

UNIVERSIDADE ESTADUAL PAULISTA

"Júlio de Mesquita Filho"

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

**Detecção, Análise e Descrição de Inversões Cromossômicas em
*Nyssorhynchus darlingi***

FILIPPE TRINDADE BOZONI

PAULO EDUARDO MARTINS RIBOLLA

Trabalho de Conclusão de Curso apresentado ao
Instituto de Biotecnologia, Campus de Botucatu, UNESP,
para obtenção de Bacharel em Ciências Biomédicas.

**BOTUCATU – SP
2023**

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

BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Bozoni, Filipe Trindade.

Detecção, análise e descrição de inversões cromossômicas
em *Nyssorhynchus darlingi* / Filipe Trindade Bozoni. -
Botucatu, 2023

Trabalho de conclusão de curso (bacharelado - Ciências
Biomédicas) - Universidade Estadual Paulista "Júlio de
Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Paulo Eduardo Martins Ribolla

Capes: 20201001

1. Genômica. 2. Inversão cromossômica. 3. Malária.
4. *Anopheles*.

Palavras-chave: Genômica populacional; Inversões
cromossômicas; Malária; *Nyssorhynchus darlingi*.

AGRADECIMENTOS

Em primeiro lugar agradeço a Deus, fonte de vida e graça. Agradeço pela minha vida, minha inteligência, minha família e amigos,

ao meu orientador, Prof. Dr. Paulo Eduardo Martins Ribolla que jamais deixou de me incentivar e ao doutorando Marcus Vinicius, sem a orientação, dedicação e auxílio dele o estudo aqui apresentado seria praticamente impossível,

aos meus pais Fernando e Elaine, que apesar das dificuldades enfrentadas, sempre incentivaram meus estudos,

à minha namorada Alexandra, pela parceria em toda minha trajetória no ensino superior além das demais áreas da vida,

aos docentes e demais funcionários do Instituto de Biociências da Universidade Estadual Paulista “Júlio de Mesquita Filho” pela dedicação em ensinar e alegria no atendimento, que contribuem para que a nossa universidade seja uma das melhores da América Latina,

por último, agradeço a mim, por todo o trabalho duro, persistência e dedicação, pelas noites sem dormir, por acreditar no meu sucesso e nunca desistir.

“Só sabemos com exatidão quando sabemos pouco;
à medida que vamos adquirindo conhecimento,
instala-se a dúvida.”

Goethe

SUMÁRIO

1	INTRODUÇÃO.....	09
2	METODOLOGIA.....	11
2.1	Coleta de amostras.....	11
2.2	Preparação e sequenciamento de amostras	12
2.3	Identificação da espécie.....	13
2.4	Chamada de variante e análise de dados.....	13
2.5	Detecção de inversão cromossômica e genotipagem de amostras.....	13
2.6	Genotipagem de Inversão Cromossômica.....	14
2.7	Testes de correlação cromossômica e testes de associação.....	14
3	RESULTADOS.....	15
3.1	Amostras	16
3.2	Estimativa do desequilíbrio de ligação em pares de SNPs	16
3.3	Análise de componentes principais (PCA)	17
3.4	Teste de alelos	19
3.5	Testes de correlação e associação	21
4	DISCUSSÃO.....	23
5	CONCLUSÃO.....	24
	REFERÊNCIAS.....	25
	APÊNDICE.....	28

RESUMO

A malária é a doença transmitida por artrópodes mais impactante nos países em desenvolvimento. No Brasil a incidência de malária é alta, com a maioria dos casos concentrados na floresta amazônica brasileira, tendo como principal vetor da doença no país o mosquito *Nyssorhynchus darlingi*, o qual é suscetível ao *Plasmodium* humano, é antropofílico e oportunista. Comportamentos de importância epidemiológica podem ser influenciados por polimorfismos de inversões cromossômicas. Essas variantes frequentemente apresentam desequilíbrio de ligação em populações naturais e desempenham um papel importante na adaptação local, além disso, existem fortes evidências de seleção atuando sobre as inversões, mas os alvos gênicos dentro delas são amplamente desconhecidos. Alguns desses polimorfismos já foram descritos em mosquitos do gênero *Anopheles*, associados à eficiência da capacidade de exploração de diferentes ambientes e de adaptação a mudanças de condições ambientais. O objetivo desse estudo, foi realizar a detecção, análise e descrição de polimorfismos de inversão cromossômica em *Ny. darlingi*, bem como avaliar se estão associadas à características de importância epidemiológica, utilizando ferramentas de biologia molecular e bioinformática. Para atingir nossos objetivos, amostras de mosquitos *Ny. darlingi* foram coletadas no município de Mâncio Lima - AC. Foi realizado sequenciamento genômico completo de baixa cobertura (LC-WGS) utilizando técnicas de sequenciamento de nova geração, além da genotipagem por sequenciamento. Inversões cromossômicas foram detectadas por alterações no padrão de desequilíbrio de ligação (LD), análise de componentes principais (PCA) e teste de alelos. Também foram realizados testes de correlação entre as inversões e testes de associação com comportamento de picada e horário de atividade. Como resultados, 321 amostras de Mâncio Lima foram sequenciadas e o painel de genótipos imputados resultou em 4.241.254 SNPs bialélicos. A análise de desequilíbrio de ligação (r^2) pareada de SNP mostrou picos de r^2 mediano ao longo dos cromossomos (indicativos de inversão). Resultados do PCA indicaram concentração de SNPs representativos em regiões específicas e bem delimitadas no genoma em múltiplos componentes principais. Resultados dos testes de alelos entre homozigotos indicaram as regiões genômicas específicas associadas significativamente às inversões, em concordância com resultados do LD e do PCA. As análises de correlação mostraram inversões correlacionadas quando próximas ou sobrepostas e o teste de associação mostrou uma inversão associada ao comportamento de picada.

PALAVRAS-CHAVE: Malária. *Nyssorhynchus darlingi*. Inversões cromossômicas. Genômica populacional.

ABSTRACT

Malaria is the most impactful disease transmitted by arthropods in developing countries. In Brazil, the incidence of malaria is high, with most cases concentrated in the Brazilian Amazon rainforest, with the mosquito *Nyssorhynchus darlingi* as the main vector of the disease in the country, which is susceptible to human Plasmodium, is anthropophilic and opportunistic. Behaviors of epidemiological importance can be influenced by polymorphisms of chromosome inversions. These variants often show linkage disequilibrium in natural populations and play an important role in local adaptation, moreover, there is strong evidence of selection acting on inversions, but the gene targets within them are largely unknown. Some of these polymorphisms have already been described in mosquitoes of the genus Anopheles, associated with the efficient ability to explore different environments and adapt to changes in environmental conditions. The aim of this study was to detect, analyze and describe chromosome inversion polymorphisms in *Ny. darlingi*, as well as to assess whether they are associated with characteristics of epidemiological importance, using molecular biology and bioinformatics tools. To achieve our goals, samples of *Ny. darlingi* mosquitoes were collected in the municipality of Mâncio Lima - AC. Low coverage whole genome sequencing (LC-WGS) was performed using next-generation sequencing techniques. Sequencing genotyping was performed and chromosomal inversions were detected by changes in the linkage disequilibrium (LD) pattern, principal component analysis (PCA) and allele test. Correlation tests were performed between inversions and association tests with biting behavior and time of activity. As results, 321 samples from Mâncio Lima were sequenced and the panel of imputed genotypes resulted in 4,241,254 biallelic SNPs. Paired SNP linkage disequilibrium (r^2) analysis showed median r^2 peaks across chromosomes (indicative of inversion). PCA results indicated concentration of representative SNPs in specific and well-delimited regions in the genome in multiple principal components. Results of allele test between homozygotes indicated the specific genomic regions significantly associated, in agreement with results of LD and PCA. Correlation analysis showed correlated inversions when close or overlapping and the association test showed an inversion associated with biting behavior.

KEYWORDS: Malaria. *Nyssorhynchus darlingi*. Chromosomal inversions. Population genomics.

LISTA DE ILUSTRAÇÕES

Figura 1 - Locais de coleta de <i>Nyssorhynchus darlingi</i>	12
Figura 2 - Análise do desequilíbrio de ligação janelado de pares de SNPs.....	16
Figura 3 - Análise de componentes principais	17
Figura 4 - Concentração de SNPs altamente representativos para cada componente da análise de componentes principais	18
Figura 5 - Teste de alelos referente ao cromossomo 2	19
Figura 6 - Teste de alelos referente ao cromossomo 3	20
Figura 7 - Teste de alelos referente ao cromossomo X	20
Figura 8 - Teste de correlação de Spearman par a par entre genótipos de inversões cromossômicas	21
Figura X - Análise de agrupamento genético IBS.....	28

LISTA DE TABELAS

Tabela 1 - Amostras de <i>Nyssorhynchus darlingi</i> identificadas com BLASTn e COI.....	15
Tabela 2 – Resultado do teste de associação com horário de picada	22
Tabela 3 - Resultado do teste de associação com comportamento de picada	22

LISTA DE ABREVIATURAS E SIGLAS

BLAST: Ferramenta básica de localização de alinhamento local (em inglês, Basic Local Alignment Search Tool)

DNA: Ácido desoxirribonucleico, também abreviado como ADN (em inglês, deoxyribonucleic acid)

DP: Profundidade de sequenciamento (em inglês, Depth).

GBS: Genotipagem por sequenciamento (em inglês, Genotyping by sequencing)

GQ: Qualidade do genótipo (em inglês, Genotype Quality).

LC-WGS: Sequenciamento do genoma completo de baixa cobertura (em inglês, Low Coverage Whole Genome Sequence)

MAF: Frequência do alelo menor (em inglês, Minor allele frequency).

MD: Dados ausentes (em inglês, Missing Data).

SNP: Polimorfismo de nucleotídeo único (em inglês, Single Nucleotide Polymorphism).

VCF: Formato de chamada de variantes (em inglês, Variant Call Format).

WGS: Sequenciamento de genoma completo (em inglês, Whole genome sequencing).

LISTA DE SÍMBOLOS

bp: Pares de base. kb (equivalente a kbp): quilo pares de base (ou 1.000 pares de base).

Km: Quilômetro (ou 1.000 metros).

p_{VALOR} : Nível descritivo do teste estatístico.

p_{FDR} : Nível descritivo do teste estatístico, corrigido para múltiplas comparações pelo método taxa de falso positivo (em inglês, False Discovery Rate).

r^2 : Desequilíbrio de ligação.

1 INTRODUÇÃO

A malária é a doença transmitida por artrópodes mais impactante nos países em desenvolvimento. De acordo com o último relatório mundial da Organização Mundial de Saúde (OMS), em 2020 foram estimados 241 milhões de casos de malária em 85 países, um aumento de 14 milhões comparado a 2019. No mesmo ano, a doença foi responsável pela morte de 627 mil pessoas, um aumento de 69 mil mortes em relação ao ano anterior (World malaria report, WHO, 2021)

O Brasil apresenta alta incidência de malária, no ano de 2020, foram notificados 145.205 casos, e em 2021, cerca de 139.211 casos sendo mais de 99% deles concentrados na região amazônica (Ministério da Saúde, 2022), a qual apresenta 33 das 54 espécies de *Anopheles* presentes no país, porém o principal vetor da doença no Brasil (especialmente na Amazônia), é o mosquito *Nyssorhynchus darlingi*, o qual é altamente suscetível ao *Plasmodium* humano e se reproduz

preferencialmente em coleções de águas claras e rasas, sombreadas, com vegetação e baixa concentração de sal (ROCHA, et.al, 2020).

A alta incidência de malária na região possui alguns fatores determinantes, como a preferência alimentar do *Ny. darlingi* por sangue humano (antropofilismo), o ciclo hidrológico anual e enchentes dos rios, além das atividades de desmatamento, agropecuária, assentamentos e piscicultura, que mantém diferentes condições ambientais que propiciam o processo de adaptação dos mosquitos, a transmissão da malária e dificultam o controle de vetores (Tadei et al. 1998 ; Hiwat e Bretas 2011 ; ROCHA, et.al 2020). Tais alterações ambientais e atividades antrópicas também são os principais fatores que influenciam o processo de adaptação e evolução dos anofelinos, por meio de rearranjos cromossômicos (Coghlan et al. 2005 ; Rafael et al. 2010).

As inversões cromossômicas são mutações estruturais que surgem quando um cromossomo se quebra em dois lugares e se reúne após uma rotação de 180°, as que segregam como polimorfismos possuem variados tamanhos. Inversões comuns em *Drosophila* podem cobrir uma parte substancial de um braço cromossômico e podem ter cerca de 10.000 kb (Krimbas & Powell 1992), outras inversões segregantes podem cobrir apenas alguns kb. Quando surge uma inversão em determinada espécie, indivíduos portadores de um cromossomo padrão e um invertido apresentam pouco ou nenhum crossing-over na região do rearranjo (Sturtevant, 1921), já que na meiose, somente regiões homólogas se pareiam durante o zigóteno e a formação do quiasma só ocorre em segmentos pareados. Dessa forma, elas podem se espalhar reduzindo a recombinação entre alelos que aumentam independentemente a aptidão, sem epistasia ou coadaptação (KIRKPATRICK; BARTON, 2006).

Inversões polimórficas em cromossomos politênicos de mosquitos do gênero *Anopheles* da Europa e Ásia (Sharakhova et al. 2011 ; Liang et al. 2014), sudeste do Brasil e região amazônica (Kreutzer et al. 1972 ; Tadei e Santos 1982 ; Rafael et al . 2010) estão associados à eficiência da capacidade de exploração de diferentes ambientes e de adaptação a mudanças de condições ambientais (KIRKPATRICK; BARTON, 2006). Em *Anopheles gambiae*, polimorfismos de inversão foram associados ao tamanho do corpo e resistência ao estresse térmico e de dessecação, consistentes com um papel importante das inversões na tolerância do mosquito à aridez (CHENG; TAN; HAHN; BESANSKY, 2018).

Nyssorhynchus darlingi apresenta uma composição cromossômica de $2n = 6$ (Rafael e Tadei 1998), com dois pares autossômicos: o maior (III) é submetacêntrico e o menor (II)

metacêntrico. No par sexual, X é acrocêntrico e Y puntiforme. Alguns polimorfismos de inversão já foram descritos em *Ny. darlingi* (Tadei et al. 1982 ; Rafael et al. 2010), porém, os estudos sobre a citogenética molecular de rearranjos cromossômicos de *Ny. darlingi* em populações naturais ainda são escassos.

2 METODOLOGIA

Coleta de amostras

Amostras de larvas e adultos foram capturadas em três diferentes pontos de coleta no município de Mâncio Lima, estado do Acre (Fig. 01) durante dezembro de 2016 e fevereiro, maio e setembro de 2017. Anofelinos adultos foram coletados por *Human Landing Catch* (HLC), durante coletas de 12 horas, das 18:00 às 06:00, durante dois dias para cada ponto de coleta, com dois voluntários internos e dois externos, alternando para mitigar o viés específico do coletor. Esses três pontos de coleta foram localizados ao redor de três casas, conforme ilustrado na Figura 1. As três amostras são: (A) casa relativamente distante do centro da cidade e das ruas principais, bem como de áreas de mata; (B) casa localizada próxima ao centro da cidade, ao lado de ruas asfaltadas e distante de áreas de mata; (C) casa relativamente distante do centro da cidade e próxima a áreas florestais. O comportamento de picada foi registrado e classificado como interno ou peridoméstico (externo). As distâncias lineares aproximadas entre os pontos de coleta são: 1,96 km de A a B, 3,39 km de A a C e 2,51 km de B a C.

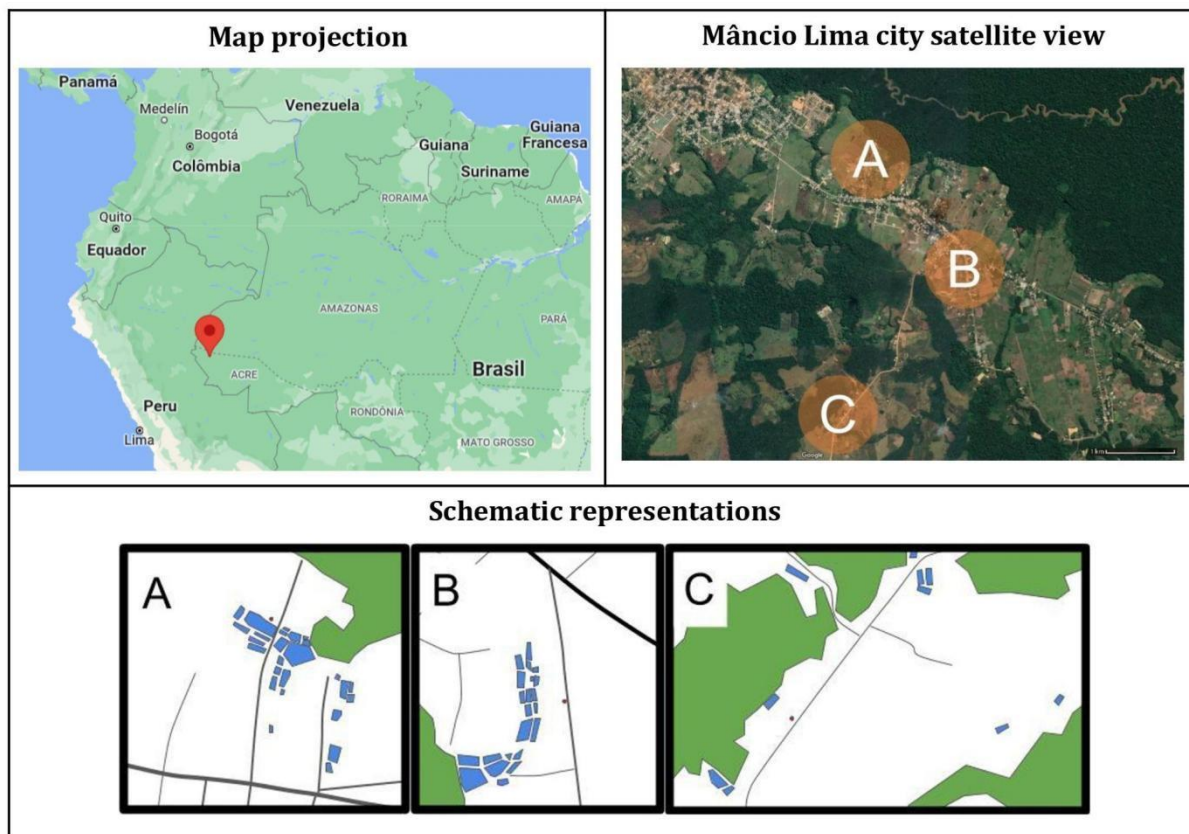


Figura 1. Locais de coleta de *Nyssorhynchus darlingi* na cidade de Mâncio Lima, Estado do Acre, Brasil. As letras representam os seguintes pontos de coleta: (A) 7°37'12.9"S 72°53'06.7"W, (B) 7°38'02.1"S 72°52'26.5"W, (C) 7°39'05.3"S 72°53'20.9"W. A representação esquemática mostra as casas onde os adultos foram capturados (pontos vermelhos), todos os criadouros analisados dentro de aproximadamente 1km de cada casa (azul), áreas florestais (verde) e as principais estradas (linhas pretas).

Preparação e sequenciamento de amostras

Para a extração do DNA, a cabeça e tórax dos mosquitos foram separados do resto do corpo com um bisturi estéril. Cada adulto (cabeça e tórax) e larva (corpo inteiro) foi extraído individualmente usando o Glass Fiber Plate DNA Extraction Kit (Canadian Center for DNA Barcoding) seguindo as recomendações do fabricante. A quantificação do DNA foi realizada por fluorometria usando Thermo Fisher Scientific QuBit dsDNA HS Assay Kit, de acordo com as recomendações do fabricante. As bibliotecas de DNA foram preparadas usando um quinto do volume total recomendado para o kit de preparação Nextera XT Library (Illumina), seguindo as recomendações do fabricante. As amostras de DNA foram multiplexadas para um total de 60 amostras por corrida e sequenciadas na plataforma NextSeq500 (Illumina) por sequenciamento do genoma completo de baixa cobertura (LC-WGS) em uma execução de 151 ciclos tipo *single-end*. A análise da qualidade da sequência foi realizada usando o programa FastQC (Cathcart R, et al. 2005) e as leituras foram usadas se os resultados de todos os módulos de análise fossem aprovados sem erros.

Identificação da espécie

Os dados de sequenciamento foram alinhados com a sequência de referência, subunidade I da citocromo oxidase (COI) de *Ny. darlingi*, (disponível em <https://www.ncbi.nlm.nih.gov/nuccore/KP193458.1/>) usando o software Burrows-Wheeler Aligner (BWA) (Li H, Durbin R. 2009). Após o alinhamento, as sequências individuais de consenso de COI foram geradas usando o pacote de software SamTools (Li H, 2011). A ferramenta BLASTn (Camacho, et al. 2009) foi usada para identificação de várias espécies usando a sequência consenso COI gerada individualmente. Apenas o resultado de correspondência mais alto do Blast foi usado. Os espécimes foram descartados se $e\text{-value} > 1e-100$, identidades < 200 , identidade $< 90\%$ e se a sequência correspondente não foi identificada como *Ny. darlingi*.

Chamada de variante e análise de dados

O genoma de referência de *Ny. darlingi* utilizado está disponível no banco de dados NCBI, número de acesso GCF_943734745.1. Os alinhamentos foram realizados com o software Burrows-Wheeler Aligner (BWA) e a variante chamada com o pacote de software SamTools. Um painel de genótipos foi gerado no formato VCF 4.2. Os SNPs foram removidos do painel de pré-imputação por frequência mínima de alelos (MAF) $< 0,1$ e dados ausentes (MD) $> 0,5$ usando o programa LCVCFTools (Alvarez MVN, 2020). Genótipos com profundidade de sequenciamento (DP) < 5 ou qualidade do genótipo de pontuação de qualidade phred (GQ) < 20 foram imputados com o software BEAGLE 4.1 (Browning BL, Browning SR, 2016) usando valores de probabilidade normalizados de genótipos (PL). Após a imputação, os SNPs foram filtrados por MAF $> 0,1$. Toda a manipulação, análise e plotagem dos dados foi realizada usando o RStudio (linguagem R), incluindo os pacotes fornecidos pelo metapacote "tidyverse" (R Core Team. 2022).

Deteção de inversão cromossômica e genotipagem de amostras

Várias abordagens foram combinadas para detectar, com alto nível de confiabilidade, inversões cromossômicas. Alvarez et al (2022) mostraram que o valor esperado de r^2 na distância de 12,57 kbp é de aproximadamente 0,1 nesta população de *Ny. darlingi*, portanto uma análise de

estimativa de desequilíbrio de ligação de pares de SNP foi realizada para pares dentro de 12 a 13 kbp de distância. A mediana r^2 janelada foi calculada para cada cromossomo em janelas de 0,1% do comprimento do cromossomo. Os picos foram levados em consideração como um indicador de inversão cromossômica. A análise de componentes principais foi realizada usando o software PLINK 1.9 e as estimativas de peso de SNP para cada componente principal foram anotadas. O número de componentes principais considerados foi estimado a partir da variância explicada de cada componente principal, estabelecendo-se um limite excludente quando a variância explicada cumulativa fosse inferior a 5%. A porcentagem de SNPs pertencentes ao 99º percentil baseado nos pesos do PCA foi calculada em janelas de 1% do comprimento do cromossomo, buscando regiões genômicas com alta representatividade para cada componente. Amostras homozigóticas estimadas foram usadas como caso/controle em um teste de alelos (Armitage, P. 1995) e regiões cromossômicas altamente associadas foram anotadas como mais um indicador de inversão cromossômica.

Genotipagem de Inversão Cromossômica

Foi aplicado teste de permutação ($N=1000$) e a distância genética entre as amostras foi estimada usando uma subamostragem de 10% de SNPs em cada iteração. A matriz de distância genética foi estimada usando SNPs altamente representativos (99º percentil das estimativas de pesos de SNPs no PCA) e os clusters genéticos foram estimados usando escala multidimensional com dimensão máxima $k = 1$, esperando um agrupamento de três grupos para os três genótipos possíveis (AA, Aa e aa). As estimativas das probabilidades dos genótipos foram filtradas usando um limiar probabilístico de 99%.

Testes de correlação cromossômica e testes de associação

Todas as inversões cromossômicas candidatas foram submetidas a um teste de correlação par a par de Spearman usando os dados estimados dos genótipos da amostra. As correlações foram consideradas estatisticamente significativas quando $P \leq 0,05$ após o ajuste da taxa de falsas descobertas (FDR; procedimento de Benjamini-Hochberg). A análise de associação foi realizada usando os genótipos das inversões e dados de comportamento de picada e horário de atividade. Aplicou-se o teste Qui-quadrado, ou Fisher para inversões menos frequentes, considerado estatisticamente significativo quando $P_{FDR} \leq 0,05$.

3 RESULTADOS

3.1 amostras

Um total de 394 amostras de Mâncio Lima foram sequenciadas. Após a identificação das espécies e alinhamento do genoma, 73 amostras de Mâncio Lima foram descartadas devido ao limite mínimo de cobertura de sequenciamento, resultado do BLAST de baixa confiança ou não-*darlingi* para identificação de espécies. As 321 amostras utilizadas na análise populacional estão descritas na Tabela 1. O painel de genótipos imputados resultou em 4.241.254 sítios bialélicos, cerca de 32,31 SNPs/Kbp com MAF > 10%

Tabela 1. Amostras de *Nyssorhynchus darlingi* identificadas com BLASTn e COI (e-value < 1e-100) utilizadas no estudo.

Fase	Ponto de Coleta	Local	Contagem
Adulta	A	Interior	12
		Exterior	35
Adulta	B	Interior	7
		Exterior	15
Adulta	C	Interior	40
		Exterior	93
Larval	A	BS 1	9
		BS 2	14
		BS 3	10
		BS 4	15
Larval	B	BS 1	4
		BS 2	10
		BS 3	11
		BS 4	13
Larval	C	BS 1	7
		BS 2	13
		BS 3	8
		BS 4	5

Os pontos de coleta são: As amostras de Mâncio Lima foram coletadas em (A) 7°37'12.9"S 72°53'06.7"W, (B) 7°38'02.1"S 72°52'26.5"W e (C) 7 °39'05.3"S 72°53'20.9"W. Fêmeas adultas foram coletadas no interior ou exterior de cada ponto de coleta. As larvas foram coletadas em torno de quatro diferentes criadouros (BS) dentro e ao redor de cada ponto de coleta.

3.2 Estimativa do desequilíbrio de ligação em pares de SNPs

A análise do desequilíbrio de ligação pareada de SNPs mostrou que, apesar do r^2 mediano na distância de 12 a 13 kbp ser 0,08, 0,06 e 0,12 para os cromossomos 2, 3 e X, respectivamente, picos atingindo quase o dobro do r^2 mediano foram encontrados em diferentes regiões ao longo dos cromossomos (figura 2)

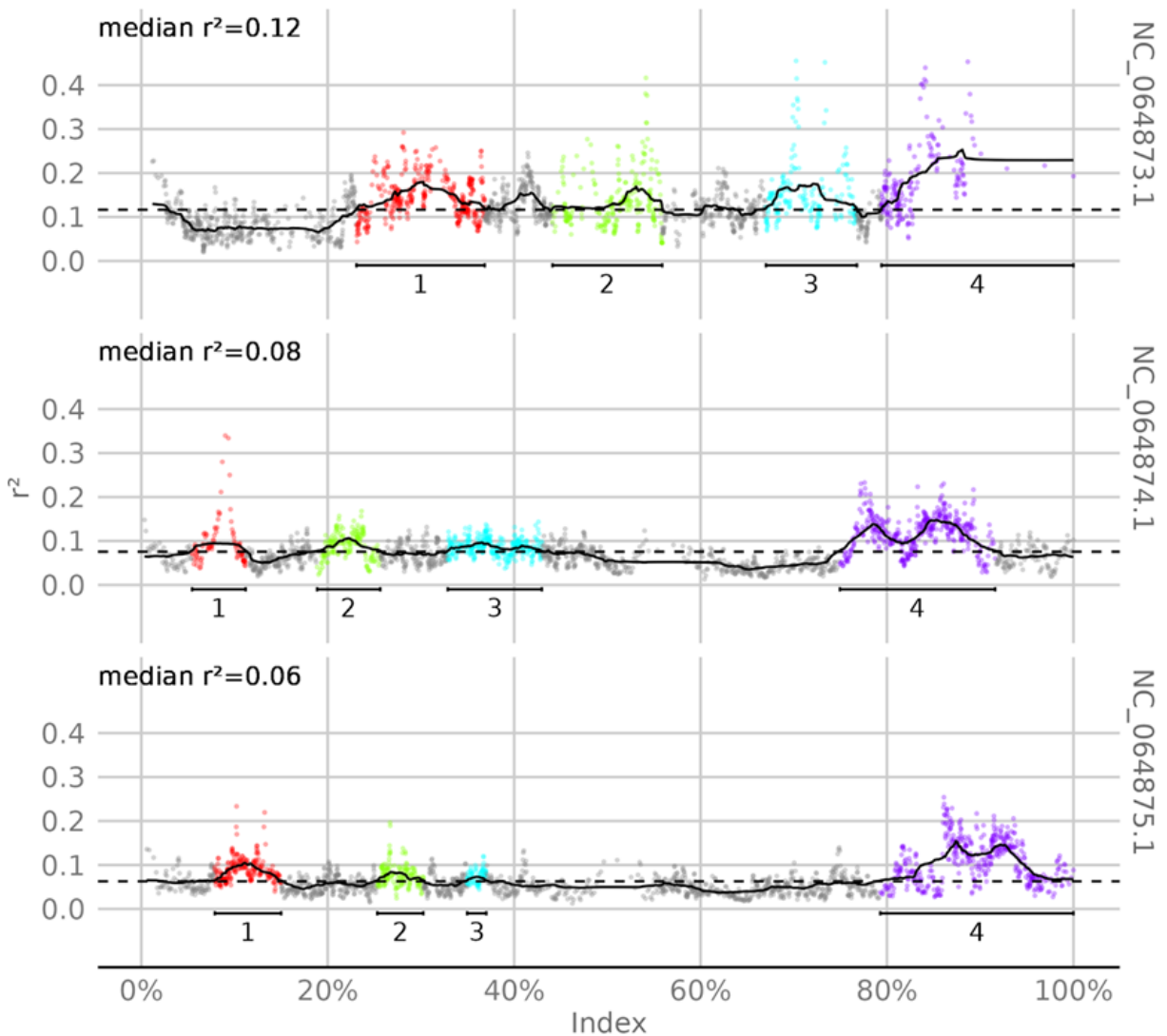


Figura 2. Análise do desequilíbrio de ligação janelado de pares de SNPs. Linha horizontal sólida preta: média polida com base em 0,1% do comprimento do respectivo cromossomo. Linha preta horizontal tracejada: r^2 mediano. Os segmentos horizontais pretos indicam regiões (ao menos 1 Mbp de tamanho) que apresentaram média polida contínua acima da mediana do r^2 e os pontos coloridos representam estimativas janeladas dentro das respectivas regiões. Os pontos cinzas representam cada estimativas de r^2 por janela.

3.3 Análise de componentes principais (PCA)

A análise PCA resultou em 5, 5 e 2 componentes principais para os cromossomos 2, 3 e X, respectivamente (figura 3). O gráfico de concentração de SNPs altamente representativos (99º percentil dos pesos do PCA) indicaram que 4 dos 5 componentes principais do cromossomo 2 mostraram picos em regiões específicas (figura 4 A), assim como para 4 dos 5 componentes principais do cromossomo 3 (figura 4 B) e todos para o cromossomo X (figura 4 C). Cada análise de agrupamento genético com base nessas regiões resultou em três agrupamentos com aproximadamente a mesma distância genética do extremo esquerdo (genótipo AA) ao meio (genótipo Aa) e do meio ao extremo direito (genótipo aa) (apêndice).

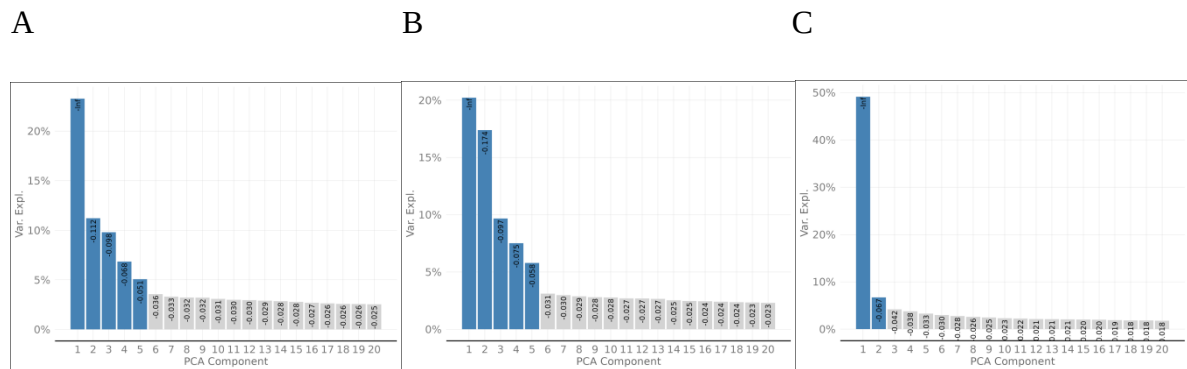
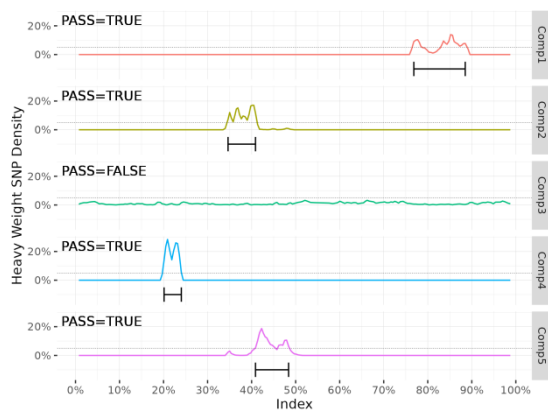
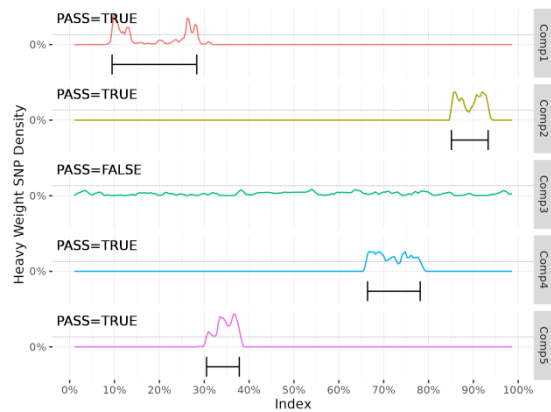


Figura 3. Variância explicada por cada componente principal na análise de PCA. Barras em destaque azul: componentes com mais de 5% da variância explicada. **A)** cromossomo 2; **B)** cromossomo 3; **C)** cromossomo X.

A



B



C



Figura 4. Concentração de SNPs altamente representativos para cada componente da análise de componentes principais (PCA). Os segmentos horizontais pretos indicam regiões cromossômicas com concentrações de SNPs que tiveram um alto peso na variação explicada para cada componente. **A)** cromossomo 2; **B)** cromossomo 3; **C)** cromossomo X.

3.4 Teste de Alelos

O teste de alelos (Figuras 5,6 e 7) revelou que SNPs com frequências estatisticamente diferentes entre homozigotos estimados, estão concentrados em regiões específicas do genoma, ou seja, SNPs diferentemente fixados entre os grupos se encontram em regiões bem delimitadas do genoma, às quais foram consideradas regiões de inversão cromossômica, visto que convergiram nas mesmas regiões de picos de desequilíbrio de ligação (figura 2) e nas regiões de concentração de SNPs altamente representativos (figura 4).

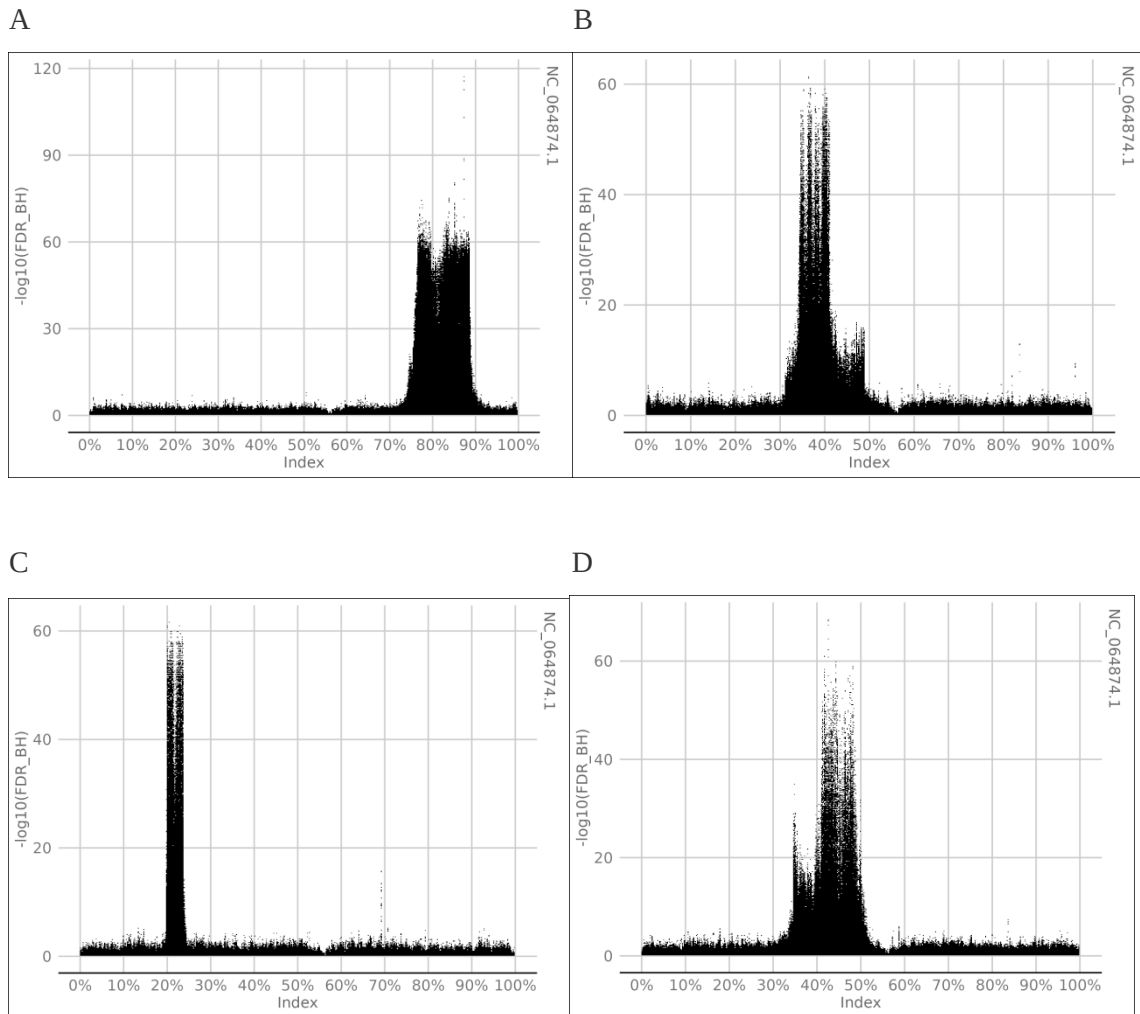


Figura 5: Teste de alelos referente ao cromossomo 2. A) componente 1; B) componente 2; C) componente 4; D) componente 5

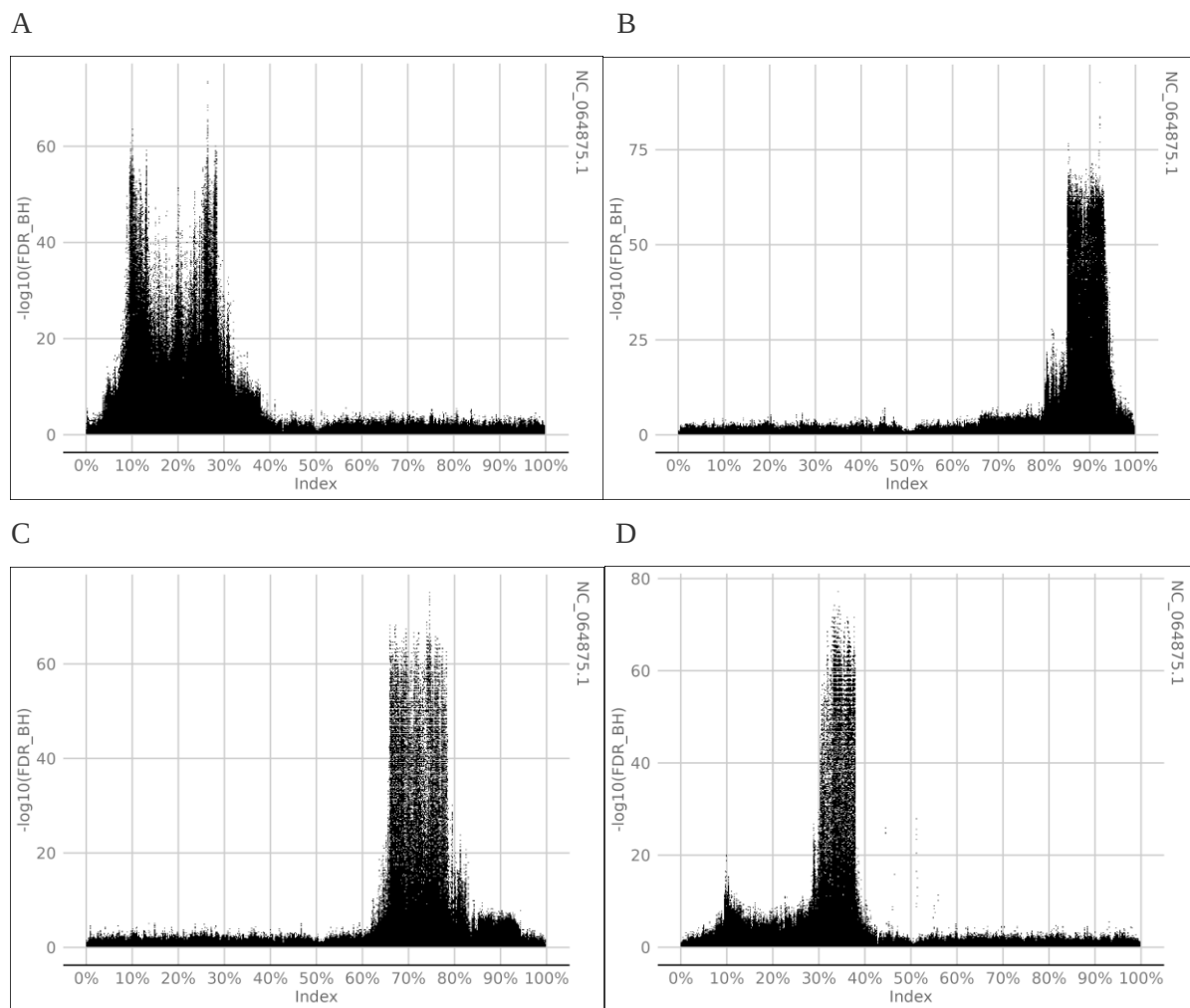


Figura 6: Teste de alelos referente ao cromossomo 3. A) componente 1; B) componente 2; C) componente 4; D) componente 5

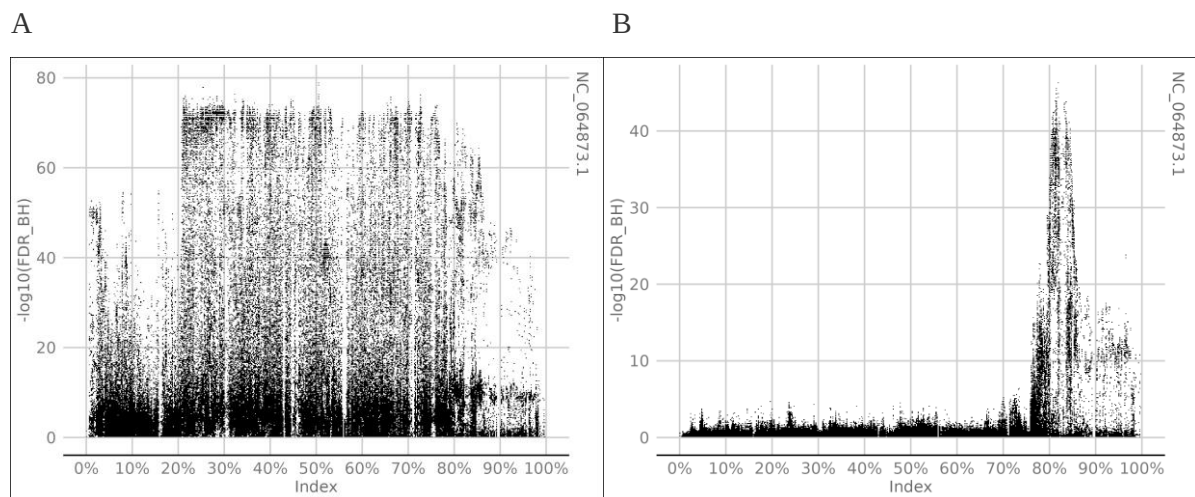


Figura 7: Teste de Alelo referente ao cromossomo X. A) componente 1; B) componente 2.

3.5 Testes de correlação e associação

As análises de correlação através do teste de Spearman mostraram quatro pares de inversões significativamente correlacionadas, como indicado na figura 5. Resultados dos testes de associação com horário de picada (tabela 2) não mostraram associações significativas ($p < 0,06$), por outro lado, encontramos uma inversão associada ao comportamento de picada (tabela 3).

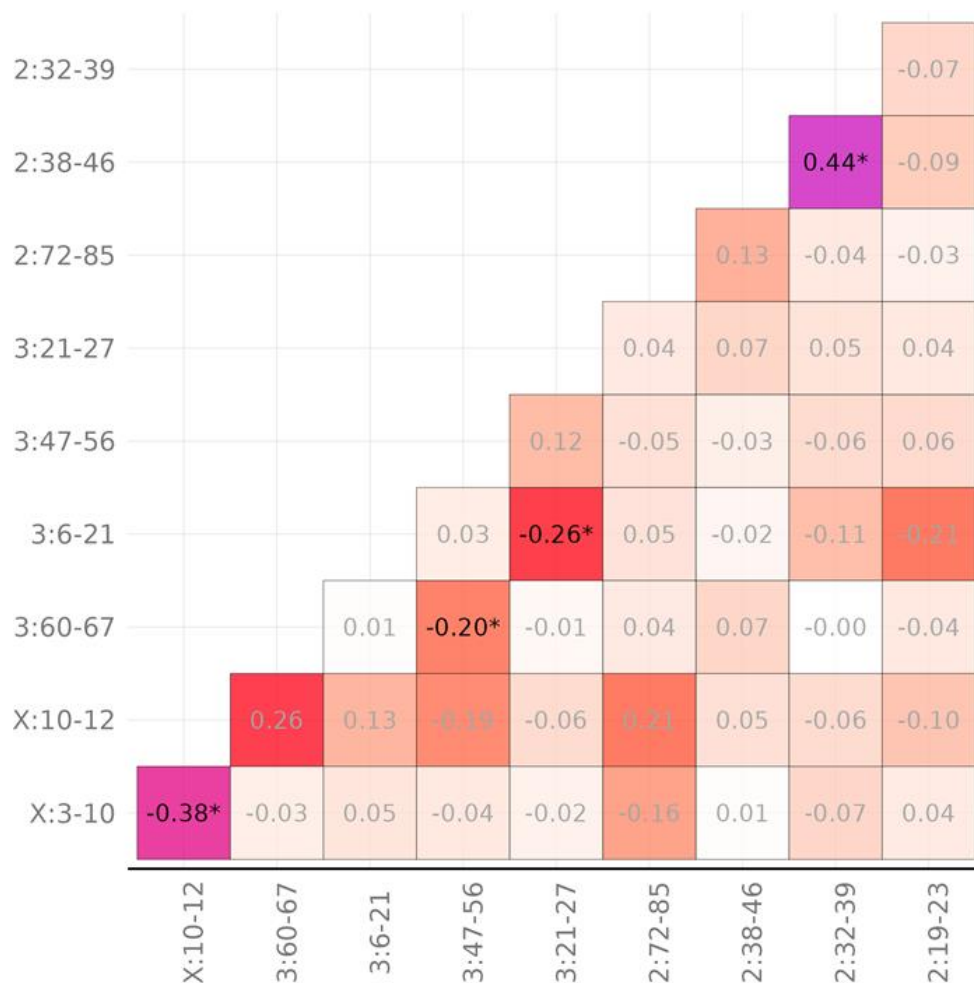


Figura 8. Teste de correlação de Spearman par a par entre genótipos de inversões cromossômicas. * Correlações estatisticamente significativas. Gradiente de cores indica a intensidade da correlação, de fraca (branco), intermediária (vermelho) e forte (roxo).

Tabela 2. Resultados dos testes de associação com horário de picada.

Id	p-valor	Test	p-adj
2:32-39	0,023	Qui-quadrado	0,235
2:19-23	0,063	Qui-quadrado	0,317
3:21-27	0,365	Fisher	0,780
3:47-56	0,452	Fisher	0,780
3:6-21	0,468	Qui-quadrado	0,780
3:60-67	0,308	Qui-quadrado	0,780
2:38-46	0,787	Fisher	0,787
2:72-85	0,557	Fisher	0,787
X:10-12	0,670	Fisher	0,787
X:3-10	0,771	Fisher	0,787

Id: identificação da inversão cromossômica, formatados como cromossomo:início-fim. p-adj: p-valor com correção para múltiplos testes de FDR. Fonte: elaborada pelos autores.

Tabela 3. Resultados dos testes de associação com comportamento de picada.

Id	p-valor	Test	p-adj
* 2:32-39	0,005	Qui-quadrado	0,050
2:38-46	0,022	Fisher	0,107
3:47-56	0,043	Fisher	0,107
3:6-21	0,053	Qui-quadrado	0,107
X:10-12	0,046	Qui-quadrado	0,107
2:72-85	0,207	Qui-quadrado	0,345
X:3-10	0,261	Fisher	0,373
3:60-67	0,421	Qui-quadrado	0,526
2:19-23	0,966	Qui-quadrado	1,000
3:21-27	1,000	Fisher	1,000

Id: identificação da inversão cromossômica, formatados como cromossomo:início-fim. p-adj: p-valor com correção para múltiplos testes de FDR. Fonte: elaborada pelos autores. * Inversão significativamente associada ao comportamento de picada.

4 DISCUSSÃO

Em *Ny. darlingi*, a dificuldade de criação em cativeiro, dentre outros motivos, dificulta o conhecimento da relação entre os polimorfismos de inversão cromossômica com o ambiente, mas sabe-se que esses fatores modificam as interações entre parasitas, vetores e hospedeiros (Tadei et al., 1998, 2016) e tendo em vista crescentes evidências relacionando inversões cromossômicas polimórficas à adaptabilidade, elas podem ser bioindicadores importantes para detecção de respostas genéticas às alterações ambientais (Hoffmann e Danborn 2007 ; Hoffmann e Willi 2008).

O efeito das inversões sobre os níveis de LD ao longo do cromossomo as tornam ótimas ferramentas para investigar a natureza dos processos adaptativos em populações naturais (Hoffman et al., 2004). Nossos resultados das estimativas de LD, análise de PCA e teste de alelos convergiram, mostrando que os picos no padrão de r^2 , a concentração de SNPs com alta porcentagem de poder explicativo no PCA e os SNPs com frequências estatisticamente diferentes entre homozigotos, estão concentrados nas mesmas regiões específicas do genoma, representando fortes indícios de inversões cromossômicas, já que confirmam as teorias genéticas sobre inversões, como seu papel na redução da probabilidade de recombinantes, promovendo o desequilíbrio de ligação (Uno, Fabiana, 2018). Além disso, nossa análise de agrupamento genético IBS com base nos componentes principais resultou em três agrupamentos com aproximadamente a mesma distância genética do extremo esquerdo ao meio e do meio ao extremo direito (Figura X), reforçando a ideia de que se tratam dos 3 possíveis genótipos das inversões (AA, Aa, aa).

Em uma população de *Ny. darlingi*, capturada na estação chuvosa entre Manaus e Boa Vista, no estado do Amazonas, Tadei et al. (1982) avaliaram a variação na frequência de inversões caracterizadas como altamente polimórficas, com até 60% dos indivíduos apresentando de 3 a 5 inversões. A frequência de heterozigotos para as inversões foi substancialmente superior, segundo os autores, inversões em heterozigose encontradas na estação chuvosa na Amazônia conferem ao *Ny. darlingi* maior adaptabilidade ao ambiente, isso acontece pois indivíduos portadores de um cromossomo padrão e um invertido apresentam pouco ou nenhum crossing-over na região do rearranjo (Sturtevant, 1921), já que na meiose, somente regiões homólogas se pareiam durante o zigóteno e a formação do quiasma só ocorre em segmentos pareados (Uno, Fabiana, 2018). Quando determinada população de *Ny. darlingi* do norte do Brasil foi comparada com populações do sudeste, significativas diferenças quanto às frequências de inversões heterozigóticas foram detectadas, sendo superiores no norte graças ao ambiente mais heterogêneo da região amazônica (Bridi e Rafael. 2016 ; Tadei et al. 1982).

Um fator que contribui para essa heterogeneidade é o ciclo hidrológico anual, que corresponde a uma estação seca (julho a outubro) e uma estação chuvosa (novembro a junho) (Tadei et al. 2010, 2016 ; Wolfarth et al. 2013 ; Musset et al. 2014), entre esses períodos, a taxa reprodutiva dos mosquitos varia, aumentando nos períodos de enchentes e diminuindo nos períodos mais secos (Tadei et al. 2016).

Nossos resultados convergem com os de Tadei, et al. 2016, que realizaram um estudo de diversidade genética em populações de *Ny. darlingi* das cidades de Coari e Manaus e revelaram que mosquitos que entram nas casas para se alimentarem (comportamento endofílico) apresentaram maiores taxas de polimorfismos, quando comparados à grupos extradomiciliares, em ambas as regiões. Nosso teste de associação, demonstrou que a inversão 2:32-39 não está igualmente distribuída entre os grupos intra e extradomiciliares de *Ny. darlingi* de Mâncio Lima, estando associada ao comportamento de picada nesta região. Essa maior variabilidade genética pode ser uma resposta adaptativa à pressão seletiva gerada pelo elevado uso de inseticidas nos ambientes internos, o qual também permanece por mais tempo impregnado nas paredes das residências, enquanto no ambiente externo se dispersam rapidamente, corroborando a influência de pressão seletiva na modulação das frequências de inversões cromossômicas, reforçando o papel desses rearranjos na evolução adaptativa.

5 CONCLUSÃO

O LC-WGS é uma abordagem alternativa que permite identificar com alta resolução e relativo baixo custo, regiões de polimorfismos de inversões cromossômicas. Os resultados das estimativas de LD, análise de PCA e teste de alelo convergiram, mostrando que os picos no padrão de r^2 , concentração de SNPs com alta porcentagem de poder explicativo no PCA e os SNPs com frequências estatisticamente diferentes entre homozigotos, estão concentrados nas mesmas regiões específicas do genoma. Assim, com o processamento adequado, representaram uma ferramenta poderosa para detecção e caracterização de inversões cromossômicas, apresentando fortes indícios de 10 inversões no genoma de *Ny.darlingi*, das quais, 4 se mostraram correlacionadas com significância inversamente proporcional à distância entre elas. Além disso, o teste de associação foi eficaz no estudo de processos comportamentais, revelando os achados de que as fêmeas que adentram as casas para se alimentar de sangue podem estar relacionadas à inversão polimórfica 2:32-39 como um processo adaptativo.

Aprovação ética e consentimento para participação

A revisão ética e a aprovação foram dispensadas para este estudo porque apenas os autores treinados profissionalmente conduziram HLC (Human Landing Catch) e o uso deste método para coletar mosquitos anofelinos foi considerado uma questão de gerenciamento de risco. Todas as precauções normais de segurança foram tomadas. Os autores envolvidos no HLC foram plenamente informados sobre os detalhes dos procedimentos, riscos potenciais e plano de mitigação e foram acompanhados por médicos por duas semanas após a coleta.

REFERÊNCIAS BIBLIOGRÁFICAS

Alvarez, MVN, Alonso, DP, Kadri, SM *et al.* Estudos do genoma de *Nyssorhynchus darlingi* relacionados à dispersão microgeográfica e comportamento de busca de sangue. **Parasites Vectors** 15 , 106 (2022). <https://doi.org/10.1186/s13071-022-05219-5>

Alvarez MVN. LCVCFTools v1.0.0-alpha. 2020 [cited 2020 Nov 26]; Available from: <https://zenodo.org/record/4243800>

ARMITAGE, Seth. EVENT STUDY METHODS AND EVIDENCE ON THEIR PERFORMANCE. **Journal Of Economic Surveys**, [S.L.], v. 9, n. 1, p. 25-52, mar. 1995. Wiley. <http://dx.doi.org/10.1111/j.1467-6419.1995.tb00109.x>.

Boletim epidemiológico/Secretaria de vigilância em saúde/Ministério da Saúde/Panorama epidemiológico da malária em 2021: buscando o caminho para a eliminação da malária no Brasil, <https://www.gov.br/saude/pt-br/centrais-deconteudo/publicacoes/boletins/epidemiologicos/edicoes/2022/boletim-epidemiologico-vol-53-no17.pdf>, Maio de 2022, Acessado em 08 de dezembro de 2022

BRIDI, L. C.; RAFAEL, M. S.. GNP domain of *Anopheles darlingi*: are polymorphic inversions and gene variation related to adaptive evolution?. **Genetica**, [S.L.], v. 144, n. 1, p. 99-106, 14 jan. 2016. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10709-016-9881-6>.

BROWNING, Brian L.; BROWNING, Sharon R.. Genotype Imputation with Millions of Reference Samples. **The American Journal Of Human Genetics**, [S.L.], v. 98, n. 1, p. 116-126, jan. 2016. Elsevier BV. <http://dx.doi.org/10.1016/j.ajhg.2015.11.020>.

CAMACHO, Christiam; COULOURIS, George; AVAGYAN, Vahram; MA, Ning; PAPADOPOULOS, Jason; BEALER, Kevin; MADDEN, Thomas L. BLAST+: architecture and applications. **Bmc Bioinformatics**, [S.L.], v. 10, n. 1, p. 253-287, dez. 2009. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1471-2105-10-421>.

CATHCART, Rachael; ROBERTS, Amanda. Evaluating Google Scholar as a Tool for Information Literacy. **Internet Reference Services Quarterly**, [S.L.], v. 10, n. 3-4, p. 167-176, 27 ago. 2005. The Haworth Press. http://dx.doi.org/10.1300/j136v10n03_15.

KRIMBAS, Costas B.; POWELL, Jeffrey R. *Drosophila* inversion polymorphism. **CRC press**, 1992.

CHENG, Changde; TAN, John C.; HAHN, Matthew W.; BESANSKY, Nora J.. Systems genetic analysis of inversion polymorphisms in the malaria mosquito *Anopheles gambiae*. **Proceedings Of The National Academy of Sciences**, [S.L.], v. 115, n. 30, p. 101-129, 9 jul. 2018. <http://dx.doi.org/10.1073/pnas.1806760115>.

COGHLAN, Avril; EICHLER, Evan E.; OLIVER, Stephen G.; PATERSON, Andrew H.; STEIN, Lincoln. Chromosome evolution in eukaryotes: a multi-kingdom perspective. **Trends In Genetics**, [S.L.], v. 21, n. 12, p. 673-682, dez. 2005. Elsevier BV. <http://dx.doi.org/10.1016/j.tig.2005.09.009>.

GIRALDO-CALDERÓN, Gloria I.; EMRICH, Scott J.; MACCALLUM, Robert M.; MASLEN, Gareth; DIALYNAS, Emmanuel; TOPALIS, Pantelis; HO, Nicholas; GESING, Sandra; MADEY, Gregory; COLLINS, Frank H.. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. **Nucleic Acids Research**, [S.L.], v. 43, n. 1, p. 707-713, 15 dez. 2014. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gku1117>.

HIWAT, Hélène; BRETAS, Gustavo. Ecology of Anopheles darlingi Root with respect to vector importance: a review. **Parasites & Vectors**, [S.L.], v. 4, n. 1, p. 134-221, 16 set. 2011. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1756-3305-4-177>.

HOFFMANN, Ary A.; DABORN, Phillip J.. Towards genetic markers in animal populations as biomonitors for human-induced environmental change. **Ecology Letters**, [S.L.], v. 10, n. 1, p. 63-76, 6 nov. 2006. Wiley. <http://dx.doi.org/10.1111/j.1461-0248.2006.00985.x>.

HOFFMANN, Ary A.; WILLI, Yvonne. Detecting genetic responses to environmental change. **Nature Reviews Genetics**, [S.L.], v. 9, n. 6, p. 421-432, jun. 2008. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/nrg2339>.

A HOFFMANN,; SGRO, C; A WEEKS,. Chromosomal inversion polymorphisms and adaptation. **Trends In Ecology & Evolution**, [S.L.], v. 19, n. 9, p. 482-488, set. 2004. Elsevier BV. <http://dx.doi.org/10.1016/j.tree.2004.06.013>.

KIRKPATRICK, Mark; BARTON, Nick. Chromosome Inversions, Local Adaptation and Speciation. **Genetics**, [S.L.], v. 173, n. 1, p. 419-434, 1 maio 2006. Oxford University Press (OUP). <http://dx.doi.org/10.1534/genetics.105.047985>.

KREUTZER RD, Kitzmiller JB, Ferreira E. Inversion polymorphism in the salivary gland chromosomes of Anopheles darlingi root. **Mosquito News**, v. 32, n. 4, p. 555-565, 1972.

LIANG, J.; SHARAKHOVA, M. V.; LAN, Q.; ZHU, H.; SHARAKHOV, I. V.; XIA, A.. A standard cytogenetic map for Anopheles sinensis and chromosome arm homology between the subgenera Anopheles and Cellia. **Medical And Veterinary Entomology**, [S.L.], v. 28, n. 1, p. 26-32, ago. 2014. Wiley. <http://dx.doi.org/10.1111/mve.12048>.

LI, H.. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. **Bioinformatics**, [S.L.], v. 27, n. 21, p. 2987-2993, 8 set. 2011. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btr509>.

LI, H.; DURBIN, R.. Fast and accurate short read alignment with Burrows-Wheeler transform. **Bioinformatics**, [S.L.], v. 25, n. 14, p. 1754-1760, 18 maio 2009. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btp324>.

MUSSET, Lise; PELLEAU, Stéphane; GIROD, Romain; ARDILLON, Vanessa; CARVALHO, Luisiane; DUSFOUR, Isabelle; GOMES, Margarete Sm; DJOSSOU, Félix; LEGRAND, Eric. Malaria on the Guiana Shield: a review of the situation in french guiana. **Memórias do Instituto Oswaldo Cruz**, [S.L.], v. 109, n. 5, p. 525-533, 13 ago. 2014. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/0074-0276140031>.

R Core Team (2022). R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria. URL <https://www.R-project.org/>.

RAFAEL, Míriam S.; ROHDE, Cláudia; TADEI, Wanderli P.; BRIDI, Letícia C.; GAIESKY, Vera Lúcia da Silva Valente. Salivary Polytene Chromosome Map of Anopheles darlingi, the Main Vector of Neotropical Malaria. **The American Journal Of Tropical Medicine And Hygiene**, [S.L.], v. 83, n. 2, p. 241-249, 5 ago. 2010. American Society of Tropical Medicine and Hygiene. <http://dx.doi.org/10.4269/ajtmh.2010.09-0481>.

ROCHA, Elerson Matos; KATAK, Ricardo de Melo; OLIVEIRA, Juan Campos de; ARAUJO, Maisa da Silva; CARLOS, Bianca Cechetto; GALIZI, Roberto; TRIPET, Frederic; MARINOTTI, Osvaldo; SOUZA-NETO, Jayme A.. Vector-Focused Approaches to Curb Malaria Transmission in the Brazilian Amazon: an overview of current and future challenges and strategies. **Tropical Medicine And Infectious Disease**, [S.L.], v. 5, n. 4, p. 161, 20 out. 2020. MDPI AG. <http://dx.doi.org/10.3390/tropicalmed5040161>.

SHARAKHOVA, Maria V; XIA, Ai; LEMAN, Scotland C; SHARAKHOV, Igor V. Arm-specific dynamics of chromosome evolution in malaria mosquitoes. **Bmc Evolutionary Biology**, [S.L.], v. 11, n. 1, p. 327-455, 7 abr. 2011. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1471-2148-11-91>.

STURTEVANT, A. H.. A Case of Rearrangement of Genes in Drosophila. **Proceedings Of The National Academy Of Sciences**, [S.L.], v. 7, n. 8, p. 235-237, ago. 1921. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.7.8.235>.

TADEI, W. P.; RODRIGUES, I. B.; RAFAEL, M. S.; SAMPAIO, R. T. M.; MESQUITA, H. G.; PINHEIRO, V. C. S.; ZEQUI, J. A. C.; ROQUE, R. A.; SANTOS, J. M. M. dos. Adaptative processes, control measures, genetic background, and resilience of malaria vectors and environmental changes in the Amazon region. **Hydrobiologia**, [S.L.], v. 789, n. 1, p. 179-196, 22 set. 2016. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10750-016-2960-y>.

TADEI, Wanderli Pedro; SANTOS, Joselita Maria Mendes dos; RABBANI, Mohammad Ghulan. Biologia de Anofelinos amazônicos. V. Polimorfismo cromossômico de Anopheles darlingi Root (Diptera, Culicidae). **Acta Amazonica**, [S.L.], v. 12, n. 2, p. 353-369, jun. 1982. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/1809-43921982122353>.

TADEI, W P; SCARPASSA, V M; THATCHER, B D; SANTOS, J M; RAFAEL, M s; RODRIGUES, I B. Ecologic observations on anopheline vectors of malaria in the Brazilian Amazon. **The American Journal Of Tropical Medicine And Hygiene**, [S.L.], v. 59, n. 2, p. 325-335, 1 ago. 1998. American Society of Tropical Medicine and Hygiene. <http://dx.doi.org/10.4269/ajtmh.1998.59.325>.

UNO, F. Desequilíbrio de ligação e coadaptação de inversões cromossômicas em Drosophila mediopunctata. **Dissertação** (mestrado em Genética e Biologia Molecular) – Universidade Estadual de Campinas, Instituto de Biologia. Campinas, SP. p. 15. 2018.

WOLFARTH, Bruna Raquel; FILIZOLA, Naziano; TADEI, Wanderli Pedro; DURIEUX, Laurent. Epidemiological analysis of malaria and its relationships with hydrological variables in four municipalities of the State of Amazonas, Brazil. **Hydrological Sciences Journal**, [S.L.], v. 58, n. 7, p. 1495-1504, out. 2013. Informa UK Limited. <http://dx.doi.org/10.1080/02626667.2013.831977>.

World malaria report 2021. **World Health Organization**; <https://www.who.int/publications/i/item/9789240040496>, Acessado em 08 de dezembro de 2022

APÊNDICE

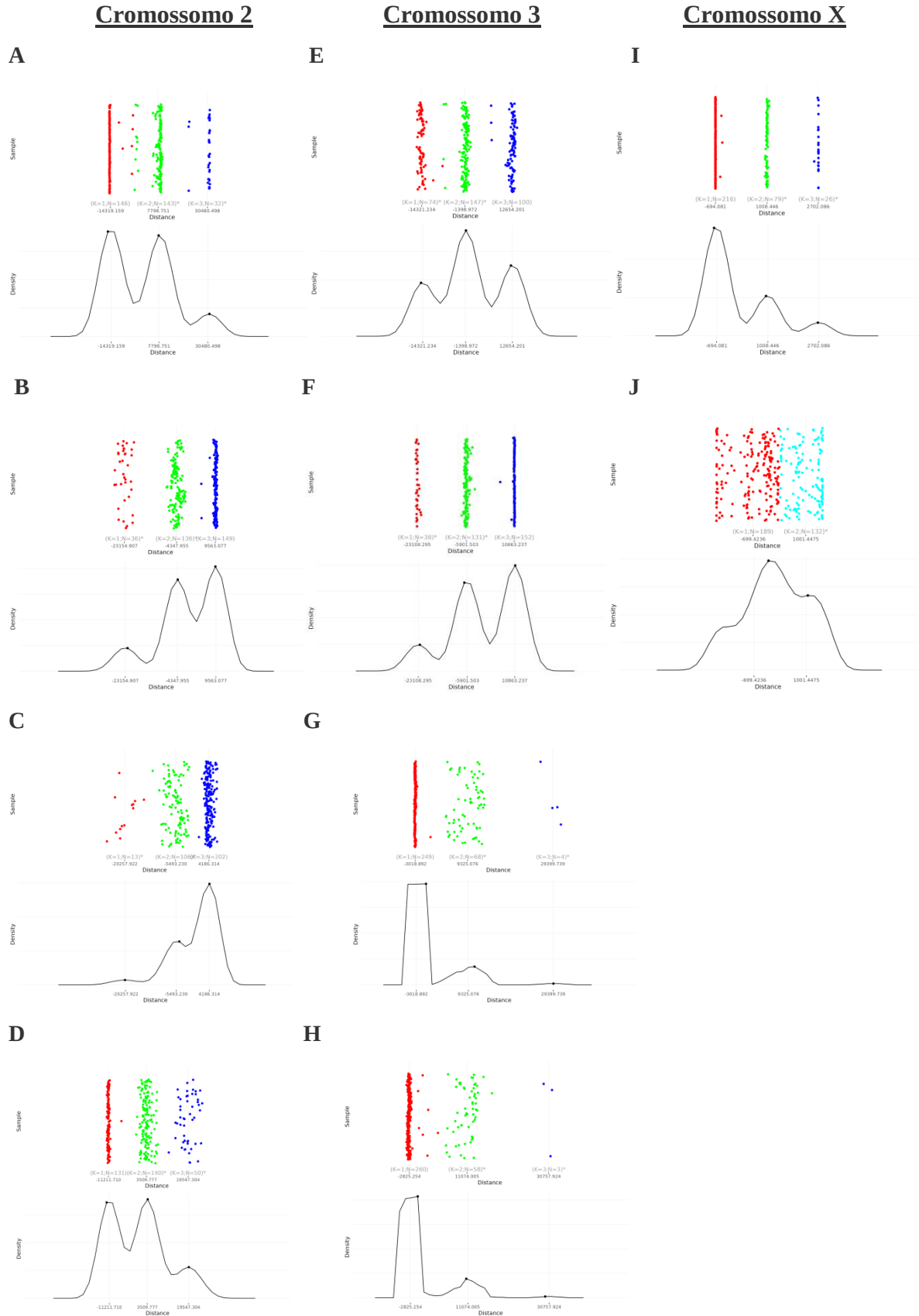


Figura X: análise de agrupamento genético. A,B,C e D: componentes 1,2,4 e 5 do cromossomo 2, respectivamente. E,F,G e H: componentes 1,2,4 e 5 do cromossomo 3, respectivamente. I e J: componentes 1 e 2 do cromossomo X, respectivamente.