

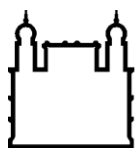
MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Mestrado em Biologia Computacional e Sistemas

**Análise do comportamento de oviposição do mosquito
Aedes aegypti em Paquetá/RJ**

LETÍCIA FERREIRA LIMA

Rio de Janeiro
Março de 2017



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

LETÍCIA FERREIRA LIMA

Análise do comportamento de oviposição do mosquito *Aedes aegypti* em Paquetá/RJ

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

Orientadora: Dr^a. Renata Schama
Laboratório de Biologia Computacional e Sistemas
IOC/Fiocruz

RIO DE JANEIRO

Março de 2017

Lima, Letícia Ferreira.

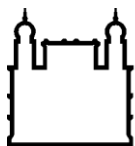
Análise do comportamento de oviposição do mosquito *Aedes aegypti* em Paquetá/RJ / Letícia Ferreira Lima. - Rio de Janeiro, 2017.
119 f.

Dissertação (Mestrado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2017.

Orientadora: Renata Schama.

Bibliografia: f. 63-80

1. *Aedes aegypti*. 2. Análise de parentesco. 3. Comportamento de oviposição. 4. Microsatélites. 5. Simulações computacionais. I. Título.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: Leticia Ferreira Lima

Análise do comportamento de oviposição do mosquito *Aedes aegypti* em Paquetá/RJ

ORIENTADORA **Dr^a. Renata Schama**

Laboratório de Biologia Computacional e Sistemas
IOC/Fiocruz

Aprovada em: 22/03/2017

EXAMINADORES:

Prof. Dr^a. Ana Carolina Paulo Vicente - **Presidente** (IOC/Fiocruz)

Prof. Dr. Luiz Guilherme Soares da Rocha Bauzer (IOC/Fiocruz)

Prof. Dr. Cristiano Valentim da Silva Lazoski (Departamento de Genética/UFRJ)

Prof. Dr. Marcio Galvão Pavan (IOC/Fiocruz)

Prof. Dr. Rodrigo Jardim Monteiro da Fonseca (IOC/Fiocruz)

Rio de Janeiro, 22 de Março de 2017

Dedico este trabalho aos meus pais José Carlos e Laura pelo esforço na minha educação e formação como pessoa e à minha orientadora Renata Schama que me ajudou a conduzir esse trabalho e foi importante para minha formação profissional.

AGRADECIMENTOS

Agradeço a todos do Laboratório de Biologia Computacional e Sistemas (LBCS), principalmente a Dr. Renata Schama pela orientação, dedicação e contribuição para conclusão do trabalho. Às amigos de laboratório, André Torres, (R)angeline Azevedo e Melise Chaves pela companhia durante esses dois anos e aos tecnólogos Fabio Motta e Rodrigo Jardim pela disponibilidade em ajudar.

À todos do Laboratório de Bioquímica da UERJ, principalmente o Dr. Rodolpho Albano e a bióloga Denise de Oliveira que colaboraram com a genotipagem e leitura das amostras.

Ao Dr. José Bento Lima e todos amigos do Laboratório de Fisiologia e Controle de Artrópodes Vetores (LAFICAVE), que tiveram participação durante o período de coleta em Paquetá, junto com os agentes de saúde, e me ajudaram na compreensão e importância do monitoramento de insetos vetores. Cynara Rodovalho, Diogo Bellinato, Raquel Santos, Paula Figliuolo, Luana Carrara e Ademir Martins.

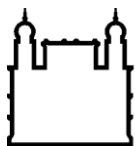
Às amigas não biólogas, que foram fundamentais em todo processo, Micaela Bella e Mayra Passigate obrigada pelos momentos de distração.

A todos os amigos biólogos da UERJ que permanecem juntos, dando suporte mesmo a distância, Alessandro, Barbara, Elaine, George, Letícia Caires, Tais, e a própria UERJ por me ensinar a importância da educação, na formação de bons professores e educadores e principalmente me ajudar a transferir todos os dias o pouco que sei e estar sempre disposta a aprender novas coisas a ensinar.

Agradeço também à minha família, Laura, Carlos e Felipe, pela dedicação, companheirismo e contribuição em todos os momentos da vida e a todos os primos e tios que mesmo sem saberem me apoiaram e me deram suporte.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo financiamento da bolsa de mestrado.

À todas as pessoas que contribuíram de alguma maneira para realização deste trabalho, e que direta ou indiretamente são responsáveis pela conclusão desta etapa da minha vida e pelo meu crescimento. Muito Obrigada !!



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

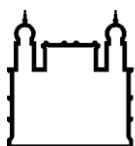
Análise do comportamento de oviposição do mosquito *Aedes aegypti* em Paquetá/RJ

RESUMO

DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Letícia Ferreira Lima

Aedes aegypti é um dos mais importantes insetos vetores, sendo responsável pela transmissão dos vírus da dengue, chikungunya e Zika no Brasil, causando atualmente grande impacto para a saúde pública. O mosquito é tradicionalmente controlado através do uso de inseticidas químicos, gerando uma pressão de seleção que faz com que o número de mosquito resistentes aumentem, comprometendo os programas de controle vetorial. Com isto, novas metodologias de controle estão ganhando impulso, o que torna importante o conhecimento do tamanho estrutura populacionais e do comportamento do vetor, como o comportamento de oviposição. Diversos estudos sobre esta espécie sugerem a existência da oviposição em saltos. Porém, em algumas populações estudadas o comportamento de oviposição em saltos não foi observado, e por isso, uma melhor avaliação se torna necessária. A seleção de um local de oviposição pela fêmea pode influenciar na sobrevivência da espécie e aumentar a dispersão do mosquito, sendo fundamental para a epidemiologia das arbovirose. Estudos de parentesco genético, são muito comuns para análise de comportamento em diversas espécies. Este trabalho, busca analisar o comportamento de oviposição do mosquito *Ae. aegypti*, através da análise de parentesco feita em indivíduos de uma mesma armadilha e entre armadilhas diferentes. Para essa inferência diversas metodologias se encontram disponíveis, mas, em sua maioria, são voltadas para determinar parentesco em populações naturais, onde ambos os indivíduos parentais e a prole estão na amostra, o que não é possível no estudo com mosquitos. Portanto, buscou-se avaliar, através de simulação, dois diferentes programas de reconstrução de parentesco para determinar qual melhor se adequa aos dados aqui trabalhados. Para isso, armadilhas foram instaladas no bairro de Paquetá/RJ, e as amostras coletadas foram analisadas com o uso de doze *loci* microssatélites. Os programas de inferência de parentesco Colony e ML-Relate foram comparados através da avaliação de dados simulados. Para as simulações, sete diferentes cenários foram criados, cada um com um total de 250 indivíduos e diferentes porcentagens de meio-irmão. Assim, foram inferidas a proporção de falsos positivos e negativos, possibilitando uma verificação da acurácia de cada programa. As simulações dos dados mostram que muitos indivíduos são categorizados de maneira incorreta nos diferentes cenários testados, mostrando que a presença de meio irmãos na amostra dificultam a categorização correta das díades. O programa ML-Relate apresentou uma acurácia de até 95%, mas taxas de classificação corretas de irmãos completos baixas. O programa Colony apresentou maior acurácia (98%), menores taxas de erro e altas taxas de classificações corretas. Portanto o programa Colony foi usado para a inferência do parentesco das amostras coletadas em Paquetá onde foi observada a presença de irmãos completos entre diferentes armadilhas instaladas no verão, indicando o comportamento de oviposição em saltos nas fêmeas de *Ae. aegypti*. O mesmo não foi observado nas amostras coletadas no inverno, o que indica uma menor dispersão das fêmeas neste período.



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Analysis of the oviposition behavior of *Aedes aegypti* mosquito in Paquetá / RJ

ABSTRACT

MASTER DISSERTATION IN COMPUTATIONAL BIOLOGY AND SYSTEMS

Letícia Ferreira Lima

Aedes aegypti is one of the most important insect vectors, being responsible for the transmission of dengue, chikungunya and Zika viruses in Brazil, causing great impact in public health. The mosquito population is traditionally controlled with chemical insecticides, generating a selection pressure that causes the number of resistant mosquitoes to increase compromising vector control programs. New control methodologies are being developed and require knowledge of the population size, structure and behavior of this vector, such as oviposition behavior. Several studies suggest that this mosquito species has a skip oviposition behavior, which can increase its dispersion. However, in some studied populations this type of behavior was not observed, demonstrating variability among populations for this trait. The selection of an oviposition site by the female can influence the survival of the species and is important in the epidemiology of arbovirus diseases. Genetic kinship studies are commonly used to investigate behavior in several species. The present work aims to analyze *Ae. aegypti* oviposition behavior through kinship analysis of individuals sampled within and between different traps in Paquetá, Brazil. For kinship inference several methodologies are available but most rely on both parental individuals and offspring being sampled together, which is not possible in mosquito studies. Therefore, here we evaluate the performance of two different kinship reconstruction programs for the data studied here. Traps were installed in the neighborhood of Paquetá / RJ, and the collected samples were analyzed with the use of twelve microsatellite *loci*. The programs Colony and ML-Relate were evaluated with simulated data. Seven different simulation scenarios were created, each with a total of 250 individuals and different half-sibling percentages. The number of false positives and negatives and accuracy of each program was compared. The simulations show that many individuals are categorized incorrectly in the different scenarios tested. Our results show that the number of half siblings dyads in comparison to full siblings is a problem in this type of analysis. The ML-Relate program showed an accuracy of up to 95% but low correct classification rates of full siblings. The program Colony showed higher accuracy (98%), lower error rates, and high correct classification rates. Therefore the program Colony was used to infer the relatedness of our sampled individuals. The presence of full siblings within and between different traps was observed in Paquetá during the summer, which indicates skip oviposition behavior of *Ae. aegypti*. During the winter no full sibling dyad was observed between different traps indicating smaller female dispersion.

ÍNDICE

RESUMO	vi
ABSTRACT	vii
1 INTRODUÇÃO	1
1.1 Patógenos transmitidos pelo mosquito <i>Aedes aegypti</i> (Linnaeus, 1762) e seu controle	1
1.2 Principais aspectos do mosquito <i>Aedes aegypti</i> Linnaeus, 1762	4
1.2.1 Comportamento de oviposição	8
1.2.2 Oviposição em saltos	10
1.3 Marcadores genéticos e o estudo de comportamento	12
1.4 Análise de parentesco através das análises genéticas	15
1.4.1 Simulações computacionais	19
2 OBJETIVOS	22
2.1 Objetivo geral.....	22
2.2 Objetivos específicos.....	22
3 METODOLOGIA	22
3.1 Coleta das amostras.....	22
3.2 Local de estudo.....	24
3.3 Extração de DNA	26
3.4 Amplificação do DNA e análise de dados genotipados.....	27
3.5 Análises de parentesco.....	30
3.6 Criação de dados através de simulação computacional	33
3.7 Análise comparativa dos programas de identificação de parentesco	34
4 RESULTADOS	36
4.1 Análise genética geral da população de Paquetá	36
4.2 Análise comparativa dos programas de identificação de parentesco	39
4.3 Análise de parentesco dos dados reais	44
4.3.1 Estrutura genética dos indivíduos capturados nas coletas de verão e inverno	44
4.3.2 Determinação do parentesco dos indivíduos capturados intra e inter armadilhas.....	48

5	DISCUSSÃO	52
5.1	Avaliação dos <i>loci</i> microssatélites usados.....	52
5.2	Dados simulados e comparação entre os programas.....	54
5.3	Comportamento de oviposição do mosquito <i>Aedes aegypti</i>.....	59
6	CONCLUSÕES	62
7	REFERÊNCIAS BIBLIOGRÁFICAS	63
8	ANEXO I	81
9	ANEXO II	84

ÍNDICE DE FIGURAS

Figura 1.1. Mapa da distribuição global de <i>Ae. aegypti</i> . As cores indicam a probabilidade de 0 a 1 (azul a vermelho) para ocorrência do mosquito. Retirado de Kraemer et al. (2015).....	5
Figura 1.2. Esquema mostrando o ciclo de transmissão das arboviroses em humanos onde (X) representa o vetor ou homem infectado e (✓) saudável.....	6
Figura 1.3. Ciclo de vida do mosquito <i>Aedes aegypti</i> (adaptada de Bon Eco Solutions , 2016).....	7
Figura 1.4. Mosquito distribuído globalmente e vetor de diversas arboviroses, <i>Aedes aegypti aegypti</i> (direita) e a subespécie africana, <i>Aedes aegypti formosus</i> (esquerda). Retirado de Powell (2016).....	7
Figura 1.5. Esquema mostrando os dois diferentes tipos de oviposição da fêmea de <i>Ae. aegypti</i> : (a) oviposição em reservatório preferencial e (b) oviposição em saltos.....	10
Figura 1.6. Esquema representando o sequenciamento de um locus microssatélite, mostrando o tipo de repetição (CA) (a), as diferenças no número de unidades de repetições de cada alelo (b), e a visualização da amplificação de cada alelo. Em (c), o eixo vertical a esquerda representa a escala de fluorescência e o eixo horizontal acima, o tamanho de cada alelo, em pares de base.....	15
Figura 1.7. Gráfico do coeficiente k esperado para diferentes categorias de relacionamento. U – não relacionados; HS – meio irmãos; 1C – primos de primeiro grau; 2C – primos de segundo grau; FS – irmãos completos; PO – pai e filho; M – gêmeos monozigóticos. O grau de relacionamento r aumenta para a direita e ao norte do gráfico. Adaptado de Kalinowski et al. (2006).	19
Figura 3.1. Mapa da ilha de Paquetá, Rio de Janeiro. Pontos rosa são os locais onde foram instaladas 50 armadilhas para coleta de ovos destinados a análise de parentesco e os pontos verdes somados ao rosa são locais onde ovitrampas foram instaladas as 90 armadilhas para coleta dos ovos destinados a análise da variabilidade da população.	24
Figura 3.2. Temperatura e precipitação mensal acumulada na ilha de Paquetá, de janeiro a dezembro de 2015. Fonte: INMET.....	26
Figura 3.3. Exemplo de matriz 5 x 5, com 250 indivíduos na prole e 10 % de meio irmãos. Diferentes configurações de relacionamentos possíveis para os indivíduos analisados. Exemplo de meio irmãos em roxo, dois grupos de larvas da prole onde os indivíduos são considerados meio irmãos por compartilharem a mesma mãe (fêmea 5). Exemplos de irmãos	

completos em vermelho, todas as 55 larvas são filhas do mesmo macho (1) e mesma fêmea (1).	34
Figura 4.1. Distribuição de r_{xy} nas comparações par a par para a categoria de relacionamento irmãos completos em todos cenários simulados (diferentes proporções de meio irmãos nas matrizes).	40
Figura 4.2. Erros tipo I e II para a categoria de irmãos completos encontrados em cada cenário (diferentes proporções de meio irmão) nos dois programas avaliados.	41
Figura 4.3. Proporção de irmãos completos (FS) verdadeiros classificados erroneamente em outras categorias e taxa de classificação correta para a categoria de irmãos completos em cada cenário (diferentes proporções de meio irmão) nos dois programas avaliados.	42
Figura 4.4. Acurácia de cada programa avaliado para os sete cenários simulados.	43
Figura 4.5. Grupos de irmãos completos distribuídos (indicados pelas diferentes cores) nas diferentes armadilhas coletadas no inverno em Paquetá e analisados no programa ML-Relate. Os números coloridos indicam quantos indivíduos de cada grupo estão na armadilha e as cores de cada círculo indicam a armadilha analisada sendo: preto (armadilha 4); roxo (armadilha 5); azul (armadilha 12), rosa (armadilha 13), vermelho (armadilha 28) e laranja (armadilha 45).	49
Figura 4.6. Grupos de irmãos completos distribuídos (indicados pelas diferentes cores) nas diferentes armadilhas coletadas no verão em Paquetá e analisados no programa Colony. Os números coloridos indicam quantos indivíduos de cada grupo estão na armadilha e as cores de cada círculo indicam a armadilha analisada sendo: azul (armadilha 19), preto (armadilha 20), branco (armadilha 21), vermelho (armadilha 22), rosa (armadilha 31), laranja (armadilha 32), amarelo (armadilha 37), roxo (armadilha 50).	51

LISTA DE TABELAS

Tabela 3.1. Marcadores microssatélites utilizados na população estudada. Nome do marcador, sequência do iniciador, grupo do PCR multiplex, cor da fluorescência usada para detecção na genotipagem e fonte bibliográfica: 1- Brown <i>et al.</i> (2011); 2 - Slotman <i>et al.</i> (2007).....	29
Tabela 3.2. Lista de coeficientes k e r para diversas categorias de relacionamento entre indivíduos, modificado de Blouin (2003).....	32
Tabela 4.1. Frequências alélicas (nomes dos alelos são seus tamanhos em pares de base no gel de sequenciamento) nos 12 <i>loci</i> de microssatélites analisados em <i>Ae. aegypti</i> . N - número de mosquitos analisados em cada <i>locus</i> ; Freq – frequência alélica.....	37
Tabela 4.2. Medidas de variabilidade encontradas para cada <i>loci</i> avaliado e valores do coeficiente de endogamia, F_{IS} , na população de <i>Ae. aegypti</i> de Paquetá. F_{IS} em negrito foram significativos e demonstram desvio para o equilíbrio de Hardy-Weinberg. (Al) número de alelos; (Na) riqueza alélica; (Ne) número efetivo de alelos; (Ho) heterozigosidade observada; (He) heterozigosidade esperada.....	38
Tabela 4.3. Número das armadilhas avaliadas no verão e inverno e quantidade de larvas genotipadas (total de larvas encontradas em cada armadilha). Os números das armadilhas correspondem aos números da figura 3.1.	44
Tabela 4.4. Frequências alélicas (nomes dos alelos são seus tamanhos em pares de base) nos 12 <i>loci</i> de microssatélites em <i>Ae. aegypti</i> coletados em diferentes estações do ano (inverno e verão) em Paquetá. N - número de mosquitos analisados em cada <i>locus</i>	46
Tabela 4.5. Medidas de variabilidade encontradas para cada <i>loci</i> avaliado e valores do coeficiente de endogamia, F_{IS} , na população de <i>Ae. aegypti</i> de Paquetá coletada no inverno (I) e verão (V). F_{IS} em negrito foram significativos e demonstram desvio para o equilíbrio de Hardy-Weinberg. (Al) número de alelos; (Na) riqueza alélica; (Ne) número efetivo de alelos; (Ho) heterozigosidade observada; (He) heterozigosidade esperada.....	47

LISTA DE SIGLAS E ABREVIATURAS

1C	Primos de primeiro grau
2C	Primos de segundo grau
AFLP	<i>Amplified Fragment Length Polymorphism</i>
AI	Número de alelos
DNA	Ácido desoxiribonucleico
FS	Irmãos completos
F_{is}	Coefficiente de endogamia
HS	Meio irmãos
H_0	Heterozigosidade observada
H_e	Heterozigosidade esperada
IBD	<i>Identity by Descent</i>
IBEX	Instituto de Biologia do Exército
LBCS Sistemas	Laboratório de Biologia Computacional e Sistemas
LAFICAVE Artrópodes Vetores	Laboratório de Fisiologia e Controle de Artrópodes Vetores
M	Gêmeos monozigóticos
N_a	Riqueza alélica
N_e	Número efetivo de alelos
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana de Saúde
PCR	Reação em cadeia da polimerase
PO	Pais e filhos
RAPD	<i>Random Amplified Polymorphic DNA</i>
RFLP	<i>Restriction Fragment Length Polymorphism</i>

SNPs

Single Nucleotide Polymorphisms

SSRs

Simple Sequence Repeats

U

Indivíduos não relacionados

1 INTRODUÇÃO

1.1 Patógenos transmitidos pelo mosquito *Aedes aegypti* (Linnaeus, 1762) e seu controle

No Brasil e no mundo diversas doenças que possuem um grande impacto para a saúde pública são causadas por patógenos transmitidos por artrópodes vetores, sendo mosquitos os principais agentes (Hemingway & Ranson 2000, Hotez *et al.* 2008). Dentre estes, a Zika, que tornou-se um problema de preocupação internacional para Organização Mundial da Saúde (WHO) devido à associação com casos de microcefalia (Klase *et al.* 2016, Mlakar *et al.* 2016) e distúrbios neurológicos (Araujo *et al.* 2016, Watrin *et al.* 2016). Nesse mesmo ano a transmissão do vírus que causa a Zika foi documentada em 48 países (WHO 2016), incluindo os Estados Unidos, onde até novembro de 2016 foram documentados 4.496 casos (CDC 2016). No Brasil, diversos casos de microcefalia foram reportados, além de 108 mortes desde o primeiro caso em 2015 (Campos *et al.* 2015, WHO 2016). Outro vírus da família Flaviviridae, transmitido pelo mosquito, é o da dengue, que é hoje uma doença de grande impacto na economia e saúde pública em áreas endêmicas (WHO 2009). Estima-se que mais de 40 % da população mundial esteja em risco de infecção por um ou mais sorotipos de dengue (Guzman *et al.* 2010). Hoje os quatro sorotipos (DENV1, DENV2, DENV3 e DENV4) encontram-se amplamente distribuídos no Brasil (Pontes & Ruffino-Netto 1994). O Brasil é especialmente vulnerável a epidemias de dengue, com dez vezes mais casos do que em outros países latino-americanos (Brathwaite Dick *et al.* 2012). No ano de 2016, o país apresentou uma incidência de 731,9

casos / 100 mil habitantes, sendo a região sudeste do país com maior número de casos confirmados, 57,6% do total do país (MS/SVS 2017).

O principal vetor dos vírus da dengue e da Zika no Brasil é o mosquito *Aedes aegypti*, que também é vetor para o vírus da febre amarela urbana (Consoli & Lourenço-de-Oliveira 1994) e chikungunya (WHO 2009, Coffey *et al.* 2014). Estas doenças são principalmente problemáticas em países tropicais e subtropicais, lugares onde o clima em conjunto com a pluviosidade e a temperatura são considerados fatores limitantes para a distribuição, e como consequência possui forte influência na dinâmica populacional do *Ae. aegypti* (Luz *et al.* 2009, Couret *et al.* 2014). A incidência das arboviroses pode ser influenciada por fatores que podem agir em quaisquer um dos seus três elementos: o hospedeiro vertebrado, o agente etiológico e o vetor. Devido à falta de uma vacina eficaz, atualmente os programas de controle de doenças como dengue, Zika e chikungunya dependem quase exclusivamente dos esforços de controle do vetor (Guzman *et al.* 2010, Maciel-de-Freitas *et al.* 2012). Para o controle do *Ae. aegypti* no Brasil, programas que estimulam o controle mecânico, ou seja, eliminação dos potenciais sítios de oviposição do mosquito são considerados os mais eficientes (MS/SVS 2009). Foi esta a metodologia utilizada em uma das primeiras campanhas implementadas em 1902 pelo sanitarista Oswaldo Cruz, buscando eliminação da febre amarela (Lowy 1990, Braga & Valle 2007a). Em 1958, a Organização Mundial da Saúde (OMS) em conjunto com a Organização Pan-Americana de Saúde (OPAS) conseguiu eliminar o mosquito em parte das Américas, inclusive no Brasil. No entanto, após cerca de 17 anos, houve a reintrodução do vetor no país (Claro *et al.* 2004, Monteiro *et al.* 2014) e os surtos epidêmicos de dengue vem aumentando em escala e periodicidade (Rigau-Pérez *et al.* 1998, Brady *et al.* 2012,

Brathwaite Dick *et al.* 2012), assim como novas doenças transmitidas por este mesmo vetor vem sendo introduzidas, como mencionado anteriormente.

Hoje, o controle das arboviroses no Brasil ocorre de forma integrada abrangendo vários aspectos da doença, como a assistência básica ao paciente, políticas de saneamento básico e combate ao vetor (Funasa 2002). Inseticidas químicos, principalmente piretroides e organofosforados, foram por muitos anos usados como o principal método de controle do *Ae. aegypti* após sua reintrodução no país. Porém, o uso intensivo destes produtos faz com que indivíduos resistentes, capazes de metabolizar ou tolerar estes compostos, sejam selecionados, comprometendo os esforços empregados. A resistência a inseticidas piretroides e organofosforados já foi detectada em diversas populações brasileiras do vetor (Braga & Valle 2007b, Montella *et al.* 2007, Linss *et al.* 2014, Bellinato *et al.* 2016). O comprometimento das estratégias de controle através do uso de inseticidas torna necessário o desenvolvimento de novas técnicas com o objetivo de diminuir as populações desta espécie (McGraw & O'Neill 2013). O uso em conjunto de diferentes métodos de controle se mostra como a forma mais eficaz de eliminação do vetor. No Brasil, os métodos de controle hoje compreendem o controle mecânico, o químico e o biológico, além de novas metodologias como a manipulação genética do vetor e o uso de bactérias endossimbiontes (Moreira *et al.* 2009, Maciel-de-Freitas *et al.* 2012, Carvalho *et al.* 2015, Dutra *et al.* 2015). Independente do método empregado, seu sucesso requer uma compreensão detalhada do tamanho, estrutura e taxas de migração das populações alvo. Portanto, o conhecimento sobre comportamento e a variabilidade e estruturação genética das populações de mosquito é fundamental para o desenvolvimento de programas racionais de controle do vetor.

1.2 Principais aspectos do mosquito *Aedes aegypti* Linnaeus, 1762

O principal vetor dessas arboviroses, o mosquito *Ae. aegypti*, é uma espécie invasora no país, muito bem sucedida que está amplamente distribuída pelo mundo com prevalência em regiões tropicais e subtropicais (Figura 1.1) (Consoli & Lourenço-de-Oliveira 1994). Segundo Braks *et al.* (2004), a densidade e distribuição da espécie está correlacionada à presença humana e por isso é predominante em áreas urbanas e periurbanas, com cobertura vegetal baixa, sendo considerada uma espécie antropofílica e adaptada a ambientes modificados (Consoli & Lourenço-de-Oliveira 1994). Essa estreita relação com os humanos e o fato de seus ovos serem muito resistentes à dessecação facilitou a dispersão do mosquito ao redor do mundo (Consoli & Lourenço-de-Oliveira 1994, Brown *et al.* 2011, Powell & Tabachnick 2013, Brown *et al.* 2014). Porém, não se pode dizer que sua distribuição seja cosmopolita pois é irregular e favorecida em regiões de clima quente e úmido, principalmente entre os paralelos 45N e 35S (Consoli & Lourenço-de-Oliveira 1994, Rigau-Pérez *et al.* 1998). Trata-se de uma espécie exótica, africana, que foi trazida inicialmente para as Américas e depois levada para a Ásia através do transporte passivo dentro de embarcações durante o período das colonizações (Christophers 1960, Powell & Tabachnick 2013, Brown *et al.* 2014, Gloria-Soria *et al.* 2016).

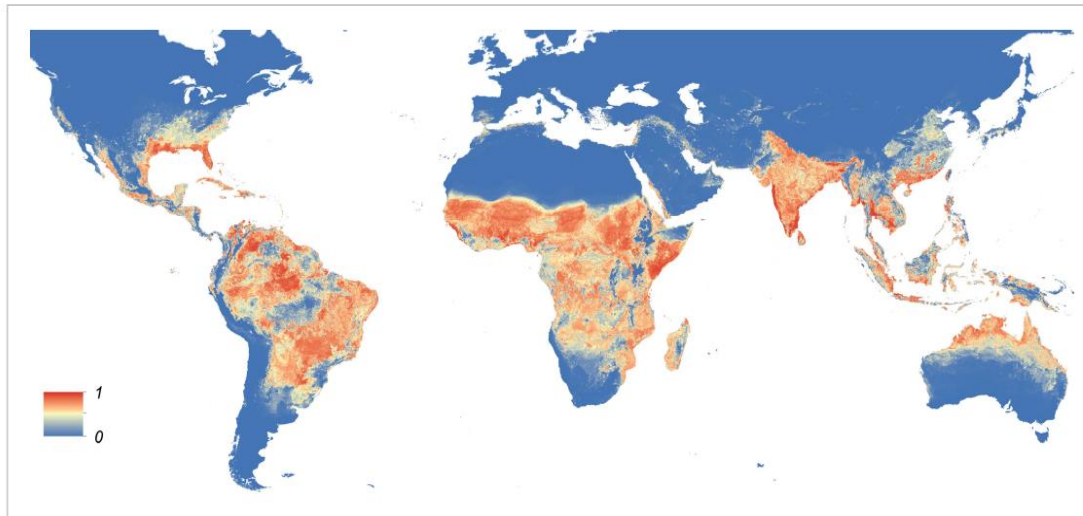


Figura 1.1. Mapa da distribuição global de *Ae. aegypti*. As cores indicam a probabilidade de 0 a 1 (azul a vermelho) para ocorrência do mosquito. Retirado de Kraemer *et al.* (2015).

A presença do mosquito vetor é essencial para que o ciclo de transmissão das arboviroses ocorra (Figura 1.2) (Christophers 1960). Como em outros holometábolos, o ciclo de vida do mosquito inclui quatro estágios de desenvolvimento: ovo, larva, pupa e adulto, sendo o estágio de larva composto de 4 instares (Figura 1.3). Na espécie *Ae. aegypti* este ciclo dura em torno de 14 dias dependendo das condições de temperatura e pluviosidade (Consoli & Lourenço-de-Oliveira 1994). Os ovos, com forma alongada e fusiforme, são sempre depositados próximos à superfície da água pela fêmea na parte interna de reservatórios artificiais ou naturais. Quando ocorre o contato dos ovos com a água esses eclodem e passam para a fase larval, período importante de alimentação e crescimento. Esta fase aquática é composta de quatro instares com duração de quatro dias dependendo da temperatura e disponibilidade de alimento (Farnesi *et al.* 2009), que consiste de microrganismos e detritos orgânicos (Forattini 1962). A fase de pupa é um período de transição, quando ocorre a metamorfose para o estágio de adulto e ocorre a

mudança do ambiente aquático para o terrestre. Nesse período de dois a três dias o inseto não se alimenta (Consoli & Lourenço-de-Oliveira 1994). Os adultos da espécie são alados e possuem características morfológicas marcantes, com todo o corpo ornamentado por manchas escuras alternadas com branco-prateadas e o escudo também ornamentado por escamas branco-prateadas formando um desenho de lira (Figura 1.4). O macho se diferencia da fêmea por possuir antenas plumosas e palpos longos (Christophers 1960, Consoli & Lourenço-de-Oliveira 1994, Funasa 2001).

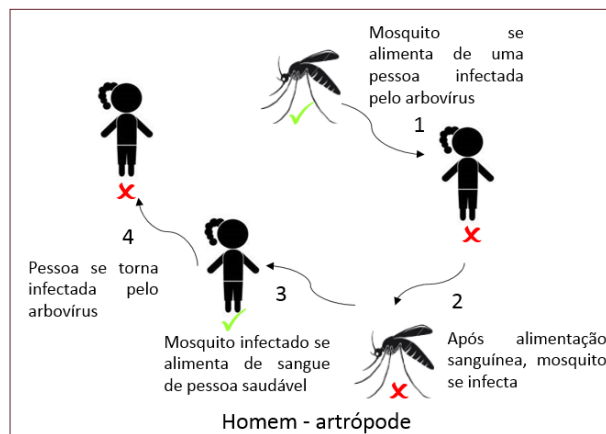


Figura 1.2. Esquema mostrando o ciclo de transmissão das arboviroses em humanos onde (X) representa o vetor ou homem infectado e (✓) saudável

Na natureza o espécime adulto pode viver por um período de 30 a 35 dias e utiliza a seiva como fonte de alimento (Consoli & Lourenço-de-Oliveira 1994). As fêmeas, no entanto, também fazem alimentação sanguínea, necessária para a maturação de seus ovos. É durante essa fase que ocorre a transmissão de diversos patógenos de importância para saúde pública. Em geral, após três dias do repasto sanguíneo ocorre a oviposição (Forattini 1962, Funasa 2001, MS/SVS 2009). As atividades de repasto e oviposição são realizadas no período da manhã ou próximo ao crepúsculo (Nelson 1986, Gentile *et al.* 2006, Gentile *et al.* 2009).

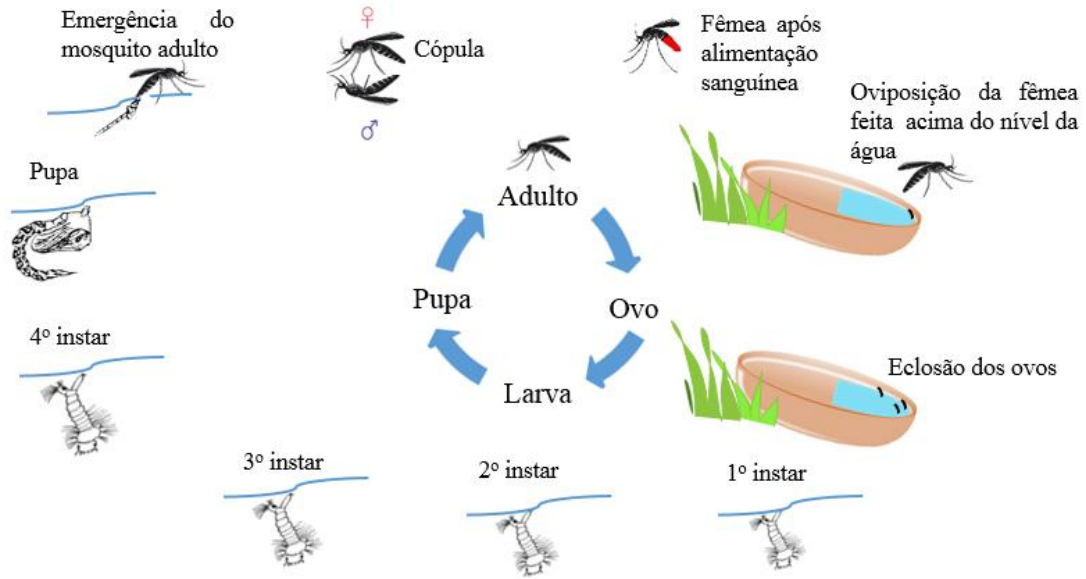


Figura 1.3. Ciclo de vida do mosquito *Aedes aegypti* (adaptada de *Bon Eco Solutions*, 2016).

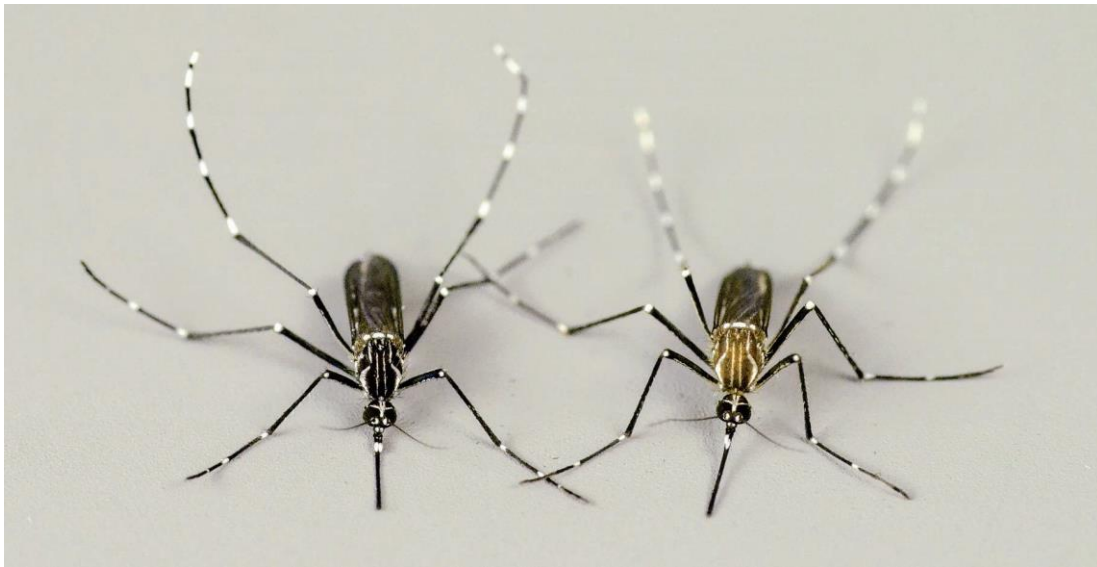


Figura 1.4. Mosquito distribuído globalmente e vetor de diversas arbovirose, *Aedes aegypti aegypti* (direita) e a subespécie africana, *Aedes aegypti formosus* (esquerda). Retirado de Powell (2016).

1.2.1 *Comportamento de oviposição*

A cada ciclo gonotrófico, período entre o repasto e a oviposição, as fêmeas depositam em torno de 100 a 200 ovos (Chadee *et al.* 2002). Uma fêmea pode passar por quatro a seis ciclos gonotróficos durante sua vida (Briegel *et al.* 2002, Chadee *et al.* 2002). Após o período embrionário, que dura cerca de 48 horas, os ovos adquirem a capacidade de resistir a longos períodos de dessecação, que podem durar até 450 dias (Rezende *et al.* 2008). Como os quatro estágios larvais e o de pupa ocorrem no sítio onde os ovos são colocados, a escolha deste sítio de oviposição é de extrema relevância para a sobrevivência da prole (Christophers 1960). Uma escolha ruim implica na perda de espécimes juvenis e de seu investimento reprodutivo anterior (Zahiri & Rau 1998). Portanto, a seleção de um local para oviposição por uma fêmea é um componente comportamental importante para a sobrevivência da espécie (Bentley & Day 1989) e com aspectos relevantes para a epidemiologia das arboviroses (Reiter 2007, Wong *et al.* 2011, Juliano *et al.* 2014).

As fêmeas podem aumentar a sobrevivência e desenvolvimento de sua prole escolhendo sítios de oviposição de forma seletiva em relação à presença de predadores, quantidade de larvas co-específicas presentes e comida disponível, entre outros aspectos do recipiente (Chadee 2009, Oliva *et al.* 2013, Panigrahi *et al.* 2014, Rey & O'Connell 2014). Populações de *Ae. aegypti formosus* (Figura 1.4), subespécie silvestre presente na África, costumam depositar seus ovos em buracos de árvores, axilas de folhas e cascas de frutos (Christophers 1960, Lounibos 1981). No entanto, a maioria das populações distribuídas mundialmente pertencem à subespécie *Ae. aegypti aegypti*, com hábitos antropofílicos, onde a fêmea utiliza recipientes frequentemente encontrados em ambientes humanos, como potes, pneus, garrafas, e vasilhames temporários (Powell 2016). No

entanto, foi observado uma preferência à oviposição no peridomicílio quando comparado ao intradomicílio na Austrália (Ritchie *et al.* 2003).

Vários fatores químicos, biológicos e físicos dos diferentes sítios potenciais de oviposição estão associados à escolha dos possíveis criadores (Harrington *et al.* 2008, Ponnusamy *et al.* 2008, Navarro-Silva *et al.* 2009, Wong *et al.* 2011). Por exemplo, Harrington *et al.* (2008) demonstraram que atributos físicos do criadouro, como maior volume e diâmetro estão correlacionados com um maior número de ovos encontrados. Os autores deste estudo sugerem que isso ocorre pois recipientes de maior volume podem armazenar uma maior quantidade de alimento e provavelmente são ambientes mais permanentes. Sítios também podem se tornar mais atrativos quando possuem a presença de bactérias que produzem metabólitos voláteis (Ponnusamy *et al.* 2008), outros ovos de mosquitos (Ganesan *et al.* 2006) e estágios imaturos da mesma espécie (Zahiri & Rau 1998, Wong *et al.* 2011). Porém, altas concentrações de larvas no ambiente podem se tornar um fator crítico devido à competição por alimento, o confinamento e a possível existência de patógenos no ambiente, passando a ser um fator de estresse, repelindo novas oviposições (Zahiri *et al.* 1997, Zahiri & Rau 1998). Realmente, alguns estudos mostram que uma concentração de larvas alta demais pode ser um fator que cause impacto negativo na sobrevivência das larvas (Braks *et al.* 2004, Wong *et al.* 2011). Este tipo de comportamento seletivo pode influenciar na distribuição das populações e na dinâmica populacional do vetor.

1.2.2 Oviposição em saltos

Além dos fatores relacionados ao sítio de oviposição, o próprio comportamento da fêmea pode acarretar uma maior ou menor distribuição de sua prole e conseqüentemente sua própria dispersão. Isto porque como uma fêmea pode colocar até 200 ovos por oviposição (ver acima) estes ovos podem ser colocados em apenas um sítio ou distribuídos em vários. Este tipo de comportamento é denominado de “oviposição em saltos” (Figura 1.5). A oviposição em saltos foi primeiro descrita para o mosquito da espécie *Wyeomyia smithii*, que habita o reservatório de água de plantas carnívoras (Mogi & Mokry 1980).

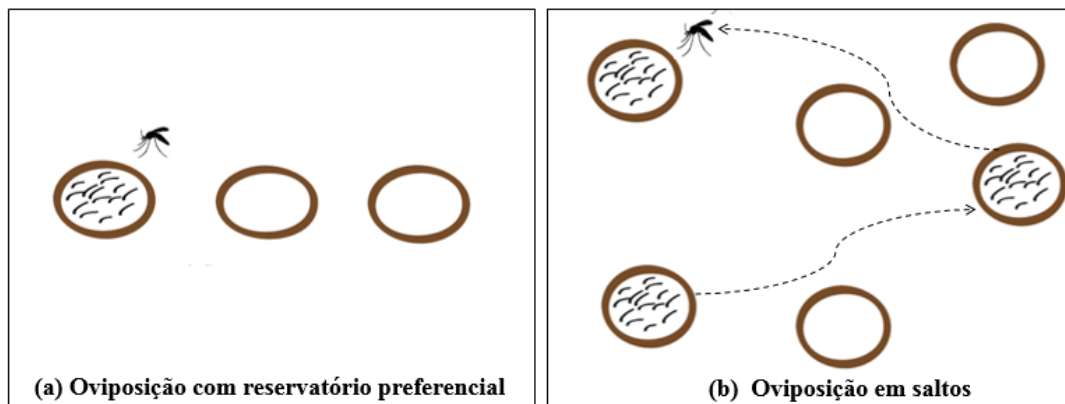


Figura 1.5. Esquema mostrando os dois diferentes tipos de oviposição da fêmea de *Ae. aegypti*: (a) oviposição em reservatório preferencial e (b) oviposição em saltos.

Na espécie *Ae. aegypti*, estudos genéticos (Apostol *et al.* 1993, Apostol *et al.* 1994, 1996, Colton *et al.* 2003) e de comportamento em laboratório (Oliva *et al.* 2013, Panigrahi *et al.* 2014) sugerem que fêmeas grávidas espalham seus ovos por diversos pontos (oviposição em saltos), aumentando as taxas de sobrevivência e dispersão da população (Reiter 2007). Por exemplo, Apostol *et al.* (1994) utilizando técnicas moleculares para verificar o número de famílias do mosquito *Aedes aegypti* dentro de uma armadilha observaram que, em geral, as fêmeas colocavam ovos em mais de uma armadilha. Neste

estudo os autores basearam o tamanho das famílias no número de irmãos completos, que possuem mesmo pai e mãe, encontrados dentro de cada armadilha. Foi observado que em 29 % dos casos as famílias dentro de uma única armadilha eram compostas de apenas um ou dois indivíduos, indicando que as fêmeas colocavam o restante dos seus ovos em outras armadilhas (Apostol *et al.* 1994). Colton *et al.* (2003), também utilizaram marcadores genéticos para determinar o número de irmãos completos dentro e entre armadilhas em Trinidad. Os autores observaram que, dos 35 grupos de irmãos completos encontrados, 19 estavam espalhados por mais de uma armadilha, mostrando que as fêmeas distribuíram seus ovos por mais de um sítio de oviposição. Foi observado em alguns casos que a escassez de criadouros adequados pode ser um dos fatores responsáveis pela maior dispersão dessas fêmeas (Edman *et al.* 1998).

No entanto, há indícios de que na Tailândia, por exemplo, fêmeas de *Ae. aegypti* não fazem a postura em diversos criadouros (Harrington & Edman 2001), o que evidencia variação no comportamento entre populações desta espécie. Além disso, nas diferentes populações avaliadas que possuíam o comportamento de oviposição em salto, uma grande gama de variação comportamental foi encontrada. Dependendo da região geográfica em que o estudo foi feito um número maior ou menor de criadouros foi visitado pelas diferentes fêmeas. Em um estudo feito em laboratório, utilizando populações brasileiras do vetor, foi possível observar que as fêmeas tendem a colocar cerca de 50 % de seus ovos em apenas um recipiente mas mantêm a oviposição em salto posteriormente (Oliva *et al.* 2013). Um resultado muito parecido foi encontrado por Abreu *et al.* (2015), estudando populações em condições de semi-campo em Minas Gerais, onde as fêmeas depositaram pelo menos 40 % dos seus ovos em um sítio chamado “predileto” e o resto era espalhado por outros sítios.

Toda esta gama de variação comportamental e a complexidade sobre a escolha dos sítios de oviposição faz com que uma melhor avaliação deste comportamento em diferentes populações de campo seja necessária.

1.3 Marcadores genéticos e o estudo de comportamento

Marcadores moleculares pode ser definido como um fenótipo molecular originado de algum gene específico do DNA ou de um gene expresso. Os marcadores podem ser herdados e apresentam uma variação entre diferentes indivíduos, sendo assim muito utilizados para diferenciar indivíduos geneticamente e são utilizados nos mais diversos estudos de ecologia, conservação e genética (Ferreira & Grattapaglia, 1998).

Um importante papel dos estudos de genética de populações é diminuir a “distância” entre a genética molecular e estudos ecológicos e comportamentais. A compreensão de maneira integrada desses assuntos é importante, por exemplo, para o conhecimento das taxas de dispersão, estudos de acasalamento e distribuição geográfica de espécies ameaçadas e assim, ampliando o entendimento da biologia evolutivas de diferentes espécies (Blouin 2003; Jones et al 2010, Wang, 2014). Para isso, diversos marcadores, cada qual com seus pontos positivos e negativos, foram desenvolvidos: RAPD (*Random Amplified Polymorphic DNA*), RFLP (*Restriction Fragment Length Polymorphism*), AFLP (*Amplified Fragment Length Polymorphism*), SSRs (*Simple Sequence Repeats*), SNPs (*Single Nucleotide Polymorphisms*) entre outros (Avisé 1994). A escolha do marcador utilizado vai depender de fatores como o custo, a tecnologia disponível e sua confiabilidade. Os mais usados atualmente para genética de populações são o sequenciamento direto de genes, SNPs (*Single Nucleotide Polymorphisms*) e microssatélites (ou SSRs, *Simple Sequence Repeats*).

Dados de microssatélites já foram usados em diversos estudos em diferentes espécies de mosquito e se mostraram valiosos para responder perguntas de difícil solução em populações naturais. Por exemplo, já se mostraram ótimos aliados para estudos relacionados com comportamento de oviposição em *Anopheles gambiae* (Chen *et al.* 2006) e *Anopheles arabiensis* (Chen *et al.* 2008). Com este tipo de marcador é possível, através da análise de parentesco das larvas encontradas dentro e entre armadilhas de oviposição, saber quantas fêmeas contribuíram para um determinado criadouro. Para o *Ae. aegypti*, uma variedade de *loci* microssatélites já foram identificados (Chambers *et al.* 2007, Slotman *et al.* 2007, Lovin *et al.* 2009) e avaliados em populações brasileiras (Monteiro *et al.* 2014).

Os microssatélites foram descobertos na década de 80 em pesquisas que observaram a presença em gel de duas bandas após um processo de centrifugação com cloreto de cério de um DNA genômico. Foi observado que uma das bandas possuía genes e a outra, com densidade diferente, DNA repetitivo (Tautz & Renz 1984, Tautz 1989). Os microssatélites são sequências de DNA com unidades de dois a seis pares de bases que se repetem sequencialmente. Estas regiões repetitivas possuem alto nível de polimorfismo e são abundantes, sendo encontrados em genomas de eucariotos em geral e estão localizadas tanto em regiões codificantes como em regiões intrônicas. Podem ser classificados de acordo com o tipo de sequência de repetição que apresentam, podendo ser: perfeitos, quando não apresentam interrupção em sua sequência (ex: TATATATATATAT) e imperfeitos, quando a sequência de repetição é interrompida por um par de bases ou uma pequena sequência (ex: TATATATACTATATAT) (Oliveira *et al.* 2006).

Os *loci* de microssatélites apresentam uma herança mendeliana e evoluem a partir de processos mutacionais decorrentes do deslizamento da polimerase e/ou do *crossing over* desigual, que levam a uma inserção ou deleção das sequências repetitivas no genoma, durante o processo de replicação (Ellegren 2004). Apresentam uma alta taxa de mutação que pode variar de 10^{-2} a 10^{-6} por geração, sendo mais alta que a taxa de substituição de bases em outras regiões do genoma (Li *et al.* 2002). Esse marcador possui vantagens quando comparados a outros marcadores genéticos (SNPs ou sequenciamento direto, por exemplo) pois são altamente polimórficos, polialélicos, sua herança é codominante (o que permite a diferenciação de homocigotos e heterocigotos), abundantes e usualmente neutros (Tautz 1989, Ellegren 2004). Para serem usados como marcadores genéticos, as regiões contendo as sequências repetidas são amplificadas utilizando iniciadores específicos complementares às sequências flanqueadoras dos SSRs. Os diferentes alelos de um *locus* possuem números diferentes de repetições e, portanto, são identificados por tamanho em eletroforese com gel de sequenciamento (Figura 1.6). As propriedades do DNA microssatélite e, principalmente, seu caráter polimórfico, os tornam úteis para estudos de mapeamento genético, diagnóstico de doenças, estrutura genética das populações, genética forense e relações de parentesco (Ellegren 2004).

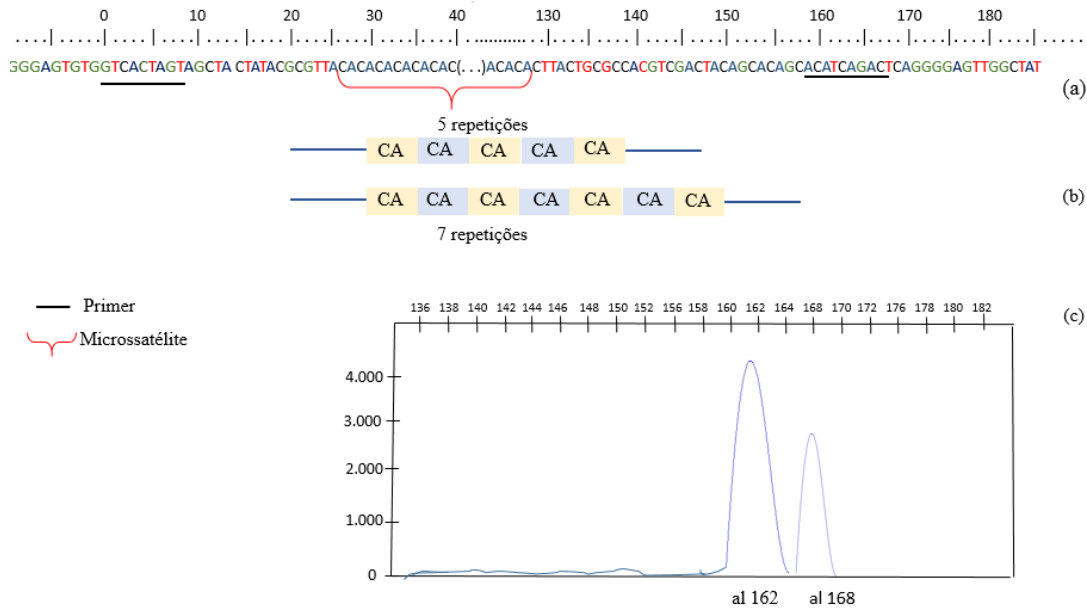


Figura 1.6. Esquema representando o sequenciamento de um *locus* microsatélite, mostrando o tipo de repetição (CA) (a), as diferenças no número de unidades de repetições de cada alelo (b), e a visualização da amplificação de cada alelo. Em (c), o eixo vertical a esquerda representa a escala de fluorescência e o eixo horizontal acima, o tamanho de cada alelo, em pares de base.

1.4 Análise de parentesco através das análises genéticas

As análises de parentesco têm ajudado a resolver questões importantes relacionadas à ecologia, evolução e comportamento de diferentes espécies. Ferramentas genéticas para análise de parentesco são importantes pois na maioria das vezes a genealogia dos grupos em estudo não é conhecida. Vários marcadores genéticos têm sido usados em estudos sobre diversidade ecológica, seleção sexual, sucesso reprodutivo, padrões de dispersão e biologia da conservação, entre outros (para revisões veja Blouin 2003, Jones *et al.* 2010). Além disso, diversas metodologias de identificação de parentesco genético estão disponíveis atualmente (Blouin *et al.* 1996, Blouin 2003, Jones *et al.* 2010). O parentesco genético

entre indivíduos pode ser calculado através de estimativas feitas a partir dos padrões de identidade por descendência ou IBD (*Identity by Descent*) onde dois alelos que descendem de um mesmo alelo ancestral são chamados de idênticos por descendência (Blouin 2003).

As inferências de parentesco podem ser de dois tipos: 1) os genótipos observados podem ser usados para distinguir e categorizar pares de indivíduos entre um conjunto de relacionamentos possíveis ou 2) os genótipos podem ser usados para se estimar o grau de parentesco genético através do cálculo de coeficientes específicos. Por exemplo, em análises de paternidade busca-se determinar quais pares de indivíduos podem ser categorizados como pais e filhos e diferenciá-los de pares de indivíduos não relacionados ou com outras categorias de relacionamento (como irmãos ou primos). Para tal, podemos calcular um coeficiente de parentesco (r) entre os pares de indivíduos, uma medida contínua, que nos diz a proporção de alelos compartilhados que são idênticos por descendência. De uma maneira simples, este coeficiente de parentesco nos diz a similaridade entre um par de indivíduos comparados com a média da população. Esta medida pode ser usada para se calcular estimativas de verossimilhança e com isso categorizar os pares de indivíduos. Diversas metodologias, tanto para o cálculo do coeficiente r quanto para a categorização dos pares de indivíduos, foram desenvolvidas (Blouin 2003, Weir *et al.* 2006, Jones *et al.* 2010, Jones & Wang 2010).

Para esse tipo de análise, o número de marcadores genéticos usados e sua variabilidade são de extrema importância para que as estimativas de parentesco sejam confiáveis (Blouin *et al.* 1996, Wang 2004, Ashley *et al.* 2009, Harrison *et al.* 2013b). Por exemplo, Ashley *et al.* (2009) utilizando populações simuladas, compararam cinco metodologias de reconstrução de parentesco e mostram que a acurácia dos programas

testados aumenta quando o número de *loci* é maior e quando o marcador possui mais alelos por *locus* (6-10). Harrison *et al.* (2013b), também utilizando dados simulados, mostrou que além do número de *loci* e diversidade alélica, a presença de erros de genotipagem podem influenciar a inferência de parentesco com diferentes metodologias. Como visto anteriormente, microssatélites já foram usados em vários estudos para inferência de parentesco genético em diferentes organismos. Realmente, seu poder de reconstrução de genealogias vai depender, principalmente, do número de *loci*, do número de alelos por *locus* e a diversidade encontrada (Ashley *et al.* 2009, Harrison *et al.* 2013b, Kopps *et al.* 2015).

Além do marcador utilizado, características da espécie sendo analisada, como tipo de acasalamento (poligâmico ou monogâmico) e história demográfica, também tem influência na confiabilidade das análises de parentesco (Robinson *et al.* 2013). Isto porque algumas categorias de relacionamento são mais fáceis de serem preditas pelos dados do que outras e o tipo de acasalamento e a existência de sobreposição de gerações vão influenciar na composição genealógica da população (Van de Casteele *et al.* 2001, Wang 2004, Csillery *et al.* 2006, Kopps *et al.* 2015).

As metodologias de inferência de parentesco foram criadas com foco principal em se determinar a estrutura familiar em populações naturais para ecologia da conservação, na qual tanto indivíduos parentais quanto filiais estão presentes nas amostras. No entanto, quando não temos sobreposição de gerações a categoria de relacionamento “pais e filhos” pode não estar presente na amostra, como é o caso da coleta de larvas de mosquito. Como estamos coletando apenas os estágios juvenis e não o mosquito adulto, não teremos os genótipos parentais para nos ajudar a determinar os grupos familiares. Este tipo de análise

torna mais importante a utilização de marcadores altamente polimórficos (Painter 1997, Harrison *et al.* 2013b).

A precisão da análise de parentesco genético também pode diminuir nos casos em que parentes muito próximos como primos ou até mesmo tios são inseridos na amostra (Olsen *et al.* 2001, Stadele & Vigilant 2016). Outro fator que pode afetar a inferência de parentesco é a presença de poligamia entre os indivíduos, pois os possíveis grupos familiares se tornam muito maiores, fazendo a análise computacionalmente mais exaustiva e aumentando as chances de erro (Wang & Santure 2009, Dexter & Brown 2013). De acordo com Helinski *et al.* (2012), *Ae. aegypti* possui um comportamento poliândrico. O acasalamento de fêmeas com mais de um macho também foi relatado na espécie do mesmo gênero *Ae. albopictus* (Boyer *et al.* 2012). Tanto a falta de parentais quanto a presença de meio irmãos e parentes distantes na amostra são fatores que podem dificultar e diminuir a precisão das metodologias de inferência de parentesco (Blouin *et al.* 1996, Olsen *et al.* 2001, Wang 2004, Csillery *et al.* 2006, Kopps *et al.* 2015).

Blouin *et al.* (1996) observaram, a partir de dados simulados, que a probabilidade de um par de indivíduos meio irmãos serem classificados erroneamente como irmãos completos seria de 15-16 %, enquanto que para classificar erroneamente irmãos completos como não relacionados seria de apenas 2,5 %. Realmente, categorias de relacionamento mais próximas na superfície do gráfico de r (Figura 1.7) são mais difíceis de serem distinguidas (Kalinowski *et al.* 2006). Por isso é importante testar, através de simulações, qual das metodologias de identificação de parentesco disponíveis atualmente (e seus pressupostos) melhor se adapta ao tipo de dados obtidos das populações naturais. Cada vez mais estudos sobre inferência de parentesco em populações naturais vêm mostrando a

importância de se estimar através de simulações a adequação da metodologia utilizada para a espécie e características dos marcadores sendo usados (Kopps *et al.* 2015, Taylor *et al.* 2015). É sempre importante determinar se a diversidade genética e características da história natural espécie específicas encontradas na população em estudo não atrapalham a inferência de parentesco pretendida. E se atrapalham é importante saber como.

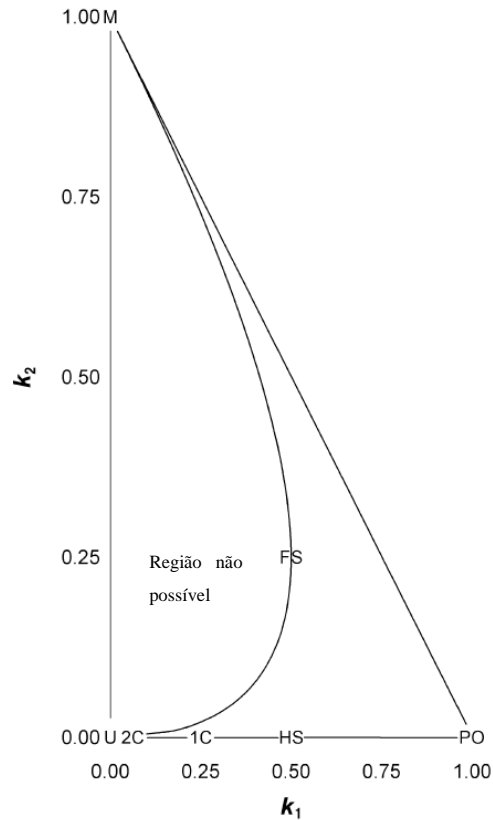


Figura 1.7. Gráfico do coeficiente k esperado para diferentes categorias de relacionamento. U – não relacionados; HS – meio irmãos; 1C – primos de primeiro grau; 2C – primos de segundo grau; FS – irmãos completos; PO – pai e filho; M – gêmeos monozigóticos. O grau de relacionamento r aumenta para a direita e ao norte do gráfico. Adaptado de Kalinowski *et al.* (2006).

1.4.1 Simulações computacionais

As duas razões básicas para se simular dados genéticos são: (1) maior conhecimento dos efeitos que parâmetros demográficos e mutacionais podem ter nos dados

reais observados ou (2) criar grupos de dados teste, ou cenários, para assim estimar o poder estatístico de métodos alternativos (Balloux 2001, Hoban *et al.* 2012). Na análise de parentesco, ao criar dados simulados com características parecidas com as populações reais, podemos comparar diferentes metodologias ou programas computacionais de análise de dados. Como nos dados simulados sabemos o relacionamento de parentesco entre os indivíduos criados é possível testar e comparar a acurácia, a eficiência e quanto robusto são os programas de parentesco e assim estimar o poder dos métodos avaliados (Hoban *et al.* 2012, Harrison *et al.* 2013a, Wang 2013).

Em simulações computacionais, dados com diferentes cenários a serem testados são gerados e, para que as simulações tenham um significado biológico, é importante que os dados gerados reflitam os dados reais. Para tanto, os conjuntos de dados criados devem considerar diversos aspectos biológicos das espécies a serem analisadas, especialmente características demográficas como tipo de acasalamento e sobreposição de gerações (Hoban *et al.* 2012). Características biológicas como separação dos sexos (dioica ou monoica) e poliandria podem dificultar as análises de inferência de parentesco e devem ser levadas em consideração quando dados simulados são usados para se determinar se as metodologias empregadas se adequam ao tipo de dados disponível (Wang & Santure 2009, Dexter & Brown 2013). Além disso, incluir dados sobre os marcadores usados como variabilidade, taxa de erro e de mutação e número de *loci* também são importantes para que possamos averiguar a validade da metodologia empregada para inferência de parentesco nos dados reais sendo analisados, como mencionado anteriormente.

No presente trabalho testamos quais metodologias atualmente empregadas para se estimar parentes dentro de uma amostra populacional melhor se adequam aos nossos dados.

Esses resultados serão importantes não somente para os dados de *Ae. aegypti* coletados no presente estudo, mas também para quaisquer estudos de comportamento de postura de ovos em mosquitos que possuam aspectos biológicos parecidos com a espécie em estudo. Em diversos vetores, como espécies dos gêneros *Phlebotomus* e *Anopheles*, um maior conhecimento deste componente comportamental das populações, que implica em diferentes taxas de sobrevivência e migração, tem se mostrado importante para uma melhor coordenação dos programas de controle de vetor (Le Menach *et al.* 2005, Kumar *et al.* 2013). Le Menach *et al.* (2005) utilizando modelos matemáticos, demonstraram que a oviposição é um fator essencial na distribuição de espécies do gênero *Anopheles* e, conseqüentemente, da epidemiologia da malária. Isto porque fêmeas deste mosquito tendem a se manter próximas dos sítios de oviposição e do hospedeiro para alimentação sanguínea. O conhecimento do comportamento do inseto pode ajudar na localização dos sítios de oviposição, e assim, auxiliar os programas de controle.

2 OBJETIVOS

2.1 Objetivo geral

Avaliar o comportamento de oviposição das fêmeas de *Aedes aegypti* em Paquetá, Rio de Janeiro, a partir da identificação do grau de parentesco das larvas e analisar através de simulação computacional qual programa de identificação de parentesco melhor se aos dados trabalhados aqui.

2.2 Objetivos específicos

- Determinar a variabilidade genética geral da população em questão (Paquetá/RJ);
- Genotipar e determinar o parentesco de indivíduos capturados inter-e intra-armadilhas;
- Inferir o comportamento sazonal de oviposição da espécie *Ae. aegypti*;
- Avaliar o efeito espacial e climático sobre comportamento de postura das fêmeas de *Ae. aegypti*;
- Criar dados com padrões similares de variabilidade genética e grau de parentesco (proporção maior de irmãos do que não parentes e sem indivíduos parentais nas amostras) através de simulação computacional;
- Comparar os programas Colony e ML-Relate quanto à adequação de cada um aos dados e ao nível de variabilidade e grau de parentesco encontrado na população estudada.

3 METODOLOGIA

3.1 Coleta das amostras

Larvas do mosquito *Ae. aegypti* foram coletadas em duas épocas do ano (Março e Julho, 2015) no bairro de Paquetá, Rio de Janeiro, utilizando armadilhas do tipo

ovitrampas, que é composta por um vaso preto (9 cm x 12 cm) e uma paleta de eucatex (20 cm x 5 cm), onde as fêmeas depositam seus ovos, e 200 ml de água com atrativo, levedo de cerveja a 0.04% (Fay & Eliason 1966, Braga *et al.* 2000) instaladas nos pontos mostrados no mapa (Figura 3.1). Para a análise da estrutura geral da população todas as 90 armadilhas permaneceram em campo por um período de sete dias. Para as análises de parentesco somente 50 armadilhas foram instaladas e coletadas após um período de três dias em campo. Esta distinção teve que ser feita devido à grande infestação de mosquitos na ilha. Isto porque, para a análise de parentesco, queríamos analisar todas as larvas de uma mesma armadilha, em mais de uma armadilha. Quando as armadilhas ficavam em campo por muitos dias, centenas de larvas eram coletadas em uma mesma armadilha, o que fazia com que o número de larvas a serem analisadas ficasse grande demais. As paletas positivas foram levadas para o Laboratório de Fisiologia e Controle de Artrópodes Vetores (LAFICAVE/IOC), onde foram imersas individualmente em copos descartáveis de 200 mL, contendo água desclorada e ração para gato triturada (Friskies®, Purina, São Paulo/SP). As larvas foram criadas até o 4º instar larvar (L4), identificadas e armazenadas em etanol 90%.

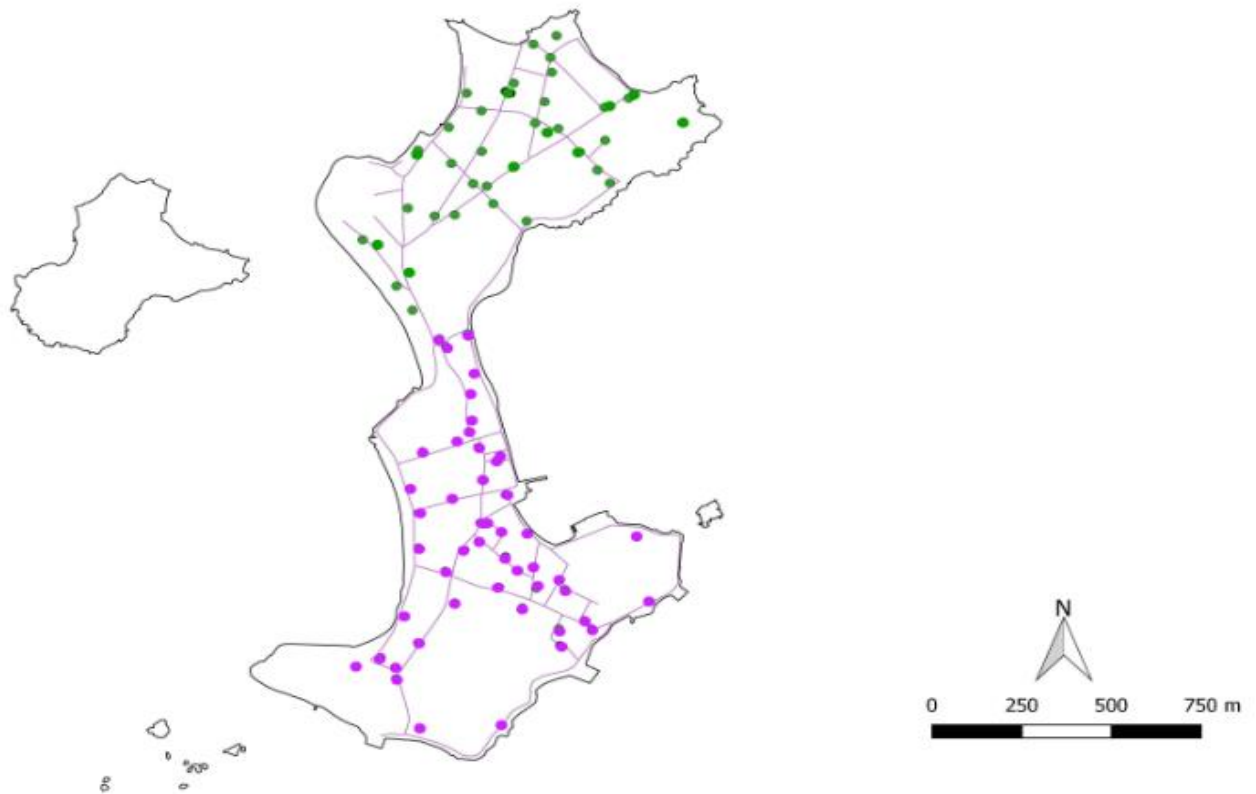


Figura 3.1. Mapa da ilha de Paquetá, Rio de Janeiro. Pontos rosa são os locais onde foram instaladas 50 armadilhas para coleta de ovos destinados a análise de parentesco e os pontos verdes somados aos pontos rosas são locais onde ovitrampas foram instaladas as 90 armadilhas para coleta dos ovos destinados a análise da variabilidade da população.

3.2 Local de estudo

Localizada no interior nordeste da baía de Guanabara, a ilha de Paquetá apresenta um formato longitudinal irregular com duas áreas mais amplas conectadas por uma massa de terra mais estreita (Figura 3.1). Possui 1,2 Km² de área e 8 Km de perímetro, com aproximadamente 58 % de área urbana (Prefeitura do Rio de Janeiro 2016). Historicamente é dividida em duas áreas: a região sul, chamada de Ponte, voltada para o Rio de Janeiro e a região norte, chamada de Campo, voltada para o fundo da Baía de Guanabara e Serra dos Órgãos. A ilha pertence à cidade do Rio de Janeiro desde 1975, vinculada à área de

planejamento 1 (AP-1) da prefeitura, em conjunto com os bairros do centro histórico da cidade (Prefeitura do Rio de Janeiro 2016). O bairro possui um índice de desenvolvimento humano de 0,822, considerado alto dentro da AP-1, com 3.361 habitantes e 2.223 domicílios, onde 90 % destes são contemplados com abastecimento de água e 99 % com esgoto sanitário (IBGE 2010).

Em 1999, pelo seu valor histórico, paisagístico e cultural, Paquetá foi transformada em Área de Proteção do Ambiente Cultural (Rio de Janeiro 1999). O bairro é livre de asfalto, e não são permitidos carros e motos particulares. Por isso, e outras razões, tornou-se um grande atrativo com fluxo turístico intenso. Além disso, muitos moradores trabalham ou estudam na região metropolitana o que torna grande a movimentação de pessoas na ilha.

Segundo o INMET (Instituto Nacional de Meteorologia, 2015), o ano de 2015 teve temperatura média alta e os níveis de precipitação acumulada mensal abaixo do esperado na maioria dos meses, inclusive no trimestre em que níveis menores de precipitação já são esperados (junho, julho e agosto). Durante o ano os valores de precipitação em Paquetá variaram de 3,2 mm a 55 mm nos meses secos e 44 mm a 183 mm nos meses chuvosos (Figura 3.2), sendo a média anual para os períodos de 32,93 mm e 100,26 mm respectivamente.

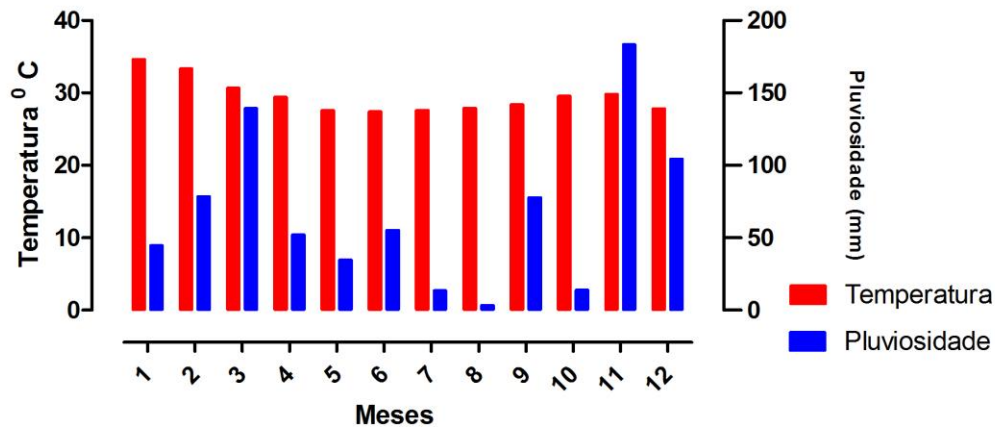


Figura 3.2. Temperatura e precipitação mensal acumulada na ilha de Paquetá, de janeiro a dezembro de 2015. Fonte: INMET.

3.3 Extração de DNA

Para análise dos genótipos, as amostras foram escolhidas de maneira diferente dependendo do objetivo a ser alcançado. Para a análise da estruturação genética geral, uma larva de cada armadilha coletada foi escolhida aleatoriamente. Para as análises de parentesco, todas as larvas dentro de uma mesma armadilha tiveram seus DNAs genômicos extraídos. Nos dois casos, as extrações foram realizadas para cada larva separadamente, segundo o modelo estabelecido no Laboratório de Biologia Computacional e Sistemas (LBCS/IOC), como descrito a seguir. Larvas no 4^o estágio foram maceradas separadamente em microtubos de 1,5 ml, contendo 150 µl de tampão CTAB 2 % (100 mM Tris, 20 mM EDTA, 1,42 M NaCl e 2 % CTAB) e 2 µl de proteinase K (concentração inicial de 10 µg/µl). Após a maceração, os tubos foram incubados em banho maria a 56 °C por 1 hora. Em seguida foi adicionad 150 µl de clorofórmio: álcool isoamílico (24:1 v/v) e os tubos foram agitados manualmente por 5 minutos seguido de 5 minutos de centrifugação a 13.000 rpm. A fase aquosa foi retirada para um novo tubo de 0,5 ml e acrescentou-se 5 µl de cloreto

de sódio (5 M) e 200 µl de etanol absoluto gelado. As amostras foram então colocadas no freezer durante a noite para a precipitação do DNA genômico. Após o período de precipitação, as amostras foram centrifugadas por 20 minutos a 14.000 rpm, o sobrenadante retirado e o pellet lavado com 400 µl de etanol diluído à 70 % gelado e os tubos centrifugados por mais 10 minutos a 14.000 rpm. O etanol foi então retirado dos tubos e o pellet deixado para secar por 20 minutos à temperatura ambiente. O pellet de cada amostra foi ressuspensão com 60 µl de água destilada e congelado até seu uso.

3.4 Amplificação do DNA e análise de dados genotipados

O DNA das amostras analisadas foi amplificado por meio da reação em cadeia da polimerase (PCR), com iniciadores marcados segundo o protocolo econômico de Schuelke (2000). Esta metodologia utiliza uma sequência M13 adicional nos iniciadores *forward*. Com isto, e a utilização de iniciadores M13 marcados com diferentes fluorescências (HEX, 6-FAM), conseguimos amplificar as regiões de microssatélites e marcar os fragmentos finais com diferentes cores sem precisarmos marcar diretamente os iniciadores de cada *locus*. Este procedimento reduz consideravelmente os custos envolvidos na confecção dos iniciadores. As diferentes cores de fluorescência permitem a genotipagem através de PCRs multiplex e posterior análise em conjunto em eletroforese de capilar. Os doze *loci* escolhidos para a análise estão entre os disponíveis na literatura (Slotman *et al.* 2007, Brown *et al.* 2011) e foram amplificados em três conjuntos (reações multiplex, Tabela 3.1). A reação foi preparada para um volume final de 10 µl contendo: 5 µl (concentração final 1x) de Type-It Multiplex PCR Master Mix (Qiagen), 0,025 µl (ou 0,025 µM) de iniciador *forward*, 0,25 µl (ou 0,25 µM) de iniciador *reverse*, 0,05 µl (ou 0,5 µM) do iniciador M13 com fluorescência, 3,4 µl de água mili-Q e 1 µl do DNA genômico diluído 1:10 (com água

mili-Q). Cada um dos conjuntos multiplex amplificados, além do DNA das larvas extraído possuía controles negativo e positivos, amostras utilizadas em outro estudo com os mesmos marcadores, para verificação posterior da qualidade e comparação dos genótipos analisados. A reação foi feita em um termociclador (Applied Biosystems - Veriti) nas seguintes condições: desnaturação inicial a 94 °C por 8 minutos, seguida por 35 ciclos de 30 segundos a 94 °C para desnaturação, 30 segundos a 54 °C para anelamento dos iniciadores e 30 segundos a 72 °C para extensão, seguidos de um ciclo de 5 minutos a 72 °C para extensão final. Os produtos amplificados foram genotipados em um sequenciador automático ABI3500 (Applied Biosystems) no Laboratório Genoma da Universidade do Estado do Rio de Janeiro (UERJ).

Tabela 3.1. Marcadores microssatélites utilizados na população estudada. Nome do marcador, sequência do iniciador, grupo do PCR multiplex, cor da fluorescência usada para detecção na genotipagem e fonte bibliográfica: 1- Brown *et al.* (2011); 2 - Slotman *et al.* (2007).

LOCUS	PRIMER (5'-3')	MULTIPLEX	FLUORESCÊNCIA	FONTE
A9	F-TCCCAGTCACGACGTGCAGCATGCACTTCACATTT R-CGAATGGCATCTGATTCAAG	1	6-Fam	1
AC4	F-TCCCAGTCACGACGTGCGAATCGGTTCCCATAGTA R-CTTTATCGATCGACGCCATT	1	6-Fam	2
B2	F-TCCCAGTCACGACGTGGAAACACTTGCAGGGACAT R-GCAGATGGTGGCAGTAGTGA	1	HEX	1
B3	F-TCCCAGTCACGACGTGCAAGTTGCAAAGTGCTCAA R-ACCCACCGTTTGCTTTGTAG	1	HEX	1
AC1	F-TCCCAGTCACGACGTTCCGGTGGGTTAAGGATAGA R-ACTTCACGCTCCAGCAATCT	2	6-Fam	2
AC5	FTCCCAGTCACGACGTTGGATTGTTCTTAACAAACACGAT R-CGATCTCACTACGGGTTTCG	2	6-Fam	2
AG1	F-TCCCAGTCACGACGTAATCCCCACACAAACACACC R-GGCCGTGGTGTACTCTCTC	2	HEX	2
AG5	F-TCCCAGTCACGACGTTGATCTTGAGAAGGCATCCA R-CGTTATCCTTTCATCACTTGTTTG	2	HEX	2
A1	F-TCCCAGTCACGACGTGACGTAACCGAGTGGGAGA R-GCATTTAACCGCGCTAGAAC	3	6-Fam	1
AC2	F-TCCCAGTCACGACGTAATAACAACGCGATCGACTCC R-AACGATTAGCTGCTCCGAAA	3	6-Fam	2
CT2	F-TCCCAGTCACGACGTGCGCAGTAGGCGATATTCGTT R-ACCACCACCAACACCATTCT	3	HEX	2
AG2	F-TCCCAGTCACGACGTTCCCCTTTCAAACCTAATGG R-TTTGCCCTCGTATGCTCTCT	3	HEX	2

O programa GeneMapper (Applied Biosystems) foi usado para a identificação dos alelos. Para conferir estatisticamente a confiabilidade dos marcadores utilizados, os *loci* foram testados para o desvio do equilíbrio de Hardy-Weinberg (HW) usando o programa Genepop v1.2 (Raymond & Rousset 1995, Rousset 2008). A independência dos *loci* foi avaliada através do teste de desequilíbrio de ligação (LD) usando o mesmo programa. Para ambos os testes foram usados os parâmetros de 10.000 demorizações, 1.000 *batches* e 10.000 iterações. Foi usada a correção de Bonferroni para testes múltiplos (Rice 1989). O programa GenAlEx 6.5 (Peakall & Smouse 2012) foi usado para os cálculos da frequência alélica, número efetivo de alelos e diversidade genética (heterozigosidades). O número efetivo de alelos nos diz a quantidade de alelos com a mesma frequência que seriam necessários para que a mesma medida de heterozigosidade fosse encontrada em um *locus*. Alelos com frequência baixa contribuem muito pouco para esta estimativa e com isto podemos ver se existem muitos alelos em baixa frequência (se houver uma diferença grande entre essa estimativa e o número de alelos encontrados). A presença de alelos nulos foi verificada com o programa ML-Relate (Kalinowski *et al.* 2006). O programa HP-Rare v1.0 (Kalinowski 2004, Kalinowski 2005) foi usado para o cálculo da riqueza alélica. A riqueza alélica calculada por esse programa leva em consideração o tamanho da amostra e, portanto, é uma medida menos enviesada da diversidade alélica nos diferentes *loci*.

3.5 Análises de parentesco

Para avaliar o comportamento de oviposição das fêmeas de *Ae. aegypti*, é necessário identificar as larvas que vieram de uma mesma fêmea, ou seja, a categoria de relacionamento de irmãos. Como esta espécie possui comportamento poligâmico, para se ter certeza de que estamos vendo irmãos que vieram da mesma fêmea é necessário identificarmos os irmãos completos, que possuem mesmo pai e mãe. Como

todos os programas de análise de parentesco possuem um certo grau de incerteza em suas inferências (Jones *et al.* 2010) é importante a utilização de mais de um programa. Por isso foram escolhidos dois programas, ML-Relate (Kalinowski *et al.* 2006) e Colony v. 2.0.5.9 (Jones & Wang 2010). Ambos realizam a reconstrução de parentesco somente com análise da prole, ou seja, sem a necessidade do conhecimento prévio dos genótipos da geração parental e levam em conta erros de genotipagem e mutacionais. Por isso são muito usados em análises de inferência de parentesco em diversos organismos como insetos, peixes, mamíferos entre outros (Hauser *et al.* 2011, Aglieri *et al.* 2014, Larson *et al.* 2015, Ottmann *et al.* 2016).

O programa ML-Relate calcula, por verossimilhança, estimativas de parentesco entre duas amostras a partir do cálculo do coeficiente de parentesco r , que é igual a $\frac{1}{2} k_1 + k_2$. O coeficiente k diz a probabilidade de dois alelos serem idênticos por descendência (são iguais não devido à mutação mas sim pelo fato de terem sido herdados de ancestrais comuns). Então k_0 , k_1 e k_2 são a probabilidade de dois indivíduos terem zero, um ou dois alelos que sejam idênticos por descendência (IBD) em um determinado *locus*. Se dois indivíduos são irmãos completos (possuem o mesmo pai e mãe), k_0 será igual a 0,25, k_1 será 0,5 e k_2 será 0,25 (neste caso, r será igual a 0,5). Essas probabilidades são diferentes para as diferentes categorias de relacionamento como podemos ver na tabela 3.2. O programa calcula qual a categoria de relacionamento entre dois indivíduos tem maior verossimilhança multiplicando essas probabilidades para cada *locus* avaliado. Com isso ele classifica os pares de indivíduos em: não relacionados (U), irmãos completos (FS), meio irmãos (HS), pais e filhos (PO). A significância de cada categoria testada entre os pares de indivíduo é obtida a partir de simulações e apenas categorias que sejam suportadas pelos dados são listadas.

Tabela 3.2. Lista de coeficientes k e r para diversas categorias de relacionamento entre indivíduos, modificado de Blouin (2003).

Categoria de relacionamento	k_0	k_1	k_2	r
Pais e filhos	0	1	0	0,50
Irmãos completos	0,25	0,50	0,25	0,50
Meio irmãos/ avós / tios / sobrinhos / netos	0,50	0,50	0	0,25
Primos de 1 ^o grau	0,75	0,25	0	0,125
Não relacionados	1	0	0	0

O programa Colony também usa verossimilhança para estimar o relacionamento de parentesco entre os indivíduos de uma população. Este é um programa mais rebuscado que foi criado com o intuito de avaliar estatisticamente as relações de parentesco entre indivíduos levando em conta toda a árvore genealógica dos dados e não apenas comparações par a par. Com este tipo de estimativa o programa é capaz de levar em conta se a espécie é poligâmica, fato que já foi observado em *Ae. aegypti*. Como o programa leva em conta toda a árvore genealógica dos dados, é mais fácil detectar indivíduos com graus diferentes de parentesco como meio irmãos e primos. O programa considera inicialmente todos os indivíduos como não relacionados e calcula a verossimilhança deste padrão genealógico. Conforme os relacionamentos são modificados novas verossimilhanças são calculadas, o novo padrão de relacionamento pode ser rejeitado ou aceito se a verossimilhança encontrada for maior que a do padrão anterior. Esse processo continua até que todos os parentescos tenham sido solucionados. Para que não fique preso em um pico de verossimilhança local (e não chegue ao pico global) o programa faz uso de um algoritmo que prevê a possibilidade

de não aceitação de verossimilhanças melhores, algumas vezes (“simulated annealing”, Kirkpatrick *et al.* 1983).

3.6 Criação de dados através de simulação computacional

Dados com padrões similares aos dados reais foram criados no programa Colony v. 2.0.5.9. Para isso, os resultados obtidos na estruturação geral da população como o número de *loci*, as frequências alélicas e número de alelos por *loci* foram inseridos como parâmetros iniciais. Outros parâmetros relacionados às características biológicas da espécie em estudo também são usados. Neste caso, para a espécie *Ae. aegypti*, as características usadas foram: dioica (sexos separados), diploide e com sistema de acasalamento poligâmico para machos e fêmeas, sem existência de clones. A metodologia escolhida foi a de máxima verossimilhança e as corridas foram feitas com tamanho médio, alta precisão e foi considerada uma taxa de erro de genotipagem de 0,001 (Hoffman & Amos 2005). O programa utiliza esses parâmetros e informações sobre a estrutura familiar para produzir os genótipos da prole final. A estrutura familiar é inserida através de matrizes, dando ao programa informações sobre o total de indivíduos a serem analisados e como estes estão relacionados (Figura 3.3). Foram criados sete cenários (matrizes) com diferentes proporções de meio irmãos (0%, 5%, 10%, 20%, 30%, 40% e 50%). Cada matriz com um total final de 250 indivíduos, distribuídos por 12 representantes paternos e maternos (matriz 12 x 12).

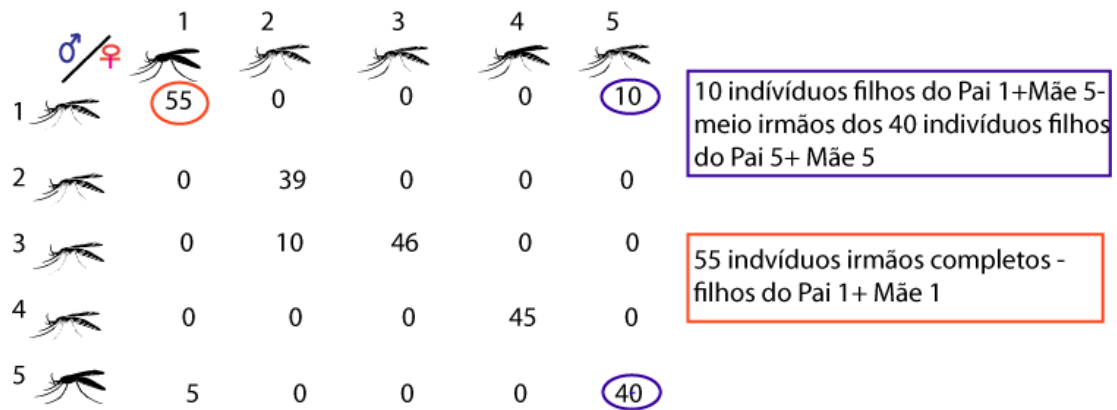


Figura 3.3. Exemplo de matriz 5 x 5, com 250 indivíduos na prole e 10 % de meio irmãos. Diferentes configurações de relacionamentos possíveis para os indivíduos analisados. Exemplo de meio irmãos em roxo, dois grupos de larvas da prole onde os indivíduos são considerados meio irmãos por compartilharem a mesma mãe (fêmea 5). Exemplos de irmãos completos em vermelho, todas as 55 larvas são filhas do mesmo macho (1) e mesma fêmea (1).

3.7 Análise comparativa dos programas de identificação de parentesco

Para testar qual dos programas usados anteriormente melhor se adequa aos dados estudados, grupo de dados simulado com diferentes cenários feitos no programa Colony foram analisados nos programas Colony e ML-Relate. Os resultados dos dados simulados foram analisados com os mesmos critérios usados para os dados reais. A partir das matrizes de estruturas familiares inseridas em cada um dos cenários simulados é possível identificar o real grau de parentesco de cada combinação par a par dos 250 indivíduos simulados. Com isto é possível calcular os erros e a acurácia de cada programa. Além disso, avaliamos o quanto os programas Colony e ML-Relate classificam os pares de indivíduos corretamente nas diferentes categorias e ainda em quais cenários os dois programas realizam a categorização dos indivíduos de maneira mais incorreta.

A acurácia e as taxas de erro foram calculadas buscando avaliar cada programa em relação à capacidade de identificar irmãos completos corretamente, já que esta é a categoria que melhor nos indicará o comportamento de oviposição das fêmeas, pois indivíduos que são meio irmãos podem ser filhos de fêmeas diferentes com o mesmo macho. Acurácia, ou proporção de erros, diz quanto próximo de um determinado resultado verdadeiro o programa realiza a categorização. Ou seja, ela é calculada a partir do total de erros classificatórios sobre o total de comparações para a par. O total de erros é calculado a partir do número de comparações de indivíduos que são irmãos completos mas não são corretamente categorizados (erro do tipo II, falso negativos) acrescido o número de indivíduos que não são irmãos completos mas são categorizados como tal (erro do tipo I, falsos positivos). Calculamos também a taxa de classificação correta de cada programa, ou seja, o número de indivíduos verdadeiramente irmãos completos sobre o número total de irmãos completos determinado por cada programa e a taxa de irmãos completos classificados erroneamente, ou o total de indivíduos classificados de maneira errada, como irmãos completos, sobre o total de irmãos encontrados (Wang 2004, Wang & Santure 2009).

4 RESULTADOS

4.1 Análise genética geral da população de Paquetá

Um total de 75 indivíduos escolhidos aleatoriamente foram genotipados para 12 *loci* de microssatélites em uma análise da variabilidade genética geral da população de Paquetá. Isto porque os dados sobre variabilidade dos marcadores usados poderiam ficar com viés se apenas amostras genealógicamente aparentadas fossem usadas. Este é o caso das amostras coletadas para a determinação do objetivo geral do trabalho. As frequências alélicas para cada *locus* estão na tabela 4.1. Nesta tabela os números amostrais para cada *locus* (N na tabela 4.1) podem variar devido a problemas de genotipagem de determinados indivíduos em alguns *loci*. Isto é comum em genotipagens e, se a proporção de genótipos ausentes não for grande, como foi o caso, não atrapalha as análises. Em média a população teve uma riqueza alélica de 4,2, com o *locus* AG2 possuindo o maior número de alelos (13). A heterozigosidade total observada foi de 0,42 e a esperada de 0,48. Após a correção de Bonferroni, foram observados desvios para o equilíbrio de Hardy-Weinberg para os *loci* A9, B3, AC5, devido à um déficit de heterozigotos nos três casos. Os números de coeficiente de endogamia (F_{IS}) podem ser observados na tabela 4.2. O programa ML-Relate constatou a presença de alelos nulos para os *loci* A9, B3, AC1 e AC5. A tabela 4.2 resume as principais medidas de variabilidade encontradas para cada *locus*. Esses dados foram usados para a criação dos dados simulados no programa Colony.

Tabela 4.1. Frequências alélicas (nomes dos alelos são seus tamanhos em pares de base no gel de sequenciamento) nos 12 *loci* de microssatélites analisados em *Ae. aegypti*. N - número de mosquitos analisados em cada *locus* (pode ser diferente em cada *locus* devido a erros na genotipagem); Freq – frequência alélica.

<i>Locus</i>	Alelos	Freq	<i>Locus</i>	Alelos	Freq
A9	N	75	AG1	N	39
	182	0,053		117	0,500
	188	0,607		119	0,154
	191	0,340		121	0,346
AC4	N	75	AG5	N	35
	128	0,827		166	0,057
	130	0,173		168	0,343
B2	N	75	AG2	170	0,157
	95	0,080		172	0,043
	107	0,033		174	0,014
	113	0,887		176	0,271
B3	N	75	AC2	178	0,114
	166	0,807		N	66
	172	0,100		184	0,356
AC1	N	35	CT2	186	0,091
	195	0,257		N	67
	197	0,229		184	0,731
AC5	N	35	AG2	188	0,269
	207	0,014		N	67
	209	0,500		115	0,246
	155	0,014		117	0,022
	156	0,086		119	0,455
	157	0,014		149	0,007
	158	0,100		153	0,022
159	0,643	155	0,022		
A1	N	69	157	0,007	
	160	0,143	159	0,082	
	156	0,007	161	0,007	
	158	0,014	164	0,007	
	162	0,116	167	0,060	
	168	0,855	169	0,037	
	174	0,007	171	0,022	

Tabela 4.2. Medidas de variabilidade encontradas para cada *locus* avaliado e valores do coeficiente de endogamia, F_{IS} , na população de *Ae. aegypti* de Paquetá. F_{IS} em negrito foram significativos e demonstram desvio para o equilíbrio de Hardy-Weinberg. (Al) número de alelos; (Na) riqueza alélica; (Ne) número efetivo de alelos; (Ho) heterozigosidade observada; (He) heterozigosidade esperada.

<i>Locus</i>	A9	AC4	B2	B3	AC1	AC5	AG1	AG5	A1	AC2	AG2	CT2
Al	3	2	3	3	4	6	3	7	5	3	13	2
Na	2,99	2	2,95	2,99	4	6	3	7	3,77	2,99	10,63	2
Ne	2,06	1,40	1,26	1,50	2,71	2,21	2,54	4,27	1,34	2,27	3,55	1,65
Ho	0,13	0,35	0,19	0,20	0,60	0,55	0,61	0,73	0,25	0,59	0,73	0,39
He	0,52	0,29	0,21	0,33	0,63	0,56	0,61	0,78	0,26	0,56	0,72	0,30
F_{IS}	0,74	-0,20	0,10	0,40	0,06	0,49	-0,002	0,04	0,10	-0,05	-0,01	0,02

4.2 Análise comparativa dos programas de identificação de parentesco

Foram criadas sete matrizes de 250 indivíduos para a geração de dados simulados onde conhecemos o relacionamento de parentesco entre os indivíduos para a avaliação dos programas ML-Relate e Colony. As matrizes de estrutura familiar utilizadas para a geração dos dados no programa Colony podem ser encontradas no Anexo I, assim como os parâmetros escolhidos. Como explicado anteriormente, dados com diferentes proporções de meio irmãos foram criados e analisados nos programas ML-Relate e Colony. Scripts escritos em Perl foram criados para que os resultados dos programas fossem analisados de maneira mais eficiente. Isto porque o número de comparações par a par que devem ser avaliadas para calcularmos as taxas de erro e acurácia de cada programa é grande (31.125). Os scripts completos podem ser encontrados no Anexo II.

O programa ML-Relate calcula o coeficiente de relacionamento r para categorizar os pares de indivíduos analisados. A figura 4.1 mostra a distribuição de r_{xy} nas comparações par a par nos dados simulados para a categoria de relacionamento irmãos completos. As médias das distribuições de r encontradas nas diferentes matrizes ficaram entre 0,48 e 0,57, bem próximas da esperada para esta categoria, que seria de 0,5 (Tabela 3.2).

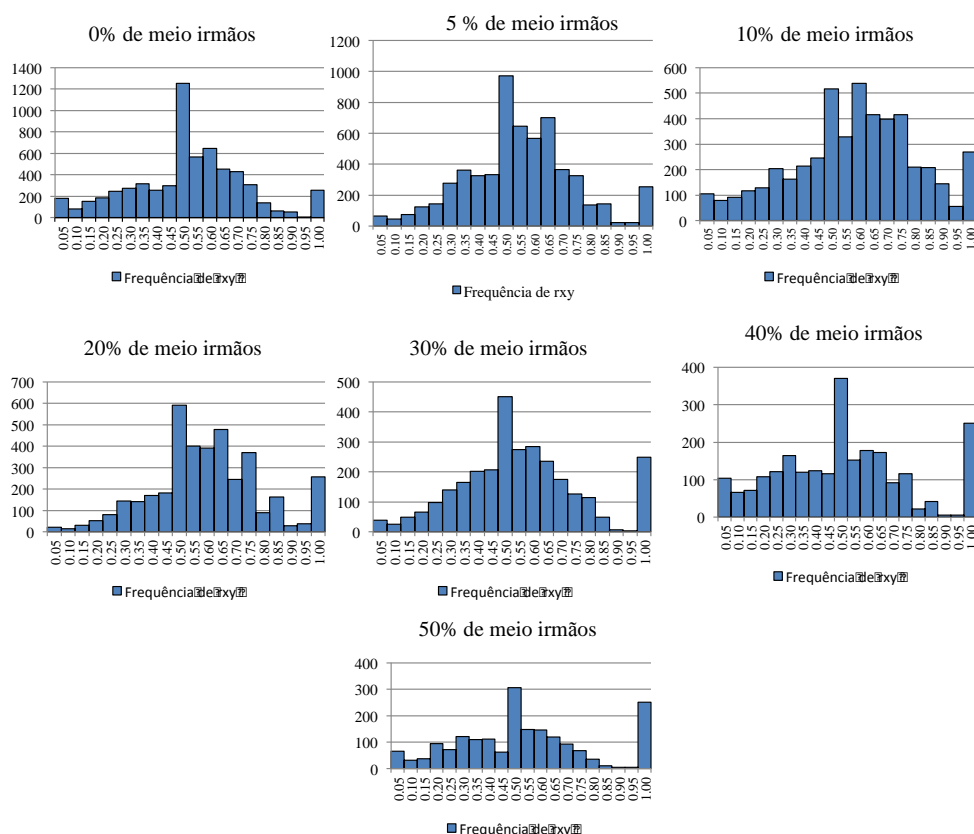


Figura 4.1. Distribuição de r_{xy} nas comparações par a par para a categoria de relacionamento irmãos completos em todos cenários simulados (diferentes proporções de meio irmãos nas matrizes).

Após a análise de cada cenário em ambos programas, observou-se que, em muitos casos, as díades categorizadas pelos programas não eram condizentes com o parentesco real dos indivíduos. Para a avaliação do comportamento de oviposição das fêmeas, poder identificar irmãos completos dentro e entre as armadilhas é de extrema importância. Por isso focamos os cálculos de erro para esta categoria, irmãos completos. O programa ML-Relate apresentou uma proporção do erro tipo I (classificar outras categorias de relacionamento como sendo irmãos completos), entre 0,017 e 0,037, uma quantidade de erro baixa, mas que se elevou de maneira moderada, a medida que a proporção de meio irmãos nas amostras aumentava (Figura 4.2). A proporção do erro tipo II (classificar irmãos completos como outra categoria) também foi baixa, entre

0,016 e 0,057, e diminuiu com o aumento da proporção de meio irmãos nas amostras. O erro do tipo II foi maior que o erro do tipo I até o cenário com 30 % de meio irmãos, quando esta tendência se inverteu (Figura 4.2). O programa Colony apresentou uma proporção de erro tipo I mais baixa que o ML- Relate, entre 0 e 0,00035. Neste programa a proporção de erro tipo II foi mais alta (entre 0,014 e 0,034) que o erro de tipo I, mas em geral mais baixa que no ML-Relate.

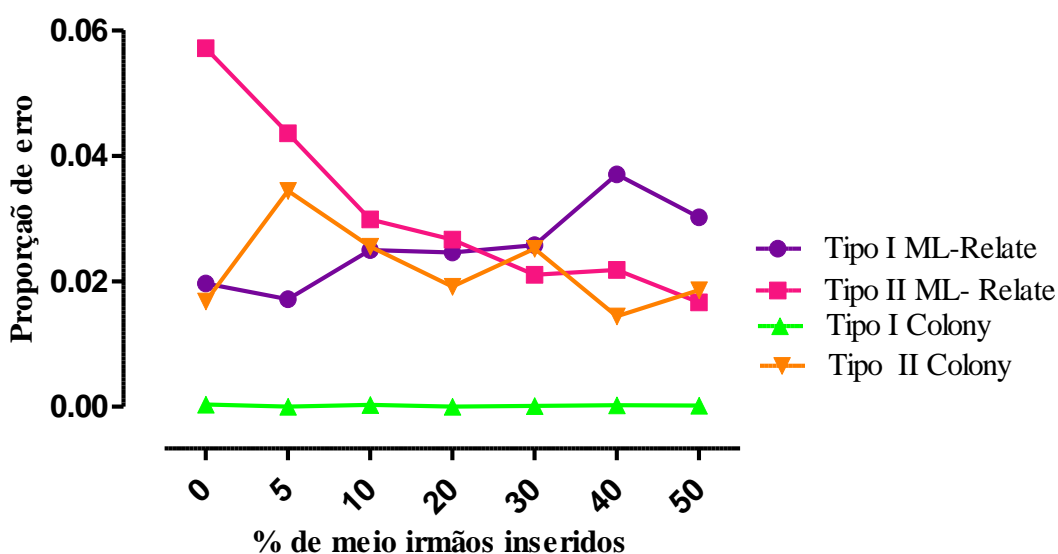


Figura 4.2. Erros tipo I e II para a categoria de irmãos completos encontrados em cada cenário (diferentes proporções de meio irmão) nos dois programas avaliados.

A taxa de classificação correta variou entre 24 e 73 % para o ML-Relate e 97 e 100 % para o Colony (Figura 4.3). Esta taxa mostra que quando o programa Colony classifica indivíduos como irmãos completos podemos confiar nele. No entanto, o programa, em muitos casos, deixa de classificar irmãos completos verdadeiros como tal, como mostra a figura 4.3. Mesmo alta, a proporção de irmãos completos verdadeiros classificados erradamente pelo programa Colony foi apenas mais alta do que a do ML-Relate nos cenários com 30 e 50 % de meio irmãos. Um resultado

interessante é que, para o programa Colony, este tipo de erro é bem menor para o cenário onde não existem meio irmãos na população. Ou seja, quando a espécie é monogâmica, o que não ocorre para o ML-Relate. Os dois programas possuem uma menor taxa deste erro nos cenários com proporções intermediárias de meio irmãos, onde o número de comparações par a par entre meio irmãos ainda é menor do que a de irmãos completos. No cenário com 30 % de meio irmãos, o número de comparações par a par entre meio irmãos já é duas vezes maior que a entre irmãos completos, com 40 % é três vezes e com 50 % quase cinco vezes maior. A grande quantidade de meio irmãos parece atrapalhar a correta classificação dos irmãos completos.

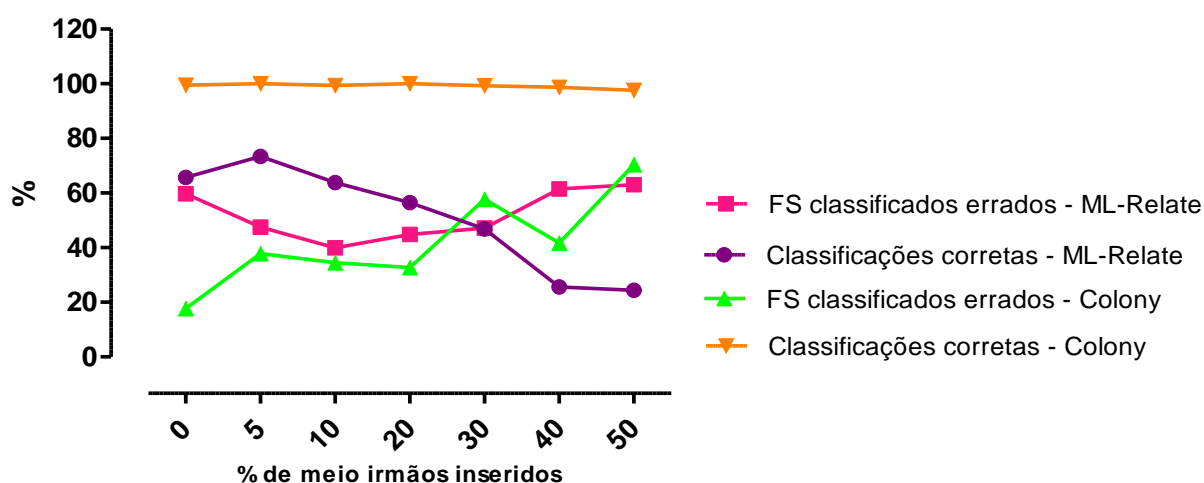


Figura 4.3. Proporção de irmãos completos (FS) verdadeiros classificados erroneamente em outras categorias e taxa de classificação correta para a categoria de irmãos completos em cada cenário (diferentes proporções de meio irmão) nos dois programas avaliados.

A acurácia do programa ML – Relate variou de 92 a 95 % e do programa Colony de 96 a 98 %. Um ponto interessante é a diferença em performance entre os dois programas para os dados simulados sem poligamia (Figura 4.4). O programa Colony parece ter uma melhor performance na ausência de poligamia, tendo melhor desempenho que o programa ML-Relate em todos os casos. O programa ML-Relate

parece ter uma melhor performance quando existem meio irmãos nos dados, em pequenas proporções.

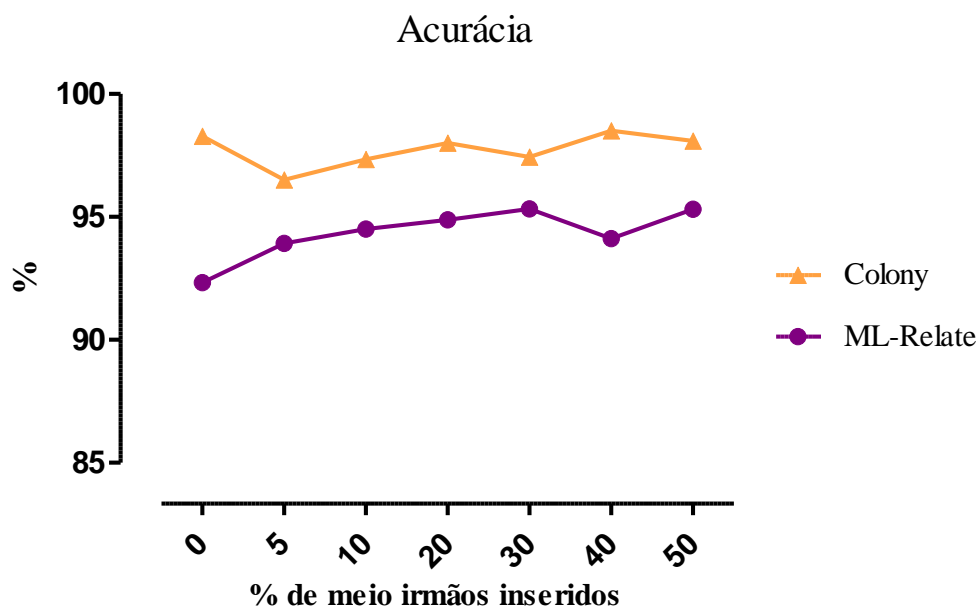


Figura 4.4. Acurácia de cada programa avaliado para os sete cenários simulados.

O programa ML-Relate calcula automaticamente uma categoria de relacionamento chamada de Pais e Filhos (PO) que é de difícil diferenciação da de irmãos completos por ter o mesmo número de r_{xy} , apenas diferenças na quantidade de alelos IBD esperada (k , Tabela 3.2). Realmente esse programa categoriza entre 22 e 37% de irmãos completos (FS) verdadeiros erradamente como sendo da categoria PO. Seria tentador, quando sabemos que a geração parental não está presente na amostra, passarmos esta categoria para FS diretamente. No entanto, estaríamos aumentando drasticamente o nosso erro do tipo I, pois de 21 a 30 % de meio irmãos (HS) também são erroneamente classificados como PO. Esse programa classifica de 10 a 21 % de FS como HS e de 2 a 14 % como não relacionados (U). O programa Colony não possui a categoria PO mas também classifica erradamente cerca de 2 a 19% de FS como HS e 17 a 51% como U.

4.3 Análise de parentesco dos dados reais

4.3.1 Estrutura genética dos indivíduos capturados nas coletas de verão e inverno

A coleta da estação chuvosa foi feita em março (verão) e a da estação seca em julho (inverno) de 2015. Um total de 1.125 e 753 ovos foram coletados em cada estação. Após a eclosão das ovos e identificação das larvas, obtivemos 426 e 471 larvas de *Ae. aegypti* para o verão e inverno respectivamente. A partir das 50 ovitrampas instaladas, foram escolhidas 12 ovitrampas (213 indivíduos) coletadas no verão e seis ovitrampas (108 indivíduos) no inverno (Tabela 4.3).

Tabela 4.3. Número das armadilhas avaliadas no verão e inverno e quantidade de larvas genotipadas (total de larvas encontradas em cada armadilha).

Verão		Inverno	
Armadilhas	Larvas	Armadilhas	Larvas
15	17	4	24
19	28	5	17
20	8	12	7
21	4	13	6
22	12	28	28
27	14	45	26
28	12		
30	5		
31	30		
32	14		
37	37		
50	32		
Total	213		108

Em média encontramos uma riqueza alélica de 4,5, muito parecida com a encontrada na análise geral da população. Em cinco *loci* (A9, AC1, AC2, AC5 e AC4) um número maior de alelos foi encontrado nessas coletas quando comparadas com as análises da população geral. Em todos os casos, os alelos estavam em baixa frequência e, apesar do grau de parentesco ser maior nestas amostras, é possível que esta diferença em número de alelos encontrados seja devido ao maior número de indivíduos avaliados neste caso (321 contra 75 da análise populacional geral). A heterozigosidade observada total foi de 0,44 e esperada de 0,50, parecidas com as encontradas na análise da população geral. O *locus* AG2 novamente foi onde encontramos o maior número de alelos (11). A tabela 4.4 mostra as frequências alélicas dos indivíduos coletados no verão e inverno separadamente.

Após a correção de Bonferroni, foram observados desvios para o equilíbrio de Hardy-Weinberg em nove *loci* nas amostras de verão e em dez dos 12 *loci* nas amostras de inverno. Os números de F_{IS} para cada estação podem ser observados na tabela 4.5. A grande quantidade de *loci* em desequilíbrio provavelmente se deve ao fato de estarmos analisando indivíduos com um alto grau de parentesco. A tabela 4.5 resume as principais medidas de variabilidade encontradas para cada *loci* e estação analisada. Como podemos ver pelas tabelas 4.4 e 4.5, não há muita diferença entre as frequências alélicas nem medidas de diversidade entre as estações de um mesmo ano.

Tabela 4.4. Frequências alélicas (nomes dos alelos são seus tamanhos em pares de base) nos 12 *loci* de microssatélites em *Ae. aegypti* coletados em diferentes estações do ano (inverno e verão) em Paquetá. N - número de mosquitos analisados em cada *locus*.

<i>Locus</i>	Alelos	Inverno	Verão	<i>Locus</i>	Alelos	Inverno	Verão
AC4	N	100	198	B2	N	101	204
	126	0,050	0,020		95	0,188	0,029
	128	0,695	0,907		107	0,000	0,167
	130	0,255	0,073		113	0,812	0,804
B3	N	100	203	AG1	N	101	208
	166	0,840	0,702		117	0,426	0,481
	172	0,135	0,182		119	0,193	0,130
	175	0,025	0,116		121	0,381	0,389
A1	N	107	211	AG5	N	102	206
	156	0,019	0,005		166	0,127	0,066
	158	0,005	0,009		168	0,368	0,291
	162	0,107	0,201		170	0,250	0,143
	166	0,005	0,005		172	0,029	0,063
	168	0,860	0,770		174	0,000	0,024
	170	0,005	0,007		176	0,118	0,323
	176	0,000	0,002		178	0,108	0,090
A9	N	97	186	AC5	N	97	204
	180	0,000	0,013		147	0,041	0,000
	182	0,015	0,121		155	0,046	0,000
	185	0,010	0,003		156	0,046	0,135
	188	0,794	0,731		157	0,010	0,020
	191	0,170	0,132		158	0,201	0,108
	194	0,010	0,000		159	0,619	0,613
AC1	N	103	207	AG2	N	107	205
	195	0,340	0,324		160	0,036	0,115
	197	0,199	0,135		161	0,000	0,010
	203	0,005	0,000		114	0,005	0,000
	207	0,078	0,017		115	0,439	0,393
	209	0,379	0,524		117	0,070	0,054
AC2	N	107	210	119	0,407	0,390	
	178	0,000	0,005	153	0,000	0,007	
	184	0,393	0,393	155	0,000	0,007	
	186	0,140	0,043	157	0,005	0,010	
	188	0,467	0,560	159	0,019	0,107	
CT2	N	106	205	161	0,014	0,012	
	184	0,717	0,739	167	0,014	0,020	
	188	0,283	0,261	169	0,028	0,000	

Tabela 4.5. Medidas de variabilidade encontradas para cada loci avaliado e valores do coeficiente de endogamia, F_{IS} , na população de *Ae. aegypti* de Paquetá coletadas no inverno (I) e verão (V). F_{IS} em negrito foram significativos e demonstram desvio para o equilíbrio de Hardy-Weinberg. (Al) número de alelos; (Na) riqueza alélica; (Ne) número efetivo de alelos; (Ho) heterozigosidade observada; (He) heterozigosidade esperada.

	<i>Locus</i>	A9	AC4	B2	B3	AC1	AC5	AG1	AG5	A1	AC2	AG2	CT2
Al	I	5	3	2	3	5	7	3	6	6	3	9	2
	V	5	3	3	3	4	6	3	7	7	4	9	2
Na	I	5,00	3,00	2,00	3,00	4,94	7,00	3,00	6,00	5,72	3,00	8,81	2,00
	V	4,50	2,99	3,00	3,00	4,00	5,92	3,00	7,00	5,63	3,71	8,60	2,00
Ne	I	1,52	1,82	1,44	1,38	3,28	2,32	2,75	4,16	1,33	2,55	2,74	1,68
	V	1,76	1,21	1,48	1,85	2,51	2,39	2,50	4,41	1,58	2,13	3,11	1,63
Ho	I	0,02	0,51	0,24	0,10	0,76	0,31	0,85	0,67	0,22	0,67	0,60	0,43
	V	0,10	0,15	0,39	0,19	0,68	0,39	0,60	0,73	0,38	0,59	0,69	0,35
He	I	0,34	0,45	0,31	0,28	0,69	0,57	0,64	0,76	0,25	0,61	0,64	0,41
	V	0,43	0,17	0,32	0,46	0,60	0,58	0,60	0,77	0,37	0,53	0,68	0,39
F_{IS}	I	0,94	-0,13	0,23	0,64	-0,08	0,46	-0,33	0,14	0,10	-0,10	0,06	-0,06
	V	0,76	0,15	-0,20	0,58	-0,13	0,33	0,01	0,06	-0,03	-0,10	-0,02	0,10

4.3.2 *Determinação do parentesco dos indivíduos capturados intra e inter armadilhas.*

Para avaliar o comportamento das fêmeas de *Ae. aegypti*, os genótipos dos indivíduos de cada estação foram analisados separadamente nos programas ML-Relate e Colony. Assim conseguimos identificar os grupos de irmãos completos (que possuem mesmo pai e mãe) existentes dentro e entre armadilhas. Para o programa Colony, além da informação dos genótipos, características da espécie estudada e o tipo de marcador usado também são inseridos. Os seguintes parâmetros foram usados: indivíduos diploides e dioicos, machos e fêmeas poligâmicos, marcadores codominantes e taxa de erro de genotipagem de 0,001. As análises foram realizadas por verossimilhança com comprimento médio de corrida e alta precisão (Jones & Wang 2010). No programa ML-Relate apenas os genótipos são inseridos.

A partir dos 321 indivíduos analisados, foram realizadas 22.578 análises par a par nas amostras de verão e 5.778 nas de inverno. A análise das duas coletas do ano mostrou que existe presença de irmãos completos e meio irmãos dentro de armadilhas nos dois períodos. Em muitos casos a categoria de relacionamento de irmão completo (FS) também foi encontrada em análises par a par entre armadilhas diferentes indicando um comportamento de oviposição em saltos para esta espécie em Paquetá.

Para o programa ML-Relate, no verão, não foi possível selecionar grupos de irmãos completos, já que em todos havia indivíduos repetidos. Isto acontece, pois este programa faz comparações par a par entre todos os indivíduos e calcula a partir do coeficiente de r e k a probabilidade desses indivíduos pertencerem a uma determinada categoria de relacionamento. Com isto, em muitos casos ocorrem inconsistências. Por exemplo, o programa infere que o indivíduo A é irmão completo do indivíduo B e este,

por sua vez é irmão completo do indivíduo C mas C e A não são relacionados. Este tipo de problema não ocorre no programa Colony que infere a genealogia completa que melhor explica os dados. No inverno, para o programa ML-Relate, puderam ser distinguidos 6 grupos de irmãos completos, todos espalhados em mais de uma armadilha. Desses grupos de irmãos completos a maioria das larvas estava bem distribuída pelas armadilhas e em apenas um grupo uma armadilha continha muito mais larvas (13 indivíduos) do que as outras indicando sítio “predileto” (Figura 4.5). A menor distância entre as armadilhas que possuem irmãos completos é de 107 m e a maior de 900 m.

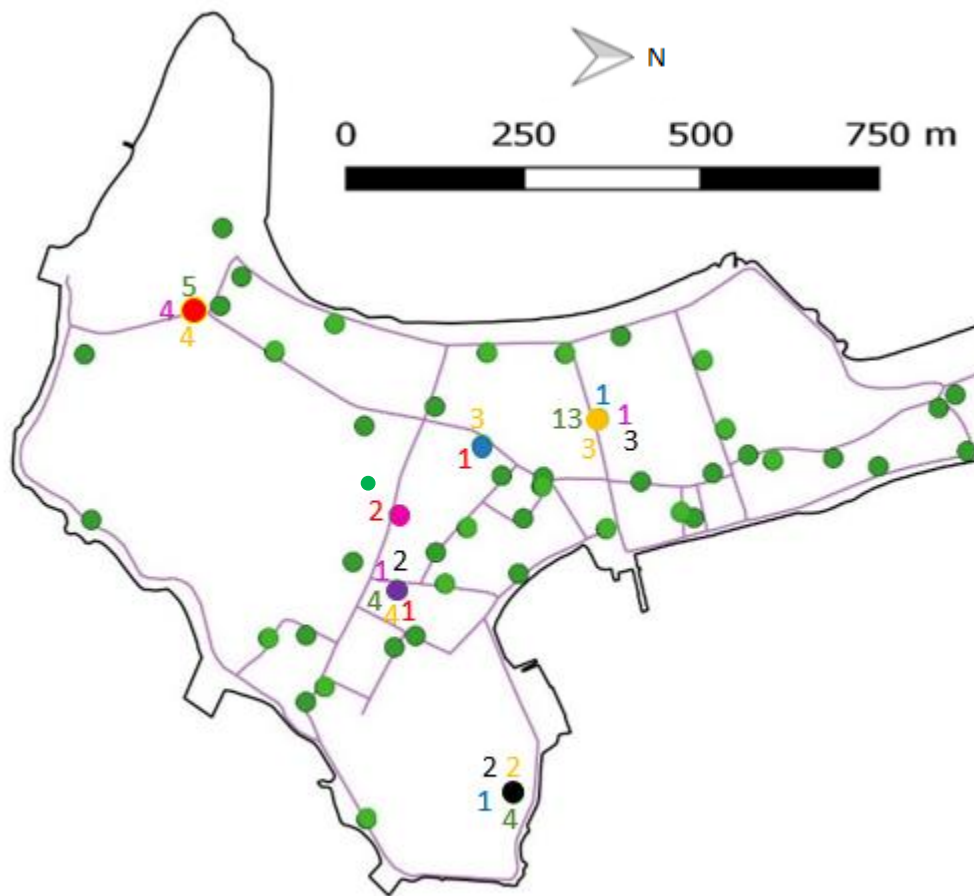


Figura 4.5. Grupos de irmãos completos distribuídos (indicados pelas diferentes cores) nas diferentes armadilhas coletadas no inverno em Paquetá e analisados no programa ML-Relate. Os números coloridos indicam quantos indivíduos de cada grupo estão na

armadilha e as cores de cada círculo indicam a armadilha analisada sendo: preto (armadilha 4); roxo (armadilha 5); azul (armadilha 12), rosa (armadilha 13), vermelho (armadilha 28) e laranja (armadilha 45).

Já para o programa Colony, no verão foram encontrados 15 grupos de irmãos completos e 4 desses estavam espalhados em mais de uma armadilha (Figura 4.6). Três desses grupos de irmãos completos possuem apenas uma ou duas larvas em cada armadilha e em apenas um a maioria das larvas (33 indivíduos) foi encontrada em uma das armadilhas (Figura 4.6). É possível que, para os três grupos onde apenas poucos indivíduos foram encontrados não tenhamos encontrado o sítio “predileto” por acaso. Mas é possível também que as fêmeas de *Ae. aegypti* em Paquetá não possuam sempre este comportamento. Dos 11 grupos de irmãos completos restantes, dois indivíduos (V1501 e V1503) foram encontrados na armadilha 15, 11 indivíduos (V1903, V1905, V1906, V19013, V19016, V19018, V19022, V19023, V19024, V1925 e V1928) na armadilha 19, quatro indivíduos (V2202, V2203, V22012 e V2206) na armadilha 22, seis indivíduos (V2701, V2705, V2706, V2709, V27010, V27013) na armadilha 27 e sete indivíduos (V3103, V31012, V31017, V31022, V31024, V31025 e V31026) na armadilha 31. No inverno também foram encontrados 15 grupos de irmãos completos mas todos dentro das diferentes armadilhas. Dezesete indivíduos (I401, I403, I404, I407, I408, I409, I4010, I4011, I4012, I4013, I4015, I4016, I4017, I4018, I4019, I4021, e I4024) foram encontrados na armadilha 4, quatro indivíduos (I1203, I1204, I1205 e I1206) na armadilha 12, dois indivíduos (I1302 e I1305) na armadilha 13, 16 indivíduos (I2801, I2803, I2806, I28010, I28011, I28012, I28014, I28015, I28016, I28017, I28018, I28023, I28026, I28027, I28028 e I28029) na armadilha 28 e 23 indivíduos (I4501, I4503, I4504, I4505, I4506, I4508, I4509, I45010, I45011, I45012, I45013, I45014, I45015, I45016, I45018, I45019, I45020, I45021, I45022, I45023, I45024,

I45025 e I45026) na armadilha 45. A menor distância entre armadilhas que compartilhassem irmãos completos no verão foi de 140 m e a maior de 700 m. Nas análises do programa Colony não foi observada a presença de irmãos completos entre diferentes armadilhas no inverno, o que pode indicar a baixa dispersão da fêmea neste período.

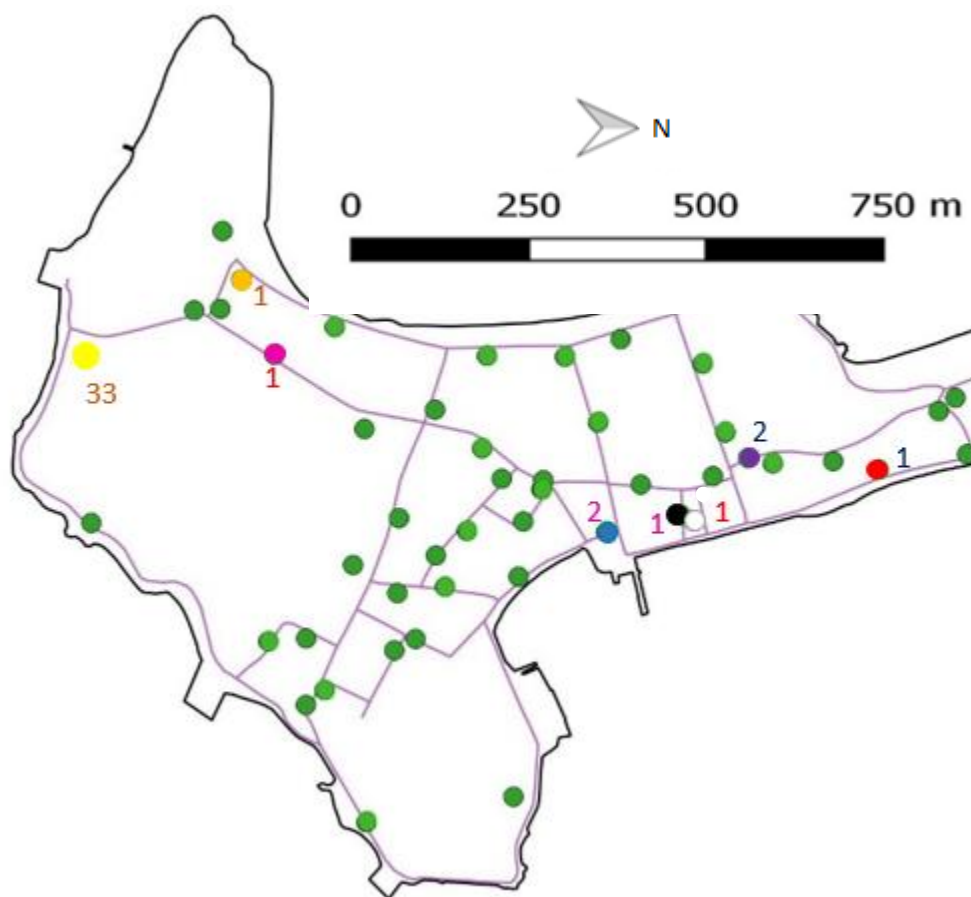


Figura 4.6. Grupos de irmãos completos distribuídos (indicados pelas diferentes cores) nas diferentes armadilhas coletadas no verão em Paquetá e analisados no programa Colony. Os números coloridos indicam quantos indivíduos de cada grupo estão na armadilha e as cores de cada círculo indicam a armadilha analisada sendo: azul (armadilha 19), preto (armadilha 20), branco (armadilha 21), vermelho (armadilha 22), rosa (armadilha 31), laranja (armadilha 32), amarelo (armadilha 37), roxo (armadilha 50).

5 DISCUSSÃO

5.1 Avaliação dos *loci* microssatélites usados

As taxas de classificação correta em uma análise de parentesco são muito influenciadas pelo número de *loci*, a diversidade genética encontrada e a confiabilidade dos marcadores usados (Wang 2004, Harrison *et al.* 2013b, Kopps *et al.* 2015). Marcadores confiáveis são aqueles onde erros de genotipagem são pequenos e os *loci* analisados encontram-se em equilíbrio. Erros de genotipagem em marcadores microssatélites são baixos (Hoffmam & Amos 2005) e os *loci* avaliados neste trabalho são bem conhecidos. Estes mesmos *loci* foram usados em diversos estudos de genética de populações de *Ae. aegypti* distribuídas mundialmente (Brown *et al.* 2011, Gloria-Soria *et al.* 2016), incluindo o Brasil (Monteiro *et al.* 2014). Nossos resultados de número de alelos e suas frequências são parecidos com resultados anteriores que utilizaram esses mesmos *loci* em populações desta espécie em outro bairro da cidade do Rio de Janeiro (Monteiro *et al.* 2014). Além disso, os programas de inferência de parentesco usados levam em conta este tipo de erro e a possibilidade de erros mutacionais (Kalinowski *et al.* 2006, Jones & Wang 2010).

Neste estudo, a avaliação dos marcadores na população de Paquetá indicou que apenas três (A9, B3, AC5) dos 12 *loci* analisados estava em desequilíbrio de Hardy-Weinberg e nenhum par em desequilíbrio de ligação. O desequilíbrio de Hardy-Weinberg pode acontecer devido a diversas causas como endocruzamento, seleção, mutação e presença de alelos nulos. Realmente, alelos nulos são comuns em marcadores microssatélites em insetos (Huang *et al.* 2008, Kobayashi *et al.* 2011, Babin *et al.* 2012, Monteiro *et al.* 2014, Gloria-Soria *et al.* 2016) e em nossa população os *loci* em desequilíbrio de Hardy-Weinberg encontrados possuíam um déficit de heterozigotos,

comum quando alelos nulos estão presentes. O programa ML-Relate indicou a presença de alelos nulos para estes mesmos três *loci* e mais um quarto (AC1). A presença de alelos nulos nas amostras pode ser uma importante fonte de erro em análises de parentesco, mas possui maior influência em análises de exclusão parental (Wagner *et al.* 2006) e nem tanto em análises de verossimilhança de outras categorias. Isto porque pais e filhos necessariamente possuem um alelo IBD em cada *locus* (se não houver mutações), mas irmãos completos tem, em média, metade de seus alelos IBD, mas em um *locus* específico podem ter 0, 1 ou 2 alelos IBD.

Tanto o programa ML-Relate, quanto o Colony podem levar em consideração a presença de alelos nulos (Kalinowski *et al.* 2006, Jones & Wang 2010). No entanto, em nossas análises, não testamos este tipo de erro já que, mesmo com este viés nos marcadores, as taxas de classificação correta para os dois programas foram altas nos dados simulados (Figura 4.3). Este resultado indica que a inferência da categoria de irmãos completos, importante para que saibamos o comportamento de oviposição das fêmeas, é confiável a partir do uso desses marcadores. A presença de alelos nulos não afetará a taxa de classificação correta mas poderá aumentar o erro do tipo II, aumentando o número de irmãos completos classificados errados e diminuindo a acurácia.

Neste trabalho foram usados 12 *loci* de microsatélites que continham de dois a 13 alelos por *locus*. Houve uma média de seis alelos por *locus*, um pouco mais baixa do que gostaríamos para análises de parentesco. No entanto, Ashley *et al.* (2009) mostraram com dados simulados que o número de alelos por *locus* aumenta a acurácia dos programas testados, mas que de seis a 10 seriam suficientes. As médias de diversidade genética encontradas, heterozigosidade observada e esperada (0,44 e 0,48),

foram intermediárias. Blouin *et al.* (1996), observaram que com apenas 10 *loci* a distinção entre irmãos completos e indivíduos não relacionados ainda é alta, com 90 % de acurácia. No entanto, os *loci* avaliados pelos autores possuíam um nível de heterozigosidade mais alto ($H_e=0,75$) do que o encontrado neste estudo. Kopps *et al.* (2015) investigaram o número mínimo de marcadores necessários para se diferenciar as categorias de relacionamento e viram que este variava de acordo com características da história natural da espécie (como tipo de reprodução e sobreposição de gerações). Estes autores consideraram que dez *loci* são suficientes para categorizar irmãos completos corretamente na ausência de poligamia, no entanto, a taxa de falsos positivos é extremamente alta.

O comportamento poliandrico da espécie *Ae. aegypti* (Helinski *et al.* 2012, Richardson *et al.* 2015), torna a análise de parentesco mais difícil, ao aumentar a quantidade de meio irmãos e dificultar as atribuições de irmãos completos (Wang & Santure 2009, Dexter & Brown 2013, Kopps *et al.* 2015). No entanto, nossas análises de simulações indicam que o programa Colony, apesar de ter uma melhor performance quando não existem meio irmãos nas amostras, também é confiável para a determinação de irmãos completos mesmo quando o número de meio irmãos é alto. O erro do tipo I (falsos positivos) encontrado para este programa é sempre baixo e suas taxas de classificação correta altas (Figuras 4.2 e 4.3).

5.2 Dados simulados e comparação entre os programas

O uso de simulação de dados com características similares aos reais foi importante para identificarmos os tipos de erros que poderiam ser encontrados e quais pontos poderíamos confiar nos programas usados. Para a criação dos dados simulados além das características do marcador, como número de *loci* e frequências dos alelos

também indicamos características da espécie como o comportamento poligâmico de ambos os sexos. Isto faz com que os dados simulados gerados contenham o mesmo grau de dificuldade encontrado na análise dos dados reais da população de Paquetá. Realmente diversos trabalhos recentes com inferência de parentesco em populações naturais vem ressaltando a necessidade de avaliação dos programas para o tipo de marcador usado e características da espécie sendo avaliada (Kopps *et al.* 2015, Taylor *et al.* 2015). Tem-se observado que as análises rebuscadas de inferência de parentesco são uma importante fonte de informação sobre a genealogia de espécies cujos relacionamentos não seriam passíveis de identificação sem dados genéticos. No entanto, é sempre necessário identificar os pontos fortes e fracos dos dados sendo avaliados e qual metodologia e programa de inferência de parentesco melhor se adequa a eles.

Neste trabalho estamos interessados em identificar o comportamento de oviposição da fêmea do mosquito *Ae. aegypti*. Como esta espécie possui comportamento poliandrico precisamos identificar os indivíduos que são irmãos completos dentro das amostras. Isto porque meio irmãos podem possuir a mesma mãe mas pais diferentes. Por isso, nossas avaliações de erro e acurácia dos programas avaliados foram baseadas nesta categoria de relacionamento. Kopps *et al.* (2015) mostraram que quando a espécie é poligâmica, a categoria de relacionamento de irmãos completos fica muito pequena e mesmo com um número grande de marcadores o aumento de falsos positivos torna a análise pouco confiável. Em nossas análises com diferentes proporções de meio irmãos nos dados observamos que apesar dos dois programas avaliados terem maior dificuldade de identificação de irmãos completos quando o número de díades destes é muito menor que de meio irmãos a taxa de falsos positivos aumenta apenas para o programa ML-Relate (Figura 4.2).

Para este programa, quando o número de díades de meio irmãos fica maior que a de irmãos completos (cenários com mais de 30 % de meio irmãos) o número de falsos positivos (erro tipo I) fica maior que o de falsos negativos (erro tipo II). O aumento de falsos positivos é um dos maiores problemas em nossa análise do comportamento de oviposição das fêmeas. Isto porque precisamos ter certeza da classificação dos irmãos completos sendo avaliados indicando que este não é um bom programa para a identificação de irmãos completos quando existe poligamia. O programa Colony não sofre este tipo de problema. As taxas de falsos positivos e de classificações corretas (Figuras 4.2 e 4.3) mostram que podemos confiar na identificação de irmãos completos feita por este programa. No entanto, com o aumento do número de díades de meio irmãos na amostra, o número de irmãos completos classificados erradamente pelo programa aumenta consideravelmente (Figura 4.3), indicando que muitos irmãos completos não são identificados pelo programa. Este resultado nos mostra que a identificação do número real de ovos depositados por uma mesma fêmea em determinado criadouro não é confiável. Para os dados reais será possível comprovar o comportamento de oviposição em saltos se observarmos irmãos completos entre armadilhas diferentes, e apesar dos falsos positivos encontrados, observamos que uma das armadilhas (armadilha 37) possuía 33 indivíduos o que traz a possibilidade da existência no local de oviposição em reservatório predileto. Porém, a falta de identificação de irmãos completos entre outras armadilhas não comprova que esta categoria de relacionamento não esteja presente nos dados. Além disso, a determinação da presença e utilização de um criadouro “predileto” pelas fêmeas também fica comprometida.

Apesar da classificação errada de irmãos completos aumentar quando o número de díades de meio irmãos é maior nas amostras nos dois programas, o destino desses

irmãos completos varia entre eles. Olhando apenas estes três últimos cenários dos dados simulados, o programa ML-Relate classificou 26, 28 e 37 % dos irmãos completos como pertencentes à categoria pais e filhos e 16, 20 e 21 % como meio irmãos. O programa Colony não possui a categoria pais e filhos mas classificou 14 e 19 % de irmãos completos como meio irmãos. Essas porcentagens são similares às encontradas por Blouin *et al.* (1996) onde 17 % de irmãos completos foram classificados como meio irmãos em dados simulados, sem poligamia. Como dito anteriormente, o aumento do erro entre parentes próximos pode ocorrer devido a uma proximidade na superfície do gráfico do coeficiente k (Figura 1.7) (Kalinowski *et al.* 2006). Por exemplo, para a diferenciação entre irmãos completos e meio irmãos, cerca de 40 *loci* são necessários para se chegar a um nível de 90 % de acurácia (Blouin *et al.* 1996). Além disso, as categorias irmãos completos e pais e filhos possuem o mesmo grau de r (0,5) mas diferenças no coeficiente k (Tabela 3.2) o que pode explicar a confusão entre estas duas categorias (Blouin 2003).

Realmente em Blouin *et al.* (1996) apenas 2 a 3 % de irmãos completos foram classificados como não relacionados, categorias com graus diferentes de r e bem distantes no gráfico de k . Mas neste trabalho os dados não foram criados simulando poligamia. Em nossos resultados a classificação errada de irmãos completos como indivíduos não relacionados é mais alta (8 % no ML-Relate e 17 % para o Colony) mesmo para o cenário sem poligamia. O que pode ser explicado pela menor variabilidade do marcador usado; $H_e=0,48$ em comparação com $H_e=0,75$ em Blouin *et al.* (1996). Para os três cenários com alto número de díades de meio irmãos estes números são ainda mais altos. O programa ML-Relate classifica 6, 14 e 10 % de irmãos completos como não relacionados e o programa Colony 22, 43 e 52 %. Estes resultados indicam que os irmãos completos classificados erradamente pelo programa Colony

acabam entrando nesta categoria, já que a categoria de pais e filhos não existe. Podemos observar neste programa como a presença de poligamia atrapalha a classificação correta de irmãos completos. Sem poligamia o programa classifica 82 % das díades de irmãos completos corretamente, mas com apenas 5 % de meio irmãos nos dados este número cai para 62 % (Figura 4.3).

Apesar dos erros de classificação encontrados, o programa Colony mostrou categorizar corretamente de 97 a 100 % dos irmãos completos e com um número extremamente baixo de falsos positivos (Figura 4.2 e Figura 4.3). Uma porcentagem similar foi observada por Ackerman *et al.* (2016), que mostrou uma categorização correta de 95 % para esta categoria. No programa ML-Relate as classificações corretas foram bem menores (24 a 73 %) e, como dito anteriormente, para os cenários com um número maior de díades de meio irmãos o número de falsos positivos é maior que o de falsos negativos, um problema para as nossas análises. Em todos os cenários o programa Colony apresentou uma proporção de falsos positivos mais baixa que o ML-Relate. Hauser *et al.* (2011), ao comparar esses dois programas, também mostrou que o Colony teve uma proporção de irmãos completos classificados corretamente mais alta que o ML-Relate (77 contra 39 %). O programa também se mostrou eficiente em outros cenários, com diferentes números de *loci*, prole por casal e tamanho da amostra (Sheikh *et al.* 2008).

Além disso, o programa Colony leva em consideração diversos parâmetros sobre a espécie estudada (como o tipo de acasalamento). O ML-Relate é um programa simples e rápido. É bastante utilizado (Mino *et al.* 2011, Aglieri *et al.* 2014), porém a inferência de parentesco é feita de forma independente para cada par de indivíduos podendo resultar em grande volume de combinações (Ackerman *et al.* 2016).

O programa Colony usa mais informações sobre os dados (Walling *et al.* 2010) considerando um conjunto maior de relacionamentos ao determinar a relação entre uma díade, pois verifica toda a genealogia dos dados (Waples & Waples 2011). O que pode resultar num maior poder de inferência em circunstâncias variadas, como as analisadas aqui (alto grau de poligamia), do que programas que realizam a verossimilhança par a par (Jones *et al.* 2010, Walling *et al.* 2010). O programa Colony é frequentemente utilizado em diversos trabalhos e se mostrou eficiente tanto para dados empíricos quanto simulados (Ashley *et al.* 2009, Mino *et al.* 2011, Harrison *et al.* 2013b).

5.3 Comportamento de oviposição do mosquito *Aedes aegypti*

Os dois programas, ML-Relate e Colony, mostraram resultados diferentes nas análises dos dados empíricos de irmãos completos entre diferentes armadilhas. Enquanto o primeiro programa identificou a existência de irmãos completos dentro e entre as armadilhas nas duas estações; o Colony só encontrou esta categoria entre diferentes armadilhas no verão. Mas, os resultados das simulações onde o programa Colony mostrou uma taxa de classificação correta maior e falsos positivos menor que o ML-Relate, preferimos levar em consideração este programa para a avaliação dos dados empíricos. Essa proporção de falsos positivos mais baixa encontrada no programa, nos diz que um número menor de indivíduos de outras categorias é classificado errado como irmãos completos, categoria importante da determinação do comportamento, o que somado as altas taxas de classificação correta da categoria nos dá uma maior confiança na classificação e determinação do comportamento.

Durante as duas estações do ano analisadas, verão e inverno, observou-se que os valores de temperatura foram muito altos e sem grandes variações. No entanto, apesar da pluviosidade acumulada ter sido abaixo do esperado nas duas estações, esta

foi um pouco maior no verão. Foi somente nessa estação de maior pluviosidade que o programa Colony identificou quatro grupos familiares entre as armadilhas. Este resultado comprova o comportamento de oviposição em saltos das fêmeas de *Ae. aegypti* neste período em Paquetá. A ausência de irmãos completos entre armadilhas no inverno indica que a fêmea tende a migrar e visitar outros criadouros com mais frequência na estação chuvosa. Estes resultados são contrários a outros trabalhos encontrados na literatura que mostraram uma maior dispersão da fêmea nos períodos de escassez de criadouros, ou seja, estação seca (Reiter *et al.* 1995, Edman *et al.* 1998). No entanto, como mostrado neste trabalho e por Kopps *et al.* (2015), o comportamento de poligamia e a grande presença de meio irmãos na amostra é um fator que pode atrapalhar a categorização de irmãos completos. Como vimos, apesar das baixas taxas de falsos positivos, quanto maior a proporção de díades de meio irmãos quando comparadas com a de irmãos completos nos dados maior é a taxa de classificação errada de irmãos completos (Figura 4.3). Esta taxa mostra que o número de indivíduos desta categoria é subestimado em situação de poligamia. Se este for o caso, a subestimativa de díades de irmãos completos no inverno pode ser a causa da falta desta categoria entre armadilhas. Isto porque muitos irmãos completos são categorizados incorretamente pelo programa, ou seja, são colocados em outras categorias de parentesco.

Além disso, esta deve ter sido a razão para que a existência de um sítio “predileto” com a oviposição em saltos não possa ser confirmada, como em outros estudos (Oliva *et al.* 2013, Abreu *et al.* 2015). Em apenas um caso um número maior de larvas do mesmo grupo familiar foi encontrado dentro de uma única armadilha (Figura 4.6). No entanto o fato dos grupos familiares encontrados possuírem poucos indivíduos dentro e entre as armadilhas já indica que as fêmeas espalharam seus ovos

por outras armadilhas já que a cada ciclo gonotrófico estas podem depositar cerca de 100 a 200 ovos (Chadee *et al.* 2002).

6 CONCLUSÕES

- O número de *loci* e suas características, com número de alelos, heterozigosidade e erros (de genotipagem e mutacionais) foram suficientes para os objetivos das análises de parentesco pretendido. No entanto, um marcador com heterozigosidade mais alta seria mais informativo.
- Os dados simulados foram importantes para que pudéssemos observar como a existência de poligamia e diversidade genética da espécie sendo avaliada afetam as estimativas de parentesco.
- Entre os programas de parentesco avaliados com os dados simulados, o programa Colony se mostrou mais robusto, com maior acurácia, menores taxas de erro (especialmente do tipo I) e altas taxas de classificação corretas.
- Apesar do programa Colony de ter uma performance satisfatória para a determinação dos grupos de irmãos completos nos dados simulados, foi possível observar que a grande presença de meio irmãos nas amostras podem aumentaras taxas de classificação erradas de irmãos completos nos dois programas testados,
- As fêmeas do mosquito *Ae. aegypti* na população de Paquetá, no Rio de Janeiro, possuem comportamento de oviposição em saltos. As fêmeas parecem ter dispersado menos no inverno, pois nenhum grupo de irmãos completos foi encontrado entre diferentes armadilhas, apenas dentro da mesma armadilha.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- Abreu F V, Morais M M, Ribeiro S P & Eiras A E (2015). Influence of breeding site availability on the oviposition behaviour of *Aedes aegypti*. Memorias do Instituto Oswaldo Cruz **110**(5): 669-676.
- Ackerman M W, Hand B K, Waples R K, Luikart G, Waples R S, Steele C A, *et al.* (2016). Effective number of breeders from sibship reconstruction: empirical evaluations using hatchery steelhead. Evolutionary applications.
- Aglieri G, Papetti C, Zane L, Milisenda G, Boero F & Piraino S (2014). First evidence of inbreeding, relatedness and chaotic genetic patchiness in the holoplanktonic jellyfish *Pelagia noctiluca* (Scyphozoa, Cnidaria). PloS one **9**(6): e99647.
- Apostol B L, Black IV W C, Miller B R, Reiter P & Beaty B J (1993). Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applications to the mosquito *Aedes aegypti*. Theoretical and Applied Genetics **86**: 991-1000.
- Apostol B L, Black IV W C, Reiter P & Miller B R (1994). Use of randomly amplified polymorphic DNA amplified by polymerase chain reaction markers to estimate the number of *Aedes aegypti* families at oviposition sites in San Juan, Puerto Rico. American Journal of Tropical Medicine and Hygiene **51**(1): 89-97.
- Apostol B L, Black IV W C, Reiter P & Miller B R (1996). Population genetics with RAPD-PCR markers: the breeding structure of *Aedes aegypti* in Puerto Rico. Heredity **76**: 325-334.
- Araujo A Q, Silva M T & Araujo A P (2016). Zika virus-associated neurological disorders: a review. Brain: a journal of neurology **139**(Pt 8): 2122-2130.

- Ashley M V, Caballero I C, Chaovalitwongse W, Dasgupta B, Govindan P, Sheikh S I, *et al.* (2009). KINALYZER, a computer program for reconstructing sibling groups. Molecular ecology resources **9**(4): 1127-1131.
- Avise J C (1994). Molecular Markers, Natural History, and Evolution. New York, Chapman & Hall.
- Babin R, Fenouillet C, Legavre T, Blondin L, Calatayud C, Risterucci A M, *et al.* (2012). Isolation and characterization of twelve polymorphic microsatellite *Loci* for the cocoa mirid bug *Sahlbergella singularis*. International journal of molecular sciences **13**(4): 4412-4417.
- Balloux F (2001). EASYPOP (Version 1.7): A computer program for population genetics simulation. Journal of Heredity **92**(3): 301-302.
- Bellinato D F, Viana-Medeiros P F, Araujo S C, Martins A J, Lima J B & Valle D (2016). Resistance Status to the Insecticides Temephos, Deltamethrin, and Diflubenzuron in Brazilian *Aedes aegypti* Populations. BioMed research international **2016**: 8603263.
- Bentley M D & Day J F (1989). Chemical ecology and behavioral aspects of mosquito oviposition. Annual Review in Entomology **34**: 401-421.
- Blouin M S, Parsons M, Lacaille V & Lotz S (1996). Use of microsatellite loci to classify individuals by relatedness. Molecular ecology **5**: 393-401.
- Blouin M S (2003). DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. Trends in Ecology & Evolution **18**(10): 503-511.
- Boyer S, Toty C, Jacquet M, Lemperiere G & Fontenille D (2012). Evidence of multiple inseminations in the field in *Aedes albopictus*. PloS one **7**(8): e42040.

Bon Eco Solutions. About Tiger mosquitoes. [Acesso em 12 de novembro de 2016].

Disponível em:

<http://www.bonecosolutions.com/ENG/MOSQ20ENG/abouttigeremosqui.html>

Brady O J, Gething P W, Bhatt S, Messina J P, Brownstein J S, Hoen A G, *et al.* (2012).

Refining the global spatial limits of dengue virus transmission by evidence-based consensus. PLoS neglected tropical diseases **6**(8): e1760.

Braga I A, Gomes A C, Nelson M J, Mello R d C, Bergamaschi D P & Souza J M P

(2000). Comparação entre pesquisa larvária e armadilha de oviposição, para detecção de *Aedes aegypti*. Revista da Sociedade Brasileira de Medicina Tropical **33**(4): 347-353.

Braga I A & Valle D (2007a). *Aedes aegypti*: Histórico de Controle no Brasil.

Epidemiol Serv Saúde **16**: 113-118.

Braga I A & Valle D (2007b). *Aedes aegypti*: inseticidas, mecanismos de ação e

resistência. Epidemiol Serv Saúde **16**: 277-281.

Braks M A H, Honório N A, Lounibos L P, Lourenço-de-Oliveira R & Juliano S A

(2004). Interspecific Competition Between Two Invasive Species of Container Mosquitoes, *Aedes aegypti* and *Aedes albopictus* (Diptera: Culicidae), in Brazil. Annals of the Entomological Society of America **97**(1): 130-139.

Brathwaite Dick O, San Martin J L, Montoya R H, del Diego J, Zambrano B & Dayan

G H (2012). The history of dengue outbreaks in the Americas. The American journal of tropical medicine and hygiene **87**(4): 584-593.

Briegel H, Hefti M & DiMarco E (2002). Lipid metabolism during sequential

gonotrophic cycles in large and small female *Aedes aegypti*. Journal of Insect Physiology **48**: 547-554.

- Brown J E, McBride C S, Johnson P, Ritchie S, Paupy C, Bossin H, *et al.* (2011). Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases. Proceedings. Biological sciences / The Royal Society **278**(1717): 2446-2454.
- Brown J E, Evans B R, Zheng W, Obas V, Barrera-Martinez L, Egizi A, *et al.* (2014). Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. Evolution; international journal of organic evolution **68**(2): 514-525.
- Campos G S, Bandeira A C & Sardi S I (2015). Zika Virus Outbreak, Bahia, Brazil. Emerging Infectious Diseases **21**(10): 1885-1886.
- Carvalho D O, McKemey A R, Garziera L, Lacroix R, Donnelly C A, Alphey L, *et al.* (2015). Suppression of a Field Population of *Aedes aegypti* in Brazil by Sustained Release of Transgenic Male Mosquitoes. PLoS neglected tropical diseases **9**(7): e0003864.
- Centres for Disease Control and Prevention-CDC. Case count in the US. [Acesso em: 10 de novembro de 2016]. Disponível em: www.cdc.gov/zika/geo/united-states.html.
- Chadee D D, Beier J C & Mohammed R T (2002). Fast and slow blood-feeding durations of *Aedes aegypti* mosquitoes in Trinidad. Journal of Vector Ecology **27**(2): 172-177.
- Chadee D D (2009). Oviposition strategies adopted by gravid *Aedes aegypti* (L.) (Diptera: Culicidae) as detected by ovitraps in Trinidad, West Indies (2002-2006). Acta tropica **111**(3): 279-283.
- Chambers E W, Meece J K, McGowan J A, Lovin D D, Hemme R R, Chadee D D, *et al.* (2007). Microsatellite isolation and linkage group identification in the yellow fever mosquito *Aedes aegypti*. The Journal of heredity **98**(3): 202-210.

- Chen H, Fillinger U & Yan G (2006). Oviposition behavior of female *Anopheles gambiae* in Western Kenya inferred from microsatellite markers. American Journal of Tropical Medicine and Hygiene **75**(2): 246-250.
- Chen H, Minakawa N, Cui L & Yan G (2008). Conspecific Sharing of Breeding Sites by Anopheline Female Mosquitoes (Diptera: Culicidae) Inferred from Microsatellite Markers. Journal of Insect Behavior **21**(1): 24-33.
- Christophers S R (1960). *Aedes aegypti* (L) the yellow fever mosquito. Cambridge, Cambridge University Press.
- Claro L B L, Tomassini H C B & Rosa M L G (2004). Dengue prevention and control: a review of studies on knowledge, beliefs, and practices. Cadernos de Saúde Pública **20**(6): 1447-1457.
- Coffey L L, Failloux A B & Weaver S C (2014). Chikungunya virus-vector interactions. Viruses **6**(11): 4628-4663.
- Colton Y M, Chadee D D & Severson D W (2003). Natural skip oviposition of the mosquito *Aedes aegypti* indicated by codominant genetic markers. Medical and Veterinary Entomology **17**: 195-204.
- Consoli R & Lourenço-de-Oliveira R (1994). Principais mosquitos de importância sanitária no Brasil. Rio de Janeiro, Brazil, Ed. Fiocruz.
- Couret J, Dotson E & Benedict M Q (2014). Temperature, larval diet, and density effects on development rate and survival of *Aedes aegypti* (Diptera: Culicidae). PloS one **9**(2): e87468.
- Csillery K, Johnson T, Beraldi D, Clutton-Brock T, Coltman D, Hansson B, *et al.* (2006). Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. Genetics **173**(4): 2091-2101.

- Dexter D & Brown D G (2013). Fast half-sibling population reconstruction: theory and algorithms. Algorithms for Molecular Biology **8**(20): 1-12.
- Dutra H L, Dos Santos L M, Caragata E P, Silva J B, Villela D A, Maciel-de-Freitas R, *et al.* (2015). From lab to field: the influence of urban landscapes on the invasive potential of Wolbachia in Brazilian *Aedes aegypti* mosquitoes. PLoS neglected tropical diseases **9**(4): e0003689.
- Edman J D, Scott T W, Costero A, Morrison A C, Harrington L C & Clark G G (1998). *Aedes aegypti* (Diptera: Culicidae) movement influenced by availability of oviposition sites. Journal of Medical Entomology **35**(4): 578-583.
- Ellegren H (2004). Microsatellites: simple sequences with complex evolution. Nature reviews. Genetics **5**(6): 435-445.
- Farnesi L C, Martins A J, Valle D & Rezende G L (2009). Embryonic development of *Aedes aegypti* (Diptera: Culicidae): influence of different constant temperatures. Memórias do Instituto Oswaldo Cruz **104**(1): 124-126.
- Fay R W & Eliason D A (1966). A Preferred oviposition site as a surveillance method for *Aedes aegypti*. Mosquito News **26**(4): 531-535.
- Ferreira, M E ; Grattapaglia, D. (1988). Introdução ao uso de marcadores moleculares em análise genética. 3.ed. Brasília: EMBRAPA-CENARGEN. 220p.
- Forattini O P (1962). Entomologia Médica. São Paulo, Faculdade de Higiene e Saúde Pública.
- Fundação Nacional de Saúde (2001). Dengue instruções para o pessoal de combate ao vetor: manual de normas técnicas. Brasília, 3. ed., 84p. 1–84.
- Fundação Nacional de Saúde (2002). Programa Nacional de Controle da Dengue.1.32.
- Ganesan K, Mendki M J, Suryanarayana M V S, Prakash S & Malhotra R C (2006). Studies of *Aedes aegypti* (Diptera: Culicidae) ovipositional responses to newly

- identified semiochemicals from conspecific eggs. Australian Journal of Entomology **45**(1): 75-80.
- Gentile C, Meireles-Filho A C A, Britto C, Lima J B P, Valle D & Peixoto A A (2006). Cloning and daily expression of the timeless gene in *Aedes aegypti* (Diptera:Culicidae). Insect Biochemistry and Molecular Biology **36**(11): 878-884.
- Gentile C, Rivas G B, Meireles-Filho A C, Lima J B & Peixoto A A (2009). Circadian expression of clock genes in two mosquito disease vectors: cry2 is different. Journal of biological rhythms **24**(6): 444-451.
- Gloria-Soria A, Ayala D, Bheecarry A, Calderon-Arguedas O, Chadee D D, Chiappero M, *et al.* (2016). Global genetic diversity of *Aedes aegypti*. Molecular ecology **25**(21): 5377-5395.
- Guzman M G, Halstead S B, Artsob H, Buchy P, Farrar J, Gubler D J, *et al.* (2010). Dengue: a continuing global threat. Nature reviews. Microbiology **8**(12 Suppl): 7-16.
- Harrington L C & Edman J D (2001). Indirect evidence against delayed "Skip-Oviposition" behavior by *Aedes aegypti* (Diptera: Culicidae) in Thailand. Journal of Medical Entomology **38**(5): 641-645.
- Harrington L C, Ponlawat A, Edman J D, Scott T W & Vermeulen F (2008). Influence of container size, location, and time of day on oviposition patterns of the dengue vector, *Aedes aegypti*, in Thailand. Vector borne and zoonotic diseases **8**(3): 415-423.
- Harrison H B, Saenz-Agudelo P, Planes S, Jones G P & Berumen M L (2013a). On minimizing assignment errors and the trade-off between false positives and negatives in parentage analysis. Molecular ecology **22**: 5738-5742.

- Harrison H B, Saenz-Agudelo P, Planes S, Jones G P & Berumen M L (2013b). Relative accuracy of three common methods of parentage analysis in natural populations. Molecular ecology **22**(4): 1158-1170.
- Hauser L, Baird M, Hilborn R, Seeb L W & Seeb J E (2011). An empirical comparison of SNPs and microsatellites for parentage and kinship assignment in a wild sockeye salmon (*Oncorhynchus nerka*) population. Molecular ecology resources **11 Suppl 1**: 150-161.
- Helinski M E, Valerio L, Facchinelli L, Scott T W, Ramsey J & Harrington L C (2012). Evidence of polyandry for *Aedes aegypti* in semifield enclosures. The American journal of tropical medicine and hygiene **86**(4): 635-641.
- Hemingway J & Ranson H (2000). Insecticide resistance in insect vectors of human disease. Annual Review in Entomology **45**: 371-391.
- Hoban S, Bertorelle G & Gaggiotti O E (2012). Computer simulations: tools for population and evolutionary genetics. Nature reviews. Genetics **13**(2): 110-122.
- Hoffman, J I & Amos W (2005). Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology **14**: 59-612.
- Hotez P J, Bottazzi M E, Franco-Paredes C, Ault S K & Periago M R (2008). The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. PLoS neglected tropical diseases **2**(9): e300.
- Huang S, Molaei G & Andreadis T G (2008). Genetic Insights into the Population Structure of *Culex pipiens* (Diptera: Culicidae) in the Northeastern United States by Using Microsatellite Analysis. American Journal of Tropical Medicine and Hygiene **79**(4): 518-527.

Instituto Brasileiro de Geografia e Estatística. [Acesso em 01 de novembro de 2016].

Disponível em:

<http://www.censo2010.ibge.gov.br/sinopse/index.phpuf=33dados=1> .

Jones A G, Small C M, Paczolt K A & Ratterman N L (2010). A practical guide to methods of parentage analysis. Molecular ecology resources **10**(1): 6-30.

Jones O R & Wang J (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular ecology resources **10**(3): 551-555.

Juliano S A, Ribeiro G S, Maciel-de-Freitas R, Castro M G, Codeco C, Lourenco-de-Oliveira R, *et al.* (2014). She's a femme fatale: low-density larval development produces good disease vectors. Memorias do Instituto Oswaldo Cruz **109**(8): 1070-1077.

Kalinowski S T (2004). Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conservation Genetics Resources **5**: 539-543.

Kalinowski S T (2005). Hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes **5**(1): 187-189.

Kalinowski S T, Wagner A P & Taper M L (2006). ML-Relate: a computer program for maximum likelihood estimation of relatedness and relationship. Molecular Ecology Notes **6**(2): 576-579.

Kirkpatrick S, Gelatt C D & Vecchi M P (1983). Optimization by simulated annealing. Science **220**(4598): 671-680.

Klase Z A, Khakhina S, Schneider Ade B, Callahan M V, Glasspool-Malone J & Malone R (2016). Zika Fetal Neuropathogenesis: Etiology of a Viral Syndrome. PLoS neglected tropical diseases **10**(8): e0004877.

- Kobayashi T, Sakurai T, Sakakibara M & Watanabe T (2011). Multiple origins of outbreak populations of a native insect pest in an agro-ecosystem. Bulletin of entomological research **101**(3): 313-324.
- Kopps A M, Kang J, Sherwin W B & Palsboll P J (2015). How Well Do Molecular and Pedigree Relatedness Correspond, in Populations with Diverse Mating Systems, and Various Types and Quantities of Molecular and Demographic Data? G3 **5**(9): 1815-1826.
- Kraemer M U, Sinka M E, Duda K A, Mylne A Q, Shearer F M, Barker C M, *et al.* (2015). The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. eLife **4**: e08347.
- Kumar V, Rama A, Kesari S, Bhunia G S, Dinesh D S & Das P (2013). Oviposition behaviour of *Phlebotomus argentipes*--a laboratory-based study. Memorias do Instituto Oswaldo Cruz **108**(8): 1065-1067.
- Larson S, Farrer D, Lowry D & Ebert D A (2015). Preliminary Observations of Population Genetics and Relatedness of the Broadnose Sevengill Shark, *Notorynchus cepedianus*, in Two Northeast Pacific Estuaries. PloS one **10**(6): e0129278.
- Le Menach A, McKenzie F E, Flahault A & Smith D L (2005). The unexpected importance of mosquito oviposition behaviour for malaria: non-productive larval habitats can be sources for malaria transmission. Malaria journal **4**: 23.
- Li Y-C, Korol A B, Fahima T, Beiles A & Nevo E (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Molecular ecology **11**: 2453-2465.
- Linss J G, Brito L P, Garcia G A, Araki A S, Bruno R V, Lima J B, *et al.* (2014). Distribution and dissemination of the Val1016Ile and Phe1534Cys Kdr

- mutations in *Aedes aegypti* Brazilian natural populations. Parasites & vectors **7**: 25.
- Lounibos L P (1981). Habitat segregation among African treehole mosquitoes. Ecological Entomology **6**: 129-154.
- Lovin D D, Washington K O, deBruyn B, Hemme R R, Mori A, Epstein S R, *et al.* (2009). Genome-based polymorphic microsatellite development and validation in the mosquito *Aedes aegypti* and application to population genetics in Haiti. BMC genomics **10**: 590.
- Lowy I (1990). Yellow fever in Rio de Janeiro and the Pasteur Institute mission (1901-1905): The transfer of science to the periphery. Medical History **34**: 144-163.
- Luz P M, Codeco C T, Medlock J, Struchiner C J, Valle D & Galvani A P (2009). Impact of insecticide interventions on the abundance and resistance profile of *Aedes aegypti*. Epidemiology and infection **137**(8): 1203-1215.
- Maciel-de-Freitas R, Aguiar R, Bruno R V, Guimarães M C, Lourenço-de-Oliveira R, Sorgine M H, *et al.* (2012). Why do we need alternative tools to control mosquito-borne diseases in Latin America? . Memórias do Instituto Oswaldo Cruz **107**: 828-829.
- McGraw E A & O'Neill S L (2013). Beyond insecticides: new thinking on an ancient problem. Nature reviews. Microbiology **11**(3): 181-193.
- Ministério da Saúde/ Sistema de Vigilância em Saúde (2009). Diretrizes Nacionais para a Prevenção e Controle de Epidemias de Dengue. Série A. Normas e Manuais Técnicos. 160 .
- Ministério da Saúde/ Sistema de Vigilância em Saúde (2017). Monitoramento de casos de dengue, febre de chikungunya e febre pelo vírus Zika até Semana epidemiológica 51, 2016. Boletim Epidemiológico. **48** (2).

- Mino C I, Russello M A, Mussi Goncalves P F & Del Lama S N (2011). Reconstructing genetic mating systems in the absence of parental information in colonially breeding waterbirds. BMC evolutionary biology **11**: 196.
- Mlakar J, Korva M, Tul N, Popovic M, Poljsak-Prijatelj M, Mraz J, *et al.* (2016). Zika Virus Associated with Microcephaly. The New England journal of medicine **374**(10): 951-958.
- Mogi M & Mokry J (1980). Distribution of *Wyeomyia smithii* (Diptera, Culicidae) eggs in pitcher plants in Newfoundland, Canada. Tropical Medicine **22**(1): 1-12.
- Monteiro F A, Schama R, Martins A J, Gloria-Soria A, Brown J E & Powell J R (2014). Genetic diversity of Brazilian *Aedes aegypti*: patterns following an eradication program. PLoS neglected tropical diseases **8**(9): e3167.
- Montella I R, Martins A J, Viana-Medeiros P F, Lima J B, Braga I A & Valle D (2007). Insecticide resistance mechanisms of Brazilian *Aedes aegypti* populations from 2001 to 2004. The American journal of tropical medicine and hygiene **77**(3): 467-477.
- Moreira L A, Iturbe-Ormaetxe I, Jeffery J A, Lu G, Pyke A T, Hedges L M, *et al.* (2009). A *Wolbachia* symbiont in *Aedes aegypti* limits infection with dengue, Chikungunya, and Plasmodium. Cell **139**(7): 1268-1278.
- Navarro-Silva M A, Marques F A & Duque-L J E (2009). Review of semiochemicals that mediate the oviposition of mosquitoes: a possible sustainable tool for the control and monitoring of Culicidae. Revista Brasileira de Entomologia **53**(1): 1-6.
- Nelson M J (1986). *Aedes aegypti*: Biologia Y Ecologia. Washington, Organização Panamericana de Saúde.

- Oliva L O, Correia J C & Albuquerque C M R (2013). How Mosquito Age and the Type and Color of Oviposition Sites Modify Skip-Oviposition Behavior in *Aedes aegypti* (Diptera: Culicidae)? Journal of Insect Behavior **27**(1): 81-91.
- Oliveira E J, Pádua J G, Zucchi M I, Vencovsky R & Vieira M L C (2006). Origin, evolution and genome distribution of microsatellites. Genetics and Molecular Biology **29**(2): 294-307.
- Olsen J B, Busack C, Britt J & Bentzen P (2001). The Aunt and Uncle effect: an empirical evaluation of the confounding influence of full sibs of parents on pedigree reconstruction. Journal of Heredity **92**(243-247).
- Ottmann D, Grorud-Colvert K, Sard N M, Huntington B E, Banks M A & Sponaugle S (2016). Long-term aggregation of larval fish siblings during dispersal along an open coast. Proceedings of the National Academy of Sciences of the United States of America **113**(49): 14067-14072.
- Painter I (1997). Sibship reconstruction without parental information. Journal of Agricultural, Biological, and Environmental Statistics **2**(2): 212-229.
- Panigrahi S K, Barik T K, Mohanty S & Tripathy N K (2014). Laboratory Evaluation of Oviposition Behavior of Field Collected *Aedes* Mosquitoes. Journal of Insects **2014**: 1-8.
- Peakall R & Smouse P E (2012). GenA1Ex 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics **28**(19): 2537-2539.
- Prefeitura do Rio de Janeiro. Empresa de Turismo do Rio de Janeiro. Ilha de Paquetá. [Acesso em 04 de novembro de 2016]. Disponível em: <http://www.rio.rj.gov.br/web/riotur/exibeconteudo?id=157598>.

- Ponnusamy L, Xu N, Nojima S, Wesson D M, Schal C & Apperson C S (2008). Identification of bacteria and bacteria-associated chemical cues that mediate oviposition site preferences by *Aedes aegypti*. Proceedings of the National Academy of Sciences of the United States of America **105**(27): 9262-9267.
- Pontes R J S & Ruffino-Netto A (1994). Dengue in urban locality of Southeastern, Brazil: epidemiological aspects. Revista de Saúde Pública **28**(3): 218-227.
- Powell J (2016). Mosquitoes on the Move. Spread and hybridization of *Aedes aegypti* raise the risk of Zika, dengue, and other viral epidemics. Science **354**(6315): 971-972.
- Powell J R & Tabachnick W J (2013). History of domestication and spread of *Aedes aegypti*--a review. Memorias do Instituto Oswaldo Cruz **108 Suppl 1**: 11-17.
- Raymond M & Rousset F (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. Journal of Heredity **86**: 248-249.
- Reiter P, Amador M A, Anderson R A & Clark G G (1995). Short Report: Dispersal of *Aedes aegypti* in an urban area after blood feeding as demonstrated by rubidium-marked eggs. American Journal of Tropical Medicine and Hygiene **52**(2): 177-179.
- Reiter P (2007). Oviposition, dispersal, and survival in *Aedes aegypti*: implications for the efficacy of control strategies. Vector borne and zoonotic diseases **7**(2): 261-273.
- Rey J R & O'Connell S M (2014). Oviposition by *Aedes aegypti* and *Aedes albopictus*: Influence of congeners and oviposition site characteristics. Journal of Vector Ecology **39**(1): 190-196.

- Rezende G L, Martins A J, Gentile C, Farnesi L C, Pelajo-Machado M, Peixoto A A, *et al.* (2008). Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. BMC developmental biology **8**: 82.
- Rice W R (1989). Analyzing Tables of Statistical Tests. Evolution; international journal of organic evolution **43**(1): 223-225.
- Richardson J B, Jameson S B, Gloria-Soria A, Wesson D M & Powell J (2015). Evidence of limited polyandry in a natural population of *Aedes aegypti*. The American journal of tropical medicine and hygiene **93**(1): 189-193.
- Rigau-Pérez J G, Clark G G, Gubler D J, Reiter P, Sanders E J & Vance Vorndam A (1998). Dengue and dengue haemorrhagic fever. The Lancet **352**(9132): 971-977.
- Ritchie S A, Long S, Hart A, Webb C E & Russel R C (2003). An adulticidal sticky ovitrap for sampling container-breeding mosquitoes. Journal of the American Mosquito Control Association **19**(3): 235-242.
- Rio de Janeiro. Decreto Rio n ° 46 19 de maio de 1999. Tomba definitivamente os bens que menciona, transforma a ilha de Paquetá em área de proteção do ambiente cultural (APAC), e dá outras providências. Diário Oficial da União, 19 de maio de 1999.1-3.
- Robinson S P, Simmons L W & Kennington W J (2013). Estimating relatedness and inbreeding using molecular markers and pedigrees: the effect of demographic history. Molecular ecology **22**(23): 5779-5792.
- Rousset F (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Molecular ecology resources **8**: 103-106.
- Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology **18**: 233-234.

- Sheikh S I, Berger-Wolf T Y, Ashley M V, Caballero I C, Chaovalitwongse W & DasGupta B (2008). Error Tolerant Sibship Reconstruction in Wild Populations. Computational Systems Bioinformatics Conference **7**: 273-284.
- Slotman M A, Kelly N B, Harrington L C, Kitthawee S, Jones J W, Scott T W, *et al.* (2007). Polymorphic microsatellite markers for studies of *Aedes aegypti* (Diptera: Culicidae), the vector of dengue and yellow fever. Molecular Ecology Notes **7**(1): 168-171.
- Stadele V & Vigilant L (2016). Strategies for determining kinship in wild populations using genetic data. Ecology and evolution **6**(17): 6107-6120.
- Tautz D & Renz M (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Research **12**(10): 4127-4138.
- Tautz D (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Research **17**(16): 6463-6471.
- Taylor H R, Kardos M D, Ramstad K M & Allendorf F W (2015). Valid estimates of individual inbreeding coefficients from marker-based pedigrees are not feasible in wild populations with low allelic diversity. Conservation Genetics **16**(4): 901-913.
- Van de Casteele T, Galbusera P & Matthysen E (2001). A comparison of microsatellite-based pairwise relatedness estimators. Molecular ecology **10**(1539-49).
- Wagner A P, Creel S & Kalinowski S T (2006). Estimating relatedness and relationships using microsatellite loci with null alleles. Heredity **97**(5): 336-345.
- Walling C A, Pemberton J M, Hadfield J D & Kruuk L E (2010). Comparing parentage inference software: reanalysis of a red deer pedigree. Molecular ecology **19**(9): 1914-1928.
- Wang J (2004). Sibship reconstruction from genetic data. Genetics **166**: 1963-1979.

- Wang J & Santure A W (2009). Parentage and sibship inference from multilocus genotype data under polygamy. Genetics **181**(4): 1579-1594.
- Wang J (2013). A simulation module in the computer program COLONY for sibship and parentage analysis. Molecular ecology resources **13**(4): 734-739.
- Waples R S & Waples R K (2011). Inbreeding effective population size and parentage analysis without parents. Molecular ecology resources **11 Suppl 1**: 162-171.
- Watrln L, Ghawche F, Larre P, Neau J P, Mathis S & Fournier E (2016). Guillain-Barre Syndrome (42 Cases) Occurring During a Zika Virus Outbreak in French Polynesia. Medicine **95**(14): e3257.
- Weir B S, Anderson A D & Hepler A B (2006). Genetic relatedness analysis: modern data and new challenges. Nature reviews. Genetics **7**(10): 771-780.
- Wong J, Stoddard S T, Astete H, Morrison A C & Scott T W (2011). Oviposition site selection by the dengue vector *Aedes aegypti* and its implications for dengue control. PLoS neglected tropical diseases **5**(4): e1015.
- Zahiri N, Rau M E & Lewis D J (1997). Starved Larvae of *Aedes aegypti* (Diptera: Culicidae) Render Waters Unattractive to Ovipositing Conspecific Females. Environmental Entomology **26**(5): 1087-1090.
- Zahiri N & Rau M E (1998). Oviposition attraction and repellency of *Aedes aegypti* (Diptera: Culicidae) to waters from conspecific larvae subjected to crowding, confinement, starvation, or infection. Journal of Medical Entomology **35**(5): 782-787.
- World Health Organization. (2009). Dengue: Guidelines for diagnosis, treatment prevention and control. Geneva: WHO Press. 1-146.

World Health Organization (2016). Zika virus, microcephaly and guillain-barré syndrome. Situation Report. World Health Organization :1–10.

8 ANEXO I

Parâmetros usados para criação dos dados simulados. Espécie dióica e diploide; sem consanguinidade (endocruzamento), com sistema de acasalamento poligâmico, matrizes de acasalamento com 12 representantes maternos e 12 paternos. Método de análise escolhido foi o de máxima verossimilhança, com uma corrida de tamanho médio, alta precisão, uma réplica e distribuição da frequência alélica conhecida. As frequências alélicas usadas foram obtidas na análise geral da população de Paquetá com erro de genotipagem igual a 0,001. As matrizes de acasalamento com 12 representantes maternos e paternos inseridas no programa Colony para criação dos dados simulados estão abaixo. As matrizes variam de acordo com a porcentagem de meio irmãos no total dos 250 indivíduos criados ao final.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
P1	19	0	0	0	0	0	0	0	0	0	0	0
P2	0	29	0	0	0	0	0	0	0	0	0	0
P3	0	0	16	0	0	0	0	0	0	0	0	0
P4	0	0	0	15	0	0	0	0	0	0	0	0
P5	0	0	0	0	25	0	0	0	0	0	0	0
P6	0	0	0	0	0	17	0	0	0	0	0	0
P7	0	0	0	0	0	0	23	0	0	0	0	0
P8	0	0	0	0	0	0	0	32	0	0	0	0
P9	0	0	0	0	0	0	0	0	40	0	0	0
P10	0	0	0	0	0	0	0	0	0	11	0	0
P11	0	0	0	0	0	0	0	0	0	0	14	0
P12	0	0	0	0	0	0	0	0	0	0	0	9

(a) 0% de meio irmão

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
P1	19	0	0	0	0	0	1	0	0	1	0	0
P2	0	29	0	1	0	1	0	0	0	0	0	0
P3	0	0	16	0	0	0	1	1	1	0	0	0
P4	0	0	0	15	0	0	0	0	1	1	0	0
P5	0	0	0	0	25	0	1	1	0	1	0	0
P6	0	0	0	0	0	17	0	0	0	0	0	0
P7	0	0	0	0	0	0	23	0	0	1	0	0
P8	0	0	0	0	0	0	0	32	0	0	0	0
P9	0	0	0	0	0	0	0	0	40	0	0	0
P10	0	0	0	0	0	0	0	0	0	6	0	0
P11	0	0	0	0	0	0	0	0	0	0	8	0
P12	0	0	0	0	0	0	0	0	0	0	0	7

(b) 5% de meio irmão

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
P1	10	1	1	1	0	0	1	1	0	1	0	0
P2	0	14	0	1	1	1	0	1	1	0	0	0
P3	0	0	31	0	1	0	1	1	1	0	0	0
P4	0	0	0	15	0	0	1	0	0	1	0	0
P5	0	0	0	0	25	0	1	1	1	1	0	0
P6	0	0	0	0	0	17	0	1	1	1	0	0
P7	0	0	0	0	0	0	23	0	0	1	0	0
P8	0	0	0	0	0	0	0	22	0	0	0	0
P9	0	0	0	0	0	0	0	0	12	0	0	0
P10	0	0	0	0	0	0	0	0	0	12	0	0
P11	0	0	0	0	0	0	0	0	0	0	13	0
P12	0	0	0	0	0	0	0	0	0	0	0	31

(c) 10% de meio irmão

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
P1	22	1	1	1	1	1	1	1	0	0	0	0
P2	1	20	0	1	1	1	1	1	1	1	0	0
P3	0	0	31	0	1	1	1	1	1	1	0	0
P4	0	0	0	14	0	1	1	1	1	1	0	0
P5	0	0	0	0	18	0	1	1	1	1	0	0
P6	1	0	0	0	0	5	1	1	1	1	0	0
P7	0	1	0	1	0	0	12	0	1	1	0	0
P8	1	0	0	0	1	0	0	10	0	1	0	0
P9	0	0	0	1	0	0	1	0	20	0	0	0
P10	1	0	1	0	1	0	1	0	0	21	0	0
P11	0	0	0	0	0	0	0	0	0	0	12	0
P12	0	0	0	0	0	1	0	0	0	0	0	15

(d) 20% de meio irmão

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
P1	12	1	0	1	0	1	0	1	1	1	1	0
P2	1	19	1	1	1	0	1	1	0	1	0	1
P3	1	1	29	1	1	1	1	1	1	1	1	1
P4	0	1	1	14	1	1	0	1	1	0	0	1
P5	1	1	1	1	18	1	0	0	1	0	0	0
P6	0	0	1	1	1	9	1	1	0	1	0	0
P7	1	0	1	0	1	1	10	1	1	1	0	0
P8	1	1	1	1	0	1	1	10	1	1	0	0
P9	1	0	0	1	1	1	1	1	12	1	0	1
P10	1	1	0	1	1	0	1	1	1	13	0	0
P11	0	0	0	0	0	0	0	0	0	0	12	0
P12	0	0	0	0	0	0	0	0	0	0	0	17

(e) 30% de meio irmão

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
P1	14	1	1	1	1	1	1	1	1	1	1	1
P2	1	30	1	1	1	1	1	1	1	1	1	1
P3	1	1	9	1	1	1	1	1	1	1	1	1
P4	1	1	1	10	1	1	1	1	1	1	1	1
P5	1	1	1	1	12	1	1	1	1	1	1	1
P6	1	1	1	1	1	13	1	1	1	1	0	0
P7	1	1	1	1	1	1	10	1	1	1	0	0
P8	1	1	1	1	0	1	1	15	1	1	0	0
P9	1	1	1	1	1	1	1	1	5	1	0	0
P10	1	1	1	1	1	1	1	1	1	13	0	0
P11	0	0	0	0	0	0	0	0	0	0	12	0
P12	0	0	1	0	0	0	0	0	0	0	0	7

(f) 40% de meio irmão

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
P1	25	1	1	1	1	1	1	1	1	1	1	1
P2	1	11	1	1	1	1	1	1	1	1	1	1
P3	1	1	9	1	1	1	1	1	1	1	1	1
P4	1	1	1	7	1	1	1	1	1	1	1	1
P5	1	1	1	1	11	1	1	1	1	1	1	1
P6	1	1	1	1	1	5	1	1	1	1	1	1
P7	1	1	1	1	0	1	12	1	1	1	1	1
P8	1	1	1	1	1	0	1	4	1	1	1	1
P9	1	1	1	1	1	0	1	1	2	0	1	0
P10	1	1	1	1	1	0	1	1	1	6	0	1
P11	1	1	1	1	1	1	1	1	1	1	17	1
P12	1	1	1	1	1	1	1	1	1	1	1	16

(f) 50% de meio irmão

9 ANEXO II

Scripts usados para a melhor e mais eficiente avaliação dos resultados obtidos.

```
#####MLERROS.PL#####  
#####  
  
#!/usr/bin/perl  
  
use strict;  
  
use warnings;  
  
#mlerros.pl - para separar e contar cada erro existente nas simulações  
do programa ml relate.  
  
#usage: ./script.pl list.txt  
  
my $pop_file = $ARGV[0];  
  
my @line;  
  
my $pai1;  
  
my $pai2;  
  
my $mae1;  
  
my $mae2;  
  
#primeira condição  
  
my @igual_FS;  
  
my @igual_HS;  
  
my @igual_U;  
  
my @igual_PO;
```

```
#erros
```

```
my @no_sib;
```

```
my @error2;
```

```
my @errorhs;
```

```
my @erroru;
```

```
my @error1;
```

```
#contadores
```

```
my $I_FS_cont = "0";
```

```
my $I_HS_cont = "0";
```

```
my $errorfs_cont = "0";
```

```
my $errorhs_cont = "0";
```

```
my $erroru_cont = "0";
```

```
my $error1_cont = "0";
```

```
my $error2_cont = "0";
```

```
my $half_cont = "0";
```

```
my $no_cont = "0";
```

```
my $total;
```

```
open FILE, '<', $pop_file or die "Error: $!";
```

```
while (<FILE>){
```

```
next if /^M*+F*+C/;
```

```

@line = split/^t/;
$line[0] =~ m/M(.+)F(.+)C/;
$pai1 = $1;
$mae1 = $2;
#print "pais 1 $pai1 .... $mae1\n";
$line[1] =~ m/M(.+)F(.+)C/;
$pai2 = $1;
$mae2 = $2;
#print "pais 2 $pai2 .... $mae2\n";
$line[2] =~ m/^w+/;
#    print "$line[2]";
    if (($pai1 == $pai2) and ($mae1 == $mae2)){
        if($line[2] eq "FS"){
            push (@igual_FS, $_);
            $I_FS_cont++;
        }
        else{
            push(@error2, $_);
            $error2_cont++;
        }
    }
    if (($pai1 != $pai2) and ($mae1 == $mae2)){
        if($line[2] eq "HS"){
            push (@igual_HS, $_);

```

```

        $I_HS_cont++;
    }
else{
    push(@errorhs, $_);
    $errorhs_cont++;
}
}
if (($pai1 == $pai2) and ($mae1 != $mae2)){
    if($line[2] eq "HS"){
        push (@igual_HS, $_);
        $I_HS_cont++;
    }
else{
    push(@errorhs, $_);
    $errorhs_cont++;
}
}
if (($pai1 != $pai2) and ($mae1 != $mae2)){
    if($line[2] eq "U"){
        push (@no_sib, $_);
        $no_cont++;
    }
}
}

```

```

if (($pai1 != $pai2) and ($mae1 != $mae2)){
    if($line[2] eq "FS"){
        push (@error1, $_);
        $error1_cont++;
    }
}

if (($pai1 != $pai2) and ($mae1 != $mae2)){
    if($line[2] eq "HS"){
        push (@erroru, $_);
        $erroru_cont++;
    }
}

if (($pai1 != $pai2) and ($mae1 != $mae2)){
    if($line[2] eq "PO"){
        push (@erroru, $_);
        $erroru_cont++;
    }
}

}

close FILE;

```

```
    $total =
    $I_FS_cont+$error2_cont+$I_HS_cont+$errorhs_cont+$no_cont+$erroru_con
    t+$error1_cont++;
```

```
    open LOG, '>LOG.txt', or die "Error: $!";
```

```
    print LOG "full sib: $I_FS_cont\n full sib error 2:$error2_cont\n half sib:
    $I_HS_cont\n half sib error: $errorhs_cont\n non relatives: $no_cont\n non
    relatives error: $erroru_cont\n non relatives error 1: $error1_cont\n\n Total:
    $total\n\n";
```

```
    close LOG;
```

```
    open WRITE, '>full_sib_ML.txt', or die "Error: $!";
```

```
    print WRITE "##### full sib ML- RELATE#####\n";
```

```
    print WRITE "full sib:$I_FS_cont\n";
```

```
    print WRITE "full sib error:$error2_cont\n";
```

```
    foreach (@igual_FS){
```

```
        print WRITE "$_";
```

```
    }
```

```
    close WRITE;
```

```
    open WRITE1, '>full_sib_error2_ML.txt', or die "Error: $!";
```

```
    print WRITE1 "##### full sib error ML- RELATE
    #####\n";
```

```
    print WRITE1 "full sib:$I_FS_cont\n";
```



```

print WRITE1 "full sib error:$error2_cont\n";

foreach (@error2){
print WRITE1 "$_";
}

close WRITE1;

open WRITE2, '>half_sib.txt', or die "Error: $!";
print WRITE2 "##### half sib ML- RELATE
#####\n";
print WRITE2 "half sib:$I_HS_cont\n";
print WRITE2 "half sib error:$errorhs_cont\n";

foreach (@igual_HS){
print WRITE2 "$_";
}

close WRITE2;

open WRITE3, '>half_sib_error.txt', or die "Error: $!";
print WRITE3 "##### half sib error ML-
RELATE#####\n";
print WRITE3 "half sib:$I_HS_cont\n";
print WRITE3 "half sib error:$errorhs_cont\n";

```

```

foreach (@errorhs){
    print WRITE3 "$_";
}

close WRITE3;

open WRITE4, '>no_relatives.txt', or die "Error: $!";
print WRITE4 "##### non relatives ML-
RELATE#####\n";

print WRITE4 "non relatives :$no_cont\n";
print WRITE4 "non relatives error:$erroru_cont\n";

foreach (@no_sib){
    print WRITE4 "$_";
}

close WRITE4;

open WRITE5, '>no_relatives_error.txt', or die "Error: $!";
print WRITE5 "##### non relatives ML-
RELATE#####\n";

print WRITE5 "non relatives :$no_cont\n";
print WRITE5 "non relatives error:$erroru_cont\n";
foreach (@erroru){
    print WRITE5 "$_";
}

```

```

}

close WRITE5;

open WRITE6, '>no_relatives_error1.txt', or die "Error: $!";

print WRITE6 "##### non relatives ML- RELATE erro
1#####\n";

print WRITE6 "non relatives :$no_cont\n";

print WRITE6 "non relatives error 1:$error1_cont\n";

foreach (@error1){

print WRITE6 "$_";

}

close WRITE6;

exit;

#####ERRORFS.PL#####
#####

#!/usr/bin/perl

use strict;

use warnings;

#mlerros.pl - para separar e contar cada entrada fs classificada em outra
categoria nas simulacoes do programa ml relate, com o argumento
fullsib_error.txt

#usage: ./script.pl list.txt

my $pop_file = $ARGV[0];

my @line;

```

```
my $pai1;
my $pai2;
my $mae1;
my $mae2;
#primeira condicao
my @igual_FS;
my @igual_HS;
my @igual_U;
my @igual_PO;
#erros
my @error2;
my @erroru;
my @error1;
my @error3;
my @error;
#contadores
my $I_FS_cont = "0";
my $I_HS_cont = "0";
my $errorfs_cont = "0";
my $errorhs_cont = "0";
my $erroru_cont = "0";
my $error1_cont = "0";
my $error2_cont = "0";
my $half_cont = "0";
```

```

my $no_cont = "0";

my $error3_cont = "0";

my $total;

my $error_cont = "0";

open FILE, '<', $pop_file or die "Error: $!";

while (<FILE>){

next if /^M*+F*+C/;

@line = split/\t/;

$line[0] =~ m/M(.+)F(.+)C/;

$pai1 = $1;

$mae1 = $2;

#print "pais 1 $pai1 .... $mae1\n";

$line[1] =~ m/M(.+)F(.+)C/;

$pai2 = $1;

$mae2 = $2;

#print "pais 2 $pai2 .... $mae2\n";

$line[2] =~ m/\w+/;

#    print "$line[2]";

    if (($pai1 == $pai2) and ($mae1 == $mae2)){

        if($line[2] eq "FS"){

            push (@igual_FS, $_);

            $I_FS_cont++;

        }

}

```

```

}
if (($pai1 == $pai2) and ($mae1 == $mae2)){
    if($line[2] eq "HS"){
        push (@error1, $_);
        $error1_cont++;
    }
}
}
if (($pai1 == $pai2) and ($mae1 == $mae2)){
    if($line[2] eq "U"){
        push (@error2, $_);
        $error2_cont++;
    }
}
}
if (($pai1 == $pai2) and ($mae1 == $mae2)){
    if($line[2] eq "PO"){
        push (@error3, $_);
        $error3_cont++;
    }
}
}
#else{
#    push(@error, $_);
#    $error_cont++;
#    }
}

```

```

}

close FILE;

$total = $I_FS_cont+$error1_cont+$error2_cont+$error3_cont;

open WRITE, '>error1FS_ML.txt', or die "Error: $!";

print WRITE "##### full sib error ML- RELATE RESULTADO
HS #####\n";

print WRITE "TOTAL:$total\n";

print WRITE "full sib error:$error1_cont\n";

foreach (@error1){

print WRITE "$_";

}

close WRITE;

open WRITE1, '>erro2FS_ML.txt', or die "Error: $!";

print WRITE1 "##### full sib error ML- RELATE RESULTADO U
#####\n";

print WRITE1 " error:$error2_cont\n";

foreach (@error2){

print WRITE1 "$_";

```

```

}

close WRITE1;

open WRITE2, '>erro3FS_ML.txt', or die "Error: $!";

print WRITE2 "##### Full sib error ML- RELATE
RESULTADO PO#####\n";

print WRITE2 " error:$error3_cont\n";

foreach (@error3){
print WRITE2 "$_";
}

close WRITE2;

#open WRITE3, '>verificar.txt', or die "Error: $!";

#print WRITE3 "##### Full sib error ML-
RELATE#####\n";

#print WRITE3 " error:$error_cont\n";

#foreach (@error){
# print WRITE3 "$_";
#}

#close WRITE3;

```



```
exit;
```

```
#####ERROSHS.PL#####  
#####
```

```
#!/usr/bin/perl
```

```
use strict;
```

```
use warnings;
```

```
#erros.pl - para separar e contar cada entrada hs classificados em  
outra categoria nas simulacoes do programa ml relate, com o argumento  
halfsib_erro.txt
```

```
#usage: ./script.pl list.txt
```

```
my $pop_file = $ARGV[0];
```

```
my @line;
```

```
my $pai1;
```

```
my $pai2;
```

```
my $mae1;
```

```
my $mae2;
```

```
#primeira condicao
```

```
my @igual_FS;
```

```
my @igual_HS;
```

```
my @igual_U;
```

```
my @igual_PO;
```

```
#erros
```

```
my @error2;
my @erroru;
my @error1;
my @error3;
my @error;

#contadores

my $I_FS_cont = "0";
my $I_HS_cont = "0";
my $errorfs_cont = "0";
my $errorhs_cont = "0";
my $erroru_cont = "0";
my $error1_cont = "0";
my $error2_cont = "0";
my $half_cont = "0";
my $no_cont = "0";
my $error3_cont = "0";
my $total;
my $error_cont = "0";

open FILE, '<', $pop_file or die "Error: $!";
while (<FILE>){
next if /^M*+F*+C/;

@line = split/\t/;

$line[0] =~ m/M(.+)F(.+)C/;
```

```

$pai1 = $1;

$mae1 = $2;

#print "pais 1 $pai1 .... $mae1\n";

$line[1] =~ m/M(.+)F(.+)C/;

$pai2 = $1;

$mae2 = $2;

#print "pais 2 $pai2 .... $mae2\n";

$line[2] =~ m/\w+/;

#    print "$line[2]";

    if (($pai1 == $pai2) and ($mae1 != $mae2)){

        if($line[2] eq "HS"){

            push (@igual_HS, $_);

            $I_HS_cont++;

        }

    }

    if (($pai1 == $pai2) and ($mae1 != $mae2)){

        if($line[2] eq "FS"){

            push (@error1, $_);

            $error1_cont++;

        }

    }

    if (($pai1 == $pai2) and ($mae1 != $mae2)){

        if($line[2] eq "U"){

            push (@error2, $_);


```

```

        $error2_cont++;
    }
}
if (($pai1 == $pai2) and ($mae1 != $mae2)){
    if($line[2] eq "PO"){
        push (@error3, $_);
        $error3_cont++;
    }
}
if (($pai1 != $pai2) and ($mae1 == $mae2)){
    if($line[2] eq "HS"){
        push (@igual_HS, $_);
        $I_HS_cont++;
    }
}
if (($pai1 != $pai2) and ($mae1 == $mae2)){
    if($line[2] eq "FS"){
        push (@error1, $_);
        $error1_cont++;
    }
}
if (($pai1 != $pai2) and ($mae1 == $mae2)){
    if($line[2] eq "U"){

```

```

        push (@error2, $_);
        $error2_cont++;
    }
}
if (($pai1 != $pai2) and ($mae1 == $mae2)){
    if($line[2] eq "PO"){
        push (@error3, $_);
        $error3_cont++;
    }
}
}
close FILE;

$total = $I_FS_cont+$error1_cont+$error2_cont+$error3_cont;

open WRITE, '>error1HS_ML.txt', or die "Error: $!";
print WRITE "#####full sib error ML- RELATE RESULTADO
FS #####\n";
print WRITE "TOTAL:$total\n";
print WRITE " error:$error1_cont\n";

foreach (@error1){

```

```

print WRITE"$_";
}

close WRITE;

open WRITE1, '>erro2HS_ML.txt', or die "Error: $!";
print WRITE1 "##### error ML- RELATE RESULTADO U
#####\n";
print WRITE1 " error:$error2_cont\n";

foreach (@error2){
print WRITE1 "$_";
}

close WRITE1;

open WRITE2, '>erro3HS_ML.txt', or die "Error: $!";
print WRITE2 "#####error ML- RELATE RESULTADO PO####\n";
print WRITE2 " error:$error3_cont\n";

foreach (@error3){
print WRITE2 "$_";
}

```

```

close WRITE2;

open WRITE3, '>verificar.txt', or die "Error: $!";

print WRITE3 "##### Full sib error ML-
RELATE#####\n";

print WRITE3 " error:$I_HS_cont\n";

foreach (@igual_HS){
print WRITE3 "$_";
}

close WRITE3;

exit;

#####ERROSU.PL#####
#####

#!/usr/bin/perl

use strict;

use warnings;

#errou.pl - para separar e contar cada erro da categoria não relacionado
existente nas simulacoes do programa ml relate com argumento
no_relatives_error.txt

#usage: ./script.pl list.txt

my $pop_file = $ARGV[0];

```

```
my @line;

my $pai1;

my $pai2;

my $mae1;

my $mae2;

#primeira condicao

my @igual_FS;

my @igual_HS;

my @igual_U;

my @igual_PO;

#erros

my @error2;

my @erroru;

my @error1;

my @error3;

my @error;

#contadores

my $_FS_cont = "0";

my $_HS_cont = "0";

my $errorfs_cont = "0";

my $errorhs_cont = "0";

my $erroru_cont = "0";

my $error1_cont = "0";

my $error2_cont = "0";
```



```

my $half_cont = "0";

my $no_cont = "0";

my $error3_cont = "0";

my $total;

my $error_cont = "0";

my $I_U_cont = "0";

open FILE, '<', $pop_file or die "Error: $!";

while (<FILE>){

next if /^M*+F*+C/;

@line = split/^t/;

$line[0] =~ m/M(.+)F(.+)C/;

$pai1 = $1;

$mae1 = $2;

#print "pais 1 $pai1 .... $mae1\n";

$line[1] =~ m/M(.+)F(.+)C/;

$pai2 = $1;

$mae2 = $2;

#print "pais 2 $pai2 .... $mae2\n";

$line[2] =~ m/^w+;/

#    print "$line[2]";

    if (($pai1 != $pai2) and ($mae1 != $mae2)){

        if($line[2] eq "U"){

            push (@igual_U, $_);

            $I_U_cont++;


```

```

    }
}

if (($pai1 != $pai2) and ($mae1 != $mae2)){
    if($line[2] eq "HS"){
        push (@error1, $_);
        $error1_cont++;
    }
    else{
        push(@error2, $_);
        $error2_cont++;
    }
}

}

close FILE;

$total = $I_FS_cont+$error1_cont+$error2_cont+$error3_cont;

open WRITE, '>error1U_ML.txt', or die "Error: $!";

print WRITE "##### error ML- RELATE RESULTADO
HS#####\n";

print WRITE "TOTAL:$total\n";

print WRITE "unrelated error1:$error1_cont\n";

```

```

foreach (@error1){

print WRITE"$_";

}

close WRITE;

open WRITE1, '>erro2U_ML.txt', or die "Error: $!";

print WRITE1 "##### error ML- RELATE RESULTADO PO
#####\n";

print WRITE1 " unrelated error2:$error2_cont\n";

foreach (@error2){

print WRITE1 "$_";

}

close WRITE1;

open WRITE2, '>erro3U_ML.txt', or die "Error: $!";

print WRITE2 "##### error ML- RELATE RESULTADO U
VERIFICAR #####\n";

print WRITE2 "unrelated error3:$I_U_cont\n";

foreach (@igual_U){

print WRITE2 "$_";

```

```
}
```

```
close WRITE2;
```

```
exit;
```

```
#####HISTOGRAMA.pl#####  
#####  
##
```

```
#!/usr/bin/perl
```

```
use strict;
```

```
use warnings;
```

```
#mlhistograma.pl - para separar todos os irmaos completos do ml
```

```
#usage: ./script.pl list.txt
```

```
my $pop_file = $ARGV[0];
```

```
my @line;
```

```
my $pai1;
```

```
my $pai2;
```

```
my $mae1;
```

```
my $mae2;
```

```
#primeira condicao
```

```
my @igual_FS;
```

```
my @igual_HS;
```

```
my @igual_U;
```

```
my @igual_PO;
```

```
#erros
```

```
my @no_sib;
```

```
my @error2;
```

```
my @errorhs;
```

```
my @erroru;
```

```
my @error1;
```

```
#contadores
```

```
my $I_FS_cont = "0";
```

```
my $I_HS_cont = "0";
```

```
my $errorfs_cont = "0";
```

```
my $errorhs_cont = "0";
```

```
my $erroru_cont = "0";
```

```
my $error1_cont = "0";
```

```
my $error2_cont = "0";
```

```
my $half_cont = "0";
```

```
my $no_cont = "0";
```

```
my $total;
```

```
open FILE, '<', $pop_file or die "Error: $!";
```

```
while (<FILE>){
```

```

#next if /^M*+F*+C/;

@line = split/+/;

$line[0] =~ m/M(.+)F(.+)C/;

$pai1 = $1;

$mae1 = $2;

#print "pais 1 $pai1 .... $mae1\n";

$line[1] =~ m/M(.+)F(.+)C/;

$pai2 = $1;

$mae2 = $2;

#print "pais 2 $pai2 .... $mae2\n";

$line[2] =~ m/^(d+)/;

#    print "$line[2]";

    if (($pai1 == $pai2) and ($mae1 == $mae2)){

        if($line[2] >= 0){

            push (@igual_FS, $_);

            $I_FS_cont++;

        }

        else{

            push(@error2, $_);

            $error2_cont++;

        }

    }

}

close FILE;

```

```

$total = $I_FS_cont+$error2_cont;

open LOG, '>LOG.txt', or die "Error: $!";

print LOG "full sib: $I_FS_cont\n full sib error 2:$error2_cont\n Total:
$total\n\n";

close LOG;

open WRITE, '>full_sib_todos_ML.txt', or die "Error: $!";

print WRITE "##### TODOS full sib ML-
RELATE#####\n";

print WRITE "full sib:$I_FS_cont\n";

print WRITE "full sib error:$error2_cont\n";

foreach (@igual_FS){

print WRITE "$_";

}

close WRITE;

exit;

#####COLONY#####
#####

#!/usr/bin/perl

use strict;

```

```
use warnings;
```

```
#para separar no arquivo fullsibdyad e halvesib dyad os pares que foram  
considerados fs(maiores que 0.35) sem ser
```

```
#argumento: fullsibdyad.txt ou halvesibdyad.txt
```

```
#usage: ./script.pl list.txt
```

```
my $pop_file = $ARGV[0];
```

```
my @line;
```

```
my $pai1;
```

```
my $pai2;
```

```
my $mae1;
```

```
my $mae2;
```

```
my @full_sib;
```

```
my @half_sib;
```

```
my @no_sib;
```

```
my @error1;
```

```
my @error2;
```

```
my @error3;
```

```
#contadores
```

```
my $full_cont = "0";
```

```
my $half_cont = "0";
```

```
my $no_cont = "0";
```

```
my $error1_cont = "0";
```



```

my $error2_cont = "0";

my $error3_cont = "0";

my $total;

open FILE, '<', $pop_file or die "Error: $!";

while (<FILE>){

next if /^M*+F*+C/;

@line = split/,/;

$line[0] =~ m/M(.+)F(.+)C/;

$pai1 = $1;

$mae1 = $2;

#print "pais 1 $pai1 .... $mae1\n";

$line[1] =~ m/M(.+)F(.+)C/;

$pai2 = $1;

$mae2 = $2;

#print "pais 2 $pai2 .... $mae2\n";

$line[2] =~ m/\d+;/;

#    print "$line[2]";

    if (($pai1 == $pai2) and ($mae1 == $mae2)){

        if($line[2] >= 0.35){

            push (@full_sib, $_);

            $full_cont++;

        }

        else{

```

```

        push(@error1, $_);
        $error1_cont++;
    }
}
if (($pai1 != $pai2) and ($mae1 == $mae2)){
    if($line[2] >= 0.35){
        push (@half_sib, $_);
        $half_cont++;
    }
    else{
        push(@error2, $_);
        $error2_cont++;
    }
}
if (($pai1 == $pai2) and ($mae1 != $mae2)){
    if($line[2] >= 0.35){
        push (@half_sib, $_);
        $half_cont++;
    }
    else{
        push(@error2, $_);
        $error2_cont++;
    }
}
}

```

```

if (($pai1 != $pai2) and ($mae1 != $mae2)){
    if($line[2] >= 0.35){
        push (@no_sib, $_);
        $no_cont++;
    }
    else{
        push(@error3, $_);
        $error3_cont++;
    }
}
}

```

```
close FILE;
```

```
$total=$full_cont+$error1_cont+$half_cont+$error2_cont+$no_cont+$error3_cont;
```

```
open LOG, '>LOG_colony.txt', or die "Error: $!";
```

```
print LOG "full sib_colony: $full_cont\n full sib error_colony:
$error1_cont\n half sib_colony: $half_cont\n half sib error_colony:
$error2_cont\n non relatives_colony: $no_cont\n non relatives error_colony:
$error3_cont\n\n Total: $total\n\n";
```

```
close LOG;
```

```
open WRITE, '>full_sib_colony.txt', or die "Error: $!";
```

```

print WRITE "##### full sib colony #####\n";

print WRITE "full sib_colony:$full_cont\n";

print WRITE "full sib error:$error1_cont\n";

foreach (@full_sib){

print WRITE "$_";

}

close WRITE;

open WRITE1, '>full_sib_error_colony.txt', or die "Error: $!";

print WRITE1 "##### full sib error colony #####\n";

print WRITE1 "full sib:$full_cont\n";

print WRITE1 "full sib error:$error1_cont\n";

foreach (@error1){

print WRITE1 "$_";

}

close WRITE1;

open WRITE2, '>half_sib_colony.txt', or die "Error: $!";

print WRITE2 "##### half sib colony #####\n";

print WRITE2 "half sib:$half_cont\n";

```

```

print WRITE2 "half sib error:$error2_cont\n";

foreach (@half_sib){
print WRITE2 "$_";
}

close WRITE2;

open WRITE3, '>half_sib_error_colony.txt', or die "Error: $!";
print WRITE3 "##### half sib error colony #####\n";
print WRITE3 "half sib:$half_cont\n";
print WRITE3 "half sib error:$error2_cont\n";

foreach (@error2){
print WRITE3 "$_";
}

close WRITE3;

open WRITE4, '>no_relatives_colony.txt', or die "Error: $!";
print WRITE4 "##### non relatives colony #####\n";
print WRITE4 "non relatives:$no_cont\n";
print WRITE4 "non relatives error:$error3_cont\n";

```

```
foreach (@no_sib){
print WRITE4 "$_";
}
close WRITE4;

open WRITE5, '>no_relatives_error_colony.txt', or die "Error: $!";
print WRITE5 "##### non relatives colony #####\n";
print WRITE5 "non relatives:$no_cont\n";
print WRITE5 "non relatives error:$error3_cont\n";
foreach (@error3){
print WRITE5 "$_";
}
close WRITE5;

exit;
```