



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Câmpus de São José do Rio Preto

Samara Videira Zorzato

**Filogeografia de *Drosophila sturtevanti* (Diptera:  
Drosophilidae) em biomas Neotropicais**

São José do Rio Preto  
2020

Samara Videira Zorzato

**Filogeografia de *Drosophila sturtevanti* (Diptera:  
Drosophilidae) em biomas Neotropicais**

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Biociências, junto ao Programa de Pós-Graduação em Biociências, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: FAPESP – Proc. 2014/14059-0  
FAPESP – Proc. 2016/ 11994-5  
CAPES  
PDSE / CAPES – Proc. 88881.  
190070/2018-01

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Lilian Madi-Ravazzi  
Coorientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Adriana Coletto Morales

São José do Rio Preto  
2020

Z89f

Zorzato, Samara Videira

Filogeografia de *Drosophila sturtevantii* (Diptera: Drosophilidae) em biomas Neotropicais / Samara Videira Zorzato. -- São José do Rio Preto, 2020

97 p. : il., tabs., mapas

Tese (doutorado) - Universidade Estadual Paulista (Unesp), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto

Orientadora: Lilian Madi-Ravazzi

Coorientadora: Adriana Coletto Morales

1. Adaptação. 2. Grupo saltans. 3. Região Neotropical. 4. mtDNA. 5. Filogeografia. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Samara Videira Zorzato

**Filogeografia de *Drosophila sturtevanti* (Diptera:  
Drosophilidae) em biomas Neotropicais**

Tese apresentada como parte dos requisitos para obtenção do título de Doutora em Biociências – Genética, junto ao Programa de Pós-Graduação em Biociências, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: FAPESP – Proc. 2014/14059-0  
FAPESP – Proc. 2016/ 11994-5  
CAPES  
PDSE / CAPES – Proc. 88881.  
190070/2018-01

**Comissão Examinadora**

Prof<sup>a</sup>. Dr<sup>a</sup>. Lilian Madi-Ravazzi  
UNESP – Câmpus de São José do Rio Preto  
Orientador

Prof<sup>a</sup>. Dr<sup>a</sup>. Cláudia Marcia Aparecida Carareto  
UNESP – Câmpus de São José do Rio Preto

Prof<sup>a</sup>. Dr<sup>a</sup>. Maura Helena Manfrin  
FFCLRP – USP – Campus de Ribeirão Preto

Prof<sup>a</sup>. Dr<sup>a</sup>. Luciana Paes de Barros Machado  
Universidade Estadual do Centro-Oeste – Campus de Guarapuava

Prof. Dr. Luís Gustavo da Conceição Galego  
Universidade Federal do Triângulo Mineiro – Uberaba

São José do Rio Preto  
05 de julho de 2019

A todos que me acompanharam e apoiaram durante esses anos, especialmente meus pais, **João Roberto** e **Sueli** e minha irmã **Vitória**.  
Sem vocês nada disso seria possível.

## AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Esse é o resultado de mais de 4 anos de trabalho e eu posso dizer que se cheguei até aqui, não foi apenas graças à minha dedicação, mas a inúmeras pessoas que me ajudaram ao longo do caminho e eu não seria capaz de agradecer a todos de forma apropriada.

À minha orientadora, Profa. Dra. Lilian Madi-Ravazzi, por ter aceitado me orientar e por todo o aprendizado durante esses anos e à minha coorientadora, Profa. Dra. Adriana C. M. Corrêa e Castro pelo auxílio inicial em compreender “a tal da filogeografia”, por toda a colaboração e ensinamentos.

Ao Dr. Amir Yassin por toda a dedicação e paciência em me ensinar durante o tempo em que me supervisionou no estágio na França, nada disso seria possível sem você.

Aos membros do Institut de Systématique, Evolution, Biodiversité (ISYEB) do Muséum National d’Histoire Naturelle em Paris por me receberem tão bem durante o tempo que lá passei. Amir Yassin, Claudie Doums, Thierry Wirth, Stefano Mona, Romain Nattier, Stephane Prigent, Marie Christine Maurel, Arnaud Suwalski, Pascaline Chifflet-Belle, Karen Salazar, Lory-Anne Baker, Basile Finand, Yohann Vigoureux, Hugo Lain e Erina Ferreira. Especialmente à Erina pela amizade e por todos os momentos que passamos juntas, Stephane por todas as conversas sobre *Drosophila* e Hugo pelas conversas políglotas misturando português, francês e inglês.

Ao Programa de Pós-Graduação em Biociências pela oportunidade de realização da pesquisa e competência em ajudar sempre que necessitei.

À CAPES pela concessão de bolsa de doutorado e pela bolsa PDSE, que me possibilitou uma oportunidade única de aprendizado durante os seis meses que passei no Museu de História Natural em Paris.

À Fundação de Amparo à Pesquisa do Estado de São Paulo, FAPESP, pelo financiamento do projeto sob os processos nº 2014/14059-0 e 2016/11994-5.

Às coordenadoras do Programa de Pós Graduação em Biociências durante os anos que aqui estive, Profa. Dra. Claudia Carareto e Profa. Dra. Claudia Bonini, por sempre serem prestativas e competentes na realização de suas funções, facilitando (e muito!) esses quatro anos de trabalho.

Aos professores Claudia Rohde, Daniela de Toni e Hermes J. Schmitz pela disponibilização de material biológico que foi essencial para uma boa amostragem geográfica.

Aos membros da banca de qualificação, Marjorie da Silva e Lilian Castiglioni, e, de defesa, que trouxeram valiosas contribuições, ajudando na lapidação da tese. Em especial à Marjorie e ao Luís por toda a paciência e pela ajuda durante o processo, sempre dispostos a tirar as mil dúvidas que surgiram no meio do caminho.

A todos os professores que contribuíram para minha formação acadêmica e se mostraram verdadeiros amigos nos momentos críticos. À Luciana Paes de Barros Machado, por ser meu primeiro exemplo e mãe acadêmica, mesmo à distância nunca me esquecerei de todos os ensinamentos e do primeiro gel de agarose que você me ensinou a correr (e da sua reação quando apliquei a amostra não tão delicadamente e perfurei o gel rsrs), à Maura e Dora pelas valiosas discussões sobre os meus dados e pelo auxílio em rodar o BAPS.

Ao Rodrigo e Beto, motoristas da Unesp, por todo o auxílio nas coletas e por literalmente se embrenharem no mato e pela relação de amizade criada a partir disso.

Aos colegas de laboratório Bruna Emília, Bruna Trava, Luis Fernando Segala, Gabriel Tripode, Rafael Miyasaki, Larissa Gonçalves, Jéssica Paixão e Natália Francielle, por todos os momentos compartilhados, pelas risadas, pelo auxílio constante e por me ensinarem a importância do trabalho em equipe. Ao Tião, técnico do Laboratório de Meio de Cultura, por toda a atenção, pelas conversas e pela pipoca de fim de tarde...

Especialmente à Bruna Emília, com quem convivi diariamente dentro e fora do laboratório e que é uma das melhores pessoas que já conheci. Você é de longe a primeira pessoa que eu penso em Rio Preto e vou morrer de saudades! Também à Larissa, por todas as nossas conversas e por sempre conseguir me acalmar e mostrar uma solução para os problemas. Obrigada por me aceitarem e apoiarem. Vocês tornaram minha vida mais doce e alegre e, se hoje cheguei onde estou, esse mérito também é de vocês!

Ao pessoal do laboratório de Anatomia, Bruno Valverde, Maysa Succi, Wadson Resende e Gabriel da Cunha, por todos os momentos. Obrigada por estarem sempre presentes e ao Bruno por sempre trazer comidas gostosas. À Julcimara, Damaris, Joyce e demais funcionários do departamento de Biologia e do Ibilce por toda a infraestrutura que permitiram um ambiente propício para a execução do trabalho.

Ao pessoal do laboratório de Evolução Molecular, em especial à Maryanna Simão e Cecília Ártico, que além de me ajudarem em várias questões acadêmicas se tornaram verdadeiras amigas. À Cecília pelo apoio durante o tempo que passei na França, tornando a adaptação e a saudade de falar português uma tarefa muito mais fácil.

À Nara Chiarini por me ajudar desde o início do doutorado e por se tornar uma amiga que quero levar comigo onde estiver. À Nara e ao Ivã Guidini Lopes por levarem minhas amostras para sequenciamento em Jaboticabal.

À Equipe da STAEPE por todo o auxílio nas prestações de conta e pela amizade dentro e fora do ambiente acadêmico. Em especial ao André Domingues, Eliane Gonçalves e à

Rosemar Rosa, que foram essenciais, me auxiliando em todos os momentos, vibrando a cada conquista e enxugando minhas lágrimas sempre que necessário.

Ao meu amigo desde a época de Guarapuava, William Menezes, por ser o mais próximo de um irmão que eu poderia ter. Obrigada pelo carinho incondicional e por me acolher em sua casa após meu retorno da França.

Aos amigos do lado francês, Laurent Alaguin, Fatim Sidibé, Dorine Souap, Anass, Ly KimSounaï, Kroa, Marie-Ange, Marie-Simone e Marie-Antoinette, por todo o tempo que passamos juntos, por todas as experiências e pela troca diária de conhecimento, por tornarem a França um segundo lar.

Às minhas amigas de longa data, Cássia Ribeiro e Fernanda Fuzzi, por me acompanharem nessa jornada e acompanharem meu crescimento pessoal e profissional, não tenho palavras para descrever o que sinto por vocês; aos amigos que o Ibilce me proporcionou, Gabriela Castro, Gabriela Mastelari, Jaqueline Vieira, Ana Beatriz Bortolotto e Beatriz Ricardo e a todos que algum dia já me ouviram falando sobre minhas mosquinhas.

À Manu Gustani, pela ajuda com vários aspectos da tese, por ser um exemplo de profissional e por me mostrar que 'tudo passa'. Obrigada por me ajudar a manter a calma e torcer pela minha felicidade e realização!

À Simone Izidoro, Vanessa Sovrani, Dani Lima, Gabriela Cebulski, Thaianne Prolo, Celina Likes, Cristian Zwetzch e demais amigos de Guarapuava por todo o apoio e por acreditarem em mim quando eu não acreditava mais.

Um agradecimento especial à Lara Talhaferro, a pessoa que me ensinou o significado da palavra sororidade e que me ajuda a cada dia a ser um ser humano melhor. Obrigada pelas conversas na cantina do Gil à base de pão de queijo e cappuccino.

Por fim, agradeço à minha família. A minha mãe Sueli e meu pai João Roberto, por me ensinarem valores que faculdade nenhuma poderia ter ensinado. E a minha irmã Vitória, por ser meu porto seguro. Agradeço a compreensão por todas as vezes que estive ausente e por todo o apoio, por me acompanharem aonde quer que eu fosse porque sabiam o quanto isso tudo significa para mim.

Obrigada a todos que contribuíram para que isso se tornasse possível. Com certeza essa é uma obra falha, foi feita por uma pessoa falha e isso não podia ser diferente. Mas aqui eu coloquei todo meu ATP, meu suor, minhas lágrimas e meu amor à pesquisa. E isso me basta e faz feliz.



*"Eu sou é eu mesmo. Diverjo de todo o mundo... Eu quase que nada não sei. Mas desconfio de muita coisa. O senhor concedendo, eu digo: para pensar longe, sou cão mestre - o senhor solte em minha frente uma idéia ligeira, e eu rastreio essa por fundo de todos os matos, amém!"*

(João Guimarães Rosa, 1986, p.15)

## RESUMO

As mudanças climáticas ocorridas no Pleistoceno afetaram a distribuição e os padrões de diversificação populacionais e são conhecidas como um dos potenciais impulsionadores da alta diversidade encontrada na região Neotropical. *Drosophila sturtevantii* pertence ao subgrupo *sturtevantii* do grupo *saltans* de *Drosophila* e é uma das espécies com distribuição geográfica mais ampla dentro do grupo, ocorrendo em áreas florestais de diferentes biomas. Além disso, essa espécie apresenta altos níveis de variabilidade genética para diversos marcadores, constituindo um excelente modelo para estudos de estrutura populacional, diversificação e adaptação em diferentes cenários. O objetivo desse estudo foi analisar a diferenciação genética, usando os genes Citocromo Oxidase subunidades I e II em 163 indivíduos provenientes de 21 populações de *D. sturtevantii* coletados em fragmentos florestais da América Central ao sul do Brasil, com ênfase na Mata Atlântica. Nossos resultados moleculares mostram que as populações de *D. sturtevantii* estão estruturadas na América do Sul e sua distribuição mostra uma correlação significativa com variáveis geográficas e climáticas, principalmente temperatura e umidade. Além disso, os eventos de divergência entre os grupos geográficos sugerem associação significativa com as oscilações paleoclimáticas ocorridas desde o Pleistoceno tardio e, esses eventos, provavelmente moldaram a distribuição geográfica dos haplogrupos de acordo com variáveis ambientais, principalmente temperatura e umidade.

**Palavras-chave:** Adaptação, Grupo *saltans*, Região Neotropical, mtDNA, Filogeografia, Paleoclima

## ABSTRACT

The Pleistocene climatic changes have affected population distribution and diversification patterns and it is known as one of the potential drivers to the high diversity found in the Neotropics. *Drosophila sturtevantii* belongs to the *sturtevantii* subgroup of the *Drosophila saltans* group and it is one of the most geographically wide species within the group, occurring on forest areas of different biomes. Besides, it presents high levels of genetic variability for several markers, thus constituting an excellent model to study population structure, diversification and adaptation over different scenarios. The aim of this study was to analyze the genetic differentiation using the Cytochrome Oxidase I and II genes of 163 individuals from 21 populations of *D. sturtevantii* collected forest fragments from Central America to South Brazil, with emphasis on the Brazilian Atlantic Forest. Our molecular results showed that *D. sturtevantii* populations were structured in South America and that their mitochondrial haplotypes had a significant correlation with geography and climatic variables, mainly temperature and humidity. Moreover, the divergence events between the geographical groups suggest significant association with paleoclimate oscillations occurred since the late Pleistocene. The results hence support a role of paleoclimatic events on cytogenomic diversity of present-day widespread Neotropical species.

**Keywords:** Adaptation, *saltans* group, Neotropics, mtDNA, Phylogeography, Paleoclimate

## LISTA DE ILUSTRAÇÕES

- Figura 1.** Mapa das regiões zoogeográficas terrestres atualizadas por Holt et al., (2013). Em azul escuro encontra-se a região Neotropical, que se estende do México ao sul da América do Sul..... 18
- Figura 2A-B.** Mapas de distribuição do bioma da Mata Atlântica brasileira. **A.** Sub-regiões biogeográficas baseadas nas principais áreas de endemismo de aves, borboletas e primatas, como proposto por Silva e Casteleti (2003). **B.** Remanescentes florestais na Mata Atlântica brasileira. ....22
- Figura 3.** Mapa dos Biomas Brasileiros. Em cores preenchidas estão os biomas que fazem fronteira com a Mata Atlântica. ....24
- Figura 4A-B.** **A.** Relação evolutiva proposta dentro do subgênero *Sophophora*, adaptado de O’Grady e Kidwell (2002) e, em **B.** relação evolutiva dentro do grupo *saltans*, como enfoque no subgrupo *sturtevanti*, baseada nos resultados obtidos por Roman (2018) por meio de uma análise particionada por Inferência Bayesiana. ....39
- Figura 5.** Imagem de uma fêmea de *D. sturtevanti*.....41

## Capítulo

- Figure 1.** Haplotype network of concatenated data to *D. sturtevanti* populations. Each line of the network represents a single mutation step. Haplotypes are represented by circles with sizes which are proportional to the number of individuals. Hypothetical haplotypes are represented by small blue circles and the outgroup is represented by the white circle on the left represented by stv-like sample.....52
- Figure 2.** BI tree of *D. sturtevanti* based on concatenated dataset (*COI* and *COII* sequences). *D. sturt-like* strain from French Guyana was utilized as the outgroup. The colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively. ....53
- Figure 3.** Geographical distribution of the haplogroup frequencies to the sampled localities of *D. sturtevanti*. 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 (<5). The frequencies colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively.....55

**Figure 4.** Principal Component Analysis (PCA) of the 13 climatic variables of 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 (<5). The frequencies colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively. ....56

**Figure 5.** Neighbor Joining tree inferred from the Fst pairwise distances between 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 (<5). The frequencies colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively. ....57

**Figure 6.** Geographical distribution of the haplogroup frequencies to the sampled localities of *D. sturtevantii* splitted 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 (<5). On the right, the divergence time plot generated by IMA2. A. Clusters generated by BAPS, each color corresponds to the same groups used to IMA2. ....58

## LISTA DE TABELAS

### Capítulo

<b>Table 1.</b> Location, code and summary descriptive statistics for the sampled populations of <i>D. sturtevantii</i> using the <i>COI</i> and <i>COII</i> concatenated data. ....	51
<b>Table 2.</b> Molecular Variance Analysis (AMOVA) for the Cytochrome Oxidase I ( <i>COI</i> ) and II ( <i>COII</i> ) concatenated genes of <i>Drosophila sturtevantii</i> . $\Phi_{st}$ = covariance within localities; $\Phi_{sc}$ = covariance among localities within groups; $\Phi_{ct}$ = covariance among groups. ....	54

## LISTA DE ABREVIATURAS E SIGLAS

**AG:** Aguaí

**APA:** Área de Proteção Ambiental

**BEL:** Belém

**BET:** Betary

**BRA:** Brasília

**CANT:** Cantareira

**CE:** Central East populations

**COI:** Citocromo c oxidase subunidade I

**COII:** Citocromo c oxidase subunidade II

**CUR:** Curió

**CW:** Central West populations

**DNA:** Ácido desoxirribonucleico

**dNTP:** Desoxirribonucleotídeo Fosfatado

**DOU:** Dourados

**ELD:** Eldorado do Sul

**FAPESP:** Fundação de Amparo à Pesquisa do Estado de São Paulo

**FES:** Floresta Estadual Semidecidual

**FG:** Guiana Francesa

**FOD:** Floresta Ombrófila Densa

**GUA:** Guaribas

**IBAMA:** Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis

**IBD:** Isolation-by-geography

**IBE :** Isolation-by-Ecology

**IBGE:** Instituto Brasileiro de Geografia e Estatística

**INEMA:** Instituto Do Meio Ambiente E Recursos Hídricos

**LGM:** Last Glacial Maximum

**MAT:** Matão

**MMA:** Ministério do Meio Ambiente

**mtDNA:** DNA mitocondrial

**NG:** Nova Granada

**NGS:** *New Generation Sequencing*

**NJ:** Neighbor-Joining

**NO:** North populations

**PCA:** Principal Component Analysis

**PCR:** Polymerase Chain Reaction (reação em cadeia da polimerase)

**PER:** Pernambuco

**PETAR:** Parque Estadual Turístico do Alto do Ribeira

**PI:** Pirai

**PIC:** Picinguaba

**PRA:** Pratigi

**RI:** Ribeirão da Ilha

**SEMA RS:** Secretaria do Meio Ambiente do Rio Grande do Sul

**SEMA:** Áreas de preservação na Reserva Biológica Guaribas

**SEMACE:** Secretaria do Meio Ambiente do Ceará

**SMQ:** Maquiné

**SPN:** Panamá

**SO:** South populations

**TIR:** Tirimbina

**UMG:** Último Máximo Glacial

**VIT:** Vitória

**Ya:** years ago



## LISTA DE SÍMBOLOS

**%:** Porcentagem

**bp:** Par de bases

**μL:** Microlitros

**ng:** Nanograma

**pmol:** Picomol

**Mm:** Milimolar

**°C:** graus Celsius

**π:** Diversidade nucleotídica

**Hd:** Diversidade haplotípica

**S:** Número de sítios polimórficos

**h:** Número de haplótipos

**N:** Número de indivíduos

**Km:** quilômetro

**<sup>14</sup>C:** Isótopo de Carbono 14

## Sumário

1. Introdução .....	17
<b>1.1. Biodiversidade na região neotropical: diferentes biomas e fragmentação de habitat .....</b>	<b>17</b>
<b>1.2. Biogeografia e Filogeografia: passado e presente.....</b>	<b>30</b>
<b>1.3. Marcadores moleculares: do sequenciamento clássico à obtenção de dados genômicos.....</b>	<b>34</b>
<b>1.4. <i>Drosophila</i>: um organismo versátil .....</b>	<b>37</b>
2. Objetivos.....	42
<b>2.1. Objetivo geral.....</b>	<b>42</b>
<b>2.2. Objetivos específicos.....</b>	<b>42</b>
3. Materiais e Métodos.....	43
4. Resultados.....	43
<b>4.1. Capítulo: Signature of climatic differentiation on mitochondrial DNA of <i>Drosophila sturtevanti</i>.....</b>	<b>44</b>
5. Discussão.....	67
6. Conclusões.....	70
Referências .....	71
Apêndice A – Customized perl scripts used to IBE analysis.....	83
Apêndice B – Climatic variables and Distance matrices .....	94

## **1. Introdução**

### **1.1. Biodiversidade na região neotropical: diferentes biomas e fragmentação de habitat**

A região Neotropical é conhecida como uma das regiões mais diversas do globo e se estende do México ao sul da América do Sul, incluindo o Caribe e, sua fauna e flora ficaram quase completamente isoladas desde a separação da Gondwana, há 65 milhões de anos, até o soerguimento do istmo do Panamá, por volta de três milhões de anos atrás. (JOLY, 2008; MORRONE, 2014; CARVALHO; ALMEIDA, 2016). A definição formal da região Neotropical foi primeiramente introduzida por Sclater em 1858, ao dividir a Terra em seis regiões zoogeográficas, baseado na distribuição de aves e, em 1876, o termo também foi aceito por Wallace, que o aplicou a outros táxons de vertebrados terrestres e de água doce e publicou o primeiro mapa com a distribuição das regiões, delineadas pelo que hoje conhecemos como placas continentais (HOLT et al., 2013). Holt e colaboradores (2013), integrando dados de distribuição e relações filogenéticas de mais de 20 mil espécies entre anfíbios, aves e mamíferos não-marinhos, identificaram um total de 20 regiões zoogeográficas abrigadas em 11 regiões maiores (Figura 1). Mais recentemente, Morrone (2014) sugeriu a subdivisão do continente sul-americano em duas regiões, Andina e Neotropical, e uma zona de transição que apresenta elementos bióticos neotropicais e andinos de forma sobreposta (MORRONE, 2014; CARVALHO; ALMEIDA, 2016).

A região apresenta uma grande amplitude latitudinal, de 12° de latitude Norte a 56° de latitude Sul, estando, dessa forma, sujeita a uma grande variação climática. Esse atributo, associado à alta complexidade do terreno, resulta em um mosaico intrincado, incluindo diversas formações vegetais, com uma grande variedade de biomas e habitats que varia desde zonas áridas, regiões montanhosas e campos abertos a extensas áreas de floresta úmida, como a Mata Atlântica e a Amazônia (ANTONELLI et al., 2018).

**Figura 1.** Mapa das regiões zoogeográficas terrestres atualizadas por Holt et al., (2013). Em azul escuro encontra-se a região Neotropical, que se estende do México ao sul da América do Sul.



Fonte: HOLT et al., 2013.

Dessa forma, a alta diversidade encontrada na região tem sido alvo da curiosidade científica desde o século XIX, quando foi notada por naturalistas como Darwin e, as explicações propostas de sua origem tem sido controversas desde então (RULL, 2011). Com o aumento no número de estudos sobre a distribuição da biodiversidade na América do Sul foi possível o reconhecimento de padrões biogeográficos complexos e diversas hipóteses têm sido formuladas ao longo dos anos para explicar a origem de tal diversidade (HAFFER, 1969, 1993, 1997; HAFFER; PRANCE, 2002; CARNAVAL; MORITZ, 2008, CARVALHO; ALMEIDA, 2016). As hipóteses para explicar a alta diversidade e endemismo na região neotropical com frequência recorrem à heterogeneidade climática no espaço e tempo, geralmente associada à expectativa de que espécies tropicais tenham tolerâncias fisiológicas restritas. Essa heterogeneidade climática (histórica e contemporânea) afetaria diretamente a distribuição das linhagens por meio de restrições fisiológicas espécie-específicas e indiretamente por meio da aptidão a parasitas, patógenos, predadores e competidores (CARNAVAL et al., 2014). Outro padrão bem estabelecido é o Gradiente Latitudinal de Biodiversidade, caracterizado pelo declínio dos índices de biodiversidade em relação à latitude e existem diversas causas potenciais para sua ocorrência e, os modelos utilizam proposições baseadas nas taxas de origem e extinção de espécies na tentativa de compreender tal padrão (RULL, 2011). De acordo com a literatura, as baixas taxas de extinção resultantes de

uma maior estabilidade climática no hemisfério Sul poderiam ter contribuído para esse padrão permitindo que espécies pertencentes a clados ancestrais persistissem através do tempo (SANDEL et al., 2011; HOLT et al., 2013) Ainda assim, a questão permanece sem solução embora os testes de hipóteses estejam cada vez mais detalhados e forneçam cada vez mais informações sobre a complexidade da região (CARVALHO; ALMEIDA, 2016).

Outro componente importante a se ressaltar quando se estuda a diversidade na região Neotropical é a presença do istmo do Panamá. O istmo corresponde a uma via terrestre que conecta a América do Sul e as Américas Central e do Norte que foi formada ao longo de milhões de anos (Ma) (25 até cerca de 2,7 Ma atrás), quando finalmente se fechou completamente no Plioceno e possibilitou um intenso intercâmbio biótico (Grande Intercâmbio Americano) entre as porções de terra que se encontravam isoladas há milhões de anos. Assim, reforça-se a complexa formação histórica da América do Sul, moldada durante milhões de anos por elementos florísticos e faunísticos autóctones ou provenientes de outros continentes, ao longo de diferentes épocas e fases climáticas (O'DEA et al., 2016; CARVALHO; ALMEIDA, 2016).

Dentre os diferentes biomas encontrados na região Neotropical, um que merece lugar de destaque é a Mata Atlântica. A Mata Atlântica brasileira está entre os 25 *hotspots* de biodiversidade do mundo, abrigando uma grande diversidade de animais, plantas e tipos de habitats (MITTERMEIER et al., 2005; MYERS et al., 2000). Ocupa regiões costeiras do leste do Brasil atingindo sua extensão máxima no leste do Paraguai e nordeste da Argentina. Na porção meridional alcança a bacia do Rio Paraná e é limitada pela Caatinga no norte e pelo Cerrado na região central. Dada sua vasta extensão geográfica, apresenta uma grande diversidade florística, com várias fitofisionomias regionais, composta por formações florestais nativas (Figura 2a), a saber: Floresta Ombrófila Densa; Floresta Ombrófila Mista, também denominada de Mata de Araucárias; Floresta Ombrófila Aberta; Floresta Estacional Semidecidual; e Floresta Estacional Decidual), e ecossistemas associados (manguezais, vegetações de restingas, campos de altitude, brejos interioranos e encaves florestais do Nordeste) (OLIVEIRA-FILHO; FONTES, 2000; RIBEIRO et al., 2009; MINISTÉRIO DO MEIO AMBIENTE, 2019), além de uma grande variação na umidade e temperatura ao longo de sua extensão.

Embora haja alguma controvérsia quanto a sua idade exata, a Mata Atlântica é considerada a mais antiga floresta brasileira, sendo formada por uma mistura de espécies que evoluíram a partir da vegetação nativa do final do Cretáceo, era em que a América do Sul estava se separando da África, há 65 milhões de anos (RIZZINI, 1997; COLOMBO; JOLY, 2010).

A alta diversidade e distribuição da Mata Atlântica foram influenciadas por eventos históricos, como as mudanças climáticas ocasionadas pelos vários eventos glaciais que ocorreram durante os períodos Terciário e Quaternário. Essas alterações nas condições de temperatura e umidade podem ter sido o fator determinante na diferenciação e padrão de distribuição das espécies durante os eventos de expansão e retração da floresta (HAFFER, 1969, 1993). Essa teoria, denominada Teoria dos Refúgios Florestais, foi primeiramente sugerida para explicar a diversificação na Amazônia, mas tornou-se também o principal paradigma para explicar a alta diversidade encontrada na Mata Atlântica (HAFFER, 1969, 1997; PRANCE, 1982) o que também foi proposto por Carnaval e Moritz (2008) baseado em modelagem ecológica de nichos, que sugeriram que durante o Último Máximo Glacial a floresta úmida teria persistido nas porções norte e central do bioma (refúgios), enquanto a porção sul da Mata Atlântica deveria apresentar sinais de expansão demográfica como consequência da expansão pós-glacial nessa região.

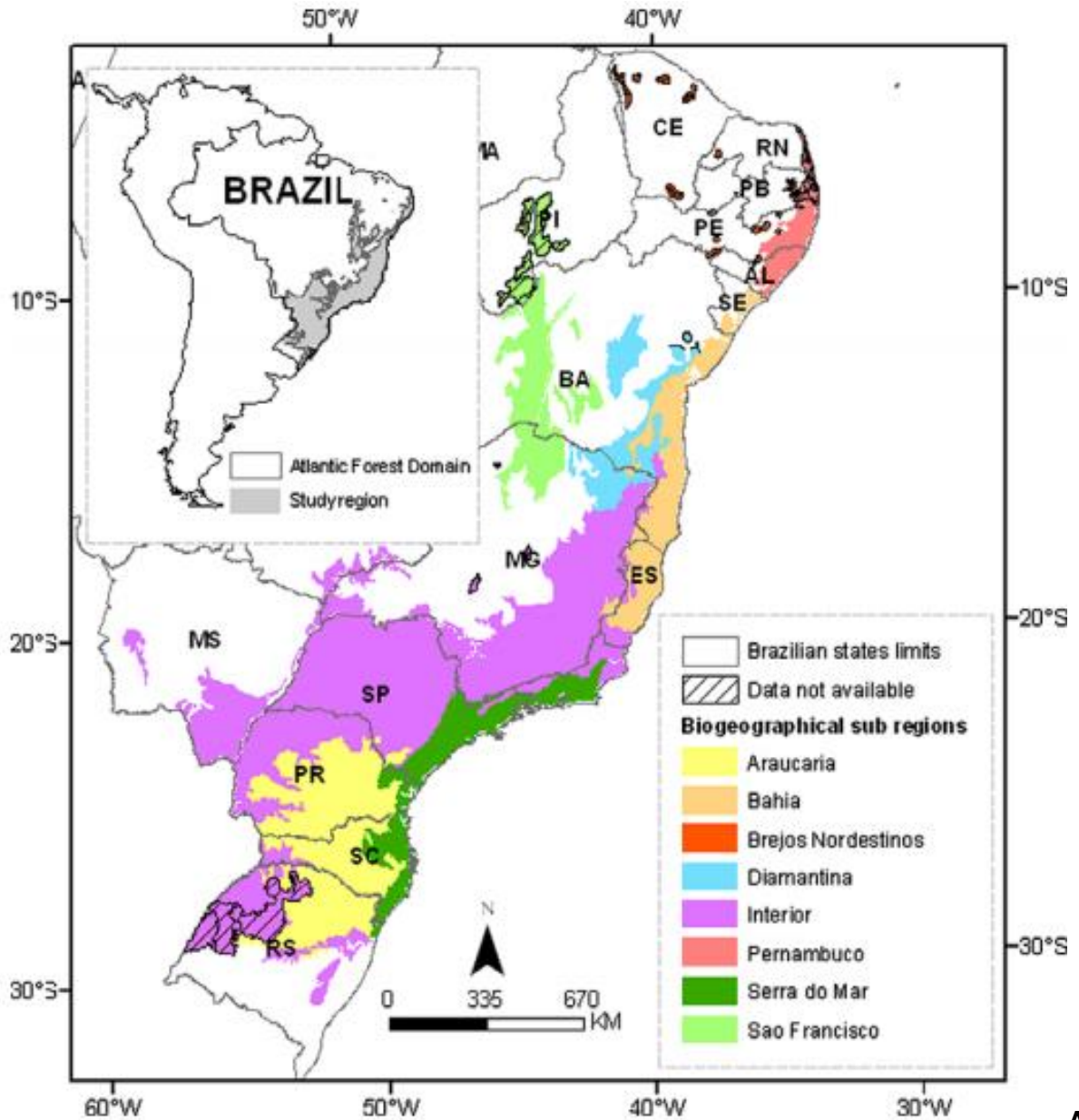
De acordo com Carnaval e Moritz (2008), os refúgios da Mata Atlântica resultantes das glaciações do Pleistoceno ficaram restritos ao limite norte deste bioma, principalmente no litoral, enquanto as florestas foram praticamente eliminadas nas latitudes mais altas. A floresta Atlântica do sul teria sido colonizada de refúgios do norte durante o Holoceno, por meio de expansões de populações do sul. No entanto, o trabalho de Stefenon et al., (2019) sugere a existência de refúgios glaciais do sul da Mata Atlântica, que originou os grupos de haplótipos de Araucárias do sul do Brasil, de áreas altas, independentemente dos grupos genéticos do norte, no Último Máximo Glacial, há 20 mil anos.

Dessa forma, embora haja trabalhos que corroborem a hipótese proposta por Carnaval e Moritz (2008), outros mostram diferentes padrões, o que indica que apesar de alguns organismos partilharem uma história comum de diversificação, parece haver um complexo conjunto de padrões de diversificação e processos

dentro da Mata Atlântica (BATALHA-FILHO et al., 2012; BATALHA-FILHO; MIYAKI, 2016). Assim, são necessários estudos com espécies endêmicas para se entender os mecanismos que levaram à diversificação que observamos atualmente (BATALHA-FILHO; MIYAKI, 2016; LEITE et al., 2016).

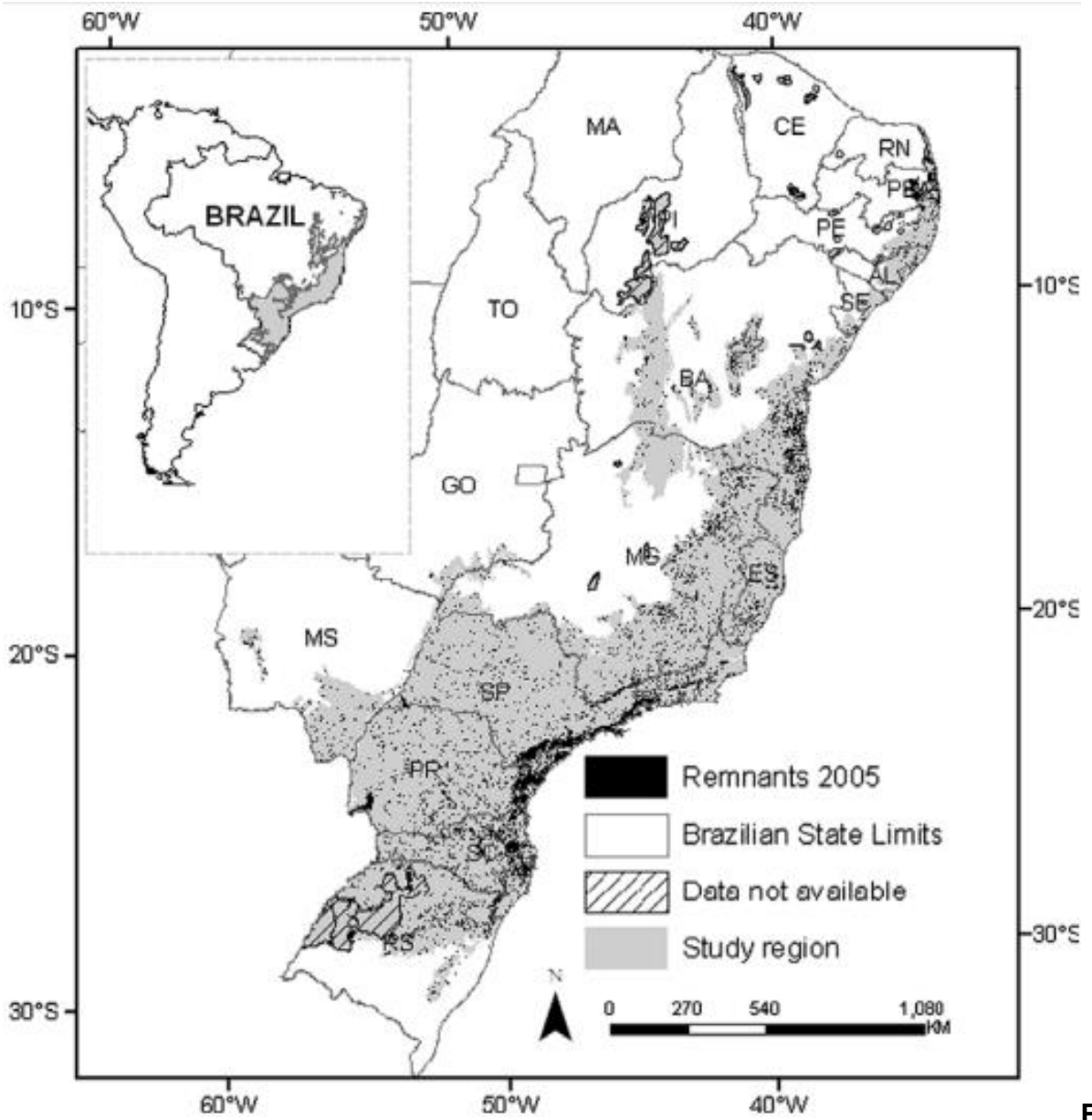
Antes da destruição e conseqüente fragmentação da Mata Atlântica, que teve início há cinco séculos com a chegada dos colonizadores europeus, a Mata Atlântica brasileira cobria 1,3 milhões de quilômetros quadrados, distribuídos em 17 estados do território nacional. Entretanto, devido à localização do bioma, que se estende pelas regiões mais habitadas do país, sofrendo forte pressão populacional e grande utilização de seus recursos, atualmente, a Mata Atlântica encontra-se altamente fragmentada e os remanescentes estão reduzidos a cerca de 22% da cobertura vegetal e desses, apenas 7% são fragmentos com mais de 100 hectares e em bom estado de conservação (Figura 2b) (RIBEIRO et al., 2009; COLOMBO; JOLY, 2010; MINISTÉRIO DO MEIO AMBIENTE, 2019). Estima-se que esse bioma abrigue cerca de 35% das espécies vegetais existentes no Brasil, incluindo diversas espécies endêmicas e ameaçadas de extinção, e sua riqueza de espécies é maior que a de alguns continentes.

**Figura 2A-B.** Mapas de distribuição do bioma da Mata Atlântica brasileira. **A.** Sub-regiões biogeográficas baseadas nas principais áreas de endemismo de aves, borboletas e primatas, como proposto por Silva e Casteleti (2003). **B.** Remanescentes florestais na Mata Atlântica brasileira.



A



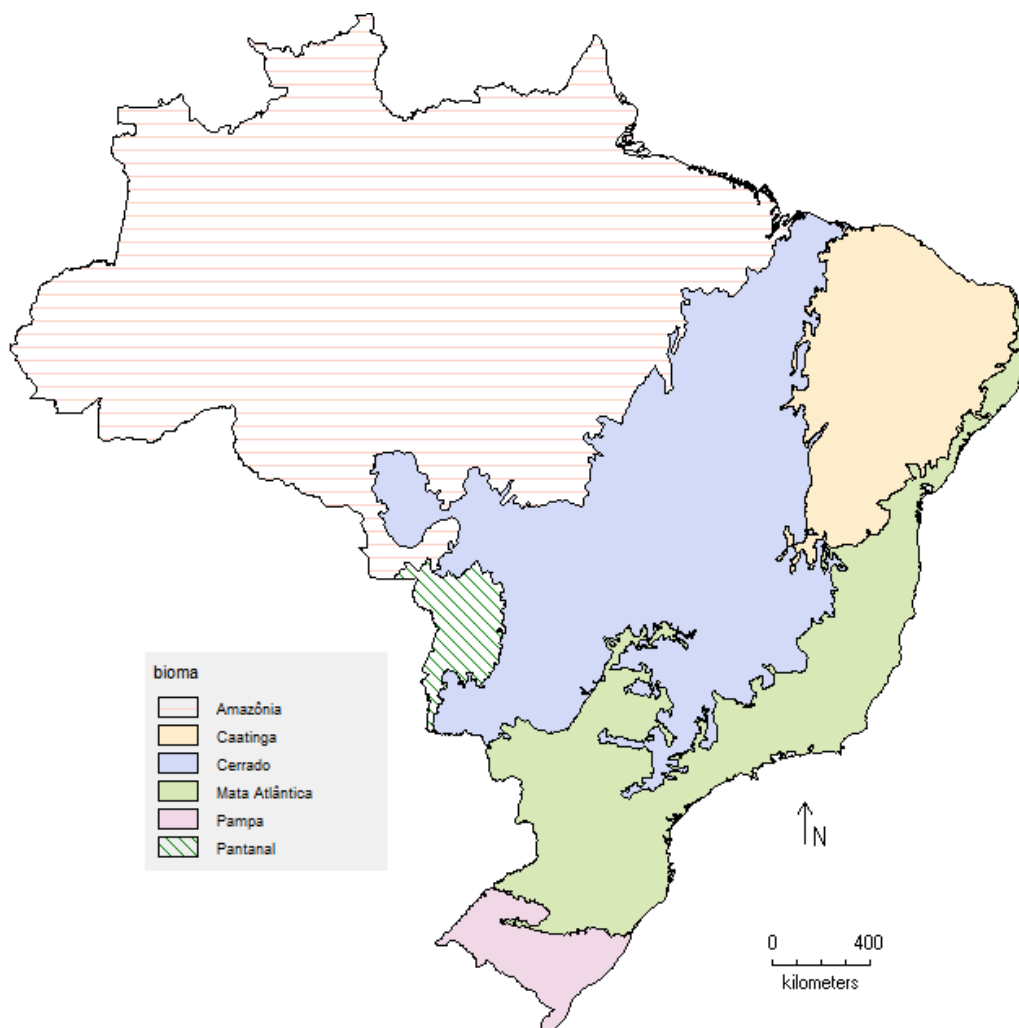


Fonte: Ribeiro et al., 2009.

A perda de habitat é a principal ameaça para as espécies que ocupam o bioma de Mata Atlântica (MYERS et al., 2000), sendo a conservação dos remanescentes a longo prazo uma tarefa desafiadora. Essa gestão dos remanescentes deveria priorizar a transformação dos mesmos em reservas de conservação, em que se reestabeleça a conectividade entre os fragmentos, além de aumentar a área das unidades de conservação para evitar os efeitos de borda (RIBEIRO et al., 2009).

Outros biomas de grande importância e representatividade que permeiam a região neotropical e fazem fronteira com a Mata Atlântica são o Cerrado, a Caatinga e o Pampa (Figura 3). Os dois primeiros são biomas de vegetação aberta que ocorrem sob as mesmas condições de estresse hídrico, apresentam vegetações adaptadas a essas condições climáticas, biotas únicas e distribuição complexa em forma de mosaicos e, tem recebido menos atenção científica que as florestas úmidas tropicais (WERNECK, 2011), enquanto o Pampa exibe paisagens naturais variadas, de serras a planícies, de morros rupestres a coxilhas, com o predomínio de campos nativos, sendo ainda um ambiente pouco explorado que só foi reconhecido como bioma em 2004 (MINISTÉRIO DO MEIO AMBIENTE, 2019).

**Figura 3.** Mapa dos Biomas Brasileiros. Em cores preenchidas estão os biomas que fazem fronteira com a Mata Atlântica.



Fonte: Elaborado pela autora.

O Cerrado, também conhecido como a Savana brasileira, é o segundo maior bioma na região Neotropical e também um importante *hotspot* de biodiversidade no mundo (MYERS, 2000). O bioma é composto por savanas e campos abertos em meio a áreas florestais secas e úmidas (DAMASCO et al., 2018) e está localizado na região central do Brasil, cobrindo uma extensão de aproximadamente dois milhões de quilômetros quadrados, sendo parte da “diagonal de formações abertas” (VANZOLINI, 1963), que separa a floresta Amazônica ao norte e a Mata Atlântica no sudeste, fazendo fronteira com a Caatinga no nordeste e com o Pantanal no sudoeste (WERNECK et al., 2012).

A alta biodiversidade e importância do Cerrado, a savana mais rica do mundo, pode ser destacada pelas mais de 12 mil espécies vegetais ali encontradas, sendo um terço endêmicas. Além disso, o bioma desempenha um importante papel na manutenção dos rios na América do Sul, sendo considerado o ‘berço das águas’ devido a grande quantidade de aquíferos existentes na região e ao fato que as raízes das plantas no Cerrado são mais profundas e responsáveis por absorver a água das chuvas e transportá-las para os aquíferos (MINISTÉRIO DO MEIO AMBIENTE, 2019; DAMASCO et al., 2018). Apesar da importância do bioma para a manutenção do potencial hídrico, nas últimas décadas, mais de 50% da vegetação original do Cerrado foi modificada e apenas 2,2% estão estritamente protegidas (KLINK; MACHADO, 2005), pesquisadores acreditam que a quantidade de água disponível nos aquíferos tenha atingido níveis mínimos, o que pode modificar toda a dinâmica climática na região e reforça a necessidade de proteção do Cerrado (DAMASCO et al., 2018).

Já a Caatinga ocupa uma área de cerca de 850 mil quilômetros quadrados, o equivalente a 11% do território nacional e é o maior bioma semiárido da América do Sul e o mais biodiverso do mundo (MINISTÉRIO DO MEIO AMBIENTE, 2019; WORLD WILD LIFE, 2019). Esse é também um dos menos conhecidos e menos conservados biomas brasileiros, com apenas 1,5% de sua extensão abrangida por unidades de proteção integral e mais de 80% de seus ecossistemas naturais alterados, sendo muitas vezes apontado como um bioma pobre quanto à biodiversidade e importância biológica (VANZOLINI et al., 1980; ANDRADE-LIMA, 1982; PRANCE, 1987, MINISTERIO DO MEIO AMBIENTE, 2019). No entanto, dados atuais indicam uma grande riqueza de ambientes e espécies, muitas das

quais são endêmicas da Caatinga, e também sua importância para a conservação da biodiversidade brasileira (LEAL et al., 2003, 2005). Outro fato importante a ser considerado é que a conservação da Caatinga está estreitamente relacionada ao combate à desertificação. No Brasil, aproximadamente 95% das áreas susceptíveis à desertificação estão no território da Caatinga e grande parte já está em estado avançado de degradação, o que torna ainda mais urgente a conscientização sobre a importância do bioma.

Já o Pampa ocupa uma área de 176.496 quilômetros quadrados do território brasileiro onde está restrito ao estado do Rio Grande do Sul e, uma área total de 700 mil quilômetros quadrados, dividida em quatro países - Argentina, Brasil, Paraguai e Uruguai (MINISTÉRIO DO MEIO AMBIENTE, 2019). Devido à sua formação geológica, em muitas regiões do bioma é verificado solo arenoso, frágil e sensível à erosão fluvial e eólica (ROESCH et al., 2009). Ele abriga uma grande e rica biodiversidade, com flora e fauna próprias e, estimativas indicam valores em torno de 3000 espécies vegetais, mais de 100 mamíferos e quase 500 espécies de aves (MINISTÉRIO DO MEIO AMBIENTE, 2019).

Dada a importância dos fragmentos florestais na manutenção da biodiversidade, segue uma síntese sobre algumas das áreas de coleta amostradas neste estudo, obtida por meio de dados oficiais ou pesquisas realizadas nas áreas.

- A Reserva Biológica Tirimbina (TIR) é um refúgio de vida selvagem de 345 hectares localizado em La Virgen de Sarapiquí, província de Heredia, na bacia do rio Sarapiquí, na Costa Rica, formada por floresta tropical úmida, sendo uma das mais biologicamente diversas e relativamente bem estudadas áreas do país (ANGULO et al., 2017).
- A Área de Relevante Interesse Ecológico do Sítio Curió (CUR) abriga 57,35 hectares, sendo o último enclave de Mata Atlântica na zona urbana de Fortaleza, no estado do Ceará. Foi criada por meio do Decreto Estadual nº 28.333, de 28 de julho de 2006, com o objetivo de preservar a diversidade biológica e o meio ambiente especialmente em áreas da Região Metropolitana, ainda dotadas de valiosa biodiversidade de flora e fauna (SEMACE, 2019).

- A Reserva Biológica Guaribas (GUA) está localizada na Mesorregião da Mata Paraibana e na Microrregião do Litoral norte, sendo a região mais úmida do litoral do estado. Engloba três áreas descontínuas, denominadas SEMA 01, SEMA 02 e SEMA 03, abrangendo cinco municípios. A Reserva constitui um dos últimos remanescentes de Floresta Atlântica do Estado da Paraíba, abrigando espécies raras, endêmicas e ameaçadas de extinção e, de acordo com o Plano de Manejo, Fase 2, da Reserva Biológica Guaribas (IBAMA/MMA, 2019), as formações vegetais encontradas na REBIO Guaribas apresentam forte ligação com a Caatinga, sendo que nas florestas estacionais da Reserva encontram-se gêneros característicos da Caatinga e elementos florísticos das florestas estacionais encontradas na área “core” do Cerrado, devido à localização da reserva em uma região limitada pela Caatinga e com a presença de formações savânicas.
- A Área de Proteção Ambiental do Pratigi (PRA) é dividida em três ecopolos e está inserida no Mosaico de APAS do Baixo Sul da Bahia, com uma área de mais de 85 mil hectares. Estende-se por quatro municípios e é formada por remanescentes de Floresta Ombrófila Densa (FOD) e formações pioneiras, estando entre as áreas de maior prioridade para a conservação da biodiversidade do Corredor Central da Mata Atlântica (INEMA, 2019).
- O Parque Estadual da Fonte Grande, em Vitória, no estado do Espírito Santo (VIT) é a última área contígua de grande porte com vegetação característica de encostas da Mata Atlântica. Apresenta uma área de 217 hectares em um relevo acidentado, incluindo vales e pontões (PREFEITURA DE VITÓRIA, 2019).
- O Núcleo Picinguaba (PIC) é uma área de conservação situada em Ubatuba, no estado de São Paulo. É um dos dez núcleos do Parque Estadual da Serra do Mar, criado em 1979, e possui área de 47 mil hectares. Além da área florestal, o núcleo conta com restingas, manguezais, praias e costões rochosos que passaram a fazer parte do cenário protegido do Parque Estadual Serra do Mar (SÃO PAULO, 2019).

- O Parque da Cantareira (CANT) está localizado na cidade de São Paulo e possui uma das maiores áreas de mata tropical nativa do mundo situada em uma região metropolitana. Seus 7.900 hectares são formados por remanescentes, que abrigam diversas espécies ameaçadas de Mata Atlântica, fazendo parte do Cinturão Verde da cidade de São Paulo, que está incluído na Reserva da Biosfera (SÃO PAULO, 2019).
- A Fazenda Cambuhy, no município de Matão (MAT) é um dos remanescentes florestais amostrados na região noroeste do estado de São Paulo. O fragmento conta com 2189,6 hectares de Floresta Estacional Semidecidual (FES) em diferentes estágios de sucessão ecológica, tendo como matriz adjacente ao fragmento, o plantio de café e laranja (NECCHI JUNIOR, 2012).
- A Fazenda São João, no município de Nova Granada (NG), assim como o fragmento de MAT, localiza-se no noroeste do estado de São Paulo, sendo um dos fragmentos aqui amostrados. Conta com uma área de 1796,9 hectares de Floresta Estadual Semidecidual (FES), também em diferentes estágios de sucessão e apresenta como matriz adjacente o plantio de laranja (NECCHI JUNIOR, 2012).
- O município de Dourados (DOU), no estado do Mato Grosso do Sul ainda mantém 2% dos remanescentes da Mata Atlântica brasileira. Atualmente, são 42 hectares de mata, mas de acordo com dados do governo a vegetação já chegou a ocupar 31 mil quilômetros quadrados no estado, que começou a ser desmatado na década de 70 com a expansão da fronteira agrícola para o cultivo de soja. A fitofisionomia encontrada na região é de Floresta Estacional Semidecidual (FES) e representa um ecótono com o Cerrado (PADOVAN et al., 2018).
- A Reserva Ecológica do IBGE, localizada em Brasília (BRA), no Distrito Federal, é uma área protegida de interesse científico sob a gestão do Instituto Brasileiro de Geografia e Estatística (IBGE). Apresenta uma área de 1.391,25 hectares e abriga uma grande heterogeneidade de ecossistemas. A área está inserida no bioma do Cerrado, contando com mais de 4000 espécies

inventariadas, com ocorrência de 55 espécies ameaçadas de extinção e 91 espécies endêmicas do Cerrado (RECOR, 2019).

- A Reserva Betary (BET) situa-se na região com a maior superfície contínua de Mata Atlântica do país, contando com uma área de 60 hectares. A bacia do rio Ribeira de Iguape, também conhecida como Vale do Ribeira, fica ao sul do estado de São Paulo, na divisa com o Paraná. Os principais maciços florestais nativos estão protegidos por grandes unidades de conservação, como o Parque Estadual Turístico do Alto do Ribeira (Petar), conhecido por suas enormes cavernas, o Parque Estadual de Jacupiranga, entre outros (PREFEITURA DE IPORANGA, 2019; IPBio, 2019).
- A região do Piraí (PI) localiza-se no município de Joinville, no estado de Santa Catarina e abriga remanescentes de Mata Atlântica. A região está inserida na Área de Proteção Ambiental (APA) Serra Dona Francisca, que conta com uma área total de mais de 40 mil hectares, englobando os mananciais dos rios Cubatão e Piraí. A APA está situada na Zona Núcleo e de Amortecimento da Reserva da Biosfera da Mata Atlântica, integrando um dos mais importantes corredores ecológicos que conectam Unidades de Conservação ao longo das serras litorâneas brasileiras, sendo de ocorrência na região do Piraí a Floresta Ombrófila Densa (FOD) (JOINVILLE, 2019).
- O Ribeirão da Ilha (RI) é um bairro localizado ao sul do município de Florianópolis, em Santa Catarina em que está presente a vegetação de Mata Atlântica. Por estar afastado do centro da cidade, o bairro funciona como um bairro-dormitório, onde não existem prédios e as residências ainda mantém a característica do período colonial e como a Constituição de 1988 passou a considerar a vegetação de Mata Atlântica patrimônio nacional em toda a costa brasileira, a ilha conta com vários parques distribuídos ao longo de sua extensão, de forma que existem vários fragmentos da Mata Atlântica espalhados nos 433 quilômetros quadrados da ilha, que correspondem a 12,7% da sua cobertura original (GUIA FLORIPA, 2019).
- A Reserva Biológica Estadual do Aguaí (AG) está localizada nos contrafortes da Serra Geral, em altitudes que variam de 200 a 1470 metros, e abrange os

municípios de Morro Grande, Nova Veneza, Siderópolis e Treviso, no estado de Santa Catarina. Foi criada em 1º de julho de 1983, por meio do Decreto nº 19.635, protege uma área de 7.672 hectares e está inserida no bioma da Mata Atlântica. A criação da Reserva Biológica do Aguai justificou-se pelo seu relevo acidentado, a presença de diversos *canyons*, pela riqueza de ecossistemas e pela grande variedade de espécies de plantas e animais, que fazem da região um cenário valioso para a conservação da biodiversidade (FATMA SC, 2019).

- Eldorado do Sul (ELD) é um município na região metropolitana de Porto Alegre, no Rio Grande do Sul e integra a área de preservação ambiental do Delta do Jacuí, onde é encontrado um sistema de transição entre os biomas da Mata Atlântica e do Pampa (PREFEITURA DE ELDORADO, 2019; SEMA RS, 2019).

## **1.2. Biogeografia e Filogeografia: passado e presente**

O efeito da fragmentação de habitats sobre as espécies que nele habitam pode ser avaliado por meio da relação entre os dados genéticos e o contexto ecológico, buscando compreender os padrões e mecanismos evolutivos que conectam a genética do organismo e o ambiente no qual os mesmos estão inseridos (TIDON, 2006). A importância dessa correlação genética/ambiente pode ainda ser destacada pelo fato de que o potencial evolutivo das populações de áreas fragmentadas é comprometido devido à perda de variabilidade genética pela redução demográfica da população. A diminuição do tamanho populacional torna as espécies mais susceptíveis à depressão endogâmica e leva à redução da variação intrapopulacional pela ação da deriva genética e do fluxo gênico, o que aumenta a diferenciação genética entre diferentes populações e os efeitos associados ao pequeno tamanho populacional persistirão mesmo se a causa da fragmentação for solucionada (BROWN; KODRIC-BROWN, 1977; TEMPLETON et al., 1990, 2001; RATHCKE; JULES, 1993; FRANKHAM, 2003). Dessa forma, torna-se indispensável



o entendimento sobre os fatores históricos e contemporâneos que levaram à distribuição atual das espécies e da variabilidade genética.

Em vista disso, a fim de identificar e compreender as relações existentes entre os fatores responsáveis pela distribuição das espécies e sua variabilidade genética, a abordagem proveniente de áreas da ciência tais como a Biogeografia, a Genética de Populações, a Filogenia e a Ecologia, aplicadas de forma integrada, podem fornecer uma fonte confiável de informações que ajudariam na resolução de questões biológicas como a acima citada.

A biogeografia é a ciência que estuda a distribuição geográfica dos seres vivos no espaço através do tempo buscando compreender os padrões atuais de distribuição e os processos que geraram tais padrões. Para isso, existem três componentes principais que devem ser avaliados: o espaço, que corresponde à área geográfica na qual os organismos estão distribuídos, o tempo, no qual se inserem todos os eventos históricos que influenciaram os padrões atuais e a forma, que corresponde ao grupo de organismos a ser estudados. Assim, a Biogeografia é uma disciplina complexa que integra diversas ciências, tais como a Geografia, Geologia e Ecologia e, que vêm progredindo e sendo cada vez mais utilizada no cenário atual (CRISCI, 2001; CRISCI et al., 2003; GILLUNG, 2011).

A biogeografia fornece a perspectiva histórica necessária tanto para a compreensão da evolução das biotas como também da evolução geológica do planeta. O desenvolvimento da biogeografia enquanto ciência ao longo dos últimos três séculos apresenta uma forte relação com o desenvolvimento e refinamento de sistemas classificatórios estáveis em paralelo ao aumento do conhecimento da distribuição geográfica das espécies e da geologia do planeta (CARVALHO; ALMEIDA, 2016). A obtenção de informações sobre a distribuição dos táxons possibilitou a formulação de questões cruciais, como a descoberta de padrões biogeográficos e a regionalização do espaço, ainda assim, estudar o resultado de eventos históricos é uma tarefa desafiadora (CRISCI et al., 2003; CARVALHO; ALMEIDA, 2016). Existem inúmeros fatores, históricos e contemporâneos responsáveis pela presença de uma espécie ou comunidade em determinada área, o que torna a reconstrução complexa, exigindo muito mais que um simples método analítico. Nesse contexto, a abordagem biogeográfica difere das ciências

experimentais, pois volta-se para o passado na tentativa de estabelecer conexões entre eventos cujas situações não podem ser repetidas experimentalmente. Devido à natureza dos eventos biológicos e geológicos, existe um processo de eliminação de evidências, tornando ainda mais árdua a compreensão dos processos que originaram determinados padrões (CARVALHO; ALMEIDA, 2016).

Assim, dois processos importantes para a compreensão da história e biogeografia são a dispersão e a vicariância. Esses processos são responsáveis por moldar os padrões de distribuição dos organismos, sendo o primeiro o resultado do deslocamento e conquista de novas áreas, ultrapassando barreiras pré-existentes, de forma que as duas populações, isoladas pela barreira, se diferenciam e podem se tornar diferentes espécies que não mais se reconhecem em caso de novo contato, devido à restrição ao fluxo gênico e acúmulo de diferenças em cada um dos grupos. Já na vicariância, os indivíduos de uma população ancestral são isolados por meio do surgimento de uma barreira, seja ela geográfica ou ecológica, que também pode levar à diferenciação e surgimento de duas populações que, embora compartilhem um ancestral comum, não mais se reconhecem como um único deme. O problema quando se compara a distribuição atual de um grupo separado por uma barreira geográfica é determinar o momento do surgimento da barreira em relação ao evento de divergência e à origem do grupo (CARVALHO; ALMEIDA, 2016).

Praticamente todas as espécies apresentam certo grau de estrutura populacional. A estrutura consiste em um conjunto de características genéticas e demográficas resultantes da ação e interações entre uma série de mecanismos evolutivos e ecológicos (MARTINS, 1987). A caracterização da estrutura populacional pode ser obtida por meio de diversos parâmetros ecológicos e/ou genéticos. A abordagem genética e evolutiva busca quantificar a variabilidade existente entre os indivíduos, a fim de compreender quais fatores são responsáveis pelo padrão observado. Nesse contexto, a Genética de Populações é o ramo da Genética que investiga a dinâmica dos genes nas populações naturais, buscando elucidar quais os mecanismos alteram sua composição gênica ou as frequências genotípicas pelo aumento de homozigose (BEIGUELMAN, 2008).

A Filogeografia, por sua vez, é a área capaz de fazer a distinção entre os fatores evolutivos recorrentes e históricos responsáveis pela distribuição

demográfica das espécies e, para isso, busca integrar a Filogenia e a Genética de Populações, a fim de investigar as relações entre os processos macro e microevolutivos e compreender como os eventos históricos ajudaram a formar a distribuição geográfica atual de genes, populações e espécies (AVISE, 2000, 2009). O contexto geográfico da análise filogeográfica é examinado pelos processos históricos que podem ter influenciado a demografia populacional, diferenciação e especiação (KINGMAN, 1982; FU; LI, 1999; AVISE, 2000, 2004, 2009; BEHEREGARAY, 2008).

O poder de resolução desta área se deve a investigação tanto dos componentes espaciais quanto temporais envolvidos no processo de coalescência, que é baseada em um complexo conjunto de modelos matemáticos que acomodam fenômenos biológicos que refletem dados genômicos e identificam uma genealogia ancestral por meio de uma amostra contemporânea (KINGMAN, 1982). A teoria da coalescência, por sua vez, é baseada em um complexo conjunto de modelos matemáticos que acomodam fenômenos biológicos que refletem dados genômicos e identificam uma genealogia ancestral por meio de uma amostra contemporânea (KINGMAN, 1982).

A filogeografia pode ser separada em duas partes. Uma baseada em demonstrações gráficas, árvores explicativas da hipótese da história biogeográfica dos organismos, sustentada por métodos de reconstrução filogenética, rede de haplótipos e métodos de coalescência e, a segunda, fundamentada na estatística e na matemática para elucidar a demografia e estrutura populacional, baseada na teoria da coalescência, em fenômenos estocásticos e estatística computacional, em que os modelos de análises utilizam métodos fundamentados na verossimilhança, particularmente por inferência Bayesiana (KNOWLES; MADDISON, 2002; NIELSEN; BEAUMONT, 2009).

O termo Filogeografia foi, pela primeira vez, cunhado por John Avise e colaboradores em 1987 para descrever uma disciplina com raízes técnicas e conceituais ligadas ao campo ainda em fase inicial da genética molecular nos anos 70 e, desde então, o número de estudos filogeográficos aumentou significativamente nos últimos 30 anos (BEHEREGARAY, 2008).

Apesar do crescente número de trabalhos com abordagens filogeográficas e da grande necessidade de estudos com espécies endêmicas, ainda é uma tarefa desafiadora trabalhar com filogeografia em países do hemisfério sul, como é o caso do Brasil. Beheregaray, em seu artigo de revisão de 2008, aponta o crescente número de estudos na área 20 anos depois do seu surgimento, mas também nos mostra que a maior parte dos artigos publicados em filogeografia está em países do hemisfério norte, o que contrasta com a escassez de estudos em ambientes megadiversos, como é o caso da Mata Atlântica. Esses dados, embora sejam de 2008, ainda hoje refletem nossa realidade e reforçam a necessidade de estudos como o aqui proposto.

### **1.3. Marcadores moleculares: do sequenciamento clássico à obtenção de dados genômicos**

O surgimento, no final da década de 70, do método de Sanger, conhecido como método de terminação da cadeia, permitiu a determinação das sequências de DNA e foi essencial para o surgimento da era genômica da Biologia, sendo o único método amplamente utilizado durante os 30 anos seguintes (SANGER et al., 1977; SCHUSTER, 2007). Esse período foi marcado por grande desenvolvimento técnico e tecnológico, o que permitiu o sequenciamento de genomas completos de diversos organismos (SCHUSTER, 2007).

Porém, mesmo com vários aperfeiçoamentos realizados durante esse período, as limitações do método automatizado de Sanger mostravam a necessidade de técnicas de sequenciamento mais rápidos, econômicos e eficientes, o que culminou no surgimento do chamado Sequenciamento de Nova Geração (do inglês, *New Generation Sequencing*, NGS). Vale ressaltar que o método automatizado de Sanger é considerado a tecnologia de “primeira geração” e, todos os novos métodos constituem o chamado NGS. Essas novas tecnologias compõem várias estratégias que contam com a combinação da preparação de moldes, sequenciamento, visualização e métodos de alinhamento e montagem (METKZER, 2010). Assim, em 2005, surgiu o que seria um marco no sequenciamento em larga escala capaz de gerar um volume muito maior de dados, com o desenvolvimento

baseado na tecnologia do sequenciamento por síntese, desenvolvido pela *454 Life Sciences* (MARGULIES et al., 2005) que usava como estratégia uma enorme redução no volume das reações, enquanto aumentava o número de reações por meio da utilização de várias centenas de milhares de moldes em cada placa, de forma que elas podiam ser analisadas em paralelo, um enorme avanço em relação ao método de Sanger, que permitia a análise de no máximo 96 amostras por vez. Hoje, já na Quarta Geração das tecnologias de Sequenciamento, dentre toda a gama de tecnologias disponíveis, é possível o sequenciamento direto de uma molécula inteira de DNA, em escala de tempo real por meio da tecnologia Nanopore (FENG et al., 2015).

Dessa forma, com o imenso volume de dados, se tornava necessário o desenvolvimento de algoritmos e sistemas computacionais capazes de comportar e analisar tais resultados. Nesse aspecto, a bioinformática apresenta papel fundamental na obtenção da sequência completa de DNA e na transformação da informação genética em conhecimento aplicável.

Mas tão importante quanto a metodologia para obtenção e análise do DNA é a definição dos melhores genes e *primers* para o objetivo que se deseja alcançar. Os marcadores moleculares são as ferramentas mais adequadas para qualificar e quantificar a variabilidade genética das populações naturais, assim como para determinar as possíveis causas da distribuição desta variação. O DNA mitocondrial representa apenas uma pequena fração do genoma total de um indivíduo e, ainda assim, tem sido o marcador mais popular e amplamente utilizado para avaliar a diversidade molecular animal nas últimas trinta décadas e, todo estudo molecular na área de genética e sistemática molecular emprega seu uso em algum estágio (GALTIER et al., 2009). As razões para sua alta aceitabilidade e escolha devem se ao fato de que, experimentalmente, o DNA mitocondrial é relativamente fácil de amplificar, pois apresenta múltiplas cópias na célula. Além disso, seu conteúdo é fortemente conservado entre os animais, com muito poucas duplicações, ausência de íntrons e, regiões intergênicas muito curtas (GISSI et al., 2008). Devido à sua alta taxa de mutação, o DNA mitocondrial é altamente variável em populações naturais, o que pode gerar algum sinal sobre a história populacional em curtos intervalos de tempo e, em contrapartida, as regiões variáveis são geralmente flanqueadas por regiões altamente conservadas, o que permite o desenho de *primers* e sua

consequente amplificação (GALTIER et al., 2009). Outra característica a ser considerada é quanto à neutralidade do marcador. Marcadores mitocondriais têm sido considerados, por muito tempo, marcadores evolutivos neutros, entretanto, um número cada vez maior de estudos tem sugerido que uma quantidade considerável de variação genética dentro do genoma mitocondrial é sensível à seleção natural e exerce fortes efeitos sobre o fenótipo, o que levou à hipótese de “adaptação climática mitocondrial”, proposta em 2017 por Camus e colaboradores. Dados obtidos por meio de um experimento de evolução experimental por Lajbner e colaboradores (2018) para testar diretamente essa hipótese demonstraram evidência experimental de que a seleção termal estaria moldando a evolução do genoma mitocondrial e sugerem que a distribuição da variação mitocondrial em populações naturais de metazoários pode refletir adaptação ao ambiente climático.

Nesse contexto, e dadas todas as características acima apresentadas, o DNA mitocondrial tem se mostrado um instrumento muito útil na reconstrução de filogenias e dos eventos da história demográfica, devido à sua capacidade de caracterizar a variabilidade inter e intraespecífica; poder identificar espécies e populações; inferir relações filogenéticas (o que pode esclarecer possíveis eventos históricos); e principalmente pela possibilidade de simular cenários geológicos, padrões de fluxo gênico histórico, isolamento histórico da diversidade genética, eventos de expansão demográfica e contração populacional (AVISE et al., 1984). Todos estes aspectos, que podem ser abordados utilizando este marcador, devem-se ao fato de sua análise ser realizada pela reconstrução da linhagem materna, e também pela taxa de evolução mais rápida que sequências nucleares, devido, entre outros fatores, ao tempo de coalescência mais rápido (BIRKY et al., 1983; TAJIMA, 1989; AVISE; BALL, 1990; AVISE, 2000; HOFFMAN et al., 2009).

Além de todas as características e aplicabilidades de marcadores mitocondriais no geral, o uso dos genes mitocondriais que codificam para as subunidades I e II do gene Citocromo Oxidase (*COI* e *COII*, respectivamente) apresentam algumas características extras, que os tornam particularmente adequados para aplicações em estudos evolutivos. Por se tratar de uma enzima chave no metabolismo aeróbico celular, ambas as subunidades têm sido estudadas quanto a características bioquímicas e, seu tamanho e estruturas parecem ser conservados ao longo dos organismos investigados (SARASTE, 1990; LIU;

BECKENBACH, 1992; LUNT et al., 1996). A subunidade I é a maior das três subunidades que compõem o gene, seguida pela subunidade II e é o mais conservado em termos de evolução em nível de aminoácidos (SIMON et al., 1994), além de ser considerado um dos genes relevantes para identificação de animais, chamado de códigos de barras genético (HEBERT et al., 2003).

#### **1.4. *Drosophila*: um organismo versátil**

A família Drosophilidae foi um dos primeiros modelos a serem utilizados em estudos genéticos, servindo de base para o desenvolvimento do conhecimento nas mais diversas áreas. Moscas do gênero *Drosophila*, em especial *Drosophila melanogaster*, têm sido empregadas nos mais diversos tipos de experimentos no último século, devido a diversas características que as tornam um excelente organismo modelo, como o fato de serem facilmente manejadas em laboratório, apresentarem grande número de descendentes e um curto tempo de geração. Além disso, o genoma é compacto e de fácil manipulação genética, o que resultou em *D. melanogaster* sendo o segundo genoma multicelular sequenciado (ADAMS et al., 2000) e, desde então, o número de genomas disponíveis nos bancos de dados aumentou para a casa das centenas de indivíduos, provenientes de diferentes regiões geográficas na última década além do sequenciamento do genoma de outras 11 espécies de *Drosophila* por meio do *Drosophila 12 Genomes Consortium* et al., (2007) (HALES et al., 2015).

A família Drosophilidae inclui atualmente cerca de 4000 espécies válidas (TIDON; ALMEIDA, 2016) distribuídas em mais de 70 gêneros, dos quais 16 gêneros e 305 espécies estão representados no Brasil (O'GRADY; DE SALLE, 2018) e apesar dessa grande diversidade de espécies, poucos são os trabalhos disponíveis que usam esses organismos para estudos filogeográficos. Espécies cactofílicas do grupo *repleta* foram avaliadas para o gene *COI* em um amplo estudo de Franco e Manfrin (2013), no qual os resultados obtidos sugerem uma congruência entre as flutuações no tamanho populacional e as flutuações paleoclimáticas ocorridas no Quaternário. Outros trabalhos utilizando *D. maculifrons* (CENZI DE RÉ et al., 2014) e *D. ornatifrons* (GUSTANI et al., 2015), duas espécies dos grupos *guaramunu* e

*guarani*, respectivamente, também foram realizados, nos quais se verificou expansão populacional após o Último Máximo Glacial, o que corrobora o padrão encontrado para outros grupos. Barrios-Leal e colaboradores (2018a, 2018b) utilizando espécies do grupo *repleta* também estudaram a estrutura populacional e história evolutiva, encontrando também a relação dos eventos climáticos ocorridos durante o Pleistoceno e Plioceno na diversificação das mesmas, reforçando assim, a importância do uso de espécies neotropicais do gênero *Drosophila* como organismos modelo para estudos filogeográficos (MANFRIN; SENE, 2006).

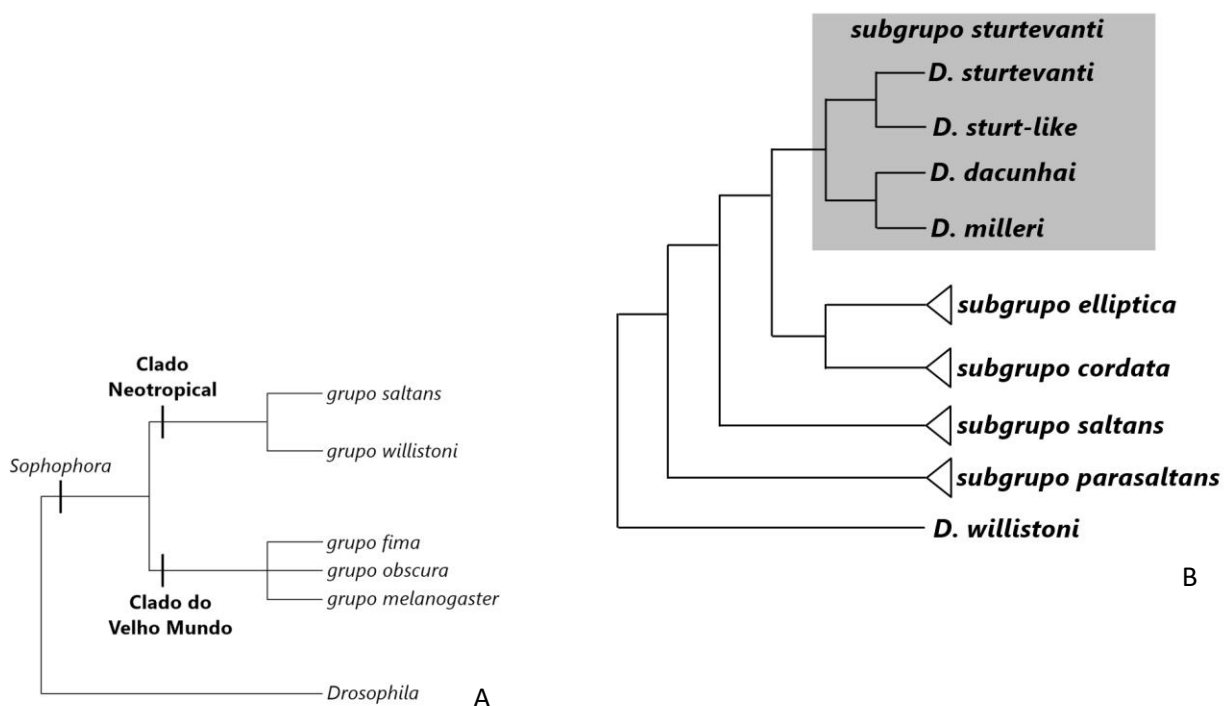
Um dos grupos do gênero *Drosophila* é o *saltans*, que está inserido no subgênero *Sophophora*, sendo um dos sete grupos que o compõe, destacando-se os grupos *melanogaster*, *obscura*, *willistoni* e *saltans* (Figura 4-A). O grupo de espécies de *saltans* e *willistoni* representa o clado neotropical do subgênero *Sophophora*, sendo proximamente relacionadas entre si e mais distantes do restante dos grupos de *Sophophora* (THROCKMORTON, 1975; O'GRADY; DESALLE, 2018; O'GRADY; KIDWELL, 2002). O grupo *saltans* (Figura 4-B) é composto por 21 espécies descritas até o momento divididas em cinco subgrupos: *cordata*, *elliptica*, *sturtevantii*, *parasaltans* e *saltans* (MAGALHÃES; BJÖRNBERG, 1957, MAGALHÃES, 1962) e, apresenta uma grande quantidade de incongruências em sua filogenia, de acordo com a literatura. O primeiro trabalho publicado sobre a filogenia do grupo foi realizado por Throckmorton e Magalhães (1962), que executaram a primeira análise filogenética do grupo *saltans* por meio de comparações anatômicas internas e externas, colocando o subgrupo *cordata* como o mais basal, seguido por *elliptica*, *sturtevantii*, *parasaltans* e *saltans*. Com base neste estudo, Throckmorton (1975) propôs que o grupo se originou na América do Norte, devido à localização atual do subgrupo mais primitivo e, a partir daí os subgrupos *sturtevantii*, *parasaltans* e *saltans* teriam se diversificado na América do Sul antes da formação do istmo do Panamá.

O subgrupo *sturtevantii* abriga, segundo a literatura, cinco espécies: *D. sturtevantii*, *D. milleri*, *D. dacunhai*, *D. magalhaesi* e *D. rectangularis* (MAGALHÃES, 1962; MOURÃO; BICUDO, 1967). Entretanto, a inserção de *D. rectangularis* nesse subgrupo parece ser equivocada, pois pela análise de caracteres morfológicos (como por exemplo, Índice Costal e edeago) esta espécie difere das demais pertencentes a este subgrupo (comunicação pessoal, Lilian Madi-Ravazzi). Além



disso, o grupo conta com uma nova espécie a ser descrita, proveniente da Guiana Francesa e denominada *D. sturt-like* (MADI-RAVAZZI et al., em preparação).

**Figura 4A-B.** **A.** Relação evolutiva proposta dentro do subgênero *Sophophora*, adaptado de O’Grady e Kidwell (2002) e, em **B.** relação evolutiva dentro do grupo *saltans*, como enfoque no subgrupo *sturtevantii*, baseada nos resultados obtidos por Roman (2018) por meio de uma análise particionada por Inferência Bayesiana.



Fonte: Elaborado pela autora.

*Drosophila sturtevantii* (Figura 5) pertence ao subgrupo *sturtevantii* e é a espécie com distribuição geográfica mais abrangente dentro do subgrupo *sturtevantii* e mesmo dentro do grupo *saltans* (MAGALHÃES, 1962; BICUDO, 1979). A espécie está incluída em quase toda a área de distribuição do grupo, do México ao sul do Brasil e nas ilhas Caribenhas e, Hosaki-Kobayashi e Bicudo (1997), baseando-se no estudo de inversões cromossômicas em populações naturais recentemente coletadas e linhagens da espécie mantidas em laboratório, sugeriram que a diferenciação de *D. sturtevantii* provavelmente se deu na região amazônica devido ao alto polimorfismo de inversões cromossômicas encontrado na região, que

apresenta uma grande diversidade ecológica. Além disso, Carareto (1994) e Hosaki-Kobayashi e Bicudo (1998) observaram incompatibilidades reprodutivas entre diferentes linhagens de *D. sturtevantii* provenientes do Brasil e outros países das Américas Central e do Sul, o que demonstra diferentes graus de isolamento e é indicativo que essas populações podem apresentar mais de um status taxonômico e, isto merece ser investigado. Em contraste à ampla distribuição de *D. sturtevantii*, as outras espécies desse subgrupo mostram uma distribuição geográfica restrita. Magalhães (1962) registrou a ocorrência de *D. milleri* exclusivamente em El Yunque, Porto Rico, e *D. rectangularis* em Orizaba e Tixtla, México. *D. magalhaesi* foi registrada em Eldorado/RS, Brazil e *D. dacunhai* em Kingston e Ocho Rios na Jamaica (MOURÃO; BICUDO, 1967). Essas espécies são crípticas, sendo distinguidas por pequenas diferenças morfológicas, principalmente na estrutura dos edeagos (MAGALHÃES; BJORNBERG, 1957; MAGALHÃES, 1962; MOURÃO; BICUDO, 1967; SOUZA et al.; 2014).

As populações de *D. sturtevantii* (subgrupo *sturtevantii*, grupo *saltans*) têm sido avaliadas de longa data em diversos estudos quanto ao isolamento reprodutivo (DOBZHANSKY, 1944; BICUDO, 1979; HOSAKI-KOBAYASHI; BICUDO, 1994, 1997), polimorfismos de inversões (BICUDO, 1973; KOBAYASHI; BICUDO, 1997), valor adaptativo (CARARETO et al., 1999), morfologia (MAGALHAES; BJORNBERG, 1957; MAGALHAES, 1962), esterases (NASCIMENTO; BICUDO, 2002, 2006; BERNARDO; BICUDO, 2009), elementos de transposição (CLARK et al. 1995; CLARK; KIDWELL, 1997, ALMEIDA; CARARETO, 2002; SILVA; KIDWELL, 2004; CASTRO; CARARETO, 2004a, 2004b; SETTA et al., 2007a, 2007b), sequências moleculares de genes estruturais (O'GRADY et al., 1998, RODRIGUEZ-TRELLES et al., 1999) e mais recentemente morfologia da terminália masculina (SOUZA et al., 2014), estruturação populacional por meio de locos de microssatélites (TRAVA, 2018) e morfologia geométrica de asas e edeagos (SEGALA, 2019). Os resultados obtidos por Trava (2018) mostraram alto polimorfismo para os locos de microssatélites analisados, sendo o grau de polimorfismo observado maior que o obtido para qualquer espécie do grupo *saltans*, além de estrutura populacional moderada, embora não tenha sido verificada correlação com as distâncias geográficas. Já os dados de Segala (2019) indicaram a existência de estruturação latitudinal tanto para o tamanho e a forma da asa como

para a forma do edeago e refletem a diferenciação genética em curso nas populações de *D. sturtevanti*, sendo acentuada para as populações provenientes da América Central e as do Sul do Brasil e, indicando a ocorrência de diferenciação populacional que poderia ser reflexo da existência de novas espécies deste subgrupo.

A ampla distribuição geográfica e alta abundância em fragmentos florestais dessa espécie faz com que ela seja um excelente organismo para estudos filogeográficos, o que nos motivou avaliar a sua estrutura populacional com marcadores moleculares, e verificar se há um modelo de distribuição biogeográfica para o bioma da Mata Atlântica. Além disso, a alta variabilidade genética verificada nas populações avaliadas para a maioria dos marcadores considerados e o fato de estudos entre populações geográficas mostrarem um grau de isolamento reprodutivo incipiente instigou a necessidade de se avaliar se isso também se reflete nos marcadores de DNA mitocondrial, tornando esse o primeiro estudo a avaliar sob o contexto filogeográfico uma espécie pertencente ao grupo *Sophophora* na região Neotropical.

**Figura 5.** Imagem de uma fêmea de *D. sturtevanti*.



Créditos da imagem: Gabriel Acciolly Tripode.

## 2. Objetivos

### 2.1. Objetivo geral

O presente trabalho teve por objetivo realizar uma análise filogeográfica da espécie *Drosophila sturtevantii*, por meio de sequências parciais dos genes mitocondriais *COI* e *COII* em populações geográficas da espécie em fragmentos florestais na região neotropical. Neste estudo foi avaliada a existência de estruturação genética nas populações, bem como a identificação se existe relação entre a estruturação genética e a distribuição geográfica das populações. Além disto, avaliamos a existência de eventos históricos que poderiam estar relacionados ao padrão filogeográfico observado.

### 2.2. Objetivos específicos

- Obter as sequências de bases dos genes mitocondriais *COI* e *COII* de *D. sturtevantii* para caracterização genética dos indivíduos das populações nas diferentes localidades estudadas;
- Determinar as variantes genéticas existentes nestas populações, por meio da identificação e quantificação de sítios polimórficos, diversidade e diferenças nucleotídicas e haplotípicas;
- Determinar a distância geográfica e grau de estruturação populacional por meio dos marcadores nas populações estudadas;
- Verificar os eventos recorrentes, como fluxo gênico e/ou históricos tais como vicariância ou expansão de área que contribuem para a estrutura genética destas populações;
- Estabelecer inferências filogeográficas para auxiliar na compreensão da evolução da espécie;
- Estimar o tempo de divergência das populações que se encontram estruturadas e inferir possíveis eventos geográficos associados a esse padrão;

- Utilizar as informações obtidas para inferir um modelo de distribuição biogeográfica para a espécie na região de domínio Mata Atlântica, bem como utilizá-las junto a outras pesquisas para compreender a evolução dos biomas que permeiam a região Neotropical e estabelecer estratégias para a conservação destes biomas.

### **3. Materiais e Métodos**

A metodologia utilizada para obtenção e análise dos dados será apresentada no capítulo “Signature of climatic differentiation on mitochondrial DNA of *Drosophila sturtevantii*”.

### **4. Resultados**

Esta tese está organizada em um capítulo, intitulado “Signature of climatic differentiation on mitochondrial DNA of *Drosophila sturtevantii*”, apresentado na forma de artigo científico e submetido à revista *Journal of Biogeography*, em co-autoria com o Dr. Amir Yassin, pesquisador do Centre National de la Recherche Scientifique, no Muséum National d’Histoire Naturelle em Paris e, portanto, já formatado de acordo com as normas exigidas pela revista.

#### 4.1. **Capítulo: 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</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.

#### **Acknowledgments**

To Prof. Dr. Daniela de Toni, Prof. Dr. Claudia Rohde and Prof. Dr. Hermes J. Schmitz for supplying some strains of *D. sturtevanti*. To FAPESP for the support of the research project (process number 2014/14059-0, 2016/11994-5); to CAPES for support to 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).

#### **Abstract**

**Aim** The aim of this study is to analyze the genetic differentiation of *Drosophila sturtevanti*, a Neotropical species belonging to the *Drosophila saltans* group which is one of the most geographically wide species within the group, occurring on forest areas of different biomes in order to highlight the diversification patterns and the role of Pleistocene climatic changes on the Brazilian Atlantic Forest and South America biodiversity.

**Location** Atlantic Forest and rainforest fragments from Central America to South Brazil.

**Methods** The nucleotide information obtained from a total of 1383bp, resulted from the concatenated data of the cytochrome C oxidase subunits I and II genes of 163 individuals from 21 populations of *D. sturtevanti*, was used to perform statistical analyses to infer the phylogeographical history of the species and to estimate the split of time between the populations, besides to correlate the molecular data with climatic variables in order to perform Isolation By Ecology tests.

**Results** Our molecular results showed that *D. sturtevanti* populations were structured in South America and that their mitochondrial haplotypes had a significant correlation with geography and climatic variables, mainly temperature and humidity. Moreover, the divergence events between the geographical groups were associated with paleoclimate oscillations occurred since the late Pleistocene.

**Main conclusions** The results support a role of paleoclimatic events on genomic diversity of present-day widespread Neotropical species, besides to suggest a shared distribution pattern with other species, congruent with the existence of genetic barriers and the correlation with the climatic variables analyzed.

**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, Stefenon et al., 2019). 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*, a species belonging to one of the two groups that represent the Neotropical clade of *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). It is widely distributed in many different environments and geographical conditions, which makes it a suitable candidate and informative model to study population dynamics and demographic history in South America. Thus we aimed, for the first time, to investigate the population diversification and demographic history in a Neotropical species of subgenus *Sophophora*, *D. sturtevantii*, using haplotypes of Cytochrome Oxidase I and II in order to infer ecological and climate association with the observed distribution patterns and, this way, we expect to contribute to elucidate the evolutionary history of the species and to understand the diversification patterns in the Neotropical region.



## Material and methods

### Sampling

In this study, we used 163 individuals from 21 localities distributed in South America (Table 1). Seventeen collections were performed between 2015 and 2018 (FAPESP 2014/14059-0; 2016/11994-5), using suspended closed traps made from plastic bottles with bananas 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. Besides, we also used isofemale strains from other four localities maintained in the laboratory on banana medium and biological yeast (*Saccharomyces cerevisiae*).

### Molecular Data and Statistical Analysis

Genomic DNA of adult males from *Drosophila sturtevanti* was extracted using Wizard Genomic DNA Purification kit (Promega, Madison, WI, USA). The isolation and amplification of fragments 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 fragments of *COII* (687 bp) was performed using TL2-J-3037 forward and TK-N-3785 reverse primers (Simon et al., 1994). PCR reaction 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<sub>2</sub>, and 1 U of Taq DNA polymerase (Promega). All reactions were performed in a Verity Termocycler under the following conditions: 94°C for 2 minutes, followed by 40 cycles of 94°C for 30 seconds, (56°C for *COI* and 58°C for *COII*) for 30 seconds, 72°C for 1 minute, then 72°C for 5 minutes. PCR products were visualized after electrophoresis in 2% agarose gel stained with ethidium bromide and purified using an Illustra GFX kit (GE Healthcare, Little Chalfont, UK) following manufacturer's instructions. Amplicons sequencing was performed using BigDye Terminator Cycle Sequencing Ready Reaction kit (Perkin-Elmer, Foster City, CA, USA) in an automatic sequencer ABI 3739 XL DNA Analyzer (Applied Biosystems) with the same primers and conditions used for PCR

amplification. The forward and reverse sequences were aligned using ClustalW Multiple Alignment 1.8 (Thompson et al., 1994) and edited in BioEdit v7.1.9 (Hall, 1999).

### **Population genetics analysis**

Due to the fact that Cytochrome Oxidase I and II are both mitochondrial genes, thus sharing the same inheritance with no 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), number of polymorphic sites (S) and number of haplotype (h) were computed using DNAsp 5.10 (Rozas et al., 2003; Librado & Rozas, 2009). We also performed neutrality tests Tajima's D (Tajima, 1989), Fu's Fs (Fu, 1997) and R2 test (Ramos-Onsins & Rozas, 2002) based on sequences polymorphism in order to verify which factors could be acting on the sequences evolution using DNAsp (Rozas et al., 2003; Librado & Rozas, 2009). JModelTest 3.0 (Posada & Crandall, 1998; 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.

The haplotype network, was generated using Haplotype Viewer (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 v.1.8 (Drummond & Rambaut, 2007; Drummond et al., 2012) using a HKY+I model, calculated by JModelTest 3.0 and the *sturtevantii*-like strain from French Guiana was utilized as the outgroup. 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 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 by BAPS6 (Corander et al., 2008; Corander et al., 2013). We performed the analysis using population structure model followed by admixture analysis with K values ranging from 1 to 10, 200 interactions and 10 repetitions and then we selected the solution with the highest log likelihood. 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 in order to infer the divergence time among the population groups inferred by the clustering analyses, using an Isolation-with-Migration model framework. The phylogenetic relationships between the population groups was inferred based on DNAsp-estimated  $F_{ST}$  pairwise distances.

### Isolation by ecology analysis

DNAsp was used to estimate  $F_{ST}$  between pairs of populations with 157 sampled individuals and a neighbor-joining (NJ) tree was inferred from the  $F_{ST}$  pairwise distances (hereafter the genetic 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) wherein the correlation between  $F_{ST}$ -based genetic matrix and matrices summarizing the geographical and ecological distances is simultaneously estimated through the partial equation:

$$r_{ExG|D} = \frac{r_{ExG} - r_{GxD}r_{ExD}}{\sqrt{1 - r_{GxD}^2} \sqrt{1 - r_{ExD}^2}}$$

Wherein G, D and E refer to the genetic, geographical and environmental distance matrices respectively and  $r$  to the coefficient of correlation. While the simple equation  $r_{ExG}$  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, using the linear distances from the decimal coordinates. For the ecological distance matrix E,

we downloaded 13 climatic variables (see Appendix B in Supporting Information) for the 17 geographical localities from the NASA POWER Project Data Sets, available at <https://power.larc.nasa.gov/> as in Bastide et al. (2014). We then used the PAST software package (Hammer & Harper, 2006) to run a Principal Component Analysis (PCA) and to convert the climatic variables into a distance matrix using Euclidean method. Coefficient of correlation between G, D and E and significance of the IBE coefficient were estimated using a customized perl script (see Appendix A in Supporting Information) and after three matrix permutations.

## Results

### Low overall nucleotide diversity of mitochondrial DNA in *D. sturtevanti*

A total of 1383bp, resulted from the concatenated data of *COI* (696pb) and *COII* (687pb) genes, corresponding to positions 1483-2178 – *COI* and 3083-3769 – *COII* of the *Drosophila yakuba* reference genome (accession number NC\_001322) were obtained for 163 sequences of *D. sturtevanti*. Overall, we observed 47 polymorphic sites, resulting in 48 haplotypes, distributed in three haplogroups (Fig. 1). The haplotype diversity was high ( $Hd = 0.864$ ), while the nucleotide diversity was low ( $\pi = 0.00138$ ).

**Table 1.** Location, code and summary descriptive statistics for the sampled populations of *D. sturtevantii* using the *COI* and *COII* concatenated data.

Location	Code	N	h	S	Hd	$\pi$	Fu's FS	Tajima's D	R2
Tirimina-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	BAPS	33	11	13	0.633	0.00084	<b>-7.186</b>	<b>-2.05883</b>	<b>0.0592</b>
Central Weast	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>

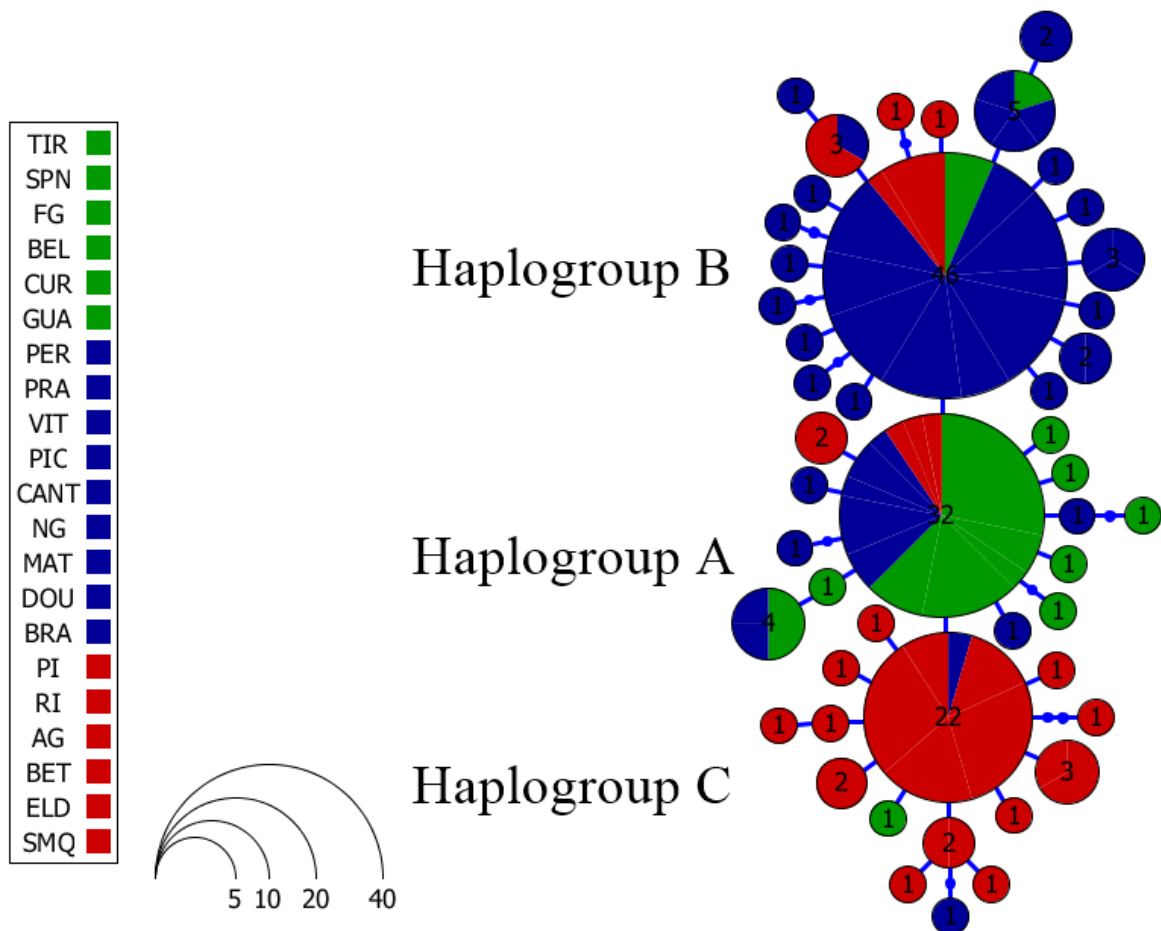
Code, laboratory code; N, number of individuals; *h*, number haplotypes; S, number of variable sites; *Hd*, haplotype diversity;  $\pi$ , nucleotide diversity. To neutrality tests, bold values should be considered significant ( $p < 0.05$ ).

\* Tajima's D Statistical significance: Not significant,  $0.10 > p > 0.05$ .

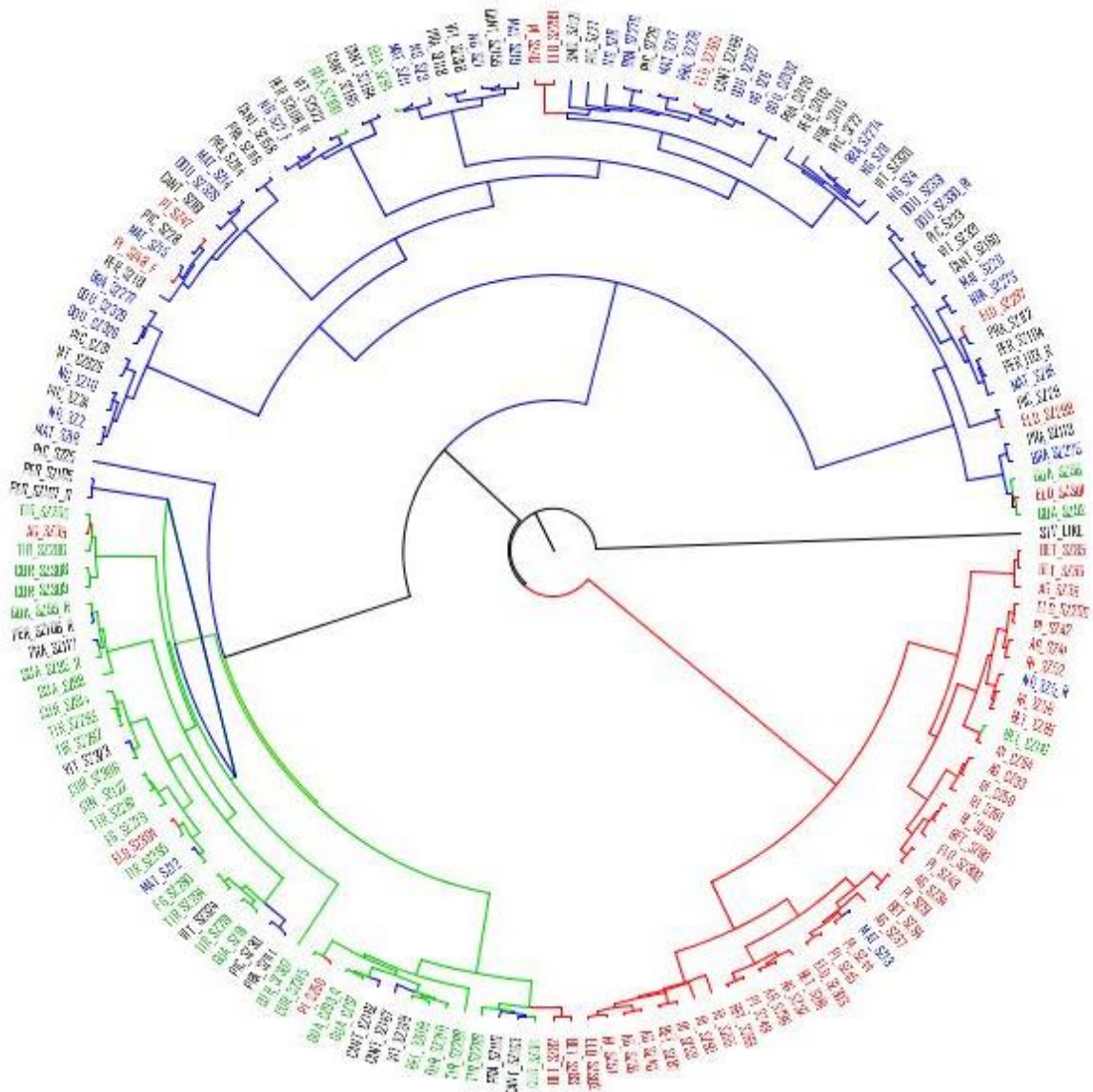
### Mitochondrial haplotypes cluster within three major haplogroups

The 48 mitochondrial haplotypes clustered in three major haplogroups depicted in different colors and denoted as A, B and C in Figure 1. Haplogroup A (green) seems to represent the ancestral sequence. It is connected to the *sturtevantii*-like strain (Figure 2) 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. Haplotypes are represented by circles with sizes which are proportional to the number of individuals. Hypothetical haplotypes are represented by small blue circles and the outgroup is represented by the white circle on the left represented by stv-like sample.



**Figure 2.** BI tree of *Drosophila sturtevantii* based on concatenated dataset (*COI* and *COII* sequences). *D. sturt-like* strain from French Guyana was utilized as the outgroup. The colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively.



### 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 3).

The degree of genetic variation also differed within each of the 17 populations (Table 1). Neutrality tests (Table 1) provided significant positive results concordant with the overall low nucleotide diversity, which could either be due to a selective pressures on or recent demographic expansions of these populations. The optimal

genetic structure according to BAPS analysis was considering 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 (Figure 6A).

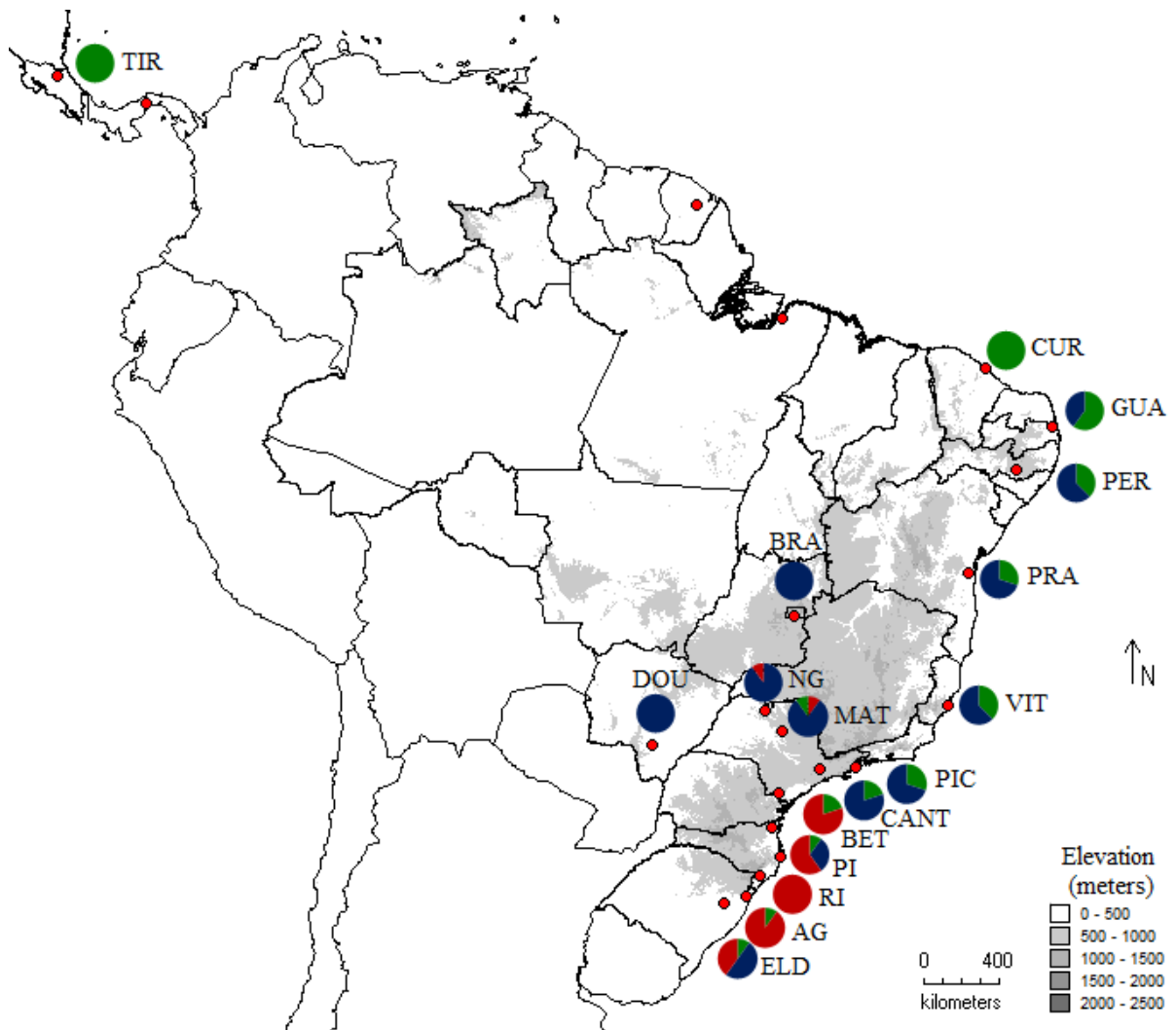
AMOVA results (Table 2) showed similar values to scenarios with three and four groups, based on BAPS clustering, so here we are going to present the results using  $K=4$ , as generated by BAPS because it seems to be more realistic scenario to explain *D. sturtevantii* diversification.  $F_{CT}$  and  $F_{ST}$  values indicate high level of genetic differentiation (0.26528 and 0.32415, respectively) among groups and within populations and moderate differentiation among populations within groups ( $F_{SC} = 0.08012$ ) thus, the AMOVA results are congruent with the haplotype network and reinforce BAPS structure. Isolation-by-geography (IBD), as measured by the simple  $r_{GxD}$  between both geographical and  $F_{ST}$ -inferred genetic distances (Matrix 2, see Appendix B in Supporting Information) produced 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 3.** Geographical distribution of the haplogroup frequencies to 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 (<5). The frequencies colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively.

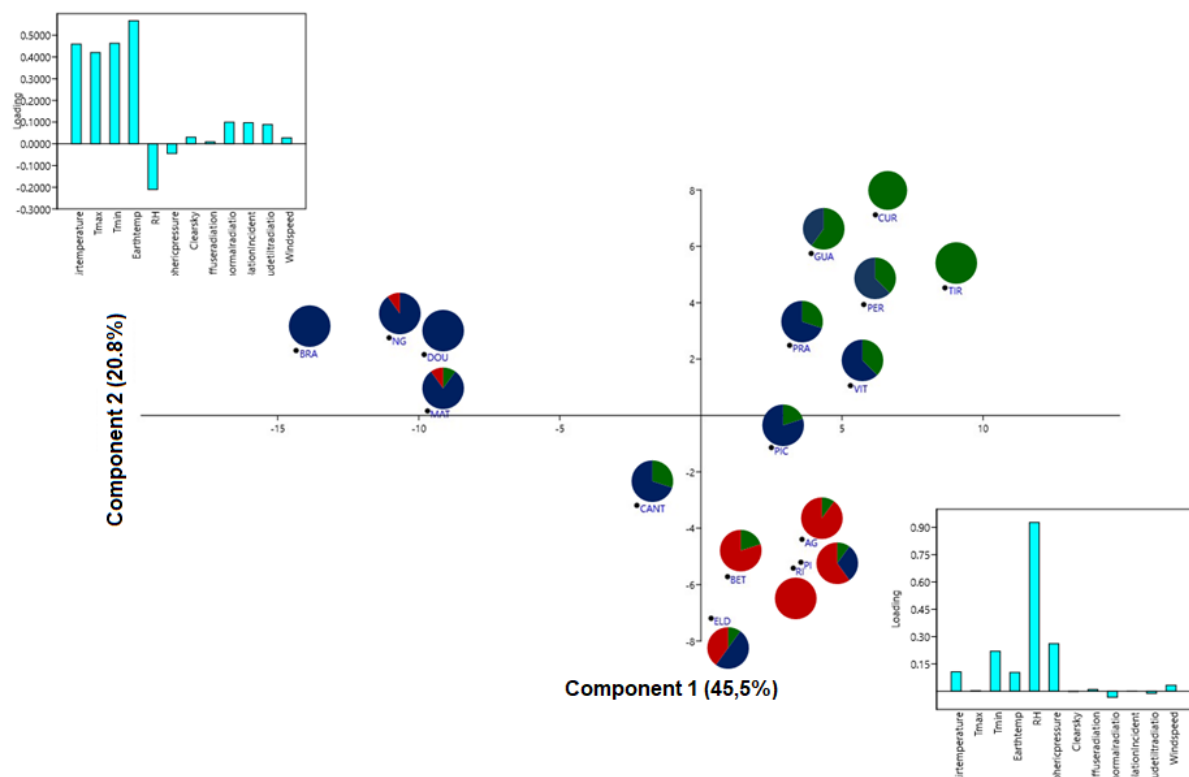


### Genetic differentiation strongly correlates with climatic variables

Principal components analysis (PCA) of the 13 climatic variables of the 17 geographical localities indicated a strong differentiation of these localities (Figure 4). The two first components (PC1 and PC2) explained 45.5 and 20.8% of the total variance. PC1 strongly and positively 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 4). 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 Isolation-by-ecology (IBE)  $r = 0.28352$   $P = 0.0222$  and eco-spatial correlation ( $r = 0.33457$   $P = 0.008$ ), with significant results.

**Figure 4.** Principal Component Analysis (PCA) of the 13 climatic variables of 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 (<5). The frequencies colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively.

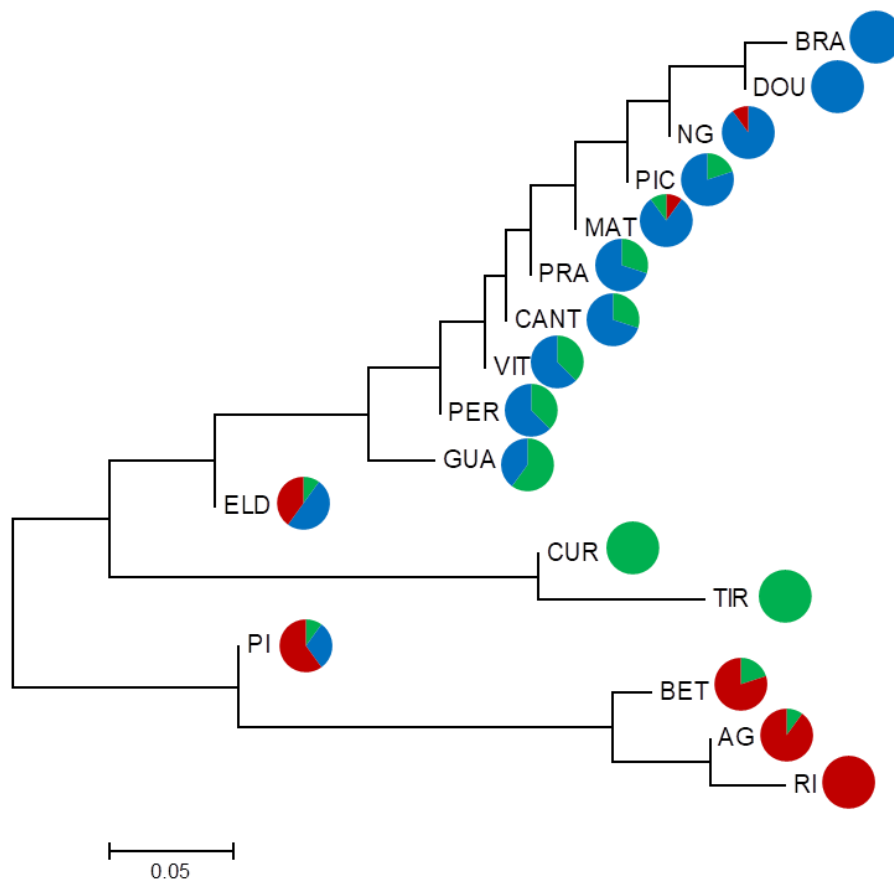


### Phylogeographic history of *D. sturtevantii* mitochondrial DNA

A tree inferred from the degree of genetic differentiation (Matrix 2, see Appendix B in Supporting Information) supports that a closer affinity between Northern and Central populations, predominated by the haplogroups A and B, respectively, than to the Southern population (except for ELD) (Figure 2 and 5). Together with the haplogroup network shown in 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

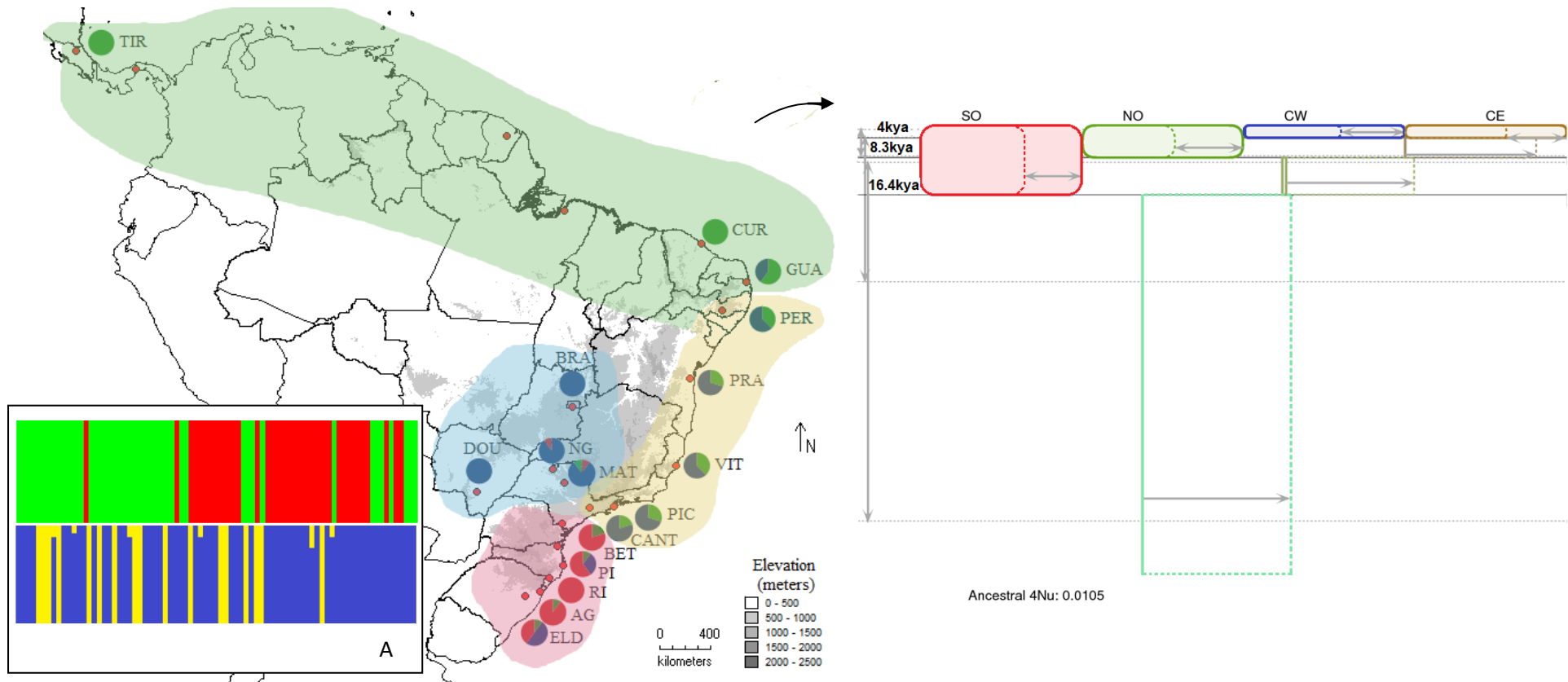
appearing as the most distant from the presumably ancestral stem of the North (Figure 5).

**Figure 5.** Neighbor Joining tree inferred from the  $F_{st}$  pairwise distances between 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 (<5). The frequencies colors (green, blue and red) are the same showed in Figure 1 and represent haplogroups A, B and C, respectively.



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 6, left), and conducted an isolation-with-migration analysis using IMA2. Assuming a mitochondrial DNA mutation rate of  $6.2 \times 10^{-9}$  (Haag-Liautard et al., 2008) and 10 generations per year, the results (Figure 6, 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 6.** 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 (<5). On the right, the divergence time plot generated by IMA2. A. Clusters generated by BAPS, each color corresponds to the same groups used to IMA2.



## Discussion

Our work suggests three main results in *D. sturtevantii*: 1) there is a high mitochondrial haplotype diversity which contrasts with a very low mitochondrial nucleotide diversity, 2) the mitochondrial 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 will discuss the bearing of each of these points on Neotropical phylogeography.

The diversity values found in *D. sturtevantii* populations ( $H_d = 0.864$ ) are higher than the average values found to other Neotropical *Drosophila* species occurring in Atlantic Forest fragments ( $H_d = 0.206$ ) and also to cactophilic species occurring in the seasonally dry tropical forests in 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). Also, the nucleotide diversity found to *D. sturtevantii* is very low, which associated with the star-like shape of the haplotype network (Figure 1) and the neutrality tests give support to the population expansion hypothesis.

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's fitness, so it could also be affecting the diversity indices in *D. sturtevantii*. The haplotype diversity can be considered high when compared with other species collected in forest fragments or moderate to low when compared with 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) results showed no evidence that *Wolbachia* infections could be decreasing the diversity of mtDNA in *D. willistoni* natural populations, even with 55% of infection rate, so we are not going to consider this factor for now.

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 *Drosophila sturtevantii* using 14 populations here

analyzed. The data to aedeagi variation found for Segala (2019) presented a population structure similar to the one we found based on mtDNA data. However, besides the literature assumption that the male terminalia exhibits rapid and divergent evolution, aedeagus analysis was not able to differentiate the Central groups (CE and CW) which present the most recent diversification. Furthermore, aedeagi analysis also evidenced a southernmost cluster which comprises ELD population, and it presented a high percentage of the blue haplogroup in our findings. This population is located in a transition area between the Atlantic Forest and the Pampa grasslands which could withstand a greater variation than more homogeneous environments. Also, Trava (2018) results using the same populations of *D. sturtevantii* evidenced moderate population structure to microsatellite data in spite of no correlation with geographic distances.

Moreover, the geographic distribution of *D. sturtevantii* is quite similar to the one evidenced to *D. meridionalis* for Barrios-Leal et al., (2018a), where they also had an inland and two coastal groups, in spite of the coastal groups to *D. meridionalis* seemed to have 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.

Concerning to the divergence times, the first split observed in the South region 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 to BET (the first geographical population in the South group) has been dominated by forest at least from ~20,000 years ago (ya) and this suggests a change toward more humid and wet conditions which could have favored the expansion and split to *D. sturtevantii* populations to a more suitable environment, following the forest expansion after the Last Glacial Maximum (LGM). It is also interesting to reinforce the location of all the populations belonging to the South group down the tropic of Capricorn, which associated with the physical geography are the main static elements influencing the climate and vegetation and, here, apparently, the haplotype distribution (Sevegnani et al., 2016), 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, (240km from VIT). Then, the more recent divergence event occurred around 4,000 ya, which could be influenced by the establishment of the modern seasonal climate around 4,000 ya and the return of the small gallery forest and arboreal Cerrado (Brazilian savannas) 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. sturtevantii* 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.

## 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. *PNAS*. 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 *sturtevantii* subgroup. *Rev. Brasil. Genet.* 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*. 2ª 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. *J. South Amer. 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 jararacacomplex* (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. *Biodivers. Conserv.* 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*. v. 16: 175-206.

Hales, K.G., Korey, C.A., Larracuenta, A.M. & Roberts, D.M. (2015). *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. *PNAS*, 104 : 2785–2790.

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.

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

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. **Tese de doutorado**. Programa de Pós-Graduação em Biociências. 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.

Stefenon, V.M., Klabunde, G., Lemos, R.P.M., Rogalski, M. & Nodari, R.O. (2019). Phylogeography of plastid DNA sequences suggests post-glacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. *Scientific Reports*. 9: 2752.

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*. v 34: 1294-1295.

## Biosketches

**Samara Videira Zorzato** graduated as a biologist in Biosciences with emphasis on evolutionary genetics under the supervision of L.M.R and carried out a PhD internship under A.Y. supervision. Her research interests include population genetics, phylogeographical history and evolutionary process related to *Drosophila* species differentiation especially in the Brazilian Atlantic Forest.

**Amir Yassin** is a CNRS research scientist working on *Drosophila* evolutionary genetics and systematics. His main focus is to investigate the genomic basis and evolutionary mechanisms underlying phenotypic divergence during speciation.

**Lilian Madi-Ravazzi** is interested in evolutionary biology of drosophilid, with emphasis on the Neotropical *saltans* group of *Drosophila*. She is currently working on researches using natural populations of *D. sturtevantii* with molecular and morphological markers in order to better understand the genetic variability and

structure of the populations, besides other ongoing projects focusing on phylogeny and phylogeography of other species belonging to the group.

### **Supporting Information**

A – Customized perl scripts used to IBE analysis

B – Ecological variables and Distance matrices

## 5. Discussão

Nosso trabalho sugere três resultados principais para *D. sturtevantii*. 1) existe uma alta diversidade haplotípica mitocondrial o que contrasta com uma diversidade nucleotídica muito baixa, 2) a variação mitocondrial está geograficamente estruturada e correlaciona com variáveis climáticas, e 3) a diferenciação eco-geográfica do DNA mitocondrial ocorreu primeiro no Pleistoceno (para as populações subtropicais do Sul) e depois durante o Holoceno (para as populações tropicais do Norte e Centrais). Abaixo, discutiremos a relevância de cada um desses pontos para a filogeografia neotropical.

Os valores de diversidade encontrados para as populações de *D. sturtevantii* ( $H_d = 0,864$ ) são superiores aos valores médios encontrados para outras espécies neotropicais de *Drosophila* ocorrendo em fragmentos de Mata Atlântica ( $H_d = 0,206$ ) e também para espécies cactofílicas com ocorrência em florestas tropicais sazonalmente secas (SDTF) na América do Sul ( $H_d = 0,554$ ) (MANFRIN; SENE, 2006; FRANCO; MANFRIN, 2013; CÉNZI DE RÉ et al., 2014; GUSTANI et al., 2015; BARRIOS-LEAL et al., 2018a, b). Além disso, a diversidade nucleotídica encontrada em *D. sturtevantii* é muito baixa e, este resultado associado à forma de estrela da rede de haplótipos (Figura 1) e aos testes de neutralidade dão suporte à hipótese de expansão populacional.

Há diferentes causas para a ocorrência de variação do mtDNA, uma destas e com importância significativa é a infecção por *Wolbachia*. Esse microrganismo é um simbionte intracelular co-herdado com as mitocôndrias e é responsável pela maioria das infecções que ocorrem em invertebrados. Sua presença pode levar a distorções da proporção sexual e, sob algumas condições, pode aumentar a aptidão do hospedeiro e, portanto, poderia também afetar os índices de diversidade em *D. sturtevantii*. A diversidade de haplótipos em *D. sturtevantii* pode ser considerada alta quando comparada com outras espécies coletadas em fragmentos florestais ou moderadas a baixas quando comparadas com espécies cactofílicas. Já sabemos que as populações de *D. sturtevantii* são naturalmente infectadas com *Wolbachia* (MILLER; RIEGLER, 2006), no entanto, ainda temos que investigar a associação entre os haplótipos mitocondriais e a presença de *Wolbachia*, para entender se a mesma está interferindo na variação observada. Os resultados de Müller e

colaboradores (2012) não mostraram evidências de que as infecções por *Wolbachia* pudessem estar diminuindo a diversidade de mtDNA em populações naturais de *D. willistoni*, mesmo com 55% da taxa de infecção, entretanto, é algo que merece ser investigado no futuro em *D. sturtevanti*.

Com relação à distribuição geográfica dos haplogrupos e estimativas de tempo, é importante enfatizar as semelhanças de distribuição entre nossos resultados e aqueles encontrados para dados morfológicos em *D. sturtevanti* utilizando 14 das populações aqui analisadas. Os dados para a variação de edeagos encontrados por Segala (2019) apresentaram uma estrutura populacional semelhante à que encontramos com base nos dados de mtDNA. No entanto, além da premissa da literatura de que a terminália masculina apresenta evolução rápida e divergente, a análise de edeagos não foi capaz de diferenciar os grupos centrais (CE e CW) que apresentaram diversificação mais recente. Além disso, a análise de edeagos também evidenciou um *cluster* mais ao sul que compreende a população de ELD, e apresentou uma alta porcentagem do haplogrupo B (em azul) em nossos achados. Essa população está localizada em uma área de transição entre a Mata Atlântica e o Pampa, que poderia suportar uma variação maior do que ambientes mais homogêneos. Além disso, os resultados de Trava (2018) usando as mesmas populações de *D. sturtevanti* evidenciaram uma estrutura populacional moderada por meio de microssatélites espécie-específicos, embora essas análises não tenham evidenciado correlação com as distâncias geográficas.

Além disso, a distribuição geográfica de *D. sturtevanti* é semelhante à evidenciada para *D. meridionalis* por Barrios-Leal e colaboradores (2018a), onde também se observou um grupo no interior e dois grupos costeiros, apesar dos grupos costeiros de *D. meridionalis* parecerem apresentar uma dinâmica mais complexa afetada por expansões recorrentes e contrações com isolamento e mistura de populações, os grupos de *D. sturtevanti* parecem ser mais isolados, evidenciados pelas baixas taxas de haplótipos compartilhados.

Com relação aos tempos de divergência, a primeira divisão observada na região Sul está de acordo com a análise de isótopos e datação de  $^{14}\text{C}$  de fragmentos de carvão enterrados e da fração húmica da matéria orgânica do solo (SAIA et al., 2018). De acordo com esses achados, a área perto de BET (a primeira população geográfica no grupo do Sul) tem sido dominada pela floresta, pelo menos a partir de aproximadamente 20.000 anos atrás, e isso sugere uma mudança em direção a

condições de maior umidade e precipitação que podem ter favorecido a expansão e divisão das populações de *D. sturtevantii* para um ambiente mais adequado, seguindo a expansão da floresta após o Último Máximo Glacial (UMG). Também é interessante reforçar a localização de todas as populações pertencentes ao grupo do Sul abaixo do trópico de Capricórnio, que associado à geografia física são os principais elementos estáticos influenciando o clima e a vegetação e, aqui, aparentemente, a distribuição haplotípica (SEVEGNANI et al., 2016), além disso, a ocorrência de uma barreira genética previamente sugerida entre os estados de São Paulo e Paraná, próximo ao rio Paranapanema, poderia sugerir vicariância (GRAZZIOTIN et al., 2006).

O segundo evento de divisão, entre as populações Norte e Central, está estimado há cerca de 8.300 anos, em concordância com o aumento da diversidade baseada em evidência palinológica (estudo de grãos de pólen) no Parque Estadual do Rio Doce, em Minas Gerais (240 km de VIT). Então, o evento de divergência mais recente ocorreu há cerca de 4.000 anos, o que poderia ser influenciado pelo estabelecimento do clima sazonal moderno em torno de 4.000 anos e o retorno da pequena floresta de galeria e do Cerrado arbóreo à região central do Brasil ~ 3.500 anos, com um clima quente e semi-úmido, semelhante ao atual (FERRAZ-VICENTINI; SALGADO-LABORIAU; 1996; BUSO et al., 2013). A evidência paleoclimática destaca como a diversificação e expansão das populações de *D. sturtevantii* ocorreram em diferentes ambientes ao longo do tempo e também pode explicar, pelo menos em parte, a distribuição atual dos haplogrupos relacionados ao clima e à variação ecológica.

## 6. Conclusões

Este é o primeiro trabalho a avaliar a filogeografia e história evolutiva de uma espécie pertencente ao subgênero *Sophophora* na região Neotropical e os resultados obtidos demonstraram que a espécie apresenta estruturação populacional e que a datação dos eventos de divergência e expansão populacional observados corroborou com dados paleoclimáticos disponíveis na literatura. Além disso, parece haver uma associação entre a distribuição geográfica dos haplótipos observados para *D. sturtevanti* e as variáveis climáticas testadas que merece ser melhor investigada.



## Referências

- ADAMS, M. D., CELNIKER, S. E.; HOLT, R. A.; EVANS, C. A.; GOCAYNE, J. D. et al., The genome sequence of *Drosophila melanogaster*. **Science**. v. 287, p. 2185–2195, 2000.
- ALMEIDA, L.M.; CARARETO, C.M.A. Gonadal Hybrid dysgenesis in *Drosophila sturtevantii* (Diptera, Drosophilidae). **Iheringia, Sér. Zool.**, Porto Alegre, v. 92(2), p. 71-79, 2002.
- ANDRADE-LIMA, D. DE. Present-day forest refuges in northeastern Brazil. in G. T. Prance, editor. Biological diversification in the tropics. **Columbia University Press**, New York, p. 245–251, 1982.
- ANGULO, A.; NARANJO-ELIZONDO, B.; ROJAS, E.; LEY-LÓPEZ, J.M. Fishes from the Tirimbina Biological Reserve, La Virgen de Sarapiquí, Heredia, Costa Rica. **Check List**. v. 13, p. 683-702, 2017.
- ANTONELLI, A.; ZIZKA, A.; CARVALHO, F.A.; SCHARN, R.; BACON, C.D.; SILVESTRO, D.; CONDAMINE, F.L. Amazonia is the primary source of Neotropical biodiversity. **PNAS**. v. 115, p. 6034-6039, 2018.
- AVISE, J.C. Molecular Markers, Natural History, and Evolution. 2ed. **Sinauer**, Sunderland, MA. 684 p, 2004.
- AVISE, J.C. Phylogeography: history and formation of species. 1 ed. Harvard: **Harvard University Press**, 2000.
- AVISE, J.C. Phylogeography: retrospect and prospect. **Journal of Biogeography**, v.36, p.3-15, 2009.
- AVISE, J.C., ARNOLD, J., BALL, R.M., JR, BERMINGHAM, E., LAMB, T., NEIGEL, J.E., REEB, C.A., SAUNDERS, N.C. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. **Annual Review of Ecology and Systematics**, v.18, p.489–522, 1987.
- AVISE, J.C.; BALL, R.M. Principles of genealogical concordance in species concepts and biological taxonomy. In: Futuyama, D.; Antonovics, J.(Org.). *Evolutionary Biology*. Oxford: **Oxford University Press**. p. 45–67, 1990.
- AVISE, J.C.; NEIGEL, J.E.; ARNOLD, J. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. **J. Mol. Evol.**, v. 20, p. 99-105, 1984.
- BARRIOS-LEAL, D.Y.; FRANCO, F.F.; SILVA, E.C.C.; SANTOS, C.K.B.; SENE, F.M.; MANFRIN, M.H. Deep intraspecific divergence in *Drosophila meridionalis*, a

cactophilic member of the New World *Drosophila repleta* group. **Biological Journal of the Linnean Society**. v. 123, p. 163–178, 2018a.

BARRIOS-LEAL, D.Y.; NEVES-DA-ROCHA, J.; MANFRIN, M.H. Genetics and Distribution Modeling: The Demographic History of the Cactophilic *Drosophila buzzatii* Species Cluster in Open Areas of South America. **Journal of Heredity**. v. 20, p. 1-12, 2018b.

BATALHA-FILHO, H.; CABANNE, G. S.; MIYAKI, C. Y.; Phylogeography of an Atlantic forest passerine reveals demographic stability through the last glacial maximum. **Molecular Phylogenetics and Evolution**, v. 65, p. 892-902, 2012.

BATALHA-FILHO, H.; MIYAKI, C. Y.; Late Pleistocene divergence and postglacial expansion in the Brazilian Atlantic Forest: multilocus phylogeography of *Rhopias gularis* (Aves: Passeriformes). **Journal of Zoological Systematics and Evolutionary Research**, v.54, p. 137-147, 2016.

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

BEIGUELMAN, B. **Genética de populações humanas**. Ribeirão Preto: SBG, 235p, 2008.

BERNARDO, A.A., BICUDO, H.E.M.C. Variability of esterase patterns in adult flies of the *saltans* species group of *Drosophila* (subgenus *Sophophora*). **Genetica** v. 137, p. 111-124, 2009.

BICUDO, H.E.M.C. Reproductive isolation in the *saltans* group of *Drosophila*. IV. The *sturtevantii* subgroup. **Rev. Brasil. Genet.** v. 4, p. 247-258, 1979.

BICUDO, H.E.M.C. Reproductive isolation in the *saltans* group of *Drosophila*. I. The *saltans* subgroup. **Genetica** v. 44, p. 313-329, 1973.

BIRKY, C. W.; MARUYAMA, T.; FUERST, P. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts and some results. **Genetics**, v.103, p. 513–527, 1983.

BROWN, J. H., KODRIC-BROWN, A. Turnover rates in insular biogeography: effect of immigration on extinction. **Ecology**, v. 58, p. 445-449, 1977.

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. Late Pleistocene and Holocene Vegetation, Climate Dynamics, and Amazonian Taxa in the Atlantic Forest, Linhares, SE Brazil. **Radiocarbon**. v. 55, p. 1747–1762, 2013.

CAMUS, M.F.; WOLFF, J.N.; SGRÒ, C.M.; DOWLING, D.K. Experimental evidence that thermal selection has shaped the latitudinal distribution of mitochondrial haplotypes in Australian fruit flies. **Mol. Biol. Evol.** v.34, p. 2600-2612, 2017.

CARARETO, C.M.A. Incipient ethological isolation among strains of *Drosophila sturtevantii*. **Naturalia**, v. 19, p.185-198, 1994.

CARARETO, C.M.A.; LOURENÇO, M.F.; MOURÃO, C.A. Sensitivity of fitness to variation in its components in *Drosophila sturtevantii* (Diptera, Drosophilidae), **Iheringia**, Série Zoologia, v. 87, p. 37-48, 1999.

CARNAVAL, A. C.; MORITZ, C. Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. **Journal of Biogeography**, v. 35, p. 1187-1201, 2008.

CARNAVAL, A.C.; WALTARI, E.; RODRIGUES, M.T.; ROSAUER, D.; VANDERWAL, J.; DAMASCENO, R.; PRATES, I.; STRANGAS, M.; SPANOS, Z.; RIVERA, D.; PIE, M.R.; FIRKOWSKI, C.R.; BORNSCHEIN, M.R.; RIBEIRO, L.F.; MORITZ, C. Prediction of phylogeographic endemism in an environmentally complex biome. **Proc. R. Soc. B.**, v. 281, p. 1-8, 2014.

CARVALHO, C.J.B.; ALMEIDA, E.A.B. Biogeografia da América do Sul: Análise de tempo, espaço e forma. 2ª edition. Rio de Janeiro: **Roca**. 324p. 2016.

CASTRO, J.P.; CARARETO, C.M.A. Canonical *P* elements are transcriptionally active in the *saltans* group of *Drosophila*. **J Mol Evol** v. 59, p. 31-40, 2004b.

CASTRO, J.P.; CARARETO, C.M.A. *P* elements in the *saltans* group of *Drosophila*: a new evaluation of their distribution and number of genomic insertion sites. **J Mol Evol** v. 32, p. 383-387, 2004a.

CENZI de RÉ, F.; GUSTANI, E. C.; OLIVEIRA, A. P. F.; MACHADO, L. P. B.; MATEUS, R. P.; LORETO, E. L. S.; ROBE, L. J.; 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**, v. 112, p. 55-66, 2014.

CLARK, J.B.; ALTHEIDE, T.K.; SCHLOSSER, M.J.; KIDWELL, M.G. Molecular evolution of *P* transposable elements in genus *Drosophila*. I. The *saltans* and *willistoni* species groups. **Mol Biol Evol** v. 12, p. 902-913, 1995.

CLARK, J.B.; KIDWELL, M.G. Phylogenetic perspective on *P* transposable elements evolution in *Drosophila*. **Proc Natl Acad Sci USA** v. 94, p. 11428-11433, 1997.

COLOMBO, A. F.; JOLY, C. A. Brazilian Atlantic Forest lato sensu: the most ancient Brazilian forest, and a biodiversity hotspot, is highly threatened by climate change. **Brazilian Journal of Biology**, v. 70, n. 3, p. 697-708, 2010.

- CRISCI, J. V. The Voice of Historical Biogeography. **Journal of Biogeography** v. 28 (2), p. 157-168, 2001.
- CRISCI, J.V., KATINAS, L. E POSADAS, P. Historical Biogeography: an introduction. Cambridge, **Harvard University Press**, 2003.
- DAMASCO, G.; FONTES, C.; FRANÇOSO, R.; HAIDAR, R. The Cerrado Biome: A Forgotten Biodiversity Hotspot. **Frontiers for Young Minds**, v. 6, p.1-9, 2018.
- DOBZHANSKY, T. Experiments on sexual isolation in *Drosophila* III. Geographic strains of *Drosophila sturtevantii*. **Proc. Natl. Acad. Sci. USA** v. 30, p. 335-339, 1944.
- DROSOPHILA 12 GENOMES CONSORTIUM *et al.*, Evolution of genes and genomes on the *Drosophila* phylogeny. **Nature**, v. 450, p. 203–218, 2007.
- FATMA SC. **Reserva Biológica Estadual do Aguaí**. Disponível em: <http://ima.sc.gov.br/index.php/ecosistemas/unidades-de-conservacao/reserva-biologica-estadual-do-aguai> Acesso em 18 de abril de 2019.
- FENG, Y.; ZHANG, Y.; YING, C.; WANG, D; DU, C. Nanopore-based Fourth-generation DNA Sequencing Technology. **Genomics Proteomics Bioinformatics**. v. 13, p. 4-16, 2015.
- FERRAZ-VICENTINI, K.R.; SALGADO-LABOURIAU, M.L. Palynological analysis of a palm swamp in Central Brazil. *J. South Amer.* **Earth Sciences**, v. 9, p. 207-219, 1996.
- FRANCO, F. F; MANFRIN, M. H.; Recent demographid history of cactophilic *Drosophila* species can be related to Quaternary paleoclimatic changes in South America. **Journal of Biogeography**, v. 40, p. 142-154, 2013.
- FRANKHAM, R. Genetics and conservation biology. **Comptes Rendus Biologies**, v. 326, p. S22-S29, 2003.
- FU, Y.; LI, W. Coalescing into the 21st Century: An Overview and Prospects of Coalescent Theory. **Theory in Population Biology**, v. 56, p. 1-10, 1999.
- GALTIER, N.; NABHOLZ, B.; GLÉMIN, S.; HURST, G.D.D. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, v. 18, p. 4541-4550, 2009.
- GILLUNG, J.P. Biogeography: the history of life on Earth. **Revista da Biologia**. p. 1-5, 2011.
- GISSI, C.; IANNELLI, F.; PESOLE, G. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. **Heredity**, v.101, p. 301-320, 2008.

GRAZZIOTIN, F. G.; MONZEL, M.; ECHEVERRIGARAY, S.; BONATTO, S. L. Phylogeography of the Bothrops jararacacomplex (Serpentes: Viperidae): past fragmentation and island colonization in the Brazilian Atlantic Forest. **Molecular Ecology**, v.15, p. 3969–3982, 2006.

GUIA FLORIPA. **Informações gerais.** Disponível em: <http://www.guiafloripa.com.br/cidade/informacoes-gerais-sobre-florianopolis/geografia> Acesso em 18 de abril de 2019.

GUIMARÃES ROSA, J. **Grande sertão: veredas.** Rio de Janeiro: Nova Fronteira, 1986.

GUSTANI, E. C.; OLIVEIRA, A. P. F.; SANTOS, M. H.; MACHADO, L. P. B.; MATEUS, R. P.; Demographic Structure and Evolutionary History of *Drosophila ornatifrons* (Diptera, Drosophilidae) from Atlantic Forest of Southern Brazil. **Zoological Science**, v. 32, p. 141-150, 2015.

HAFFER, J. Amazonian forest birds. **Science**, v. 165, p. 131-137, 1969.

HAFFER, J. Time's cycle and time's arrow in the history of Amazonia. **Biogeographica**, v. 69, 15-45, 1993.

HAFFER, J.; Alternative models of vertebrate speciation in Amazonia: An overview. **Biodivers. Conserv**, v. 6, p. 451-476, 1997.

HAFFER, J.; PRANCE, G.T. Impulsos climáticos da evolução na Amazônia durante o Cenozóico: sobre a teoria dos Refúgios da diferenciação biótica. **Estudos Avançados**. v. 16, p. 175-206, 2002.

HALES, K.G.; KOREY, C.A.; LARRACUENTE, A.M.; ROBERTS, D.M. **Genetics**. v. 201, p. 815-842, 2015.

HEBERT, P.D.N.; CYWINSKA, A.; BALL, S.L.; DEWAARD, J.R. Biological identifications through DNA barcodes. **Proc. R. Soc. Lond. B**. v. 270, p. 313-321, 2003.

HOFFMAN, J. L. I.; DASMAHAPATRA, K.K.; AMOS, W.; PHILIPS, C.D.; GELAT, T.S.; BICKMHAM, J.W. Contrasting patterns of genetic diversity at three different genetic markers in a marine mammal metapopulation. **Molecular Ecology**, v. 18, p. 2961-2978, 2009.

HOLT, B.G.; LESSARD, J.P.; BORREGAARD, M.K.; FRITZ, S.A.; ARAÚJO, M.B.; DIMITROV, D.; FABRE, P.H.; GRAHAM, C.H.; GRAVES, G.R.; JØNSSON, K. A.; NOGUÉS-BRAVO, D.; WANG, Z.; WHITTAKER, R.J.; FJELDSÅ, J.; RAHBEK, C. An Update of Wallace's Zoogeographic Regions of the World. **Science**. v. 339, p. 74-77, 2013.

HOSAKI-KOBAYASHI, M.; BICUDO, H.E.M.C. Inversion polymorphism in laboratory strains and natural samples of *Drosophila sturtevantii* (saltans group, sturtevantii subgroup). **Cytobios** v. 89, p. 7-20, 1997.

HOSAKI-KOBAYASHI, M.; BICUDO, H.E.M.C. Productivity in mass crosses of *Drosophila sturtevantii*: a comparative study of laboratory stocks and recently collected flies. **Ver. Brasil. Genet.** v. 17, p. 393-399, 1994.

HOSAKI-KOBAYASHI, M.; BICUDO, H.E.M.C. Reproductive incompatibilities among *Drosophila sturtevantii* laboratory stocks in mass mating and pair mating crosses. **SBPN Scientific Journal**. v.1, n.1, p.16-30, 1998.

IBAMA/MMA. **REBIO Guaribas.** Disponível em: [http://www.icmbio.gov.br/portal/images/stories/imgs-unidades/coservacao/rebio\\_guaribas.pdf](http://www.icmbio.gov.br/portal/images/stories/imgs-unidades/coservacao/rebio_guaribas.pdf) Acesso em 18 de abril de 2019.

INEMA. **Área de Proteção Ambiental do Pratigi.** Disponível em: <http://www.inema.ba.gov.br/gestao-2/unidades-de-conservacao/apa/apa-do-pratigi/> Acesso em 18 de abril de 2019.

INSTITUTO DE PESQUISAS DA BIODIVERSIDADE. **Reserva Betary.** Disponível em: <https://ipbio.org.br/ipbio/> Acesso em 18 de abril de 2019.

JOINVILLE. **Plano de Manejo Executivo.** Disponível em: [http://sistemaspmj.joinville.sc.gov.br/documentos\\_vivacidade/Plano%20de%20Manejo%20APA/Resumo%20Executivo/resumo.pdf](http://sistemaspmj.joinville.sc.gov.br/documentos_vivacidade/Plano%20de%20Manejo%20APA/Resumo%20Executivo/resumo.pdf) Acesso em 18 de abril de 2019.

JOLY, C.A. Biodiversity and climate change in the Neotropical region. **Biota Neotrop.** v. 8, p. 1-2, 2008.

KINGMAN, J.F.C. The coalescent. **Stoch. Process. Their Appl.**, v.13, p.235-248, 1982.

KLINK, C.A.; MACHADO, R.B. Conservation of the Brazilian Cerrado. **Conservation Biology**, v. 19, p. 707–713, 2005.

KNOWLES, L.; MADDISON, W. P. Statistical phylogeography. **Molecular Ecology**, v. 11, p.2623-2635, 2002.

LAJBNER, Z., PNINI, R., CAMUS, M.F., MILLER, J. & DOWLING, D.K. Experimental evidence that thermal selection shapes mitochondrial genome evolution. **Scientific Reports**. v. 9500, p.1-12, 2018.

LEAL, I.R., TABARELLI, M. SILVA, J.M.C. Ecologia e conservação da Caatinga. **Editora Universitária**, Universidade Federal de Pernambuco, Recife, Brasil (in Portuguese), 2003.

LEAL, I.R.; DA SILVA, J.M.C.; TABARELLI, M.; LACHER JR, T.E. Changing the Course of Biodiversity Conservation in the Caatinga of Northeastern Brazil. **Conservation Biology**. v. 19, p. 701-706, 2005.

LEITE, Y. L. R.; COSTA, L. R.; LOSS, A. C.; ROCHA, R. G.; BATALHA-FILHO, H.; BASTOS, A. C.; QUARESMA, V. S.; FAGUNDES, V.; PARESQUE, R.; PASSAMANI, M.; PARDINI, R. Neotropical forest expansion during the last glacial period challenges refuge hypothesis. **Proceedings of the National Academy of Sciences of the USA**, v. 113, p. 1108-1013, 2016.

LIU, H.; BECKENBACH, A. T. Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. **Molecular Phylogenetics and Evolution**, v.1(1), p. 41-52, 1992.

LUNT, D. H.; ZHANG, D.-X.; SZYMURA, J. M.; HEWLTT, O. M. The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. **Insect Molecular Biology**, v.5(3), p. 153-165, 1996.

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

MAGALHÃES, L.E.; BJÖRNBERG A.J.S. Estudo da genitalia masculina de *Drosophila* do grupo *saltans* (Diptera). **Revista Brasileira de Biologia**, v.17, p. 435-450, 1957.

MANFRIN, M. H; SENE, F. M.; Cactophilic *Drosophila* in South America: a model for evolutionary studies. **Genetica**, v. 126, p. 57-75, 2006.

MARGULIES, M.; EGHOLM, M.; ALTMAN, W. E.; ATTIYA, S.; BADER, J. S.; BEMBEN, L. A.; ...; CHEN, Z. Genome sequencing in microfabricated high-density picolitre reactors. **Nature**, v. 437(7057), p. 376-380, 2005.

MARTINS, P.S. Estrutura populacional, fluxo gênico e conservação "in situ". **IPEF**. v. 35, p.71-78, 1987.

METZKER, M.L. Sequencing technologies – the next generation. **Nature Reviews Genetics**. v. 11, p. 31-46, 2010.

MILLER, W.J.; RIEGLER, M. Evolutionary Dynamics of wAu-Like *Wolbachia* Variants in Neotropical *Drosophila* spp. **Applied and Environmental Microbiology**. v.72, p. 826-835, 2006.

MINISTÉRIO DO MEIO AMBIENTE. **Mata Atlântica**. Disponível em <http://www.mma.gov.br/biomas/mata-atlantica> Acesso em 18 de abril de 2019.

MITTERMEIER, R. A.; GIL, P. R.; HOFFMANN, M.; PILGRIM, J.; BROOKS, J.; MIITERMEIER, C. G.; LAMOURUX, J.; FONSECA, G. A. B. Hotspots Revisited:

Earth's Biologically Richest and Most Endangered Terrestrial Ecoregions. **Cemex**, Washington, 2005.

MORRONE, J.J. Biogeographical regionalisation of the Neotropical region. **Zootaxa**. v. 3782, 110p., 2014.

MOURÃO, C. A.; BICUDO, H. E. M. C. Duas novas espécies de *Drosophila* do grupo "saltans" (Drosophilidae Diptera) **Papeis Avulsos Zool.** S. Paulo. v. 20, p. 123-134, 1967.

MÜLLER, M.J.; VON MÜHLEN, C.; VALIATI, V.H.; VALENTE, V.L.S. *Wolbachia pipientis* is associated with different mitochondrial haplotypes in natural populations of *Drosophila willistoni*. **Journal of Invertebrate Pathology**. v. 109, p. 152-155, 2012.

MYERS, N.; MITTERMEIER, R. A.; MITTERMEIER, C. G.; DA FONSECA, G. A. B.; KENT, J. Biodiversity hotspots for conservation priorities. **Nature**, v. 403, p. 853–858, 2000.

NASCIMENTO, A.P.; BICUDO, H.E.M.C. Esterase patterns and phylogenetic relationships of *Drosophila* species in the *saltans* subgroup (*saltans* group). **Genetica** v. 114, p. 41-51, 2002.

NASCIMENTO, A.P.; BICUDO, H.E.M.C. Further study on the esterase patterns of sibling species in the *Drosophila saltans* subgroup (*saltans* group): intraspecific and interespecific variations in the development. **Genetica** v. 126, p. 265-276, 2006.

NECCHI-JUNIOR, O. (Ed.). Fauna e flora dos fragmentos florestais remanescentes da região norte do estado de São Paulo. Ribeirão Preto: **Holos**, 2012. 301 p.

NIELSEN, R.; BEAUMONT, M. Statistical inferences in phylogeography. **Molecular Ecology**, v.18, p.1034-1047, 2009.

O' GRADY, P.M.; CLARK, J.B.; KIDWEL, M.G. Phylogeny of the *Drosophila saltans* species group based on combined analysis of nuclear and mitochondrial DNA sequences. **Mol Biol Evol** v. 15, p. 656-664, 1998.

O'GRADY, P.; DE SALLE, R. Phylogeny of the Genus *Drosophila*. **Genetics**. v. 209, p. 1-25, 2018.

O'GRADY, P.M.; KIDWELL, M.G. Phylogeny of the Subgenus *Sophophora* (Diptera: Drosophilidae) Based on Combined Analysis of Nuclear and Mitochondrial Sequences. **Molecular Phylogenetics and Evolution**. v. 22, p. 442–453, 2002.

O'DEA, A.; LESSIOS, H.A.; COATES, A.G.; EYTAN, R.I.; RESTREPO-MORENO, S.A.; CIONE, A.L.; COLLINS, L.S.; DE QUEIROZ, A.; FARRIS, D.W.; NORRIS, R.D.; STALLARD, R.F.; WOODBURNE, M.O.; AGUILERA, O.; AUBRY, M.P.; BERGGREN, W.A.; BUDD, A.F.; COZZUOL, M.A.; COPPARD, S.E.; DUQUE-CARO,



H.; FINNEGAN, S.; GASPARINI, G.M.; GROSSMAN, E.L.; JOHNSON, K.G.; KEIGWIN, L.D.; KNOWLTON, N.; LEIGH, E.G.; LEONARD-PINGEL, J.S.; MARKO, P.B.; PYENSON, N.D.; RACHELLO-DOLMEN, P.G.; SOIBELZON, E.; SOIBELZON, L.; TODD, J.A.; VERMEIJ, G.J.; JACKSON, J.B. Formation of the Isthmus of Panama. **Sci Adv.** v. 17, p. 1-11, 2016.

OLIVEIRA-FILHO, A. T.; FONTES, M. A. Patterns of floristic differentiation among Atlantic Forests in Southeastern Brazil and the influence of climate. **Biotropica**, v. 32, p. 793–810, 2000.

PADOVAN, M.P.; PEREIRA, Z.V.; FERNANDES, S.S.L. pioneer native tree species in biodiverse agroforestry systems. **Revista GeoPantanal.** v.24, p. 53-68, 2018.

PRANCE, G. T.; Biological Diversification in the Tropics. **Columbia University Press**, New York, 1982.

PRANCE, G.T. "Biogeography of Neotropical plants". In: WHITMORE, T.C.; G.T. PRANCE (eds.): *Biogeography and Quaternary history in tropical America*. **Oxford Monogr. Biogeogr.** v.3, p. 46-65, 214 p., 1987.

PREFEITURA DE ELDORADO DO SUL. **Dados do Município**. Disponível em: [http://eldorado.rs.gov.br/pagina/85\\_Dados-do-Municipio.html](http://eldorado.rs.gov.br/pagina/85_Dados-do-Municipio.html) Acesso em 18 de abril de 2019.

PREFEITURA DE IPORANGA. **Reserva Betary**. Disponível em: <http://iporanga.sp.gov.br/ipotur/reserva-betary/> Acesso em 18 de abril de 2019.

PREFEITURA DE VITÓRIA. Parques: Parque Estadual da Fonte Grande. Disponível em: <http://www.vitoria.es.gov.br/cidade/parques> Acesso em 18 de abril de 2019.

RATHCKE, J. L.; JULES, E. S. Habitat fragmentation and plant-pollinator interactions. **Current Science**, v. 65, p. 273-277, 1993.

RECOR. **Reserva Biológica do IBGE**. Disponível em: <https://recor.org.br/> Acesso em 18 de abril de 2019.

RIBEIRO, M. C.; METZGER, J. P.; MARTENSEN, A. C.; PONZONI, F. J.; HIROTA, M. M. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distribution? Implications for conservation. **Biological Conservation**, v. 142, p. 1141-1153, 2009.

RIZZINI, C. T. Tratado de fitogeografia do Brasil. 2 ed. **Ambito Cultural Edições Ltda**, 1997.

RODRÍGUEZ-TRELLES, F.; TARRÍO, R.; AYALA, F.J. Molecular evolution and phylogeny of the *Drosophila saltans* species group inferred from the *Xdh* gene. **Mol Phylogenet Evol** v. 13, p. 110-121, 1999.

ROESCH, L.F.W.; VIEIRA, F.C.B.; PEREIRA, V.A.; SCHÜNEMANN, A.L.; TEIXEIRA, I.F.; SENNA, A.J.T.; STEFENON, V.M. The Brazilian Pampa: A Fragile Biome. **Diversity**. v. 1, p. 182-198, 2009.

ROMAN, B.E. Reconstrução filogenética do grupo saltans de *Drosophila* utilizando marcadores morfológicos e moleculares. Dissertação de Mestrado. Programa de Pós-Graduação em Biociências. UNESP, 78 p., 2018.

RULL, V. Neotropical biodiversity : timing and potential drivers. **Trends Ecol Evol**. v. 26, p. 508-513, 2011.

SAIA, S.E.M.G.; PESSEDA, L.C.R.; GOUVEIA, S.E.M.; ARAVENA, R.; BENDASSOLLI, J.A. Last glacial maximum (LGM) vegetation changes in the Atlantic Forest, southeastern Brazil. **Quaternary International**. v. 184, p. 195–201, 2008.

SANDEL, B.; ARGE, L.; DALSGAARD, B.; DAVIES, R. G.; GASTON, K. J.; SUTHERLAND, W. J.; SVENNING, J.-C. The Influence of Late Quaternary Climate-Change Velocity on Species Endemism. **Science**, v. 334(6056), p. 660–664, 2011.

SANGER, F.; NICKLEN, S.; COULSON, A.R. DNA sequencing with chain-terminating inhibitors. **Proc Natl Acad Sci U S A**. v. 74, p. 5463–5467, 1977.

SÃO PAULO. **Parque Estadual da Cantareira**. Disponível em: <http://www.saopaulo.sp.gov.br/conhecasp/parques-e-reservas-naturais/parque-estadual-da-cantareira/> Acesso em 18 de abril de 2019.

SÃO PAULO. **Picinguaba**. Disponível em: <http://www.parqueestadualserradomar.sp.gov.br/pesm/nucleos/picinguaba/> Acesso em 18 de abril de 2019.

SARASTE, M. Structural features of cytochrome oxidase. **Quaterly Reviews of Biophysics** v. 23, p. 331–366, 1990.

SCHUSTER, S. C. Next-generation sequencing transforms today's biology. **Nature Methods**, v. 5(1), p. 16–18, 2007.

SCLATER, P.L. On the general Geographical Distribution of the Members of the Class Aves, **Zoological Journal of the Linnean Society**, V.2, p. 130–136, 1858. <https://doi.org/10.1111/j.1096-3642.1858.tb02549.x>

SEGALA, L.F. Diferenciação populacional em *Drosophila sturtevantii* (subgrupo *sturtevantii*, grupo *saltans*) avaliada por morfometria geométrica da asa e do edeago. **Tese de doutorado**. Programa de Pós-Graduação em Biociências. UNESP, 106 p., 2019.

SEMA RS. **Parque Estadual Delta do Jacuí**. Disponível em: <https://www.sema.rs.gov.br/parque-estadual-delta-do-jacui> Acesso em 18 de abril de 2019.

SEMACE. **Área de Relevante Interesse Ecológico do Sítio Curió**. Disponível em: <https://www.semace.ce.gov.br/2010/12/09/area-de-relevante-interesse-ecologico-do-sitio-curio/> Acesso em 19 abril 2019.

SETTA, N.; COSTA, A.P.P.; LOPES, F.R.; CARARETO, C.M.A. 2007b. Transposon display supports transpositional activity of *P* elements in species of the *saltans* group of *Drosophila*. **Journal of Genetics**, v. 86,1, 2007b.

SETTA, N.; LORETO, E.L.S.; CARARETO, C.M.A. Is the Evolutionary History of the O-Type *P* Element in the *saltans* and *willistoni* Groups of *Drosophila* Similar to That of the Canonical *P* Element? **J. Mol. Evol.** v. 65, p. 715-724, 2007a.

SEVEGNANI, L.; UHLMANN, A.; GASPER, A.L.; MEYER, L.; VIBRANS, A.C. Climate affects the structure of mixed rain forest in southern sector of Atlantic domain in Brazil. **Acta Oecologica**. v. 77, p.109-117, 2016.

SILVA J. C., KIDWELL M. G. Evolution of *P* elements in natural populations of *Drosophila willistoni* and *D. sturtevantii*. **Genetics**. v. 168, p. 1323-1335, 2004.

SILVA, J.M.C.; CASTELETI, C.H.M. Status of the biodiversity of the Atlantic Forest of Brazil. In C. Galindo-Leal & I. G. Câmara (Eds.). The Atlantic Forest of South America: biodiversity status, threats, and outlook. **CABS & Island Press**, Washington. p. 43-59, 2003.

SIMON, C.; FRATI, F.; BECKENBACH, A.; CRESPI, B.; LIU, H.; FLOOK, P. 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**, v. 87, p. 651-702, 1994.

SOUZA, T.A.J.; NOLL, F.B.; BICUDO, H.E.M.D.C.; MADI-RAVAZZI, L. Scanning Electron Microscopy of Male Terminalia and Its Application to Species Recognition and Phylogenetic Reconstruction in the *Drosophila saltans* Group. **PLoS ONE** v. 9(6): e97156, 2014.

STEFENON, V.M.; KLABUNDE, G.; LEMOS, R.P.M.; ROGALSKI, M.; NODARI, R.O. Phylogeography of plastid DNA sequences suggests post-glacial southward demographic expansion and the existence of several glacial refugia for *Araucaria angustifolia*. **Scientific Reports**. v.9, 2019.

TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, v.123, p. 585-595, 1989.

TEMPLETON, A. R.; ROBERTSON, R. J.; BRISSON, J.; STRASBURG, J. Disrupting evolutionary processes: The effect of habitat fragmentation on collared lizards in the Missouri Ozarks. **Proceedings of the National Academy of Sciences**, v. 98, p. 5426-5432, 2001.

TEMPLETON, A. R.; SHAW, K.; ROUTMAN, E.; DAVIS, S. K. The genetic consequences of habitat fragmentation. **Annals of the Missouri Botanical Garden**, v. 77, p. 13-27, 1990.

THROCKMORTON, L.H. The phylogeny, ecology, and geography of *Drosophila* [A]. In: King RC. **Handbook of Genetics**. Plenum Press, v. 3, p. 421-469, 1975.

THROCKMORTON, L.H.; MAGALHAES, LE. Changes with evolution of pteridine accumulations in species of the saltans group of the genus *Drosophila*. **University of Texas Publications**, v. 6205, p. 489-505, 1962.

TIDON, R. Relationships between drosophilids (Diptera, Drosophilidae) and the environment in two contrasting tropical vegetations. **Biological Journal of the Linnean Society**. v. 87, p. 233-247, 2006.

TIDON, R.; ALMEIDA, J.M.D. Family Drosophilidae. **Zootaxa**, v. 4122(1), p. 719-751, 2016.

TRAVA, B.M. 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., 2018.

VANZOLINI, P. E.; COSTA A. M.; E VITT, L. Répteis das Caatingas. Rio de Janeiro: **Academia Brasileira de Ciências**, 161p, 1980.

VANZOLINI, P.E. Problemas faunísticos do Cerrado. Simpósio Sobre o Cerrado (ed. by M.G. Ferri. **Editora da Universidade de São Paulo**, São Paulo), pp. 267–280, 1963.

WALLACE, A. Frontmatter. In *The Geographical Distribution of Animals: With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface* (Cambridge Library Collection - Zoology, pp. I-IV). Cambridge: **Cambridge University Press**. (1876, 2011)

WERNECK, F.P. The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. **Quaternary Science Reviews**, v. 30, p. 1630–1648, 2011.

WERNECK, F.P.; NOGUEIRA, C.; COLLI, G.R.; SITES JR, J.W.; COSTA, G.C. Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. **Journal of Biogeography**. doi:10.1111/j.1365-2699.2012.02715.x, 2012.

WWF **Deserts and xeric shrublands: Northern South America: Northeastern Brazil** Disponível em: <https://www.worldwildlife.org/ecoregions/nt1304> Acesso em 18 de abril de 2019.

## Apêndice A – Customized perl scripts used to IBE analysis

```
#!/usr/bin/perl -w
#Partial Mantel test (isolation-by-ecology)
#use strict;

my $Outputfile = "IBEchr.txt";
open O, ">>$Outputfile";

my $chromosome='3R';
my $environment='UV index';
print "chr\tenvironment\ttrGD\ttrED\ttrGE\tIBE\tp\n";
#print O "chr\tenvironment\ttrGD\ttrED\ttrGE\tIBE\tp\n";
print "$chromosome\t$environment\t";
print O "$chromosome\t$environment\t";

my $chr = "MeanFst3R.txt";
open(CHR, $chr);
my $geo = "geodist.txt";
open(GEO, $geo);
my $eco = "UVdist.txt";
open(ECO, $eco);

#Read the genetic matrix
my $c=0;
my @C = ();
while (<CHR>){
chomp;
my @c = split '\t',$_;
$c = scalar(@c);
push @C, [@c];
@c=();
}
#Transform the genetic matrix into a vector
my $i=0;
my $j=0;
my @chrdist=();
for ($i = 1; $i < $c; $i++){
for ($j = 1; $j < $c; $j++){
if ($i < $j){
push @chrdist, $C[$i][$j];
}
}
}

#Read the geographical distance matrix
my $g=0;
my @G = ();
while (<GEO>){
chomp;
my @g = split '\t',$_;
$g = scalar(@g);
push @G, [@g];
```

```

@g=();
}
#Transform the geographical matrix into a vector
my @geodist=();
for ($i = 1; $i < $g; $i++){
for ($j = 1; $j < $g; $j++){
if ($i < $j){
push @geodist, $G[$i][$j];
}
}
}
}

```

```

#Read the ecological distance matrix
my $e=0;
my @E = ();
while (<ECO>){
chomp;
my @e = split '\t',$_;
$e = scalar(@e);
push @E, [@e];
@e=();
}
#Transform the geographical matrix into a vector
my @ecodist=();
for ($i = 1; $i < $e; $i++){
for ($j = 1; $j < $e; $j++){
if ($i < $j){
push @ecodist, $E[$i][$j];
}
}
}
}

```

```

#Size
my $NX = scalar(@chrdist);
my $NY = scalar(@geodist);
my $NZ = scalar(@ecodist);
my $N = ($NX + $NY + $NZ)/3;

```

```

#Mean
my $SumX = 0;
foreach (@chrdist){
$SumX += $_;
}
my $MeanX = ($SumX/$NX);
my $SumY = 0;
foreach (@geodist){
$SumY += $_;
}
my $MeanY = ($SumY/$NY);
my $SumZ = 0;
foreach (@ecodist){
$SumZ += $_;
}
}

```

```

my $MeanZ = ($SumZ/$NZ);

#Covariance
my @CovarianceXY;
for ($i = 0; $i < $N; $i++) {
push @CovarianceXY, ($chrdist[$i] - $MeanX) * ($geodist[$i] - $MeanY);
}
my $SumcovXY = 0;
foreach (@CovarianceXY){
$SumcovXY += $_;
}
my @CovarianceYZ;
for ($i = 0; $i < $N; $i++) {
push @CovarianceYZ, ($geodist[$i] - $MeanY) * ($ecodist[$i] - $MeanZ);
}
my $SumcovYZ = 0;
foreach (@CovarianceYZ){
$SumcovYZ += $_;
}
my @CovarianceXZ;
for ($i = 0; $i < $N; $i++) {
push @CovarianceXZ, ($chrdist[$i] - $MeanX) * ($ecodist[$i] - $MeanZ);
}
my $SumcovXZ = 0;
foreach (@CovarianceXZ){
$SumcovXZ += $_;
}

#Variance
my @Dev2X;
foreach (@chrdist){
push @Dev2X, ($_ - $MeanX) ** 2;
}
my $Sumdev2X = 0;
foreach (@Dev2X){
$Sumdev2X += $_;
}
my $VarianceX = ($Sumdev2X/($NX-1));
my @Dev2Y;
foreach (@geodist){
push @Dev2Y, ($_ - $MeanY) ** 2;
}
my $Sumdev2Y = 0;
foreach (@Dev2Y){
$Sumdev2Y += $_;
}
my $VarianceY = ($Sumdev2Y/($NY-1));
my @Dev2Z;
foreach (@ecodist){
push @Dev2Z, ($_ - $MeanZ) ** 2;
}
my $Sumdev2Z = 0;
foreach (@Dev2Z){
$Sumdev2Z += $_;
}

```

```
my $VarianceZ = ($Sumdev2Z/($NZ-1));
```

```
#Standard deviation
```

```
my $StdevX = $VarianceX ** 0.5;
```

```
my $StdevY = $VarianceY ** 0.5;
```

```
my $StdevZ = $VarianceZ ** 0.5;
```

```
#Correlation
```

```
my $rGD = 0;
```

```
$rGD = $SumcovXY / (($N - 1)*$StdevX*$StdevY);
```

```
print "$rGD\t";
```

```
print O "$rGD\t";
```

```
my $rED = 0;
```

```
$rED = $SumcovYZ / (($N - 1)*$StdevY*$StdevZ);
```

```
print "$rED\t";
```

```
print O "$rED\t";
```

```
my $rGE = 0;
```

```
$rGE = $SumcovXZ / (($N - 1)*$StdevX*$StdevZ);
```

```
print "$rGE\t";
```

```
print O "$rGE\t";
```

```
my $IBE = 0;
```

```
$IBE = ($rGE-($rGD*$rED)) / (((1-($rGD ** 2)) ** 0.5) * ((1-($rED ** 2)) ** 0.5));
```

```
print "$IBE\t";
```

```
print O "$IBE\t";
```

```
#Transform @chr dist to a symmetrical matrix
```

```
my $k=-1;
```

```
my @row;
```

```
my @chr matrix;
```

```
my @pop =
```

```
('CO','ED','GA','GU','KR','NG','RG','TZ','UG','ZI','ZS','EA','EB','EF','EM','ER','EZ','KN','SB','SD','SE','SF','SP','UK','UM','ZK');
```

```
for ($i = 0; $i < @pop; $i++){
```

```
for ($j = 0; $j < @pop; $j++){
```

```
if ($i < $j){
```

```
$k++;
```

```
push @row, $chr dist[$k];
```

```
}
```

```
elsif ($i > $j){
```

```
push @row, $chr matrix[$j][$i];
```

```
}
```

```
else{
```

```
push @row, 0;
```

```
}
```

```
}
```

```
push @chr matrix, [@row];
```

```
@row=();
```

```
}
```

```
my @original = (0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25);
```

```
my $permt = 10000;
```

```
my $count = 0;
```

```
my $rGpD = 0;
```

```
my $rGpE = 0;
```

```
my $IBEp = 0;
```



```

for (1..$permt){
#Matrix permutation
use List::Util qw( shuffle );
my @permutated = shuffle @original;
my @mpr=();
my @matrixp=();
for ($i = 0; $i < @pop; $i++){
for ($j = 0; $j < @pop; $j++){
push @mpr, $chrmatrix[$permutated[$i]][$permutated[$j]];
}
push @matrixp, [@mpr];
@mpr=();
}

#Transform the permutated matrix into a vector
my @chrdistp=();
for ($i = 0; $i < @pop; $i++){
for ($j = 0; $j < @pop; $j++){
if ($i < $j){
push @chrdistp, $matrixp[$i][$j];
}
}
}

#Size
my $NXp = scalar(@chrdistp);

#Mean
my $SumXp = 0;
foreach (@chrdistp){
$SumXp += $_;
}
my $MeanXp = ($SumXp/$NXp);

#Covariance
my @CovarianceXpY;
for ($i = 0; $i < $NXp; $i++) {
push @CovarianceXpY, ($chrdistp[$i] - $MeanXp) * ($geodist[$i] - $MeanY);
}
my $SumcovXpY = 0;
foreach (@CovarianceXpY){
$SumcovXpY += $_;
}
my @CovarianceXpZ;
for ($i = 0; $i < $NXp; $i++) {
push @CovarianceXpZ, ($chrdistp[$i] - $MeanXp) * ($ecodist[$i] - $MeanZ);
}
my $SumcovXpZ = 0;
foreach (@CovarianceXpZ){
$SumcovXpZ += $_;
}

#Variance
my @Dev2Xp;
foreach (@chrdistp){

```

```

push @Dev2Xp, ($_ - $MeanXp) ** 2;
}
my $Sumdev2Xp = 0;
foreach (@Dev2Xp){
$Sumdev2Xp += $_;
}
my $VarianceXp = ($Sumdev2Xp/($NXp-1));

#Standard deviation
my $StdevXp = $VarianceXp ** 0.5;

#Correlation
$rGpD = $SumcovXpY / (($NXp - 1)*$StdevXp*$StdevY);
$rGpE = $SumcovXpZ / (($NXp - 1)*$StdevXp*$StdevZ);
$IBEp = ($rGpE-($rGpD*$rED)) / (((1-($rGpD ** 2)) ** 0.5) * ((1-($rED ** 2)) ** 0.5));
if ($IBEp >= $IBE){
$count++;
}
}
my $p = $count/$permt;
print "$p\n";
print O "$p\n";

```

```

exit;

```

```

#!/usr/bin/perl -w
#Mantel test (isolation-by-distance)
#use strict;

```

```

my $Outputfile = "IBDchreco.txt";
open O, ">>$Outputfile";

my $chromosome='3L';
print "chr\t\r\tp\n";
#print O "chr\t\r\tp\n";
print "$chromosome\t";
print O "$chromosome\t";

#Read the genetic matrix
my $chr = "gendist1.txt";
open(CHR, $chr);
my $c=0;
my @C = ();
while (<CHR>){
chomp;
my @c = split '\t',$_;
$c = scalar(@c);
push @C, [@c];
@c=();
}
#Transform the genetic matrix into a vector
my @chrdist=();
for ($i = 1; $i < $c; $i++){
for ($j = 1; $j < $c; $j++){
if ($i < $j){
push @chrdist, $C[$i][$j];
}
}
}
#print @chrdist, "\n";

#Read the geographical distance matrix
my $geo = "geodist1.txt";
open(GEO, $geo);
my $g=0;
my @G = ();
while (<GEO>){
chomp;
my @g = split '\t',$_;
$g = scalar(@g);
push @G, [@g];
@g=();
}
#Transform the geographical matrix into a vector
my $i=0;
my $j=0;
my @geodist=();
for ($i = 1; $i < $g; $i++){
for ($j = 1; $j < $g; $j++){
if ($i < $j){
push @geodist, $G[$i][$j];
}
}
}

```

```

}
}
#print @geodist, "\n";
#print M "@geodist\n";

#Size
my $NX = scalar(@chrdist);
#print "NX = $NX\n";
my $NY = scalar(@geodist);
#print "NY = $NY\n";
my $N = ($NX + $NY)/2;

#Mean
my $SumX = 0;
foreach (@chrdist){
$SumX += $_;
}
#print "SumX = $SumX\n";
my $MeanX = ($SumX/$NX);
#print "MeanX = $MeanX\n";
my $SumY = 0;
foreach (@geodist){
$SumY += $_;
}
#print "SumY = $SumY\n";
my $MeanY = ($SumY/$NY);
#print "MeanY = $MeanY\n";

#Covariance
my @Covariance;
for ($i = 0; $i < $N; $i++) {
push @Covariance, ($chrdist[$i] - $MeanX) * ($geodist[$i] - $MeanY);
}
#print "Covariance = @Covariance\n";
my $Sumcov = 0;
foreach (@Covariance){
$Sumcov += $_;
}
#print "Sumcov = $Sumcov\n";

#Variance
my @Dev2X;
foreach (@chrdist){
push @Dev2X, ($_ - $MeanX) ** 2;
}
my $Sumdev2X = 0;
foreach (@Dev2X){
$Sumdev2X += $_;
}
my $VarianceX = ($Sumdev2X/($NX-1));
#print "VarianceX = $VarianceX\n";
my @Dev2Y;
foreach (@geodist){
push @Dev2Y, ($_ - $MeanY) ** 2;
}

```

```

my $Sumdev2Y = 0;
foreach (@Dev2Y){
$Sumdev2Y += $_;
}
my $VarianceY = ($Sumdev2Y/($NY-1));
#print "VarianceY = $VarianceY\n";

#Standard deviation
my $StdevX = $VarianceX ** 0.5;
#print "Standard deviation X = $StdevX\n";
my $StdevY = $VarianceY ** 0.5;
#print "Standard deviation Y = $StdevY\n";

#Correlation
my $r = 0;
$r = $Sumcov / (($N - 1)*$StdevX*$StdevY);
print "$r\t";
print O "$r\t";

#Transform @chr dist to a symmetrical matrix
my $k=-1;
my @row;
my @chr matrix;
my @pop =
('TIR','CUR','GUA','PER','PRA','BRA','VIT','MAT','NG','DOU','PIC','CANT','BET','PI','RI','AG','E
LD');
for ($i = 0; $i < @pop; $i++){
for ($j = 0; $j < @pop; $j++){
if ($i < $j){
$k++;
push @row, $chr dist[$k];
}
elseif ($i > $j){
push @row, $chr matrix[$j][$i];
}
else{
push @row, 0;
}
}
}
#print "@row\n";
#print M "@row\n";
push @chr matrix, [@row];
@row=();
}

my @original = (0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16);
my $permt = 10000;
my $count = 0;
my $rp = 0;
for (1..$permt){
#Matrix permutation
use List::Util qw( shuffle );
my @permutated = shuffle @original;
#print "@permutated\n";
#print M "@permutated\n";
}

```

```

#my @popp;
#for ($i = 0; $i < @pop; $i++){
#push @popp, $pop[$permutated[$i]];
#}
#print "@popp\n";
#print M "@popp\n";
my @mpr=();
my @matrixp=();
for ($i = 0; $i < @pop; $i++){
for ($j = 0; $j < @pop; $j++){
push @mpr, $chrmatrix[$permutated[$i]][$permutated[$j]];
}
#print "@mpr\n";
#print M "@mpr\n";
push @matrixp, [@mpr];
@mpr=();
}
#Transform the permuted matrix into a vector
my @chrdistp=();
for ($i = 0; $i < @pop; $i++){
for ($j = 0; $j < @pop; $j++){
if ($i < $j){
push @chrdistp, $matrixp[$i][$j];
}
}
}
#print @chrdistp, "\n";
#print M "@chrdistp\n";

#Size
my $NXp = scalar(@chrdistp);
#print "NXp = $NXp\n";
my $Np = ($NXp + $NY)/2;

#Mean
my $SumXp = 0;
foreach (@chrdistp){
$SumXp += $_;
}
#print "SumX = $SumX\n";
my $MeanXp = ($SumXp/$NXp);
#print "MeanXp = $MeanXp\n";

#Covariance
my @Covariancep;
for ($i = 0; $i < $Np; $i++) {
push @Covariancep, ($chrdistp[$i] - $MeanXp) * ($geodist[$i] - $MeanY);
}
#print "Covariancep = @Covariancep\n";
my $Sumcovp = 0;
foreach (@Covariancep){
$Sumcovp += $_;
}
#print "Sumcovp = $Sumcovp\n";

```

```

#Variance
my @Dev2Xp;
foreach (@chrdistp){
push @Dev2Xp, ($_ - $MeanXp) ** 2;
}
my $Sumdev2Xp = 0;
foreach (@Dev2Xp){
$Sumdev2Xp += $_;
}
my $VarianceXp = ($Sumdev2Xp/($NXp-1));
#print "VarianceXp = $VarianceXp\n";

#Standard deviation
my $StdevXp = $VarianceXp ** 0.5;
#print "Standard deviation Xp = $StdevXp\n";

#Correlation
$rp = $Sumcovp / (($Np - 1)*$StdevXp*$StdevY);
#print "rp = $rp\n";
if ($rp >= $r){
$count++;
}
}
my $p = $count/$permt;
print "$p\n";
print O "$p\n";

exit;

```

## Apêndice B – Climatic variables and Distance matrices

Climatic variables as obtained from the NASA POWER Project Data Sets, available at <https://power.larc.nasa.gov/>

1. Airtemperature (air temperature)
2. Tmax (maximum temperature)
3. Tmin (minimum temperature)
4. Earthtemp (temperature on Earth surface)
5. RH (relative humidity)
6. Nosun
7. Atmosphericpressure
8. Clearsky
9. Diffuseradiation
10. Directnormalradiation
11. InsolationIncident
12. Latitudetilradiation
13. Windspeed





ELD	14.979379	15.573728	13.584046	12.754184	10.496118	18.165341	10.130933	12.77939	15.259839
	13.939591	7.2659961	6.4063406	5.0526924	4.7640634	4.2518467	5.3150353	0	

### Matrix 2 - Fst-based genetic matrix

	TIR	CUR	GUA	PER	PRA	BRA	VIT	MAT	NG	DOU	PIC	CANT	BET	PI	RI	AG	ELD
TIR	0.0000	0.0000	0.2424	0.2902	0.3590	0.7895	0.2916	0.3660	0.5333	0.5983	0.5185	0.3704	0.5128	0.2778	0.6013	0.5139	0.2540
CUR	0.0000	0.0000	0.1524	0.2064	0.2393	0.5217	0.1905	0.2569	0.3647	0.4308	0.3363	0.2497	0.3419	0.1975	0.4350	0.3654	0.1632
GUA	0.2424	0.1524	0.0000	0.0000	0.0351	0.3415	0.0315	0.0994	0.1770	0.2634	0.1705	0.0495	0.4242	0.1641	0.4957	0.4217	0.0882
PER	0.2902	0.2064	0.0000	0.0000	0.0000	0.1513	0.0000	0.0057	0.0378	0.1177	0.0541	0.0000	0.4189	0.1550	0.4795	0.4142	0.0571
PRA	0.3590	0.2393	0.0351	0.0000	0.0000	0.0690	0.0000	0.0000	0.0000	0.0486	0.0000	0.0000	0.4722	0.2014	0.5318	0.4735	0.0574
BRA	0.7895	0.5217	0.3415	0.1513	0.0690	0.0000	0.1029	0.0191	0.0000	0.0000	0.0000	0.1333	0.7042	0.4086	0.7309	0.6833	0.2424
VIT	0.2916	0.1905	0.0315	0.0000	0.0000	0.1029	0.0000	0.0000	0.0052	0.0506	0.0000	0.0000	0.4404	0.1712	0.5016	0.4440	0.0353
MAT	0.3660	0.2569	0.0994	0.0057	0.0000	0.0191	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.4276	0.1598	0.4745	0.4252	0.0389
NG	0.5333	0.3647	0.1770	0.0378	0.0000	0.0000	0.0052	0.0000	0.0000	0.0000	0.0000	0.0107	0.5355	0.2475	0.5794	0.5250	0.0915
DOU	0.5983	0.4308	0.2634	0.1177	0.0486	0.0000	0.0506	0.0000	0.0000	0.0000	0.0207	0.0990	0.6014	0.3417	0.6343	0.5933	0.1900
PIC	0.5185	0.3363	0.1705	0.0541	0.0000	0.0000	0.0000	0.0000	0.0000	0.0207	0.0000	0.0419	0.5604	0.2557	0.6032	0.5521	0.1111
CANT	0.3704	0.2497	0.0495	0.0000	0.0000	0.1333	0.0000	0.0000	0.0107	0.0990	0.0419	0.0000	0.4701	0.1949	0.5259	0.4640	0.0779
BET	0.5128	0.3419	0.4242	0.4189	0.4722	0.7042	0.4404	0.4276	0.5355	0.6014	0.5604	0.4701	0.0000	0.0869	0.0741	0.0388	0.2055
PI	0.2778	0.1975	0.1641	0.1550	0.2014	0.4086	0.1712	0.1598	0.2475	0.3417	0.2557	0.1949	0.0869	0.0000	0.1026	0.0580	0.0000
RI	0.6013	0.4350	0.4957	0.4795	0.5318	0.7309	0.5016	0.4745	0.5794	0.6343	0.6032	0.5259	0.0741	0.1026	0.0000	0.0035	0.2751
AG	0.5139	0.3654	0.4217	0.4142	0.4735	0.6833	0.4440	0.4252	0.5250	0.5933	0.5521	0.4640	0.0388	0.0580	0.0035	0.0000	0.2125
ELD	0.2540	0.1632	0.0882	0.0571	0.0574	0.2424	0.0353	0.0389	0.0915	0.1900	0.1111	0.0779	0.2055	0.0000	0.2751	0.2125	0.0000

### Matrix 3 – Geographical distances (km)

Geography	TIR	CUR	GUA	PER	PRA	BRA	VIT	MAT	NG	DOU	PIC	CANT	BET	PI	RI	AG	ELD
TIR	0	5300	5750	5650	5640	4960	5900	5280	5140	4840	5700	5560	5480	5580	5720	5720	5710
CUR	5300	0	490	580	1120	1700	1850	2250	2200	2700	2280	2355	2555	2740	2870	3000	3210
GUA	5750	490	0	305	915	1725	1615	2190	2170	2725	2120	2230	2460	2630	2730	2900	3095
PER	5650	580	305	0	615	1430	1330	1880	1870	2420	1820	1930	2150	2325	2430	2600	2800
PRA	5640	1120	915	615	0	955	730	1300	1890	1210	1320	1550	1720	1820	1960	2180	
BRA	4960	1700	1725	1430	955	0	940	630	540	1010	880	850	970	1160	1310	1420	1605
VIT	5900	1850	1615	1330	730	940	0	850	935	1520	580	740	980	1110	1170	1315	1540
MAT	5280	2250	2190	1880	1300	630	850	0	140	660	420	280	330	530	680	790	980

NG	5140	2200	2170	1870	1300	540	935	140	0	600	555	420	450	635	795	890	1065
DOU	4840	2700	2725	2420	1890	1010	1520	660	600	0	1030	850	690	750	880	890	935
PIC	5700	2280	2120	1820	1210	880	580	420	555	1030	0	190	410	530	615	750	980
CANT	5560	2355	2230	1930	1320	850	740	280	420	850	190	0	240	400	510	645	860
BET	5480	2555	2460	2150	1550	970	980	330	450	690	410	240	0	190	350	460	660
PI	5580	2740	2630	2325	1720	1160	1110	530	635	750	530	400	190	0	165	265	470
RI	5720	2870	2730	2430	1820	1310	1170	680	795	880	615	510	350	165	0	140	370
AG	5720	3000	2900	2600	1960	1420	1315	790	890	890	750	645	460	265	140	0	203
ELD	5710	3210	3095	2800	2180	1605	1540	980	1065	935	980	860	660	470	370	203	0