

RESEARCH ARTICLE

Minimal genetic differentiation of the malaria vector *Nyssorhynchus darlingi* associated with forest cover level in Amazonian Brazil

Catharine Prussing^{1*}, Kevin J. Emerson², Sara A. Bickersmith³, Maria Anice Mureb Sallum⁴, Jan E. Conn^{1,3*}

1 University at Albany, State University of New York, School of Public Health, Department of Biomedical Sciences, Albany, NY, United States of America, **2** Department of Biology, St. Mary's College of Maryland, St. Mary's City, MD, United States of America, **3** Wadsworth Center, New York State Department of Health, Albany, NY, United States of America, **4** Departamento de Epidemiologia, Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, SP, Brasil

* cprussing@gmail.com (CP); jan.conn@health.ny.gov (JC)



OPEN ACCESS

Citation: Prussing C, Emerson KJ, Bickersmith SA, Sallum MAM, Conn JE (2019) Minimal genetic differentiation of the malaria vector *Nyssorhynchus darlingi* associated with forest cover level in Amazonian Brazil. PLoS ONE 14(11): e0225005. <https://doi.org/10.1371/journal.pone.0225005>

Editor: Nicholas C. Manoukis, United States Department of Agriculture, UNITED STATES

Received: June 11, 2019

Accepted: October 25, 2019

Published: November 14, 2019

Copyright: © 2019 Prussing et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Raw Illumina sequences were deposited in the NCBI Sequence Read Archive (SRA; BioProject ID PRJNA545461). All other relevant data are within the manuscript and its Supporting Information files.

Funding: This study was funded by National Institutes of Health-National Institute of Allergy and Infectious Diseases (NIH-NIAID) R01AI110112 to JEC and MAMS (<https://www.niaid.nih.gov>), Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) Grant No. 2014/26229-7 to MAMS

Abstract

The relationship between deforestation and malaria in Amazonian Brazil is complex, and a deeper understanding of this relationship is required to inform effective control measures in this region. Here, we are particularly interested in characterizing the impact of land use and land cover change on the genetics of the major regional vector of malaria, *Nyssorhynchus darlingi* (Root). We used nextera-tagmented, Reductively Amplified DNA (nextRAD) genotyping-by-sequencing to genotype 164 *Ny. darlingi* collected from 16 collection sites with divergent forest cover levels in seven municipalities in four municipality groups that span the state of Amazonas in northwestern Amazonian Brazil: São Gabriel da Cachoeira, Presidente Figueiredo, four municipalities in the area around Cruzeiro do Sul, and Lábrea. Using a dataset of 5,561 Single Nucleotide Polymorphisms (SNPs), we investigated the genetic structure of these *Ny. darlingi* populations with a combination of model- and non-model-based analyses. We identified weak to moderate genetic differentiation among the four municipality groups. There was no evidence for microgeographic genetic structure of *Ny. darlingi* among forest cover levels within the municipality groups, indicating that there may be gene flow across areas of these municipalities with different degrees of deforestation. Additionally, we conducted an environmental association analysis using two outlier detection methods to determine whether individual SNPs were associated with forest cover level without affecting overall population genetic structure. We identified 14 outlier SNPs, and investigated functions associated with their proximal genes, which could be further characterized in future studies.

Introduction

In the Amazon region, there is an increasing understanding that land use and land cover (LULC) changes caused by agricultural activity, logging, and road construction are modifying

(<http://www.fapesp.br>), and Conselho Nacional de Pesquisa (CNPq) no. 301877/2016-5 to MAMS (<http://cnpq.br>). The Biodefense and Emerging Infectious Disease training fellowship grant T32AI05532901 provided partial support for CP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

the risk of human malaria infection [1–6]. In Amazonian Brazil, malaria rates have been shown to increase during the earliest stages of settlement, when rainforest is first being cleared, humans are settling near the forest fringe, and immunity is low [7, 8]. Though a direct, positive relationship between deforestation and malaria rates in the Amazon has been reported in numerous studies [1, 2, 5, 6], other studies have found the opposite result [4, 9]. These discrepancies could stem from inconsistencies in defining forest cover and deforestation, and highlight the complex relationship between landscape change and malaria risk [4]. Furthermore, it is possible that such discrepancies stem from differences in the effects of landscape changes on the vectorial capacity of genetically distinct mosquito populations [10].

In the Amazon, the major vector of malaria is *Nyssorhynchus darlingi* (Root, also known as *Anopheles darlingi* Root; we have followed the recommendation in [11] to elevate the subgenus *Nyssorhynchus* to a genus). Deforested areas have been associated with an increased *Ny. darlingi* human biting rate [12, 13] and increased larval habitat suitability for *Ny. darlingi* [14–16]. LULC changes have also been associated with changes in mosquito species composition, with *Ny. darlingi* abundance generally highest in human-modified and deforested landscapes [17–20]. Deforestation might additionally impact the population genetic structure of malaria vectors. Genetic differentiation between and within members of *Anopheles gambiae* Giles s.l. has been associated with agricultural activity and the degree of urban/built environment landscapes [21, 22]. Similarly, population structure was detected between *Ny. darlingi* collected from two Brazilian settlements separated by 60 km with very different forest cover levels [23]. The possibility of microgeographic genetic differentiation of *Ny. darlingi* among areas with different forest covers has implications for malaria transmission, as subpopulations of *Ny. darlingi* may differ in vector competence or vectorial capacity, or in their response to vector control activities. Differences in *Plasmodium* infection rates have previously been identified across genetically distinct populations of Anophelinae species [24–26].

Brazil accounts for ~25% of malaria cases reported from the Americas, and has reported 100,000–200,000 cases annually since 2013 [27]. Over 99% of the malaria burden in Brazil is concentrated in the Amazon Basin, particularly in the states of Amazonas and Acre [28, 29]. Though the majority of malaria cases in these states, as in all of Brazil, are caused by *Plasmodium vivax*, these states also have a persistently and disproportionately high proportion of cases caused by *Plasmodium falciparum*, which is associated with higher morbidity and mortality [29]. Within these states, malaria transmission is heterogeneous [30, 31], with pockets of high malaria infection rates in human and mosquito populations [32].

In Brazil, nextera-tagmented Reductively Amplified DNA (nextRAD) genotyping-by-sequencing detected genetic clustering of *Ny. darlingi* into three groups by biogeographical region (southeast, west Atlantic, and Amazon) [33]. These results are consistent with previous studies based on microsatellites [34, 35] and COI sequencing [36] that have shown little genetic differentiation of *Ny. darlingi* across Western and Central Amazonian Brazil. At the microgeographic level, low levels of genetic differentiation have been detected between *Ny. darlingi* collected in different seasons [37] and habitats [13] using microsatellites. Despite evidence that deforestation may impact the population genetic structure of *Ny. darlingi* [23], this question has not been investigated at a broader geographic scale. To address this, we used nextRAD genotyping-by-sequencing to determine whether forest cover level is associated with microgeographic genetic structuring of *Ny. darlingi* in four municipality groups in Amazonas and Acre states in Amazonian Brazil.

Methods

Study site selection and adult *Ny. darlingi* collections

Adult female *Ny. darlingi* were collected in 2015–2017 from 16 sites spread across seven municipalities in Amazonas and Acre States (Fig 1). For the purposes of this study, the geographically proximal municipalities of Cruzeiro do Sul, Rodrigues Alves, Mâncio Lima, and Guajará were grouped together as one municipality group (Cruzeiro do Sul area), and Lábrea, Presidente Figueiredo, and São Gabriel da Cachoeira were each considered a separate municipality group, for a total of four municipality groups. The collection methods for this study are described in detail in [32]. Briefly, houses at least 2.5km apart were selected as collection sites within each municipality. Mosquitoes were collected by human landing catch (HLC) in the peridomestic area, within ~5m of each house. Collections were done from 6pm to 12am during the dry season and wet-dry season transition (between April and November). Mosquitoes were euthanized with ethyl acetate in the field, stored on silica gel, and identified morphologically to species using entomological keys [38].

All necessary permits were obtained for the field collections. Collections were made under permanent permit number 12583–1 from Instituto Chico Mendes de Conservação da Biodiversidade (ICMBio, SISBIO). Specific permission was not required for these locations as permission to collect was granted under the permanent permit. The collection locations were not

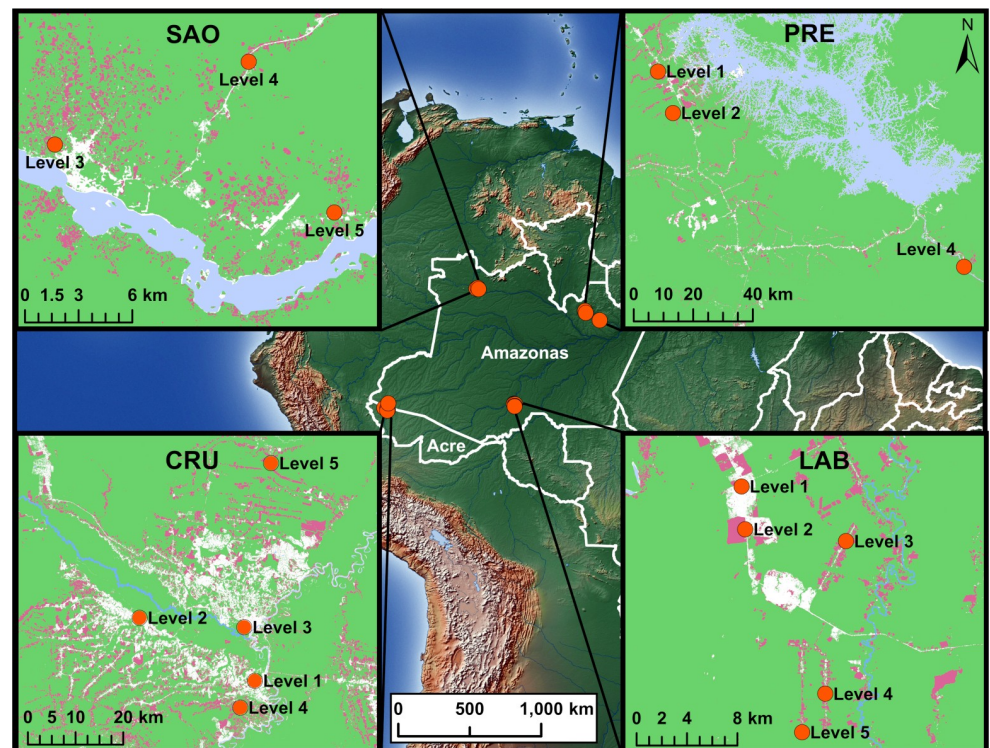


Fig 1. Map of *Ny. darlingi* collection sites in Amazonas and Acre States, Brazil, with insets for each municipality group: SAO: São Gabriel da Cachoeira, PRE: Presidente Figueiredo, CRU: Cruzeiro do Sul area, LAB: Lábrea. Each collection site within each municipality group is indicated by an orange dot, with Levels 1–5 in the municipality insets indicating the forest cover level (Level 1 = ~0–20% forest cover in a 1km radius, Level 2 = ~20–40% forest cover, etc.). Green color on inset map indicates areas of tree canopy cover as of the year 2000, pink color indicates forest loss during the period 2000–2014 (Hansen/UMD/Google/USGS/NASA, <http://data.globalforestwatch.org/>); and blue color indicates bodies of water (ESRI World Water Bodies/World Linear Water layers). Basemap: Natural Earth (<https://www.naturalearthdata.com/>).

<https://doi.org/10.1371/journal.pone.0225005.g001>

Table 1. Forest cover, collection date, and sample size for the 16 collection sites.

Municipality	Forest Cover Level	% Forest Cover in 1km radius	Collection Date	# <i>Ny. darlingi</i> genotyped	# <i>Ny. darlingi</i> in final analysis ²
Rodrigues Alves ¹	1	24%	6/30/2017	15	12
Mâncio Lima ¹	2	43%	5/28/2015	10	9
Cruzeiro do Sul ¹	3	58%	4/19/2015	10	7
Rodrigues Alves ¹	4	70%	6/26/2017	10	9
Guajará ¹	5	83%	7/14/2017	15	14
Lábrea	1	19%	8/9/2015	15	13
Lábrea	2	31%	8/10/2015	10	9
Lábrea	3	59%	8/2/2015	10	10
Lábrea	4	77%	8/5/2015	10	10
Lábrea	5	87%	8/4/2015	15	14
Presidente Figueiredo	1	18%	8/20/2017	15	12
Presidente Figueiredo	2	30%	8/17/2017	10	8
Presidente Figueiredo	4	72%	8/19/2017	10	6
São Gabriel da Cachoeira	3	40%	11/13/2017	10	9
São Gabriel da Cachoeira	4	69%	11/9/2017	10	9
São Gabriel da Cachoeira	5	78%	11/18/2017	15	13
Total				190	164

¹Cruzeiro do Sul area

²Includes only individuals matching to at least 10,000 catalog loci during STACKS analysis.

<https://doi.org/10.1371/journal.pone.0225005.t001>

privately owned or protected, and the field studies did not involve protected or endangered species.

For each collection site, forest cover in a 1 km radius around the site was calculated; this radius was selected to reflect the approximate maximum flight range of *Ny. darlingi* [39]. Forest cover was calculated using European Space Agency (ESA) Sentinel 2A satellite imagery from the closest possible date to the field collection, minimizing cloud coverage. Radiometric and atmospheric corrections were performed using the ESA's Sen2Cor software v2.4.0 [40]. Maximum likelihood supervised classification was used to assign each pixel as forest or non-forest using spectral bands 2, 3, and 4. The percentage forest cover was calculated for each collection site using the formula $\% \text{ Forest Cover} = \frac{\sum_{j=1}^n a_{ij}}{A} (100)$, where a_{ij} corresponds to the forest area and A corresponds to the total area of the landscape.

For each municipality group, collection sites where at least 10 *Ny. darlingi* were collected were selected to cover the maximum possible range of forest cover percentage within each municipality group. Collection sites were split approximately into quintiles by forest cover percentage (level 1 = ~0–20% forest cover, level 2 = ~20–40% forest cover etc.). *Nyssorhynchus darlingi* were available from all 5 forest cover levels for two municipality groups (Lábrea and Cruzeiro do Sul-area), and from 3 levels for the remaining two municipality groups (Presidente Figueiredo and São Gabriel da Cachoeira), for a total of 16 collection sites across the four municipality groups. Genomic DNA was extracted from individual *Ny. darlingi* using the Qiagen DNeasy Blood & Tissue Kit (Hilden, Germany), and DNA concentrations measured using a Qubit Fluorometer (Life Technologies, Carlsbad, CA, USA). For genotyping, 15 *Ny. darlingi* with DNA concentration ≥ 0.5 ng/ μ L were selected from each level 1 and level 5 collection site, and 10 from each level 2, 3, and 4 site, for a total of 190 *Ny. darlingi* (Table 1).

nextRAD genotyping-by-sequencing

Genomic DNA from the 190 individuals was converted into nextRAD genotyping-by-sequencing libraries by SNPsaurus, LLC as in [41]. Genomic DNA was first fragmented with Nextera reagent (Illumina, Inc., San Diego, CA, USA), which also ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled down to fragment 3 ng of genomic DNA (the kit is optimized to fragment 50 ng). Fragmented DNA was then amplified using the Phusion Hot Start Flex DNA Polymerase (New England Biolabs, Inc., Ipswich, MA) for 25 cycles at 75°C, with one of the primers matching the adapter and extending 8 nucleotides into the genomic DNA with the selective sequence TGCAGGAG. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer will be efficiently amplified. The nextRAD libraries were sequenced on a HiSeq 4000 with two lanes of 150 bp reads (University of Oregon). All sequencing reads were uploaded to the NCBI SRA database (BioProject ID: PRJNA545461).

Sequence processing

Raw sequence reads were analyzed using STACKS v2.3b [42, 43]. Briefly, the STACKS pipeline collects raw sequencing reads together into matching stacks, then builds a catalog of putative consensus RAD loci, which span the length of the amplified RAD fragments, by combining stacks from multiple individuals. The pipeline then matches individuals against the catalog of loci, and calls SNPs for each individual at each locus based on a maximum likelihood framework. Finally, the set of loci is filtered based on their frequencies in the study populations. Low-quality reads were dropped using the STACKS *process_radtags* program, and *ustacks* was used to align reads into stacks, with the minimum depth of coverage required to create a stack set to 3, the maximum distance allowed between stacks set to 4, and the maximum distance allowed to align secondary reads to primary stacks set to 6. A catalog of putative loci was built using 24 representative *Ny. darlingi* individuals (S3 File) collected from Brazil and Peru, including 5 individuals from the current study, 4 from [44], 2 from [33], and the remainder from Chu *et al.* (manuscript submitted). All 24 individuals were sequenced, and the reads processed in *ustacks*, using the methods described above. The catalog was built using the STACKS *cstacks* program, allowing 4 mismatches between stacks, with gapped alignments enabled. The catalog loci were mapped to the *Ny. darlingi* genome (AdarC3) [45] using BWA MEM with default parameters [46]; only loci mapping to the genome were retained. After processing in *ustacks*, stacks from the 190 individuals in the current study were searched against the catalog, and SNPs called, with the STACKS *sstacks*, *tsv2bam*, and *gstacks* programs using default settings.

To control for quality and sequence coverage variation among individuals, only individuals that matched to at least 10,000 catalog loci were included in the analysis ($n = 164$, including at least 6 individuals from each municipality group/deforestation level combination; Table 1). The STACKS *populations* program was used to select the first SNP from each RAD locus found in all 4 municipality groups, and in at least 50% of the individuals in each municipality group, with the minimum minor allele frequency set to 0.02 and the maximum observed heterozygosity set to 0.7. A bash script including the STACKS parameters used is included as S4 File, and the STRUCTURE file used for subsequent analyses is included as S5 File.

Population structure analysis

STRUCTURE v2.3.4 [47] was run using the program StrAuto [48] for 10 replicates each of $K = 1$ to 8, with a burn-in of 100,000 generations and an MCMC chain of 1,000,000 generations. The Evanno method [49] implemented in STRUCTURE Harvester [50] was used to

determine the optimal number of genetic clusters. CLUMPP v1.1.2 [51] was run using default settings, and STRUCTURE plots visualized, using the R v3.5.2 [52] package *pophelper* v2.2.7 [53]. As a less computationally intensive method to investigate substructure within municipality groups, fastStructure [54] analysis was run for the full dataset (to confirm that results were comparable to STRUCTURE results) and then for each municipality group separately using default settings for five replicates each of $K = 1$ to 10. The replicate runs were combined using CLUMPP and visualized in *pophelper* as above.

Principal components analysis (PCA) was performed using the R package *ade4* v1.7–13 [55] `dudi.pca()` function, and PCA plots were created using the R package *factoextra* v1.0.5 [56] `fviz_pca_ind()` function. Discriminant Analysis of Principal Components (DAPC) [57] was performed using the R package *adegenet* v2.1.1 [58].

A hierarchical Analysis of Molecular Variance (AMOVA), with individuals grouped into forest cover levels within municipality groups, was calculated using the `poppr.amova()` function in the R package *poppr* v2.8.2 [59]. Pairwise F_{ST} values with confidence intervals using 999 bootstrap samples were calculated using the `stamppFst()` function in the R package *StAMPP* v1.5.1 [60]. Isolation by distance was examined by plotting the geographic distance between each collection site vs. Prevosti's genetic distance (calculated using the `dist.genpop()` function in *adegenet*) and calculating a Mantel test to compare the two distance matrices using the `mantel.randtest()` function in *ade4*.

Outlier analysis

Two methods were used to identify SNPs associated with forest cover. First, latent-factor mixed modelling (LFMM) [61] was run using the R package *LEA* v2.4.0 [62]. In preparation for LFMM, a sparse non-negative matrix factorization (sNMF) [63] analysis completed using *LEA* was run with ten repetitions of $K = 1$ through 10. The optimal number of populations, where the cross-entropy curve was at a minimum, was four. The sNMF results were used to impute missing genotypes using the *LEA* `impute()` function. Five repetitions of LFMM were run on the imputed dataset, with a burn-in of 5000 and 10000 iterations for each, adjusting for four latent factors (as suggested by optimal number of populations from the sNMF analysis), with forest cover percentage as the explanatory variable. The z-scores from the five runs were combined, and the p -values adjusted for multiple testing using the Benjamini-Hochberg procedure as recommended by the authors [61]. SNPs with an adjusted p -value of 0.05 were considered outliers.

As a second method to identify SNPs associated with forest cover, *bayenv2* [64] was run, following conversion of the `vcf` file in PGDSpider v2.1.1.5 [65], with each collection site as a population. Three replicate covariance matrices were computed, using 100,000 iterations. As no significant differences between the three matrices were detected using the `cortest()` function in the R package *psych* v1.8.12 [66], the first matrix was used. *bayenv2* was run using the correlation matrix and the standardized forest cover percentage for each population, for 100,000 iterations, using the `-c` flag to include non-parametric tests. SNPs that were within both the top 5% of Bayes factors and the top 10% of Spearman's ρ values were considered outliers. The final set of outlier SNPs includes only SNPs identified using both LFMM and *bayenv2*.

Gene ontology analysis

Genes in the annotated *Ny. darlingi* genome scaffolds [67] located within 100 kb of each outlier SNP were investigated for gene function. A wide search window was selected because this was intended to be a broad, exploratory investigation with the goal of hypothesis generation. *Drosophila* orthologs of all genes were investigated. In addition, a gene ontology (GO) enrichment

analysis was performed using the R package *topGO* v2.34.0 [68] to determine whether particular GO terms were enriched among these genes compared to the rest of the genome. Separate Fisher tests were calculated for each sub-ontology (BP: biological process, CC: cellular component, MF: molecular function) using the weight01 algorithm and a cut-off *p*-value of 0.01.

Results

An average of 3,191,681 quality-filtered reads were sequenced from each of the 190 individual *Ny. darlingi* (range 359,366–7,845,737). From these, an average of 30,025 stacks per individual (range 259–409,973) made up of 643,458 total reads (range 2,035–6,430,931) matched to the catalog (S2 File). The final dataset includes one biallelic SNP from each of 5,561 loci from 164 individuals meeting the filtering constraints described in the Methods (S5 File). The average sequencing depth across all loci and individuals was 35X (range per locus: 7X–150X, SD 16X; per individual: 8X–97X, SD 19X).

Population structure

STRUCTURE Harvester analysis determined that the optimal number of genetic clusters was two, though the estimated natural log probability of the data ($\ln\Pr(X|K)$) started leveling off at $K = 5$ (Fig A in S1 File). STRUCTURE results for $K = 2$ –6 are shown in Fig 2, with individuals grouped by municipality group and forest cover level. Overall, there is evidence for structure among the four municipality groups, particularly between the Cruzeiro do Sul area and the other three municipalities, but no evidence of substructure among forest cover levels within each municipality group. fastStructure results were consistent with the STRUCTURE results (Fig B Panel A in S1 File), and separate fastStructure analyses for each municipality group confirmed that there was no evidence for microgeographic structure within each municipality group (Fig B Panels B–E in S1 File).

A PCA similarly showed separation among CRU, PRE, and LAB/SAO along the first two dimensions, and between LAB and SAO along the third dimension (Fig C in S1 File). Though the Bayesian Information Criterion (BIC) of the *k*-means clustering algorithm used in preparation for DAPC indicated that the optimal number of clusters was one, the Akaike Information Criterion (AIC) indicated that the optimal number of clusters was three or four (Fig D in S1 File). Setting the number of clusters to four discriminated perfectly among the four municipality groups, while setting the number of clusters to three collapsed SAO, LAB, and one individual from PRE into a single cluster (Fig 3).

By hierarchical AMOVA (Table 2), 95% of the total genetic variation was within or between individuals, 4.5% was between municipality groups, and only 0.5% was among forest cover levels within municipality groups, supporting weak genetic differentiation between the municipality groups and very little differentiation among forest cover levels. Pairwise F_{ST} values between municipality groups ranged from 0.026 (Lábrea vs. São Gabriel) to 0.075 (Presidente Figueiredo vs. Cruzeiro do Sul area) (Table 3). There was evidence for isolation by distance (Mantel $r = 0.84$, $p = 0.001$, Fig E in S1 File).

Outlier and gene ontology analyses

To determine whether there were individual SNPs associated with forest cover percentage, we conducted environmental association analyses using the 5,561 SNP dataset. LFMM detected 67 SNPs associated with forest cover percentage, and bayenv2 detected 139. Of these, 14 SNPs were detected by both methods (Figs F–G in S1 File). In the *Ny. darlingi* genome scaffolds, 150 genes are located within 100 kb of these 14 SNPs. These genes have a variety of functions in numerous structural and cell signaling pathways (S6 File). One (ADAC006142) encodes a

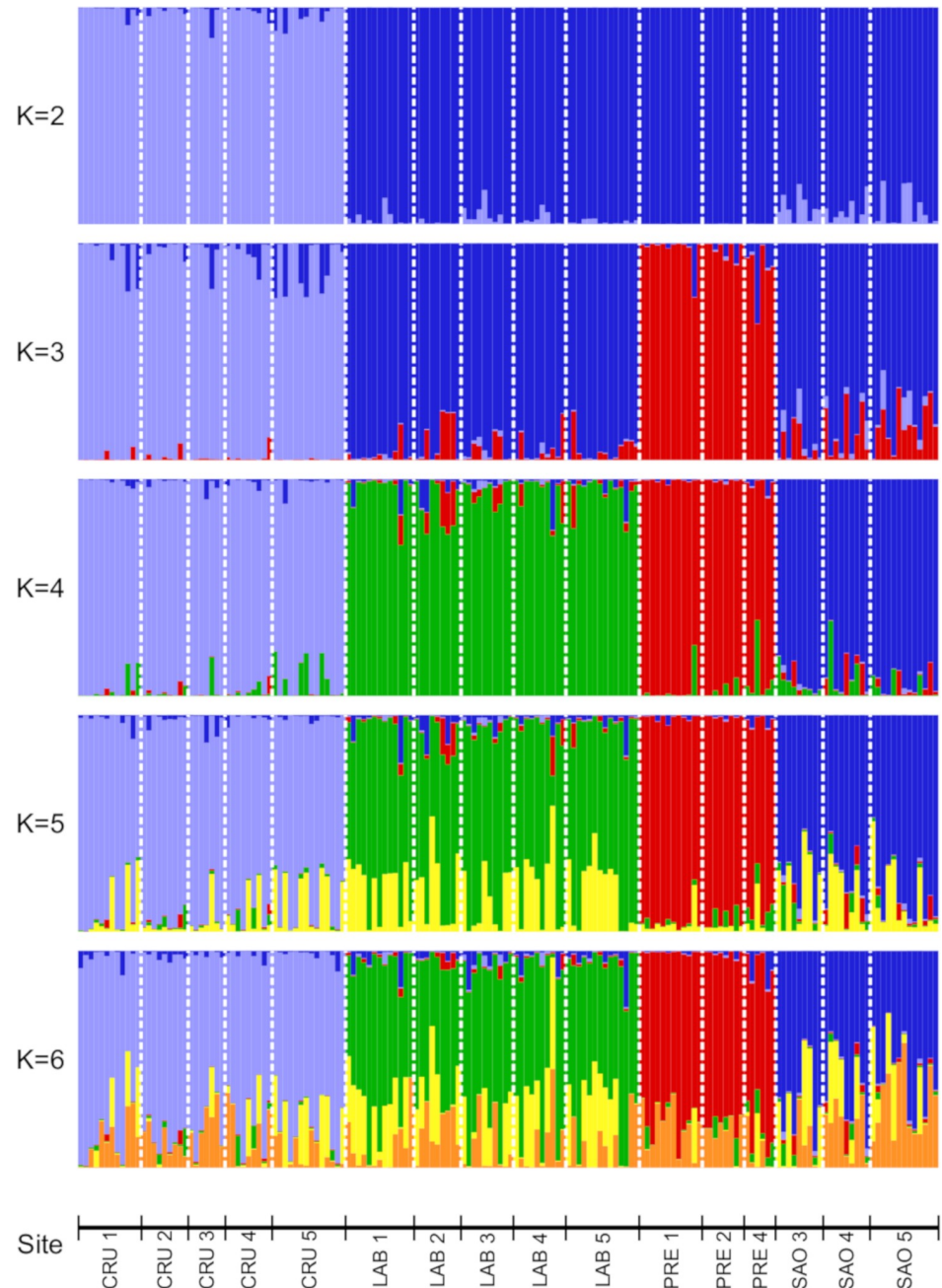


Fig 2. Results of STRUCTURE analysis of the 5,561 SNP dataset, comparing *Ny. darlingi* collected from different municipality groups (CRU: Cruzeiro do Sul area, LAB: Lábrea, PRE: Presidente Figueiredo, SAO: São Gabriel da Cachoeira) and forest cover levels (1–5), depicting $K = 2$ –6 inferred clusters.

<https://doi.org/10.1371/journal.pone.0225005.g002>

venom allergen. The topGO analysis found six GO terms significantly enriched among these 150 genes compared to the rest of the *Ny. darlingi* genome (Table A in [S1 File](#)): GO:0035278 (miRNA mediated inhibition of translation), GO:0018149 (peptide cross-linking), GO:0005576 (extracellular region), GO:0003810 (protein-glutamine gamma-glutamyltransferase activity), GO:0052689 (carboxylic ester hydrolase activity), and GO:0035091 (phosphatidylinositol binding).

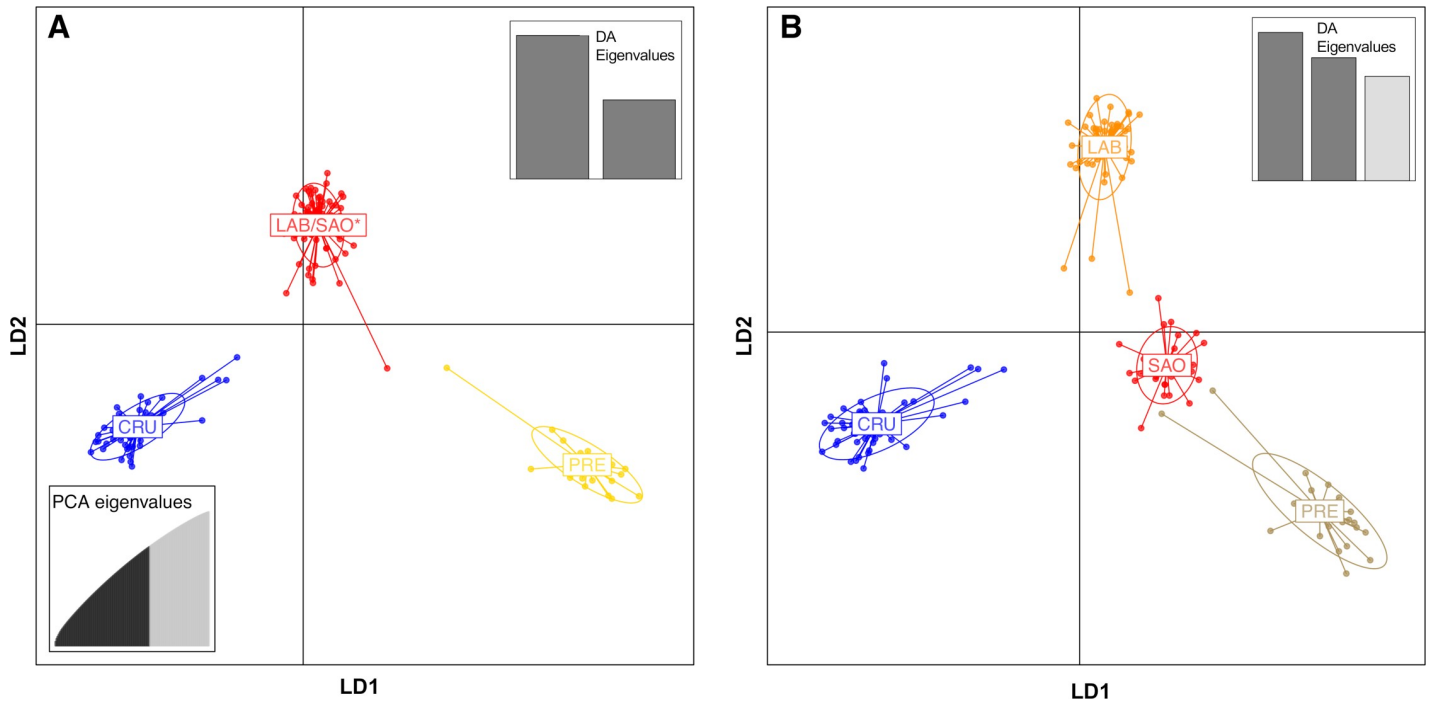


Fig 3. Results of Discriminant Analysis of Principal Components (DAPC) using the 5,561 SNP dataset. Ordination plots for DAPCs of three (A) and four (B) clusters, with insets showing the distribution of eigenvalues for the principal components analysis (PCA) and discriminant analysis (DA). For both analyses, 100 principal components were retained in the PCA, accounting for 73.8% of the variance in the dataset. CRU: Cruzeiro do Sul area, LAB: Lábrea, PRE: Presidente Figueiredo, SAO: São Gabriel da Cachoeira. *SAO/LAB cluster in (A) also includes one individual from PRE.

<https://doi.org/10.1371/journal.pone.0225005.g003>

Discussion

We did not find evidence of microgeographic genetic structure among 164 *Ny. darlingi* collected from multiple forest cover levels within four Amazonian Brazil municipality groups based on model- and non-model based analyses of genotypes at 5,561 SNP loci. This could indicate that there are high levels of gene flow across populations of *Ny. darlingi* in areas with different forest cover levels, consistent with studies in other insects in South America [69, 70] (but see [71]). Our results are not consistent with a previous study in *Ny. darlingi* [23], that found microgeographic structure between two municipalities in Brazil with divergent forest cover levels based on analysis of ~2,000 SNP loci generated using ddRADseq. It is possible that the structure between the two municipalities in the Campos *et al.* study was due to unmeasured differences between the two municipalities not related to forest cover, such as differences in breeding site ecology or vector control activities. The current study more comprehensively

Table 2. Analysis of molecular variance (AMOVA), with individual *Ny. darlingi* nested within forest covers nested within municipality groups.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation	p-value*
Among municipality groups	3	17634.98	54.83	4.54	0.001
Among forest cover levels within municipality groups	12	18259.70	6.54	0.54	0.001
Among individuals within forest cover levels	148	205587.54	242.10	20.03	0.001
Within individuals	164	148406.03	904.91	74.89	0.001
Total	327	389888.25	1208.38	100.00	

*based on 999 Monte Carlo permutation tests

<https://doi.org/10.1371/journal.pone.0225005.t002>

Table 3. Pairwise F_{ST} values between municipality groups (CRU: Cruzeiro do Sul area, LAB: Lábrea, PRE: Presidente Figueiredo, SAO: São Gabriel da Cachoeira) with 95% confidence intervals based on 999 bootstrap samples (below the diagonal), and geographic distances between the centroid of collection sites in each municipality group (above the diagonal).

	CRU	LAB	PRE	SAO
CRU	-	884 km	1558 km	1044 km
LAB	0.042 (0.038, 0.045)	-	820 km	849 km
PRE	0.075 (0.071, 0.079)	0.052 (0.048, 0.055)	-	808 km
SAO	0.043 (0.040, 0.046)	0.026 (0.024, 0.028)	0.055 (0.051, 0.059)	-

<https://doi.org/10.1371/journal.pone.0225005.t003>

investigates the relationship between forest cover and population structure, as it includes intermediate forest cover levels, as well as four replicate municipality groups. It is also possible that the adaptive response to forest cover differences varies among *Ny. darlingi* populations in different locations due to other factors in the external environment.

Anthropogenic changes in forest cover may produce adaptive phenotypic changes in mosquito populations not reflected in genomic studies, particularly among *Ny. darlingi*, as this species has been shown to display a high degree of plasticity in life history traits [72] and biting behavior [44]. Deforestation has been shown to affect the survival and reproductive fitness of *An. gambiae* [73], *An. arabiensis* [74], and *An. minimus* [75]. Additionally, it is possible that whole genome sequencing, or identification of structural variants [76, 77] could identify genetic adaptation to forest cover within *Ny. darlingi* populations not reflected in our SNP dataset.

We identified weak to moderate genetic differentiation ($F_{ST} = 0.026$ – 0.075) among four municipality groups across Amazonian Brazil separated by 800–1,600 km. This is consistent with a previous study using microsatellites that found similar F_{ST} values comparing *Ny. darlingi* collected from localities in central and western Amazonian Brazil separated by comparable geographic distances to the current study [34]. However, it contrasts with previous findings that *Ny. darlingi* from Amazonian Brazil belong to a single genetic population [33, 35, 36]. This discrepancy could be the result of a combination of increased resolution provided by nextRAD genotyping-by-sequencing [23], different geographic scales of analyses of these studies, and the use of different collection sites. It is clear that more research into the population genetic structure of *Ny. darlingi* at both continental and regional scales is needed.

The differentiation that we identified between the municipality groups could be the result of isolation by distance (IBD), isolation by barrier (IBB), isolation by resistance (IBR), or isolation by environment (IBE). It is not possible to differentiate between these possibilities with the sampling scheme of the current study because of the lack of intermediate sampling points between these municipality groups. The overall low level of differentiation between municipality groups, with the highest F_{ST} (0.075) between the two most geographically separated municipality groups (Presidente Figueiredo and Cruzeiro do Sul area), in combination with a significant Mantel test, is suggestive of IBD. Several previous population genetics studies of *Ny. darlingi* using nuclear and mitochondrial markers detected IBD [35, 36, 78, 79], including one study that genotyped *Ny. darlingi* at eight microsatellite loci from municipalities in Amazonian Brazil at a similar geographic scale to the current study [34]. However, other studies of *Ny. darlingi* have not found evidence of IBD [33, 80, 81]. Additionally, there are geographic barriers between the municipality groups in the current study, including rivers, primary forest, and extensive areas of unsuitable habitats. It is possible that the particular pattern of barriers

separating the municipality groups could explain the fact that individuals from SAO and LAB are more similar than other pairs of municipality groups that are separated by a similar distance (Table 3). Future studies could explore the effects of these barriers on genetic divergence between these *Ny. darlingi* populations.

We identified 14 SNPs associated with forest cover percentage using two outlier detection methods. These SNPs were located within 100 kb of 150 genes, among which 6 GO terms were over-represented compared with the rest of the *Ny. darlingi* genome. We acknowledge that this type of analysis is limited both by the highly fragmented nature of the current *Ny. darlingi* genome assembly [45], and by the possibility of false positives [82]. Therefore, we present these results cautiously, with the intention of generating hypotheses for future studies.

Conclusions

Using nextRAD genotyping-by-sequencing, we report weak genetic structure among four municipality groups in the Amazonian Brazil, but a lack of microgeographic structure across forest cover levels within these municipality groups. These results do not preclude an adaptive response of *Ny. darlingi* to deforestation in the Amazon, but indicate that such an adaptive response was not associated with genome-wide differentiation. Additional studies using whole genome sequencing and an improved *Ny. darlingi* genome assembly should be undertaken to further explore this topic.

Supporting information

S1 File. Contains Figures A-G and Table A:

- Fig A. Selection of optimal number of clusters for STRUCTURE analysis using the Evanno method.
- Fig B. Results of fastSTRUCTURE analysis of 5561 SNP dataset.
- Fig C. Principal components analysis of 5561 SNP dataset, with shapes and colors indicating the locality and forest cover level of *Ny. darlingi* collection sites.
- Fig D. Selection of optimal number of clusters for DAPC.
- Fig E. Isolation by distance plot depicting Prevosti's absolute genetic distance vs. geographic distance in km for pairs of 16 collection sites.
- Fig F. Selection of candidate SNPs using LFMM and bayenv2.
- Fig G. Frequency of major allele for each candidate SNP within each collection site, plotted by the forest cover percentage at each site.
- Table A. GO terms significantly enriched (Fisher's $p < 0.01$) within 100kb of candidate SNPs in *Ny. darlingi* genome.

(PDF)

S2 File. Sample information table, including collection information, coordinates, sequencing and STACKS data, and SRA accession number for 190 sequenced *Ny. darlingi*.

(XLSX)

S3 File. Collection details of 24 *Ny. darlingi* used to build STACKS catalog, including collection date, location, and associated publication.

(XLSX)

S4 File. Bash script including all Stacks commands used to build loci and call SNPs from HiSeq reads.

(TXT)

S5 File. Structure file of 5561 SNPs genotyped in 164 individual *Ny. darlingi* used for all analyses.

(STR)

S6 File. Annotation of 150 genes in *Ny. darlingi* genome located within 100 kb of 14 candidate SNPs, including GO terms, OrthoDB identifiers for Diptera level orthology, *D. melanogaster* gene names and FlyBase ID numbers were available.

(CSV)

Acknowledgments

The authors are indebted and thankful to Eduardo Sterlino Bergo, Leonardo Suveges Moreira Chaves, and Gabriel Zorello Laporta for their valuable help during field collections; and to the staff of the municipalities responsible for malaria control for providing useful information that guided field collections in Rodrigues Alves, Cruzeiro do Sul, Mâncio Lima, Guajará, Lábrea, Presidente Figueiredo and São Gabriel da Cachoeira. We greatly appreciate the inhabitants of the rural settlements, and the staff of the Vector Malaria Control of the municipalities, who generously helped in contacting the communities where field collections were conducted.

Author Contributions

Conceptualization: Catharine Prussing, Kevin J. Emerson, Maria Anice Mureb Sallum, Jan E. Conn.

Data curation: Catharine Prussing, Kevin J. Emerson, Sara A. Bickersmith.

Formal analysis: Catharine Prussing, Kevin J. Emerson.

Funding acquisition: Maria Anice Mureb Sallum, Jan E. Conn.

Investigation: Catharine Prussing, Kevin J. Emerson, Sara A. Bickersmith, Maria Anice Mureb Sallum, Jan E. Conn.

Supervision: Maria Anice Mureb Sallum, Jan E. Conn.

Visualization: Catharine Prussing.

Writing – original draft: Catharine Prussing.

Writing – review & editing: Kevin J. Emerson, Sara A. Bickersmith, Maria Anice Mureb Sallum, Jan E. Conn.

References

1. Olson SH, Gangnon R, Silveira GA, Patz JA. Deforestation and malaria in Mancio Lima County, Brazil. *Emerg Infect Dis*. 2010; 16:1108–1115. <https://doi.org/10.3201/eid1607.091785> PMID: 20587182.
2. Stefani A, Dusfour I, Correa AP, Cruz MC, Dessay N, Galardo AK, et al. Land cover, land use and malaria in the Amazon: A systematic literature review of studies using remotely sensed data. *Malar J*. 2013; 12:192. <https://doi.org/10.1186/1475-2875-12-192> PMID: 23758827.
3. Hahn MB, Gangnon RE, Barcellos C, Asner GP, Patz JA. Influence of deforestation, logging, and fire on malaria in the Brazilian Amazon. *PLoS One*. 2014; 9:e85725. <https://doi.org/10.1371/journal.pone.0085725> PMID: 24404206.
4. Tucker Lima JM, Vittor A, Rifai S, Valle D. Does deforestation promote or inhibit malaria transmission in the Amazon? A systematic literature review and critical appraisal of current evidence. *Philos Trans R Soc L B Biol Sci*. 2017; 372. <https://doi.org/10.1098/rstb.2016.0125> PMID: 28438914.
5. Chaves LSM, Conn JE, López RVM, Sallum MAM. Abundance of impacted forest patches less than 5 km² is a key driver of the incidence of malaria in Amazonian Brazil. *Sci Rep*. 2018; 8:7077. <https://doi.org/10.1038/s41598-018-25344-5> PMID: 29728637.

6. Santos AS, Almeida AN. The impact of deforestation on malaria infections in the Brazilian Amazon. *Ecol Econ*. 2018; 154:247–256. <https://doi.org/10.1016/j.ecolecon.2018.08.005>
7. de Castro MC, Monte-Mor RL, Sawyer DO, Singer BH. Malaria risk on the Amazon frontier. *Proc Natl Acad Sci U S A*. 2006; 103:2452–2457. <https://doi.org/10.1073/pnas.0510576103> PMID: 16461902.
8. da Silva-Nunes M, Codeco CT, Malafronte RS, da Silva NS, Juncansen C, Muniz PT, et al. Malaria on the Amazonian frontier: Transmission dynamics, risk factors, spatial distribution, and prospects for control. *Am J Trop Med Hyg*. 2008; 79:624–635. PMID: 18840755.
9. Valle D, Clark J. Conservation efforts may increase malaria burden in the Brazilian Amazon. *PLoS One*. 2013; 8:e57519. <https://doi.org/10.1371/journal.pone.0057519> PMID: 23483912.
10. Afrane YA, Little TJ, Lawson BW, Githeko AK, Yan G. Deforestation and vectorial capacity of *Anopheles gambiae* Giles mosquitoes in malaria transmission, Kenya. *Emerg Infect Dis*. 2008; 14:1533–1538. <https://doi.org/10.3201/eid1410.070781> PMID: 18826815.
11. Foster PG, de Oliveira TMP, Bergo ES, Conn JE, Sant'Ana DC, Nagaki SS, et al. Phylogeny of Anophelinae using mitochondrial protein coding genes. *R Soc Open Sci*. 2017; 4:170758. <https://doi.org/10.1098/rsos.170758> PMID: 29291068.
12. Vittor AY, Gilman RH, Tielsch J, Glass G, Shields T, Lozano WS, et al. The effect of deforestation on the human-biting rate of *Anopheles darlingi*, the primary vector of falciparum malaria in the Peruvian Amazon. *Am J Trop Med Hyg*. 2006; 74:3–11. <https://doi.org/10.4269/ajtmh.2006.74.3> PMID: 16407338.
13. Lainhart W, Bickersmith S, Nadler K, Moreno M, Saavedra M, Chu VM, et al. Evidence for temporal population replacement and the signature of ecological adaptation in a major Neotropical malaria vector in Amazonian Peru. *Malar J*. 2015; 14:375. <https://doi.org/10.1186/s12936-015-0863-4> PMID: 26415942.
14. Vittor AY, Pan W, Gilman RH, Tielsch J, Glass G, Shields T, et al. Linking deforestation to malaria in the Amazon: Characterization of the breeding habitat of the principal malaria vector, *Anopheles darlingi*. *Am J Trop Med Hyg*. 2009; 81:5–12. PMID: 19556558.
15. de Barros FS, Honorio NA, Arruda ME. Temporal and spatial distribution of malaria within an agricultural settlement of the Brazilian Amazon. *J Vector Ecol*. 2011; 36:159–169. <https://doi.org/10.1111/j.1948-7134.2011.00153.x> PMID: 21635654.
16. de Barros FS, Honorio NA. Deforestation and malaria on the Amazon frontier: Larval clustering of *Anopheles darlingi* (Diptera: Culicidae) determines focal distribution of malaria. *Am J Trop Med Hyg*. 2015; 93:939–953. <https://doi.org/10.4269/ajtmh.15-0042> PMID: 26416110.
17. Tadei WP, Thatcher BD, Santos JM, Scarpassa VM, Rodrigues IB, Rafael MS. Ecologic observations on anopheline vectors of malaria in the Brazilian Amazon. *Am J Trop Med Hyg*. 1998; 59:325–335. <https://doi.org/10.4269/ajtmh.1998.59.325> PMID: 9715956.
18. Reinbold-Wasson DD, Sardelis MR, Jones JW, Watts DM, Fernandez R, Carbajal F, et al. Determinants of *Anopheles* seasonal distribution patterns across a forest to periurban gradient near Iquitos, Peru. *Am J Trop Med Hyg*. 2012; 86:459–463. <https://doi.org/10.4269/ajtmh.2012.11-0547> PMID: 22403317.
19. Adde A, Roux E, Mangeas M, Dessay N, Nacher M, Dusfour I, et al. Dynamical mapping of *Anopheles darlingi* densities in a residual malaria transmission area of French Guiana by using remote sensing and meteorological data. *PLoS One*. 2016; 11:e0164685. <https://doi.org/10.1371/journal.pone.0164685> PMID: 27749938.
20. Martins LMO, David MR, Maciel-de-Freitas R, Silva-do-Nascimento TF. Diversity of *Anopheles* mosquitoes from four landscapes in the highest endemic region of malaria transmission in Brazil. *J Vector Ecol*. 2018; 43:235–244. <https://doi.org/10.1111/jvec.12307> PMID: 30408291.
21. Kamdem C, Tene Fossog B, Simard F, Etouna J, Ndo C, Kengne P, et al. Anthropogenic habitat disturbance and ecological divergence between incipient species of the malaria mosquito *Anopheles gambiae*. *PLoS One*. 2012; 7:e39453. <https://doi.org/10.1371/journal.pone.0039453> PMID: 22745756.
22. Caputo B, Nwakanma D, Caputo FP, Jawara M, Oriero EC, Hamid-Adiamoh M, et al. Prominent intra-specific genetic divergence within *Anopheles gambiae* sibling species triggered by habitat discontinuities across a riverine landscape. *Mol Ecol*. 2014; 23:4574–4589. <https://doi.org/10.1111/mec.12866> PMID: 25040079.
23. Campos M, Conn JE, Alonso DP, Vinetz JM, Emerson KJ, Ribolla PE. Microgeographical structure in the major Neotropical malaria vector *Anopheles darlingi* using microsatellites and SNP markers. *Parasit Vectors*. 2017; 10:76. <https://doi.org/10.1186/s13071-017-2014-y> PMID: 28193289.
24. Riehle M, Guelbeogo W, Gnome A, Eiglmeier K, Holm I, Bischoff E, et al. A cryptic subgroup of *Anopheles gambiae* is highly susceptible to human malaria parasites. *Sci*. 2011; 331.

25. Fryxell RTT, Nieman CC, Fofana A, Lee Y, Traoré SF, Cornel AJ, et al. Differential *Plasmodium falciparum* infection of *Anopheles gambiae* s.s. molecular and chromosomal forms in Mali. *Malar J.* 2012; 11:133. <https://doi.org/10.1186/1475-2875-11-133> PMID: 22540973
26. Sanford MR, Cornel AJ, Nieman CC, Dinis J, Marsden CD, Weakley AM, et al. *Plasmodium falciparum* infection rates for some *Anopheles* spp. from Guinea-Bissau, West Africa. *F1000Res.* 2014; 3. <https://doi.org/10.12688/f1000research.5485.2> PMID: 25383188.
27. WHO, Organization WH. World malaria report 2018. Geneva: World Health Organization; 2018.
28. Ferreira MU, Castro MC. Challenges for malaria elimination in Brazil. *Malar J.* 2016; 15:284. <https://doi.org/10.1186/s12936-016-1335-1> PMID: 27206924.
29. Carlos BC, Rona LDP, Christophides GK, Souza-Neto JA. A comprehensive analysis of malaria transmission in Brazil. *Pathog Glob Heal.* 2019; 113:1–13. <https://doi.org/10.1080/20477724.2019.1581463> PMID: 30829565.
30. Canelas T, Castillo-Salgado C, Ribeiro H. Analyzing the local epidemiological profile of malaria transmission in the Brazilian Amazon between 2010 and 2015. *PLoS Curr.* 2018; 10. <https://doi.org/10.1371/currents.outbreaks.8f23fe5f0c2052bfaaa648e6931e4e1a> PMID: 29623243.
31. Conn JE, Sallum M, Correa MM, Grillet ME. Malaria transmission in South America—present status and prospects for elimination. In: Manguin S, Dev V, editors. *Towards Malaria Elimination—A Leap Forward.* Intech Open; 2018. <https://doi.org/10.5772/intechopen.76964>
32. Sallum MAM, Conn JE, Bergo ES, Laporta GZ, Chaves LSM, Bickersmith SA, et al. Vector competence, vectorial capacity of *Nyssorhynchus darlingi* and the basic reproduction number of *Plasmodium vivax* in agricultural settlements in the Amazonian Region of Brazil. *Malar J.* 2019; 18:117. <https://doi.org/10.1186/s12936-019-2753-7> PMID: 30947726.
33. Emerson KJ, Conn JE, Bergo ES, Randel MA, Sallum MAM. Brazilian *Anopheles darlingi* (Diptera: Culicidae) clusters by major biogeographical region. *PLoS One.* 2015; 10:e0130773. <https://doi.org/10.1371/journal.pone.0130773> PMID: 26172559.
34. Scarpassa VM, Conn JE. Population genetic structure of the major malaria vector *Anopheles darlingi* (Diptera: Culicidae) from the Brazilian Amazon, using microsatellite markers. *Mem Inst Oswaldo Cruz.* 2007; 102:319–327. <https://doi.org/10.1590/s0074-02762007005000045> PMID: 17568937.
35. Mirabello L, Vineis JH, Yanoviak SP, Scarpassa VM, Povoá MM, Padilla N, et al. Microsatellite data suggest significant population structure and differentiation within the malaria vector *Anopheles darlingi* in Central and South America. *BMC Ecol.* 2008; 8:3. <https://doi.org/10.1186/1472-6785-8-3> PMID: 18366795.
36. Pedro PM, Sallum MAM. Spatial expansion and population structure of the neotropical malaria vector, *Anopheles darlingi* (Diptera: Culicidae). *Biol J Linn Soc London.* 2009; 97:854–866. <https://doi.org/10.1111/j.1095-8312.2009.01226.x>
37. Angella AF, Salgueiro P, Gil LH, Vicente JL, Pinto J, Ribolla PE. Seasonal genetic partitioning in the neotropical malaria vector, *Anopheles darlingi*. *Malar J.* 2014; 13:203. <https://doi.org/10.1186/1475-2875-13-203> PMID: 24885508.
38. Linthicum KJ. A revision of the *Argyrytarsis* Section of the subgenus *Nyssorhynchus* of *Anopheles* (Diptera: Culicidae). *Mosq Syst.* 1988; 20:99–271.
39. Hiwat H, Bretas G. Ecology of *Anopheles darlingi* Root with respect to vector importance: a review. *Parasit Vectors.* 2011; 4:177. <https://doi.org/10.1186/1756-3305-4-177> PMID: 21923902.
40. Mueller-Wilm U, Devignot O, Pessiot L. Sen2Cor. European Space Agency; 2017. Available from: <https://step.esa.int/main/third-party-plugins-2/sen2cor/>.
41. Russello MA, Waterhouse MD, Etter PD, Johnson EA. From promise to practice: pairing non-invasive sampling with genomics in conservation. *Fonseca D, editor. PeerJ.* 2015; 3:e1106. <https://doi.org/10.7717/peerj.1106> PMID: 26244114.
42. Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. Stacks: building and genotyping loci de novo from short-read sequences. *G3 Genes, Genomes, Genet.* 2011; 1:171–182. <https://doi.org/10.1534/g3.111.000240> PMID: 22384329.
43. Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. Stacks: an analysis tool set for population genomics. *Mol Ecol.* 2013; 22:3124–3140. <https://doi.org/10.1111/mec.12354> PMID: 23701397.
44. Prussing C, Moreno M, Saavedra MP, Bickersmith SA, Gamboa D, Alava F, et al. Decreasing proportion of *Anopheles darlingi* biting outdoors between long-lasting insecticidal net distributions in peri-lquitos, Amazonian Peru. *Malar J.* 2018; 17:86. <https://doi.org/10.1186/s12936-018-2234-4> PMID: 29463241.
45. Marinotti O, Cerqueira GC, De Almeida LGP, Ferro MIT, Da Silva Loreto EL, Zaha A, et al. The genome of *Anopheles darlingi*, the main neotropical malaria vector. *Nucleic Acids Res.* 2013; 41:7387–7400. <https://doi.org/10.1093/nar/gkt484> PMID: 23761445.

46. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009; 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324> PMID: 19451168.
47. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. PMID: 10835412.
48. Chhatre VE, Emerson KJ. StrAuto: Automation and parallelization of STRUCTURE analysis. *BMC Bioinformatics*. 2017; 18:192. <https://doi.org/10.1186/s12859-017-1593-0> PMID: 28340552.
49. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol*. 2005; 14:2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> PMID: 15969739.
50. Earl DA, vonHoldt BM. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour*. 2012; 4:359–361. <https://doi.org/10.1007/s12686-011-9548-7>
51. Jakobsson M, Rosenberg NA. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*. 2007; 23:1801–1806. <https://doi.org/10.1093/bioinformatics/btm233> PMID: 17485429.
52. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. Available from: <https://www.r-project.org/>.
53. Francis RM. pophelper: an R package and web app to analyse and visualize population structure. *Mol Ecol Resour*. 2017; 17:27–32. <https://doi.org/10.1111/1755-0998.12509> PMID: 26850166.
54. Raj A, Stephens M, Pritchard JK. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*. 2014; 197:573–589. <https://doi.org/10.1534/genetics.114.164350> PMID: 24700103.
55. Dray S, Dufour A-B. The ade4 package: Implementing the duality diagram for ecologists. *J Stat Softw*. 2007; 22:20. <https://doi.org/10.18637/jss.v022.i04>
56. Kassambara A, Mundt F. factoextra: Extract and visualize the results of multivariate data analyses. R package. 2017. Available from: <https://cran.r-project.org/package=factoextra>.
57. Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 2010; 11:1–15. <https://doi.org/10.1186/1471-2156-11-1> PMID: 20051104.
58. Jombart T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008; 24:1403–1405. <https://doi.org/10.1093/bioinformatics/btn129> PMID: 18397895.
59. Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. Souza V, editor. *PeerJ*. 2014; 2:e281. <https://doi.org/10.7717/peerj.281> PMID: 24688859.
60. Pembleton LW, Cogan NOI, Forster JW. StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Mol Ecol Resour*. 2013; 13:946–952. <https://doi.org/10.1111/1755-0998.12129> PMID: 23738873.
61. Frichot E, Schoville SD, François O, Bouchard G. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol*. 2013; 30:1687–1699. <https://doi.org/10.1093/molbev/mst063> PMID: 23543094.
62. Frichot E, François O. LEA: An R package for landscape and ecological association studies. *Methods Ecol Evol*. 2015; 6:925–929. <https://doi.org/10.1111/2041-210x.12382>
63. Frichot E, Mathieu F, Trouillon T, Bouchard G, François O. Fast and efficient estimation of individual ancestry coefficients. *Genetics*. 2014; 196:973–983. <https://doi.org/10.1534/genetics.113.160572> PMID: 24496008.
64. Günther T, Coop G. Robust identification of local adaptation from allele frequencies. *Genetics*. 2013; 195:205–220. <https://doi.org/10.1534/genetics.113.152462> PMID: 23821598.
65. Lischer HEL, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 2011; 28:298–299. <https://doi.org/10.1093/bioinformatics/btr642> PMID: 22110245.
66. Revelle W. psych: Procedures for personality and psychological research. R package. Northwestern University, Evanston, Illinois, USA; 2018. Available from: <https://cran.r-project.org/package=psych>.
67. VectorBase. *Anopheles darlingi*, AdarC3.8. 2018. Available from: <https://www.vectorbase.org/organisms/anopheles-darlingi/coari/adarc38>.
68. Alexa A, Rahnenfuhrer J. topGO: Enrichment analysis for gene ontology. R package. 2018.
69. Jaffe R, Pope N, Acosta AL, Alves DA, Arias MC, De la Rúa P, et al. Beekeeping practices and geographic distance, not land use, drive gene flow across tropical bees. *Mol Ecol*. 2016; 25:5345–5358. <https://doi.org/10.1111/mec.13852> PMID: 27662098.

70. Landaverde-González P, Enríquez E, Ariza MA, Murray T, Paxton RJ, Husemann M. Fragmentation in the clouds? The population genetics of the native bee *Partamona bilineata* (Hymenoptera: Apidae: Meliponini) in the cloud forests of Guatemala. *Conserv Genet.* 2017; 18:631–643. <https://doi.org/10.1007/s10592-017-0950-x>
71. Soare TW, Kumar A, Naish KA, O'Donnell S. Genetic evidence for landscape effects on dispersal in the army ant *Eciton burchellii*. *Mol Ecol.* 2014; 23:96–109. <https://doi.org/10.1111/mec.12573> PMID: 24372755.
72. Chu VM, Sallum MAM, Moore TE, Lainhart W, Schlichting CD, Conn JE. Regional variation in life history traits and plastic responses to temperature of the major malaria vector *Nyssorhynchus darlingi* in Brazil. *Sci Rep.* 2019; 9:5356. <https://doi.org/10.1038/s41598-019-41651-x> PMID: 30926833.
73. Afrane YA, Zhou G, Lawson BW, Githeko AK, Yan G. Effects of microclimatic changes caused by deforestation on the survivorship and reproductive fitness of *Anopheles gambiae* in western Kenya highlands. *Am J Trop Med Hyg.* 2006; 74:772–778. PMID: 16687679.
74. Afrane YA, Zhou G, Lawson BW, Githeko AK, Yan G. Life-table analysis of *Anopheles arabiensis* in western Kenya highlands: effects of land covers on larval and adult survivorship. *Am J Trop Med Hyg.* 2007; 77:660–666. PMID: 17978067.
75. Wang X, Zhou G, Zhong D, Wang X, Wang Y, Yang Z, et al. Life-table studies revealed significant effects of deforestation on the development and survivorship of *Anopheles minimus* larvae. *Parasit Vectors.* 2016; 9:323. <https://doi.org/10.1186/s13071-016-1611-5> PMID: 27267223.
76. Ayala D, Zhang S, Chateau M, Fouet C, Morlais I, Costantini C, et al. Association mapping desiccation resistance within chromosomal inversions in the African malaria vector *Anopheles gambiae*. *Mol Ecol.* 2019; 28:1333–1342. <https://doi.org/10.1111/mec.14880> PMID: 30252170.
77. Wellenreuther M, Mérot C, Berdan E, Bernatchez L. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Mol Ecol.* 2019; 28:1203–1209. <https://doi.org/10.1111/mec.15066> PMID: 30834648.
78. Conn JE, Rosa-Freitas MG, Luz SL, Momen H. Molecular population genetics of the primary neotropical malaria vector *Anopheles darlingi* using mtDNA. *J Am Mosq Control Assoc.* 1999; 15:468–474. PMID: 10612610.
79. Conn JE, Vineis JH, Bollback JP, Onyabe DY, Wilkerson RC, Povoas MM. Population structure of the malaria vector *Anopheles darlingi* in a malaria-endemic region of eastern Amazonian Brazil. *Am J Trop Med Hyg.* 2006; 74:798–806. PMID: 16687683.
80. Naranjo-Diaz N, Conn JE, Correa MM. Behavior and population structure of *Anopheles darlingi* in Colombia. *Infect Genet Evol.* 2016; 39:64–73. <https://doi.org/10.1016/j.meegid.2016.01.004> PMID: 26792711.
81. Rosero CY, Jaramillo GI, Gonzalez R, Cardenas H. Genetic differentiation of Colombian populations of *Anopheles darlingi* Root (Diptera: Culicidae). *Neotrop Entomol.* 2017; 46:487–498. <https://doi.org/10.1007/s13744-017-0488-0> PMID: 28229354.
82. Pavlidis P, Jensen JD, Stephan W, Stamatakis A. A critical assessment of storytelling: gene ontology categories and the importance of validating genomic scans. *Mol Biol Evol.* 2012; 29:3237–3248. <https://doi.org/10.1093/molbev/mss136> PMID: 22617950.