Observations on the distribution of anophelines in Suriname with particular reference to the malaria vector *Anopheles darlingi*. *Mem. Inst. Oswaldo Cruz*, 85: 221-34.
- SANTOS** JB, DOS SANTOS F, MACÊDO V 1999 Variation of anopheles density with deltamethrin-impregnated mosquito nets in an endemic malaria area of the Brazilian Amazon. *Cad Saúde Pública*, 15:281-292.
- SCARPASSA** VM & CONN JE 2007 Population genetic structure of the major malaria vector *Anopheles darlingi* (Diptera: Culicidae) from the Brazilian Amazon, using microsatellite markers. *Mem. Inst. Oswaldo Cruz*, 102(3): 319-327.

- SINKA** ME, RUBIO-PALIS Y, MANGUIN S, PATIL AP, TEMPERLEY WH, GETTING PW, VAN BOECKEL T, KABARIA CW, HARBACH RE, HAY SI 2010 The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic Precis. *Parasites and Vectors*, 3:72.
- SINKA** ME, BANGS MJ, MANGUIN S, RUBIO-PALIS Y, CHAREONVIRIYAPHAP T, COETZEE M, MBOGO CM, HEMINGWAY J, PATIL AP, TEMPERLEY WH et al. 2012 A global map of dominant malaria vectors. *Parasit. Vectors*, 5, 69.
- SOLIGNAC** M, MOUGEL F, VAUTRIN D, MONNEROT M, CORNUET JM 2007 A third-generation microsatellite-based linkage map of the honey bee, *Apis mellifera*, and its comparison with the sequence-based physical map. *Genome Biology* May 8:R66.
- SOUZA** EF 2002 Análises Comparativas dos Custos Totais de Produção dos PAD's Boa Esperança e Humaita. Monografia apresentada ao curso de Graduação em Economia da Universidade Federal do Acre como requisito parcial para obtenção do Grau em Bacharel em Ciências Econômicas. Acre. Pag: 11-15.
- SOUZA-SANTOS** R 2002 Seasonal distribution of malaria vectors in Machadinho d'Oeste, Rondônia State, Amazon Region, Brazil. *Cad Saúde Pública*, 18:1813-1818.
- STAEDKE** SG, NOTTINGHAM EW, COX J, KAMYA MR, ROSENTHAL PJ, DORSEY G 2003 Short report: proximity to mosquito breeding sites as a risk factor for clinical malaria episodes in an urban cohort of Ugandan children. *American Journal Tropical Medicine Hygiene*;69:244-6.
- SUN** JX, MULLIKIN JC, PATTERSON N, REICH DE 2009 Microsatellites Are Molecular Clocks That Support Accurate Inferences about History. *Molecular Biology and Evolution* 26(5):1017-1027.
- SUTHERLAND** CJ, TANOMSING N, NOLDER D, et al. 2010 Two nonrecombining sympatric forms of the human malaria parasite *Plasmodium ovale* occur globally. *J. Infect Dis*: 201(10):1544-1550.
- TADEI** WP, THATCHER BD, SANTOS JMM, SCARPASSA VM, BRANDÃO RODRIGUES I, RAFAEL MS 1998 Ecological observations on Anopheline vectors of malaria in the Brazilian Amazon. *Am J Trop Med Hyg*, 59:325-345.
- TAUIL** P & DANIEL-RIBEIRO C 1998 Some aspects of epidemiology and control of malaria in Brazil. *Res Rev Parasitol*, 58:163-167.

- TAUTZ** D 1989 Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl Acids Res. 17: 6463-71.
- VALLE** D & LIMA JMT 2014 Large-scale drivers of malaria and priority areas for prevention and control in the Brazilian Amazon region using a novel multi-pathogen geospatial model. Malaria Journal, 13:443.
- VAN** OOSTERHOUT C, HUTCHINSON WF, WILLS DPM, SHIPLEY PF 2004 Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Mol Ecol Notes, 4: 535-538.
- VIEIRA** ICG, TOLEDO PM, SILVA JMC, HIGUCHI H 2008 Deforestation and threats to the biodiversity of Amazonia Braz. J. Biol., 68(4, Suppl.): 949-956.
- VITRAL** CL, SILVA-NUNES M, PINTO MA, OLIVEIRA JM, GASPAR AMC, PEREIRA RCC, FERREIRA MU 2014 Hepatitis A and E seroprevalence and associated risk factors: a community-based cross-sectional survey in rural Amazonia. BMC Infectious Diseases, 14:458.
- VITTOR** AY, GILMAN RH, TIELSCH J, GLASS G, SHIELDS T, LOZANO WS, PINEDO-CANCINO V, PATZ JA 2006 The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian Amazon. Am. J. Trop. Med. Hyg. 74, 3–11.
- VITTOR** AY, PAN W, GILMAN RH, TIELSCH J, GLASS G, SHIELDS T, SANCHEZ-LOZANO W, PINEDO VV, SALAS-COBOS E, FLORES S, PATZ JA 2009 Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, *Anopheles darlingi*. Am. J. Trop. Med. Hyg. 81, 5–12.
- VOORHAM** J 2002 Intra-population plasticity of *Anopheles darlingi*'s (Diptera, Culicidae) biting activity patterns in the state of Amapá, Brazil. Rev Saúde Pública;36(1):75-80.
- WARNES** GR 2011 Gplots: Various R Programming Tools for Plotting Data. <http://cran.r-project.org/web/packages/gplots/index.html>.
- WEISSENBAACH** J 1992 A second-generation linkage map of the human genome. Nature. 359: 749-801.

WORLD MALARIA ORGANIZATION 2014 World Malaria Report 2015 [http://www.who.int/malaria/publications/world_malaria_report_2014/en/].

WORLD MALARIA ORGANIZATION 2015 World Malaria Report 2015 [<http://www.who.int/malaria/publications/world-malaria-report-2015/en/>].

YASUOKA J & LEVINS R 2007 Impact Of Deforestation And Agricultural Development On Anopheline Ecology And Malaria Epidemiology. *Am. J. Trop. Med. Hyg.*, 76(3), pp. 450-460.

ZIMMERMAN RH, RIBEIRO CALARDO AK, LOUNIBOS LP, ARRUDA M, WIRTZ R 2006 Bloodmeal Hosts of *Anopheles* species (Diptera: Culicidae) in a Malaria- endemic area of the Brazilian Amazon. *J Med Entomol*, 43:947-956.

ZIMMERMAN RH, LOUNIBOS LP, NISHIMURA N, GALARDO AKR, GALARDO CD, ARRUDA ME 2013 Nightly biting cycles of malaria vectors in a heterogeneous transmission area of eastern Amazonian Brazil. *Malaria Journal*, 12:262.

8 Anexos

1. Anexo- Tabela 1. Indexação das amostras de *An. darlingi* para a biblioteca 1 sequenciada no Miseq Desktop Sequencer.
2. Anexo- Tabela 2. Indexação das amostras de *An. darlingi* para a biblioteca 2 sequenciada no Miseq Desktop Sequencer.
3. Anexo- Tabela 3. Indexação das amostras de *An. darlingi* para a biblioteca 3 sequenciada no Miseq Desktop Sequencer.
4. Anexo- Tabela 4. Indexação das amostras de *An. darlingi* para a biblioteca 4 sequenciada no Miseq Desktop Sequencer.
5. Anexo- Tabela 5. Indexação das amostras de *An. darlingi* para a biblioteca 1 sequenciada no Hiseq 2500.
6. Anexo- Tabela 6. Indexação das amostras de *An. darlingi* para a biblioteca 2 sequenciada no Hiseq 2500.
7. Anexo- Tabela 7. Indexação das amostras de *An. darlingi* para a biblioteca 2 sequenciada no Hiseq 2500.
8. Anexo- Tabela 8. Resultado do sequenciamento ddRADseq para *An. darlingi* das três localidades. Detalhes por indivíduo do número de sequências genotipadas.
9. Anexo- Tabela 9. Resultado do sequenciamento ddRADseq para *An. darlingi* de Granada e Remansinho em Março de 2012. Detalhes por indivíduo do número de sequências genotipadas.
10. Anexo- Tabela 10. Resultado do sequenciamento ddRADseq para *An. darlingi* de Granada e Remansinho em Maio de 2012. Detalhes por indivíduo do número de sequências genotipadas.

11. Anexo- Tabela 11. Resultado do sequenciamento ddRADseq para *An. darlingi* endofágicos e exofágicos em Remansinho. Detalhes por indivíduo do número de sequências genotipadas.
12. Anexo- Tabela 12. Resultado do sequenciamento ddRADseq para *An. darlingi* exofágicos em Granada. Detalhes por indivíduo do número de sequências genotipadas.
13. Anexo- Tabela 13. Resultado do sequenciamento ddRADseq para *An. darlingi* endofágicos em Granada. Detalhes por indivíduo do número de sequências genotipadas.
14. Anexo- Texto 1. *Script* utilizado para o processamento dos dados com Stacks.
15. Anexo- Texto 2. *Script* utilizado para a conversão de uma lista de genes para uma lista de GO.

Anexo- Tabela 1. Indexação das amostras de *An. darlingi* para a biblioteca 1 sequenciada no Miseq Desktop Sequencer.

Biblioteca 1 Miseq		
Amostra	Índice Nextera I7	Índice Nextera I5
R1	N701	N501
R2	N701	N502
R3	N702	N501
R4	N702	N502
R5	N703	N501
R6	N703	N502
R7	N704	N501
R8	N704	N502
R9	N705	N501
R10	N705	N502
R11	N706	N501
R12	N706	N502
G1	N707	N501
G2	N707	N502
G3	N708	N501
G4	N708	N502
G5	N709	N501
G6	N709	N502
G7	N710	N503
G8	N710	N504
G9	N711	N503
G10	N711	N504
G11	N712	N503
G12	N712	N504

Anexo- Tabela 2. Indexação das amostras de *An. darlingi* para a biblioteca 2 sequenciada no Miseq Desktop Sequencer.

Biblioteca 2 Miseq		
Amostra	Índice Nextera I7	Índice Nextera I5
G13	N701	N501
G14	N701	N502
G15	N701	N503
G16	N701	N504
G17	N702	N501
G18	N702	N502
G19	N702	N503

G20	N702	N504
G21	N703	N501
G22	N703	N502
G23	N703	N503
G24	N703	N504
G25	N704	N501
G26	N704	N502
G27	N704	N503
G28	N704	N504
G29	N705	N501
G30	N705	N502
CS1	N705	N503
CS2	N705	N504
CS3	N706	N501
CS4	N706	N502
CS5	N706	N503
CS6	N706	N504

Anexo- Tabela 3. Indexação das amostras de *An. darlingi* para a biblioteca 3 sequenciada no Miseq Desktop Sequencer.

Biblioteca 3 Miseq		
Amostra	Índice Nextera I7	Índice Nextera I5
R13	N701	N501
R14	N702	N501
R15	N703	N501
R16	N704	N501
R17	N705	N501
R18	N706	N501
R19	N707	N501
R20	N708	N501
R21	N709	N501
R22	N710	N501
R23	N711	N501
R24	N712	N501
R25	N701	N502
R26	N702	N502
R27	N703	N502
CS7	N704	N502
CS8	N705	N502
CS9	N706	N502

CS10	N707	N502
CS11	N708	N502
CS12	N709	N502
CS13	N710	N502
CS14	N711	N502
CS15	N712	N502

Anexo- Tabela 4. Indexação das amostras de *An. darlingi* para a biblioteca 4 sequenciada no Miseq Desktop Sequencer.

Biblioteca 4 Miseq		
Amostra	Índice Nextera I7	Índice Nextera I5
CS16	704	502
CS17	705	502
CS18	706	502
CS19	707	502
CS20	708	502
CS21	709	502
CS22	710	502
CS23	711	502
CS24	712	502
CS25	701	501
CS26	702	501
CS27	703	501
CS28	704	501
CS29	705	501
CS30	706	501

Anexo- Tabela 5. Indexação das amostras de *An. darlingi* para a biblioteca 1 sequenciada no Hiseq 2500.

Biblioteca 1 Hiseq		
Amostra	Índice Nextera I7	Índice Nextera I5
A11R001	N701	N501
A11R002	N702	N501
A11R003	N703	N501
A11R004	N704	N501
A11R005	N705	N501
A11R006	N706	N501
A11R007	N707	N501
A11R008	N708	N501

A11R009	N709	N501
A11R010	N710	N501
A11R011	N711	N501
A11R012	N712	N501
A11R013	N701	N502
A11R014	N702	N502
A11R015	N703	N502
A11R016	N704	N502
A11R137	N705	N502
C11G008	N706	N502
C11G009	N707	N502
C11G010	N708	N502
C11G011	N709	N502
C11G012	N710	N502
C11G013	N711	N502
C11G014	N712	N502
A11R145	N701	N503
C11G015	N702	N503
C11G016	N703	N503
C11G017	N704	N503
C11G018	N705	N503
A11R150	N706	N503
A11R151	N707	N503
A11R152	N708	N503
A11R153	N709	N503
A11R154	N710	N503
A11R155	N711	N503
A11R156	N712	N503
A11R157	N701	N504
A11R158	N702	N504
A11R159	N703	N504
A11R160	N704	N504
A11R161	N705	N504
A11R162	N706	N504
A11R163	N707	N504
A11R164	N708	N504
A11R165	N709	N504
A11R166	N710	N504
A11R167	N711	N504
A11R168	N712	N504
A11R169	N701	N505

A11R170	N702	N505
A11R171	N703	N505
A11R172	N704	N505
A11R173	N705	N505
A11R174	N706	N505
A11R175	N707	N505
A11R176	N708	N505
C11G001	N709	N505
C11G002	N710	N505
C11G003	N711	N505
C11G004	N712	N505
A11G293	N701	N506
A11G216	N702	N506
A11G294	N703	N506
A11G218	N704	N506
A11G295	N705	N506
A11G296	N706	N506
A11G297	N707	N506
A11G298	N708	N506
A11G223	N709	N506
A11G299	N710	N506
A11G225	N711	N506
A11G226	N712	N506
A11G300	N701	N507
A11G228	N702	N507
A11G229	N703	N507
A11G230	N704	N507
A11G231	N705	N507
A11G301	N706	N507
A11G302	N707	N507
A11G303	N708	N507
A11G304	N709	N507
A11G236	N710	N507
A11G305	N711	N507
A11G306	N712	N507
A11G307	N701	N508
A11G308	N702	N508
A11G309	N703	N508
C11G005	N704	N508
C11G006	N705	N508
C11G007	N706	N508

A11G287	N707	N508
A11G288	N708	N508
A11G289	N709	N508
A11G291	N711	N508
A11G290	N710	N508
A11G292	N712	N508

Anexo- Tabela 6. Indexação das amostras de *An. darlingi* para a biblioteca 2 sequenciada no Hiseq 2500.

Biblioteca 2 Hiseq		
Amostra	Índice Nextera I7	Índice Nextera I5
C11R016	N701	N501
C11R017	N702	N501
C11R018	N703	N501
C11R019	N704	N501
C11R020	N705	N501
A12R011	N706	N501
A12R012	N707	N501
A12R013	N708	N501
A12R014	N709	N501
A12R015	N710	N501
A12R016	N711	N501
A12R017	N712	N501
A12R018	N701	N502
A12R019	N702	N502
A12R020	N703	N502
A12R021	N704	N502
C11R001	N705	N502
A12R022	N706	N502
A12R023	N707	N502
A12R024	N708	N502
A12R025	N709	N502
A12R026	N710	N502
A12R027	N711	N502
A12R028	N712	N502
C11R002	N701	N503
A12R029	N702	N503
A12R030	N703	N503
A12R031	N704	N503
A12R032	N705	N503

C11R003	N706	N503
A12R033	N707	N503
A12R034	N708	N503
A12R035	N709	N503
A12R036	N710	N503
A12R037	N711	N503
A12R038	N712	N503
A12R039	N701	N504
A12R040	N702	N504
A12G002	N703	N504
A12G003	N704	N504
A12G008	N705	N504
A12G009	N706	N504
A12G010	N707	N504
A12G011	N708	N504
A12G012	N709	N504
A12G013	N710	N504
A12G014	N711	N504
A12G015	N712	N504
A12G016	N701	N505
A12G017	N702	N505
A12G018	N703	N505
A12G019	N704	N505
A12G020	N705	N505
A12G021	N706	N505
A12G022	N707	N505
A12G023	N708	N505
A12G024	N709	N505
B12G001	N710	N505
B12G003	N711	N505
B12G004	N712	N505
B12G005	N701	N506
C11R004	N702	N506
B12G006	N703	N506
C11R005	N704	N506
B12G007	N705	N506
B12G008	N706	N506
B12G009	N707	N506
B12G010	N708	N506
C11R006	N709	N506
B12G011	N710	N506

C11R007	N711	N506
C11R008	N712	N506
B12G028	N701	N507
C11R009	N702	N507
C11R010	N703	N507
C11R011	N704	N507
C11R012	N705	N507
B12G029	N706	N507
B12G030	N707	N507
B12G031	N708	N507
B12G032	N709	N507
C11R013	N710	N507
B12G033	N711	N507
B12G034	N712	N507
B12G035	N701	N508
B12G036	N702	N508
B12G038	N703	N508
B12G039	N704	N508
B12G040	N705	N508
B12G041	N706	N508
B12G042	N707	N508
B12G043	N708	N508
B12R019	N709	N508
C11R014	N710	N508
B12R020	N711	N508
C11R015	N712	N508

Anexo- Tabela 7. Indexação das amostras de *An. darlingi* para a biblioteca 2 sequenciada no Hiseq 2500.

Biblioteca 3 Hiseq		
Amostra	Índice Nextera I7	Índice Nextera I5
B12R021	N701	N501
B12R022	N702	N501
B12R023	N703	N501
B12R024	N704	N501
B12R025	N705	N501
B12R026	N706	N501
B12R027	N707	N501
B12R029	N708	N501
B12R030	N709	N501

B12R031	N710	N501
B12R032	N711	N501
B12R034	N712	N501
B12R035	N701	N502
B12R036	N702	N502
A11G170	N703	N502
A11G171	N704	N502
A11G172	N705	N502
A11G173	N706	N502
A11G174	N707	N502
A11G175	N708	N502
A11G176	N709	N502
A11G177	N710	N502
A11G178	N711	N502
A11G179	N712	N502
A11G180	N701	N503
A11G181	N702	N503
A11G182	N703	N503
A11G066	N704	N503
A11G067	N705	N503
A11G068	N706	N503
A11G183	N707	N503
A11G184	N708	N503
A11G185	N709	N503
A11G186	N710	N503
A11G187	N711	N503
A11G188	N712	N503
A11G069	N701	N504
A11G190	N702	N504
A11G191	N703	N504
A11G070	N704	N504
A11G193	N705	N504
A11G194	N706	N504
A11G071	N707	N504
A11G072	N708	N504
A11G197	N709	N504
A11G198	N710	N504
A11G073	N711	N504
A11G200	N712	N504
A11G201	N701	N505
A11G074	N702	N505

A11G075	N703	N505
A11G204	N704	N505
A11G205	N705	N505
A11G206	N706	N505
A11G076	N707	N505
A11G077	N708	N505
A11G078	N709	N505
A11G079	N710	N505
A11G080	N711	N505
A11G101	N712	N505
A11G102	N701	N506
A11G216	N702	N506
A11G217	N703	N506
A11G218	N704	N506
A11G103	N705	N506
A11G104	N706	N506
A11G221	N707	N506
A11G222	N708	N506
A11G223	N709	N506
A11G105	N710	N506
A11G225	N711	N506
A11G226	N712	N506
A11G227	N701	N507
A11G228	N702	N507
A11G229	N703	N507
A11G230	N704	N507
A11G231	N705	N507
A11G232	N706	N507
A11G233	N707	N507
A11G234	N708	N507
A11G235	N709	N507
A11G236	N710	N507
A11G237	N711	N507
A11G238	N712	N507
A11G106	N701	N508
A11G107	N702	N508
A11G108	N703	N508
A11G109	N704	N508
A11G110	N705	N508
A11G111	N706	N508

Anexo- Tabela 8. Resultado do sequenciamento ddRADseq para *An. darlingi* das três localidades. Detalhes por indivíduo do número de sequências genotipadas.

Amostra	Total de Sequências	Sequências Ambíguas/Baixa Qualidade	Sequências Retidas	% Sequências Alinhadas	Sequências Únicas	Loci Polimórfico	% Loci Polimórfico	SNPs Encontrados
G1	739640	28337	711303	32,17	13087	3122	23,86	5175
G2	1042838	371585	671253	30,64	8725	1945	22,29	3200
G3	860774	81858	778916	28,87	7419	1564	21,08	2509
G4	954722	296864	657858	32,08	12114	2889	23,85	4702
G5	1215308	63222	1152086	31,06	14474	3362	23,23	5523
G6	1554466	504712	1049754	28,35	9517	2128	22,36	3516
G7	865802	98008	767794	27,64	7961	1744	21,91	2846
G8	1244804	84773	1160031	30,09	11936	2717	22,76	4460
G9	821506	32422	789084	31,89	14898	3609	24,22	5943
G10	941918	90739	851179	28,47	8946	1883	21,05	3126
G11	677158	48355	628803	29,77	8891	2133	23,99	3439
G12	749370	43637	705733	31,55	14363	4478	31,18	7668
R1	805138	83061	722077	31,77	14228	3535	24,85	5845
R5	904304	44682	859622	32,96	14866	3822	25,71	6295
R8	931754	233033	698721	33,57	13070	3280	25,10	5422
R13	1633808	637729	996079	34,04	17168	4228	24,63	7045
R14	1573448	727205	846243	33,18	16749	4209	25,13	7051
R15	1723638	971992	751646	29,04	14631	3601	24,61	5796
R16	1125340	503984	621356	33,06	14508	3491	24,06	5713
R17	1809918	778097	1031821	30,69	17125	5985	34,95	10183
R18	1178518	569470	609048	27,09	11618	2724	23,45	4457
R20	1387330	679004	708326	33,00	13458	3254	24,18	5287
R21	1306100	630948	675152	33,12	14112	2963	21,00	4847
R23	1221992	548214	673778	33,23	13452	3175	23,60	5261
R24	1132644	495291	637353	32,87	13884	3328	23,97	5489
R27	983976	270336	713640	32,33	14880	3539	23,78	5751
CS16	1588022	53443	1534579	32,92	24366	5669	23,27	9512
CS17	1630668	43683	1586985	33,12	27046	6224	23,01	10357
CS18	2841410	89047	2752364	33,78	29527	6796	23,02	11425
CS20	1123202	44471	1078731	33,42	23683	5362	22,64	8855
CS21	1467178	47555	1419623	36,77	33428	7205	21,55	11810
CS22	1404994	35273	1369721	35,98	28991	6877	23,72	11266
CS23	1332746	36131	1296615	35,23	28269	6420	22,71	10649
CS24	2069126	73434	1995692	33,94	30371	6638	21,86	10993
CS25	1285728	205913	1079815	34,99	20468	4516	22,06	7414
CS26	1629530	306425	1323105	33,48	25786	5323	20,64	8638
CS27	1758560	281691	1476869	30,97	22777	4638	20,36	7514

CS28	979740	151960	827780	34,53	19155	4053	21,16	6733
CS29	1305716	275191	1030525	32,26	15439	3267	21,16	5452
CS30	1382422	262851	1119569	32,92	25253	5360	21,23	8853
Média	1279631,40	270615,65	1009015,72	32,17	17265,97	4026,4	23,47	6650
Desvio Padrão	423559,18	258555,07	433045,16	2,23	6977,24	1556,93	2,65	2603,23
Min	677158	28337	609048	27,09	7419	1564	20,36	2509
Max	2841410	971992	2752364	36,77	33428	7205	34,94	11810
Erro Padrão	28556,35	17431,77	29195,89	0,35	470,40	104,96	0,18	175,50

Anexo- Tabela 2. Resultado do sequenciamento ddRADseq para *An. darlingi* de Granada e Remansinho em Março de 2012. Detalhes por indivíduo do número de sequências genotipadas.

	Amostra	Total de Sequências	Sequências Ambíguas/Baixa Qualidade	Sequências Retidas	% Sequências Alinhadas	Sequências únicas	Loci Polimórficos	% Loci Polimórficos	SNPs Encontrados
Granada	A12G013	1032342	281891	750451	28,97	10947	4396	40,16	7002
	A12G014	1160256	292396	867860	31,80	12526	5595	44,67	8981
	A12G015	1462112	314999	1147113	32,70	16898	8377	49,57	13672
	A12G016	1302342	257728	1044614	32,03	15508	7593	48,96	12253
	A12G017	1199492	232218	967274	33,05	14418	6859	47,57	11064
	A12G018	1212746	235230	977516	33,50	14324	6504	45,41	10384
	A12G019	854694	170186	684508	33,02	10343	4794	46,35	7385
	A12G020	700138	130310	569828	32,36	8599	3902	45,38	6091
	A12G022	977500	188003	789497	34,22	12102	5398	44,60	8553
	A12G023	895912	182425	713487	35,41	10749	4241	39,45	6673
Remansinho	A12R011	962416	165712	796704	35,86	14964	4250	28,40	7302
	A12R012	1021264	179363	841901	37,17	15303	4248	27,76	7082
	A12R014	671234	115238	555996	36,57	11991	3618	30,17	6032
	A12R015	881228	159109	722119	36,03	13513	3788	28,03	6296
	A12R016	1307338	232064	1075274	34,34	16715	4798	28,70	8163
	A12R020	1022774	175224	847550	36,24	15240	4425	29,04	7443
	A12R021	785294	130575	654719	34,98	12522	3821	30,51	6384
	A12R022	694708	112427	582281	35,97	11858	3440	29,01	5666
	A12R024	937214	162034	775180	37,94	13905	4123	29,65	6914
	A12R026	1099312	186515	912797	34,54	16149	4876	30,19	8202
	A12R028	798908	135083	663825	36,02	12278	3502	28,52	5912
	A12R030	797328	138365	658963	33,70	12338	3567	28,91	6157
	A12R031	871308	144422	726886	33,04	13187	3811	28,90	6532
	A12R033	1508602	246142	1262460	33,64	16783	4818	28,71	8271

Média	1006519,25	190319,12	816200,12	34,30	13465	4781	35,77	7850,58
Desvio Padrão	234704,99	58037,50	186430,28	2,04	2222,75	1329,05	8,42	2090,87
Min	671234	112427	555996	28,97	8599	3440	27,75	5666
Max	1508602	314999	1262460	37,94	16898	8377	49,57	13672
Erro Padrão	47908,95	11846,85	38054,92	0,42	453,71	271,29	1,72	426,79

Anexo- Tabela 3. Resultado do sequenciamento ddRADseq para *An. darlingi* de Granada e Remansinho em Maio de 2012. Detalhes por indivíduo do número de sequências genotipadas.

	Amostra	Total de Sequências	Sequências Ambíguas/ Baixa Qualidade	Sequências Retidas	% Sequências Alinhadas	Sequências únicas	Loci Polimórficos	% Loci Polimórficos	SNPs Encontrados
Granada	B12G004	1544108	666145	877963	32,89	10817	3351	30,98	5638
	B12G005	1101752	435090	666662	35,13	9175	2727	29,72	4572
	B12G006	2343330	863314	1480016	27,27	15369	7482	48,68	12594
	B12G009	1592710	579277	1013433	29,25	13476	3902	28,96	6578
	B12G010	1192662	462903	729759	31,71	11008	3277	29,77	5458
	B12G029	1469440	572374	897066	36,02	12269	3671	29,92	6213
	B12G032	2482554	864938	1617616	22,99	13482	3765	27,93	6330
	B12G034	2756898	1062446	1694452	24,19	13525	3735	27,62	6362
	B12G035	1830182	724609	1105573	28,01	12156	3535	29,08	5996
	B12G036	1614986	653287	961699	30,87	11946	3344	27,99	5781
	B12G038	2151912	910121	1241791	27,78	12290	3420	27,83	5740
	B12G042	1676572	672921	1003651	27,68	12154	3724	30,64	6127
	B12G043	1870188	779402	1090786	24,80	11534	3188	27,64	5377
Remansinho	B12R020	2066510	492578	1573932	26,67	17934	8107	45,20	13501
	B12R021	2287746	337225	1950521	32,95	20852	6464	31,00	10963
	B12R022	2269458	385148	1884310	36,68	22872	10124	44,26	16964
	B12R023	2872556	355138	2517418	32,29	19831	5910	29,80	9914
	B12R024	2556684	303118	2253566	31,83	17921	5300	29,57	8911
	B12R025	2118638	279166	1839472	31,42	11567	3731	32,26	6032
	B12R026	2398338	362229	2036109	32,41	24337	9073	37,28	15260
	B12R027	2332556	326653	2005903	37,83	22662	9076	40,05	15139
	B12R029	4681170	723510	3957660	34,30	30790	13665	44,38	23564
	B12R030	1807220	197782	1609438	33,20	13961	3947	28,27	6774
	B12R031	2828864	396820	2432044	32,80	24597	10212	41,52	17363
	B12R032	3184864	499515	2685349	33,61	25865	11923	46,10	20708
	B12R035	2907916	651756	2256160	34,05	21817	9063	41,54	15344
	B12R036	3633000	722549	2910451	33,36	28222	14385	50,97	24734
	Média	2280474,59	565926,44	1714548,15	31,18	17127	6300,04	34,78	10664,33

Desvio Padrão	773535,64	221948,11	772475,32	3,86	6141,58	3505,10	7,67	6034,04
Min	1101752	197782	666662	22,99	9175	2727	27,62	4572
Max	4681170	1062446	3957660	37,83	30790	14385	50,97	24734
Erro Padrão	148867	42713,93	148662,94	0,74	1181,95	674,56	1,48	1161,25

Anexo- Tabela 4. Resultado do sequenciamento ddRADseq para *An. darlingi* endo/exofágicos em Remansinho. Detalhes por indivíduo do número de sequências genotipadas.

	Amostra	Total de Sequências	Sequências Ambíguas/ Baixa Qualidade	Sequências Retidas	% Sequências Alinhadas	Sequências Únicas	Loci Polimórficos	% Loci Polimórficos	SNPs Encontrados
Exofágico	A11R001	766700	154761	611939	40,73	11798	3542	30,02	5971
	A11R002	1188546	223243	965303	73,35	4644	1485	31,98	2417
	A11R003	880266	168676	711590	41,06	13115	3883	29,61	6566
	A11R004	805712	170683	635029	49,56	7889	2437	30,89	4089
	A11R005	1104246	187852	916394	36,98	14875	4222	28,38	7054
	A11R006	1275964	247970	1027994	38,39	16958	4962	29,26	8337
	A11R007	920928	167732	753196	52,24	11219	3389	30,21	5653
	A11R008	750862	147932	602930	36,9	10252	3328	32,46	5531
	A11R009	745762	132644	613118	38,66	11688	3553	30,40	5969
	A11R010	993456	213665	779791	46,9	12817	3957	30,87	6663
	A11R011	987798	201724	786074	62,04	8108	2485	30,65	4140
	A11R012	1014508	191331	823177	53,12	9916	2995	30,20	5048
	A11R013	856148	138137	718011	48,75	11283	3306	29,30	5547
	A11R014	938760	165558	773202	44,85	12092	3570	29,52	6097
	A11R015	986842	154243	832599	43,94	13334	3759	28,19	6351
	A11R016	961332	147922	813410	53,57	11808	3456	29,27	5949
Endofágico	A11R151	1207012	223748	983264	34,41	9124	2458	26,94	4239
	A11R152	1649652	298997	1350655	22,91	11031	3073	27,86	5323
	A11R153	1073860	171662	902198	21,46	11043	3065	27,76	5277
	A11R154	1154636	203058	951578	36,3	11412	3186	27,92	5375
	A11R155	1435670	251998	1183672	34,18	17601	4910	27,90	8281
	A11R156	1507744	236887	1270857	31,49	14953	4216	28,20	7329
	A11R157	1355672	252438	1103234	28,56	14239	4102	28,81	6994
	A11R158	1064252	232355	831897	28,55	9079	3361	37,02	5580
	A11R159	1243968	214058	1029910	27,8	13222	4317	32,65	7327
	A11R160	1142804	195638	947166	30,5	16090	4593	28,55	7854
	A11R161	1273096	267922	1005174	29,45	9176	3147	34,30	5344
	A11R162	1622878	291280	1331598	26,65	15745	4509	28,64	7691
	A11R163	1918922	343755	1575167	30,49	16089	4365	27,13	7470
	A11R164	1999590	349187	1650403	30,98	17717	4735	26,73	8138
	A11R165	1642172	277986	1364186	37,99	19180	5266	27,46	9018
	A11R166	1617696	279475	1338221	27,11	16269	4708	28,94	8043

Média	1190232,94	215766,16	974466,78	38,75	12617,69	3698,13	29,62	6270,78
Desvio Padrão	335195,67	58147,54	280458,25	11,71	3324,90	855,46	2,22	1477,06
Min	745762	132644	602930	21,46	4644	1485	26,73	2417
Max	1999590	349187	1650403	73,35	19180	5266	37,02	9018
Erro Padrão	59254,78	10279,13	49578,48	2,07	587,77	151,23	0,39	261,11

Anexo- Tabela 5. Resultado do sequenciamento ddRADseq para *An. darlingi* exofágicos em Granada. Detalhes por indivíduo do número de sequências genotipadas.

	Amostra	Total de Sequências	Sequências Ambíguas/Baixa Qualidade	Sequências Retidas	% Sequências Alinhadas	Sequências únicas	Loci Polimórfico	% Loci Polimórfico	SNPs Encontrados
	A11G066	394064	76535	317529	21,53	3988	550	13,79	937
	A11G067	502104	109288	392816	27,19	4712	376	7,98	606
	A11G068	406356	79670	326686	28,34	4636	824	17,77	1405
	A11G069	2184654	312508	1872146	24,74	9376	1765	18,82	3048
	A11G070	1387186	183272	1203914	26,39	11221	2170	19,34	3709
	A11G071	1109734	214357	895377	40,64	5983	707	11,82	1166
	A11G072	1371834	213739	1158095	35,42	8044	1074	13,35	1771
	A11G073	980908	151474	829434	34,88	11213	2148	19,16	3690
	A11G074	609658	133205	476453	28,27	7518	1389	18,48	2370
	A11G075	611494	130521	480973	34,74	7289	1267	17,38	2156
	A11G076	726138	160578	565560	29,86	7659	1380	18,02	2338
	A11G077	1050010	247388	802622	28,33	7191	956	13,29	1514
	A11G078	1252044	281244	970800	33,13	4154	296	7,13	441
	A11G079	564856	111824	453032	27,92	6401	1226	19,15	1981
	A11G080	1312714	195392	1117322	28,08	13123	4228	32,22	7245
	A11G101	437750	109077	328673	29,9	3764	323	8,58	515
	A11G102	1449390	163190	1286200	25,5	14612	3270	22,38	5628
	A11G103	886436	118386	768050	32,17	10673	2297	21,52	3929
	A11G105	2930560	439499	2491061	33,31	11275	1793	15,90	2962
	A11G107	385234	68945	316289	30,65	2521	173	6,86	262
Crepúsculo	A11G108	339608	49691	289917	42,76	4883	578	11,84	918
	A11G109	517180	102373	414807	30,45	2573	163	6,34	263
	A11G110	351424	54240	297184	33,8	4466	301	6,74	502
	A11G111	289200	54908	234292	32,04	2294	139	6,06	210
Amanhecer	A11G170	531320	70454	460866	45,51	4542	338	7,44	548
	A11G171	723206	90696	632510	51,68	8368	1619	19,35	2746
	A11G172	496474	57256	439218	42,29	6891	955	13,86	1639
	A11G173	655670	79062	576608	30,58	7720	3174	41,11	5276
	A11G174	981246	102772	878474	47,78	12259	3233	26,37	5452

A11G175	851246	98695	752551	47,99	11693	2927	25,03	4987
A11G176	676984	71143	605841	47,51	9095	1828	20,10	3129
A11G177	747092	93070	654022	37,24	10258	1952	19,03	3309
A11G178	826530	158889	667641	33,79	9671	1742	18,01	2985
A11G179	871338	101617	769721	37,55	11747	2173	18,50	3617
A11G180	584488	60331	524157	36,03	9521	2063	21,67	3446
A11G181	848830	83847	764983	50,96	10621	2271	21,38	3857
A11G182	821150	129448	691702	41,69	11415	2671	23,40	4520
A11G183	595204	57790	537414	49,96	7890	2203	27,92	3625
A11G184	332040	42446	289594	37,01	4498	1307	29,06	2158
A11G185	339536	32810	306726	44,57	5303	1545	29,13	2594
A11G186	267846	35050	232796	39,15	3107	866	27,87	1415
A11G187	508960	52538	456422	23,13	4069	1142	28,07	1877
A11G188	325324	30776	294548	14,45	1176	366	31,12	582
A11G190	639238	60479	578759	53,81	7129	1901	26,67	3192
A11G191	482588	53279	429309	31,61	6439	1669	25,92	2771
A11G193	322134	35094	287040	43,45	4581	1237	27,00	2042
A11G194	380736	47711	333025	27,78	4746	1220	25,71	2058
A11G197	268450	31251	237199	35,5	3905	1135	29,07	1788
A11G198	433170	54368	378802	31,85	6660	1863	27,97	3058
A11G200	424056	45935	378121	37,69	6225	1706	27,41	2884
A11G201	320432	40254	280178	31,83	4306	1149	26,68	1926
A11G204	474732	48579	426153	41,34	7035	1966	27,95	3286
A11G205	578188	67444	510744	36,13	8264	2250	27,23	3742
A11G206	446400	55808	390592	34,01	5873	1650	28,09	2736
Média	718613,78	106485,11	612128,67	35,26	7121,78	1509,52	20,41	2533,54
Desvio Padrão	484114,40	79759,41	411641,94	8,32	3141,72	902,48	8,03	1543,21
Min	267846	30776	232796	14,45	1176	139	6,06	210
Max	2930560	439499	2491061	53,81	14612	4228	41,11	7245
Erro Padrão	65879,63	10853,88	56017,37	1,13	427,53	122,81	1,09	210

Anexo- Tabela 6. Resultado do sequenciamento ddRADseq para *An. darlingi* no intradomicílio em Granada. Detalhes por indivíduo do número de sequências genotipadas.

	Amostra	Total de Sequências	Sequências Ambíguas/ Baixa Qualidade	Sequências Retidas	% Sequências Alinhadas	Sequências únicas	Loci Polimórficos	% Loci Polimórficos	SNPs Encontrados
Crepúsculo	A11G216	983342	151268	832074	32,08	13366	3557	26,61	6128
	A11G217	951378	129676	821702	12,38	6226	1431	22,98	2387
	A11G218	864416	133858	730558	23,3	10243	2797	27,31	4652
	A11G221	967144	90825	876319	23,52	10702	2779	25,97	4700
	A11G222	569726	63226	506500	21,69	7408	1746	23,57	2891

	A11G223	948518	138775	809743	25,93	11363	3134	27,58	5347
	A11G225	976272	163146	813126	25,51	10469	3013	28,78	5127
	A11G226	1009386	153420	855966	22,92	10360	2779	26,82	4765
	A11G227	807746	82350	725396	24,99	8139	1976	24,28	3303
	A11G228	968662	140020	828642	28,33	10976	2868	26,13	4860
	A11G229	912996	134263	778733	28,58	10422	2785	26,72	4790
	A11G230	1163968	182217	981751	25,29	11771	3231	27,45	5604
	A11G231	1310522	205120	1105402	22,6	12026	3209	26,68	5454
	A11G232	1130066	123454	1006612	16,08	8658	2191	25,31	3678
	A11G233	958264	95993	862271	21,5	7312	1870	25,57	3169
	A11G235	1407442	122690	1284752	13,2	8476	2158	25,46	3601
	A11G236	738372	123330	615042	22,83	8345	2309	27,67	3829
	A11G237	578534	54064	524470	23,56	4655	1091	23,44	1745
	A11G238	538472	53019	485453	22,45	6064	1580	26,06	2605
Amanhecer	A11G287	864574	151399	713175	36,59	13027	3732	28,65	6454
	A11G288	782516	132395	650121	34,08	13350	3860	28,91	6481
	A11G289	707680	111634	596046	37,71	11801	3528	29,90	5870
	A11G290	964964	154083	810881	35,27	13271	3576	26,95	6087
	A11G291	1114978	179765	935213	37,49	14561	4184	28,73	7213
	A11G292	678758	112635	566123	39,38	10988	3111	28,31	5259
	A11G293	1598850	253831	1345019	39,21	16091	4409	27,40	7508
	A11G294	1268794	222619	1046175	36,51	14313	3934	27,49	6715
	A11G295	1298098	221040	1077058	39,69	14984	4178	27,88	7040
	A11G296	3675958	666013	3009945	32,35	19231	5220	27,14	9027
	A11G297	1234226	201059	1033167	38,2	14484	3925	27,10	6664
	A11G298	1572840	287610	1285230	41,58	13845	3781	27,31	6437
	A11G299	1120990	184105	936885	36,14	15069	4115	27,31	7032
	A11G300	1807618	287881	1519737	39,35	16287	4474	27,47	7773
	A11G301	1088334	184823	903511	43,51	12598	3443	27,33	5804
	A11G302	1111120	200835	910285	42,1	11717	3476	29,67	5910
	A11G303	1118010	180086	937924	37,02	12242	3271	26,72	5498
	A11G304	1825226	276677	1548549	41,29	15208	5070	33,34	8771
	A11G305	1122844	178599	944245	33,17	13634	3678	26,98	6179
	A11G306	1086662	178528	908134	37,08	13411	4220	31,47	7139
	A11G307	1164368	196852	967516	45,56	13609	4058	29,82	6812
	A11G308	1836158	303653	1532505	34,47	19974	5386	26,97	9213
	A11G309	1119720	210264	909456	44,79	14769	4271	28,92	7189
	Média	1141631,24	176597,62	965033,62	31,41	12034,4	3319,14	27,29	5635,95
	Desvio Padrão	509852,35	98770,54	415621,81	8,79	3331,72	1020,81	1,97	1776,84
	Min	538472	53019	485453	12,38	4655	1091	22,98	1745
	Max	3675958	666013	3009945	45,56	19974	5386	33,34	9213
	Erro Padrão	78671,93	15240,62	64131,84	1,36	514,10	157,51	0,30	274,17

Anexo- Texto 1. *Script* utilizado para o processamento dos dados com Stacks v.1 (Catchen et al. 2013).

```
# Filtrar as sequências com o process_radtags

mkdir $src/filtered_seqs
process_radtags -f $src/raw_sequences -o $src/filtered_seqs \
  -i fastq -c -q -r

# Alinhar os resultados com o genoma de An. darlingi
## Arquivos SAM

mkdir $src/rad_alignments
for f in $src/filtered_seqs/*.fastq
do
  echo "Running bowtie alignments on: $f"
  bowtie /genomes/Anopheles_darlingi/AdarC3 $f -p 30 --sam >
  ./rad_alignments/${basename $f |
    sed 's/\.fastq//'} .sam 2>>./rad_alignments/alignments.log
done

# Correr o programa ref_map.pl

ref_map.pl -o $stacks -n 0 -T 30 -m 5 -O popmap -b 1 -S \
  -B darlingi_radtags -D 'Melinas RAD database' \
## Preparar essa função para cada amostra em *

-s ./rad_alignment/*.sam \

# Correr o modo de correção rxstacks

rxstacks -b 1 -P $src/stacks -o ./stacks_cor --prune_haplo \
  --lnl_lim -10.0 -t 8 --verbose --lnl_dist

cstacks -b 1 -p 15 -o ./stacks_cor \

## Preparar essa função para cada amostra *
-s ./stacks_cor/* \

while read f; do
  sstacks -p 60 -b 1 -c ./stacks_cor/batch_1 \
  -s ./stacks_cor/${f} \
  -o ./stacks_cor/ &>> ./stacks_cor/Log
done < files

# Consider todas as amostras como uma única população

populations -b 1 -P ./stacks_cor -M ./popmap -t 20 -r 0.5 -m 5 --fstats \
  --write_single_snp --phylip --structure --log_fst_comp --vcf --genepop

# Correr o Structure usando runStructure.py

grep -v "#" $src/stacks_cor/batch_1.structure.str $src/structure_input.str
./runStructure.p -i $src/structure_input.str -o $src/structure -p 10 -b 100000
\
  -m 1000000 -x 40 -k 40 -n 60
```

Anexo- Texto 2. *Script* utilizado para a conversão de uma lista de genes para uma lista de GO.

```
#!/usr/bin/env python
#
# geneID2GO

import sys

# main()
#
def main(cmdline=None):
    infile = open(sys.argv[1], "r")

    # Cria um dicionario com chaves de geneIDs e valores como a lista de termos
    GO

    d = {}
    for line in infile:
        line = line.strip("\n").split("\t")
        if line[0] in d:
            d[line[0]].append(line[1])
        elif line[1] != "":
            d[line[0]] = [line[1]]
    infile.close()

    # Print dicionario de maneira apropriada.
    for geneID in d:
        print "{}\t{}".format(geneID, ",".join(d[geneID]))
    return 0

# make_parser()
#

if __name__ == "__main__":
    sys.exit(main(sys.argv[1:]))
```
