



**HAL**  
open science

# Signature of climatic differentiation on mitochondrial DNA of *Drosophila sturtevantii*

Samara Videira Zorzato, Amir Yassin, Lilian Madi-Ravvazzi

► **To cite this version:**

Samara Videira Zorzato, Amir Yassin, Lilian Madi-Ravvazzi. Signature of climatic differentiation on mitochondrial DNA of *Drosophila sturtevantii*. *Mitochondrial DNA Part A*, 2022, 10.1080/24701394.2021.2024814 . hal-03864794

**HAL Id: hal-03864794**

**<https://hal.science/hal-03864794v1>**

Submitted on 22 Nov 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

**Signature of climatic differentiation on mitochondrial DNA of *Drosophila sturtevanti***

**Running title: Phylogeography of *Drosophila sturtevanti***

**SAMARA VIDEIRA ZORZATO<sup>1,2</sup>; AMIR YASSIN<sup>2,3</sup>; LILIAN MADI-RAVAZZI<sup>1</sup>**

<sup>1</sup> São Paulo State University, UNESP IBILCE, Biology, Department São José do Rio Preto, SP, Brazil

<sup>2</sup> Institut de Systématique, Evolution, Biodiversité (ISYEB), Centre National de la Recherche Scientifique, MNHN, EPHE, Sorbonne Université, Univ. des Antilles, 57 rue Cuvier, CP50, 75005 Paris, France.

<sup>3</sup> Laboratoire Évolution, Génomes, Comportement et Écologie, CNRS, IRD, Université Paris-Saclay, 91198 Gif-sur-Yvette, France

**Abstract**

Pleistocene climatic changes have played a major role on the evolution Brazilian Atlantic Forest and South America biodiversity but their impacts on the genetic structure of widely distributed species remain unclear. Here, we investigate mitochondrial DNA (mtDNA) diversity in 21 geographical populations of *Drosophila sturtevanti*. Nucleotide sequences of the cytochrome C oxidase subunits I and II genes (*COI* and *COII*, respectively) from 163 individuals, showed a significant north-south structure, in spite of an overall low level of variation. The haplotypes clustered in three groups that showed strong correlations with geographical and climatic variables, suggesting that local adaptations might have contributed to differentiation within the species. Coalescent-based analyses indicated that the three clusters have differentiated nearly ~17.000 years ago, suggesting a major role for Pleistocene changes in shaping current day distributions and differentiation of widespread Neotropical species.

**Keywords:** adaptation, mtDNA, Neotropics, paleoclimate, phylogeography, *saltans* group

## Introduction

The incredibly high biodiversity found on the Neotropics is, since the 19<sup>th</sup> century, an important topic in evolution and it is still very controversial. Several hypotheses have been proposed to explain it, however the problem remains unsolved (Haffer, 1969; 1993; 1997; Haffer & Prance, 2002; Tuomisto, 2007; Carnaval & Moritz, 2008; Rull, 2011; Carvalho & Almeida, 2016; Antonelli et al., 2018). In this context, the Brazilian Atlantic Forest, located in the Neotropical region, is one of the largest and most diverse rainforests in the world and its history and diversification patterns are complex and reflect on the high diversity and endemism levels generally found (Mittermeier et al., 2005; Myers et al., 2000). Despite all the importance of the Atlantic Forest as one of the 25 terrestrial hotspots of biodiversity in Earth, it is also one of the most threatened biomes, with the habitat loss due to deforestation being the main threat to the Atlantic Forest species (Myers et al., 2000). In this context, phylogeographic studies using endemic species could help us to understand rainforest species ecology and the historic and contemporary patterns that shaped their current distribution and variability, besides to provide us more information about the complex diversification patterns in South America. However, phylogeographic studies are still scarce in southern hemisphere countries, such as Brazil, which makes knowledge of biodiversity and conservation a challenging task (Beheragaray, 2008; Ribeiro et al., 2009). One of the most used molecular markers in phylogeographic approaches is the mitochondrial DNA (mtDNA) due to the fact that it is usually maternally inherited and it was considered a neutral marker for a long time, which allows us to trace its origin over space and time. Nonetheless, growing evidence over the past decades has suggested that mitochondrial genomes were sensitive to natural selection and thermal variation and that the distribution of haplotypes might reflect adaptation to climatic environments instead of just dispersion across populations over time (Lajbner et al., 2018).

For more than a century, *Drosophila* genus flies have been used as a model organism and they enabled great advances in several fields in biology. In spite of all the relevance and high number of studies using *Drosophila* so far, most of them focus only on a few species and groups. This practice, as it is performed, prevents researchers from building knowledge on the biodiversity and variability of natural populations (Hales et al., 2015; O'Grady & De Salle, 2018).

In this study, we analyzed natural populations of *Drosophila sturtevantii* Duda, 1927, a species belonging to one of the two groups that represent the Neotropical clade of the *Sophophora* subgenus (Throckmorton, 1975; O'Grady & DeSalle, 2018). *D. sturtevantii* is the most widespread species in the *sturtevantii* subgroup and even in the *saltans* group, occurring from Central America to South Brazil (Magalhães, 1962; Bicudo, 1979; Madi-Ravazzi et al. 2021). It is widely distributed in many different habitat types and geographical conditions, which makes it a suitable candidate and informative model to study population dynamics and demographic history in South America. We investigated for the first time the population diversification and demographic history of *D. sturtevantii* using Cytochrome Oxidase I and II genes in order to infer ecological and climate associations with the observed distribution patterns. We expect to contribute to elucidating the evolutionary history of the species and to understand its diversification patterns in the Neotropical region.

## **Materials and Methods**

### **Sampling**

We used 163 individuals from 21 localities distributed across South America (Table 1). Seventeen collections were performed in 2015-2018 (collection codes: FAPESP 2014/14059-0, 2016/11994-5), using suspended closed traps made from plastic bottles with bananas (*Musa acuminata* Colla) and fresh yeast (*Saccharomyces cerevisiae*) baits. The traps were left for 3 days on the collecting points and then transported to the laboratory for identification. All flies were initially separated based on morphological features. The females were placed into individual vials, and then the male offspring and the wild-caught males were identified through aedeagus morphology and preserved in absolute ethanol for subsequent analyses. We also used isofemale strains from other four localities maintained in the laboratory on banana medium and yeast (*Saccharomyces cerevisiae*).

### **Molecular Data and Statistical Analysis**

Genomic DNA of adult males from *D. sturtevantii* was extracted using Wizard Genomic DNA Purification kit (Promega). The isolation and amplification of the first half of *COI* (696 bp) were performed via polymerase chain reaction (PCR), using TY-

J-1460 forward and C1-N-2190 reverse (Simon et al., 1994) primers, while the amplification of a portion of *COII* (687 bp) was performed using TL2-J-3037 forward and TK-N-3785 reverse primers (Simon et al., 1994). PCR solution for both genes was performed in a 25  $\mu$ L total volume containing 6.4 ng of DNA, 0.01 pmol of each dNTP, 0.01 pmol of each primer, 5  $\mu$ L of 5x buffer, 2,5Mm of  $MgCl_2$ , and 1 U of Taq DNA polymerase (Promega). All reactions were performed in a Veriti Thermocycler (Applied Biosystems) under the following conditions: 94°C for 2 min, followed by 40 cycles of 94°C for 30 s, (56°C for *COI* and 58°C for *COII*) for 30 s, 72°C for 1 min, then 72°C for 5 min. PCR products were visualized after electrophoresis in 2% w/v agarose gel stained with ethidium bromide and purified using an Illustra GFX kit (GE Healthcare) following the manufacturer's instructions. Sanger sequencing was performed using BigDye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer) in an automated sequencer ABI 3739x/ DNA Analyzer (FisherThermo) with the same primers and conditions used for PCR amplification. The forward and reverse sequences were aligned using ClustalW2 (Larkin et al., 2007) and edited in BioEdit v7.1.9 (Hall, 1999). Sequences were deposited in GenBank under accession numbers MZ323470-MZ323633.

### **Population genetics analysis**

Due to the fact that *COI* and *COII* are both mitochondrial genes, thus sharing the same inheritance mechanism without recombination, and the results we found are similar and concordant, all analyses were performed using the concatenated sequences.

Standard indices of nucleotide diversity ( $\pi$ ), haplotype diversity (Hd), the number of polymorphic sites (S) and the number of haplotypes (h) were computed using DnaSP 5.10 (Librado and Rozas, 2009). Neutrality test statistics were used to examine the evolutionary forces that potentially have acted on the mtDNA loci, such as Tajima's D (Tajima, 1989), Fu's  $F_s$  (Fu, 1997) and  $R_2$  (Ramos-Onsins & Rozas, 2002). JModelTest 2.0 (Darriba et al. 2012) was used to determine the nucleotide substitution model using the Akaike (AIC) and Bayesian information (BIC) criteria, and these models were assumed in further statistics analysis.

A haplotype network was generated using HaploViewer (Salzburger et al., 2011) (available in <http://www.cibiv.at/~greg/haploviewer>) assuming as input a Bayesian Inference tree in Newick format generated in BEAST v1.8 (Drummond & Rambaut, 2007; Drummond et al., 2012) using a HKY+I model, calculated by JModelTest 2.0. The strain of *D. lehrmanae* Madi-Ravazzi et al. from French Guiana was utilized as the outgroup (Madi-Ravazzi et al. 2021). We performed a run with 50 million generations, sampling every 2000 steps using a Birth-Death tree prior. We checked for stationary posterior distributions, effective sample size (ESS) above 200, and convergence between runs by examining parameter traces using Tracer 1.7 (Rambaut et al., 2018). Then, a consensus maximum clade credibility tree was generated with TreeAnnotator 1.8 (Drummond & Rambaut, 2007; Drummond et al., 2012) and the Newick tree file was created using FigTree v1.4.2 (Drummond & Rambaut, 2007; Drummond et al., 2012).

### **Reconstructing of the demographic history**

We first estimated population structure using Bayesian inference with the concatenated *COI* and *COII* data using BAPS6 (Corander et al., 2008; Corander et al., 2013). A population structure model with admixture was used with K values (number of genetic clusters) ranging from 1 to 10, 200 interactions and 10 repetitions. After this, we performed an Analysis of Molecular Variance (AMOVA) using different hierarchical groups, based on geographic, climatic, and phylogenetic criteria, in Arlequin 3.5.2.2 (Excoffier & Lischer, 2010). IMA2 (Hey, 2010; Hey & Nielsen, 2007) was used to estimate the divergence time among the population groups inferred from the clustering analyses, using an isolation-with-migration modeling framework.

### **Isolation-by-ecology analysis**

DnaSP was used to calculate  $F_{ST}$  in population pairs ( $n=157$ ) and a neighbor-joining (NJ) tree was inferred from the  $F_{ST}$  pairwise distances (hereafter the genetic distance matrix). In order to evaluate the relative effects of geographic and ecological differences on the levels of genetic differentiation, we followed the isolation-by-ecology analysis (IBE) (Shafer & Wolf, 2013) wherein the correlation between  $F_{ST}$

based genetic matrix and matrices summarizing the geographical and ecological distances is simultaneously estimated through the partial equation:

$$r_{G|D,E} = \frac{r_{G,D} - r_{G,E}r_{D,E}}{\sqrt{1 - r_{D,E}^2}}$$

Wherein G, D and E refer to the genetic, geographical and environmental distance matrices, respectively, and  $r$  is the correlation coefficient. While the simple equation  $r_{G,D}$  directly measures the level of isolation-by-ecology, the use of the partial equation controls for the eco-spatial autocorrelation by controlling for the geographical distance, D (Shafer & Wolf, 2013).

The geographical distance matrix D was obtained from DIVA-GIS (Hijmans et al. 2012), using the linear distances from the decimal coordinates. For the ecological distance matrix, E, we downloaded 13 climatic variables (see Appendix S2) for the 17 geographical localities from the NASA POWER Project Data Sets (<https://power.larc.nasa.gov>) as in Bastide et al. (2014). We then used PAST (Hammer & Harper, 2006) to run a principal components analysis (PCA) and to convert the climatic variables into a distance matrix using the Euclidean method, wherein the distance ( $d$ ) between populations  $j$  and  $k$  with a value  $x$  for  $i$  climatic variables is estimated as:

$$d_{jk} = \sqrt{\sum_i (x_{ji} - x_{ki})^2}$$

Coefficients of correlation between G, D and E, and the significance of the IBE coefficient were estimated using a customized Perl script (see Appendix S1), following three matrix permutations.

## Results

### Low overall mtDNA diversity in *D. sturtevantii*

A total of 1383bp was obtained for  $n=163$  individuals of *D. sturtevantii*, resulting from the concatenated sequence data of *COI* (696 bp) and *COII* (687 bp) genes, corresponding to positions 1483-2178 (*COI*) and 3083-3769 (*COII*) of the *D. yakuba* reference sequence (GenBank accession NC\_001322). Overall, we observed 47 polymorphic sites, resulting in 48 haplotypes, distributed across three haplogroups



(Fig. 1). Haplotype diversity was high ( $Hd = 0.864$ ), while nucleotide diversity was low ( $\pi = 0.00138$ ).

**Table 1.** Population information and population genetics statistics using the *COI* and *COII* concatenated sequence data.

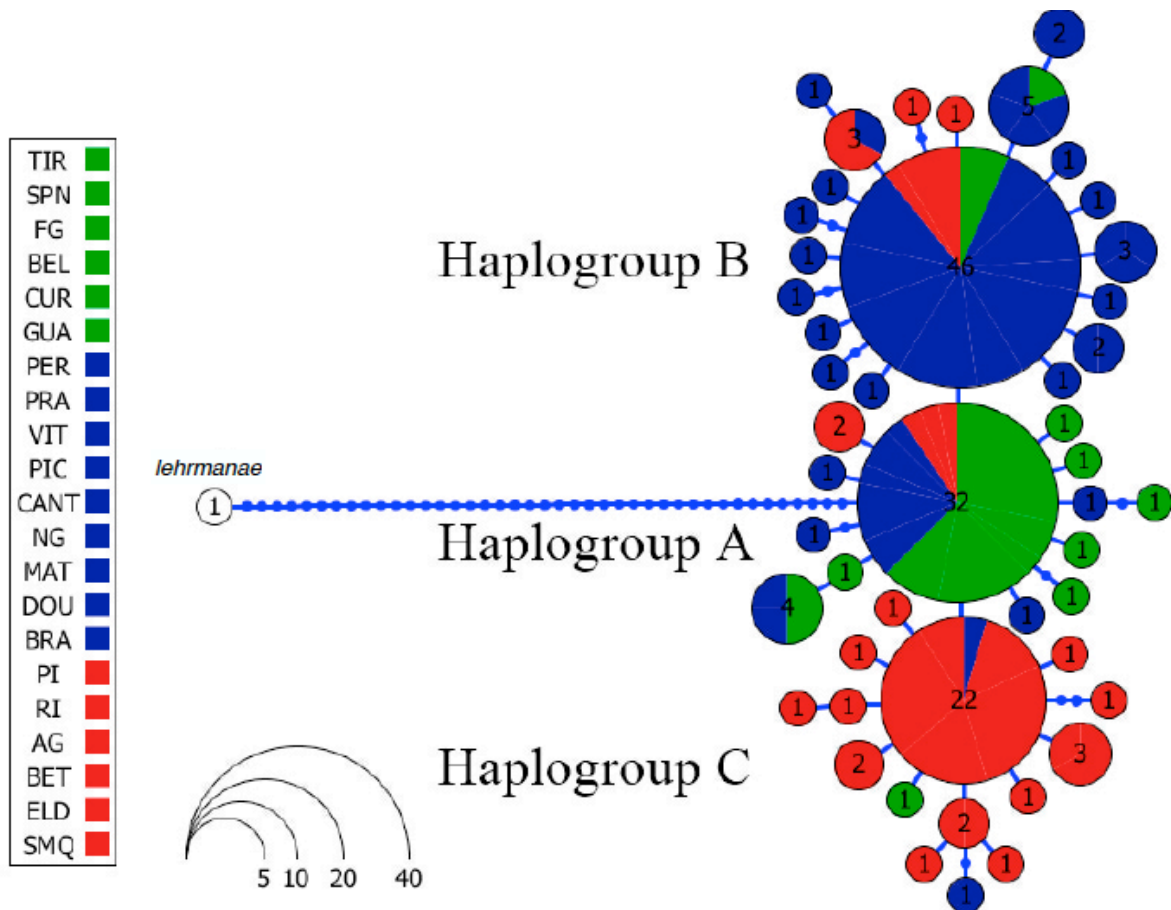
Location	Code	<i>N</i>	<i>h</i>	<i>S</i>	<i>Hd</i>	$\pi$	Fu's <i>F<sub>s</sub></i>	Tajima's <i>D</i>	<i>R<sub>2</sub></i>
Tirimbina-Costa Rica	TIR	10	2	1	0.200	0.00014	-0.339	-1.11173	0.3000
Gamboa-Panama	SPN	1	1	---	---	---	---	---	---
Nouragues-French Guyana	FG	2	1	0	0	---	---	---	---
Belém-PA	BEL	2	2	2	1.000	0.00145	---	---	---
Floresta do Curió/Fortaleza-CE	CUR	8	4	6	0.643	0.00108	-0.422	-1.63982	0.1816
Reserva Ecológica Guaribas/ Mamanguape-PB	GUA	10	5	3	0.844	0.00106	-1.320	1.37817	0.2444
Pernambuco-PE	PER	8	6	6	0.893	0.00150	<b>-2.534</b>	-0.48874	<b>0.1221</b>
APA do Pratigi/Igrapiúna-BA	PRA	10	5	6	0.756	0.00106	-1.320	-1.27616	0.1399
Parque Estadual da Fonte Grande/Vitória-ES	VIT	8	5	5	0.857	0.00111	0.967	-0.92337	0.1512
Parque Estadual da Serra do Mar/Picinguaba-SP	PIC	10	5	4	0.667	0.00069	<b>-2.377</b>	-1.24468	<b>0.1229</b>
Parque Estadual da Cantareira/SP	CANT	10	6	5	0.889	0.00122	<b>-2.290</b>	-0.17819	0.1517
Reserva Ecológica do IBGE/Brasília- DF	BRA	6	2	1	0.333	0.00024	-0.003	-0.93302	0.3727
Fazenda São João/Nova Granada-SP	NG	10	6	6	0.778	0.00087	<b>-3.294</b>	<b>-1.79631</b>	<b>0.1106</b>
Fazenda Cambuhy/Matão-SP	MAT	10	6	9	0.778	0.00141	-1.892	-1.68719	0.1402
Dourados-MS	DOU	7	4	4	0.714	0.00083	-1,217	-1.43414	0.1821
Reserva Betary/Iporanga-SP	BET	10	3	3	0.622	0.00077	0,603	0.02107	0.1778
Pirai/Joinville-SC	PI	10	7	6	0.911	0.00153	<b>-3.099</b>	-0.01912	0,1574
Ribeirão da Ilha/Florianópolis-SC	RI	10	4	5	0.644	0.00084	-0.575	-1.38818	0.1867
Aguaí Santuário Ecológico/Siderópolis-SC	AG	10	7	6	0.867	0.00098	<b>-4.644</b>	-1.49289	<b>0.0884</b>
Eldorado do Sul/RS	ELD	10	6	6	0.844	0.00137	-1.985	-0.45258	0.1430
Maquiné-RS	SMQ	1	---	---	---	---	---	---	---
North	Eco	59	18	22	0.774	0.00109	<b>-13.686</b>	<b>-2.14521</b>	<b>0.0364</b>
Central	PCA	53	18	17	0.717	0.00093	<b>-16.357</b>	<b>-2.02929</b>	<b>0.0374</b>
South	PCA	51	19	20	0.820	0.00128	<b>-14.504</b>	<b>-1.91256</b>	<b>0.0431</b>
North		33	11	13	0.633	0.00084	<b>-7.186</b>	<b>-2.05883</b>	<b>0.0592</b>
Central West	BAPS	33	13	14	0.672	0.00090	<b>-10.132</b>	<b>-2.09443</b>	<b>0.0478</b>
Central East	BAPS	36	14	15	0.835	0.00119	<b>-9.008</b>	-1.77686*	<b>0.0511</b>
South	BAPS	51	19	20	0.820	0.00128	<b>-14.504</b>	<b>-1.91256</b>	<b>0.0431</b>
<b>Total</b>		<b>163</b>	<b>48</b>	<b>46</b>	<b>0.864</b>	<b>0.00136</b>	<b>-62.095</b>	<b>-2.32760</b>	<b>0.0193</b>

Code, laboratory code; *N*, number of individuals; *h*, number of haplotypes; *S*, number of variable sites; *Hd*, haplotype diversity;  $\pi$ , nucleotide diversity. Values in bold type are statistically significant ( $p < 0.05$ ). \*Not significant,  $0.10 > p > 0.05$ .

### Mitochondrial haplotypes cluster within three major haplogroups

The 48 mitochondrial haplotypes clustered in three major haplogroups (Figure 1). Haplogroup A (green, Fig. 1) seems to represent the ancestral sequence, as it is nested in the center of the network. It is connected to *D. lehrmanae* from French Guiana through 42 mutations, and separated from each of the two other haplogroups by a single mutation, *i.e.* T1827C (*COI*) and A3256G (*COII*) from haplogroups B and C, respectively.

**Figure 1** Haplotype network of concatenated data to *D. sturtevantii* populations. Each line of the network represents a single mutation step. Circles size is proportional to the number of individuals. Hypothetical haplotypes are represented by small light blue circles. Populations are abbreviated as in Table 1.



**The three haplogroups show a significant geographical structure**

Frequencies of the three haplogroups differed among populations in a strict geographical pattern. Northern populations had predominantly haplogroup A, whereas Central and Southern populations had high frequencies of haplogroups B and C, respectively (Figure 2).

The degree of genetic variation also differed within each of the 17 populations (Table 1). Neutrality tests (Table 1) yielded significant results concordant with the overall low nucleotide diversity, which could either be due to selective pressure or recent demographic expansions following a bottleneck of these populations. The optimal genetic structure according to BAPS suggested populations grouped into four groups ( $K = 4$ ).

The first analysis with all 163 sequences, showed a clear division among the coastal and the inland populations, whereas further analyses, using each one of the clusters generated by the first inspection, showed a division among the North and South coastal groups, formed respectively by the haplogroups A and C, while the inland populations were also divided into 2 groups, a heterogeneous one, formed by the Haplogroups A and B, and a more homogenous one, formed mostly by haplogroup B.

AMOVA results showed similar values to scenarios with three and four groups, based on BAPS clustering. A scenario of 4 groups seems more realistic to explain *D. sturtevantii* diversification (Table 2).  $F_{CT}$  and  $F_{ST}$  values indicated high levels of genetic differentiation (0.26528 and 0.32415, respectively) among groups and within populations, and moderate differentiation among populations within groups ( $F_{SC} = 0.0801$ ). Thus, the AMOVA results are congruent with the haplotype network and reinforce the BAPS structure results. Isolation-by-distance (IBD), as measured by the simple  $r_{G \times D}$  between both geographical and  $F_{ST}$ -inferred genetic distances (Matrix 2, see Appendix S2) yielded significant results ( $r = 0.29869$ ,  $p = 0.047$ ).

**Table 2.** Molecular Variance Analysis (AMOVA) for the Cytochrome Oxidase I (COI) and II (COII) concatenated genes of *Drosophila sturtevantii*.  $\Phi_{ST}$  = covariance within localities;  $\Phi_{SC}$  = covariance among localities within groups;  $\Phi_{CT}$  = covariance among groups.

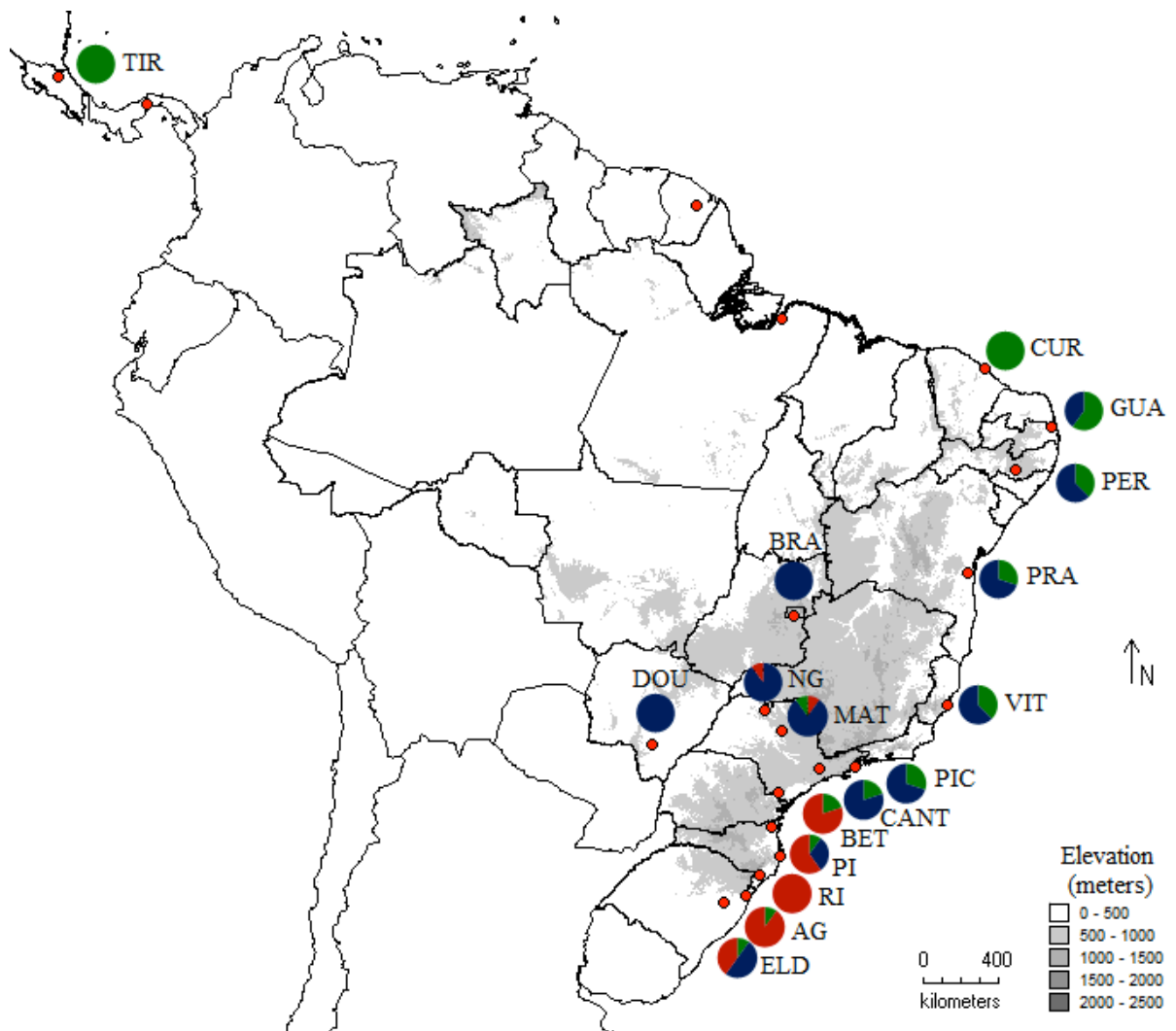
Source of variation	Variance components	Percentage of variation	$\Phi$ Statistics
Among groups	0.26933	26.53	$\Phi_{CT} = 0.26528$ $p < 0.05$
Among populations within groups	0.05976	5.89	$\Phi_{SC} = 0.08012$ $p < 0.05$
Within populations	0.68618	67.59	$\Phi_{ST} = 0.32415$

---

p < 0.05

---

**Figure 2.** Geographical distribution of the haplogroup frequencies across the sampled localities of *D. sturtevantii*. The codes correspond to the localities specified in Table 1. The red dots without frequencies represent SPN, FG, BEL and SMQ, with insufficient number of individuals (n<5). The colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively.

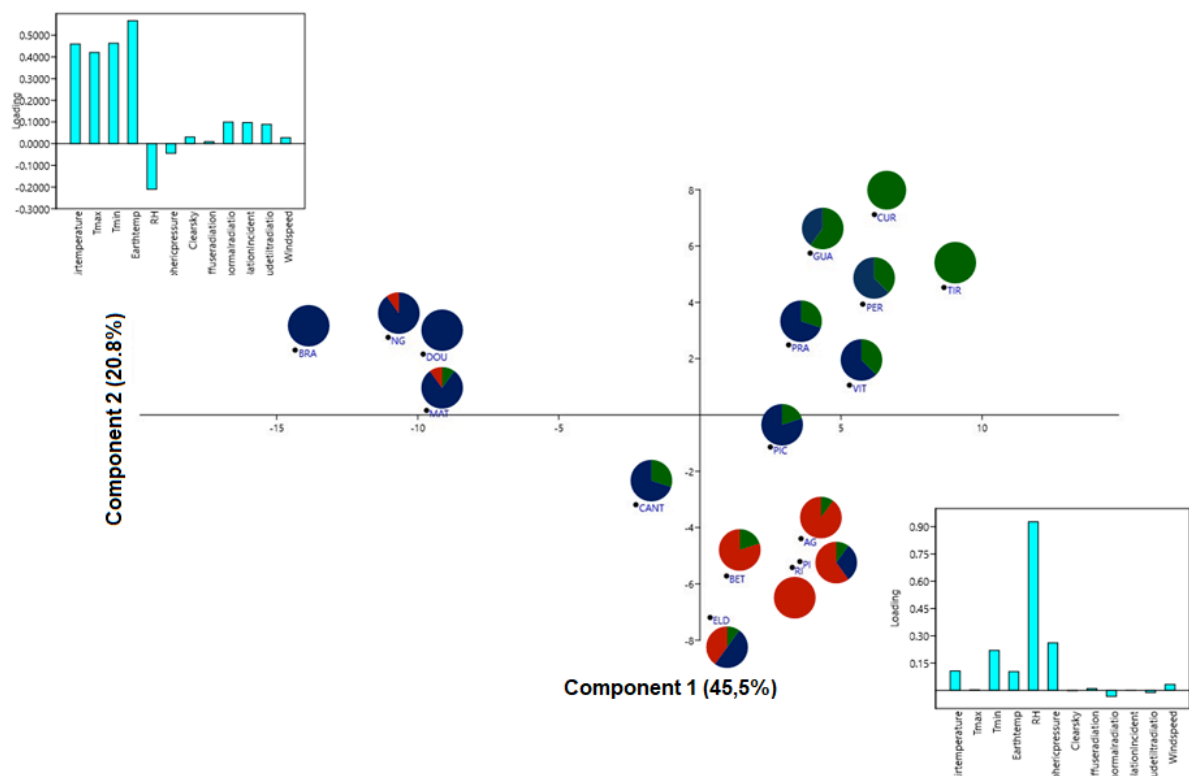


### Genetic differentiation strongly correlates with climatic variables

Principal components analysis (PCA) of the 13 climatic variables indicated a strong differentiation across the 17 localities (Figure 3). The two first components

(PC1 and PC2) explained 45.5 and 20.8% of the total variance, respectively. PC1 strongly correlated with relative humidity, whereas PC2 correlated positively with temperature variables and negatively with relative humidity. Interestingly, the frequency of the three haplogroups followed this climatic differentiation (Figure 3). Haplogroup A dominated in Northern equatorial populations, characterized by high humidity and temperature. Haplogroup B dominated mostly in inland, western Central populations characterized by low humidity but moderately high temperature. Haplogroup C dominated in Southern temperate populations characterized by high humidity and low temperature. This also corroborates significant isolation-by-ecology ( $r=0.28352$ ,  $p=0.0222$ ) and eco-spatial correlation ( $r=0.33457$ ,  $p=0.008$ ).

**Figure 3.** Principal Component Analysis (PCA) of the 13 climatic variables across the 17 geographical localities of *D. sturtevantii*. The codes correspond to the localities specified in Table 1. SPN, FG, BEL and SMQ, were not included in the analysis due to insufficient number of individuals ( $n<5$ ). The colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively. The bar plots in the insets refer to the correlation between each climatic variable and principal component 1 (low right) and 2 (top left).

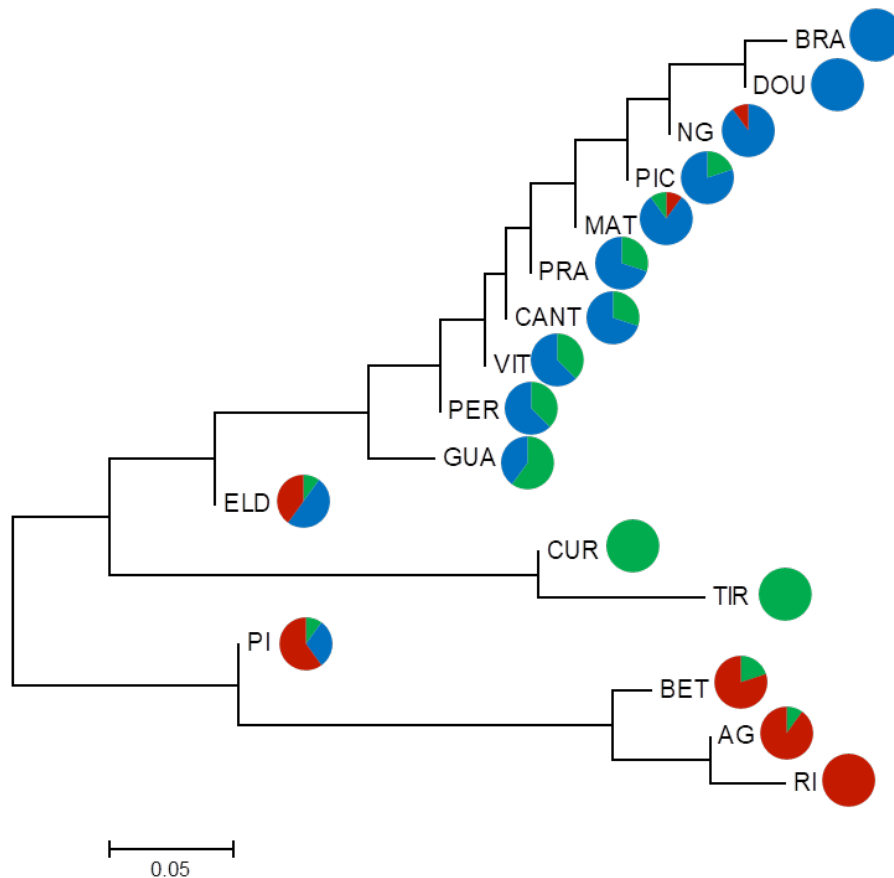


### Phylogeographic history of *D. sturtevantii* mitochondrial DNA

Tree clustering inferred from the degree of genetic differentiation (Matrix 2, see Appendix S2) supports a closer affinity between Northern and Central

populations, predominated by haplogroups A and B, respectively, than to the Southern population (except for ELD) (Figure 4). Together with the haplogroup network (Figure 1), this pattern suggests that Southern populations have diverged first, followed by a gradual and more recent split between Northern and Central populations, with inland, Western Central populations appearing as the most distant from the presumably ancestral stem of the North (Figure 4).

**Figure 4.** Neighbor Joining tree inferred from the  $F_{ST}$  pairwise distances among 17 geographical localities of *D. sturtevantii*. The codes correspond to the localities specified in Table 1. SPN, FG, BEL and SMQ, were not included in the analysis due to the insufficient number of individuals ( $n < 5$ ). The colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively. Scale bar refers to  $F_{ST}$ .

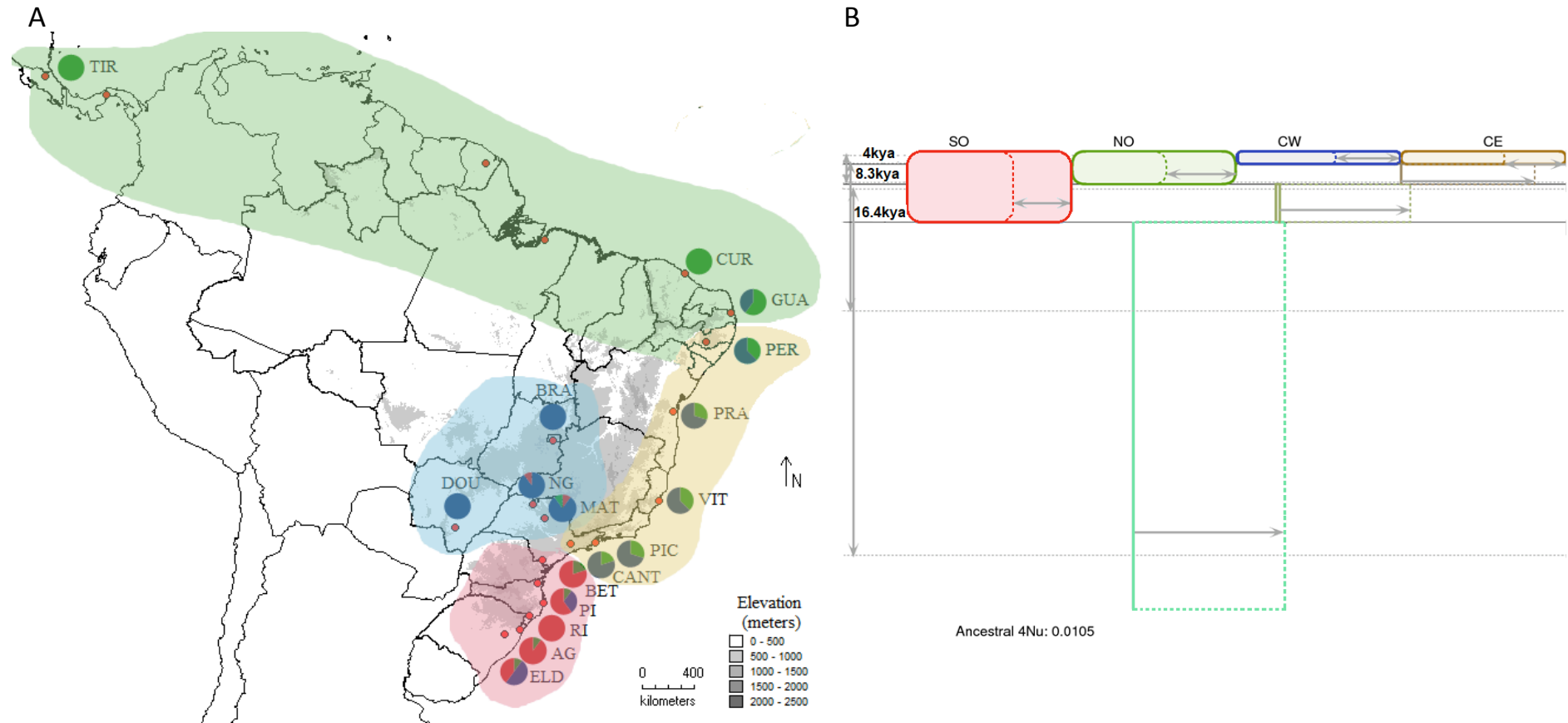


In order to test the possible times of such events we split the 17 populations into four populations groups: South, North, Central East and Central West (Figure 5, left), and conducted an isolation-with-migration analysis using IMA2. Assuming a mtDNA mutation rate of  $6.2 \times 10^{-9}$  per site (Haag-Liautard et al., 2008) and 10 generations per year, the results (Figure 5, right) suggest that the split of the Southern populations took place around 16,500 year ago (ya). This was followed by

a split between Northern and Central populations around 8,300 ya. The inland Western Central populations seem to diverge from the Eastern Atlantic populations nearly 4,000 ya.



**Figure 5.** IMa2-inferred demographic history of *D. sturtevantii* populations. A) Geographical distribution of the haplogroup frequencies to the sampled localities of *D. sturtevantii* split into four population groups: South (SO), North (NO), Central East (CE) and Central West (CW). The codes correspond to the localities specified in Table 1. The red dots without frequencies represent SPN, FG, BEL and SMQ, with insufficient number of individuals ( $n < 5$ ). B) population size and divergence time plot generated by IMa2.



## Discussion

Our work suggests three main findings in *D. sturtevantii*: 1) high mitochondrial haplotype diversity contrasts a very low mitochondrial nucleotide diversity, 2) mtDNA variation is geographically structured and correlates with ecological variables, and 3) eco-geographical differentiation of mtDNA took place first in the Pleistocene (for Southern subtropical populations) and then during the Holocene for Northern and Central tropical populations). Below we discuss the bearing of each of these points on Neotropical phylogeography as it relates to *Drosophila*.

The diversity values found in *D. sturtevantii* populations ( $H_d = 0.864$ ) are higher than the average values found in other Neotropical *Drosophila* species occurring in Atlantic Forest fragments ( $H_d = 0.206$ ) and also in cactophilic species occurring in the seasonally dry tropical forests (SDTF) of South America ( $H_d = 0.554$ ) (Manfrin & Sene, 2006; Franco & Manfrin, 2013; Cenzi de Ré et al., 2014; Gustani et al., 2015; Barrios-Leal et al., 2018a, b). Furthermore, mtDNA diversity in *D. sturtevantii* is very low, which is in line with the star-like shape of the haplotype network (Figure 1) and the neutrality test statistics supporting a population expansion scenario.

Among the different causes of mtDNA variation, an important one to be considered is the *Wolbachia* infection. *Wolbachia* is an intracellular symbiont co-inherited with mitochondria and it is responsible for most of the infections in invertebrates. Its presence can lead to sex ratio distortions and under some conditions, it can increase host fitness, so it could also be affecting the diversity in *D. sturtevantii* (Miller & Riegler, 2006). Haplotype diversity can be considered high when compared to other species collected in forest fragments, or moderate to low when compared to cactophilic species. We already know that *D. sturtevantii* populations are naturally infected with *Wolbachia* (Miller & Riegler, 2006), however, we still have to investigate the association between the mitochondrial and *Wolbachia* haplotypes, to understand if it is interfering in the observed variation. Müller et al. (2012) showed no evidence that *Wolbachia* infections could be decreasing the diversity of mtDNA in *D. willistoni* natural populations, even with an infection rate as high as 55%.

Regarding the geographic distribution of the haplogroups and time estimates, it is important to emphasize the distribution similarities between our

results and those found to morphological data to *D. sturtevantii* using 14 populations here analyzed. Segala (2019) presented a population structure similar to the one we found here, based on morphological variation (aedeagus). Aedeagus analysis also supported a southernmost cluster which comprises the ELD population, and it contained a large portion of the blue haplogroup in our findings. This population is located in a transition ecotone between the Atlantic Forest and the Pampa grasslands which could withstand a greater variation than more homogeneous environments. Also, Trava (2018) examined the same populations of *D. sturtevantii* and found moderate population structure with microsatellite data in spite of a lack of correlation with geographic distances.

Moreover, the geographic distribution of *D. sturtevantii* is quite similar to the one supported for *D. meridionalis* by Barrios-Leal et al. (2018a), where they also had one inland and two coastal groups, in spite of the coastal groups having a more complex dynamic affected by recurrent expansions and contractions with isolation and admixture of populations, *D. sturtevantii* groups seem to be more isolated, evidenced by the low rates of shared haplotypes.

Regarding the timing of population divergence, the first split observed in the South is in agreement with isotope analysis and  $^{14}\text{C}$  dating of the buried charcoal fragments and humin fraction of soil organic matter (Saia et al., 2018). According to their findings, the area near BET (the first geographical population in the South group) has been dominated by forests for at least 20,000 years. This suggests a change toward more humid and wet conditions which could have favored the expansion and split of *D. sturtevantii* populations to a more suitable environment, following the forest expansion after the LGM (Last Glacial Maximum). Moreover, the occurrence of a previously suggested genetic barrier between the São Paulo and Paraná states, near the Paranapanema River, could suggest vicariance (Grazziotin et al., 2006). The second split event, between the North and Central populations is estimated around 8,300 ya, in concordance with an increase in diversity based in palynological evidence in the Parque Estadual do Rio Doce, Minas Gerais, (240 km from VIT). Then, the more recent divergence event occurred around 4,000 ya, which could have been influenced by the establishment of the modern seasonal climate and the return of the small gallery forest and arboreal Cerrado (Brazilian savanna biome) to central Brazil ~3,500 ya, with a warm semi-humid climate, similar to

the present one (Ferraz-Vicentini & Salgado-Laboriau; 1996; Buso et al., 2013). The paleoclimatic evidence highlights how the diversification and expansion of *D. sturtevanti* populations took place in different environments over time and can also explain, at least in part, the current distribution of the haplogroups related to the climate and ecological variation.

### **Acknowledgments**

We thank Prof. Dr. Daniela de Toni, Prof. Dr. Claudia Rohde and Prof. Dr. Hermes J. Schmitz for supplying some strains of *D. sturtevanti*, FAPESP for support of this project (process number 2014/14059-0, 2016/11994-5), CAPES for supporting the Postgraduate Program in Biosciences of Unesp/Ibilce and CAPES for the doctoral and PDSE scholarship granted to Samara Videira Zorzato (PDSE/CAPES – Proc. 88881.190070/2018-01).

### **Supporting Information**

S1 – Customized Perl scripts used in IBE analysis

S2 – Ecological variables and distance matrices

### **References**

- Antonelli, A., Zizka, A., Carvalho, F.A., Scharn, R., Bacon, C.D., Silvestro, D. & Condamine, F.L. (2018). Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences of the U S A* 115: 6034-6039.
- Barrios-Leal, D.Y., Franco, F.F., Silva, E.C.C., Santos, C.K.B., Sene, F.M. & Manfrin, M.H. (2018a). Deep intraspecific divergence in *Drosophila meridionalis*, a cactophilic member of the New World *Drosophila repleta* group. *Biological Journal of the Linnean Society* 123: 163-178.
- Barrios-Leal, D.Y., Neves-da-Rocha, J. & Manfrin, M.H. (2018b). Genetics and distribution modeling: the demographic history of the cactophilic *Drosophila buzzatii* species cluster in open areas of South America. *Journal of Heredity* 20:1-12.
- Bastide, H., Yassin, A., Johanning, E.J. & Pool, J.E. (2014). Pigmentation in *Drosophila melanogaster* reaches its maximum in Ethiopia and correlates most

strongly with ultra-violet radiation in sub-Saharan Africa. *BMC Evolutionary Biology* 179: 1-14.

Beheregaray, L.B. (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern Hemisphere. *Molecular Ecology* 17: 3754-3774.

Bicudo, H.E.M.C. (1979). Reproductive isolation in the *saltans* group of *Drosophila*. IV. The *sturtevanti* subgroup. *Revista Brasileira de Genética* 4: 247-258.

Buso Junior, A.A., Ruiz Pessenda, L.C., de Oliveira, P.E., Fonseca Giannini, P.C., Lisboa Cohen, M.C., Volkmer-Ribeiro, C. & Siqueira, G.S. (2013). Late Pleistocene and Holocene vegetation, climate dynamics, and Amazonian taxa in the Atlantic Forest, Linhares, SE Brazil. *Radiocarbon* 55: 1747-1762.

Carnaval, A.C. & Moritz, C. (2008). Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography* 35: 1187-1201.

Carvalho, C.J.B. & Almeida, E.A.B. (2016). *Biogeografia da América do Sul: Análise de tempo, espaço e forma*. 2<sup>a</sup> edition. Rio de Janeiro: Roca. 324p.

Cenzi de Ré, F., Gustani, E.C., Oliveira, A.P.F., Machado, L.P.B., Mateus, R.P., Loreto ELS & Robe, L.J. (2014). Brazilian populations of *Drosophila maculifrons* (Diptera, Drosophilidae): low diversity levels and signals of a population expansion after the Last Glacial Maximum. *Biological Journal of the Linnean Society* 112: 55-66.

Corander, J., Cheng, L., Marttinen, P., Sirén, J. & Tang, J. (2013). *BAPS: Bayesian Analysis of Population Structure*, Manual. 6: 1-28.

Corander, J., Marttinen, P. & Tang, J. (2008). Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics* 9: 1-14.

Darriba, D., Taboada, G.I., Doallo, R. & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9(8): 772.

Drummond, A.J. & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology*. 7: 214.

Drummond, A.J., Suchard, M.A., Xie, D. & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* 29: 1969-1973.

Excoffier, L. & Lischer, H.E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

- Ferraz-Vicentini, K.R. & Salgado-Labouriau, M.L. (1996). Palynological analysis of a palm swamp in Central Brazil. *Journal of South American Earth Sciences* 9: 207-219.
- Franco, F.F. & Manfrin, M.H. (2013). Recent demographic history of cactophilic *Drosophila* species can be related to Quaternary palaeoclimatic changes in South America. *Journal of Biogeography* 40: 142-154.
- Fu, Y. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147: 915-925.
- Grazziotin, F. G., Monzel, M., Echeverrigaray, S., & Bonatto, S. L. (2006). Phylogeography of the *Bothrops jararaca* complex (Serpentes: Viperidae): past fragmentation and island colonization in the Brazilian Atlantic Forest. *Molecular Ecology* 15: 3969-3982.
- Gustani, E.C., Oliveira, A.P.F., Santos, M.H., Machado, L.P.B. & Mateus, R.P. (2015). Demographic Structure and evolutionary History of *Drosophila ornatifrons* (Diptera, Drosophilidae) from Atlantic Forest of Southern Brazil. *Zoological Science* 32: 141-150.
- Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B. & Keightley, P.D. (2008) Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biology* 6: 1706-1714.
- Haffer, J. (1969). Amazonian forest birds. *Science* 165: 131-137.
- Haffer, J. (1993). Time's cycle and time's arrow in the history of Amazonia. *Biogeographica* 69: 15-45.
- Haffer, J. (1997). Alternative models of vertebrate speciation in Amazonia: An overview. *Biodiversity and Conservation* 6: 451-476.
- Haffer, J. & Prance, G.T. (2002). Impulsos climáticos da evolução na Amazônia durante o Cenozóico: sobre a teoria dos Refúgios da diferenciação biótica. *Estudos Avançados* 16: 175-206.
- Hales, K.G., Korey, C.A., Larracuenta, A.M. & Roberts, D.M. (2015). Genetics on the fly: A primer on the *Drosophila* model system *Genetics* 201: 815-842.
- Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/ 98/NT. *Nucleic Acids Symposium Series* 41: 95-98.
- Hammer, Ø. & Harper, D.A.T. (2006). *Paleontological Data Analysis*. Oxford: Blackwell Publishing.
- Hey, J. (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*. 27: 905-920.
- Hey, J. & Nielsen, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics.

*Proceedings of the National Academy of Sciences of the U S A* 104 : 2785-2790.

Hijmans, R.J., Guarino, L. & Mathur, P. (2012) DIVA-GIS version 7.5, Manual available at [www.diva-gis.org](http://www.diva-gis.org).

Lajbner, Z., Pnini, R., Camus, M.F., Miller, J. & Dowling, D.K. (2018). Experimental evidence that thermal selection shapes mitochondrial genome evolution. *Scientific Reports* 9500: 1-12.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

Librado, P. & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451-1452.

Madi-Ravazzi, L., Segala, L.F., Roman, B.E., Alevi, K.C.C., Prediger, C., Yassin, A., Hua-Van, A. & Miller, W. (2021) Integrative taxonomy and a new species description in the *sturtevanti* subgroup of the *Drosophila saltans* group (Diptera: Drosophilidae). *Zootaxa* 4980:269-292.

Magalhães, L.E. (1962). Notes on the taxonomy, morphology and distribution of saltans group of *Drosophila*, with description of four new species. *UT Publications* 6205: 135-154.

Manfrin, M.H. & Sene, F.M. (2006). Cactophilic *Drosophila* in South America: a model for evolutionary studies. *Genetica* 126: 57-75.

Miller, W.J & Riegler, M. (2006). Evolutionary Dynamics of wAu-Like *Wolbachia* variants in Neotropical *Drosophila* spp. *Applied and Environmental Microbiology* 72: 826-835.

Mittermeier, R.A., Gil, P.R., Hoffmann, M., Pilgrim, J., Brooks, J., Mittermeier, C.G., Lamourux, J. & Fonseca, G.A.B. (2005). *Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions*. Cemex, 392p.

Müller, M.J., von Mühlen, C., Valiati, V.H. & Valente, V.L.S. (2012). *Wolbachia pipientis* is associated with different mitochondrial haplotypes in natural populations of *Drosophila willistoni*. *Journal of Invertebrate Pathology* 109: 152-155.

Myers, N., Mittermeier, R.A., Mittermeier, C.G., Da Fonseca, G.A.B. & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*. 403 : 853-858.

O'Grady, P. & DeSalle, R. (2018). Phylogeny of the genus *Drosophila*. *Genetics* 209: 1-25.

Posada, D. & Crandall, K.A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Rambaut, A., Drummond, A.J., Xie, D., Baele, G. & Suchard, M.A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* 37: 901-904.

Ramos-Onsins, S.E. & Rozas, J. (2002). Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* 19: 2092-2100.

Ribeiro, M.C., Metzger, J.P., Martensen, A.C., Ponzoni, F.J. & Hirota, M.M. (2009). The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distribution? Implications for conservation. *Biological Conservation* 142: 1141-1153.

Rozas, J., Sánchez-Delbarrio, J.C., Messeguer, X. & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by coalescent and other methods. *Bioinformatics* 19: 2496–2497

Rull, V. (2011). Neotropical biodiversity: timing and potential drivers. *Trends in Ecology & Evolution* 26: 508-513.

Saia, S.E.M.G., Pessenda, L.C.R., Gouveia, S.E.M., Aravena, R. & Bendassolli, J.A. (2008). Last glacial maximum (LGM) vegetation changes in the Atlantic Forest, southeastern Brazil. *Quaternary International* 184:195–201.

Salzburger, W., Ewing, G.B. & Von Haeseler, A. (2011). The performance of phylogenetic algorithms in estimating haplotype genealogies with migration. *Molecular Ecology*. 20: 1952-1963.

Segala, L.F. (2019). Diferenciação populacional em *Drosophila sturtevantii* (subgrupo *sturtevantii*, grupo *saltans*) avaliada por morfometria geométrica da asa e do edeago. Doctoral dissertation. Postgraduate Program in Biosciences. UNESP, 106 p.

Sevegnani, L., Uhlmann, A., Gasper, A.L., Meyer, L. & Vibrans, A.C. (2016). Climate affects the structure of mixed rain forest in southern sector of Atlantic domain in Brazil. *Acta Oecologica* 77: 109-117.

Shafer, A.B.A. & Wolf, J.B.W. (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters* 16: 940-950.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* 87: 651-701.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585-595.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence



weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 22: 4673–4680.

Throckmorton, L.H. (1975). The phylogeny, ecology and geography of *Drosophila*, p. 421–469 in *Handbook of Genetics*, edited by R. King. Plenum Publishing Corporation, New York.

Trava, B.M. (2018). Estrutura populacional de *Drosophila sturtevantii* (subgrupo *sturtevantii*; grupo *saltans*) por meio de microssatélites espécie-específicos e biodiversidade de drosofilídeos em domínios da Mata Atlântica. **Tese de doutorado**. Programa de Pós-Graduação em Biociências. UNESP, 125p.

Tuomisto, H. (2007). Interpreting the biogeography of South America. *Journal of Biogeography* 34: 1294-1295.