UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

JAQUELINE REGINATO KOSER

Estrutura populacional e filogeografia de *Drosophila antonietae* Tidon & Sene

RIBEIRÃO PRETO, SP.

- 2015 -

JAQUELINE REGINATO KOSER

Estrutura populacional e filogeografia de *Drosophila antonietae* Tidon & Sene

Dissertação apresentada à Faculdade de
Medicina de Ribeirão Preto da Universidade de
São Paulo para obtenção do título de Mestre em
Ciências.

Área de concentração: GENÉTICA

Orientadora: Prof^a. Dr^a. Maura Helena Manfrin

RIBEIRÃO PRETO, SP.

- 2015 -

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Koser, Jaqueline Reginato

Estrutura populacional e filogeografia de *Drosophila antonietae*

Tidon & Sene

Ribeirão Preto, 2015.

XXp. : il. Xcm

Dissertação de Mestrado, apresentada à Faculdade de
Medicina de Ribeirão Preto/USP. Área de concentração: Genética.

Orientador: Manfrin, Maura Helena

1. *Cluster Buzattii*.
2. Dinâmica populacional.
3. Núcleo de *Missiones*.
4. mtCOI.
5. Gene *period*.

Nome: Koser, Jaqueline Reginato

Título: Estrutura populacional e filogeografia de *Drosophila antonietae* Tidon & Sene

Dissertação apresentada à Faculdade de
Medicina de Ribeirão Preto da Universidade de
São Paulo para a obtenção do título de Mestre
em Genética

Aprovado em:

BANCA EXAMINADORA

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

Prof. Dr. _____

Instituição: _____

Assinatura: _____

AGRADECIMENTOS

Ao Departamento de Genética da Faculdade de Medicina de Ribeirão Preto, em especial à secretária Silvia Sant'anna Consiglieri e à Susie Adriana Penha Nalon, secretária da pós-graduação.

A minha orientadora Maura Helena Manfrin, especialmente pela oportunidade concedida, e ao Fábio de Melo Sene, também presente nesta fase.

Aos orientadores do Programa de Aperfeiçoamento em Ensino – PAE, Maura Helena Manfrin e Victor Ferraz, pela oportunidade de acompanhamento e aprendizagem.

Às colegas de laboratório, Natácia Evangelista de Lima, Dora Yovanna Barrios Leal, Tais Carmona Lavagnini, Gislaine da Silva, Camila Borgonove, Camila Kokudai, Rafaela Rossetti, Rogério Picela Matheus, Luciana Paes de Barros Machado.

Ao técnico de laboratório, Paulo Epifânio, por todos os ensinamento sobre os cuidados com as moscas.

Ao Lucas Poppe, por ceder material biológico da localidade de Bossoroca – RS.

Às pessoas que auxiliaram na coleta, Rudi Laps, Felipe Barretos, Natácia e Dora. E também aos proprietários que abriram suas propriedades para as coletas.

Às pessoas que colaboraram no encontro dos locais de coleta, como o Rudi Laps, Felipe Barretos, Tatiana Souza do Amaral, Tiana Custódio, Fernando Straube e outros.

Às pessoas que colaboraram com as análises: Fernando Lopes, Danillo Silva, em especial às colegas de laboratório Tais, Dora e Natácia.

Amigos que fiz em Ribeirão Preto e aos de Blumenau. Em especial aqueles que leram as versões preliminares, Júnia Schultz, João Paulo de Maçaneiro, Thiago Sanches Ranzani da Silva, Dora e Natácia, que revisaram e corrigiram o texto com muita presteza.

À minha família, pelo apoio e suporte.

Aos órgãos financiadores, CNPq pela concessão da bolsa, e também pelos financiamentos CAPES, FAPESP e FINEP.

*Se todos aqueles cujos experimentos malograram deixassem de fazer ciência, não
teríamos ciência.*

Allan Wilson

RESUMO

Koser, J. R. (2015). *Estrutura populacional e filogeografia de Drosophila antonietae Tidon & Sene*. Dissertação de Mestrado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo.

A espécie *Drosophila antoneitae* é cactófila, ovipositando principalmente no cacto hospedeiro *Cereus hildmaniannus*. Ambas as espécies são encontradas em solos drenados ou afloramentos de rocha na região de *Misiones*, um dos núcleos de Floresta Tropical Sazonalmente Seca (FTSS), e no litoral sul-brasileiro. O núcleo de *Misiones* compreende a base dos rios Paraná-Paraguai e é uma possível área de estabilidade climática durante as oscilações do Quaternário. Para avaliar a distribuição atual da diversidade genética de *D. antonietae* e sua associação com alterações na paisagem, foram realizadas as seguintes análises: estruturação populacional, estabelecimento de hipóteses filogeográficas e de eventos demográficos, testes de neutralidade e modelagem de nicho atual e paleoecológico. O gene mitocondrial COI e nuclear *period* foram analisados. O gene *period* é pouco variável e a distribuição da variabilidade genética é homogênea. O gene COI é bastante polimórfico e há estruturação entre as populações formando três agrupamentos: um em Santa Catarina e Rio Grande do Sul, outro no Paraná e São Paulo e o terceiro no centro-oeste brasileiro. A população do centro-oeste brasileiro provavelmente é a mais antiga, coincidente com a área do estudo mais estável climaticamente. Este agrupamento forma uma rede de haplótipos separada, devido alta estruturação e isolamento, provavelmente tendo no rio Paraná uma importante barreira de fluxo gênico. Sugere-se que o provável centro de dispersão das demais populações de *D. antonietae* se localiza no sudeste brasileiro, e que houve diversos eventos de migração para as demais regiões de sua distribuição. Há indícios de polimorfismo compartilhado devido à recente diversificação das populações. Ambos os agrupamentos genéticos exibem sinais de expansão populacional, especialmente nas áreas de borda no núcleo de *Misiones*, onde o clima parece ter sido menos estável. O período de expansão demográfica é recente e coincidente com a maior extensão da vegetação seca, que também pode ter papel fundamental na estruturação das populações.

ABSTRACT

Koser, J. R. (2015). *Population structure and phylogeography of Drosophila antonietae* Tidon & Sene. Dissertação de Mestrado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo.

Drosophila antoneitae is a cactophilic species, ovipositing primarily in the host cacti *Cereus hildmaniannus*. Both species are found in drained soils or rocky outcrops in the *Misiones* region - one of the Seasonally Dry Tropical Forest (SDTF) nuclei, and in south Brazilian coast. The *Misiones* nuclei comprises the basin of the Paraná-Paraguay Rivers and it is a possible area of climatic stability during the Quaternary oscillations. To evaluate the current distribution of the genetic diversity of *D. antoneitae* and its association with landscape modifications, the following analyses were performed: populational structure, establishment of phylogeographic hypotheses and demographic events, neutrality tests and paleoecological niche modeling. The mitochondrial gene COI and the nuclear gene *period* were analyzed. The gene *period* had low genetic diversity and an homogeneity on the distribution of genetic variability. For gene COI analysis we found a high polymorphism and genetic structure among populations, forming three groups: one in Santa Catarina and Rio Grande do Sul, another in Paraná and São Paulo and the third in the midwestern Brazil. The midwestern population is probably the oldest one, coinciding with the most climatically stable area of this study. This group forms a network of separate haplotypes, due to a high structuring and isolation, probably with Paraná River acting as major barrier for gene flow. We suggest that the possible center of dispersion of the remaining populations of *D. antonietae* is located in southeastern Brazil, and there were several migration events to other regions of its distribution. There is evidence of shared polymorphism due to recent diversification of populations. Both gene clusters exhibit signs of population expansion, especially in border areas at the *Misiones* nuclei, where the climate seems less stable. The demographic expansion period is recent and coincides with the major expansion of dry vegetation, which can also play a critical role in structuring populations.

SUMÁRIO

1.	INTRODUÇÃO	11
1.1.	Domínios paisagísticos e alterações climáticas	11
1.2.	O cluster <i>Drosophila buzzatii</i> e a espécie <i>Drosophila antonietae</i>.	13
1.3.	Filogeografia e Teoria da Coalescência	15
1.4.	Marcadores Moleculares	17
1.4.1.	<i>Citocromo Oxidase I – COI</i>	17
1.4.2.	<i>Period</i>	18
1.5.	Modelagem de distribuição de espécies	19
2.	OBJETIVOS	20
3.	MATERIAL E MÉTODOS	21
3.1.	Amostragem	21
3.2.	Obtenção das sequências de DNA	24
3.3.	Análise dos dados genéticos	25
3.3.1.	<i>Leitura e Alinhamento das Sequências</i>	25
3.3.2.	<i>Estatística Descritiva</i>	25
3.3.3.	<i>Modelos de Substituição</i>	26
3.3.4.	<i>Determinação de linhagens</i>	26
3.3.5.	<i>Análises Filogeográficas</i>	26
3.3.6.	<i>Estruturação Populacional</i>	27
3.3.7.	<i>Testes Demográficos</i>	28
3.4.	Modelagem de Distribuição da Espécie	28
4.	RESULTADOS	32
4.1.	Análises das sequências do gene nuclear period	32
4.1.1.	<i>Distribuição dos Haplótipos</i>	33
4.1.2.	<i>Determinação de linhagens</i>	33
4.1.3.	<i>Análises Filogeográficas</i>	34
4.1.4.	<i>Estruturação Populacional</i>	35
4.1.5.	<i>Testes Demográficos</i>	37
4.2.	Análises das sequências do gene mitocondrial COI	38

4.2.1.	<i>Distribuição dos haplótipos</i>	39
4.2.2.	<i>Determinação de linhagens</i>	40
4.2.3.	<i>Análises Filogeográficas</i>	42
4.2.4.	<i>Estruturação Populacional</i>	48
4.2.5.	<i>Testes Demográficos</i>	50
4.3.	Modelagem de distribuição da espécie	54
4.3.1.	<i>Projeção atual</i>	54
4.3.2.	<i>Projeção Miroc</i>	55
4.3.3.	<i>Projeção CCSM</i>	55
5.	DISCUSSÃO	57
6.	CONCLUSÕES	63
7.	REFERÊNCIAS	64

1. INTRODUÇÃO

1.1. Domínios paisagísticos e alterações climáticas

A América do Sul é caracterizada pela combinação de elevada diversidade climática e geomorfológica, a qual permitiu o desenvolvimento de diversas formações vegetacionais, desde florestas úmidas até regiões semi-áridas (Hueck, 1972). No continente Sul-americano há uma faixa de vegetação sazonalmente seca e aberta, orientada na direção nordeste-sudoeste (Hueck, 1972; Ab'Saber, 2003), composta pelos domínios da Caatinga, Cerrado, do Núcleo de *Misiones* e do Chaco, entre as duas principais florestas tropicais úmidas da América do Sul: a Floresta Amazônica a noroeste e a Mata Atlântica contornando o litoral à leste.

Nos últimos anos, trabalhos vêm demonstrando que a distribuição atual destas formações, assim como de seus elementos, foi constantemente influenciada pelas alterações climáticas dos ciclos glaciais (Prado & Gibbs, 1993; Pennington, Prado, & Pendry, 2000). Estes ciclos iniciaram no final do Terciário e se estenderam com diferentes intensidades durante o Quaternário, que foi marcado pela alternância de épocas glaciais e interglaciais (Ab'Saber, 1977; Bennett, 1990; Bennett, 2004; Hewitt, 2011). Estas alterações podem ter levado a mudanças ambientais significativas e consequentemente a alterações na distribuição das espécies (Prado & Gibbs, 1993; Pennington et al., 2000; Zeisset & Beebe, 2008; Carnaval, Hickerson, Haddad, Rodrigues, & Moritz, 2009; Thomé et al., 2010).

Uma das mudanças mais marcantes é o deslocamento da vegetação em função da umidade disponível. Nos períodos interglaciais, a alta umidade favorece a expansão das florestas úmidas, que isolam os fragmentos de vegetação xeromórfica que resistem localmente em ambientes mais drenados. Já nos períodos glaciais, secos e frios, a vegetação xerófila se expande isolando as florestas úmidas em enclaves (Ab'Saber, 1977; Bennett, 2004; Hewitt, 2011). Estas oscilações alteraram, por exemplo, a distribuição das Florestas Tropicais Sazonalmente Secas (FTSS), que possivelmente são remanescentes de uma formação única, que teria se separado e se unido novamente por

diversas vezes entre 6 mil a 21 mil anos antes do presente (Ab'Saber, 1977; Carnaval et al., 2009; Mayle, 2006; Werneck, Costa, Colli, Prado, & Sites, 2011a).

As FTSS (Figura 1) são formações encontradas nos continentes da América do Sul e América Central. Dentre elas, estão os núcleos de *Misiones* e a Caatinga na América do Sul. Ambas ocorrem em locais com precipitação menor que 1600 mm/ano e que durante 5 a 6 meses recebem menos de 100 mm. O solo é fértil, o pH e o nível de nutrientes está entre moderado a alto e possui baixos teores de alumínio. Há a marcada presença de espécies conhecidas como “sempre verdes”, suculentas e os cactos, que são elementos importantes da diversidade destas florestas. Estes são ecossistemas dominados por árvores com o dossel parcialmente contínuo, caracterizada por uma vegetação, em sua maioria, decídua durante a estação seca (Pennington et al., 2000).

Figura 1. Distribuição da vegetação sazonalmente seca no Neotrópico. 1: Caatinga; 2: Núcleo de Misiones; 3: Chiquito Bolivarinao; 4: Núcleo Piedmont; 5: Vales inter-andinos do Peru e da Bolívia; 6: Costa Pacífica; 7: Vales inter-andinos da Colômbia; 8: Costa Caribenha; 9: América Central; 10: Antilhas; Ce: Cerrado; LI: Llanos; Ru: Rupununi; Ch: Chaco. Fonte: Pennington *et al.*, 2000.

A área atual do núcleo de *Misiones* compreende a vegetação ao longo do vale dos rios Paraná e Paraguai. Está inserida em ‘Floresta Atlântica Semi-Decídua’ e ‘Decídua’ quando se usa a definição ampla de ‘Floresta Atlântica’, que engloba além das florestas ombrófilas, as decíduas e semi-decíduas (Carnaval & Moritz, 2008;

Werneck, 2011). O núcleo de *Misiones* é considerado uma região que permaneceu relativamente estável climaticamente desde o Último Máximo Glacial (*Last Glacial Maximum* – LGM), há cerca de 21 mil anos, com base em simulações do paleoclima (Carnaval et al., 2009; Werneck et al., 2011a), e por estudos de perfil palinológico (Carnaval et al., 2009), sendo que a região compreendida pela Serra da Bodoquena, no Mato Grosso do Sul e no Pontal do Paranapamena, no oeste de São Paulo, são considerados refúgios naturais, ou seja, áreas onde a vegetação pode permanecer bastante estável durante as oscilações climáticas (Werneck et al., 2011a).

Diversas espécies tem a distribuição restrita ao núcleo de *Misiones*, como o cacto *Cereus hildmaniannus* (Taylor & Zappi, 2004; Bruxel & Jasper, 2005), a *Araucaria angustifolia* (Behling, 2002; Behling, Pillar & Bauermann, 2004), e a mosca *Drosophila antonietae* (Manfrin & Sene, 2006), o que faz deste ambiente uma importante área de endemismo (Amorim & Pires, 1996).

Estas alterações climáticas e, conseqüentemente, paisagísticas modificaram a distribuição geográfica das espécies vegetais neotropicais, influenciando suas estruturas populacionais e certamente de espécies animais com associação ecológica obrigatória, como as espécies cactófilas de *Drosophila* do cluster *Drosophila buzzatii*.

1.2. O cluster *Drosophila buzzatii* e a espécie *Drosophila antonietae*.

O cluster de espécies *Drosophila buzzatii* é naturalmente endêmico da América do Sul e pertencente ao grupo *Drosophila repleta*, gênero *Drosophila*. Em ambientes naturais as espécies do cluster *D. buzzatii* utilizam tecidos de cactos em decomposição como recurso alimentar para suas larvas (Starmer, Lachance, Phaff, & Heed, 1990; Manfrin & Sene, 2006). Como as cactáceas têm baixa tolerância a ambientes úmidos, a distribuição de suas populações é vulnerável às alterações climáticas. Desta forma, é esperado que a distribuição das espécies de *Drosophila* também seja influenciada pelos mesmos fatores.

A base de distribuição geográfica de *D. antonietae* está no núcleo de *Misiones* e em restingas do litoral dos estados do Rio Grande do Sul e Santa Catarina. Utiliza preferencialmente o cacto colunar *Cereus hildmaniannus* K. Schum como hospedeiro,

podendo utilizar, contudo, podendo utilizar, contudo, espécies palmadas do gênero *Opuntia* em regiões de dunas nas costas litorâneas (Ruiz, Cassian, Kuhn, Alves, & Sene, 2000; Manfrin & Sene, 2006). *D. antonietae* ocorre em maior parte alopatricamente a *D. serido*, mas populações em contato secundário com introgressão foram observadas através de análise cromossômica, enzimática e morfológica do edeago no litoral dos estados do Rio Grande do Sul e Santa Catarina (Cansian, Ruiz, & Sene, 1996; Ruiz et al., 2000; Kokudai, Sene & Manfrin, 2011).

Alguns estudos sugerem homogeneidade genética e morfológica entre as populações de *D. antonietae*. A baixa variabilidade desta espécie foi sugerida com base em dados preliminares com o gene mitocondrial COI (Manfrin, De Brito & Sene, 2001), e nucleares a partir de satélites (Franco, Sene & Manfrin., 2010) e do gene *period* (Franco, Silva-Bernardi, Sene, Hasson & Manfrin, 2010a), aloenzimas (Mateus & Sene 2007), e morfologia do edeago (Monteiro & Sene, 1995; Franco, Prado, Sene, Costa & Manfrin, 2006; Mateus, Moura, Manfrin, Monteiro & Sene, 2013). Foi encontrada certa estruturação entre populações mais ao norte da distribuição em relação às do sul (Monteiro & Sene, 1995; De Brito, Manfrin, & Sene, 2002; Mateus & Sene, 2003). A hipótese para esta estruturação é o fluxo gênico restrito pela distância, sendo este facilitado pela ocorrência do cacto hospedeiro ao longo das bacias dos rios (Mateus & Sene, 2007). A partir de dados aloenzimáticos, Mateus & Sene (2007) sugerem que as populações de *D. antonietae* sofram efeito moderado do fluxo gênico sobreposto ao efeito de deriva genética, resultando em um gradiente genético.

O fluxo gênico extensivo pela base dos rios Paraná-Paraguai, juntamente ao isolamento recente entre as populações (Manfrin & Sene 2006; Machado, Mateus, Sene & Manfrin, 2010; Mateus et al., 2013) ao gargalo genético recente causado pela contração da vegetação xerofítica depois do último período glacial (Ab'saber, 1977; Pennington et al., 2000), a manutenção do polimorfismo ancestral, e a seleção natural (Mateus & Sene, 2007; Machado et al., 2010) seriam as causas prováveis da baixa variabilidade genética e homogeneidade entre as populações, especialmente quando comparada às outras espécies do grupo.

Análises existentes utilizando informações de sequências do gene mitocondrial COI sugerem que *D. antonietae*, depois de passar por um período de equilíbrio demográfico, passou por um período de expansão populacional entre 100 a 15 mil anos atrás, sendo que o período que compreende entre 75 e 45 mil anos atrás coincide com o

período glacial Wisconsin. Após isso, há uma tendência ao declínio populacional nos últimos 15 mil anos, embora com pouco apoio estatístico (Franco & Manfrin, 2013).

1.3. Filogeografia e Teoria da Coalescência

A história evolutiva e demográfica de uma espécie pode ser acessada através de métodos filogeográficos. Filogeografia é um termo criado por Avise et al. (1987) que integra a filogenia - que visa entender a relação evolutiva entre as espécies - e a genética de populações, para compreender o curso evolutivo das populações pertencentes a uma mesma espécie. Quando obtido dados das linhagens gênicas, a análise filogeográfica colabora no entendimento de como os eventos históricos influenciariam na distribuição da diversidade genética das espécies, na elaboração de hipóteses biogeográficas, na evolução de unidades populacionais isoladas reprodutivamente e em processos sobre a origem, distribuição e manutenção da biodiversidade.

A estrutura da genealogia atual dos genes é influenciada pelos eventos demográficos históricos, sendo possível inferir sobre mudanças passadas no ambiente físico e biótico a partir dos dados genéticos atuais (Beheregaray, 2008). Estas inferências são possíveis a partir da compreensão da teoria da coalescência (Kingman, 1982), que prediz que todos os alelos atuais tiveram origem a partir de um único ancestral no passado. Esta teoria descreve como os processos genéticos populacionais determinam a forma da genealogia das sequências gênicas amostradas (Drummond, Rambaut, Shapiro, & Pybus, 2005). Dessa maneira, podem ser obtidos diversos cenários e estimar taxas de recombinação (Griffiths & Marjoram, 1996), divisão populacional (Beerli & Felsenstein, 2001) e variações no tamanho populacional e o tempo decorrido entre os eventos (Drummond et al., 2005).

Diferentemente da filogenia, onde a divergência das espécies é representada por ramos que bifurcam, em filogeografia se infere a evolução intraespecífica de genes que resultam em genealogias multifurcadas, pois haplótipos descendentes coexistem com seus ancestrais, há homoplasia e eventos de recombinação que produzem árvores reticuladas e, por isso, a abordagem usando redes e a teoria da coalescência é mais apropriada para análises populacionais. Eles dependem de fatores demográficos, como o

tamanho populacional e a estrutura geográfica, uma vez que, por definição, mutações neutras não afetam o número de descendentes ou o padrão de migração, por isso não alteram a genealogia do gene (Posada & Crandall, 2001).

Existem diversos métodos para a inferência de eventos demográficos. Os mais utilizados em filogeografia são os baseados em probabilidade e os baseados em parcimônia. O método probabilístico mais robusto atualmente é a estatística Bayesiana, através da simulação de Monte Carlo via Cadeia de Markov (MCMC), que traça a genealogia de todos os alelos amostrados até um ancestral comum mais recente (MRCA - *Most Recent Common Ancestor*), tendo como pressuposto o tamanho populacional constante e acasalamentos aleatórios (Templeton, 2006). É frequentemente usada para estimar a distribuição conjunta *a posteriori* (*Joint posterior distribution*) de probabilidades, que consiste na distribuição de todas as combinações possíveis dos valores dos parâmetros utilizados sem ter que explorar todo o espaço de parâmetros. A partir do cálculo de probabilidade de um ponto aleatório inicial no espaço de possibilidades, um novo ponto é escolhido na sua vizinhança e sua probabilidade também é avaliada. Dependendo da relação destas duas probabilidades, o novo estado é aceito ou não. Este processo é repetido por diversas vezes para explorar de maneira eficaz o espaço de parâmetros, com uma concentração maior no espaço com as maiores probabilidades (Excoffier & Heckel, 2006).

Já as inferências de redes de relações entre os haplótipos são baseadas em parcimônia estatística (Templeton, Crandall, & Sing, 1992). Um dos métodos mais utilizados é o Agrupamento de Clados Não-Enraizados (*Nested Clade Phylogeographic Analysis* – NCPA) (Templeton, Routman, & Phillips, 1995). Como a árvore não é enraizada, a teoria da coalescência prediz que os clados das pontas são mais jovens do que os clados do interior aos quais estão conectados. Contrastes entre os clados mais antigos com os clados jovens constituem um contraste temporal independente de relógios moleculares (Templeton, 2004). Em um primeiro momento, esta análise avalia a hipótese nula de que não há associação entre a variação haplotípica e a localidade geográfica, testando as observações por permutação randômica dentro do clado aninhado, e entre as localizações geográficas preservando as frequências dos clados e o tamanho amostral das localidades (Templeton, 1998; Templeton, 2004). São três os fatores mais influentes que podem causar uma associação espacial/temporal da variação haplotípica: fluxo gênico restrito, evento passado de fragmentação e expansão de área

que não são mutuamente exclusivas e cada clado pode exibir um padrão. Como os clados de ponta são mais novos que os clados de interior, usando a polaridade temporal inerente ao desenho da rede, é possível reconstruir os vários fatores que influenciaram a atual distribuição da variação genética como um processo dinâmico ao longo do tempo (Templeton, 1998).

O resultado da inferência por NCPA pode ser testado usando outras análises (Garrick, Dyer, Beheregaray, & Sunnucks, 2008). A fragmentação alopátrica pode ser testada com métodos filogenéticos, como por exemplo, Máxima Verossimilhança e *Neighbour Joining*. Contudo, se a fragmentação é um processo recente pode-se tratar de estruturação das populações, e isso pode ser verificado com Análise de Variância Molecular (AMOVA) (Excoffier, Smouse, & Quattro, 1992). A inferência de expansão de área pode ser validada por *Mismatch Distribution*, com testes de neutralidade ou com métodos baseados no modelo coalescente (Garrick et al., 2008).

1.4. Marcadores Moleculares

O uso de marcadores moleculares com modos de herança e tempos de coalescência distintos pode fornecer informações diferentes, porém muitas vezes complementares, aumentando a acurácia das inferências estatísticas sobre os processos demográficos (Hare, 2001; Templeton, 2004; Avise, 2009). Neste trabalho, analisamos sequências do gene COI mitocondrial e *period* ligado ao cromossomo X.

1.4.1. Citocromo Oxidase I – COI

O gene COI está no genoma da mitocôndria, que é uma organela herdada predominantemente de forma clonal e matrilinear. Consequentemente, o tamanho efetivo populacional dos genes deste genoma é a metade dos genes nucleares. As taxas de mutação também são maiores no DNA mitocondrial do que de genes nucleares. Estas características combinadas fazem com que o tempo de coalescência de genes mitocondriais seja menor do que de genes nucleares, fazendo destas sequências bons marcadores para eventos demográficos (Avise, 1987; Templeton, 2006), como o

estabelecimento de genealogias para o estudo filogeográfico e fonte de dados para inferências genealógicas no nível intraespecífico e microevolutivo (Avise, 2000; Avise, 2009).

A evolução do DNA mitocondrial é basicamente consequência de mutações de ponto. Isso facilita o alinhamento das sequências e compreensão dos dados nas análises populacionais. (Avise, 1987; Templeton, 2006).

Dentre os genes mitocondriais, o COI é o mais utilizado em estudos intra e interespecíficos de metazoários, possui divergência o suficiente entre as espécies (Hebert, Ratnasingham, & De Waard, 2003) e entre as populações, para que estas possam ser identificadas e ter suas relações evolutivas estabelecidas (Avise, 2000; Avise, 2009).

1.4.2. *Period*

O gene *period* está associado com a regulação do ciclo circadiano de dípteros (Konopka & Benzer, 1971; Panda, Hogenesch, & Kay, 2002; Bauzer, Souza, Ward, Kyriacou, & Peixoto, 2002), himenópteros (Toma, Bloch, Moore, & Robinson, 2000), e lepidópteros (Regier et al., 1998). Em *Drosophila*, além do ciclo circadiano, este gene está relacionado ao comportamento espécie-específico de corte dos machos, a fecundidade e ao período e duração da cópula (Kyriacou & Hall, 1980). Devido a estas associações, este gene pode estar envolvido no estabelecimento de isolamento reprodutivo entre espécies (Colot, Hall, & Rosbash, 1988; Coyne, 1992; Wang & Hey, 1996).

O gene *period* está localizado no cromossomo X de *Drosophila*, e por isso tem número efetivo populacional menor que os genes autossômicos. Este gene tem sido utilizado em inferências filogenéticas (Regier et al., 1998; Franco et al., 2010a), sendo informativo na separação de espécies relacionadas e em análises populacionais (Bauzer et al., 2002; Barr, Cui, & McPheron, 2005) com concordância na variação deste gene com a variação de comportamentos sexuais e de feromônios em *Lutzomyia longipalpis* (Diptera) (Bauzer et al., 2002).

O domínio de inibição heterodímero chamado CLOCK/CYCLE (CCID) está localizado próximo à extremidade 5' do gene e é um componente importante na maquinaria do controle do ciclo circadiano (Chang & Reppert, 2003). Por ser composta

de regiões altamente conservadas e regiões polimórficas de evolução rápida, esta região tem sido utilizada para análises comparativas em diversas escalas de tempo (Barr et al., 2005). Dentre os trabalhos que usam esta porção do gene está o de Franco et al. (2010a), que estabelece a filogenia das espécies do cluster *D. buzzatii* e sugere que este gene está sob seleção purificadora nesta linhagem.

1.5. Modelagem de distribuição de espécies

Para inferir a dinâmica espacial de uma fisionomia ambiental ou de uma espécie pode-se fazer uso de Modelos Estatísticos de Distribuição de Espécies (*Statistical Species Distribution Models - SDMs*) (Anderson, 2012). Com os registros de ocorrência de uma espécie conhecidos, pode-se projetar em um mapa a distribuição potencial das mesmas variáveis climáticas e geográficas e, por consequência, das espécies que se utilizam deste ambiente (Anderson, 2012).

Esta metodologia vem sendo largamente utilizada para inferências sobre as potenciais mudanças de distribuição das espécies devido às mudanças climáticas. Desta forma, quando a modelagem ambiental é utilizada para inferir condições paleoambientais, pode, além de ajudar a definir estratégias de amostragem, testar hipóteses sobre a dinâmica populacional obtida a partir metodologias moleculares e permitir uma melhor interpretação destes dados, o que a torna uma importante ferramenta em filogeografia (Nogués-Bravo, 2009; Chan, Brown, & Yode, 2011; Werneck, 2011).

2. OBJETIVOS

Para avaliar as hipóteses existentes e entender a história demográfica de *D. antonietae*, visando buscar os fatores causais que determinaram a distribuição da variabilidade genética de suas populações, este trabalho foi desenvolvido com uma abordagem filogeográfica usando sequências do gene COI mitocondrial e do gene *period* ligado ao cromossomo X, tendo como objetivos:

- Descrever a diversidade e estruturação genéticas das populações de *Drosophila antonietae*;
- Estabelecer hipóteses filogeográficas para *Drosophila antonietae*;
- Inferir eventos demográficos que expliquem a distribuição atual de *Drosophila antonietae*.

3. MATERIAL E MÉTODOS

3.1. Amostragem

Para o gene mitocondrial COI, foram analisados 262 machos de *Drosophila antonietae*, representantes de 36 localidades distintas (Tabela 1). Para as análises com o gene nuclear *period*, foram analisados 120 machos, coletados em 17 localidades (Tabela 2).

Tabela 1 - Amostras utilizadas para análise pelo gene COI. Estado, município, sigla utilizada, número de indivíduos (N), coordenadas geográficas das localidades amostradas e o código de coleta utilizado pelo laboratório para a análise de mtCOI. Continua.

Estado	Município	Sigla	N	Coordenadas Geográficas	Código
Mato Grosso do Sul	Bataguassu	BAT	3	21° 54' 16" S 52° 29' 35" W	R65
	Bodoquena 1	BO1	11	20° 34' 25" S 56° 43' 37" W	R63
	Bodoquena 2	BO2	10	20° 36' 4" S 56° 43' 3" W	R64
	Porto Murtinho	PMU	1	21° 33' 40" S 57° 48' 40" W	R62
São Paulo	Campinas	CAM	2	22° 55' 10" S 47° 04' 22" W	H48
	Cajuru	CAJ	6	21° 30' 00" S 47° 27' 00" W	H90
	Ibiuna	IBI	3	23° 59' 00" S 47° 20' 00" W	J2
	Itatiba	ITA	10	22° 56' 00" S 46° 55' 00" W	J13, J45
	Itirapina	ITI	14	22° 16' 00" S 47° 48' 00" W	J7, J9
	Santa Maria Serra	SMS	10	22° 34' 00" S 48° 12' 00" W	H21, J46
	Serra do Japi	SDJ	9	23° 15' 00" S 47° 48' 00" W	H18, J12
	Serrana	SER	10	21° 15' 00" S 47° 34' 00" W	H84, J38
	Sertãozinho	SET	4	21° 8' 36" S 48° 00' 25" W	H34
	Teodoro Sampaio	TSA	10	22° 30' 54" S 52° 20' 31" W	R67
Paraná	Cianorte	CIA	14	23° 37' 00" S 52° 31' 00" W	D93, J98
	Sapopema	SAP	10	23° 50' 00" S 51° 45' 00" W	J15
	Sengés	SEN	3	24° 07' 00" S 49° 23' 00" W	H99
	Tibagi - Guartelá	TGU	10	24° 32' 00" S 50° 18' 00" W	J16
	Salto Santa Rosa	TSS	10	24° 37' 00" S 50° 33' 00" W	J14, J48

Tabela 2 - Amostras utilizadas para análise pelo gene COI. Estado, município, sigla utilizada, número de indivíduos (N), coordenadas geográficas das localidades amostradas e o código de coleta utilizado pelo laboratório para a análise de mtCOI. Conclusão.

Estado	Município	Sigla	N	Coordenadas Geográficas	Código
Santa Catarina	Flop. Barra da Lagoa	FBL	1	27° 34' 25" S 48° 25' 20" W	R24
	Flop. Armação	FAR	9	27° 45' 10" S 48° 31' 4" W	J55
	Flop. Baía	FBA	11	27° 34' 02" S 48° 31' 07" W	J22
	Flop. Joaquina	FJO	6	27° 37' 37" S 48° 27' 24" W	N8, N90
	Flop. Praia Mole	FPM	1	27° 36' 08" S 48° 26' 08" W	N61
	Gov. Celso Ramos	GCR	1	27° 21' 50" S 48° 33' 43" W	N9
	Garopaba	GAR	4	28° 01' 13" S 48° 37' 24" W	J56
	Laguna	LAG	11	28° 29' 19" S 48° 46' 51" W	J57, N87, R8
Rio Grande do Sul	Arroio Teixeira	ATE	12	29° 38' 00" S 49° 58' 00" W	H41, J23, J24
	Bossoroca	BOS	2	28° 45' 01" S 54° 56' 55" W	BOS
	Capão da Canoa	CCA	2	29° 46' 00" S 50° 17' 00" W	H42
	Jaguarí	JAG	1	29° 30' 04" S 54° 41' 29" W	H46
	Osório	OSO	9	29° 53' 00" S 50° 10' 00" W	J26
	Santiago	SAN	20	29° 12' 00" S 54° 53' 00" W	H47, J27, J28
	Tramandaí	TRA	13	29° 55' 00" S 50° 13' 00" W	H43, J25
	Guaritas	GUA	5	30° 49' 16" S 53° 30' 50" W	H44
	Viamão	VIA	4	30° 06' 00" S 51° 01' 00" W	Via

Flop.: Florianópolis

Tabela 3 - Amostras utilizadas para análise pelo gene *period*. Estado, município, sigla utilizada, número de indivíduos (N), coordenadas geográficas das localidades amostradas e código de coleta utilizada pelo laboratório para a análise do gene nuclear *period*.

Estado	Município	Sigla	N	Coordenadas Geográficas	Código
São Paulo	Itatiba	ITA	4	22°56' 00" S 46° 55' 00" W	J13, J45
	Itirapina	ITI	9	22° 16' 00" S 47° 48' 00" W	J47
	Santa Maria da Serra	SMS	13	22° 34' 00" S 48° 12' 00" W	J46
	Serra do Japí	SDJ	7	23° 14' S 46° 58' W	J12
	Serrana	SER	11	21° 15' 00" S 47° 34' 00" W	J38
	Sertãozinho	SET	4	21° 8' 36" S 48° 00' 25" W	J43
Paraná	Cianorte	CIA	13	23° 37' 00" S 52° 31' 00" W	D93, J98
	Sapopema	SAP	9	23° 50' 00" S 51° 45' 00" W	J15
	Sengés	SEN	3	24° 07' 00" S 49° 23' 00" W	H99
	Tibagi – Guartelá	TGU	8	24°32' 00" S 50° 18' 00" W	J16
	Tibagi - Salto Santa Rosa	TSS	10	24° 37' 00" S 50° 33' 00" W	J14, J48
Santa Catarina	Flop. Joaquina	FJO	1	27° 37' 37" S 48° 27' 24" W	N90
	Flop. Barra da Lagoa	FBL	4	27° 34' 25" S 48° 25' 20" W	R3, R24, N82, N89
	Laguna	LAG	10	28° 29' 19" S 48° 46' 51" W	N87, R8
Rio Grande	Capão da Canoa	CCA	1	29° 44' 44" S 50° 0' 34" W	H42
do Sul	Osório	OSO	2	29° 53' 00" S 50° 10' 00" W	J26
	Santiago	SAN	11	29° 12' 00" S 54° 53' 00" W	H47, J27, J28

Flop.: Florianópolis

A maior parte das amostras de DNA utilizadas neste trabalho foi previamente extraída de indivíduos machos coletados na natureza. Este material estava armazenado em congelador a -20°C no laboratório de Genética Evolutiva do Departamento de Biologia da FFCLRP-USP. As amostras das localidades do estado do Mato Grosso do Sul e no município de Teodoro Sampaio (SP) foram amostradas nos meses de março e maio de 2014 para este estudo. Estas coletas foram realizadas com auxílio de armadilhas fechadas, contendo como isca uma mistura de banana, laranja e levedura *Saccharomyces cerevisiae*. A armadilha era depositada próxima às cactáceas, onde permaneciam por três dias. Os espécimes eram retirados das armadilhas com uso de rede entomológica. A identificação da espécie foi realizada por meio de análise

morfológica da genitália (Silva & Sene, 1991) e os indivíduos foram armazenados congelados a -20°C, até o momento da extração do DNA.

3.2. Obtenção das sequências de DNA

O DNA genômico dos indivíduos coletados foi extraído com auxílio do conjunto de reagentes *DNeasy*® tissue kit (Qiagen), de acordo com as instruções do fabricante. A amplificação dos genes de interesse foi realizada por meio da reação em cadeia da polimerase (PCR), onde cada reação foi composta de: 3µl de DNA extraído, 5µl de tampão de PCR, 1,5µl de MgCl₂ 25mM, 0,3µl *Taq* polimerase 5u/µl. (Promega), 2,5µl de dNTP (2pmol/µl), 2,5µl de cada um dos oligonucleotídeos iniciadores (*forward* e *reverse*) (5pmol/µl) e 8,7µl de água estéril para completar 25µl de volume final. A reação ocorreu em um Termociclador PE Applied Biosystems (DNA Thermal Cycler 480).

A amplificação do gene mitocondrial COI foi realizada por meio dos oligonucleotídeos TY-J-1460 e C1-J-2191 (Simon et al., 1994). As condições da PCR foram uma primeira desnaturação de 1'30" a 94°C, seguida por 25 ciclos de 94°C por 40", 48°C por 40" e 72°C por 2'. Por fim, uma última etapa de elongação por 4' a 72°C.

O par de *primers* para obtenção do fragmento do gene *period* foi desenhado por Franco et al. (2010a) e flanqueia a porção do gene contendo as regiões conservadas C4 e C5, as regiões menos conservadas n3 e n4 e o domínio CCID. As condições de amplificação tiveram como primeira etapa 1' e 30" a 94°C, seguido de 35 ciclos de 94°C por 30", 55°C por 30", e 72°C por 1', finalizando com uma última etapa de elongação por 4' a 72°C.

Os produtos de PCR foram verificados por meio da eletroforese em gel de agarose 1%, corados com Sybr Safe® Invitrogen e visualizados por meio da exposição à luz UV em transiluminador. Os produtos de PCR foram purificados com o conjunto de reagentes *QIAquick PCR Purification Kit* (Qiagen), seguindo as instruções do fabricante para a realização do sequenciamento.

A reação de sequenciamento foi realizada com o conjunto de reagentes *ABI PRISM BigDye™ Terminator Cycle Sequencing Ready Reaction Kit*, baseado no

método de Sanger, Nicklen, & Coulson (1977). A reação de sequenciamento constitui-se de 1µL de *Big Dye*, 1,5µL de tampão do *Big Dye*, 5pmol de oligonucleotídeos iniciadores, 3-5µl de solução de DNA à 14µg/µl. O sequenciamento foi realizado nos dois sentidos da fita de DNA separadamente para se obter maior confiabilidade. As condições de amplificação foram: 1' à 96°C, seguido de 39 ciclos de 15'' à 96°C, 15'' na temperatura de pareamento do *primer*, e 4' à 60°C. As amostras foram sequenciadas em sequenciador automático de DNA modelo *ABI Prism 377*.

Cento e noventa e uma sequências do gene COI foram obtidas previamente por De Brito et al. (2002) e Morales (2005), e três retiradas do GeneBank (números de acessos: HM371180.1, HM371179.1, HM371178.1). Para o gene *period*, foram obtidas oito sequências do GeneBank (número de acesso: FJ267327.1, FJ267326.1, FJ267324.1, FJ267323.1, FJ267322.1, FJ267321.1, FJ267320.1, FJ267319.1).

3.3. Análise dos dados genéticos

3.3.1. *Leitura e Alinhamento das Sequências*

A leitura e a verificação da qualidade das sequências de DNA de ambos os marcadores utilizados foram realizadas com o programa *Chromas v.2.01* Copyright© (1998-2005) Technelysium Pty Ltda. O alinhamento das sequências foi realizado com *ClustalW Multiple Alignment 1.8* (Thompson, Gibson, Plewniak, Jeanmougin, & Higgins, 1997) implementado no programa *BioEdit v. 7.0.9.0* (Hall, 1999). O resultado do alinhamento global foi ajustado de maneira a gerar arquivos de entrada adequados para cada um dos programas de análise subsequentes.

3.3.2. *Estatística Descritiva*

Os parâmetros da estatística descritiva foram obtidos com o programa *DnaSP v.5* (Librado & Rozas, 2009), visando quantificar a variação presente nas sequências de DNA do gene COI e de *period*. Estes parâmetros incluem o número de sítios polimórficos ou segregantes (S), número de haplótipos (h), diversidade haplotípica (Hd), diversidade nucleotídica (π), e o número médio de diferenças nucleotídicas (k).

3.3.3. Modelos de Substituição

O modelo de substituição nucleotídica que melhor explica a evolução das sequências dos genes COI e *period* analisadas foi escolhido utilizando-se o programa *JModelTest2* (Posada, 2008) seguindo o critério de Akaike corrigido (Akaike, 1974). O modelo escolhido foi utilizado nas análises baseados em inferência Bayesiana.

3.3.4. Determinação de linhagens

Árvores genealógicas mais prováveis de ambos os genes foram geradas com abordagem Bayesiana através do pacote *BEAST* (Drummond & Rambaut, 2007), considerando como parâmetros o tamanho populacional constante e relógio molecular relaxado (*uncorrelated lognormal*). Para isto, a taxa de substituição nucleotídica estabelecida para o gene COI foi de 1,4% entre duas linhagens por milhão de anos (Nagaraja & Ranganath, 2004) e o modelo de substituição nucleotídica selecionado foi o GTR. Este modelo é o mais próximo do modelo JC determinado pelo teste *JModelTest* que o pacote *BEAST* disponibiliza. O número de interações na MCMC foi de 200 milhões, amostradas a cada 1000 interações com um período de burn-in de 10% do total de interações. Com o programa *TreeAnnotator 1.6.2* (Drummond & Rambaut, 2007) resumiram-se as árvores usando o padrão de agrupamento *Maximum clade credibility tree-MCC tree*, que encontra a árvore com maior probabilidade *a posteriori* de todos os seus nós. A visualização da árvore de coalescência foi feita pelo programa *FigTree v.1.3.1* (<http://tree.bio.ed.ac.uk/software/figtree/>). Para o gene COI, pode-se estimar as datas dos nós e o Tempo do Ancestral Comum Mais Recente (T_{ACMR}). No entanto, para o gene *period* obteve-se apenas a topologia da árvore, sem datação, devido à ausência de taxa de mutação conhecida.

3.3.5. Análises Filogeográficas

As inferências filogeográficas foram realizadas através da análise de Agrupamento de Clados Não-Enraizados (NCPA) (Templeton et al., 1995). Uma rede haplotípica foi construída por meio do programa *TCS v. 1.21* (Clement, Posada, & Crandall, 2000) e o aninhamento dos clados realizado manualmente de acordo com a metodologia proposta por Templeton, Boerwinkle, & Sing (1987), considerando que

haplótipos diferindo por uma única mudança são conectados àqueles diferindo por duas, e depois por três e assim sucessivamente, até que todos os haplótipos sejam incluídos em uma única rede, ou até que o limite da parcimônia seja atingido (Posada & Crandall, 2001).

Por meio do programa *GeoDis* v.2.5 (Posada, Crandall, & Templeton, 2000) visou-se a identificação de significância da associação entre a ocorrência de haplótipos e sua distribuição geográfica. Os dados com significância estatística foram interpretados usando a chave de inferência filogeográfica disponível em GEODIS HOME PAGE, para inferências dos eventos e processos que influenciaram a história da espécie (Templeton, 1998).

3.3.6. Estruturação Populacional

Para determinar o melhor número de subdivisões populacionais que expliquem os dados e apontar a posição geográfica das populações geneticamente coesas usou-se o programa *Geneland* 4.0.3 (Guillot, Estoup, Mortier, & Cosson, 2005), que é baseado em estatística Bayesiana. A estruturação sugerida pelo *Geneland* foi testada através da Análise de Variância Molecular (AMOVA), por meio do programa *Arlequin* v.3.5 (Excoffier & Lischer, 2010). Para o cálculo da AMOVA, fez-se necessário a determinação das distâncias genéticas entre as populações, calculada a partir de 100 mil interações com 10% de burn-in e 0,05% de significância. Ambas as abordagens avaliam a frequência dos haplótipos, por isso localidades com amostras únicas foram excluídas destas análises.

O teste de Mantel verifica a correlação entre os elementos de uma matriz de distância genética e uma matriz de distância geográfica. Para tanto, foram calculadas as matrizes de distância genética (F_{st}) entre as populações pelo programa *Arlequin* v.3.5 (Excoffier & Lischer, 2010) e de distâncias geográficas dessas populações com o auxílio do programa *Google Earth* 7.1.2.2041. O Teste de Mantel (Mantel, 1967) foi calculado através do programa *Arlequin* v.3.5 (Excoffier & Lischer, 2010), com significância calculada a partir de 1000 permutações.

3.3.7. Testes Demográficos

Foram realizados três testes de neutralidade para os genes COI e *period*, sendo o teste *D* de Tajima (Tajima, 1989) e o teste *F_s* de Fu (Fu, 1997) por meio do programa *DnaSP* v.5 (Librado & Rozas, 2009) e *Mismatch Distribution* através do programa *Arlequin* v.3.5 (Excoffier & Lischer, 2010).

A abordagem *Bayesian Skyline Plot* (BSP), através do pacote *BEAST* (Drummond & Rambaut, 2007), foi utilizada para a obtenção de um gráfico de tamanho efetivo populacional. Para isto, a taxa de substituição estabelecida foi de 1,4% entre duas linhagens por milhão de anos (Nagaraja & Ranganath, 2004) e o modelo de substituição nucleotídica selecionado foi o GTR. Este modelo é o mais próximo do modelo JC determinado pelo teste *JModelTest* que o pacote *BEAST* disponibiliza. O número de interações na MCMC foi de 200 milhões, amostradas a cada 1000 interações com um período de burn-in de 10% do total de interações. A convergência e estabilidade desta análise foram checadas com o programa *Tracer 1.5* (Drummond & Rambaut, 2007) através dos valores do tamanho efetivo da amostra (ESS). Esta análise não foi realizada com o gene *period*, pois não há uma taxa de mutação estabelecida.

3.4. Modelagem de Distribuição da Espécie

Com o objetivo de inferir a dinâmica de distribuição da espécie, foram geradas projeções de potencial distribuição de *D. antonietae* no tempo presente e durante o LGM, há cerca de 21 mil anos. Além do relevo, as 19 camadas bioclimáticas disponíveis foram utilizadas tanto para as análises atuais quanto para as passadas (Hijmans, Cameron, Parra, Jones, & Jarvis, 2005; Thomé et al., 2010). Foram utilizadas 48 localidades ocorrência atual da espécie conhecidas através de publicações ou de coletas do Laboratório de Biologia Evolutiva da FFCL-RP (Tabela 3) e o algoritmo escolheu aleatoriamente 25% dos pontos de amostragem conhecidos para realizar as projeções.

Tabela 3 - Registros de *Drosophila antonietae* utilizados para a modelagem de distribuição das espécies no passado e no presente. Estado, localidade e coordenadas geográficas. Para coletores, foi indicada a referência publicada utilizada. Caso o registro tenha sido realizado pelo laboratório de Genética Evolutiva, foi indicado o coletor e o ano. Continua.

Estado	Localidade	Coordenada em Decimais	
Mato Grosso do Sul	Bataguassu	-21.904490, -52.493250	Registro do laboratório
	Porto Murtinho	-21.5612795, -57.81122	Registro do laboratório
	S. da Bodoquena	-20.708611, -56.793333	Tidon-Sklorz & Sene (2001)
	S. da Bodoquena	-20.573631, -56.726978	Registro do laboratório
	S. da Bodoquena	-20.601364, -56.717618	Registro do laboratório
São Paulo	Altinópolis	-21.08, -47.39	Registro do laboratório
	Cabreúva	-23.307778, -47.131944	Tidon-Sklorz & Sene (2001)
	Cajuru	-21.5, -47.45	Registro do laboratório
	Campinas	-22.919444, -47.072778	Tidon-Sklorz & Sene (2001)
	Ibiúna	-23.983333, -47.333333	Registro do laboratório
	Itatiba	-22.933333, -46.916667	Mateus & Sene (2007)
	Itirapina	-22.266667, -47.8	Mateus & Sene (2007)
	Itú	-23.25, -47.26	Tidon-Sklorz & Sene (2001)
	Santa Maria Serra	-22.566667, -48.2	Tidon-Sklorz & Sene (2001)
	Serra do Japí	-23.233333, -46.966667	Tidon-Sklorz & Sene (2001)
	Serrana	-21.25, -47.566667	Tidon-Sklorz & Sene (2001)
	Sertãozinho	-21.083333, -47.916667	Tidon-Sklorz & Sene (2001)
	Teodoro Sampaio	-22.515150, -52.342200	Registro do laboratório
Paraná	Guarapuava	-25.291611, -51.885778	Lorenci et al. (2010)
	Sapopema	-23.833333, -51.75	Registro do laboratório
	Segredo	-25.774222, -52.115444	Machado et al. (2012)
	Sengés	- 24.116667, -49.383333	Tidon-Sklorz & Sene (2001)
	Tibagi - Guartelá	-24.533333, -50.3	Mateus & Sene (2007)
	Salto Santa Rosa	-24.616667, -50.55	Mateus & Sene (2007)

Tabela 3 - Registros de *Drosophila antonietae* utilizados para a modelagem de distribuição das espécies no passado e no presente. Estado, localidade e coordenadas geográficas. Para coletores, foi indicada a referência publicada utilizada. Caso o registro tenha sido realizado pelo laboratório de Genética Evolutiva, foi indicado o coletor e o ano. Conclusão.

Estado	Localidade	Coordenada em Decimais	
Santa Catarina	Araranguá	-28.925, -49.466944	Registro do laboratório
	Florianópolis	-27.6, -48.433333	Kokudai et al., (2011)
	Garopaba	-28.020278, -48.623333	Registro do laboratório
	Gov. Celso Ramos	-27.363889, -48.561944	Registro do laboratório
	Laguna	-28.488611, -48.780833	Bizzo et al., (2011)
Rio Grande do Sul	Arroio Teixeira	-29.633333, -49.966667	Morales (2001)
	Bossoroca	-28.750278, -54.948611	Poppe et al., (2014)
	Cornélio	-29.5625, -50.0625	Morales (2001)
	Guaritas	-30.821111, -53.513889	Tidon-Sklorz & Sene (2001)
	Jaguari	-29.501111, -54.691389	Tidon-Sklorz & Sene (2001)
	Osório	-29.883333, -50.166667	Morales (2001)
	Porto Alegre	-30.033056, -51.23	Poppe et al., (2014)
	Santa Maria	-29.683333, -53.8	Poppe et al., (2014)
	Santiago	-29.2, -54.883333	Morales (2001)
	S. Luiz Gonsaga	-28.407778, -54.960833	Poppe et al., (2012)
	Tramandaí	-29.916667, -50.216667	Tidon-Sklorz & Sene (2001)
	Viamão	-30.1, -51.016667	Registro do laboratório
Argentina	Puerto Tirol	-27.451389, -58.986667	Tidon-Sklorz & Sene (2001)
	Resiliência	-27.381389, -59.086111	Tidon-Sklorz & Sene (2001)

As projeções para o presente foram calculadas utilizando mapas com resolução espacial de 30 arco-segundos, em que cada pixel do mapa corresponde a um quadrado de um quilômetro de lado, retirados do banco de dados *WorldClim v.1.4* (Hijmans et al., 2005).

As projeções de distribuição para o LGM foram realizadas por meio dos modelos fornecidos pela “Comunidade de Sistemas de Modelos Climáticos” (*Community Climate System Model* - CCSM) e pelo modelo gerado pelo grupo “Modelos para Pesquisa Interdisciplinar do Clima” (*Model for Interdisciplinary Research On Climate* – MIROC), ambos disponíveis em:

<http://www.worldclim.org/past>, com resolução de 2,5 arco-minutos, em que cada pixel corresponde a um quadrado de cinco quilômetros de lado. Os dados foram analisados através do programa *Maxent* v.3.3.3a (Phillips, Anderson, & Schapire, 2006) com o algoritmo de Entropia Máxima (*MaximumEntropy*). Tanto o mapa da distribuição presente quanto passada foram editados no programa *DIVA-GIS* v. 7.3.0, disponível em DIVA-GIS HOME PAGE (www.diva-gis.org).

4. RESULTADOS

4.1. Análises das sequências do gene nuclear *period*

Foram analisados 420 pares de bases do gene nuclear *period* de 120 indivíduos machos. Foram encontrados 28 sítios polimórficos e 19 haplótipos. Os valores de diversidade genética da amostragem total e de todas as localidades estão descritos na Tabela 4. O primeiro par de bases das sequências deste estudo correspondem à base 6.066 da sequência de referência do cromossomo X completo de *Drosophila melanogaster* (código NC_004354.3).

O modelo de substituição nucleotídica que melhor explica a evolução das sequências do gene *period* é o HKY+I (AICc = 2751.3569; weight = 0,5555).

Tabela 4 - Índices de diversidade genética de populações de *Drosophila antonietae*, utilizando sequências do gene nuclear *period*.

População	N	h	S	Hd	π	k
CCA	1	1	1			
CIA	13	6	4	0,872 (0,054)	0,00336 (0,00046)	1,41026
FBL	3	2	11	0,667 (0,314)	0,01746 (0,00823)	7,33333
FJO	1	1	1	-	-	-
ITA	4	4	3	1,000 (0,177)	0,00397 (0,00106)	1,66667
ITI	9	3	4	0,417 (0,191)	0,00212 (0,00121)	0,88889
LAG	10	6	5	0,867 (0,085)	0,00392 (0,00100)	1,64444
OSO	2	2	2	1,000 (0,500)	0,00476 (0,00238)	2,00000
SAN	10	4	4	0,733 (0,101)	0,00270 (0,00076)	1,13333
SAP	9	5	9	0,861 (0,087)	0,00661 (0,00228)	2,77778
SDJ	7	4	4	0,810 (0,130)	0,00431 (0,00116)	1,80952
SEN	3	1	0	0,000 (0,000)	0,00000 (0,00000)	-
SER	11	5	6	0,836 (0,080)	0,00433 (0,00101)	1,81818
SET	4	2	1	0,500 (0,265)	0,00119 (0,00063)	0,50000
SMS	13	5	3	0,846 (0,059)	0,00305 (0,00039)	1,28205
TGU	8	6	5	0,893 (0,111)	0,00451 (0,00099)	1,89286
TSS	10	5	5	0,756 (0,130)	0,00376 (0,00097)	1,57778
Total	120	19	28	0,817 (0,022)	0,00439 (0,00055)	1,84454

N = número de indivíduos; h = número de haplótipos; S = número de sítios polimórficos; Hd = diversidade haplotípica (desvio padrão); π = diversidade nucleotídica (desvio-padrão); k = número médio de diferenças nucleotídicas. Os códigos das populações são os mesmos apresentados na Tabela 2.

4.1.1. Distribuição dos Haplótipos

A distribuição e frequência dos 19 haplótipos do gene nuclear *period* pelas 17 localidades amostradas está representada na Figura 2.

Figura 2. Mapa mostrando a distribuição dos haplótipos de *period* entre as localidades amostradas. O tamanho da representação de cada população é proporcional ao número de indivíduos amostrados.

4.1.2. Determinação de linhagens

A topologia da árvore de consenso Bayesiana mostrada na Figura 3, mostra que há uma separação das amostras em duas linhagens com alto suporte estatístico *a posteriori*. As duas linhagens, chamadas de ‘A’ e ‘B’ estão igualmente distribuídas por todas as localidades amostradas. A árvore também aponta sinais de que as populações teriam passado por um gargalo genético seguido de expansão populacional recente, caracterizado por longos ramos entre os grupos e ramos curtos ligando os haplótipos dentro de cada um dos grupos.

Figura 3. Relações entre as linhagens do gene nuclear *period* de *Drosophila antonietae*. Valores correspondem ao suporte estatístico do nó (probabilidade a posteriori) da MCMC.

4.1.3. Análises Filogeográficas

Na Tabela 5 estão descritas as associações geográficas e genéticas significativas encontradas no clado de maior nível da rede, formado pelas sequências do gene nuclear *period*. O haplótipo h5 apresentou o maior valor de *outgroup weight* (0,202). Este haplótipo ocupa uma posição interna na rede (Figura 4), e é também um dos com maior distribuição geográfica e frequência da amostra, características esperadas para haplótipos antigos (Templeton, 1998).

Tabela 5 – Resultado significativo da NCPA para a rede obtida a partir de amostras do gene nuclear *period* de populações de *Drosophila antonietae*, de acordo com a chave de inferências filogeográficas de Templeton (2011).

Clado	Etapas da Chave de Inferências
3-1	1-2-3-4: Fluxo gênico restrito com isolamento por distância.

Figura 4. Rede haplotípica não enraizada formada através de NCPA com sequências do gene *period* de populações de *Drosophila antonietae*. Cada linha representa um passo mutacional e o círculo preto representa um haplótipo não amostrado. O tamanho dos círculos é proporcional ao número de indivíduos portadores de cada haplótipo.

4.1.4. Estruturação Populacional

A análise através do programa *Geneland* usando as sequências do gene *period* mostrou que dois grupos genéticos podem explicar melhor a distribuição da variabilidade genética de *D. antonietae* (Figura 5). Os resultados sugeridos por esta análise serviram de base para a hierarquização da AMOVA.

Figura 5. Mapa gerado pelo *Geneland* a partir de sequências de *period* mostrando as prováveis áreas de ocorrência das populações estruturadas e correspondência sobre mapa do Brasil.

Foi encontrada baixa estruturação entre as populações através da AMOVA global (Tabela 6), e mesmo assim se testou a divisão entre os dois grupos proposto pelo *Geneland*. A diferença entre as populações se mostrou muito pequena, evidenciando a homogeneidade da amostragem e a grande variabilidade genética contida dentro das populações.

Tabela 6 - AMOVA global e com dois grupos, testadas a partir de sequências do gene nuclear *period*.

	Fonte de Variação	GL	% da Variação
AMOVA Global			
$\Phi_{st} = 0,06042$ $p=0,00842$	Entre as populações	14	6,04
	Dentro das populações	103	93,96
AMOVA 2 grupos conforme sugerido por Geneland			
$\Phi_{ct} = 0,01509$ $p=0,08139$	Entre os grupos	1	1,51
$\Phi_{st} = 0,06707$ $p=0,01099$	Entre as populações	13	5,20
$\Phi_{sc} = 0,05278$ $p=0,01733$	Dentro das populações	103	93,29

GL: Graus de liberdade.

O teste de Mantel para o gene *period* não indicou correlação significativa entre as diferenças genéticas e a distância geográfica ($r^2 = -0.153$; $p > 0,05$).

4.1.5. Testes Demográficos

Todos os testes de neutralidade indicam expansão populacional (Tabela 7). O valor do teste D de Tajima é de -1,92607 e o teste Fs de Fu é de -27,64491 para a totalidade das amostras de *Drosophila antonietae*. A curva gerada pelo teste de *Mismatch distribution* é unimodal (Figura 6).

Tabela 7 - Testes de Neutralidade D de Tajima e Fs de Fu para a população total gerados por AMOVA utilizando sequências do gene *period*.

	D de Tajima	Fs de Fu	<i>Mismatch distribution</i>
População Total			Unimodal
	-1,92607	-27,64491	SSD = 0,003
	p<0,01	p<0,01	pSSD = 0,146
			$\theta_0 = 0,01$
			$\theta_1 = 71718,184$

SSD: teste do desvio da soma dos quadrados (*Sum of squared deviation*); θ_0 : tamanho populacional antes da expansão; θ_1 : tamanho populacional após a expansão.

Figura 6. Representação gráfica das diferenças par a par entre as sequências (*Mismatch distribution*) do gene nuclear *period* dos indivíduos de *Drosophila antonietae*.

4.2. Análises das sequências do gene mitocondrial COI

Foram analisados 563 pares de bases de 262 sequências de DNA do gene mitocondrial COI. Foram encontrados 89 haplótipos ($h=89$) e 140 sítios polimórficos. Os índices de diversidade genética calculados para cada uma das populações estão descritos na Tabela 8.

O primeiro par de bases das sequências corresponde à base 1.594 do genoma mitocondrial de referência de *Drosophila yacuba* (Número de acesso NCBI GenBank NC_001322).

O modelo de substituição que melhor explica o padrão de evolução do gene COI é o JC ($AICc = 19438.5747$; $weight = 1$).

Tabela 8 - Índices de diversidade genética calculados a partir de sequências obtidas do gene COI das populações analisadas de *Drosophila antonietae*.

População	N	h	S	Hd	π	k
ATE	12	7	8	0,833 (0,100)	0,00261 (0,00062)	1,46970
BAT	3	3	39	1,000 (0,272)	0,04618 (0,02122)	26
BO1	11	3	34	0,655 (0,111)	0,01996 (0,00893)	11,23636
BO2	10	3	41	0,378 (0,181)	0,01543 (0,01018)	8,68889
BOS	2	2	2	1,000 (0,500)	0,00355 (0,00178)	2
CAJ	6	1	0	0,000 (0,000)	0,00000 (0,00000)	-
CAM	2	2	5	1,000 (0,500)	0,00888 (0,00444)	5
CCA	2	2	2	1,000 (0,500)	0,00355 (0,00178)	2
CIA	10	6	10	0,604 (0,150)	0,00318 (0,00131)	1,79121
FAR	9	7	25	0,917 (0,092)	0,01401 (0,00280)	7,88889
FBA	11	8	13	0,927 (0,066)	0,00827 (0,00119)	4,65455
FBL	1	1	0	-	-	-
FJO	6	5	15	0,933 (0,122)	0,00959 (0,00258)	5,40000
FPM	1	1	0	-	-	-
GAR	4	3	10	0,833 (0,222)	0,00947 (0,00281)	5,33333
GCR	1	1	0	-	-	-
GUA	5	h	10	1,000 (0,126)	0,00782 (0,00178)	4,40000
IBI	3	2	1	0,667 (0,314)	0,00118 (0,00056)	0,66667
ITA	10	3	6	0,511 (0,164)	0,00351 (0,00140)	1,97778
ITI	14	3	2	0,275 (0,148)	0,00051 (0,00029)	0,28571
JAG	1	1	0	-	-	-
LAG	11	7	10	0,818 (0,119)	0,00562 (0,00176)	3,16364
OSO	9	4	9	0,750 (0,112)	0,00720 (0,00132)	4,05556
PMU	1	1	0	-	-	-
SAN	20	14	18	0,916 (0,055)	0,00949 (0,00128)	5,34211
SAP	10	3	7	0,511 (0,164)	0,00493 (0,00149)	2,77778
SDJ	9	2	1	0,222 (0,166)	0,000395 (0,00030)	0,2222
SEN	3	2	5	0,667 (0,314)	0,00592 (0,00279)	3,33333
SER	10	2	1	0,200 (0,154)	0,00036 (0,00027)	0,20000
SET	4	1	0	0,000 (0,000)	0,00000 (0,00000)	-
SMS	10	1	0	0,000 (0,000)	0,00000 (0,00000)	-
TGU	10	3	5	0,600 (0,131)	0,00237 (0,00115)	1,33333
TRA	13	8	8	0,885 (0,070)	0,00342 (0,00082)	1,92308
TSA	10	5	91	0,667 (0,163)	0,03951 (0,01996)	22,24444
TSS	10	5	5	0,756 (0,130)	0,00280 (0,00075)	1,57778
VIA	4	2	2	0,500 (0,265)	0,00178 (0,00094)	1
Total	262	89	140	0,864 (0,018)	0,01510 (0,00172)	8,50006

N = número de indivíduos; h= número de haplótipos; S = número de sítios polimórficos; Hd = diversidade haplotípica (desvio padrão); π = diversidade nucleotídica (desvio-padrão); k = número médio de diferenças nucleotídicas. Os códigos das populações são os mesmos apresentados na Tabela 1.

4.2.1. Distribuição dos haplótipos

A Figura 7 mostra a distribuição dos 89 haplótipos do gene mitocondrial COI pelas 31 localidades amostradas. O tamanho da representação de cada população é proporcional ao número de indivíduos amostrados.

Figura 7. Mapa mostrando a distribuição dos haplótipos do gene COI entre as localidades amostradas.

4.2.2. Determinação de linhagens

A Figura 8 mostra a árvore de consenso Bayesiana, que indica as linhagens existentes para as sequências do gene COI. Os valores dos nós *a posteriori* assim como o intervalo de confiança estão demonstrados na figura. A árvore também sugere sinais de que as populações teriam passado por um gargalo genético seguido de expansão populacional recente, caracterizado por longos ramos entre os grupos e ramos curtos ligando os haplótipos dentro de cada um dos grupos.

Figura 8. Relações entre as linhagens definidas com base na variação haplotípica do gene mitocondrial COI, dos 262 indivíduos de 36 localidades de *Drosophila antonietae*, usando análise de coalescência Bayesiana (*Maximum Clade Credibility Tree*) e mostrando o Tempo do ancestral comum mais recente (Tmrca). Os valores superiores correspondem ao suporte estatístico do nó (probabilidade *a posteriori*) da MCMC e os números sublinhados abaixo correspondem ao intervalo de confiança de 95%. Valores não exibidos não alcançaram suporte estatístico.

Na primeira ramificação, ocorre a divergência da linhagem ‘H’ a mais de 100 mil anos (probabilidade 1 *a posteriori*), que é formada pela amostra proveniente do município de Teodoro Sampaio, SP (R67-2). A segunda ramificação teria ocorrido há cerca de 70 mil anos (0,579 *a posteriori*) onde a linhagem “A” diverge das demais. A linhagem “A” é composta por indivíduos de duas localidades de Bodoquena (BO1 e BO2), Bataguassú (BAT) e Teodoro Sampaio (TSA). Geograficamente, este grupo está localizado a oeste do rio Paraná, considerando seu curso antigo, onde, diferentemente dos dias atuais, TSA estaria à esquerda rio. Há 30 mil anos a linhagem ‘G’ teria divergido (0,971 *a posteriori*) formada pela amostra proveniente de Florianópolis (FBL) (R24-21), enquanto que a linhagem ‘F’ (0,999 *a posteriori*), formada pela amostra proveniente de Garopaba, (GAR) (J56-6) teria divergido há 25 mil anos. A linhagem ‘E’ divergiu do restante a menos de 20 mil anos (0,927 *a posteriori*) e está presente apenas no litoral de Santa Catarina. A linhagem ‘B’ diverge do restante há cerca de 10 mil anos (0,586 *a posteriori*) e é encontrado predominantemente nos estados do

Paraná e do Mato Grosso do Sul, com uma ocorrência em São Paulo. Concomitantemente, as linhagens ‘C’ e ‘D’ se separam, porém, com baixa probabilidade *a posteriori* (0,193). O grupo ‘C’, que abriga a maior parte da amostragem, tem ampla distribuição em todos os estados, com exceção de Santa Catarina. Enquanto que o grupo ‘D’ é predominante nos estados do Rio Grande do Sul e Santa Catarina, sendo encontrado também em Cianorte (CIA) e Teodoro Sampaio (TSA).

4.2.3. Análises Filogeográficas

A análise por clados aninhados do gene COI gerou duas redes distintas, uma delas agrupando a maioria dos haplótipos (Figura 9) e uma menor, agrupando poucos haplótipos que são distribuídos no estado do Mato Grosso do Sul (Figura 11). Para a rede maior, foram descritos oito associações genético-geográficas significativas, porém foi possível inferir eventos demográficos apenas para cinco delas, sendo as restantes inconclusivas e com amostragem geográfica inadequada (Tabela 9).

Tabela 9 - Resultados significativos da NCPA para a rede obtida a partir de amostras do gene mitocondrial COI de populações de *Drosophila antonietae*, de acordo com a chave de inferências filogeográficas de Templeton (2004).

Clado	Etapas da Chave de Inferências
1-7	1-2-3-4: Fluxo gênico restrito com isolamento por distância.
1-33	1-2-3-4: Fluxo gênico restrito com isolamento por distância.
2-10	1-2-3-4: Fluxo gênico restrito com isolamento por distância.
3-2	1-2-3-5-6-7: Fluxo gênico restrito, mas com alguma dispersão à longa distância.
4-1	1-2-11-12: Expansão populacional contínua.
4-3	1-2: Resultado inconclusivo.
5-1	1-19-20: Amostragem geográfica inadequada.
6-1	1-2: Resultado inconclusivo.

O haplótipo 3 (h3) apresentou o maior *outgroup weight* (0.11), o que indica que este tem a maior probabilidade de ser o haplótipo ancestral da genealogia. Este haplótipo tem ampla distribuição geográfica, ocorrendo nos estados de São Paulo e Paraná, alta frequência e ocupa uma posição interna na rede, o que é esperado para haplótipos antigos (Templeton,

1998). Este haplótipo está inserido dentro do clado 4-1, onde se acredita ter ocorrido expansão populacional contínua.

O clado 3.2 é formado pelos haplótipos da porção mais a oeste da amostragem, tanto no norte da distribuição, no estado do Mato Grosso do Sul, quanto no sul, no estado do Rio Grande do Sul. O resultado obtido por NCPA sugere dispersão à longa distância, porém, pode ser explicado pela ausência de amostragem de *Drosophila antonietae* na região intermediária, por onde pode haver fluxo gênico. Isto pode ter influenciado na interpretação de amostragem inadequada do clado 5-1.

Há correspondências entre os grupos formados pela NCPA e pela árvore Bayesiana. Os haplótipos da linhagem C da análise Bayesiana são correspondentes aos aninhados nos clados 3.1, 3.3 e 2.3 (pertencente ao clado 3.2). Todos os haplótipos representantes agrupados nos clados 3.6 e 4.4 estão também agrupados na linhagem E da árvore de consenso Bayesiana. Provavelmente foram estes os haplótipos a divergirem primeiro e hoje ocorrem exclusivamente no litoral de Santa Catarina. A linhagem D da árvore Bayesiana corresponde aos clados 3.5 e o 3.4 da NCPA. O agrupamento 3.2 corresponde com poucas exceções à linhagem B, sendo que h81 corresponde à linhagem G e o h67 à linhagem F.

A formação dos clados sugere cenários de possíveis rotas de expansão de área dos grupos de haplótipos. As figuras 10A a 10H representam estas inferências.

Figura 10A. Clado 1-7: Fluxo gênico restrito com isolamento por distância. Este clado está contido nos clados de maior hierarquia 3-2, 4-1 e 5-1.

Figura 10B. Clado 3-2: Fluxo gênico restrito, mas com alguma dispersão à longa distância. O clado 3-2 está contido nos clados de maior hierarquia 4.1 e 5.1.

Figura 10C. Clado 4-1: Expansão populacional contínua. Este clado está contido no clado de maior hierarquia 5-1. A seta representa o provável sentido da migração. Este sentido foi inferido utilizando a premissa de que clados de interior são supostamente mais antigos do que clados alocados nas extremidades da genealogia.

Figura 10E. Clado 1-33: Fluxo gênico restrito com isolamento por distância.

Figura 10F. Clado 2-10: Fluxo gênico restrito com isolamento por distância. Este clado está contido nos clados 4-3 e 5-2, porém estes agrupamentos de maior hierarquia não contribuem com a adição de mais nenhuma localidade e não apresentaram resultados conclusivos pela análise NCPA.

Os haplótipos que formaram um clado isolado, demonstrado na Figura 11, têm o haplótipo 84 com o maior *outgroup weight* (0,9). Nesta rede se agruparam apenas as amostras do estado do Mato Grosso do Sul juntamente com a do município de Teodoro Sampaio (Figura 12), o mesmo grupo de amostras que formou a linhagem ‘A’ na árvore Bayesiana (Figura 9). O haplótipo h86 corresponde á amostra R67-2 que divergiu na árvore formando a linhagem ‘H’.

Figura 11. Rede haplotípica formada a partir de NCPA com sequências do gene COI de populações de *Drosophila antonietae* que não se uniram à rede principal. Cada linha representa um passo mutacional. Cada círculo preto representa um haplótipo não amostrado. O tamanho dos círculos é proporcional ao número de indivíduos portadores do haplótipo.

Figura 12. Rede com localidades que formam o 2° clado isolado da rede principal, onde a NCPA não encontrou significância estatística para realizar inferências.

4.2.4. Estruturação Populacional

A análise através do programa *Geneland* usando as sequências do gene COI mostrou que cinco grupos genéticos melhor explicam a distribuição da variabilidade genética de *D. antonietae* (Figura 13), sendo: 1) amostras do oeste do Mato Grosso do Sul; 2) amostras de ambas as margem do rio Paraná; 3) amostras dos estados de São Paulo e Paraná; 4) amostras do interior do Rio Grande do Sul e de Florianópolis e; 5) demais amostras do litoral do Rio Grande do Sul e de Santa Catarina.

Visto os agrupamentos obtidos pela NCPA e o mapa de distribuição dos haplótipos (Figura 7), optou-se por agrupar as amostras em apenas 3 populações e não houve prejuízo no valor de Φ_{st} nos testes AMOVA realizados para a confirmação (Tabela 5). Com esta nova conformação, os agrupamentos obtidos foram: 1) as localidades dos estados de São Paulo e Paraná; 2) as localidades do estado do Mato Grosso do Sul juntamente a Teodoro Sampaio (SP) e; 3) as localidades dos estados do Rio Grande do Sul e Santa Catarina (Figura 14).

Figura 13. Mapa gerado pelo *Geneland* a partir de sequências de COI mostrando as prováveis áreas de ocorrência das populações estruturadas estimadas e mapa do Brasil correspondente.

Figura 14. Mapa gerado pelo *Geneland* a partir de sequências de COI mostrando as prováveis áreas de ocorrência de populações estruturadas quando se estima três grandes populações e mapa do Brasil correspondente.

A Análise de Variância Molecular (AMOVA) global (Tabela 10) encontrou estruturação significativa entre as localidades amostradas com 52% da variação encontrada entre as populações e 47,86% da variação dentro das populações. Os valores de F_{st} significativos dos grupos hipotetizados mostram que há diferenciação entre os grupos. A

partir da AMOVA global, testaram-se alguns agrupamentos, e, dentre eles, os sugeridos pelo *Geneland* forneceram os maiores valores de diferenciações entre os grupos.

A porcentagem da variação entre os grupos é menor quando Teodoro Sampaio (TSA) é agrupado com as demais amostras do lado leste do Rio Paraná do que quando agrupada às amostras do Mato Grosso do Sul, a oeste do Rio Paraná, algo inesperado devido ao seu atual curso estar entre estas duas amostras, sugerindo que as mudanças em seu rumo exercem efeito sobre a distribuição genética de *D. antonietae*.

Tabela 10 - AMOVA global e AMOVA hierarquizada obtida a partir de sequências do gene mitocondrial COI.

	Fonte de Variação	GL	% da Variação
AMOVA Global			
$\Phi_{st} = 0,52145$ $p=0,000$	Entre as populações	30	52,14
	Dentro das populações	226	47,86
AMOVA 3 grupos, TSA agrupada às amostras às demais amostras de SP			
$\Phi_{ct} = 0,30453$ $p=0,020$	Entre os grupos	2	30,45
$\Phi_{st} = 0,57934$ $p=0,000$	Entre as populações	28	27,48
$\Phi_{sc} = 0,39515$ $p=0,000$	Dentro das populações	226	42,07
AMOVA 3 grupos, TSA agrupada às amostras à oeste do rio Paraná			
$\Phi_{ct} = 0,44187$ $p=0,000$	Entre os grupos	2	44,19
$\Phi_{st} = 0,59938$ $p=0,000$	Entre as populações	28	15,75
$\Phi_{sc} = 0,28221$ $p=0,000$	Dentro das populações	226	40,06
AMOVA 5 grupos conforme sugerido por Geneland			
$\Phi_{ct} = 0,55088$ $p=0,000$	Entre os grupos	4	55,09
$\Phi_{st} = 0,59214$ $p=0,000$	Entre as populações	26	4,13
$\Phi_{sc} = 0,09188$ $p=0,01317$	Dentro das populações	226	40,79

GL: Graus de liberdade. TSA: Teodoro Sampaio.

O teste de Mantel realizado com todas as amostras indicou correlação significativa entre as diferenças genéticas e a distância geográfica para o gene COI ($r^2 = 0.438$; $p < 0,01$).

4.2.5. Testes Demográficos

Quando a população é analisada como um todo, o valor do teste D de Tajima é de -1,914 e o teste Fs de Fu é de -24,274, ambos esperados em caso de expansão populacional. A

curva gerada pelo teste *Mismatch distribution* é multimodal, o que sugere equilíbrio demográfico.

Quando se subdivide as populações em três grupos seguindo a hierarquização que melhor separa as populações por AMOVA, o grupo de populações formado pelas localidades a oeste do Rio Paraná (Mato Grosso do Sul e Teodoro Sampaio), apresenta valor do D de Tajima de -0,146 e Fs de Fu de -17,388, concordante com uma proposta de expansão populacional. Contudo, a curva gerada pelo teste de *Mismatch distribution* é multimodal, condizente uma proposta de equilíbrio demográfico. O restante das localidades a leste do Rio Paraná (São Paulo e Paraná), geram os valores de D de Tajima de -2,094 e Fs de Fu de -28,824, indicativo de expansão, com uma curva do *Mismatch distribution* bimodal, indicando estabilidade populacional. Os valores de D de Tajima e Fs de Fu para as populações de Santa Catarina e Rio Grande do Sul são, respectivamente, -1,911 e -25,486, indicando expansão populacional, porém discordante da curva de *Mismatch distribution* que indica equilíbrio demográfico.

Desta forma, os testes de neutralidade indicam expansão populacional para a população total e para as subdivisões estabelecidas, porém não concordante com a interpretação sugerida de equilíbrio demográfico pela *Mismatch distribution*. Os valores estão descritos na Tabela 11 e as representações da *Mismatch distribution* na Figura 15.

Tabela 11 - Testes de Neutralidade D de Tajima e Fs de Fu para a população total e para as subdivisões testadas utilizando sequências do gene COI. Continua.

	D de Tajima	Fs de Fu	<i>Mismatch distribution</i>
Oeste do Rio Paraná			Multimodal
	-0,146	-17,388	SSD = 0,098
	p>0,05	p<0,01	pSSD = 0,010
			$\theta_0 = 0,000$
			$\theta_1 = 31.020$
Leste do Rio Paraná			Bimodal
	-2,094	-28,824	SSD = 0,250
	p<0,01	p<0,01	pSSD = 0,000
			$\theta_0 = 0,000$
			$\theta_1 = 1974,400$
SC e RS			Multimodal
	-1,911	-25,486	SSD = 0,007
	p<0,01	p<0,01	pSSD = 0,819
			$\theta_0 = 0,000$
			$\theta_1 = 7,155$
População total			Multimodal
	-1,914	-24,274	SSD = 0,036
	p<0,01	p<0,01	pSSD = 0,019
			$\theta_0 = 0,000$
			$\theta_1 = 12,851$

SSD: teste do desvio da soma dos quadrados (*Sum of squared deviation*); θ_0 : tamanho populacional antes da expansão; θ_1 : tamanho populacional após a expansão.

Figura 15. Representação gráfica das diferenças par a par entre as sequências (*Mismatch distribution*) do gene mitocondrial COI dos indivíduos de *Drosophila antonietae*. Linha contínua intensa representa a diferença par a par observada, enquanto que linha contínua tênue representa a diferença par a par esperada em eventos de expansão populacional. População total e subdivisões populacionais seguindo a tendência de estruturação por AMOVA.

A análise demográfica através da BSA indicou equilíbrio no número efetivo populacional, concordando com o a *Mismatch distribution*. O tempo calculado do ancestral comum mais recente (T_{MRCA}), obtido por meio da análise *Bayesian Skyline Plot* (Figura 16), com as amostras do gene COI, é de 14 mil anos atrás. A partir deste período não são observadas oscilações no tamanho efetivo populacional, apenas uma tendência de expansão, que se acentua nos últimos dois milênios.

Figura 16. Reconstrução da história demográfica, gerada por BSP, de *Drosophila antonietae*. A curva na cor preta representa a variação na média do tamanho efetivo populacional das fêmeas (Nef) em relação ao tempo em milhares de anos antes do presente. A linha tracejada vertical indica o tempo estimado para o ancestral comum mais recente da amostra (T_{ACMR}). As curvas em cinza representam o intervalo de confiança (95%). O tamanho populacional é medido em unidades de $\log Ne \cdot \tau$.

4.3. Modelagem de distribuição da espécie

Todos os modelos gerados, tanto o correspondente à provável distribuição atual da espécie quanto os dois modelos para o Último Máximo Glacial apresentaram valores de AUC (*Area Under the Curve*) acima de 0,7, correspondendo a modelos de alta precisão e alta discriminação (Newbold, Gilbert, Zalat, El-Gabbas, & Reader, 2009).

4.3.1. Projeção atual

A projeção de distribuição atual da espécie *D. antonietae* está representada na Figura 17. O AUC do modelo foi de $0,985 \pm 0,02$. O modelo sugere que as variáveis bioclimáticas que mais contribuem com a elaboração do modelo preditivo são: temperatura sazonal (Bio 4) com 34,1% de importância; precipitação do trimestre mais quente (Bio 18) com 19%; e precipitação do mês mais seco (Bio 14) com 10,7%. Cada uma das demais camadas contribui com menos de 10% para a estimativa do mapa.

Figura 19. Modelo de probabilidade de presença atual da espécie *Drosophia antonietae*. As cores esverdeadas representam a maior probabilidade de ocorrência da espécie.

4.3.2. *Projeção Miroc*

O modelo de distribuição da espécie *D. antonietae* há 21 mil anos atrás predito pelo modelo Miroc está representado na Figura 18. O AUC do modelo foi de $0,983 \pm 0,02$. O modelo sugere que as variáveis bioclimáticas que mais contribuem com a elaboração do modelo preditivo são: temperatura sazonal (Bio 4) com 39,2% de importância; temperatura média do trimestre mais frio (Bio 11) com 15,2%; e precipitação do mês mais seco (Bio 14) com 12%. Cada uma das demais camadas contribui com menos de 10% para a estimativa do mapa.

4.3.3. *Projeção CCSM*

O modelo CCSM, que também infere sobre a distribuição da espécie há 21 mil anos, está representado na Figura 19. O AUC do modelo foi de $0,987 \pm 0,02$. O modelo sugere que as variáveis bioclimáticas que mais contribuem com a elaboração do modelo preditivo são: temperatura sazonal (Bio 4) com 37,1% de importância; precipitação do mês mais seco (Bio 14) com 14,9%; e precipitação do trimestre mais quente (Bio 18) com 9,2%. Cada uma das demais camadas contribui com porcentagens menores para a estimativa do mapa.

Figura 18. Modelo de probabilidade de presença da espécie *Drosophia antonietae* há 21 mil anos atrás predito pelo modelo *Miroc*. As cores esverdeadas representam as maiores probabilidade de ocorrência da espécie.

Figura 19. Modelo de probabilidade de presença da espécie *Drosophia antonietae* há 21 mil anos atrás predito pelo modelo CCSM. As cores esverdeadas representam as maiores probabilidade de ocorrência da espécie.

5. DISCUSSÃO

A atual distribuição da diversidade genética de *D. antoneitae* tem sido explicada com diversos eventos de dispersão e expansão populacional, provavelmente tendo como origem de dispersão as populações do norte da sua distribuição atual. A estruturação genética indica que a diversidade de *D. antonietae* não é homogênea ao longo de sua distribuição geográfica como observado anteriormente (Monteiro & Sene, 1995; Manfrin et al., 2001; Franco et al., 2006; Mateus & Sene 2007; Franco et al., 2010; Franco et al., 2010a; Mateus et al., 2013), especialmente quando o gene mitocondrial COI foi analisado. Já o gene nuclear *period* demonstrou baixíssima estruturação genética, podendo ser indício de homogeneidade genética ancestral. Devido aos tempos diferenciais de coalescência dos marcadores utilizados, os resultados incongruentes encontrados podem estar indicando que o processo de estruturação das populações é incipiente, decorrente da fragmentação dos núcleos de cactos durante a expansão das florestas úmidas. As similaridades encontradas entre as populações são em boa parte explicadas pela retenção de polimorfismo ancestral (Franco et al., 2010a) e pelo curto tempo de divergência das populações (Manfrin & Sene 2006; Machado, Mateus, Sene, & Manfrin, 2010; Mateus et al., 2013), já que *D. antonietae* se trata de uma espécie relativamente recente, pois o tempo calculado de separação da espécie irmã é de 1,7 milhões de anos (Franco & Manfrin, 2013).

A homogeneidade da diversidade genética do gene *period* ao longo da distribuição geográfica de *D. antonietae* foi confirmada por todas as análises realizadas, o que reflete na pouca estruturação das populações, ocorrendo fluxo gênico restrito com isolamento por distância entre todos os haplótipos amostrados para este gene. Foram encontradas duas linhagens de haplótipos com divergência antiga, entretanto, ambas ocorrem sobrepostas e com ampla distribuição geográfica. Isto pode ser explicado por retenção de polimorfismo ancestral (Franco et al., 2010a), já verificado entre as espécies do cluster *D. buzzatii*. A similaridade dos haplótipos entre as localidades também indica o curto tempo de divergência entre as populações (Tajima, 1983; Excoffier, Foll, & Petit, 2009) e este padrão pode ser observado nos mapas de distribuição de haplótipos do gene *period*.

Os testes de neutralidade do gene *period* sugerem que as populações têm passado por expansão de área, e o compartilhamento de haplótipos entre as populações é indicativo de migração entre demes vizinhas (Ray, Currat, & Excoffier, 2003; Excoffier, 2004) ou uma expansão demográfica recente (Rogers & Harpending, 1992), que estariam de acordo com o

formato dos ramos da árvore de consenso Bayesiana. Esta expansão demográfica pode estar correlacionada ao período de expansão das florestas secas no continente (Werneck, et al., 2011a). Outro evento que poderia resultar nos valores negativos encontrados nos testes de neutralidade é a seleção purificadora, já observada para o cluster *D. buzzatii* por Franco et al. (2010a) na porção do gene *period* utilizada. Depois de um evento de seleção purificadora, muitos alelos surgem com baixa frequência (Holsinger, 2006), gerando um efeito semelhante à expansão populacional.

Devido aos diferentes tempos de coalescência entre os marcadores utilizados, espera-se que a história refletida pelo gene COI seja mais recente que a contada pelo gene *period*. Desta forma, a árvore de consenso Bayesiana do gene COI sugere que houve uma colonização inicial das localidades do Mato Grosso do Sul e de Teodoro Sampaio (TSA) (oeste do rio Paraná), há 70 mil anos pela linhagem A e pela H, o que refletiu em um grupo altamente diferenciado dos demais e coeso geneticamente, demonstrado por ambas as análises de estruturação e pela formação de uma rede de haplótipos distinta (Figura 11). A maior diversidade nucleotídica foi encontrada nesta região, confirmando que esta pode ser a localidade há mais tempo colonizada. O rio Paraná pode ser a barreira entre estas populações e as restantes, e as mudanças de curso parecem alterar o fluxo gênico entre as populações de margens opostas. Um dos sinais que levam a esta inferência é a maior similaridade da população de Teodoro Sampaio com as amostras de Bataguassu e Bodoquena, a oeste do rio Paraná, do que com populações mais próximas, a leste. Há cerca de um milhão de anos, durante a formação da Aloformação Paranaíba (Sallun, Suguio, & Stevaux, 2007), o leito do rio Paraná se localizava a direita do seu curso atual (Stevaux, 1994), sendo assim, naquela ocasião o fluxo gênico entre as populações de Bataguassu de Teodoro Sampaio era possível, mas não de Bataguassu com Bodoquena, que desde a alteração do curso do rio para o atual podem estabelecer fluxo gênico. Este evento foi concomitante ao período de aridez do Pleistoceno, que marca a formação do solo propício para a vegetação de cactos nesta região (Sallun et al., 2007).

É esperado que áreas de clima historicamente estável abriguem haplótipos mais antigos, e que durante as alterações climáticas, onde ocorrem mudanças na distribuição das espécies, estas áreas sirvam como refúgio da biodiversidade. Desta forma, propiciam maiores taxas de endemismo e manutenção da diversidade genética intraespecífica (Graham, Moritz, & Williams, 2006). De acordo com Werneck et al., (2011a) a região do Pontal do Paranapanema, representada pelas amostras de Teodoro Sampaio (TSA) e as amostras do

estado do Mato Grosso do Sul estariam em área de refúgio. Estas localidades possuem os mais altos valores de diversidade nucleotídica e são as ocupadas pelos haplótipos mais antigos.

As demais linhagens formaram uma rede haplotípica separada (Figura 9) e a árvore de consenso Bayesiana indica que a divergência destas se deu a menos de 20 mil anos. Atualmente, estas linhagens se estruturam em dois grupos, já sugeridos por De Brito et al. (2002), sendo ‘Paraná e São Paulo’ (leste do rio Paraná) e ‘Santa Catarina e Rio Grande do Sul’, e cada uma delas se formou a partir de pelo menos duas ondas de colonização, demonstrado pela árvore de consenso Bayesiana que se correlaciona com os grupos formados pela rede de haplótipos.

O grupo ‘Santa Catarina e Rio Grande do Sul’, proposto a partir da análise de estruturação, se formou a partir das linhagens E, D, C e G da árvore de consenso Bayesiana. A linhagem E avançou para o litoral de Santa Catarina e divergiu a cerca de 20 mil anos, sendo que hoje ocorre somente nesta localidade (clado 1.33, Figura 10E). A linhagem D teria divergido da C (com pouca consistência estatística) a cerca de 10 mil anos e alcançado Santa Catarina pelo Rio Grande do Sul, com fluxo gênico restrito por distância (clado 2.10, Figura 10F). Simultaneamente, a linhagem C, que é amplamente distribuída nas localidades do sudeste, teria se expandido continuamente até o Rio Grande do Sul pelo Paraguai, e colonizado Santa Catarina pelo litoral do Paraná (clado 4.1, Figura 10C). Atualmente, estas linhagens ocupam um ambiente arenoso, formado no litoral durante os últimos 10 mil anos (Villwock et al., 1986; Tanaka et al., 2009; Dillenburg et al., 2000).

Concomitantemente, os estados de São Paulo e Paraná teriam sido ocupados pela linhagem B e pela C. A linhagem B é predominante no Mato Grosso do Sul e Paraná, chegando a São Paulo com pouca representação. A maior parte dos indivíduos amostrados é pertencente da linhagem C, que também é a mais amplamente distribuída, sendo encontrada também no Rio Grande do Sul.

O compartilhamento de haplótipos entre as localidades pouco relacionadas geograficamente pode ser explicada pela retenção de polimorfismo ancestral, já demonstrado para o grupo (Franco et al., 2010a) devido ao pouco tempo de divergência da espécie, e por fluxo gênico entre as matas de galeria, sugerido por De Brito et al. (2002). O fluxo gênico entre estes grupos possivelmente ocorreu pela porção oeste da distribuição da espécie. Ambos os algoritmos de modelagem de nicho ecológico do LGM demonstram que esta área poderia ser propícia para a ocorrência de *D. antonietae*. A retenção de polimorfismo ancestral, introgressão citoplasmática (Bizzo, 2011) ou a formação de híbridos (Kokudai et al., 2011) também podem explicar a alta similaridade da linhagem H e G com outras espécies do cluster

D. buzattii. A identificação correta das linhagens foi confirmada por análise morfológica do edeago de ambas, e para a linhagem G também pela isoenzima isocitrato desidrogenase (IDH) (Bizzo, 2011).

De acordo com modelos gerados por Carnaval e Moritz (2008), a região oeste de São Paulo e Paraná e o litoral de Santa Catarina seriam áreas mais estáveis (mas não considerados refúgios) do que o restante da distribuição de *D. antonietae*. Isto pode explicar o encontro dos haplótipos mais antigos, segundo a rede maior da NCPA, em São Paulo e no Paraná e segundo a árvore de coalescência Bayesiana, no litoral de Santa Catarina. As áreas com maior diversidade nucleotídica seriam aquelas localizadas na periferia do núcleo de *Missiones* (Werneck et al., 2011a), que foram áreas de pouca estabilidade climática durante o Holoceno, e que exibem sinais de colonização recente (Excoffier et al., 2009). Este foi o padrão, confirmado pelos testes de neutralidade, encontrado no Rio Grande do Sul e em Santa Catarina, onde há excesso de haplótipos pouco distribuídos e de baixa frequência (Excoffier et al., 2009).

Tomando-se ambas as análises realizadas com o gene COI, poder-se-ia supor que o centro de dispersão e o extensivo fluxo gênico sugerido por outros autores (Manfrin & Sene 2006; Mateus & Sene 2007; Mateus et al., 2013) entre as populações de *D. antonietae* se dê pela região oeste da sua distribuição. As modelagens ecológicas do período LGM da distribuição de *D. antonietae* e de *C. hildmaniannus* (Silva, 2013) contribuem para esta hipótese. Modelagens ecológicas de 20 e 6 mil anos antes do presente indicam que as florestas secas estavam deslocadas para o brasileiro e pelos países vizinhos (Werneck, 2011; Prado & Gribbs, 1993). Esta expansão pode ter propiciado a expansão de área e demográfica de *D. antonietae*, culminando na evidente separação das linhagens mitocondriais visualizadas atualmente.

Foi demonstrado por Silva (2013) que os haplótipos cloroplastidiais mais antigos da espécie *Cereus hildamnianus*, da qual *D. antonietae* depende para oviposição, estão distribuídos nos estados de São Paulo, Paraná e oeste do Rio Grande do Sul, e que as amostras destas regiões formam um grupo bem estruturado e separado, porém com algum compartilhamento de haplótipos, das amostras da região leste do Rio Grande do Sul e Santa Catarina, seguindo um padrão parecido, porém mais evidente, do que o encontrado para *D. antonietae*. Mateus & Sene (2007) sugerem que a associação ecológica com este cacto adaptado a regiões mais úmidas, como os vale dos rios, tem resultado em isolamento lento tanto para cacto hospedeiro quanto para *D. antonietae*, criando condições que permitem fluxo gênico através dos valos dos rios por um longo período de tempo. Não é provável que isto

ocorra atualmente (Mateus et al., 2013), mas dados geológicos apontam que houveram pelo menos quatro eventos de glaciação com consequente expansão da vegetação xerofítica que permitiriam o contato secundário e a consequente homogeneização entre as populações isoladas durante o Quaternário (Ab'saber 1977).

A análise demográfica realizada com o método BSP a partir do gene COI indicou que *D. antonietae* esteve com tamanho populacional constante na maior parte do tempo, com a tendência de aumento do número efetivo populacional nos últimos 3 mil anos. Este período coincide com a extensão das FTSS que, previsto por modelagem ecológica e confirmado por registros palinológicos, está em expansão desde os últimos 6 mil anos (Werneck et al., 2011). Desta forma, o aumento de área de ocorrência da espécie *D. antonietae* poderia estar condicionada à expansão das florestas secas. Além disso, outros autores relatam que nos últimos 5 mil anos houve a expansão das matas de galeria sobre os campos no sul do Brasil, (Behling, Bauermann, & Neves, 2001; Behling, 2002; Behling *et al.*, 2005), ambiente esse que proporciona o crescimento de *C. hildmaniannus* em solo drenado ou rochoso.

Nota-se que a projeção para ocorrência de *D. antonietae* não inclui a área ocupada por vegetação do tipo Chaco. Esta vegetação é predominante na localidade de Porto Murtinho, onde apenas indivíduos de *D. koepferae* foram amostrados, com exceção de um indivíduo de *D. antonietae* identificado através da genotipagem do COI. O solo árido e salgado do Chaco é uma barreira para muitos organismos, e, apesar de padrões evolutivos não terem sido testados quanto a esta hipótese (Werneck, 2011), possivelmente este seja o motivo para o não encontro de *C. hildmaniannus* e *D. antonietae* nesta localidade. Já na Serra da Bodoquena, a cerca de 100 quilômetros de distância, cuja paisagem é formada pela vegetação tipo Cerrado em ecótono com a Floresta Sazonalmente Seca, foram encontrados muitos indivíduos de *D. antonietae* e apenas um de *D. koepferae*, evidenciando a relação da *D. antonietae* com o núcleo de *Misiones*.

A vegetação dos valos dos rios Paraná e Paraguai provavelmente são as que resistiram no período seco e que deram origem às florestas que ocuparam o espaço da vegetação xerofítica existente no Pleistoceno. As datas de divergência das linhagens amostradas são anteriores a esta expansão, sugerindo que estas florestas possivelmente abrigavam *D. antoneitae* e contribuíram para a sua expansão geográfica. Também se evidenciou retenção de haplótipos ancestrais e compartilhamento de haplótipos na porção norte da distribuição, especialmente nucleares. Os dados de expansão populacional obtidos através das várias metodologias empregadas neste estudo não corroboram a hipótese de gargalo genético recente. No entanto, a expansão recente da vegetação xerofítica pode explicar a expansão

populacional de *D. antonietae* e a baixa distribuição de alelos raros encontrados, especialmente no sul do Brasil, e o encontro de cactos em locais restritos com solo drenado pode ser um fator importante na recente estruturação genética.

6. CONCLUSÕES

A diversidade genética de *Drosophila antonietae* não é homogênea ao longo da distribuição geográfica da espécie. A não concordância entre os marcadores se deve ao tempo diferencial de coalescência e os dados obtidos são complementares no entendimento da história evolutiva de *D. antonietae*.

O gene *period* não demonstrou nenhum tipo de estruturação genética e as linhagens haplotípicas se encontram amplamente distribuídas geograficamente.

O gene COI demonstrou que há três agrupamentos genéticos, sendo que um está na região centro-oeste, o outro em Santa Catarina e Rio Grande do Sul e o terceiro em São Paulo e Paraná.

Há sinais de expansão populacional em todas as três populações, especialmente no sul, onde há excesso de haplótipos. O período de expansão da população de *D. antonietae* é coincidente com a expansão das florestas secas.

A população localizada no centro-oeste está em uma região de histórica estabilidade climática e é o mais antigo, formando uma rede de haplótipos distinta. Os outros dois agrupamentos formam uma mesma rede haplotípica e provavelmente o fluxo gênico entre elas ocorre pela porção a oeste da distribuição, hipótese corroborada por modelagem paleoecológica.

7. REFERÊNCIAS

- Ab'Saber, A. N. (1977). Espaços ocupados pela expansão dos climas secos na América do Sul, por ocasião dos períodos glaciais quaternários. *Paleoclimas*, 3, 1-17.
- Ab'Saber, A. N. (2003). *Os domínios de natureza no Brasil: potencialidades paisagísticas*. São Paulo: Ateliê Editorial.
- Amorim, D. D. S., & Pires, M. R. S. (1996). Neotropical biogeography and a method for maximum biodiversity estimation. *Biodiversity in Brazil, a first approach*. São Paulo, *CNPq*, 183-219.
- Anderson, R. P. (2012). Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions. *Annals of the New York Academy of Sciences*, 1260(1), 66-80.
- Avice, J. (2000). *Phylogeography: the history and formation of species*. Harvard: University Press.
- Avice, J. C. (2009). Phylogeography: retrospect and prospect. *Journal of Biogeography*, 36, 3-15.
- Avice, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E. et al. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual review of ecology and systematics*, 489-522.
- Barr, N. B., Cui, L., & McPherson, B. A. (2005). Molecular systematics of nuclear gene period in genus *Anastrepha* (Tephritidae). *Annals of the Entomological Society of America*, 98, 173-180.
- Bauzer, L., Souza, N., Ward, R., Kyriacou, C., & Peixoto, A. (2002). The period gene and genetic differentiation between three Brazilian populations of *Lutzomyia longipalpis*. *Insect Molecular Biology*, 11, 315-323.
- Beerli, P., & Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 4563-4568.
- Beheregaray, L. B. (2008). Twenty years of phylogeography: the state of the field and the challenges for the Southern hemisphere. *Molecular Ecology*, 17(17), 3754-3774.
- Behling, H. (2002). South and southeast Brazilian grasslands during Late Quaternary times: a synthesis. *Palaeogeography, Palaeoclimatology, Palaeoecology*, 177, 19-27.
- Behling, H., Bauermann, S. G., & Neves, P. C. P. (2001). Holocene environmental changes in the São Francisco de Paula region, southern Brazil. *Journal of South American Earth Sciences*, 14, 631-639.

- Behling, H., Pillar, V. D., & Bauermann, S. G. (2005). Late Quaternary grassland (Campos), gallery forest, fire and climate dynamics, studied by pollen, charcoal and multivariate analysis of the São Francisco de Assis core in western Rio Grande do Sul (southern Brazil). *Review of Palaeobotany and Palynology*, 133, 235 – 248.
- Bennett, K. D. (1990). Milankovitch cycles and their effects on species in ecological and evolutionary time. *Paleobiology*, 11-21.
- Bennett, K. D. (2004). Continuing the debate on the role of Quaternary environmental change for macroevolution. *Philosophical Transactions of the Royal Society*, 359, 295–303.
- Bizzo, L. E. M. (2011). Zona de contato secundário: simpatria entre duas espécies de *Drosophila*. Tese de Doutorado. Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto. Universidade de São Paulo, Ribeirão Preto.
- Bruxel, J., & Jasper, A. (2005). A família Cactaceae na Bacia Hidrográfica do Rio Taquari, RS, Brazil. *Acta Botanica Brasilica*, 19, 71-79.
- Cansian, A. M., Ruiz, A., & Sene, F. M. (1996). Inversões cromossômicas de populações naturais de *Drosophila serido* (grupo repleta) nas regiões sul e sudeste do Brasil. *Brazilian Journal of Genetics*, (Suppl.) 19, 267.
- Carnaval, A. C., & Moritz, C. (2008). Historical climate modelling predicts patterns of current biodiversity in the Brazilian Atlantic forest. *Journal of Biogeography*, 35(7), 1187-1201.
- Carnaval, A. C., Hickerson, M. J., Haddad, C. F. B., Rodrigues, M. T., & Moritz, C. (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, 323, 785.
- Chan, L. M., Brown, J. L., & Yode, A. D. (2011). Integrating statistical genetic and geospatial methods brings new power to phylogeography. *Molecular Phylogenetics and Evolution*, 59, 523-537.
- Chang, D. C., & Reppert, S. M. (2003). A novel C-terminal domain of *Drosophila* PERIOD inhibits dCLOCK: CYCLE-mediated transcription. *Current biology*, 13(9), 758-762.
- Clement, M. Posada, D., & Crandall, K. A. (2000). TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, 9, 1657–1659.
- Colot, H. V., Hall, J. C., & Rosbash, M. (1988). Interspecific comparison of the period gene of *Drosophila* reveals large blocks of non-conserved coding DNA. *The EMBO journal*, 7(12), 3929.
- Coyne, J. A. (1992). Genetics and speciation. *Nature*, 335, 511–515.
- De Brito, R. A., Manfrin, M. H., & Sene, F. M. (2002). Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. *Genetics and Molecular Biology*, 25(2), 161-171.
- Dillenburg, S. R., Laybauer, L., Mexias, A. S., Dani, N., Barboza, E. G., & Lummertz, C. N. (2000). Significado estratigráfico de minerais glauconíticos da planície costeira do Rio

- Grande do Sul, região da laguna de Tramandaí. *Revista Brasileira de Geociências*, 30(4), 649-654.
- Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC evolutionary biology*, 7(1), 214.
- Drummond, A., Rambaut, A., Shapiro, B., & Pybus, O. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution*, 22, 1185.
- Excoffier, L. (2004). Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. *Molecular Ecology*, 13(4), 853-864.
- Excoffier, L., & Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Reviews Genetics*, 7(10), 745-758.
- Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology*, 10, 564 -567.
- Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40, 481-501.
- Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479-491.
- 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.
- Franco, F. F., Prado, P. R. R., Sene, F. M., Costa, L. F., & Manfrin, M. H. (2006). Aedeagus morphology as a discriminant marker in two closely related cactophilic species of *Drosophila* (Diptera; Drosophilidae) in South America. *Annals of the Brazilian Academy of Sciences*, 78(2), 203-212.
- Franco, F. F., Sene, F. M., & Manfrin, M. H. (2010). Low satellite DNA variability in natural populations of *Drosophila antonietae* involved in different evolutionary events. *Journal of Heredity*, 101(5), 650-656.
- Franco, F. F., Silva-Bernardi, E. C. C., Sene, F. M., Hasson, E. R., & Manfrin, M. H. (2010a). Intra- and interspecific divergence in the nuclear sequences of the clock gene period in species of the *Drosophila buzzatii* cluster. *Journal of Zoological Systematics and Evolutionary Research*, 48(4), 322-331.
- Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147(2), 915-925.
- Garrick, R. C., Dyer, R. J., Beheregaray, L. B., & Sunnucks, P. (2008). Babies and bathwater: a comment on the premature obituary for nested clade phylogeographical analysis. *Molecular Ecology*, 17(6), 1401-1403.

- Graham, C. H., Moritz, C., & Williams, S. E. (2006). Habitat history improves prediction of biodiversity in rainforest fauna. *Proceedings of the National Academy of Sciences of the United States of America*, 103(3), 632-636.
- Griffiths, R. C., & Marjoram, P. (1996). Ancestral inference from samples of DNA sequences with recombination. *Journal of Computational Biology*, 3(4), 479-502.
- Guillot, G., Estoup, A., Mortier, F., & Cosson, J. F. (2005). A spatial statistical model for landscape genetics. *Genetics*, 170(3), 1261-1280.
- Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In *Nucleic acids symposium series* (Vol. 41, pp. 95-98).
- Hebert, P. D., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 270(Suppl 1), S96-S99.
- Hewitt, G. M. (2011). Quaternary phylogeography: the roots of hybrid zones. *Genetica*, 139, 617-638.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.
- Holsinger, K. E. (2006). Tajima's D, Fu's Fs, Fay and Wu's H, and Zeng et al.'s E. In K. E. Holsinger, *Lecture notes in population genetics* (pp. 239 – 244). Recuperado em 17 de novembro de 2014, de URL: <http://darwin.eeb.uconn.edu/eeb348/lecture-notes/molevol-tajima.pdf>.
- Hueck, K. (1972). *As florestas da America do Sul*. São Paulo: Poligono S.A.
- Kingman, J. F. C. (1982). The coalescent. *Stochastic Processes and their Applications*, 13, 235-248.
- Kokudai, C. B. S., Sene, F. M., & Manfrin, M. H. (2011). Sympatry and asymmetric introgression between the cactophilic species *Drosophila serido* and *Drosophila antonietae* (Diptera: Drosophilidae). *Annals of the Entomological Society of America*, 104(3), 434-442.
- Konopka, R. J., & Benzer, S. (1971). Clock mutants of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences USA*, 68, 2112.
- Kyriacou, C., & Hall, J. C. (1980). Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. *Proceedings of the National Academy of Sciences USA*, 77, 6729.
- Librado, P., & Rozas, R. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.
- Lorenci, M., Simão, D. P., Machado, L. P. B., & Mateus, R. P. (2010). Genetic variability

- analysis of two natural populations of *Drosophila antonietae* (Diptera; Drosophilidae), *Drosophila Information Service*, 93, 178-182.
- Machado, L. P. B., Mateus, R. P., Sene, F. M., & Manfrin, M. H. (2010). Microsatellite allele sequencing in population analyses of the South American cactophilic species *Drosophila antoneitae* (Diptera: Drosophilidae). *Biological Journal of the Linnean Society*, 100, 573–584.
- Machado, L. P. B., Silva, D. C., Simão, D. P., & Mateus, R. P. (2012). Spatial variation of genetic diversity in *Drosophila* species from two different South American environments. In P. M. Caliskan (Ed.), *Analysis of Genetic Variation in Animals*. Rijeka: InTech
- Manfrin, M. H., & Sene, F. M. (2006). Cactophilic *Drosophila* in South America: a model for evolutionary studies. *Genetica*, 126, 57–75.
- Manfrin, M. H., De Brito, R. O. A., & Sene, F. M. (2001). Systematics and evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) cluster using mtDNA. *Annals of the Entomological Society of America*, 94(3), 333-346.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209-220.
- Mateus, R. P., & Sene, F. M. (2003). Temporal and spatial allozyme variation in the South American cactophilic *Drosophila antonietae* (Diptera; Drosophilidae), *Biochemical Genetics*, 41(7/8), 219-233.
- Mateus, R. P., & Sene, F. M. (2007). Population genetic study of allozyme variation in natural populations of *Drosophila antonietae* (Insecta, Diptera). *Journal of Zoological Systematics and Evolutionary Research*, 45(2), 136-143.
- Mateus, R. P., Moura, M. O., Manfrin, M. H., Monteiro, S. G., & Sene, F. M. (2013). Contrasting Patterns of Within-Species Morphological Variation in Two Cactophilic *Drosophila* Species (Diptera: Drosophilidae). *Neotropical entomology*, 42(4), 384-392.
- Mayle, F. E. (2006). The late Quaternary biogeographical history of South American Seasonally Dry Tropical Forests: insights from palaeo-ecological data. In R. T. Pennington & J. A. Ratter (Eds.), *Neotropical savannas and dry forests: diversity, biogeography, and conservation*. Londres: CRC Press.
- Monteiro, S. G., & Sene, F. M. (1995). Estudo morfométrico de populações de *Drosophila serido* das regiões central e sul do Brasil. *Revista Brasileira de Genética* 18(3 suppl), 283.
- Newbold, T., Gilbert, F., Zalat, S., El-Gabbas, A., & Reader, T. (2009). Climate-based models of spatial patterns of species richness in Egypt's butterfly and mammal fauna. *Journal of Biogeography*, 36(11), 2085-2095.
- Nogués-Bravo, D. (2009). Predicting the past distribution of species climatic niches. *Ecology and Biogeography*, (*Global Ecol. Biogeogr.*), 18, 521-531.
- Panda, S., Hogenesch, J. B., & Kay, S. A. (2002). Circadian rhythms from flies to human.

Nature, 417,329–335.

- Pennington, R. T., Prado, D. E., & Pendry, C. A. (2000). Neotropical seasonally dry forests and Quaternary vegetation changes. *Journal of Biogeography*, 27,261-273.
- Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.
- Poppe, J. L., Schmitz, H. J., Grimaldi, D., & Valente, V. L. D. S. (2014). High diversity of Drosophilidae (Insecta, Diptera) in the Pampas biome of South America, with descriptions of new *Rhinoleucophenga* species. *Zootaxa*, 3779(2), 215-245.
- Poppe, J. L., Valente, V. L. S., & Schmitz, H. J. (2012). Structure of Drosophilidae assemblage (Insecta, Diptera) in Pampa biome (São Luiz Gonzaga, RS). *Papéis Avulsos de Zoologia (São Paulo)*, 52(16), 185-195.
- Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular biology and evolution*, 25(7), 1253-1256.
- Posada, D., & Crandall, K. A. (2001). Intraspecific gene genealogies: trees grafting into networks, *Trends in Ecology & Evolution*, 16(1), 37-45.
- Posada, D., Crandall, K., & Templeton, A. (2000). GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, 9, 487-488.
- Prado, D. E., & Gibbs, P. E. (1993). Patterns of species distributions in the dry seasonal forests of South America. *Annals of the Missouri Botanical Garden*, 80, 902-927.
- Ray, N., Currat, M., & Excoffier, L. (2003). Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution*, 20(1), 76-86.
- Regier, J. C., Fang, Q. Q., Mitter, C., Peigler, R. S., Friedlander, T. P., & Solis, M. A. (1998). Evolution and phylogenetic utility of the period gene in Lepidoptera. *Molecular Biology and Evolution*, 15(9), 1172-1182.
- Ruiz, A., Cassian, A. M., Kuhn, G. C. S., Alves, M. A. R., & Sene, F. M. (2000). The *Drosophila serido* speciation puzzle: putting new pieces together. *Genetica*, 108, 217–227
- Sallun, A. E. M., Suguio, K., & Stevaux, J. C. (2007). Proposição Formal do Alogruppo Alto Rio Paraná (SP, PR e MS). *Revista do Instituto de Geociências - USP*, 7(2), 49-70.
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), 5463-5467.
- Silva, A. F. G., & Sene, F. M. (1991). Morphological geographic variability in *Drosophila serido* (Diptera Drosophilidae). *Revista Brasileira de Entomologia*, 35(2), 455-468.
- Silva, G. A. R. (2013). Evolução de *Cereus hildmannianus* (Cactaceae) no sul do Brasil. Tese de Doutorado, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto.

- 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.
- Starmer, W. T., Lachance, M. A., Phaff, H. J., & Heed, W. B. (1990). The biogeography of yeasts associated with decaying cactus tissue in North America, the Caribbean, and northern Venezuela. *Evolutionary Biology*, 24, 253-296.
- Stevaux, J. C. (1994). The upper Paraná river (Brazil): Geomorphology, sedimentology and paleoclimatology, *Quaternary International*, 21, 143-161.
- Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, 105, 437-460.
- Tajima, F. (1989). The effect of change in population size on DNA polymorphism. *Genetics*, 123, 597-601.
- Tanaka, A. P. B., Giannini, P. C. F., Fornari, M., Junior, D. R. d. N., Sawakuchi, A. O., Rodrigues, S. I., Menezes, P. M. L., DeBlasis, P., & Porsani, J. L. (2009). A planície costeira holocênica de Campos Verdes (Laguna, SC): evolução sedimentar inferida a partir de georradar (GPR), granulometria e minerais pesados. *Revista Brasileira de Geociências*, 39(4), 751-767.
- Taylor, N., & Zappi, D. (2004). *Cacti of Eastern Brasil*. Royal Botanic Gardens, Kew. 511p.
- Templeton, A. R. (1998). Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, 7(4), 381-397.
- Templeton, A. R. (2004). Statistical phylogeography: methods of evaluating and minimizing inference errors. *Molecular Ecology*, 13(4), 789-809.
- Templeton, A. R. (2006). *Population genetics and microevolutionary theory*. Hoboken, N.J: Wiley-Liss.
- Templeton, A. R., Boerwinkle, E., & Sing, C. F. (1987). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, 117, 343-351.
- Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics*, 132(2), 619-633.
- Templeton, A. R., Routman, E., & Phillips, C. A. (1995). Separating population structure from population history: a cladistic analysis of geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, 140, 767-782.
- Thomé, M. T. C., Zamudio, K. R., Giovanelli, J. G. R., Haddad, C. F. B., Baldissera Jr, F. A., & Alexandrino, J. (2010). Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic Forest. *Molecular Phylogenetics and Evolution*,

55, 1018-1031.

- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 25(24), 4876–4882.
- Tidon-Sklorz, R., & Sene, F. M. (1995). Evolution of the *buzzatii* cluster (*Drosophila repleta* group) in middle South America. *Evolutionary Biology*, 9, 71-85.
- Tidon-Sklorz, R., & Sene, F. M. (2001). Two new species of the *Drosophila serido* sibling set (Diptera, Drosophilidae). *Iheringia*, (90), 141-146.
- Toma, D. P., Bloch, G., Moore, D., & Robinson, G. E. (2000). Changes in period mRNA levels in the brain and division of labor in honey bee colonies. *Proceedings of the National Academy of Sciences*, 97(12), 6914-6919.
- Villwock, J. A., Tomazelli, L. J., Loss, E. L., Dehnhardt, E. A., Horn F, N. O., Bachi, F. A. et al. (1986). Geology of the Rio Grande do Sul coastal province. In International symposium on sea-level changes and quaternary shorelines (pp. 79-97).
- Wang, R. L., & Hey, J. (1996). The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. *Genetics*, 144, 1113-1126.
- Werneck, F. P. (2011). The diversification of eastern South American open vegetation biomes: historical biogeography and perspectives. *Quaternary Science Reviews*, 30(13), 1630-1648.
- Werneck, F. P., Costa, G. C., Colli, G. R., Prado, D. E., & Sites Jr, J. W. (2011a). Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidencegeb. *Global Ecology and Biogeography*, 20(2), 272-288.
- Zeisset, I., & Beebee, T. J. C. (2008). Amphibian phylogeography: a model for understanding historical aspects of species distributions. *Heredity*, 101, 109–119.