UNIVERSIDADE DE SÃO PAULO

FFCLRP – DEPARTAMENTO DE BIOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA COMPARADA

História evolutiva de *Drosophila serido* (“cluster” *Drosophila buzzatii*)

Taís Carmona Lavagnini-Pizzo

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Doutor em Ciências, Área: BIOLOGIA COMPARADA.

RIBEIRÃO PRETO - SP

2015

UNIVERSIDADE DE SÃO PAULO

FFCLRP – DEPARTAMENTO DE BIOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA COMPARADA

História evolutiva de *Drosophila serido* (“cluster” *Drosophila buzzatii*)

Aluna: Taís Carmona Lavagnini-Pizzo

Orientadora: Profa. Dra. Maura Helena Manfrin

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Doutor em Ciências, Área: BIOLOGIA COMPARADA.

RIBEIRÃO PRETO - SP

2015

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

FICHA CATALOGRÁFICA

Lavagnini-Pizzo, Taís Carmona

História evolutiva de *Drosophila serido* (“cluster” *Drosophila buzzatii*)

Orientadora: Profa. Dra. Maura Helena Manfrin

161 p.: 11 il.

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto/USP. Departamento de Biologia. Área: Biologia Comparada.

1. Filogeografia. 2. “cluster” *Drosophila buzzatii*. 3. Caatinga. 4. Costa Atlântica. 5. Estrutura Populacional.

Taís Carmona Lavagnini-Pizzo

Título da tese: História evolutiva de *Drosophila serido* (“cluster” *Drosophila buzzatii*)

Tese apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da USP, como parte das exigências para a obtenção do título de Doutor em Ciências, Área: BIOLOGIA COMPARADA.

Aprovado em:

Banca Examinadora:

Prof.(a) Dr.(a) _____

Instituição: _____ Assinatura _____

Prof.(a) Dr.(a) _____

Instituição: _____ Assinatura _____

Prof.(a) Dr.(a) _____

Instituição: _____ Assinatura _____

Prof.(a) Dr.(a) _____

Instituição: _____ Assinatura _____

Prof.(a) Dr.(a) _____

Instituição: _____ Assinatura _____

Dedico esta Tese a pessoas muito especiais, cujo apoio e incentivo foram fundamentais nesta jornada:

Meus amados pais, Marli Carmona Lavagnini e Antonio Carlos Lavagnini, que SEMPRE acreditaram nos meus sonhos,

Ao meu companheiro, José Henrique Pizzo, por todo apoio e compreensão,

À minha orientadora Profa. Dra. Maura Helena Manfrin, por ter me proporcionado tanto aprendizado junto ao Laboratório de Genética Evolutiva.

AGRADECIMENTOS

À Profa. Dra. Maura Helena Manfrin por toda dedicação e apoio direcionados à minha formação como bióloga e pesquisadora. É incomensurável o quanto aprendi durante estes quatro anos de convivência e trabalho.

Ao Prof. Dr. Fábio de Melo Sene por ter compartilhado seus conhecimentos e experiências com relação às moscas, os cactos e o trabalho de campo.

Ao Prof. Dr. Rogério Pincela Mateus, à Profa. Dra. Maura Helena Manfrin, ao Prof. Dr. Fábio de Melo Sene, ao Prof. Dr. Fernando de Faria Franco, à doutoranda Dora Yovana Barrios Leal, ao meu pai Antonio Carlos Lavagnini e ao agente de viagens Jonas Oliveira pelo apoio na organização e realização das viagens de campo para coleta de material biológico.

Ao analista de sistemas do Departamento de Genética, Pedro Roberto Prado, pelo auxílio na utilização do sistema operacional Linux em nosso servidor.

A todos os meus familiares, especialmente meus pais, Marli Carmona Lavagnini e Antonio Carlos Lavagnini, e ao meu companheiro, José Henrique Pizzo, pelo apoio e incentivo incondicionais.

Aos colegas de laboratório Camila Borgonove, Camila Kokudai, Mateus Santos, Dora Barrios, Natácia Lima, Gislaine Silva, Rafaela Rosseti e Jaqueline Koser, pela convivência e amizade durante estes anos de trabalho.

Aos amigos de graduação Caio Silva, Cíntia Isicawa, Camila Linhares e Verônica Lebre, pela amizade e apoio.

Ao técnico do Laboratório de Genética Evolutiva, Paulo Ricardo Epifânio, pela amizade e excelente trabalho que realiza.

Às secretárias do Programa de Pós-Graduação em Biologia Comparada (FFCLRP-USP) Vera Cássia Cicilini de Lucca e Renata Andrade Cavallari pela ajuda com relação à compreensão das questões burocráticas.

Ao Programa de Pós-Graduação em Biologia Comparada, ao Departamento de Biologia e à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto pela oportunidade de realização deste doutorado.

À FAPESP pela concessão da bolsa de doutorado (2011/10499-7) e do auxílio financeiro (2011/51652-2).

À CAPES, CNPq, FINEP e USP cujas verbas mantém o Laboratório de Genética Evolutiva.

SUMÁRIO

1	Introdução.....	11
1.1	Considerações iniciais.....	11
1.2	Florestas Tropicais Sazonalmente Secas (FTSS) e Mata Atlântica.....	11
1.3	Espécies cactófilas de <i>Drosophila</i>	15
1.4	A espécie <i>Drosophila serido</i>	16
1.4.1	Populações do Nordeste.....	16
1.4.2	Populações do Litoral.....	17
1.5	Coalescência, Filogeografia e Marcadores Moleculares.....	18
1.5.1	Gene nuclear <i>kl-5</i> ligado ao cromossomo Y..	19
1.5.2	Gene mitocondrial COI.....	20
1.5.3	Gene nuclear <i>period</i> ligado ao cromossomo X.....	21
1.5.4	Genes nucleares autossômicos <i>GstD1</i> e α -esterase5.....	21
2	Objetivos.....	23
3	Material e Métodos.....	24
3.1	Material Biológico.....	24
3.2	Amplificação dos genes nucleares <i>period</i> , $\alpha E5$, <i>GstD1</i> , <i>kl-5</i> e mitocondrial COI, purificação e sequenciamento.....	33
3.3	Análise dos dados.....	38
3.3.1	Leitura e alinhamento das sequências.....	38
3.3.2	Estatística descritiva.....	38
3.3.3	Modelos de substituição nucleotídica.....	38
3.3.4	Estimativa das taxas de substituição nucleotídica.....	39
3.3.5	Análise de Variância Molecular, Teste de Mantel e Estrutura Populacional.....	39
3.3.6	Testes de neutralidade.....	40
3.3.7	Testes de seleção.....	41
3.3.8	Análises demográficas.....	42
3.3.9	Análises filogeográficas.....	44
4	Resultados.....	45
4.1	Coleta de material biológico.....	45
4.2	Gene nuclear <i>period</i>	47
4.2.1	Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica.....	47
4.2.2	Análise de Variância Molecular, teste de Mantel e estrutura populacional..	48
4.2.3	Testes de neutralidade.....	51
4.2.4	Testes de seleção.....	52
4.2.5	Análises demográficas.....	53
4.2.6	Análises filogeográficas.....	59
4.3	Gene nuclear <i>kl-5</i>	64
4.3.1	Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica.....	64
4.3.2	Análise de Variância Molecular, teste de Mantel e estrutura populacional..	66
4.3.3	Testes de neutralidade.....	69

4.3.4	Testes de seleção.....	69
4.3.5	Análises demográficas.....	71
4.3.6	Análises filogeográficas.....	74
4.4	Gene nuclear <i>GstDI</i>	77
4.4.1	Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica.....	77
4.4.2	Análise de Variância Molecular, teste de Mantel e estrutura populacional...	79
4.4.3	Testes de neutralidade.....	82
4.4.4	Testes de seleção.....	82
4.4.5	Análises demográficas.....	84
4.4.6	Análises filogeográficas.....	87
4.5	Gene nuclear <i>α-esterase5</i>	91
4.5.1	Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica.....	91
4.5.2	Análise de Variância Molecular, teste de Mantel e estrutura populacional...	93
4.5.3	Testes de neutralidade.....	96
4.5.4	Testes de seleção.....	96
4.5.5	Análises demográficas.....	98
4.5.6	Análises filogeográficas.....	101
4.6	Gene mitocondrial COI.....	109
4.6.1	Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica.....	109
4.6.2	Análise de Variância Molecular, teste de Mantel e estrutura populacional	111
4.6.3	Testes de neutralidade.....	114
4.6.4	Testes de seleção.....	114
4.6.5	Análises demográficas.....	116
4.6.6	Análises filogeográficas.....	119
5	Discussão	126
5.1	Populações do Nordeste.....	126
5.2	Populações do Litoral.....	131
5.3	Populações de Santa Catarina.....	137
5.4	Seleção.....	139
6	Conclusões	141
7	Referências	142

RESUMO

Lavagnini-Pizzo, T.C. **História evolutiva de *Drosophila serido* (“cluster” *Drosophila buzzatii*)**. 2015. 161f. Tese (Doutorado). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 2015.

O “cluster” *Drosophila buzzatii* é formado por sete espécies endêmicas da América do Sul e que apresentam relação ecológica obrigatória com cactos. Dentre estas espécies, *Drosophila serido* possui ampla distribuição geográfica, na Caatinga e ao longo da costa Atlântica, e é considerada uma espécie politípica sendo dividida em dois grupos: populações do nordeste e do litoral. Com o objetivo de compreender os processos que moldaram a distribuição atual das populações de *D. serido* foram realizadas análises com sequências dos genes nucleares *period* e *kl-5*, ligados aos cromossomos sexuais X e Y, respectivamente, genes nucleares autossômicos *GstD1* e *αE5*, e gene mitocondrial COI. Dentre os resultados obtidos, a homogeneidade genética entre as populações do Nordeste e a divisão norte-sul entre as populações da costa Atlântica foram observadas em todos os marcadores. Três padrões quanto à estruturação populacional na costa Atlântica foram observados para os diferentes marcadores. A hipótese de que a Chapada Diamantina seja o centro de dispersão para a espécie foi confirmada pelo presente trabalho, no entanto, o T_{MRCA} estimado para populações de Santa Catarina sugerem que estas sejam populações ancestrais de *D. serido*, sendo que o Nordeste teria sido colonizado a partir delas. Eventos de expansão de área e fragmentação alopátrica foram sugeridos como inferências filogeográficas para explicar o isolamento atual de populações de *D. serido* em Goiás e Minas Gerais. De acordo com as estimativas do T_{MRCA} , é possível que os eventos causais dos processos históricos inferidos estejam relacionados à influência das flutuações climáticas do Quaternário na distribuição geográfica da vegetação/cactos, afetando indiretamente as populações de moscas cactofílicas. É possível que eventos de seleção, associado aos fatores ecológicos quanto ao uso de cactos, também possam ter contribuído para o processo de diversificação populacional, uma vez que foi encontrada evidência de seleção positiva para os genes autossômicos.

Palavras-chave: Caatinga, estrutura populacional, filogeografia, Mata Atlântica.

ABSTRACT

Lavagnini-Pizzo, T.C. **História evolutiva de *Drosophila serido* (“cluster” *Drosophila buzzatii*)**. 2015. 161f. Tese (Doutorado). Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 2015.

Drosophila buzzatii cluster comprises seven species endemic of South America and that present a mandatory ecological association with cacti. Among these species, *Drosophila serido* has a wide geographical range, in Caatinga and along Atlantic coast, and is considered as a polytypic species, divided in two groups: northeast and coast populations. The purpose of this study was understand the process that shaped the current distribution of *D. serido* populations through genetic analysis using sequences of nuclear genes *period* and *kl-5*, X- and Y-linked, respectively, autosomal genes *GstD1* e *αE5*, and mitochondrial gene COI. The genetic homogeneity among Northeast populations and the north-south division among coast Atlantic populations were observed for all markers. Three patterns related to population structure in coast Atlantic were seen for the different markers. Hypothesis that Diamantina Plateau was the dispersion center for the species were confirmed at this study, although, T_{MRCA} estimated for Santa Catarina populations suggested that these ones were ancestral, and that Northeast would be colonized from them. Expansion range and allopatric fragmentation were historical events suggested as phylogeographic inferences to explain the current isolation of *D. serido* populations in Goiás and Minas Gerais. According to T_{MRCA} estimations, it is possible that causal events of historical process inferred were related to the influence of climatic fluctuations during Quaternary in the geographic distribution range of vegetation/cacti, indirectly affecting the populations of cactophilic flies. Furthermore, selection events, associated with ecological factors due to cacti use, as well as contributed to diversification process in populations, once it was found evidence of positive selection at autosomal genes.

Keywords: Atlantic Rainforest, Caatinga, phylogeography, population structure.

1 – Introdução

1.1 – Considerações Iniciais

A análise da genealogia de genes associada à distribuição geográfica, que é uma abordagem filogeográfica, permite inferências relacionadas aos eventos históricos que influenciaram a distribuição atual de determinado táxon (AVISE, 2000). Esta área vem se desenvolvendo notadamente, sendo considerada um dos campos mais integrativos da biologia (HICKERSON et al., 2010). Muito tem sido discutido com relação à escolha e utilização de genes mitocondriais e nucleares na análise filogeográfica (AVISE 2000, 2009; BRITO; EDWARDS, 2009; HARE, 2001), e às análises estatísticas (BEAUMONT et al., 2010; KNOWLES; MADDISON, 2002; TEMPLETON; ROUTMAN; PHILLIPS, 1995; TEMPLETON, 2004, 2009).

Uma radiação do gênero *Drosophila*, endêmico da América do Sul, conhecida como “cluster” *Drosophila buzzatii*, possui associação ecológica obrigatória com espécies da família Cactaceae, uma vez que suas larvas utilizam tecidos necrosados de cactos como substrato para desenvolvimento larval (MANFRIN; SENE, 2006). O fato das cactáceas serem elementos comuns no que é denominado Floresta Tropical Sazonalmente Seca (FTSS), torna este grupo de sete espécies de moscas excelente modelo para discussão de eventos históricos, como por exemplo, a influência das flutuações climáticas do Quaternário (AB’SABER, 1977) no que diz respeito à dinâmica biogeográfica dos biomas da América do Sul.

Assim, o presente trabalho teve como objetivo descrever e discutir, por meio da abordagem filogeográfica multi-*locus*, processos evolutivos relacionados à diversificação populacional de uma das espécies de maior distribuição geográfica do “cluster”, *Drosophila serido*, cuja distribuição geográfica inclui a Caatinga e enclaves de vegetação xérica ao longo da distribuição da Mata Atlântica, desde o Rio Grande do Norte até Santa Catarina.

1.2 - Florestas Tropicais Sazonalmente Secas (FTSS) e Mata Atlântica

A combinação de eventos históricos, diversidade climática e ambiental permitiu uma grande diversificação da cobertura florestal na América do Sul, englobando desde as florestas úmidas até regiões semi-áridas (HUECK, 1972). Entre as últimas, a diagonal seca, orientada na direção nordeste-sudoeste, localizada ao leste da América do Sul, entre as florestas tropicais Amazônica e Atlântica, é composta por biomas de áreas abertas denominados de Caatinga, Cerrado e Chaco (AB’SABER, 2003b).

Entre estes biomas de áreas abertas, a Caatinga é reconhecida como um componente das Florestas Tropicais Sazonalmente Secas (FTSS), que são florestas de regiões tropicais com um regime de chuvas sazonal e vários meses de secas severas (MOONEY; BULLOCK; MEDINA, 1995), incluindo: solo rico em nutrientes e com pH de alto à moderado, precipitação altamente sazonal, menor que 1600 mm/ano e um período de pelo menos cinco meses de seca (MURPHY; LUGO, 1986). A vegetação é heterogênea e apresenta maior riqueza de espécies quando comparadas às florestas temperadas, mas do ponto de vista florístico, são menos diversas quando comparadas às florestas tropicais (MAYLE, 2004; MURPHY; LUGO, 1986).

A distribuição da FTSS na América do Sul é descontínua, sendo que as três maiores áreas de ocorrência foram denominadas de “núcleos”, por Prado e Gibbs (1993) sendo: 1) núcleo Caatinga, na região nordeste do Brasil, considerado também o maior núcleo da FTSS (PRADO; GIBBS, 1993); 2) núcleo Misiones, ao longo do sistema de rios Paraná-Paraguai; e 3) núcleo Piedmont, localizado no sudoeste da Bolívia e noroeste da Argentina.

Para explicar o caráter descontínuo na distribuição geográfica da FTSS foi proposta a Hipótese do Arco Pleistocênico (PENNINGTON et al., 2000; PRADO; GIBBS, 1993; PRADO, 2000). Neste caso, os eventos de expansão e retração da vegetação durante as flutuações climáticas do Quaternário favoreceram a expansão da área de distribuição da FTSS durante períodos mais secos. Neste cenário, os núcleos Caatinga, Misiones e Piedmont, observados atualmente, estariam conectados formando um contínuo, que foi posteriormente fragmentado durante os períodos interglaciais com condições quentes e úmidas (AB’SABER, 1977; PENNINGTON et al., 2000).

A Mata Atlântica possui ampla área de distribuição geográfica, estendendo-se ao longo da costa brasileira desde o Rio Grande do Norte até o Rio Grande do Sul e é considerado o bioma mais diverso e ameaçado do planeta (FIASCHI; PIRANI, 2009; MARTINS, 2011). Por ter sua distribuição geográfica no limite leste do escudo brasileiro, este bioma constitui uma área complexa do ponto de vista topográfico, provavelmente devido à neotectônica e às mudanças no nível do mar durante o Quaternário (LIMA, 2000; SUGUIO et al., 2005). A costa brasileira pode ser dividida em seis setores geomorfológicos (AB’SABER, 2003a), sendo que a área de ocorrência da Mata Atlântica encontra-se nos últimos quatro setores:

- Litoral Equatorial Amazônico: ocorre ao longo do litoral dos Estados do Amapá, Pará e parte do Maranhão, região com grande quantidade de desembocaduras de rios e de manguezais;
- Litoral Setentrional do Nordeste: consiste na grande exceção climática, ecológica e paisagística da costa brasileira. Estende-se do nordeste do Maranhão até o litoral norte do Rio Grande do Norte. Na região que compreende o litoral do Ceará e a parte norte de Rio Grande do Norte o clima semi-árido chega até o mar;
- Litoral Oriental do Nordeste: compreende desde a região de Touros-RN (litoral leste do Rio Grande do Norte) até o delta do rio São Francisco. Esta área da costa é estreita devido aos tabuleiros que chegam como falésias no litoral. Este setor geomorfológico é também o litoral da Zona da Mata Nordestina, após esta área, em direção ao continente, há uma transição para matas secas, agreste e Caatinga;
- Litoral Leste: estende-se desde o delta do rio São Francisco até o delta do rio Doce. Em toda sua área há o predomínio de extensas restingas, com exceção do Recôncavo Baiano, devido ao acidente da linha de costa representado pela Baía de Todos os Santos;
- Litoral Sudeste: este é o macrossetor litorâneo mais diversificado e acidentado de todo o país. Possui 1500 km de extensão, desde o Espírito Santo até o Paraná, com grandes áreas de restinga e alguns manguezais. A exceção desta vegetação ocorre na faixa costeira de Cabo Frio-RJ até Macaé-RJ, em que há um reduto de Caatinga. O litoral do Rio de Janeiro e de São Paulo parecem ter sido bastante influenciados pelas flutuações no nível do mar durante o Quaternário, tanto que o Maciço da Juréia (Peruíbe-SP) tem sido interpretado como uma paleoilha que foi ressoldada ao continente por sedimentação. Na Baía de Guaratuba-PR, limite deste setor geomorfológico, há presença de manguezais;
- Litoral Sul: o início deste setor geomorfológico se dá quando terminam as escarpas florestadas da Serra do Mar, na fronteira entre Paraná e Santa Catarina. A fisionomia do litoral catarinense muda, com paleoilhas e ilhas florestadas, como a de São Francisco do Sul. Em toda a área há restingas e lagunas. Em Garopaba-SC há dunas costeiras fixadas pela vegetação.

A heterogeneidade de ambientes, relacionada às características geomorfológicas, ao longo da área de distribuição da Mata Atlântica, tem se mostrado congruente com a diferenciação populacional de grupos taxonômicos, auxiliando no entendimento dos processos envolvidos. Trabalhos com abordagens filogeográficas desenvolvidos com diferentes grupos taxonômicos, mostraram divisões de linhagens evolutivas que coincidem com áreas ao norte e ao sul da Mata Atlântica (MARTINS, 2011).

Há dados que sugerem uma dinâmica de relações entre as áreas de ocorrência de biomas como a Caatinga, Mata Atlântica e Cerrado (AB'SABER, 1977; BONATELLI et al., 2014; PRADO; GIBBS, 1993; WERNECK, 2011; WERNECK et al., 2011). Áreas de instabilidade na Caatinga, inferidas por meio de projeções paleoclimáticas realizadas para FTSS, estão mescladas com brejos de altitude, relictos de Mata Atlântica (PÔRTO; CABRAL; TABARELLI, 2004; WERNECK et al., 2011), sugerindo que houve entrelaçamento dos limites de ocorrência entre estes biomas (AB'SABER, 1977; WERNECK et al., 2011). Da mesma forma, são observadas áreas de Caatinga em meio à Mata Atlântica, como no Estado do Rio de Janeiro entre Cabo Frio e Macaé (AB'SABER, 2003a), e inseridas no Cerrado como, por exemplo, o enclave localizado no vale do Rio Paranã (WERNECK, 2011), cuja nascente encontra-se próximo ao Distrito Federal.

Diversos estudos apoiam a hipótese de que as flutuações climáticas durante o Quaternário influenciaram os eventos de expansão e retração da vegetação, incluindo estudos geomorfológicos (AB'SABER, 1977; WANG et al., 2004), palinológicos (DE OLIVEIRA; BARRETO; SUGUIO, 1999; LEDRU, 1993), com paleodunas (DE OLIVEIRA; BARRETO; SUGUIO, 1999), espeleológicos (AULER et al., 2004; WANG et al., 2004), paleomodelagem (CARNAVAL; MORITZ, 2008; CARNAVAL et al., 2009), e estudos biogeográficos (CARNAVAL; BATES, 2007; FRANCO; MANFRIN, 2013; MAGALHÃES et al., 2014; MORAES et al., 2009; PENNINGTON et al., 2000; QUIJADA-MASCAREÑAS et al., 2007). Entretanto, a extensão da influência das flutuações climáticas que ocorreram durante o Pleistoceno ainda tem sido muito questionada (MAGALHÃES et al., 2014; POELCHAU; HAMRICK, 2013; WERNECK et al., 2011).

A compreensão da história biogeográfica dos biomas da América do Sul tem sido investigada por meio da análise filogeográfica de grupos vegetais (WERNECK et al., 2001; WERNECK et al., 2012a) e animais (CARNAVAL et al., 2009; MAGALHÃES et al., 2014; THOMÉ et al., 2010). As espécies cactófilas do “cluster” *Drosophila*

buzzatii são espécies neotropicais, endêmicas da América do Sul (MANFRIN; SENE, 2006), apresentando associação ecológica obrigatória com espécies da família Cactaceae, elementos encontrados nas FTSS (MAYLE, 2004). Desta forma, estas espécies constituem excelentes modelos para estudos relacionados à dinâmica biogeográfica na América do Sul.

1.3 - Espécies cactófilas de *Drosophila*

No gênero *Drosophila* alguns grupos apresentam íntima relação ecológica obrigatória com espécies da família Cactaceae, utilizando seus tecidos necrosados como sítio para desenvolvimento larval (BARKER; STARMER, 1982). Esta associação pode ter influenciado a diferenciação das espécies, tanto em termos ecológico-adaptativos quanto históricos. Do ponto de vista ecológico-adaptativo, a diferenciação populacional pode ser influenciada pela composição química dos cactos utilizados e pela expressão gênica de genes relacionados aos processos de desintoxicação de compostos secundários (MATZKIN et al., 2006; MATZKIN, 2008, 2012). Em termos históricos, é possível que eventos subsequentes e recorrentes, porém não sincrônicos, de expansão e retração das formações vegetais abertas (árido/semi-árido), durante paleoeventos climáticos alteraram a distribuição de populações de cactos (BONATELLI et al., 2014; TAYLOR; ZAPPI, 2004) e, conseqüentemente, influenciando as populações de *Drosophila* (FRANCO; MANFRIN, 2013; MORALES, 2005; MORAES et al., 2009; SANTOS, 2011).

Na porção sudeste da América do Sul, grande diversidade de espécies da família Cactaceae é encontrada na Caatinga (TAYLOR; ZAPPI, 2004), sendo que em regiões adjacentes, populações de *Drosophila* são encontradas associadas às cactáceas de afloramentos rochosos e solos arenosos, porém com baixa diversidade de espécies. Estas regiões são adequadas para o desenvolvimento destas plantas, formando populações isoladas, que são consideradas registros de eventos de retração e expansão da vegetação xerofítica (AB'SABER, 1977).

Na América do Sul, uma radiação de espécies cactófilas de *Drosophila* forma o “cluster” *Drosophila buzzatii* (MANFRIN; SENE, 2006), que é composto por sete espécies: *D. buzzatii*, que ocorre por toda área de distribuição do “cluster”, *D. koepferae* que ocorre no domínio do Chaco; *D. antonietae*, que ocorre nas áreas da bacia dos rios Paraná-Paraguai; *D. serido*, *D. seriema*, *D. gouveai*, *D. borborema* que ocorrem em

enclaves de vegetação aberta e campos rupestres no domínio do Cerrado e pelo domínio da Caatinga e costa Atlântica.

O “cluster” *Drosophila buzzatii* é considerado um grupo monofilético, sugerido pela análise de marcadores cromossômicos (RUIZ; WASSERMAN, 1993), genes nucleares (FRANCO et al., 2010; SANTOS et al., 2009) e citoplasmáticos (MANFRIN; DE BRITO; SENE, 2001), porém as topologias das árvores filogenéticas apresentam incongruências, o que pode ser uma consequência das dificuldades intrínsecas associadas à reconstrução de filogenia para espécies com pouco tempo de divergência (MACHADO; HEY, 2003), como é o caso destas *Drosophila* cactófilas, cujos eventos cladogenéticos foram datados para o Quaternário (FRANCO, MANFRIN, 2013; FRANSACK, 2011).

1.4 - A espécie *Drosophila serido*

Drosophila serido tem ampla distribuição geográfica, ocorrendo no bioma da Caatinga, em populações isoladas nos Estados de Goiás e Minas Gerais, e em enclaves de vegetação xerófitica na costa Atlântica. As populações desta espécie são caracterizadas pela presença da inversão fixa $2x^7$ (WASSERMAN; RICHARDSON, 1987) e edeago morfotipo “A” (SILVA; SENE, 1991), embora exista variação morfológica ao longo de sua distribuição geográfica (FRANCO et al., 2008). Populações desta espécie estão associadas a várias espécies de cactos, como: *Cephalocereus piauhyensis*, *Cereus hildmannianus*, *Cereus fernambucensis*, *Opuntia ficus-indica*, *Opuntia monacantha* (MANFRIN; SENE, 2006; PEREIRA; VILELA; SENE, 1993). Análises de hidrocarbonetos da cutícula (OLIVEIRA et al., 2011), inversões cromossômicas polimórficas (RUIZ et al., 2000; TOSI; SENE, 1989), diversidade haplotípica mitocondrial (MANFRIN; DE BRITO; SENE, 2001; MORALES, 2005; FRANCO; MANFRIN, 2013) e análises cariotípicas (BAIMAI; SENE; PEREIRA, 1983) sugerem a existência de dois grupos populacionais em *D. serido*: as populações do nordeste e do litoral.

1.4.1. - Populações do Nordeste

As populações do nordeste de *D. serido* apresentam as inversões cromossômicas polimórficas $2a^8$, $2b^8$, $2c^8$ e $2d^8$ (TOSI; SENE, 1989) e são homogêneas para a maioria dos marcadores utilizados até o momento (BAIMAI; SENE; PEREIRA, 1983; FRANCO et al., 2008; RUIZ et al., 2000; SILVA; SENE, 1991). Análises

filogeográficas para estas populações, usando sequências parciais do gene mitocondrial COI (DE BRITO; MANFRIN; SENE, 2002b; FRANCO; MANFRIN, 2013), sugerem que o centro de dispersão desta espécie são localidades próximas à Chapada Diamantina, interior da Bahia, e que os eventos de expansão se iniciaram há aproximadamente 40 mil anos (FRANCO; MANFRIN, 2013). A partir desta região, expansão de área em direção ao norte explica a colonização de localidades na Chapada da Borborema-PB, e em direção ao litoral, colonizando áreas próximas ao município de Salvador-BA.

1.4.2 - Populações do Litoral

As populações ao longo da costa Atlântica apresentam outra dinâmica evolutiva, sugerida pela não congruência e politipia observada em diversos marcadores utilizados. As populações de São Paulo e Santa Catarina apresentam, sobre a inversão $2x^7$, a inversão cromossômica fixa $2y^9$ e, sobre esta, as inversões $2x^8$ e $2w^8$ (RUIZ et al., 2000). Com relação às placas metafásicas, foram descritos: cromossomos metafásicos tipo III em Arraial do Cabo-RJ, tipo IV em Peruíbe-SP, e um outro tipo em Florianópolis-SC (BAIMAI; SENE; PEREIRA, 1983; BIFFI et al., 2001). A análise da variação morfológica dos edeagos em populações de *D. serido* também evidenciou a presença de dois grupos distintos na costa atlântica: um grupo com as populações de Bertioxa-SP e Penha-SC e o outro grupo com as populações de São Sebastião-SP e Arraial do Cabo-RJ (FRANCO et al., 2008).

Análises filogeográficas para estas populações, com base em sequências parciais do gene mitocondrial COI (MORALES, 2005), sugeriram que as populações do litoral formam um grupo isolado das populações do Nordeste, agrupando as populações de São Paulo até o litoral sul da Bahia. Dentro deste grupo, foi sugerido fluxo gênico entre as populações de São Paulo e Rio de Janeiro, e isolamento dessas com as populações de Mucuri-BA. A análise da variabilidade haplotípica mitocondrial mostrou que as populações do litoral de Santa Catarina e de *D. serido* da Caatinga, formam um grupo, o que é incongruente com outros marcadores, sugerindo a colonização do litoral sul por expansão de área de populações do interior da Bahia, e isolamento em relação às outras populações do litoral, sendo, portanto, restrito devido a isolamento por distância (MORALES, 2005; KOKUDAI; SENE; MANFRIN, 2011).

1.5 – Coalescência, Filogeografia e Marcadores Moleculares

A filogeografia é um campo de estudo relativamente novo, o termo foi utilizado pela primeira vez em 1987 (AVISE et al., 1987), tendo como fundamento a associação entre a genealogia de genes e sua distribuição geográfica, em nível intraespecífico (AVISE, 2000, 2009). Por meio de estudos paleoclimáticos e informações geológicas, é possível uma abordagem multidisciplinar para inferências relacionadas aos eventos demográficos e vicariantes, fluxo gênico e migração, que influenciaram a dinâmica populacional de determinada espécie ao longo do tempo (AVISE, 2000).

A análise filogeográfica tem como premissa teórica a coalescência, proposta por Kingman (1982), em que é possível modelar o processo de deriva genética sob uma perspectiva passada (HEIN; SCHIERUP; WIUF, 2005; TEMPLETON, 2011). Este processo retrospectivo é possível quando são utilizadas sequências homólogas de DNA, pois considerando a premissa de tamanho populacional finito, dois haplótipos atuais coalescem em uma molécula ancestral comum em algum ponto do passado, caracterizando assim um evento coalescente (HEIN; SCHIERUP; WIUF, 2005; TEMPLETON, 2011).

O desenvolvimento de modelos matemáticos promoveu a associação entre o conceito de evento coalescente e o modelo populacional de Wright-Fisher, por meio da probabilidade de identidade por descendência na geração t (HEIN; SCHIERUP; WIUF, 2005; TEMPLETON, 2011).

Estudos pioneiros nesta área concentravam seus esforços na obtenção e análise de sequências de DNA mitocondrial, devido à facilidade de obtenção por sequenciamento direto. Além disso, características como elevada taxa evolutiva, ausência de recombinação e herança matrilinear, tornaram este marcador ideal para a reconstrução de genealogias de genes para inferências de história recente e parâmetros populacionais (AVISE, 2000; BRITO; EDWARDS, 2009; HARE, 2001; HICKERSON et al., 2010). Contudo, com as informações de estudos filogenéticos e filogeográficos utilizando um único marcador foi possível observar a necessidade da inclusão de outros marcadores que diferissem do DNA mitocondrial em ploidia e modo de herança, fato que levou ao início dos trabalhos com análises multi-*locus*. Assim, a concordância entre a genealogia de múltiplos *loci* caracterizam eventos históricos comuns, permitindo inferências de eventos vicariantes e evitando falsas inferências filogeográficas (BRITO; EDWARDS, 2009).

Simulações computacionais comparando a influencia do número de pares de base por sequência, número de sequências e número de *loci* sobre a acurácia na estimativa de parâmetros populacionais, elucidaram que aumentar o número de *loci* pode ser o mais relevante (CARLING; BRUMFIELD, 2007; FELSENSTEIN, 2006), especialmente no caso de *loci* não-ligados (FELSENSTEIN, 2006). Além disso, a análise comparativa de dados multi-*locus* pode reduzir possíveis erros, permitindo que haja várias réplicas dos processos evolutivos observados (PLUZHNIKOV; DONNELLY, 1996), especialmente no caso de eventos históricos que afetam o genoma como um todo.

A ploidia e o tamanho efetivo populacional do marcador molecular têm influencia direta sobre o tempo de coalescência da amostra, possibilitando a observação de diferentes cenários ao longo do tempo no processo de diversificação populacional. Sendo assim, para um gene nuclear autossômico o tempo de coalescência será quatro vezes menor do que para um gene mitocondrial (BRITO; EDWARDS, 2009) e, portanto, é esperada uma razão de 1:1:3:4 para os tempos de coalescência em genes ligados ao cromossomo Y, mitocondriais, ligados ao cromossomo X e autossômicos, respectivamente (TEMPLETON, 2011).

1.5.1 – Gene nuclear *kl-5* ligado ao cromossomo Y

O cromossomo Y no gênero *Drosophila* foi um dos primeiros a ser caracterizado genética e citologicamente, porém pouco tem sido feito para identificação de genes com cópia única, assim como quais forças evolutivas moldaram este cromossomo (CARVALHO, 2002). De forma geral, o cromossomo Y parece ser pouco conservado devido ao grande número de ganhos e perdas de genes, sendo que o ganho sugere que ele seja mais recente com relação aos demais cromossomos (KOERICH et al., 2008).

É provável que a variação genética do cromossomo Y esteja envolvida na competição entre espermatozóides, fator que pode influenciar o *fitness* dos machos (CARVALHO, 2002; CHIPPINDALE, RICE, 2001) e, portanto, selecionar determinados polimorfismos.

Entre estes fatores, o gene *kl-5* foi identificado como um gene de cópia única por Gepner e Hays (1993), codificando a proteína *dynein heavy chain*, um fator de fertilidade que está relacionado ao processo de espermatogênese. Os fatores de fertilidade *kl-1*, *kl-2*, *kl-3*, *ks-1* e *ks-2* foram descritos para o cromossomo Y e considerados como genes de cópia única (CARVALHO, 2002). Estas descobertas tem contrariado a hipótese de que o cromossomo Y teria se originado a partir do

cromossomo X, uma vez que genes parálogos dos fatores de fertilidade são autossômicos e não ligados ao X como seria esperado (CARVALHO, 2002).

Análises evolutivas dos 12 genomas de espécies de *Drosophila* revelaram excesso de substituições sinônimas com relação às não-sinônimas em sequências do gene *kl-5*, indicativo de que este gene é funcional (KOERICH et al., 2008). Este gene está ligado ao cromossomo Y apenas na ramificação *D. grimshawi* – *D. virilis*+*D. mojavensis* (KOERICH et al., 2008), sugerindo que este gene pode apresentar polimorfismos, por ser mais recente.

Resultados interessantes foram obtidos por meio de análises filogeográficas com sequências do cromossomo Y para o golfinho *Stenella longirostris* concordando com a separação em subespécies sugerida por caracteres morfológicos. O mesmo não foi evidenciado pelos marcadores mitocondriais e autossômicos. Além disso, os marcadores do cromossomo Y sugeriram presença de híbridos e foram concordantes com as variações morfológicas existentes entre as subespécies (ANDREWS et al., 2013). Como no caso de *S. longirostris*, a utilização de genes de cópia única no cromossomo Y para estudos filogeográficos pode auxiliar na compreensão do processo de diferenciação entre as linhagens evolutivas, nordeste e litoral, observadas para *D. serido*.

1.5.2 – Gene mitocondrial COI

O DNA mitocondrial foi o marcador molecular mais utilizado no início dos estudos filogeográficos, devido à elevada taxa evolutiva, ausência de recombinação e herança matrilinear (AVISE, 2000, 2009; BRITO; EDWARDS, 2009; HARE, 2001; HICKERSON et al., 2010).

Entre os genes do DNA mitocondrial, o gene mitocondrial COI tem sido extensivamente utilizado como marcador molecular em estudos filogenéticos (FRANSACK, 2011; KOKUDAI; SENE; MANFRIN, 2001; MANFRIN; DE BRITO; SENE, 2001) e filogeográficos (DE BRITO; MANFRIN; SENE, 2002a, 2002b; FRANCO; MANFRIN, 2013; MORAES et al., 2009; MORALES, 2005; SANTOS, 2011), em espécies do “cluster” *D. buzzatii*.

A subunidade I do citocromo C, codificada pelo gene mitocondrial COI, é componente fundamental da cadeia respiratória atuando no transporte de elétrons (NELSON; COX, 2006; SIMON et al., 1994).

1.5.3 – Gene nuclear *period* ligado ao cromossomo X

O gene nuclear *period* tem sido considerado um candidato para mudanças evolutivas que afetam o comportamento sexual em *Drosophila*, uma vez que está envolvido na regulação do ciclo circadiano (KONOPKA, BENZER, 1971) e na produção de sons espécie-específicos pelos machos de *Drosophila* durante a corte sexual (KYRIACOU; HALL, 1980), podendo estar envolvido no estabelecimento de isolamento reprodutivo (WANG; HEY, 1996).

Estudos envolvendo questões filogenéticas (BARR; CUI; MCPHERON, 2005; FRANCO et al., 2010; REGIER et al., 2008) e de divergência populacional (BAUZER et al., 2002; presente trabalho) foram realizados utilizando este gene como marcador molecular. Em análises filogenéticas entre as espécies do “cluster” *D. buzzatii* o gene *period* mostrou-se informativo em níveis interespecíficos, sugeriu hibridação introgressiva entre linhagens de *D. buzzatii* e *D. koepferae*, e parece estar sob forte influência de seleção purificadora (FRANCO et al., 2010). Na filogenia, para representar o ramo *D. serido* foram utilizados indivíduos de ambos os grupos, nordeste e litoral, sendo que o resultado consistiu em elevados índices de diversidade nucleotídica e ramo com baixo suporte estatístico, concordando com o status de espécie politépica (FRANCO et al., 2010).

1.5.4 – Genes nucleares autossômicos *GstD1* e α -*esterase5*

Entre os diferentes fatores seletivos relacionados ao uso de cactos como recurso alimentar, sua composição química, que varia entre as espécies e é tóxica para muitos grupos (FRANK; FOGLEMAN, 1992; MATZKIN, 2008), é de grande importância na determinação do padrão de relação entre o cacto hospedeiro e a espécie de *Drosophila* (FRANK; FOGLEMAN, 1992).

Nesta relação, diferentes *loci* estão envolvidos nos processos de desintoxicação (MATZKIN et al., 2006) como os genes pertencentes às famílias citocromo P450 (BONO et al., 2008), glutathione S-transferase (Gst) (MATZKIN et al., 2006) e α -esterase (ROBIN et al., 2000). A relação entre linhagens evolutivas e os cactos hospedeiros utilizados foi observada em *D. mojavensis* (MATZKIN, 2008) e *D. mettleri* (BONO et al., 2008) por meio da análise de genes pertencentes às famílias glutathione S-transferase (*GstD1*) e citocromo P450, respectivamente. Em análises filogenéticas entre espécies do “cluster” *D. buzzatii* o gene α -*esterase5* (*α E5*) mostrou-se informativo em

nível interespecífico e foram obtidos ramos com elevado suporte estatístico (SANTOS et al., 2009).

Ao longo da ampla distribuição geográfica das populações de *D. serido* diversos cactos são utilizados como recurso para sítio de oviposição e desenvolvimento larval (PEREIRA; VILELA; SENE, 1993), sendo que o nível de endemismo de cactos na Caatinga é maior que em áreas de restinga e afloramentos rochosos ao longo do litoral atlântico (TAYLOR; ZAPPI, 2004). Portanto, análises filogeográficas utilizando os genes *GstD1* e *αE5* podem permitir uma avaliação do processo evolutivo do ponto de vista ecológico-adaptativo, devido às diferenças na diversidade de espécies de cactos na Caatinga e litoral atlântico.

2 - Objetivos

O presente trabalho teve como objetivo geral acrescentar informações a cerca da estrutura populacional e história evolutiva das populações da espécie cactofílica *Drosophila serido* por meio de análises de sequências parciais de genes nucleares e mitocondrial. Os objetivos específicos foram:

- aumentar a amostragem de populações de *D. serido*, especialmente em áreas ainda não exploradas pelo nosso grupo de pesquisa;
- obter sequências parciais de DNA mitocondrial (gene COI) para os indivíduos coletados nas novas regiões amostradas e obtenção de sequências parciais para os genes nucleares (*period*, *kl-5*, *GstD1* e *αE5*) na amostra total;
- definir os tipos de seleção sobre as sequências obtidas de genes nucleares e mitocondrial;
- realizar análises comparativas de estrutura populacional e inferências filogeográficas utilizando sequências dos genes nucleares e mitocondrial;
- estabelecer hipóteses sobre a influencia de eventos paleoclimáticos do Quaternário na diferenciação populacional de *D. serido*;
- reavaliar a hipótese de que localidades próximas à Chapada Diamantina sejam um centro de dispersão para a espécie;
- avaliar o “status” taxonômico de *D. serido* com relação aos agrupamentos nordeste e litoral.

3 - Material e Métodos

3.1 - Material biológico

Amostras de DNA previamente extraídas de indivíduos de *D. serido* e armazenados no Laboratório de Genética Evolutiva (MORALES, 2005; FRANCO et al., 2010; FRANCO; MANFRIN, 2013), foram utilizadas para obtenção de sequências dos genes nucleares *period*, *GstD1*, *αE5*, *kl-5* e mitocondrial COI (Tabela 1, Figura 1). Além destas, foram incluídas nas análises sequências disponíveis no banco de dados do *NCBI GenBank* (gene *period*: FJ267328.1 – FJ267340.1; gene COI: JN124572.1 – JN124595.1 e JN124592.1 – JN124674.1).

Novas amostras de indivíduos de *D. serido* foram obtidas de 33 localidades, em cinco viagens, sendo elas: I – litoral e interior de Sergipe, litoral sul de Alagoas, litoral norte e interior da Bahia; II – litoral e interior do Rio Grande do Norte, litoral e interior do Ceará; III – região norte da Bahia (limite com os Estados do Piauí e Pernambuco) e região centro-sul do Piauí; IV – Peruíbe-SP, V – norte de Minas Gerais (Tabela 2, Figura 2).

As coletas foram realizadas com armadilhas fechadas, contendo como isca uma mistura de banana, laranja e fermento biológico (*Saccharomyces cerevisiae*). Estas foram colocadas próximas às cactáceas, em local sombreado e fresco, onde permaneceram por dois ou três dias. Após este período, com auxílio de rede entomológica, os indivíduos foram coletados e transferidos para frascos de vidro, contendo meio de cultura, devidamente etiquetados e acondicionados. No laboratório, os indivíduos machos foram identificados através da morfologia do edeago (VILELA, 1983; SILVA; SENE, 1991).

As amostras obtidas nestas viagens foram processadas para extração de DNA genômico, com o auxílio do conjunto de reagentes *DNeasy Blood & Tissue Kit* (Qiagen). O DNA extraído foi utilizado para amplificação e sequenciamento dos genes de interesse.

Tabela 1. Populações amostradas de *Drosophila serido* em trabalhos anteriores, cujos DNA dos indivíduos está armazenado no Laboratório de Genética Evolutiva. FFCLRP-USP.

Localidade	CL	Coletores	Data de coleta	CG	N _{period}	N _{COI}	N _{ae5}	N _{GstD1}	N _{kl-5}
Cabedelo-PB	N71	Franco FF, Santos MHS	16 a 23-II-2008	07,03°S, 34,83°W	07	07*	05	06	06
Junco do Serido-PB	N70	Franco FF, Santos MHS	16 a 23-II-2008	07,00°S, 36,71°W	08	10*	06	06	06
Pocinhos-PB	N68	Franco FF, Santos MHS	16 a 23-II-2008	07,17°S, 36,05°W	06	04*	06	06	06
Exu-PE	N67	Franco FF, Santos MHS	16 a 23-II-2008	07,65°S, 39,79°W	02	02*	01	-	02
Milagres-BA	J92	Morales AC, Kuhn GCS, Franco FF, Brisson JA	23-III-2002	11,20°S, 39,90°W	08*	06	05	04	06
Manoel Vitorino-BA	J93	Morales AC, Kuhn GCS, Franco FF, Brisson JA	23-III-2002	14,15°S, 40,24°W	07	-	05	06	06
Mucuri-BA	N32	Sene FM, Franco FF, Almeida JG	04 a 12-VII-2005	18,00°S, 39,42°W	05	05	06	06	06
Morro da Barrinha-BA	N37	Franco FF, Oliveira C, Esguícero A	05 a 14-VIII-2008	09,90°S, 40,30°W	01	10*	05	04	06
10km S de Morro do Chapéu-BA	N40	Franco FF, Oliveira C, Esguícero A	05 a 14-VIII-2008	11,65°S, 41,29°W	04	05*	04	05	04
Irecê-BA	N43	Franco FF, Oliveira C, Esguícero A	05 a 14-VIII-2008	11,22°S, 41,95°W	03	09*	06	06	06
Cach. Ferro Doido-BA	N42	Franco FF, Oliveira C, Esguícero A	05 a 14-VIII-2008	11,63°S, 41,00°W	08	10*	06	06	06

Tabela 1. Continuação.

Localidade	CL	Coletores	Data de coleta	CG	N_{period}	N_{COI}	N_{aE5}	N_{GstD1}	N_{kl-5}
Mucugê-BA	N45	Franco FF, Oliveira C, Esguícero A	05 a 14-VIII-2008	13,00°S, 41,40°W	04*	05*	04	03	04
Salvador-BA	N65	Franco FF	07 a 09-X-2007	12,93°S, 39,32°W	07	10*	06	06	05
Cabrália-BA	N35	Sene FM, Franco FF, Almeida JG	04 a 12-VII-2005	16,31°S, 39,01°W	04	05*	05	05	05
Andaraí-BA	N46	Franco FF, Oliveira C, Esguícero A	05 a 14-VIII-2008	12,86°S, 41,31°W	06	06*	05	05	06
Morro do Chapéu-BA	N39	Franco FF, Oliveira C, Esguícero A	05 a 14-VIII-2008	11,60°S, 41,20°W	01*	01*	-	-	-
Itaberaba-BA	J95	Morales AC, Kuhn GCS, Franco FF, Brisson JA	23-III-2002	12,56°S, 40,31°W	04	04	04	03	02
Itaúnas-ES	N31	Sene FM, Franco FF, Almeida JG	04 a 12-VII-2005	18,41°S, 39,70°W	05	06*	06	04	02
Arraial do Cabo-RJ	N20	Morales AC, Franco FF, Silva-Bernardi ECC	14 a 20-III-2004	23,00°S, 42,00°W	09*	06	06	06	06
Macaé-RJ	N22			22,32°S, 41,82°W	04	05	05	03	03
São Sebastião-SP	N17	Morales AC, Franco FF	05 a 09-X-2003	23,82°S, 45,42°W	06	07	05	02	03
Bertioga	H49	Sene FM, Monteiro SG, Ruiz A	10 a 18-V-1995	23,90°S, 46,10°W	05*	04	03	04	04
Guaratuba-SP	N2	Manfrin MH, Prado N	VIII-2002	23,85°S, 46,13°W	05	06	06	04	03

Tabela 1. Continuação.

Localidade	CL	Coletores	Data de coleta	CG	N _{period}	N _{COI}	N _{<i>αE5</i>}	N _{<i>GstD1</i>}	N _{<i>kl-5</i>}
Ilha de Alcatrazes-SP	Ilh			24,10°S, 45,69°W	07	-	06	06	06
Paúba-SP	N19	Morales AC, Franco FF	05 a 09-X-2003	23,80°S, 45,55°W	07	-	06	06	06
Governador Celso Ramos-SC	N9	Morales AC, Franco FF, Silva-Bernardi ECC	24 a 31-III-2003	27,37°S, 48,53°W	03	-	03	-	03
Florianópolis-SC	N88	Fransak F, Mateus RP, Bizzo L	25 a 29-IX-2008	27,56°S; 48,56°W	08	07	06	02	03
Camboriu-SC	N10	Morales AC, Franco FF, Silva-Bernardi ECC	24 a 31-III-2003	27,05°S, 48,59°W	03	-	03	-	03
Ilha de São Francisco-SC	N11	Morales AC, Franco FF, Silva-Bernardi ECC	24 a 31-III-2003	26,28°S, 48,53°W	01	-	-	--	-
Penha-SC (Praia Grande)	J52			26,75°S, 48,66°W	07	07	05	01	05
Serra do Cipó-MG	N56	Franco FF, Esguícero A	19 a 24-XI-2006	19,31°S; 43,61°W	06	08*	06	06	06
Água Fria-GO	AF	Tidon, R†		15,00°S, 47,87°W	05	06	05	04	04

CL: códigos das localidades; CG: coordenadas geográficas (grau-decimal); N_{period}: número de amostras sequenciadas para o gene nuclear *period*; N_{COI}: número de amostras sequenciadas para o gene mitocondrial COI; N_{*αE5*}: número de amostras sequenciadas para o gene nuclear *α-esterase5*; N_{*GstD1*}: número de amostras sequenciadas para o gene nuclear *GstD1*; N_{*kl-5*}: número de amostras sequenciadas para o gene nuclear *kl-5*.

† Amostras gentilmente cedidas pelo grupo de pesquisa da Profa. Rosana Tidon (Universidade de Brasília-UNB).

*Localidades em que foram utilizadas sequências disponíveis no *NCBI GenBank*.

Tabela 2. Populações amostradas de *Drosophila serido* em viagens de coleta realizadas durante o desenvolvimento do presente trabalho.

Localidade	CL	Coletores	Data de coleta	CG	N_{period}	N_{COI}	N_{aE5}	N_{GstD1}	N_{kl-5}
Viagem I									
Penedo-AL	R30	Lavagnini TC, Leal DYB, Manfrin MH, Mateus RP	15 a 22-IX-2012	10,32°S, 36,52°W	06	06	06	06	06
Monte Alegre de Sergipe-SE	R28	Lavagnini TC, Leal DYB, Manfrin MH, Mateus RP	15 a 22-IX-2012	10,13°S, 37,54°W	06	06	05	03	05
Porto das Cabras-SE	R29	Lavagnini TC, Leal DYB, Manfrin MH, Mateus RP	15 a 22-IX-2012	10,80°S, 36,93°W	06	05	04	01	04
Pacatuba-SE	R31	Lavagnini TC, Leal DYB, Manfrin MH, Mateus RP	15 a 22-IX-2012	10,46°S, 36,63°W	06	06	05	06	06
Riacho-BA	R27	Lavagnini TC, Leal DYB, Manfrin MH, Mateus RP	15 a 22-IX-2012	09,61°S, 38,22°W	05	04	05	05	05
Costa Azul-BA	R32	Lavagnini TC, Leal DYB, Manfrin MH, Mateus RP	15 a 22-IX-2012	11,61°S, 37,46°W	05	06	06	03	04
Viagem II									
Rodolfo Fernandes-RN	R33	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	05,82°S, 37,99°W	01	01	01	01	01
Jucurutu-RN	R34	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	05,95°S, 37,04°W	01	01	01	01	01

Tabela 2. Continuação.

Localidade	CL	Coletores	Data de coleta	CG	N_{period}	N_{COI}	N_{aE5}	N_{GstD1}	N_{kl-5}
Lajes Pintadas-RN	R37	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	06,17°S, 36,11°W	06	08	06	06	06
Bom Jesus-RN	R38	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	06,02°S, 35,65°W	07	06	05	06	06
Maxaranguape-RN(1)	R39	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	05,54°S, 35,27°W	06	05	06	06	06
Maxaranguape-RN(2)	R40	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	05,47°S, 35,28°W	06	06	-	-	-
Tabuleiro do Norte-CE	R35	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	05,28°S, 38,16°W	02	02	02	02	02
Russas-CE	R36	Lavagnini TC, Leal DYB, Lavagnini AC, Lavagnini MC	23 a 30-VI-2012	04,85°S, 38,04°W	06	06	04	04	06
Viagem III									
Casa Nova-BA	R41	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	09,24°S, 41,16°W	-	-	-	-	-
Remanso-BA	R42	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	09,48°S, 42,25°W	-	-	-	-	-
Sta Cruz do Piauí-PI	R43	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	07,24°S, 41,88°W	-	-	-	-	-
Wall Ferraz-PI	R44	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	07,31°S, 41,91°W	-	-	-	-	-

Tabela 2. Continuação.

Localidade	CL	Coletores	Data de coleta	CG	N_{period}	N_{COI}	N_{aE5}	N_{GstD1}	N_{kl-5}
Nova Santa Rita-PI	R45	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	08,11°S, 42,06°W	-	-	-	-	-
São Raimundo Nonato-PI	R46	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	08,90°S, 42,56°W	-	-	-	-	-
Colônia do Gurguéia-PI	R47	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	08,28°S, 43,81°W	-	-	-	-	-
Eliseu Martins-PI	R48	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	08,12°S, 43,93°W	-	-	-	-	-
Canto do Buriti-PI	R49	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	08,15°S, 43,06°W	-	-	-	-	-
Tamboriú-PI	R50	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	08,43°S, 42,83°W	-	-	-	-	-
Jurema-PI	R51	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	09,19°S, 43,15°W	-	-	-	-	-
São Braz do Piauí-PI	R52	Lavagnini TC, Leal DYB, Lavagnini AC	20 a 29-VIII-2012	09,04°S, 42,91°W	-	-	-	-	-
Viagem IV									
Peruíbe-SP	R53	Lavagnini TC, Manfrin MH, Franco FF	10 a 13-XII-2012	24,25°S; 46,90°W	-	-	-	-	-
Viagem V									
Pedra Azul-MG (1)	R54	Lavagnini TC, Leal DYB, Lavagnini AC	19 a 28-IX-2013	16,05°S; 41,23°W	-	-	-	-	-
Pedra Azul-MG (2)	R55	Lavagnini TC, Leal DYB, Lavagnini AC	19 a 28-IX-2013	16,04°S; 41,26°W	-	-	-	-	-

Tabela 2. Continuação.

Localidade	CL	Coletores	Data de coleta	CG	N_{period}	N_{COI}	N_{<i>αE5</i>}	N_{<i>GstDI</i>}	N_{<i>kl-5</i>}
Almenara-MG	R56	Lavagnini TC, Leal DYB, Lavagnini AC	19 a 28-IX-2013	16,00°S; 40,95°W	10	10	06	08	09
Itinga-MG (1)	R57	Lavagnini TC, Leal DYB, Lavagnini AC	19 a 28-IX-2013	16,60°S; 41,75°W	-	-	-	-	-
Itinga-MG (2)	R58	Lavagnini TC, Leal DYB, Lavagnini AC	19 a 28-IX-2013	16,60°S; 41,70°W	-	-	-	-	-
Jequitinhonha- MG	R59	Lavagnini TC, Leal DYB, Lavagnini AC	19 a 28-IX-2013	16,52°S; 41,30°W	01	01	-	01	01

CL: códigos das localidades; CG: coordenadas geográficas (grau-decimal); N_{period}: número de amostras sequenciadas para o gene nuclear *period*; N_{COI}: número de amostras sequenciadas para o gene mitocondrial COI; N_{*αE5*}: número de amostras sequenciadas para o gene nuclear *α-esterase5*; N_{*GstDI*}: número de amostras sequenciadas para o gene nuclear *GstDI*; N_{*kl-5*}: número de amostras sequenciadas para o gene nuclear *kl-5*.

Figura 1. Mapa topográfico do Brasil indicando as localidades amostradas. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

3.2 - Amplificação dos genes nucleares *period*, *αE5*, *GstD1*, *kl-5* e mitocondrial COI, purificação e sequenciamento

A amplificação das regiões de interesse nos genes nucleares *period* (Figura 2), *αE5* (Figura 3), *GstD1* (Figura 4), *kl-5* (Figura 5), e no gene mitocondrial COI (Figura 6) foram realizadas por meio da reação em cadeia da polimerase (PCR) em um termociclador Veriti® Thermal Cycler 96-well (Applied Biosystems).

Figura 2. Estrutura do gene nuclear *period* em *Drosophila melanogaster*. Os íntrons são representados por linhas e os éxons por retângulos. O domínio PAS (região C2) está colorida com a cor cinza. No éxon cinco estão indicadas as regiões não conservadas n2, n3 e n4, interespaçadas pelas regiões conservadas C3, C4 e C5 (COLOT et al., 1988). A faixa listrada indica a região n2, formada pela região repetitiva treonina-glicina. O retângulo negro representa o domínio inibitório CCID (região C3 até região C4). Abaixo da estrutura do gene, a barra negra indica a região do éxon cinco que foi analisada no presente trabalho. Figura reproduzida a partir de FRANCO et al.(2010).

Figura 3. Estrutura do gene nuclear *α-esterase5* em *Drosophila borborema*. Os íntrons são representados por linhas e os éxons por retângulos. Abaixo da estrutura do gene, a barra negra indica a região analisada no presente trabalho. Modificado a partir de SANTOS et al. (2009).

Figura 4. Esquema do gene nuclear *GstD1* em *Drosophila mojavensis*. O retângulo cinza representa a totalidade do gene, que não possui íntrons. Abaixo desta representação, a barra negra indica a região analisada no presente trabalho.

Figura 5. A: representação parcial da estrutura do cromossomo Y em *Drosophila melanogaster* com detalhe para a localização dos genes *kl-5*, *kl-3* e *kl-2*. **B:** detalhe para o exon 12 do gene *kl-5*, sendo que a barra negra corresponde à região amplificada e analisada no presente trabalho.

Figura 6. Esquema do genoma mitocondrial, obtido em OLIVEIRA; GARESSE; KAGUNI (2010), em *Drosophila melanogaster* com detalhe para a localização do gene COI. Os retângulos branco, cinza e negro correspondem, respectivamente, aos genes ND2, COI e COII, presentes na estrutura do genoma mitocondrial. Abaixo da estrutura do gene COI há uma barra negra que corresponde à região amplificada e analisada neste trabalho.

A concentração de cada reagente e a quantidade de DNA extraído utilizados para amplificação dos genes de interesse estão descritos na Tabela 3. Em todas as reações foi utilizada água estéril para completar o volume final de 25µl. Os oligonucleotídeos iniciadores *forward* e *reverse* para cada gene e as condições da amplificação estão descritos na Tabela 4.

Tabela 3. Concentração dos reagentes utilizados para realização da reação em cadeia da polimerase (PCR), em termociclador, para amplificação dos genes de interesse.

	<i>period</i>	<i>αE5</i>	<i>GstD1</i>	<i>kl-5</i>	COI
DNA extraído	50ng	50ng	50ng	50ng	50ng
Tampão	1x	1x	1x	1x	1x
MgCl₂	1,5mM	1,5mM	1,5mM	2,0 mM	1,5mM
Taq polimerase*	1,25U	1,25U	1,25U	1,25U	1,25U
dNTP**	200 μM	200 μM	200 μM	200 μM	200 μM
Oligonucleotídeos iniciadores	0,3 μM	0,5 μM	0,3 μM	0,3 μM	0,3 μM

*GoTaq[®] Hot Start Polymerase (Promega)

**Fermentas

Tabela 4. Oligonucleotídeos iniciadores *forward* e *reverse* e condições de PCR utilizados neste trabalho para cada gene de interesse.

Gene (referência)	Oligonucleotídeo iniciador <i>forward</i>	Oligonucleotídeo iniciador <i>reverse</i>	Condições PCR
<i>period</i> (FRANCO et al., 2010)	perCBf 5' TGGGAGGGCGAGGCGAACAA 3'	perCBr 5' GGCATGGGTTGGTACATCAT 3'	1 ciclo – 1'30" a 94°C 35 ciclos – 30" a 94°C 30" a 55°C 1' a 72°C 1 ciclo – 4' a 72°C
<i>αE5</i> (SANTOS et al., 2009)	Ester 2F 5' TGGACTGAAGGACCAGGTTT 3'	Ester CH1R 5' AGCCATGCCAGAAGATCCTA 3'	1 ciclo – 4' a 94°C 35 ciclos – 40" a 94°C 40" a 51°C 1' a 72°C 1 ciclo – 3'30" a 72°C
<i>GstD1</i> (desenhados no presente trabalho)	GstD1deg-F 5'CCCGAGTTCGTGAAGATCAAYCCNCA 3'	GstD1deg-R 5' TTCTCGTACCACTTGTTACGTTNKCR 3'	1 ciclo – 2' a 94°C 30 ciclos – 1' a 94°C 1' a 56°C 2' a 76°C 1 ciclo – 5' a 72°C
<i>kl-5</i> (DYER et al., 2011)	K15-F8 5' KTTYGARYTRCAAGGKCCAGATCC 3'	K15-R6 5' GCSGGCCAYTCRAATATCCAMAC 3'	1 ciclo – 2' a 94°C 40 ciclos – 1' a 94°C 1' a 56°C 2' a 72°C 1 ciclo – 5' a 72°C
COI (SIMON et al. 1994)	TY-J-1460 5' TACAATTTATCGCCTAAACTTCAGCC 3'	C1-N-2191 5' CCCGGTAAAATTAAAATATAAACTTC 3'	1 ciclo – 1'30" a 94°C 25 ciclos – 40" a 94°C 40" a 48°C 2' a 72°C 1 ciclo – 4' a 72°C

Os produtos de PCR foram verificados por meio da eletroforese em gel de agarose 1% e tampão de cuba TBE 1x (Tris-base 0,89M; Ácido Bórico 0,89M; EDTA 20mM). Para visualização dos produtos de PCR, o gel de agarose 1% foi corado com *SYBR Safe DNA gel stain* (Life Technologies). A eletroforese foi realizada em 70V por aproximadamente 1 hora. O DNA no gel foi visualizado por meio da exposição à luz UV em um transiluminador. Depois de checados, os produtos de PCR foram purificados com o conjunto de reagentes *GFX PCR and gel band purification kit* (GE Healthcare) seguindo-se as instruções do fabricante. Em seguida, as amostras foram analisados em um NanoDrop ND-1000-V.3 para quantificação do DNA e preparo da reação de sequenciamento.

A reação de sequenciamento foi realizada com auxílio do conjunto de reagentes *ABI PRISM BigDye™ Terminator Cycle Sequencing Ready Reaction Kit* (Life Technologies). O volume final da reação de sequenciamento era de 10µL, composta de 1µL de Big Dye, 10 µM de cada oligonucleotídeo iniciador, tampão 0,75x, até 6,5µL de produto de PCR purificado, para que a concentração fosse de 50ng/µL, e água estéril pra completar o volume final, quando necessário. Para todos os genes, no sequenciamento foram usados os mesmos oligonucleotídeos iniciadores utilizados nas reações de amplificação mencionadas anteriormente (Tabela 4). As condições gerais da reação de sequenciamento consistiram de uma etapa inicial de 1' a 96°C e 39 ciclos de 15" a 96°C, 15" a T°C e 4' a 60°C, em que "T" corresponde às temperaturas de pareamento dos oligonucleotídeos iniciadores para cada gene (Tabela 4).

Para a precipitação foram realizadas várias lavagens do produto da reação de sequenciamento. Primeiramente, foram adicionados 80µL de isopropanol 65% às amostras e estas foram deixadas em repouso por 15 minutos à temperatura ambiente. Em seguida, as amostras foram centrifugadas à 4.000 rpm por 45 minutos à 4°C e o sobrenadante descartado. Posteriormente, cada amostra recebeu 200µL de etanol 70% resfriado, sendo centrifugadas a 4.000 rpm por 10 minutos à 4°C. O sobrenadante foi descartado e esta etapa foi repetida. As amostras foram mantidas na placa aquecedora a 37°C por aproximadamente 1 hora, até a completa evaporação do álcool. As amostras foram enviadas para sequenciamento no Centro de Recursos Biológicos e Biologia Genômica – CREBIO, UNESP/FCAV, Jaboticabal – SP.

3.3 - Análise dos dados

3.3.1 - Leitura e alinhamento das sequências

A leitura e alinhamento das sequências de DNA foram realizados com auxílio dos programas Chromas v.2.01 Copyright© (1998-2005) Technelysium Pty Ltda, ClustalW v. 1.8 (THOMPSON; HIGGINS; GIBSON, 1994), implementado no programa BioEdit v. 7.1.11 (HALL, 1999), e MUSCLE v. 3.8.31 (EDGAR, 2004). Quando necessário, as sequências foram inspecionadas manualmente.

Após o alinhamento, as sequências dos genes nucleares *GstD1* e *αE5* foram analisadas no programa PHASE v.2.1 (STEPHENS; SCHEET, 2005; STEPHENS; SMITH; DONNELLY, 2001), que utiliza uma abordagem bayesiana para reconstrução de haplótipos com sítios heterozigotos. Por meio do SeqPHASE (FLOT, 2010) foi possível converter o alinhamento em formato fasta para o arquivo de entrada do programa PHASE, e converter os resultados obtidos novamente em formato fasta.

3.3.2 - Estatística descritiva

Os valores da estatística descritiva, que inclui o número de sítios polimórficos (S), diversidade haplotípica (Hd), diversidade nucleotídica (π), número médio de diferenças nucleotídicas (k), foram obtidos por meio do programa DnaSP v.5.1 (LIBRADO; ROZAS, 2009), que também foi utilizado para o cálculo do teste Rm, que avalia o número mínimo de eventos de recombinação na amostra. Além do teste Rm, o teste PHI foi realizado para avaliar recombinação (BRUEN; PHILIPPE; BRYANT, 2006), por meio do programa SplitsTree4 v. 4.13.1 (HUSON; BRYANT, 2006). Este teste indica evidência de recombinação quando o valor de $p < 0,05$.

3.3.3 – Modelos de substituição nucleotídica

Os modelos de substituição nucleotídica foram obtidos com auxílio do programa jModelTest v.2.1.2 (POSADA, 2008) e escolhidos com base no *Akaike Information Criterion* (AIC), em que o melhor modelo é indicado pelo menor valor de AIC (POSADA; BUCKLEY, 2004; POSADA; CRANDALL, 1998). Este critério escolhe o modelo que melhor se aproxima do real processo de evolução molecular dos fragmentos (POSADA; BUCKLEY, 2004).

3.3.4 – Estimativa das taxas de substituição nucleotídica

Uma calibração secundária sem grupo externo foi realizada com populações de *D. serido* para estimar as taxas de substituição nucleotídica dos genes estudados no presente trabalho. Oliveira et al. (2012) estimaram, por meio de datação molecular, que a idade do ramo de *D. serido* teria aproximadamente 720 mil anos e, portanto, esta idade foi utilizada para calibração da base do ramo das populações de *D. serido*, utilizando uma distribuição normal (HO, 2007).

As reconstruções filogenéticas e a estimativa das taxas de substituição nucleotídica foram realizadas por meio de abordagem bayesiana com Cadeias de Markov Monte Carlo (MCMC) implementados no pacote de programas BEAST v. 1.7.5 (DRUMMOND et al., 2012). As premissas utilizadas para a árvore e relógio molecular foram de modelo coalescente com tamanho populacional constante (KINGMAN, 1982) e relógio estrito com uma distribuição exponencial, respectivamente. O número de interações na MCMC variou de 30 a 60 milhões com um período de *burn-in* de 10% do total de interações.

Os arquivos gerados foram analisados pelo programa Tracer v. 1.6 (RAMBAUT et al., 2014) para avaliar se os parâmetros atingiram valores de ESS maiores que 200 e para verificar a taxa de substituição nucleotídica estimada.

3.3.5 - Análise de Variância Molecular, Teste de Mantel e Estrutura Populacional

A estruturação entre as populações foi analisada por meio da Análise de Variância Molecular (AMOVA; EXCOFFIER et al., 1992), com auxílio do programa Arlequin v.3.5 (EXCOFFIER; LISCHER, 2010). Esta análise estima a variabilidade intra- e interespecífica de cada espécie, fornecendo valores de Φ , que refletem a correlação da diversidade haplotípica em diferentes níveis hierárquicos: Φ_{st} : covariância entre indivíduos dentro de populações, Φ_{sc} : covariância entre populações dentro de um grupo e Φ_{ct} : covariância entre grupos.

Para avaliar se há correlação entre a distância genética e a distância geográfica foi realizado o teste de Mantel por meio do programa Arlequin v.3.5 (EXCOFFIER; LISCHER, 2010). A matriz de distância geográfica entre as populações foi estabelecida por meio do programa Geographic Distance Matriz Generator v.1.2.3 (ERSTS, 2006). A matriz de distância genética, F_{st} par-a-par, foi calculada pelo programa Arlequin v.3.5 (EXCOFFIER; LISCHER, 2010). O coeficiente de correlação (r) foi considerado significativo quando $p < 0,05$.

A estrutura populacional em *D. serido* também foi investigada por meio de abordagem bayesiana, utilizando o programa Structure v.2.3.4 (PRITCHARD; STEPHENS; DONNELLY, 2000). Um modelo *admixture* foi assumido, usando as localidades amostradas como premissa (LOCPRIOR), e as frequências alélicas foram consideradas correlacionadas. Para todos os genes estudados, duas réplicas foram feitas para cada corrida em cada valor de K, sendo que este variou entre 2 e 8. Cada corrida consistiu de 100 mil interações na MCMC após um *burn-in* de 10 mil interações. O número mais provável de agrupamentos genéticos (K) para cada gene foi definido pelo teste de Evanno (EVANNO; REGNAUT; GOUDET, 2005), implementado no Structure Harvester (EARL; VONHOLDT, 2012). Após este procedimento, para o K escolhido, os resultados das réplicas foram sumarizados por meio do programa CLUMPP v.1.1.2 (JAKOBSSON; ROSENBERG, 2007) e o resultado foi utilizado para gerar os gráficos com as probabilidades de agrupamento em cada grupo.

3.3.6 – Testes de neutralidade

Desvios da neutralidade foram verificados por meio dos testes de neutralidade D de Tajima, F_s de Fu e H de Fay & Wu, sendo que os dois primeiros foram realizados por meio do programa Arlequin v.3.5 (EXCOFFIER; LISCHER, 2010) e o último por meio do programa DnaSP v.5.1 (LIBRADO; ROZAS, 2009).

O teste D de Tajima foi calculado com a diferença entre o número de sítios segregantes e o número médio de diferenças nucleotídicas (TAJIMA, 1989). Este teste segue o modelo de sítios infinitos e considera que a hipótese de evolução neutra explica o polimorfismo encontrado nas amostras quando $D = 0$ ou não significativo. Porém, valores negativos e significativos ($p < 0,05$) sugerem seleção purificadora, efeito garrafa ou expansão populacional, enquanto valores positivos e significativos de D são indicativos de seleção estabilizadora ou contração populacional (TAJIMA, 1993).

O teste F_s de Fu é calculado a partir de uma distribuição de haplótipos (RAMOS-ONSINS; ROZAS, 2002; RAMÍREZ-SORIANO et al., 2008), sendo que desvios da neutralidade em que os valores de F_s são negativos e significativos ($p < 0,02$) podem ser indícios de eventos de expansão populacional (FU, 1997). Além disso, o F_s de Fu também pode ser afetado por eventos de recombinação e, portanto, não é indicado para sequências de DNA que apresentem a mínima chance de recombinação (RAMÍREZ-SORIANO et al., 2008).

O teste H de Fay e Wu baseia-se na associação entre a informação da idade relativa de um variante com dados de frequência, por isso a necessidade de um grupo externo para a classificação da variação genética em derivada e ancestral (TEMPLETON, 2011). Assim, desvios da neutralidade detectados por este teste são indicadores de seleção positiva e/ou efeito carona (FAY; WU, 2000). Para esta análise, foram utilizadas sequências de *D. buzzatii* para os genes *period*, *αE5* e COI, obtidas na base de dados do *NCBI GenBank* (FJ267311.1, DQ204660.1 e KF632603.1, respectivamente); para o gene *kl-5* a sequência foi obtida no presente trabalho e para o gene *GstD1* foi utilizada a sequência de *D. koepferae*, também sequenciada no presente trabalho, pois não foi possível obter sequências deste gene para *D. buzzatii*.

Todos os testes foram realizados para os agrupamentos detectados pela AMOVA (item 3.3.5), uma vez que a subdivisão populacional pode afetar a variância de todos os testes que assumem uma população não diferenciada em equilíbrio (ver CHARLESWORTH, 1998 e STEPHAN et al., 1998 citados em FAY; WU, 2000).

3.3.7 – Testes de seleção

Testes de seleção foram realizados para estimar a razão dN/dS ($=\omega$, em que dN é o número de substituições não-sinônimas e dS é o número de substituições sinônimas), que corresponde a uma medida da seleção natural que está atuando na proteína, uma vez que todos os genes do presente trabalho são codificantes. Esta abordagem permite verificar se as sequências de DNA estão sob seleção positiva, o que pode dar indícios de evolução adaptativa (YANG; WONG; NIELSEN, 2005).

Valores de $\omega < 1$, $= 1$ e > 1 indicam seleção purificadora, evolução neutra e seleção positiva, respectivamente. Os testes foram realizados no programa *codeml* do pacote PAML v. 4.7 (YANG, 2007), usando a abordagem de modelo de sítios, que permite que o ω varie entre os sítios (entre códons ou aminoácidos na proteína) (NIELSEN; YANG, 1998; YANG et al., 2000).

Para esta análise, as sequências de DNA foram alinhadas de modo que o primeiro nucleotídeo correspondesse à primeira posição na trinca de bases, regiões não codificantes foram excluídas, e uma árvore enraizada, para cada gene, foi obtida por meio do programa MrBayes v.3.2.2 (RONQUIST et al., 2012). Para o gene *period* foi utilizado o modelo de substituição nucleotídica GTR +I+G, para os genes COI e *αE5* foi utilizado HKY +I+G e para os genes *GstD1* e *kl-5* foi utilizado HKY +I. Estes modelos de substituição nucleotídica foram escolhidos por serem os que mais se assemelhavam

com os modelos estimados no item 3.3.3. Em todas as simulações foram realizadas 10 milhões de interações, começando com uma árvore aleatória e uma porção de 25%, correspondente ao *burn-in*, foi descartada. Após a análise, as árvores foram exportadas para o formato Newick por meio do programa FigTree v.1.4 (RAMBAUT, 2014). Como grupo externo foram utilizadas as mesmas sequências de *D. buzzatii* e *D. koepferae* apresentadas no item 3.3.5.

No programa *codeml* foram testados seis modelos: 1) M0 (*one ratio*), que estima um ω geral para o conjunto de dados; 2) M1a (*nearly neutral*), que permite duas categorias de sítios de códons, $\omega_0 < 1$ e $\omega_1 = 1$, com proporções p_0 e p_1 ($1 - p_0$); 3) M2a (*selection*), permite uma terceira categoria, comparada com o modelo M1a, $\omega_2 > 1$, com uma proporção de p_2 ($1 - p_0 - p_1$); 4) M3 (*discrete*), classifica os sítios de códons em k classes, ω_0 , ω_1 e ω_2 , com proporções p_0 , p_1 e p_2 ($1 - p_0 - p_1$); 5) M7 (beta), especifica um modelo neutro, em que ω segue uma distribuição beta com os parâmetros estimados p e q da distribuição beta; e 6) M8 (beta, ω), que é similar ao modelo M7, permitindo uma terceira categoria, $\omega_S > 1$ (YANG, 2007).

O *Likelihood Ratio Test* (LRT) foi calculado manualmente por meio da fórmula $2\Delta l = l_1 - l_0$, (em que, l_1 é o *likelihood* do modelo 1, l_0 é o *likelihood* do modelo 0 – hipótese nula – e $2\Delta l$ significa a diferença entre os *likelihoods* dos modelos que estão sendo comparados), entre os modelos neutros (M0, M1a e M7) e de seleção (M3, M2a e M8). Os resultados foram comparados com uma distribuição qui-quadrado, em que os graus de liberdade (gl) correspondem à diferença entre os parâmetros estimados entre os modelos comparados. A comparação M0 vs. M3 tem $gl = 4$, e os valores críticos do qui-quadrado são 9,49; 13,28 e 18,47 com 5%, 1% e 0.1% de significância, respectivamente. Comparações entre M1a vs. M2a e M7 vs. M8 tem $gl = 2$, e os valores críticos do qui-quadrado são 5,99; 9,21 e 13,82 com 5%, 1% e 0.1% de significância, respectivamente. LRT significativo para M0 vs. M3 mostra que há uma variação na pressão seletiva entre os sítios, mas não deve ser considerado um teste confiável para seleção positiva (YANG, 2006), para tanto é necessário LRT significativo entre os modelos M1a vs. M2a e M7 vs. M8. Uma vez detectado que há sítios sob seleção positiva, o *Naïve Empirical Bayes* (NEB) (NIELSEN; YANG, 1998; YANG et al., 2000), realizado no modelo M3, pode mostrar quais códons apresentam evidência estatisticamente significativa para seleção positiva, isso se dá por meio do cálculo da probabilidade a posteriori para as classes de sítios.

3.3.8 - Análises demográficas

Os parâmetros demográficos θ (gene *period*: $\theta = 3\mu N$; genes COI e *kl-5*: $\theta = 2\mu N$; genes *GstD1* e *$\alpha E5$* : $\theta = 4\mu N$; em que μ é o parâmetro de mutação e N é o número de unidades herdáveis na população), g (taxa de crescimento exponencial, em que valores positivos e negativos indicam crescimento ou redução populacional, respectivamente) e M (taxa de migração; $M = m/\mu$, em que m corresponde a possibilidade de uma linhagem imigrar por geração e μ é a taxa de mutação por sítio por geração) foram calculados por meio do programa Lamarc v.2.1.8 (KUHNER, 2006).

Para o gene *period*, duas réplicas foram realizadas considerando para cada uma, duas cadeias quentes, dez cadeias iniciais com 160 mil interações na MCMC e 16 mil interações como *burn-in*, e duas cadeias finais de 1,2 milhão de interações na MCMC e 6 mil interações como *burn-in*.

Para o gene COI, duas réplicas foram realizadas considerando para cada uma, quatro cadeias quentes, dez cadeias iniciais com 400 mil interações na MCMC e 40 mil interações como *burn-in*, e duas cadeias finais de 3,2 milhões de interações na MCMC e 16 mil interações como *burn-in*.

Para o gene *$\alpha E5$* , três réplicas foram realizadas considerando para cada uma, duas cadeias quentes, dez cadeias iniciais com 160 mil interações na MCMC e 16 mil interações como *burn-in*, e duas cadeias finais de 1,2 milhão de interações na MCMC e 6 mil interações como *burn-in*.

Para o gene *GstD1*, quatro réplicas foram realizadas considerando para cada uma, duas cadeias quentes, dez cadeias iniciais com 100 mil interações na MCMC e 10 mil interações como *burn-in*, e duas cadeias finais de 800 mil interações na MCMC e 4 mil interações como *burn-in*.

Para o gene *kl-5*, duas réplicas foram realizadas considerando para cada uma, duas cadeias quentes, dez cadeias iniciais com 200 mil interações na MCMC e 20 mil interações como *burn-in*, e duas cadeias finais de 1,6 milhão de interações na MCMC e 8 mil interações como *burn-in*.

A convergência e estabilidade das corridas foram checadas por meio do programa Tracer v.1.6 (RAMBAUT et al., 2014).

Simulações das histórias demográficas para cada um dos genes estudados, foram realizadas através do pacote BEAST v. 1.7.5 (DRUMMOND et al., 2012), por meio de abordagem bayesiana e MCMC. Como premissas para a árvore e relógio molecular foram assumidos o modelo coalescente *Bayesian Skyline* (Drummond et al., 2005) e o

relógio estrito com as taxas de substituição nucleotídica estimadas conforme item 3.3.4, respectivamente. O número de interações na MCMC variou de 20 milhões a 80 milhões, em que 10% foi descartado como *burn-in*. A convergência das corridas assim como o *Bayesian Skyline Plot* (BSP) foram visualizados por meio do programa Tracer v.1.6 (RAMBAUT et al., 2014).

3.3.9 - Análises filogeográficas

Foi realizada a análise de Agrupamento de Clados Não-Enraizados (*Nested Clade Phylogeographic Analysis* – NCPA) (TEMPLETON et al., 1995), que permite inferências sobre os processos históricos que determinaram a evolução e distribuição geográfica das espécies (PANCHAL, 2007). Por meio do programa TCS v. 1.21 (CLEMENT; POSADA; CRANDALL, 2000), foi construída uma rede haplotípica não-enraizada e o aninhamento dos clados foi realizado manualmente de acordo com a metodologia proposta por Templeton; Boerwinkle; Sing (1987). Com o programa GeoDis v.2.5 (POSADA; CRANDALL; TEMPLETON, 2000), foram calculados os valores de D_c , D_n e $I-T$ para os clados formados na rede haplotípica. Os índices calculados e sua significância foram interpretados por meio da chave de inferência filogeográfica disponível em GeoDis Home Page (Disponível em <<http://darwin.uvigo.es/software/geodis.html>>. Acesso em 17 de fevereiro de 2011).

4 – Resultados

4.1 – Coleta de material biológico

Os dados das coletas realizadas durante este projeto estão apresentados na Tabela 5.

Tabela 5. Resultados das coletas de material biológico realizadas durante o presente trabalho. Os códigos das localidades são os mesmos das Tabelas 1 e 2 e a distribuição geográfica das amostras encontra-se descrita na Figura 1.

Viagem	CL	Táxon	N
I – litoral e interior de Sergipe, litoral sul de Alagoas, litoral norte e interior da Bahia	R27	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	75 271
	R28	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	45 119
	R29	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	342 540
	R30	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	36 44
	R31	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	19 14
	R32	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	205 293
II – litoral e interior do Rio Grande do Norte, litoral e interior do Ceará	R33	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	01 03
	R34	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	01 02
	R35	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	02 03
	R36	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	23 78
	R37	<i>D. serido</i> <i>D. buzzatii</i> <i>D. borborema</i> “cluster” <i>D. buzzatii</i>	119 60 24 219
	R38	<i>D. serido</i> <i>D. buzzatii</i> <i>D. borborema</i> <i>D. mercatorum</i> “cluster” <i>D. buzzatii</i>	273 177 91 01 961
	R39	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	18 38
	R40	<i>D. serido</i> “cluster” <i>D. buzzatii</i>	23 85

Tabela 5. Continuação.

Viagem	CL	Táxon	N
III – região norte da Bahia (limite com os Estados do Piauí e Pernambuco) e região centro-sul do Piauí	R41	<i>D.buzzatii</i>	22
		<i>D. borborema</i>	106
		“cluster” <i>D. buzzatii</i>	416
	R42	<i>D.buzzatii</i>	04
		<i>D. borborema</i>	29
		“cluster” <i>D. buzzatii</i>	61
	R43	<i>D.buzzatii</i>	02
		“cluster” <i>D. buzzatii</i>	12
	R44	<i>D.buzzatii</i>	10
		<i>D. borborema</i>	04
		“cluster” <i>D. buzzatii</i>	23
	R45	“cluster” <i>D. buzzatii</i>	03
	R46	“cluster” <i>D. buzzatii</i>	06
IV – Peruíbe-SP	R47	*	-
	R48	“cluster” <i>D. buzzatii</i>	02
	R49	*	-
	R50	*	-
	R51	<i>D.buzzatii</i>	05
		“cluster” <i>D. buzzatii</i>	07
	R52	*	-
	R53	<i>D.buzzatii</i>	04
		<i>D. mercatorum</i>	95
		“cluster” <i>D. buzzatii</i>	291
V – norte de Minas Gerais	R54	<i>D.buzzatii</i>	01
		“cluster” <i>D. buzzatii</i>	02
	R55	<i>D. seriema</i>	01
		<i>D. mercatorum</i>	01
		<i>D.buzzatii</i>	16
		“cluster” <i>D. buzzatii</i>	19
	R56	<i>D. serido</i>	32
		<i>D. seriema</i>	03
		<i>D. mercatorum</i>	06
		<i>D.buzzatii</i>	08
		“cluster” <i>D. buzzatii</i>	55
	R57	<i>D.buzzatii</i>	12
		“cluster” <i>D. buzzatii</i>	23
	R58	<i>D.buzzatii</i>	67
		“cluster” <i>D. buzzatii</i>	70
	R59	<i>D. serido</i>	01
		<i>D.buzzatii</i>	64
		“cluster” <i>D. buzzatii</i>	70

CL: código das localidades (Tabelas 1 e 2); N: número de indivíduos amostrados.

*indica que não foram coletados indivíduos pertencentes ao gênero *Drosophila*.

4.2 – Gene nuclear *period*

4.2.1 – Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica

Foram sequenciados 423 pb em 246 indivíduos amostrados em 48 localidades. A estatística descritiva mostrou que há 62 haplótipos (h) com diversidade haplotípica (Hd) de 0,947, 41 sítios polimórficos (S), diversidade nucleotídica (π) de 0,01150 e número médio de diferenças nucleotídicas (k) de 4,185. A estatística descritiva para as populações de *D. serido* analisadas por meio de sequências do gene *period* está apresentada na Tabela 6. Os testes Rm (p=0,72) e PHI (p=0,07) não mostraram evidências estatisticamente significativas a favor da presença de eventos de recombinação.

Tabela 6. Resultados obtidos da estatística descritiva para cada população de *Drosophila serido* a partir das sequências do gene nuclear *period*.

Localidade	N	S	h	Hd	π	k
N71	07	11	07	1,00	0,01083	4,095
N70	08	13	06	0,89	0,01151	4,536
N68	06	06	04	0,80	0,00567	2,400
N67	02	10	02	1,00	0,02558	10,000
J92	08	13	08	1,00	0,01245	4,643
J93	07	14	05	0,90	0,01227	5,190
N32	05	20	05	1,00	0,02033	8,600
N37*	01	-	-	-	-	-
N40	04	06	03	0,83	0,00827	3,500
N43	03	03	02	0,68	0,00473	2,000
N42	08	07	06	0,89	0,00549	2,321
N45	04	04	03	0,83	0,00488	2,000
N65	07	10	07	1,00	0,00935	3,905
N35	04	07	04	1,00	0,00827	3,500
N46	06	10	04	0,87	0,01035	3,800
N39*	01	-	-	-	-	-
J95	04	11	03	0,83	0,01340	5,667
N31	05	08	04	0,90	0,00898	3,800
N20	09	11	09	1,00	0,00942	3,833
N22	04	06	03	0,83	0,00794	3,00
N17	06	11	03	0,73	0,01371	5,80
H49	05	05	04	0,90	0,00585	2,40
N2	05	09	03	0,70	0,01229	5,20
Ilh	07	12	06	0,95	0,01328	5,619
N19	07	08	05	0,86	0,00653	2,762
N9	03	02	02	0,68	0,00315	1,333
N10	03	12	03	1,00	0,01891	8,000
N11*	01	-	-	-	-	-

Tabela 6. Continuação.

Localidade	N	S	h	Hd	π	k
N88	08	09	04	0,75	0,00870	3,679
J52	07	15	04	0,71	0,01013	4,286
N56	06	03	02	0,33	0,00236	1,000
AF	05	04	02	0,40	0,00378	1,600
R27	05	08	04	0,90	0,00946	4,000
R28	06	08	05	0,93	0,00788	3,333
R29	06	08	06	1,00	0,00725	3,067
R30	06	10	06	1,00	0,00930	3,933
R31	06	08	05	0,93	0,00867	3,667
R32	05	08	05	1,00	0,00898	3,800
R33*	01	-	-	-	-	-
R34*	01	-	-	-	-	-
R35	02	02	02	1,00	0,00473	2,00
R36	06	06	05	0,93	0,00687	2,667
R37	06	12	06	1,00	0,01214	5,133
R38	07	14	06	0,95	0,01103	4,667
R39	06	11	06	1,00	0,01139	5,200
R40	06	10	06	1,00	0,00851	3,600
R56	10	06	08	0,96	0,00506	2,111
R59*	01	-	-	-	-	-

N = número de indivíduos, S = número de sítios polimórficos, h = número de haplótipos, Hd = diversidade haplotípica, π = diversidade nucleotídica e k = número médio de diferenças nucleotídicas. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

*Não foi possível calcular os parâmetros da estatística descritiva para populações com tamanho amostral menor que dois.

O modelo de substituição nucleotídica que melhor explica o processo de evolução dos fragmentos obtidos para o gene *period* é GTR +I +G (TAVARÉ, 1986) e sua taxa de substituição nucleotídica, estimada a partir de calibração secundária, é de 0,978 substituições/sítio/Myr.

4.2.2 – Análise de Variância Molecular, teste de Mantel e estrutura populacional

O resultado da AMOVA global mostrou que há estruturação populacional entre as amostras de *D. serido*, sendo que o Φ_{st} foi de 0,19 ($p=0,00$) (Tabela 7). A AMOVA hierárquica detectou seis agrupamentos populacionais a partir das sequências do gene *period* analisadas, sendo eles: 1) AF: localidade de Água Fria – GO; 2) N56: localidade Serra do Cipó – MG; 3) Nordeste: incluindo populações do interior e litoral da região Nordeste, e populações do Norte do Estado de Minas Gerais situadas no vale do Rio Jequitinhonha; 4) RJ-ES: populações dos Estados de Rio de Janeiro e Espírito Santo; 5) SP: populações do Estado de São Paulo; e 6) SC: populações do Estado de Santa Catarina (Tabela 7).

Tabela 7. Análise de Variância Molecular (AMOVA) utilizando sequências do gene nuclear *period* obtidas a partir de populações de *Drosophila serido*.

	Fonte de variação	Componentes de Variância	% de variação	Fixation Index
AMOVA global	Entre populações	0,45 (Va)	18,67	$\Phi_{st} = 0,19^*$ ($p=0,00$)
	Dentro das populações	1,96 (Vb)	81,33	
AMOVA hierárquica: 06 grupos	Entre grupos	0,41 (Va)	15,68	$\Phi_{ct} = 0,16^*$ ($p=0,00$)
	Entre populações dentro dos grupos	0,23 (Vb)	8,93	$\Phi_{sc} = 0,11^*$ ($p=0,00$)
	Dentro das populações	1,96 (Vc)	75,38	$\Phi_{st} = 0,25^*$ ($p=0,00$)

*valores estatisticamente significativos ($p<0,05$).

O teste de Mantel não apresentou correlação significativa entre a distância genética e a distância geográfica ($p = 1,00$) para as populações analisadas, ou seja, a distância geográfica não explica a distância genética observada.

O teste de Evanno (EVANNO; REGNAUT; GOUDET, 2005), realizado a partir dos resultados obtidos na análise de estrutura populacional obtida por meio do programa Structure, sugeriu seis agrupamentos genéticos, em concordância com a inferência obtida por meio da AMOVA hierárquica (Tabela 7). Contudo, não há uma boa resolução dos resultados, uma vez que todos os indivíduos apresentam chances similares de estarem classificados em cada um dos seis agrupamentos sugeridos pela análise, como pode ser observado na Figura 7. É possível que este resultado esteja relacionado à elevada variação genética compartilhada entre as populações, como será discutido no item 4.2.6.

Figura 7. Estrutura populacional em *Drosophila serido*, por meio da análise de classificação bayesiana para um modelo populacional com $K = 6$, utilizando sequências do gene nuclear *period*. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

4.2.3 – Testes de neutralidade

O teste de neutralidade D de Tajima (Tabela 8) não apresentou valores significativos que fossem indicativos de desvios da neutralidade em nenhum dos agrupamentos populacionais analisados pela AMOVA hierárquica (Tabela 7). Sendo assim, de acordo com este teste, a hipótese nula de evolução neutra explica os polimorfismos encontrados.

Tabela 8. Testes de neutralidade realizados a partir de sequências do gene nuclear *period* em agrupamentos populacionais de *Drosophila serido*.

Agrupamento populacional	D de Tajima	Fs de Fu	H de Fay & Wu
AF	-1,09 (p=0,13)	-3,58 (p=0,00)*	-0,03 (p=0,02)*
N56	-1,23 (p=0,11)	-6,59 (p=0,00)*	0,08 (p=0,43)
Nordeste	-1,19 (p=0,10)	-25,71 (p=0,00)*	-0,02 (p=0,44)
RJ-ES	-0,82 (p=0,20)	-18,69 (p=0,00)*	0,17 (p=0,41)
SP	0,13 (p=0,62)	-25,41 (p=0,00)*	0,02 (p=0,44)
SC	0,82 (p=0,84)	-19,01 (=0,00)*	-0,19 (p=0,47)

*valores estatisticamente significativos.

Apesar de não ser recomendado para sequências de DNA que apresentam a mínima chance de recombinação (RAMÍREZ-SORIANO et al., 2008), como é o caso do gene *period* (cromossomo X), que pode sofrer recombinação na fêmea, o teste de neutralidade Fs de Fu foi realizado (Tabela 8). Este teste apresentou valores negativos e significativos para todos os agrupamentos analisados, o que pode ser indicativo de eventos de expansão populacional (FU, 1997). Os resultados para estas populações não foram confirmados por meio da BSP (item 4.2.5), devido ao pequeno tamanho amostral.

Com relação ao teste de neutralidade H de Fay & Wu não foram observados valores estatisticamente significativos para a maioria dos agrupamentos e, portanto, a hipótese nula de evolução neutra foi aceita. A única exceção foi observada para o agrupamento AF, em que o valor de H foi negativo e significativo (Tabela 8), sugerindo um evento de varredura seletiva devido ao efeito carona possivelmente gerado por sítios sob seleção positiva. É possível que esta população esteja sob seleção, pois ela está localizada em uma matriz de habitat, um fragmento de FTSS inserido no bioma Cerrado, próximo ao vale do Rio Paranã (WERNECK, 2011).

4.2.4 – Teste de seleção

Os testes de seleção sugerem que a maioria dos sítios das sequências do gene nuclear *period* estão sob seleção purificadora, uma vez que o modelo M3 mostra que 97,8% dos sítios pertencem à classe $\omega = 0,028$ (Tabela 9). Apesar da elevada porcentagem de sítios sob seleção purificadora, o LRT apresentou evidências de pressão seletiva variável entre os sítios e de que alguns deles estariam sob seleção positiva (Tabela 9). Nesta classe de ω estão alocados 2,2% dos sítios, sendo que pelo NEB, obtido por meio do modelo M3, foi possível verificar que apenas uma trinca de bases (posição 130) está sob seleção positiva, com probabilidade a posteriori maior que 99% (Tabela 9, Figura 8). Nenhum sítio foi classificado sob seleção neutra (Tabela 9).

Tabela 9. Estimativa dos parâmetros em cada modelo, valores de ω , valores de *log likelihood* (lnL) e testes estatísticos para os modelos de sítios (LRT – $2\Delta l$), disponíveis no programa PAML, utilizando sequências do gene nuclear *period* obtidas em *Drosophila serido*.

Modelo	Parâmetros	Valores ω	#p	lnL	$2\Delta l$
M0 (one ratio)	-	0,01	1	-4220,57	M0 vs. M3
M3 (discrete)	$\omega_0 = 0,03, \omega_1 = 1,11, \omega_2 = 999,0$ $p_0 = 0,97, p_1 = 0,02, p_2 = 0,01$	5,81	5	-1987,34	4466,46***
M1a (nearly neutral)	$\omega_0 = 0,01, \omega_1 = 1,00$ $p_0 = 0,21, p_1 = 0,79$	0,79	2	-2655,11	M1a vs. M2a
M2a (selection)	$\omega_0 = 0,03, \omega_1 = 1,00, \omega_2 = 999,0$ $p_0 = 0,97, p_1 = 0,02, p_2 = 0,01$	5,80	4	-1987,36	1335,49***
M7 (beta)	$p = 0,10, q = 0,01$	0,90	2	-2691,20	M7 vs. M8
M8 (beta, ω)	$p_0 = 0,99, p = 0,15, q = 2,73$ $p_1 = 0,01, \omega_s = 999,0$	5,82	4	-1990,47	1401,47***

O LRT entre M0 vs. M3 é estatisticamente significativo quando grau de liberdade = 4 e os valores críticos de qui-quadrado são 9,49, 13,28 e 18,47 em 5%*, 1%** e 0,1%*** de significância, respectivamente. O LRT entre M1a vs. M2a e entre M7 vs. M8 é estatisticamente significativo quando grau de liberdade = 2 e os valores críticos de qui-quadrado são 5,99, 9,21 e 13,82 em 5%*, 1%** and 0,1%*** de significância, respectivamente.

Figura 8. Sítios de aminoácidos (trinca de bases) e classes de ω , obtidos a partir do modelo M3 dos testes de seleção, em sequências do gene nuclear *period* para *Drosophila serido*.

4.2.5 – Análises demográficas

Para a estimativa dos parâmetros populacionais realizada por meio do programa Lamarc as populações AF e N56, que a AMOVA hierárquica (Tabela 7) definiu como grupos independentes, tiveram que ser incluídas no agrupamento Nordeste, devido ao pequeno tamanho amostral. Portanto, foram analisados quatro em vez de seis agrupamentos populacionais.

Através da estimativa de θ para os agrupamentos populacionais é possível inferir que o Nordeste e o RJ-ES possuem o maior tamanho populacional com relação à SP e SC, que apresentaram os menores valores (Tabela 10). Com relação a este parâmetro, a taxa de crescimento exponencial (g) (Tabela 10), estimada pelo programa Lamarc, mostrou que as populações dos agrupamentos Nordeste e RJ-ES tiveram crescimento populacional, enquanto que para os demais agrupamentos houve redução populacional, uma vez que os intervalos de confiança incluem zero (Tabela 10). Por meio deste parâmetro é possível inferir que as populações do Nordeste cresceram mais rapidamente que aquelas do agrupamento RJ-ES.

Tabela 10. Valores de theta (θ) e taxas de crescimento exponencial (g) para as populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 7), utilizando sequências do gene nuclear *period*.

Grupo	Theta (θ)	Taxa de crescimento exponencial (g)
Nordeste (AF+N56)	0,12 (0,07; 0,17)	972,70 (553,18; 1000,78)
RJ-ES	0,11 (0,03; 0,41)	885,15 (378,01; 996,76)
SP	0,02 (0,01; 0,05)	386,96 (-282,38; 953,99)
SC	0,01 (0,00; 0,02)	148,88 (-341,70; 967,35)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

As taxas de migração (M) foram convertidas para o parâmetro (M') que mede o balanço entre fluxo gênico e deriva genética como forças que podem promover divergência entre populações/grupos (Tabela 11 e Figura 9). Para M'42, M'13, M'23, M'43, M'32, M'24 os valores foram menores ou iguais a 1,00, evidenciando que a migração não é grande o suficiente para sobrepor a influência da deriva genética. Tais resultados mostram que os eventos de migração foram menos prováveis, até mesmo ausentes. De forma geral, a análise sugere que tenha sido pouco provável a troca de migrantes entre RJ-ES e SC e entre RJ-ES e SP, em ambas as direções, e entre SP e SC, no sentido de que o primeiro doaria migrantes para o segundo.

Tabela 11. Taxas de migração para populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 7), utilizando sequências do gene nuclear *period*.

	Taxa de migração – M	Taxa de migração – M'
M21	97,51 (50,31; 100,24)	11,55 (3,52; 17,18)
M31	19,32 (4,17; 84,45)	2,29 (0,29; 14,48)
M41	79,93 (26,45; 98,85)	9,47 (1,85; 16,95)
M12	95,80 (27,80; 99,74)	10,08 (0,97; 41,33)
M32	9,82 (0,54; 98,23)	1,03 (0,02; 40,71)
M42	6,62 (0,04; 94,30)	0,70 (0,00; 39,08)
M13	47,71 (2,52; 98,01)	0,32 (0,01; 1,76)
M23	56,91 (3,11; 98,22)	0,38 (0,01; 1,76)
M43	92,63 (6,98; 99,24)	0,62 (0,02; 1,78)
M14	90,09 (4,15; 98,66)	1,95 (0,03; 4,55)
M24	56,56 (2,71; 97,91)	1,22 (0,02; 4,51)
M34	95,17 (31,47; 99,94)	2,06 (0,26; 4,61)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

1) Nordeste (incluindo AF e N56); 2) RJ-ES; 3) SC; 4) SP.

M21: taxa de migração de RJ-ES para o Nordeste; M31: taxa de migração de SC para o Nordeste; M41: taxa de migração de SP para o Nordeste; M12: taxa de migração do Nordeste para RJ-ES; M32: taxa de migração de SC para RJ-ES; M42: taxa de migração de SP para RJ-ES; M13: taxa de migração do Nordeste para SC; M23: taxa de migração de RJ-ES para SC; M43: taxa de migração de SP para SC; M14: taxa de migração do Nordeste para SP; M24: taxa de migração de RJ-ES para SP; M34: taxa de migração de SC para SP.

$M=m/\mu$: em que m é a chance de uma linhagem imigrar por geração, e μ é a taxa de mutação por sítio por geração.

$M'=3Nm$: em que N é o número efetivo de migrantes e m é a chance de uma linhagem imigrar por geração.

Figura 9. Taxas de migração estimadas (Tabela 11), por meio de sequências do gene nuclear *period*, entre os agrupamentos populacionais de *Drosophila serido*. Seta contínua: $1 < M' < 5$; Seta tracejada: $M' > 5$; Seta pontilhada: $M' \leq 1$.

M'_{14} , M'_{31} e M'_{34} foram os únicos parâmetros em que a força diversificadora da migração pode se sobrepor à deriva genética, portanto, parte da variação genética pode ter sido gerada por eventos migratórios entre os agrupamentos (Tabela 11 e Figura 9). Nestes casos, os sentidos dos eventos de migração são do Nordeste para SP (M'_{14}), de SC para o Nordeste (M'_{31}) e de SC para SP (M'_{34}).

Para M'_{21} , M'_{41} e M'_{12} os valores foram maiores que 5,00, o que sugere de acordo com o manual do programa Lamarc, que devem ter acontecido tantos eventos de migração que se torna difícil para o programa estimá-los. Os autores sugerem inclusive, que nestes casos, as populações/agrupamentos deveriam ser consideradas com uma unidade. Os sentidos destes eventos de migração são do Nordeste para RJ-ES, e vice-versa, e de SP para Nordeste (Tabela 11 e Figura 9).

Simulações das histórias demográficas por meio da BSP (Figura 10) foram realizadas para todos os agrupamentos observados na AMOVA hierárquica (Tabela 7) para o gene *period*, com exceção de AF e N56, devido ao pequeno tamanho amostral.

O tempo do ancestral comum mais recente (T_{MRCA} – *time of most recent common ancestor*) das populações denominadas Nordeste é de 13,16 kyr (HPD 95%: 5,73 até 22,20 kyr) (kyr – milhares de anos atrás) (Figura 10A), enquanto que para RJ-ES e SP estas idades são mais recentes, 6,90 kyr (HPD 95%: 2,69 até 12,40 kyr) e 11,18 kyr (HPD 95%: 5,11 até 17,80 kyr) (Figura 10B e 9C), respectivamente. Entretanto, para o agrupamento de SC o T_{MRCA} foi o mais antigo de todos, 14,60 kyr (HPD 95%: 7,27 até 23,10 kyr) (Figura 10D).

Figura 10. Simulações de história demográfica por meio da *Bayesian Skyline Plot* (BSP) para as populações de *Drosophila serido* estruturadas em (A) Nordeste; (B) RJ-ES; (C) SP e (D) SC, utilizando sequências do gene nuclear *period*. A curva negra corresponde à variação média na população ao longo do tempo. As curvas cinza correspondem ao HPD 95%. O tamanho populacional médio (N) é medido em unidades de $\log N_e \cdot \tau$ e o tempo em milhões de anos (Myr). A linha tracejada indica o tempo do ancestral comum mais recente da amostra (T_{MRCA}).

As populações do Nordeste apresentaram, no gráfico da BSP, discreto crescimento populacional (Figura 10A), desde o T_{MRCA} até os dias atuais, intensificando-se nos últimos 2,00 kyr. Este resultado concorda com o elevado valor de θ (Tabela 10) e sugere que as populações devem ter passado por períodos de intenso crescimento populacional. O agrupamento SP parece manter-se com tamanho populacional constante na maior parte do tempo (Figura 10B), sendo que um discreto aumento no tamanho populacional é observado nos últimos 1,00 kyr.

A linha do gráfico que descreve a história demográfica das populações do agrupamento RJ-ES (Figura 10B) apresenta inclinação, sugerindo aumento no tamanho populacional, concordando com os parâmetros θ e g (Tabela 10). Já para o agrupamento SC, as populações parecem ter se mantido com tamanho constante a maior parte do tempo, desde o T_{MRCA} até os dias atuais, em que se observa ligeiro crescimento populacional (Figura 10D).

4.2.6 – Análises filogeográficas

A rede de haplótipos não-enraizada (Figura 11), utilizando sequências do gene *period*, para as populações de *D. serido* foi usada para inferências filogeográficas por meio da NCPA. O haplótipo 34 foi considerado pela análise como o mais antigo, com o maior valor de *outgroup weight* (0,115), o que corresponde à probabilidade de ser a raiz ou grupo externo da rede haplotípica (CASTELLOE; TEMPLETON, 1994). Além disso, ele é o mais frequente na amostra, ocupando a posição mais interior na rede. Na Tabela 12 encontra-se a lista de haplótipos compartilhados pelas populações de *D. serido*. Valores significativos nas análises filogeográficas foram observados em oito clados e interpretados por meio da Chave de Inferência Filogeográfica (Tabela 13).

Para o clado 1-14 (Figura 11) foi inferido o evento filogeográfico de fluxo gênico restrito/dispersão, mas com alguma dispersão de longa distância (Tabela 13). Os haplótipos agrupados neste clado possuem ampla distribuição geográfica com populações do Nordeste, da região do vale do Rio Jequitinhonha (MG), Água Fria-GO e litoral paulista.

Expansão de área gradual e passada seguida de fragmentação ou grande expansão de área no passado seguida por extinção dos intermediários foi o evento histórico inferido para o clado 2-5 (Figura 11; Tabela 13). Neste nível hierárquico, juntamente com o

clado 1-14, são incluídas populações da Serra do Cipó-MG, Arraial do Cabo-RJ e Santa Catarina.

Para o clado 2-6 (Figura 11) foi inferido o evento filogeográfico de expansão de área contígua (Tabela 13). O haplótipo mais frequente/antigo deste clado ocorre em Penedo-AL (foz do Rio São Francisco), Salvador-BA, Itaúnas-ES e Macaé-RJ, e a partir destas localidades, eventos de expansão de área contígua podem ter ocorrido ao norte (Junco do Serido-PB, haplótipo 43 e Bom Jesus-RN, haplótipo 45) e ao sul (Paúba-SP, haplótipo 44) de sua distribuição.

Não há sobreposição de área geográfica entre os clados aninhados em 2-10 (Figura 11), portanto foi inferido evento histórico de fragmentação alopátrica (Tabela 13) entre os clados 1-22 (populações de Mucuri-BA, Cabralia-BA, Itaúnas-ES, Penha-SC, Florianópolis-SC) e 1-23 (populações de Milagres-BA, Irecê-BA, Russas-CE, Lajes Pintadas-RN, Arraial do Cabo-RJ) (Figura 11). Contudo, esta inferência filogeográfica deve ser interpretada com cautela, pois há populações amostradas entre os clados que apresentaram fragmentação alopátrica.

O clado 3-2 (Figura 15) envolve diversas populações da região Nordeste, Água Fria-GO, vale do Rio Jequitinhonha e litoral Sudeste da Costa Atlântica, sendo assim, para este clado foi inferido fluxo gênico restrito com isolamento por distância (Tabela 13), devido ao grande número de populações e haplótipos compartilhados.

O evento histórico de expansão de área contígua (Tabela 13) para o clado 3-3 (Figura 11) foi inferido devido ao compartilhamento de haplótipos entre populações do Nordeste e litoral Sudeste e Sul da Costa Atlântica. Neste clado não há representantes das populações de Goiás e Minas Gerais, indicando que este movimento não incluiu colonização de áreas em direção à região central do Brasil como sugerido anteriormente para o clado 2-5, correspondendo a uma rota independente.

Para o clado 4-1 (Figura 11) foi inferido o evento filogeográfico de fluxo gênico restrito com isolamento por distância (Tabela 13), provavelmente devido ao grande número de populações aninhadas neste clado e ao compartilhamento de haplótipos entre elas.

No clado 4-2 (Figura 11) estão agrupadas populações ao longo de toda a distribuição geográfica de *D. serido*, incluindo Nordeste e litoral Atlântico, desta forma foi inferido o evento histórico de expansão de área gradual e passada seguida de fragmentação ou grande expansão de área no passado seguida por extinção dos intermediários (Tabela 13). Tal evento é congruente com os demais eventos filogeográficos inferidos em clados

de menor nível hierárquico, evidenciando que o processo de ocupação da atual área de distribuição geográfica assim como a diferenciação populacional de *D. serido* foi composta por diversos eventos históricos.

Figura 11. Rede haplotípica não-enraizada contendo haplótipos do gene nuclear *period* para *Drosophila serido*. Cada linha da rede representa um passo mutacional. O tamanho dos círculos/quadrados que representam os haplótipos é proporcional ao número de indivíduos portando o haplótipo. Os círculos vazios representam haplótipos hipotéticos, não amostrados. A localidade geográfica de cada haplótipo pode ser visualizada na Tabela 12.

Tabela 12. Lista de haplótipos do gene nuclear *period* compartilhados por populações de *Drosophila serido* em cada localidade geográfica analisada. Os códigos das localidades são os mesmos das Tabelas 1 e 2.

CL	Haplótipos	CL	Haplótipos
AF	1xHap33; 4xHap62	R27	1xHap34; 1xHap54; 1xHap12; 1xHap16
J52	1xHap38; 4xHap41; 1xHap35; 1xHap50	R30	1xHap34; 1xHap07; 1xHap48; 1xHap01; 1xHap30; 1xHap13
N10	1xHap38; 1xHap41; 1xHap25	R31	1xHap34; 1xHap06; 1xHap07; 1xHap61; 2xHap14
Ilha	1xHap38; 2xHap41; 2xHap34; 1xHap42; 1xHap37	R32	1xHap34; 1xHap06; 1xHap54; 1xHap07; 1xHap59
N32	1xHap41; 1xHap50; 1xHap47; 1xHap19; 1xHap09	R39	1xHap34; 1xHap54; 1xHap31; 1xHap61; 1xHap10; 1xHap13
J92	1xHap41; 1xHap34; 1xHap54; 1xHap17; 1xHap07; 1xHap51; 1xHap46; 1xHap03	R40	2xHap34; 1xHap06; 1xHap54; 1xHap17; 1xHap11
J95	2xHap41; 1xHap34; 1xHap40	R56	3xHap34; 1xHap06; 3xHap17; 1xHap07; 1xHap61; 1xHap28
N17	3xHap41; 2xHap34; 1xHap21	N71	1xHap06; 1xHap17; 1xHap07; 1xHap61; 1xHap19; 1xHap02; 1xHap52
N2	3xHap41; 1xHap34; 1xHap32	R28	1xHap06; 1xHap54; 2xHap17; 1xHap07; 1xHap31
N88	1xHap50; 2xHap24; 4xHap25; 1xHap49	R33	1xHap06
N31	2xHap50; 1xHap57; 1xHap48; 1xHap56	R36	1xHap06; 1xHap17; 2xHap07; 1xHap51; 1xHap15
N35	1xHap50; 1xHap17; 1xHap57; 1xHap61	N20	1xHap54; 1xHap07; 1xHap57; 1xHap61; 1xHap58; 1xHap55; 1xHap26; 1xHap51; 1xHap10
N9	2xHap24; 1xHap25	R37	1xHap54; 1xHap07; 1xHap57; 1xHap51; 1xHap10; 1xHap18
N19	1xHap25; 1xHap34; 3xHap32; 1xHap31; 1xHap44	N56	1xHap17; 5xHap27
N11	1xHap25	N37	1xHap17
J93	2xHap34; 2xHap06; 1xHap54; 1xHap36; 1xHap17	R38	1xHap17; 2xHap07; 1xHap05; 1xHap13; 1xHap29; 1xHap45
N46	2xHap34; 1xHap54; 2xHap07; 1xHap08	R59	1xHap17
N68	3xHap34; 2xHap61; 1xHap14	N67	1xHap39; 1xHap07
N40	2xHap34; 1xHap54; 1xHap12	N45	2xHap07; 1xHap61; 1xHap05
N42	3xHap34; 1xHap17; 1xHap07; 1xHap61; 1xHap10; 1xHap02	R34	1xHap07
N70	3xHap34; 1xHap54; 2xHap07; 1xHap43; 1xHap53	R35	1xHap07; 1xHap04
N65	1xHap34; 1xHap07; 1xHap10; 1xHap48; 2xHap15; 1xHap23	N43	1xHap31; 2xHap51
R29	1xHap34; 1xHap17; 1xHap57; 1xHap61; 1xHap05; 1xHap22	N22	2xHap61; 1xHap48; 1xHap14
H49	1xHap34; 1xHap32; 2xHap31; 1xHap20	N39	1xHap60

CL: código da localidade

Tabela 13. Resultados da Chave de Inferência Filogeográfica para os clados aninhados da rede haplotípica (Figura 11), obtida com sequências do gene nuclear *period* para populações de *Drosophila serido*, que apresentaram resultados significativos na análise.

Clado	Etapa na Chave de Inferência Filogeográfica
1-14	1 – 2 – 3 – 5 – 6 – 7: Sim (Fluxo gênico restrito/dispersão mas com alguma dispersão de longa distância)
2-5	1 – 2 – 3 – 5 – 15 – 21: Não (Expansão de área gradual e passada seguida de fragmentação ou grande expansão de área no passado seguida por extinção dos intermediários)
2-6	1 – 19 – 20 – 2 – 11 – 12: Não (Expansão de área contígua)
2-10	1 – 19 – 20 – 2 – 3 – 4 – 9: Não (Fragmentação alopátrica)
3-2	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
3-3	1 – 2 – 11 – 12: Não (Expansão de área contígua)
4-1	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
4-2	1 – 2 – 3 – 5 – 15 – 21: Não (Expansão de área gradual e passada seguida de fragmentação ou grande expansão de área no passado seguida por extinção dos intermediários)

4.3 – Gene nuclear *kl-5*

4.3.1 – Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica

Foram sequenciados 1205 pb em 207 indivíduos amostrados em 45 localidades. A estatística descritiva mostrou que há 05 haplótipos (h) com diversidade haplotípica (Hd) de 0,337, 06 sítios polimórficos (S), diversidade nucleotídica (π) de 0,00107 e número médio de diferenças nucleotídicas (k) de 1,221. A estatística descritiva para as populações de *D. serido* analisadas por meio de sequências do gene *kl-5* está apresentada na Tabela 14.

Tabela 14. Resultados obtidos da estatística descritiva para cada população de *Drosophila serido* a partir das sequências do gene nuclear *kl-5*.

Localidade	N	S	h	Hd	π	K
N71	06	00	01	0,00	0,00000	0,000
N70	06	00	01	0,00	0,00000	0,000
N68	06	00	01	0,00	0,00000	0,000
N67	02	01	02	1,00	0,00083	1,000
J92	06	01	02	0,53	0,00044	0,533
J93	06	00	01	0,00	0,00000	0,000
N32	06	00	01	0,00	0,00000	0,000
N37	06	00	01	0,00	0,00000	0,000
N40	04	00	01	0,00	0,00000	0,000
N43	06	00	01	0,00	0,00000	0,000
N42	06	00	01	0,00	0,00000	0,000

Tabela 14. Continuação

Localidade	N	S	h	Hd	π	K
N45	04	00	01	0,00	0,00000	0,000
N65	05	00	01	0,00	0,00000	0,000
N35	05	00	01	0,00	0,00000	0,000
N46	06	00	01	0,00	0,00000	0,000
J95	02	00	01	0,00	0,00000	0,000
N31	02	00	01	0,00	0,00000	0,000
N20	06	00	01	0,00	0,00000	0,000
N22	03	00	01	0,00	0,00000	0,000
N17	03	00	01	0,00	0,00000	0,000
H49	04	01	02	0,50	0,00041	0,500
N2	03	00	01	0,00	0,00000	0,000
Ilh	06	00	01	0,00	0,00000	0,000
N19	06	00	01	0,00	0,00000	0,000
N9	03	01	02	0,67	0,00055	0,667
N10	03	00	01	0,00	0,00000	0,000
N88	03	00	01	0,00	0,00000	0,000
J52	05	00	01	0,00	0,00000	0,000
N56	06	00	01	0,00	0,00000	0,000
AF	04	00	01	0,00	0,00000	0,000
R27	05	00	01	0,00	0,00000	0,000
R28	05	01	02	0,40	0,00033	0,400
R29	04	00	01	0,00	0,00000	0,000
R30	06	00	01	0,00	0,00000	0,000
R31	06	00	01	0,00	0,00000	0,000
R32	03	00	01	0,00	0,00000	0,000
R33*	01	-	-	-	-	-
R34*	01	-	-	-	-	-
R35	02	00	01	0,00	0,00000	0,000
R36	06	00	01	0,00	0,00000	0,000
R37	06	01	02	0,33	0,00028	0,333
R38	06	00	01	0,00	0,00000	0,000
R39	06	00	01	0,00	0,00000	0,000
R56	09	00	01	0,00	0,00000	0,000
R59*	01	-	-	-	-	-

N = número de indivíduos, S = número de sítios polimórficos, h = número de haplótipos, Hd = diversidade haplotípica, π = diversidade nucleotídica e k = número médio de diferenças nucleotídicas. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

*Não foi possível calcular os parâmetros da estatística descritiva para populações com tamanho amostral menor que dois.

O modelo de substituição nucleotídica que melhor explica o processo de evolução dos fragmentos obtidos para o gene *kl-5* é TPM1uf +I (Kimura, 1981), contudo, na ausência deste modelo dentre as opções disponíveis nos programas utilizados, foi escolhido o modelo HKY +I (HASEGAWA; KISHINO; YANO, 1985). A taxa de substituição nucleotídica para o gene *kl-5*, estimada a partir de calibração secundária, é de 1,004 substituições/sítio/Myr.

4.3.2 – Análise de Variância Molecular, teste de Mantel e estrutura populacional

A AMOVA global realizada com sequências do gene nuclear *kl-5* mostrou que há estruturação genética entre as populações de *D. serido* (Tabela 15), sendo o Φ_{st} de 0,95 ($p=0,00$). A AMOVA hierárquica detectou quatro agrupamentos populacionais, sendo eles: 1) Nordeste: incluindo populações do interior e litoral da região Nordeste e as populações dos Estados de Minas Gerais e Goiás; 2) RJ-ES-N32: populações dos Estados de Rio de Janeiro e Espírito Santo, e inclui a população de Mucuri-BA; 3) SP: populações do Estado de São Paulo; e 4) SC: populações do Estado de Santa Catarina (Tabela 15).

Tabela 15. Análise de Variância Molecular (AMOVA) utilizando sequências do gene nuclear *kl-5* obtidas a partir de populações de *Drosophila serido*.

	Fonte de variação	Componentes de variância	% de variação	Fixation index
AMOVA global	Entre populações	0,59 (Va)	95,16	$\Phi_{st} = 0,95$ ($p=0,00$)
	Dentro das populações	0,03 (Vb)	4,84	
AMOVA hierárquica: Nordeste RJ-ES-N32 SP SC	Entre grupos	1,12 (Va)	89,11	$\Phi_{ct} = 0,89$ ($p=0,00$)
	Entre populações dentro dos grupos	0,11 (Vb)	8,48	$\Phi_{sc} = 0,78$ ($p=0,00$)
	Dentro das populações	0,03 (Vc)	2,40	$\Phi_{st} = 0,98$ ($p=0,00$)

*valores estatisticamente significativos ($p<0,05$).

O teste de Mantel não apresentou correlação significativa entre a distância genética e a distância geográfica ($p=1,00$) para as populações analisadas, ou seja, a distância geográfica não explica a distância genética observada.

Em contraste com o resultado obtido com a AMOVA hierárquica (Tabela 15), o teste de Evanno (EVANNO; REGNAUT; GOUDET, 2005), realizado por meio da análise de estruturação populacional obtida com o programa Structure, sugere três ao invés de quatro agrupamentos (Figura 12). Na Figura 12 podemos observar que as populações de Cabralia-BA, Mucuri-BA, Espírito Santo, Rio de Janeiro e Santa Catarina compõem um

dos agrupamentos sugeridos pelo Structure, representado pela cor cinza médio. Este resultado não concorda com a AMOVA hierárquica (Tabela 15), que sugere que SC seja um agrupamento independente, contudo, pode estar relacionada às premissas teóricas de cada análise estatística e à baixa diversidade genética do marcador.

As demais populações do Nordeste e SP compõem o agrupamento denominado pela cor preto (Figura 12). O agrupamento representado pela cor cinza claro possui todas as populações, contudo, com baixa probabilidade de classificação neste grupo (Figura 12).

Figura 12. Estrutura populacional em *Drosophila serido*, por meio da análise de classificação bayesiana para um modelo populacional com $K = 3$, utilizando sequências do gene nuclear *kl-5*. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

4.3.3 – Testes de neutralidade

O teste de neutralidade D de Tajima apresentou valor negativo e significativo apenas para o agrupamento Nordeste (Tabela 16), sugerindo a possibilidade de seleção purificadora, efeito carona ou expansão populacional (Tajima, 1993). Contudo, a hipótese de efeito carona pode ser rejeitada, uma vez que não foi observado nenhum valor significativo para o teste H de Fay & Wu (Tabela 16) que pudesse corroborar tal hipótese.

Tabela 16. Testes de neutralidade realizados a partir de sequências do gene nuclear *kl-5* em agrupamentos populacionais de *Drosophila serido*.

Agrupamento populacional	D de Tajima	Fs de Fu	H de Fay & Wu
Nordeste	-1,48 (p=0,03)*	>100 (p=0,00)*	0,73 (p=0,31)
RJ-ES-N32	0,00 (p=1,00)	>100 (p=1,00)	0,58 (p=0,34)
SP	-1,16 (p=0,16)	>100 (p=0,00)*	0,86 (p=0,31)
SC	-1,16 (p=0,16)	>100 (p=0,00)*	0,87 (p=0,31)

*valores estatisticamente significativos.

O teste Fs de Fu apresentou valores muito elevados, maiores que 100, possivelmente pelo baixo número de haplótipos na amostra ($h = 5$), e para os agrupamentos Nordeste, SP e SC estes valores foram considerados significativos (Tabela 16), sugerindo eventos de expansão populacionais. Entretanto, este resultado deve ser interpretado com cautela, devido à baixa variação genética observada neste marcador (Tabela 14).

O teste de neutralidade H de Fay & Wu não apresentou valor significativo para nenhum dos agrupamentos analisados (Tabela 16), sugerindo que, de acordo com as premissas teóricas deste teste, as sequências seguem o modelo de evolução neutra.

4.3.4 – Testes de seleção

Os testes de seleção sugerem que a maioria dos sítios das sequências do gene nuclear *kl-5* estão sob seleção purificadora, uma vez que o modelo M3 classificou 95,6% dos sítios na classe $\omega = 0,015$ (Tabela 17), e nenhum sítio foi classificado sob evolução neutra ($\omega = 1,00$ – Tabela 17). Apesar da elevada porcentagem de sítios classificados sob seleção purificadora, o LRT entre os modelos analisados apresentou evidências de

pressão seletiva variável entre os sítios, sugerindo que alguns deles poderiam estar sob seleção positiva (Tabela 17).

Na classe de sítios sob seleção positiva ($\omega = 55,323$ – Tabela 17), o modelo M3 classificou 4,4% dos sítios das sequências e o NEB encontrou evidência estatística para apenas uma trinca de bases, que corresponde à posição 19, com probabilidade a posteriori maior que 99%.

Tabela 17. Estimativa dos parâmetros em cada modelo, valores de ω , valores de *log likelihood* (lnL) e testes estatísticos para os modelos de sítios (LRT – $2\Delta l$), disponíveis no programa PAML, utilizando sequências do gene nuclear *kl-5* obtidas em *Drosophila serido*.

Modelo	Parâmetros	Valores ω	#p	lnL	$2\Delta l$
M0 (one ratio)	-	0,01	1	-2181,77	M0 vs. M3
M3 (discrete)	$\omega_0 = 0,01, \omega_1 = 1,49, \omega_2 = 55,32$ $p_0 = 0,96, p_1 = 0,00, p_2 = 0,04$	2,46	5	-1889,06	585,41***
M1a (nearly neutral)	$\omega_0 = 0,00, \omega_1 = 1,00$ $p_0 = 0,78, p_1 = 0,22$	0,22	2	-2057,93	M1a vs. M2a
M2a (selection)	$\omega_0 = 0,01, \omega_1 = 1,00, \omega_2 = 55,32$ $p_0 = 0,96, p_1 = 0,00, p_2 = 0,04$	2,46	4	-1889,06	337,73***
M7 (beta)	$p = 0,01, q = 0,02$	0,20	2	-2058,20	M7 vs. M8
M8 (beta, ω)	$p_0 = 0,96, p = 1,50, q = 99,00$ $p_1 = 0,04, \omega_s = 55,33$	2,46	4	-1889,07	338,26***

O LRT entre M0 vs. M3 é estatisticamente significativo quando grau de liberdade = 4 e valores críticos de qui-quadrado são 9,49, 13,28 e 18,47 em 5%*, 1%** e 0,1%*** de significância, respectivamente. O LRT entre M1a vs. M2a e entre M7 vs. M8 é estatisticamente significativo quando grau de liberdade = 2 e valores críticos de qui-quadrado são 5,99, 9,21 e 13,82 em 5%*, 1%** and 0,1%*** de significância, respectivamente.

Figura 13. Sítios de aminoácidos (trinca de bases) e classes de ω , obtidos a partir do modelo M3 dos testes de seleção, em sequências do gene nuclear *kl-5* para *Drosophila serido*.

A maioria dos sítios das sequências do gene *kl-5* classificados sob seleção purificadora sugerem que mutações sinônimas são favorecidas em detrimento das mutações não sinônimas, porém, mesmo sob seleção purificadora, a diversidade genética neste marcador é relativamente baixa, uma vez que foram observados 06 sítios polimórficos e 05 haplótipos em 207 indivíduos (item 4.3.1).

4.3.5 – Análises demográficas

Os valores de θ estimados a partir das sequências do gene nuclear *kl-5* para os agrupamentos inferidos por meio da AMOVA hierárquica (Tabela 15) sugerem que o agrupamento de maior tamanho populacional é SP, seguido por Nordeste, SC e RJ-ES-N32 (Tabela 18). Com relação à taxa de crescimento exponencial (Tabela 18) em todos os agrupamentos o intervalo de confiança inclui zero, indicando que as populações passaram por redução no tamanho populacional, fato que também explica os baixos valores de θ obtidos para os agrupamentos (Tabela 18).

Tabela 18. Valores de theta (θ) e taxas de crescimento exponencial (g) para as populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 15), utilizando sequências do gene nuclear *kl-5*.

Grupo	Theta (θ)	Taxa de crescimento exponencial (g)
Nordeste (AF+N56)	0,00026 (0,00005; 0,00068)	863,24 (-464,69; 978,60)
RJ-ES-N32	0,00002 (0,00001; 0,00042)	841,28 (-472,42; 976,32)
SP	0,00030 (0,00002; 0,00127)	869,44 (-463,38; 982,00)
SC	0,00020 (0,00002; 0,00121)	868,80 (-469,43; 980,60)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

As taxas de migração (M) foram convertidas para um parâmetro (M') que mede o balanço entre fluxo gênico e deriva genética como forças que podem promover divergência entre populações/grupos (Tabela 19 e Figura 14). Em todas as combinações possíveis entre as rotas de migração, o parâmetro M' mostrou-se menor que 1, indicando que a migração não é grande o suficiente para sobrepor a influência da deriva genética (Tabela 19 e Figura 14). Tais resultados mostram que os eventos de migração foram menos prováveis, possivelmente ausentes.

Tabela 19. Taxas de migração para populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 15), utilizando sequências do gene nuclear *kl-5*.

	Taxa de migração – M	Taxa de migração – M'
M21	91,84 (3,27; 99,15)	0,02425 (0,06742; 0,00017)
M31	93,08 (6,82; 99,65)	0,02457 (0,06776; 0,00035)
M41	92,45 (5,05; 99,58)	0,02441 (0,06771; 0,00026)
M12	91,35 (2,11; 98,62)	0,00139 (0,04122; 0,00002)
M32	91,99 (3,91; 99,29)	0,00140 (0,04150; 0,00004)
M42	89,58 (2,00; 98,27)	0,00136 (0,00084; 0,00107)
M13	70,53 (1,89; 98,33)	0,01404 (0,00228; 0,00155)
M23	92,75 (5,21; 99,55)	0,01846 (0,00629; 0,00157)
M43	63,02 (1,73; 98,47)	0,01254 (0,00209; 0,00156)
M14	93,26 (7,85; 99,70)	0,02779 (0,00998; 0,00180)
M24	67,52 (1,71; 98,28)	0,02012 (0,00217; 0,00178)
M34	40,49 (1,58; 98,33)	0,01207 (0,00201; 0,00178)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

1) Nordeste (incluindo AF e N56); 2) RJ-ES-N32; 3) SC; 4) SP.

M21: taxa de migração de RJ-ES-N32 para o Nordeste; M31: taxa de migração de SC para o Nordeste; M41: taxa de migração de SP para o Nordeste; M12: taxa de migração do Nordeste para RJ-ES-N32; M32: taxa de migração de SC para RJ-ES-N32; M42: taxa de migração de SP para RJ-ES-N32; M13: taxa de migração do Nordeste para SC; M23: taxa de migração de RJ-ES-N32 para SC; M43: taxa de migração de SP para SC; M14: taxa de migração do Nordeste para SP; M24: taxa de migração de RJ-ES-N32 para SP; M34: taxa de migração de SC para SP.

$M=m/\mu$: em que m é a chance de uma linhagem imigrar por geração, e μ é a taxa de mutação por sítio por geração.

$M'=3Nm$: em que N é o número efetivo de migrantes e m é a chance de uma linhagem imigrar por geração.

Figura 14. Taxas de migração estimadas (Tabela 19), por meio de sequências do gene nuclear *kl-5*, entre os agrupamentos populacionais de *Drosophila serido*. Seta contínua: $1 < M' < 5$; Seta tracejada: $M' > 5$; Seta pontilhada: $M' \leq 1$.

Simulações das histórias demográficas por meio da BSP (Figura 15) foram realizadas para todos os agrupamentos observados na AMOVA hierárquica (Tabela 15) utilizando sequências do gene nuclear *kl-5*, com exceção de RJ-ES-N32, devido à ausência de variação genética entre as sequências.

O T_{MRCA} estimado para o Nordeste foi muito recente, 1.156 anos (HPD 95%: 183,6 até 0,002 anos) e a reconstrução da história demográfica das populações sugere declínio no tamanho populacional com discreto crescimento próximo dos dias atuais (Figura 15A).

O gráfico da BSP gerado para o agrupamento SP sugere estabilidade no tamanho populacional e o T_{MRCA} estimado foi ainda menor que o obtido para o Nordeste, 0,0934 anos (HPD 95%: 0,025 até 2×10^{-7} anos) (Figura 15B). As amostras que representam o agrupamento SC forneceram os resultados/simulações mais parcimoniosos, o T_{MRCA} estimado foi de 14,16 kyr (HPD 95%: 7,16 até 22,20 kyr) e o gráfico da BSP mostra estabilidade no tamanho populacional durante a maior parte do tempo (Figura 15C), com ligeiro crescimento populacional nos últimos 1,00 kyr.

Figura 15. Simulações de história demográfica por meio da *Bayesian Skyline Plot* (BSP) para as populações de *Drosophila serido* estruturadas em (A) Nordeste; (B) SP e (C) SC, utilizando sequências do gene nuclear *kl-5*. A curva negra corresponde à variação média na população ao longo do tempo. As curvas cinza correspondem ao HPD 95%. O tamanho populacional médio (N) é medido em unidades de $\log N_e \cdot \tau$ e o tempo em milhões de anos (Myr). A linha tracejada indica o tempo do ancestral comum mais recente da amostra (T_{MRCA}).

4.3.6 – Análises filogeográficas

A rede de haplótipos não-enraizada, utilizando sequências do gene *kl-5*, para as populações de *D. serido* (Figura 16) foi usada para inferências filogeográficas por meio da NCPA. O haplótipo 01 foi considerado, pela análise, como o mais antigo, com o maior valor de *outgroup weight* (0,414), o que corresponde à probabilidade de ser a raiz ou grupo externo da rede haplotípica (CASTELLOE; TEMPLETON, 1994) e, além disso, ele é o mais frequente na amostra (Figura 16). Na Tabela 20 encontra-se a lista de haplótipos compartilhados pelas populações de *D. serido*. Valores significativos nas

análises filogeográficas foram observados em apenas um clado e interpretados por meio da Chave de Inferência Filogeográfica.

Para o clado 2-1 (Figura 16) (Chave de Inferência: 1 – 19: Não – Fragmentação alopátrica), que corresponde ao aninhamento total da rede haplotípica, foi inferido o evento de fragmentação alopátrica entre os clados 1-1 e 1-2. Estes clados não compartilham localidades geográficas e, além disso, as populações aninhadas nos clados 1-1 e 1-2 correspondem à estruturação sugerida pela análise do Structure (item 4.3.2), sendo o clado 1-1 composto por populações do Nordeste e SP, e o clado 1-2 com populações do sul da Bahia (Mucuri e Cabralia), RJ, ES e SC. Esta divisão populacional observada tanto para a análise do Structure (item 4.3.2) quanto para a NCPA pode ser um indício de que tenha ocorrido processos de dispersão e colonização pelo interior do continente, como sugerido para o gene *period*. Esta hipótese se baseia no fato de que, por exemplo, as populações de SP tem maior similaridade com o Nordeste do que com o restante da costa Atlântica, que são mais próximas geograficamente.

Figura 16. Rede haplotípica não-enraizada contendo haplótipos do gene nuclear *kl-5* para *Drosophila serido*. Cada linha da rede representa um passo mutacional. O tamanho dos círculos/quadrados que representam os haplótipos é proporcional ao número de indivíduos portando o haplótipo. Os círculos vazios representam haplótipos hipotéticos, não amostrados. A localidade geográfica de cada haplótipo pode ser visualizada na Tabela 20.

Tabela 20. Lista de haplótipos do gene nuclear *kl-5* compartilhados por populações de *Drosophila serido* em cada localidade geográfica analisada. Os códigos das localidades são os mesmos das Tabelas 1 e 2.

CL	Haplótipos	CL	Haplótipos
AF	4xHap01	R30	6xHap01
J52	5xHap04	R31	6xHap01
N10	3xHap04	R32	4xHap01
Ilha	6xHap01	R39	6xHap01
N32	6xHap04	R56	9xHap01
J92	4xHap01, 2xHap02	N71	6xHap01
J95	2xHap01	R28	4xHap01, 1xHap02
N29	3xHap01	R33	1xHap01
N2	3xHap01	R36	6xHap01
N88	3xHap04	N20	6xHap04
N31	2xHap04	R37	5xHap01, 1xHap02
N35	5xHap04	N56	6xHap01
N9	2xHap04, 1xHap05	N37	6xHap01
N19	6xHap01	R38	6xHap01
J93	6xHap01	R59	1xHap01
N46	6xHap01	N67	1xHap01, 1xHap03
N68	6xHap01	N45	4xHap01
N40	4xHap01	R34	1xHap01
N42	6xHap01	R35	2xHap01
N70	6xHap01	N43	6xHap01
N65	5xHap01	N22	3xHap04
R29	4xHap01	R27	5xHap01
H49	3xHap01, 1xHap02		

CL: código da localidade

4.4 – Gene nuclear *GstD1*

4.4.1 – Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica

Foram sequenciados 410 pb em 184 indivíduos amostrados em 42 localidades. Após a reconstrução dos haplótipos com sítios heterozigotos, o tamanho da amostra passou de 184 para 368 indivíduos. A estatística descritiva mostrou que há 32 haplótipos (h) com diversidade haplotípica (H_d) de 0,849, 41 sítios polimórficos (S), diversidade nucleotídica (π) de 0,11421 e número médio de diferenças nucleotídicas (k) de 4,683. A estatística descritiva para as populações de *D. serido* analisadas por meio de sequências do gene *GstD1* está apresentada na Tabela 21. O teste Rm ($p=0,1$) não encontrou evidência significativa para um número mínimo de eventos de recombinação na amostra, entretanto, o teste PHI ($p=0,002$) encontrou evidência estatisticamente significativa a favor da presença de eventos de recombinação.

Tabela 21. Resultados obtidos da estatística descritiva para cada população de *Drosophila serido* a partir das sequências do gene nuclear *GstDI*.

Localidade	N	S	h	Hd	π	k
N71	12	04	02	0,54	0,05322	2,182
N70	12	06	04	0,70	0,06135	2,515
N68	12	05	03	0,62	0,05728	2,348
J92	08	08	05	0,86	0,08014	3,286
J93	12	07	04	0,71	0,07206	2,955
N32	12	08	06	0,85	0,08167	3,348
N40	10	06	04	0,73	0,06396	2,622
N43	12	10	05	0,76	0,07761	3,182
N42	12	08	05	0,79	0,07280	2,985
N45	06	09	05	0,93	0,09919	4,067
N65	12	07	05	0,76	0,06541	2,682
N35	10	11	07	0,93	0,11707	4,800
N37	08	07	04	0,75	0,07404	3,036
N46	10	09	05	0,80	0,07859	3,222
J95	06	08	05	0,93	0,09106	3,733
N31	08	11	05	0,86	0,12631	5,179
N20	12	13	06	0,80	0,09978	4,091
N22	06	09	06	1,00	0,10894	4,467
N29	04	07	03	0,83	0,10569	4,333
H49	08	12	05	0,86	0,13850	5,679
N2	08	12	05	0,86	0,13415	5,500
Ilh	12	09	05	0,79	0,10015	4,106
N19	12	13	05	0,76	0,12454	5,106
N88	04	10	04	1,00	0,16260	6,667
J52	02	07	02	1,00	0,17073	7,000
N56	12	06	03	0,62	0,06135	2,515
AF	08	05	03	0,68	0,06185	2,536
R27	10	05	03	0,64	0,05908	2,422
R28	06	05	04	0,87	0,07154	2,933
R29	02	04	02	1,00	0,09756	4,000
R30	12	04	02	0,54	0,05322	2,182
R31	12	05	03	0,62	0,05728	2,348
R32	06	05	03	0,73	0,06667	2,733
R33	02	04	02	1,00	0,09756	4,00
R34	02	05	02	1,00	0,12195	5,00
R35	04	04	02	0,67	0,06504	2,667
R36	08	04	02	0,57	0,05575	2,286
R37	12	08	06	0,81	0,06948	2,848
R38	12	05	03	0,62	0,05728	2,348
R39	12	06	03	0,62	0,06135	2,515
R56	16	10	06	0,82	0,08638	3,542
R59	02	06	02	1,00	0,14634	6,000

N = número de indivíduos, S = número de sítios polimórficos, h = número de haplótipos, Hd = diversidade haplotípica, π = diversidade nucleotídica e k = número médio de diferenças nucleotídicas. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

O modelo de substituição nucleotídica que melhor explica o processo de evolução dos fragmentos obtidos para o gene *GstDI* é TIM2ef +I (manual programa jModelTest),

contudo, na ausência deste modelo dentre as opções dos programas utilizados, foi escolhido o modelo TN93 +I (TAMURA; NEI, 1993). A taxa de substituição nucleotídica para o gene *GstD1*, estimada a partir de calibração secundária, é de 1,071 substituições/sítio/Myr.

4.4.2 – Análise de Variância Molecular, teste de Mantel e estrutura populacional

A AMOVA global utilizando sequências do gene nuclear *GstD1* apresentou evidência para presença de estruturação populacional, sendo que o Φ_{st} foi de 0,34 ($p=0,00$) (Tabela 22). A AMOVA hierárquica detectou três grupos, sendo eles: 1) Nordeste: incluindo populações do interior e litoral da região Nordeste e dos Estados de Minas Gerais e Goiás; 2) RJ-ES: populações dos Estados de Rio de Janeiro e Espírito Santo; e 3) SP-SC: populações dos Estados de São Paulo e Santa Catarina (Tabela 22).

Tabela 22. Análise de Variância Molecular (AMOVA) utilizando sequências do gene nuclear *GstD1* obtidas a partir de populações de *Drosophila serido*.

	Fonte de variação	Componentes de Variância	% de variação	Fixation Index
AMOVA global	Entre populações	0,70 (Va)	33,76	$\Phi_{st} = 0,34^*$
	Dentro das populações	1,37 (Vb)	66,24	($p=0,00$)
AMOVA hierárquica	Entre grupos	1,57 (Va)	50,96	$\Phi_{ct} = 0,51^*$
03 grupos	Entre populações dentro dos grupos	0,14 (Vb)	4,51	$\Phi_{sc} = 0,09^*$
	Dentro das populações	1,37 (Vc)	44,53	($p=0,00$)
				$\Phi_{st} = 0,55^*$
				($p=0,00$)

*valores estatisticamente significativos ($p<0,05$).

O teste de Mantel não apresentou correlação significativa entre a distância genética e a distância geográfica ($p = 1,00$) para as populações analisadas, ou seja, a distância geográfica não explica a distância genética observada.

O teste de Evanno (EVANNO; REGNAUT; GOUDET, 2005), por meio da análise de estrutura populacional realizada com o programa Structure, sugere que existam três agrupamentos populacionais.

No gráfico gerado com as probabilidades de classificação em cada agrupamento (Figura 17) é possível observar que as populações do Nordeste tem grande

probabilidade de serem classificadas no mesmo grupo, definido pela cor cinza médio, com algumas exceções, como é o caso das populações do sul da Bahia (Mucuri e Cabrália), AF e N56, que apresentam baixas probabilidades de classificação (Figura 17).

As populações do sul da Bahia, AF e N56, juntamente com populações com distribuição ao longo do litoral Sudeste e Sul, foram classificadas com probabilidades semelhantes nos agrupamentos definidos pelas cores preto e cinza claro (Figura 17).

Figura 17. Estrutura populacional em *Drosophila serido*, por meio da análise de classificação bayesiana para um modelo populacional com $K = 3$, utilizando sequências do gene nuclear *GstD1*. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

4.4.3 – Testes de neutralidade

Em virtude do teste de recombinação PHI ter detectado evidência de eventos de recombinação (item 4.4.1), o teste de neutralidade F_s de F_u não foi realizado para o conjunto de dados de sequências do gene nuclear *GstDI*, como sugerido por RAMÍREZ-SORIANO et al. (2008).

Os testes de neutralidade D de Tajima e H de Fay & Wu não foram estatisticamente significativos (Tabela 23), para nenhum dos agrupamentos definidos pela AMOVA hierárquica (Tabela 22), portanto, a hipótese nula de evolução neutra é aceita para explicar os polimorfismos encontrados.

Tabela 23. Testes de neutralidade realizados a partir de sequências do gene nuclear *GstDI* em agrupamentos populacionais de *Drosophila serido*.

Agrupamento populacional	D de Tajima	H de Fay & Wu
Nordeste	0,25 (p=0,66)	0,03 (p=0,39)
RJ-ES	0,16 (p=0,62)	0,11 (p=0,37)
SP-SC	1,09 (p=0,88)	0,12 (0,38)

*valores estatisticamente significativos ($p < 0,05$).

4.4.4 – Testes de seleção

Os testes de seleção sugerem que a maior parte dos sítios das sequências do gene nuclear *GstDI* encontram-se sob seleção purificadora, correspondendo a cerca de 99,2% ($\omega = 0,002$), de acordo com o modelo M3 (Tabela 24). Nenhum sítio foi classificado sob evolução neutra e sob seleção positiva foram classificados 0,8% dos sítios ($\omega > 1$) (Tabela 24).

Alguns problemas foram encontrados com relação ao cálculo do LRT (Tabela 24). O valor de $2\Delta l$ para comparação dos modelos M0 e M3 foi de -1951,02, o que parece pouco provável. Problemas com valores de LRT muito menores que zero tem sido relatados no grupo de discussão do programa PAML (<https://groups.google.com/forum/#!forum/pamlsoftware>), contudo, os mediadores do grupo sugerem que isto seja um *bug* no algoritmo, em que, por algum motivo, o algoritmo não consegue atingir convergência. Sendo assim, mesmo após diversos testes, não foi possível obter resultados consistentes neste teste. A comparação entre os modelos M0 e M3 sugere se há ou não pressão seletiva variável entre os sítios, enquanto as demais comparações M1a vs. M2a, e M7 vs. M8, sugerem se há ou não sítios sob

seleção positiva. Nestes casos, apenas o primeiro mostrou-se significativo para tal evidência (Tabela 24).

Mesmo com a inconsistência dos resultados do LRT, o NEB obtido por meio do modelo M3 sugere que uma trinca de base (posição 91) esteja sob seleção positiva, com probabilidade a posteriori maior que 99%. Análises de sequências do gene *GstD1* em 12 espécies de *Drosophila* com genoma sequenciado sugeriram presença de seleção positiva, especialmente no sítio 171, mudança que pode gerar grandes alterações bioquímicas na molécula (LOW et al., 2007).

Tabela 24. Estimativa dos parâmetros em cada modelo, valores de ω , valores de *log likelihood* (lnL) e testes estatísticos para os modelos de sítios (LRT – $2\Delta l$), disponíveis no programa PAML, utilizando sequências do gene nuclear *GstD1* obtidas em *Drosophila serido*.

Modelo	Parâmetros	Valores ω	#p	lnL	$2\Delta l$
M0 (one ratio)	-	1,37	1	-4774,05	M0 vs. M3
M3 (discrete)	$\omega_0 = 0,002, \omega_1 = 1,89, \omega_2 = 999,00$ $p_0 = 0,99, p_1 = 0,01, p_2 = 0,00$	0,17	5	-5749,56	-1951,02
M1a (nearly neutral)	$\omega_0 = 0,001, \omega_1 = 1,00$ $p_0 = 0,60, p_1 = 0,40$	0,40	2	-6183,12	M1a vs. M2a
M2a (selection)	$\omega_0 = 0,002, \omega_1 = 1,00, \omega_2 = 6,29$ $p_0 = 0,97, p_1 = 0,008, p_2 = 0,022$	0,15	4	-5765,40	835.45***
M7 (beta)	$p = 0,01, q = 0,09$	0,10	2	-5770,35	M7 vs. M8
M8 (beta, ω)	$p_0 = 0,99, p = 0,01, q = 0,08$ $p_1 = 0,01, \omega_s = 6,20$	0,15	4	-5768,15	4.40

O LRT entre M0 vs. M3 é estatisticamente significativo quando grau de liberdade = 4 e valores críticos de qui-quadrado são 9,49, 13,28 e 18,47 em 5%*, 1%** e 0,1%*** de significância, respectivamente. O LRT entre M1a vs. M2a e entre M7 vs. M8 é estatisticamente significativo quando grau de liberdade = 2 e valores críticos de qui-quadrado são 5,99, 9,21 e 13,82 em 5%*, 1%** and 0,1%*** de significância, respectivamente.

Figura 18. Sítios de aminoácidos (trinca de bases) e classes de ω , obtidos a partir do modelo M3 dos testes de seleção, em sequências do gene nuclear *GstD1* para *Drosophila serido*.

4.4.5 – Análises demográficas

Os valores de θ estimados por meio do programa Lamarc mostram que o agrupamento RJ-ES possui maior tamanho populacional com relação aos demais, seguido por SP-SC e Nordeste, sendo que este último apresentou o menor tamanho (Tabela 25). Em concordância com estas estimativas, RJ-ES apresentou sinal de crescimento populacional, uma vez que os valores do intervalo de confiança da taxa de crescimento exponencial (g) não incluem zero e este resultado, para os demais agrupamentos, sugere redução no tamanho populacional (Tabela 25).

Tabela 25. Valores de theta (θ) e taxas de crescimento exponencial (g) para as populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 22), utilizando sequências do gene nuclear *GstD1*.

Grupo	Theta (θ)	Taxa de crescimento exponencial (g)
Nordeste	0,007 (0,004; 0,011)	58,70 (-383,11; 947,12)
RJ-ES	0,031 (0,013; 0,094)	472,23 (42,55; 963,46)
SP-SC	0,011 (0,005; 0,025)	98,63 (-473,00; 652,76)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

As taxas de migração (M) foram convertidas para um parâmetro (M') que mede o balanço entre fluxo gênico e deriva genética como forças que podem promover divergência entre populações/grupos (Tabela 26 e Figura 19). O parâmetro M'_{23} , que mede a taxa de migração de RJ-ES para SP-SC, apresentou valor igual a 1, indicando

que migração e mutação possuem a mesma probabilidade de ocorrência, com relação à geração de variação genética (Tabela 26 e Figura 19).

No caso dos parâmetros M'_{21} , M'_{31} e M'_{13} , os valores foram menores que 1 indicando, portanto, que a migração não é grande o suficiente para sobrepor a influência da deriva genética (Tabela 26 e Figura 19). O inverso pode ser observado nos parâmetros M'_{12} e M'_{32} em que os valores foram maiores que 1, evidenciando que a migração é grande suficiente para competir com a deriva genética em termos de geração de variação genética (Tabela 26 e Figura 19).

Tabela 26. Taxas de migração para populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 22), utilizando sequências do gene nuclear *GstD1*.

	Taxa de migração – M	Taxa de migração – M'
M21	96,32 (43,59; 100,18)	0,68 (0,17; 1,08)
M31	6,75 (0,15; 95,03)	0,05 (0,00; 1,02)
M12	94,73 (21,77; 100,14)	2,96 (0,29; 9,45)
M32	62,39 (6,03; 98,42)	1,95 (0,08; 9,29)
M13	86,29 (15,07; 99,20)	0,94 (0,08; 2,51)
M23	94,84 (18,99; 99,95)	1,03 (0,10; 2,53)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

1) Nordeste; 2) RJ-ES; 3) SP-SC.

M21: taxa de migração de RJ-ES para o Nordeste; M31: taxa de migração de SP-SC para o Nordeste; M12: taxa de migração do Nordeste para RJ-ES; M32: taxa de migração de SP-SC para RJ-ES; M13: taxa de migração do Nordeste para SP-SC; M23: taxa de migração de RJ-ES para SP-SC.

$M=m/\mu$: em que m é a chance de uma linhagem imigrar por geração, e μ é a taxa de mutação por sítio por geração.

$M'=3Nm$: em que N é o número efetivo de migrantes e m é a chance de uma linhagem imigrar por geração.

Figura 19. Taxas de migração estimadas (Tabela 26), por meio de sequências do gene nuclear *GstD1*, entre os agrupamentos populacionais de *Drosophila serido*. Seta contínua: $1 < M' < 5$; Seta tracejada: $M' > 5$; Seta pontilhada: $M' \leq 1$.

Simulações das histórias demográficas por meio da BSP foram realizadas para todos os agrupamentos observados na AMOVA hierárquica (Tabela 22) para o gene *GstD1*. Para o agrupamento Nordeste a simulação mostrou discreta redução populacional, desde seu T_{MRCA} , que foi estimado em 7,31 kyr (HPD 95%: 3,45 até 17,30 kyr) (Figura 20A).

Com relação ao agrupamento RJ-ES, a simulação demográfica sugere tamanho populacional constante desde o T_{MRCA} , estimado em 10,51 kyr (HPD 95%: 3,14 até 20,90 kyr), até aproximadamente 3,00 kyr, quando se observa aumento no tamanho populacional (Figura 20B). Já para o agrupamento SP-SC o tamanho populacional permanece constante até próximo dos dias atuais, em que as populações sofrem um declínio e rapidamente se recuperam, apresentando crescimento populacional. O T_{MRCA} para SP-SC foi estimado em 15,40 kyr (HPD 95%: 5,39 até 29,50 kyr) (Figura 20C).

Figura 20. Simulações da história demográfica por meio da *Bayesian Skyline Plot* (BSP) para as populações de *Drosophila serido* estruturadas em (A) Nordeste; (B) RJ-ES e (C) SP-SC, utilizando sequências do gene nuclear *GstD1*. A curva negra corresponde à variação média na população ao longo do tempo. As curvas cinza correspondem ao HPD 95%. O tamanho populacional médio (N) é medido em unidades de $\log N_e \cdot \tau$ e o tempo em milhões de anos (Myr). A linha tracejada indica o tempo do ancestral comum mais recente da amostra (T_{MRCA}).

4.4.6 – Análises filogeográficas

A rede de haplótipos não-enraizada (Figura 21), utilizando sequências do gene nuclear *GstD1*, para as populações de *D. serido* foi usada para inferências filogeográficas por meio da NCPA. O haplótipo 30 foi considerado pela análise como o mais antigo, com o maior valor de *outgroup weight* (0,147), o que corresponde à

probabilidade de ser a raiz ou grupo externo da rede haplotípica (CASTELLOE; TEMPLETON, 1994).

Na Tabela 27 encontra-se a lista de haplótipos compartilhados pelas populações de *D. serido*. Valores significativos nas análises filogeográficas foram observados em sete clados e interpretados por meio da Chave de Inferência Filogeográfica (Tabela 28).

Para o clado 1-11 foi inferido o evento histórico de fragmentação alopátrica (Tabela 28) entre os haplótipos 19 e 21, sendo que o primeiro é compartilhado por populações do Nordeste e vale do Rio Jequitinhonha, e o último é compartilhado pelas populações de Água Fria-GO e Serra do Cipó-MG (Figura 21). Apesar da ausência de diversos passos mutacionais entre os haplótipos 19 e 21, a ausência de haplótipos compartilhados entre as localidades apoiam a hipótese de fragmentação alopátrica.

O evento de expansão de área gradual passada seguida de fragmentação ou grande expansão de área passada com extinção dos intermediários foi inferido para o clado 1-15 (Tabela 28). Neste clado estão aninhados três haplótipos, o de número 28, compartilhado pelas populações Água Fria-GO e Serra do Cipó-MG; o haplótipo 29, compartilhado entre populações do Nordeste e vale do Rio Jequitinhonha; e o haplótipo 30, exclusivo de Arraial do Cabo-RJ (Figura 21).

Para o clado 2-1 foi inferido fluxo gênico restrito com isolamento por distância (Figura 33). Neste clado estão aninhadas as populações descritas acima para o clado 1-11, e no clado 1-5 estão agrupadas populações próximas à Chapada Diamantina. O evento inferido entre estes dois clados possivelmente revela o fluxo gênico entre populações do Nordeste, sendo mais intenso entre populações próximas, reforçando a hipótese de homogeneidade genética sugerida pela AMOVA hierárquica (Tabela 22).

O evento de fluxo gênico restrito com isolamento por distância também foi inferido para o clado 2-4 (Figura 33), contudo neste caso trata-se de populações com distribuição ao longo da costa Atlântica, localizadas no sul da Bahia (Mucuri e Cabrália), Itaúnas-ES, São Paulo (Bertioga, Paúba e Guaratuba) e Florianópolis-SC.

Para o clado 3-1 foi inferido o evento filogeográfico de expansão de área gradual passada seguida de fragmentação ou grande expansão de área passada com extinção dos intermediários (Tabela 28). Neste clado estão agrupadas populações ao longo de toda a área de distribuição geográfica de *D. serido*.

O evento de expansão de área contígua foi inferido para o clado 3-2 (Tabela 28) e, da mesma forma, neste clado estão agrupadas populações ao longo da distribuição geográfica, e de certa forma complementa o que foi inferido para o clado 3-1.

Para o clado 4-1, o maior nível hierárquico da rede haplotípica (Figura 21), não foi possível inferir um evento histórico, o resultado obtido pela chave de inferência foi inconclusivo (Tabela 28), pois não era possível definir a condição ponta ou interior dos clados aninhados (3-1 e 3-2), ambos eram clados de ponta.

Figura 21. Rede haplotípica não-enraizada contendo haplótipos do gene nuclear *GstD1* para *Drosophila serido*. Cada linha da rede representa um passo mutacional. O tamanho dos círculos/quadrados que representam os haplótipos é proporcional ao número de indivíduos portando o haplótipo. Os círculos vazios representam haplótipos hipotéticos, não amostrados. A localidade geográfica de cada haplótipo pode ser visualizada na Tabela 27.

Tabela 27. Lista de haplótipos do nuclear *GstD1* compartilhados por populações de *Drosophila serido* em cada localidade geográfica analisada. Os códigos das localidades são os mesmos das Tabelas 1 e 2.

CL	Haplótipos	CL	Haplótipos
AF	4xHap28, 4xHap21	R27	5xHap19, 5xHap29
J52	2xHap32	R30	6xHap19, 6xHap29
H49	4xHap12, 1xHap23, 1xHap32, 2xHap01	R31	6xHap19, 6xHap29
Ilha	1xHap32, 9xHap01, 1xHap15, 1xHap17	R32	3xHap19, 3xHap29
N32	2xHap26, 4xHap25, 6xHap06	R39	6xHap19, 5xHap29, 1xHap09
J92	1xHap26, 4xHap19, 3xHap29	N22	1xHap32, 1xHap26, 2xHap06, 1xHap27, 1xHap03
J95	3xHap19, 3xHap29	R56	4xHap26, 8xHap19, 3xHap29, 1xHap09
N17	4xHap01	N71	6xHap19, 6xHap29
N2	1xHap12, 2xHap32, 5xHap01	R28	2xHap19, 2xHap29, 1xHap11, 1xHap20
N88	2xHap01, 1xHap03, 1xHap24	R33	1xHap19, 1xHap29
N31	1xHap26, 2xHap25, 2xHap06, 1xHap04, 1xHap13, 1xHap05	R36	4xHap19, 4xHap29
N35	2xHap26, 1xHap25, 4xHap06, 1xHap04, 2xHap14	N20	2xHap32, 2xHap15, 3xHap16, 3xHap18, 1xHap30, 1xHap31
R29	1xHap19, 1xHap29	R37	6xHap19, 6xHap29
N19	1xHap12, 9xHap01, 1xHap02, 1xHap22	N56	6xHap28, 6xHap21
N65	6xHap19, 6xHap29	N37	4xHap19, 3xHap29, 1xHap10
J93	4xHap19, 6xHap29, 2xHap07	R38	6xHap19, 6xHap29
N46	1xHap26, 4xHap19, 4xHap29, 1xHap07	R59	1xHap26, 1xHap19
N68	6xHap19, 6xHap29	N43	5xHap19, 5xHap29, 1xHap07, 1xHap09
N40	5xHap19, 5xHap29	N45	1xHap26, 2xHap19, 1xHap29, 1xHap07, 1xHap11
N42	5xHap19, 6xHap29, 1xHap08	R34	1xHap19, 1xHap29
N70	6xHap19, 6xHap29	R35	2xHap19, 2xHap29

CL: código da localidade.

Tabela 28. Resultados da Chave de Inferência Filogeográfica para os clados aninhados da rede haplotípica (Figura 21), obtida com sequências do gene nuclear *GstD1* para populações de *Drosophila serido*, que apresentaram resultados significativos na análise.

Clado	Etapas na Chave de Inferência Filogeográfica
1-11	1 – 19: Não (Fragmentação alopátrica)
1-15	1 – 19 – 20 – 2 – 3 – 5 – 6 – 13 – 21: Não (Expansão de área gradual passada seguida de fragmentação ou grande expansão de área passada com extinção dos intermediários)
2-1	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
2-4	1 – 19 – 20 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
3-1	1 – 2 – 11 – 12 – 13 – 21: Não (Expansão de área gradual passada seguida de fragmentação ou grande expansão de área passada com extinção dos intermediários)
3-2	1 – 2 – 11 – 12: Não (Expansão de área contígua)
4-1	1 – 2: não é possível definir a condição ponta/interior (Resultado inconclusivo)

4.5 – Gene nuclear *α -esterase5*

4.5.1 – Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica

Foram sequenciados 759 pb em 212 indivíduos amostrados em 44 localidades. Após a reconstrução dos haplótipos com sítios heterozigotos o tamanho da amostra passou de 212 para 424 indivíduos. A estatística descritiva mostrou que há 117 haplótipos (h) com diversidade haplotípica (H_d) de 0,958, 145 sítios polimórficos (S), diversidade nucleotídica (π) de 0,10452 e número médio de diferenças nucleotídicas (k) de 15,155. A estatística descritiva para as populações de *D. serido* analisadas por meio de sequências do gene *$\alpha E5$* está apresentada na Tabela 29. O teste Rm ($p=1,0$) não encontrou evidência significativa para um número mínimo de eventos de recombinação na amostra, entretanto, o teste PHI ($p=0,0001$) encontrou evidência estatisticamente significativa a favor da presença de eventos de recombinação.

Tabela 29. Resultados obtidos da estatística descritiva para cada população de *Drosophila serido* a partir das sequências do gene nuclear α -esterase5.

Localidade	N	S	h	Hd	π	k
N71	10	06	03	0,64	0,01050	1,533
N70	12	32	07	0,83	0,06040	8,818
N68	12	20	07	0,88	0,03829	5,591
N67	02	00	01	0,00	0,00000	0,000
J92	10	17	08	0,93	0,04384	6,400
J93	10	22	05	0,82	0,06408	9,356
N32	12	28	09	0,91	0,05334	7,788
N37	10	14	04	0,73	0,03044	4,444
N40	08	18	05	8,6	0,05626	8,214
N43	12	27	07	0,92	0,05677	8,288
N42	12	28	10	0,97	0,07078	10,333
N45	08	22	04	0,82	0,07290	10,643
N65	12	17	07	0,86	0,03856	5,258
N35	10	11	05	0,82	0,02953	4,311
N46	10	25	07	0,93	0,06956	10,156
J95	08	36	07	0,96	0,07877	11,500
N31	12	16	07	0,77	0,03279	4,788
N20	12	15	07	0,83	0,03809	5,561
N22	10	14	07	0,91	0,04186	6,111
N29	10	11	05	0,80	0,03105	4,533
H49	06	01	02	0,53	0,00365	0,533
N2	12	02	03	0,68	0,00560	0,818
Ilh	12	23	06	0,68	0,06382	9,318
N19	12	09	06	0,85	0,02646	3,864
N9	06	01	02	0,33	0,00228	0,333
N10	06	00	01	0,00	0,00000	0,000
N88	12	02	02	0,17	0,00228	0,333
J52	10	00	01	0,00	0,00000	0,000
N56	12	08	08	0,89	0,01878	2,742
AF	10	01	02	0,36	0,00244	0,356
R27	10	15	05	0,84	0,04612	6,733
R28	10	22	08	0,95	0,04673	6,822
R29	08	13	05	0,89	0,03596	5,250
R30	12	12	08	0,89	0,02252	3,288
R31	10	16	05	0,80	0,04323	6,311
R32	12	14	05	0,67	0,02127	3,106
R33	02	12	02	1,00	0,08219	12,000
R34	02	11	02	1,00	0,07534	11,000
R35	04	04	02	0,50	0,01370	2,000
R36	08	06	03	0,61	0,01663	2,429
R37	12	22	09	0,95	0,06155	8,924
R38	10	14	06	0,89	0,02968	4,333
R39	12	14	06	0,81	0,02698	3,939
R56	12	27	08	0,89	0,05853	8,545

N = número de indivíduos, S = número de sítios polimórficos, h = número de haplótipos, Hd = diversidade haplotípica, π = diversidade nucleotídica e k = número médio de diferenças nucleotídicas. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

O modelo de substituição nucleotídica que melhor explica o processo de evolução dos fragmentos obtidos para o gene *αE5* é TPM2 +I +G (manual programa jModelTest), contudo, na ausência deste modelo dentre as opções dos programas utilizados, foi escolhido o modelo HKY +I +G (HASEGAWA; KISHINO; YANO, 1985). A taxa de substituição nucleotídica para o gene *αE5*, estimada a partir de calibração secundária, é de 1,035 substituições/sítio/Myr.

4.5.2 – Análise de Variância Molecular, teste de Mantel e estrutura populacional

A AMOVA global, realizada a partir de sequências do gene nuclear *αE5*, apresentou evidência de estruturação populacional, com valor de Φ_{st} de 0,67 ($p=0,00$) (Tabela 30). Por meio da AMOVA hierárquica foram detectados três agrupamentos populacionais, sendo eles: 1) Nordeste: incluindo as populações do interior e litoral da região Nordeste, e populações dos Estados de Goiás e Minas Gerais; 2) RJ-ES-N32: populações dos Estados de Rio de Janeiro e Espírito Santo, e a população de Mucuri-BA; e 3) SP-SC: populações dos Estados de São Paulo e Santa Catarina (Tabela 30).

Tabela 30. Análise de Variância Molecular (AMOVA) utilizando sequências do gene nuclear *αE5* obtidas a partir de populações de *Drosophila serido*.

	Fonte de variação	Componentes de Variância	% de variação	Fixation Index
AMOVA global	Entre populações	4,98 (Va)	66,75	$\Phi_{st} = 0,67^*$
	Dentro das populações	2,48 (Vb)	33,25	($p=0,00$)
AMOVA hierárquica: 03 grupos	Entre grupos	7,94 (Va)	68,70	$\Phi_{ct} = 0,69^*$ ($p=0,00$)
	Entre populações dentro dos grupos	1,13 (Vb)	9,82	$\Phi_{sc} = 0,31^*$ ($p=0,00$)
	Dentro das populações	2,48 (Vc)	21,48	$\Phi_{st} = 0,79^*$ ($p=0,00$)

*valores estatisticamente significativos ($p<0,05$).

O teste de Mantel não apresentou correlação significativa entre a distância genética e a distância geográfica ($p = 1,00$) para as populações analisadas, ou seja, a distância geográfica não explica a distância genética observada.

A análise de estruturação populacional realizada através do programa Structure, por meio do teste de Evanno (EVANNO; REGNAUT; GOUDET, 2005), sugeriu um $K = 5$

(Figura 22), maior do que o número de agrupamentos sugerido pela AMOVA hierárquica (Tabela 30).

Em concordância com a AMOVA hierárquica (Tabela 30), foi mantido o agrupamento SP-SC com elevada probabilidade de classificação destas populações no mesmo grupo (Figura 22 - cor cinza médio claro). As populações de Minas Gerais e a população de Goiás foram classificadas em grupos isolados, sendo que as primeiras foram classificadas no mesmo grupo que populações do Espírito Santo, Chapada Diamantina, interior da Bahia, Sergipe, Pernambuco, Paraíba, Rio Grande do Norte e Ceará (Figura 22 – cor preto). Para a população de Água Fria-GO um padrão diferente foi observado, no mesmo grupo estão classificadas as populações do litoral da Bahia, Sergipe, Alagoas, Paraíba e Rio Grande do Norte, além de algumas populações do interior da Paraíba, Rio Grande do Norte, Ceará e Morro da Barrinha-BA (Figura 22 – cor branco).

Figura 22. Estrutura populacional em *Drosophila serido*, por meio de classificação bayesiana para um modelo populacional com $K = 5$, utilizando sequências do gene nuclear *αE5*. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

4.5.3 – Testes de neutralidade

Em virtude do teste de recombinação PHI ter detectado evidência de eventos de recombinação (item 4.5.1) o teste de neutralidade F_s de F_u não foi realizado para o conjunto de dados de sequências do gene nuclear $\alpha E5$, como sugerido por RAMÍREZ-SORIANO et al. (2008).

O teste de neutralidade D de Tajima mostrou-se significativo apenas para o agrupamento SP-SC (Tabela 31), indicativo de seleção purificadora, efeito carona ou expansão populacional. Entretanto, a hipótese de efeito carona pode ser desconsiderada, pois o teste de neutralidade H de Fay & Wu não foi significativo (Tabela 31). Para os demais agrupamentos o D de Tajima não foi significativo, portanto a hipótese nula de evolução neutra é aceita (Tabela 31).

O teste H de Fay & Wu não foi significativo em nenhum dos agrupamentos populacionais avaliados, mostrando que não há evidência de seleção positiva e/ou efeito carona e, portanto, a hipótese nula de evolução neutra também é aceita neste caso (Tabela 31).

Tabela 31. Testes de neutralidade realizados a partir de sequências do gene nuclear $\alpha E5$ em agrupamentos populacionais de *Drosophila serido*.

Agrupamento populacional	D de Tajima	H de Fay & Wu
Nordeste	-1,19 (p=0,09)	0,17 (p=0,44)
RJ-ES-N32	-0,46 (p=0,39)	0,07 (p=0,42)
SP-SC	-1,53 (p=0,03)*	-0,01 (p=0,43)

*valores estatisticamente significativos (p<0,05).

4.5.4 – Testes de seleção

Os testes de seleção sugerem que a maior parte dos sítios das sequências do gene nuclear $\alpha E5$ estão sob seleção purificadora, uma vez que o modelo M3 mostra que 94,0% dos sítios pertencem à classe $\omega = 0,09$ (Tabela 32). Apesar disso, o LRT apresentou evidência de pressão seletiva variável entre os sítios, sugerindo inclusive que alguns deles poderiam estar sob seleção positiva (Tabela 32). Sob seleção positiva, o modelo M3 classificou 6,0% dos sítios, e nenhum sítio foi classificado sob evolução neutra (Tabela 32).

Por meio do NEB, obtido através do modelo M3, foi possível observar que oito trincas de base (posições 07, 62, 94, 105, 126, 168, 184, 196) foram estatisticamente

significativas para presença de seleção positiva, sendo que apenas uma trinca de base (posição 105) mostrou probabilidade a posteriori de 95%, em todas as demais esta probabilidade foi maior que 99%.

Tabela 32. Estimativa dos parâmetros em cada modelo, valores de ω , valores de *log likelihood* (lnL) e testes estatísticos para os modelos de sítios (LRT – 2 Δ l), disponíveis no programa PAML, utilizando sequências do gene nuclear *α E5* obtidas em *Drosophila serido*.

Modelo	Parâmetros	Valores ω	#p	lnL	2 Δ l
M0 (one ratio)	-	0,02	1	-10677,83	M0 vs. M3
M3 (discrete)	$\omega_0 = 0,09, \omega_1 = 3,41, \omega_2 = 999,00$ $p_0 = 0,94, p_1 = 0,04, p_2 = 0,02$	20,63	5	-4359,52	16126,37***
M1a (nearly neutral)	$\omega_0 = 0,01, \omega_1 = 1,00$ $p_0 = 0,15, p_1 = 0,85$	0,85	2	-2629,60	M1a vs. M2a
M2a (selection)	$\omega_0 = 0,05, \omega_1 = 1,00, \omega_2 = 999,00$ $p_0 = 0,90, p_1 = 0,08, p_2 = 0,02$	17,62	4	-2614,65	3459,84***
M7 (beta)	$p = 0,05, q = 0,01$	0,80	2	-4355,90	M7 vs. M8
M8 (beta, ω)	$p_0 = 0,96, p = 0,01, q = 0,02$ $p_1 = 0,04, \omega_S = 449,09$	19,90	4	-2645,45	3420,91***

O LRT entre M0 vs. M3 é estatisticamente significativo quando grau de liberdade = 4 e valores críticos de qui-quadrado são 9,49, 13,28 e 18,47 em 5%*, 1%** e 0,1%*** de significância, respectivamente. O LRT entre M1a vs. M2a e entre M7 vs. M8 é estatisticamente significativo quando grau de liberdade = 2 e valores críticos de qui-quadrado são 5,99, 9,21 e 13,82 em 5%*, 1%** and 0,1%*** de significância, respectivamente.

Figura 23. Sítios de aminoácidos (trinca de bases) e classes de ω , obtidos a partir do modelo M3 dos testes de seleção, em sequências do gene nuclear *α E5* para *Drosophila serido*.

4.5.5 – Análises demográficas

O maior valor de θ foi observado para o agrupamento Nordeste, seguido por RJ-ES-N32 e SP-SC (Tabela 33). Em concordância com estes resultados, a taxa de crescimento exponencial (g) para o Nordeste sugere crescimento populacional, enquanto que para os demais agrupamentos há evidência de redução populacional, pois os intervalos de confiança incluem zero (Tabela 33).

Tabela 33. Valores de theta (θ) e taxas de crescimento exponencial (g) para as populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 30), utilizando sequências do gene nuclear *αE5*.

Grupo	Theta (θ)	Taxa de crescimento exponencial (g)
Nordeste	0,068 (0,052; 0,092)	363,97 (187,17; 680,06)
RJ-ES-N32	0,026 (0,014; 0,046)	392,95 (-125,05; 945,27)
SP-SC	0,004 (0,002; 0,008)	-395,76 (-489,17; 875,71)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

As taxas de migração (M) foram convertidas para um parâmetro (M') que mede o balanço entre fluxo gênico e deriva genética como forças que podem promover divergência entre populações/grupos (Tabela 34 e Figura 24). Para os parâmetros M'_{21} e M'_{32} os valores estimados são maiores que 1, indicativo de que a força diversificadora da migração consegue sobrepor-se à deriva genética (Tabela 34 e Figura 24).

Para os demais parâmetros as estimativas apresentaram valores menores que 1 indicando que a migração não é suficientemente grande para sobrepor-se à força diversificadora da deriva genética, e que provavelmente, esta última é a principal responsável pela geração de variação genética (Tabela 34 e Figura 24).

Tabela 34. Taxas de migração para populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 30), utilizando sequências do gene nuclear $\alpha E5$.

	Taxa de migração – M	Taxa de migração – M'
M21	32,12 (0,71; 76,56)	2,20 (0,04; 7,02)
M31	1,31 (-0,14; 39,78)	0,09 (-0,01; 3,65)
M12	6,47 (0,12; 94,12)	0,17 (0,00; 4,35)
M32	92,84 (9,23; 99,35)	2,45 (0,13; 4,59)
M13	8,54 (0,84; 95,13)	0,03 (0,00; 0,74)
M23	94,09 (17,77; 100,10)	0,37 (0,04; 0,78)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

1) Nordeste; 2) RJ-ES-N32; 3) SP-SC.

M21: taxa de migração de RJ-ES-N32 para o Nordeste; M31: taxa de migração de SP-SC para o Nordeste; M12: taxa de migração do Nordeste para RJ-ES-N32; M32: taxa de migração de SP-SC para RJ-ES-N32; M13: taxa de migração do Nordeste para SP-SC; M23: taxa de migração de RJ-ES-N32 para SP-SC.

$M=m/\mu$: em que m é a chance de uma linhagem imigrar por geração, e μ é a taxa de mutação por sítio por geração.

$M'=3Nm$: em que N é o número efetivo de migrantes e m é a chance de uma linhagem imigrar por geração.

Figura 24. Taxas de migração estimadas (Tabela 34), por meio de sequências do gene nuclear $\alpha E5$, entre os agrupamentos populacionais de *Drosophila serido*. Seta contínua: $1 < M' < 5$; Seta tracejada: $M' > 5$; Seta pontilhada: $M' \leq 1$.

Simulações de histórias demográficas por meio da BSP foram realizadas com os agrupamentos observados na AMOVA hierárquica (Tabela 30) para o gene *αE5*. Dentre os agrupamentos observados, o Nordeste possui o T_{MRCA} mais antigo, 16,71 kyr (HPD 95%: 12,30 até 29,60 kyr), sendo que o gráfico apresenta curva indicativa de crescimento populacional desde o T_{MRCA} até os dias atuais (Figura 25A), em concordância com o cenário sugerido acima pela demais análises demográficas (Tabelas 33 e 34).

Já para o agrupamento RJ-ES-N32 foi observado redução no tamanho populacional constante com rápido crescimento próximo dos dias atuais, seu T_{MRCA} foi estimado em 7,19 kyr (HPD 95%: 2,55 até 12,80 kyr) (Figura 25B). Para o agrupamento SP-SC o T_{MRCA} foi estimado em 11,56 kyr (HPD 95%: 6,56 até 17,30 kyr) e seu gráfico apresenta discreta redução populacional com rápido crescimento próximo dos dias atuais (Figura 25C). Ambos estão em concordância com os resultados obtidos para θ e g (Tabela 33), que evidenciam menor tamanho populacional e possível redução em seu tamanho.

Figura 25. Simulações de história demográfica por meio da *Bayesian Skyline Plot* (BSP) para as populações de *Drosophila serido* estruturadas em (A) Nordeste; (B) RJ-ES-N32 e (C) SP-SC, utilizando sequências do gene nuclear *αE5*. A curva negra corresponde à variação média na população ao longo do tempo. As curvas cinza correspondem ao HPD 95%. O tamanho populacional médio (N) é medido em unidades de $\log N_e \tau$ e o tempo em milhões de anos (Myr). A linha tracejada indica o tempo do ancestral comum mais recente da amostra (T_{MRCA}).

4.5.6 – Análises filogeográficas

As sequências do gene nuclear *αE5* utilizadas para obtenção da rede haplotípica não-enraizada geraram duas redes individuais (Figuras 22 e 23). A primeira delas, denominada neste trabalho como Rede A (Figura 26), possui 32 haplótipos sendo que o haplótipo 32 apresentou o maior valor de *outgroup weight* (0,235), o que corresponde à probabilidade de ser a raiz ou grupo externo desta rede haplotípica (CASTELLOE;

TEMPLETON, 1994). A segunda rede haplotípica obtida possui 85 haplótipos, sendo denominada neste trabalho como Rede B (Figura 27), e o haplótipo que possui o maior valor de *outgroup weight* (0,109) é o 85. Para ambas as redes haplotípicas foi possível realizar o aninhamento hierárquico dos clados e a análise por meio da Chave de Inferência Filogeográfica (Tabelas 37 e 38). A lista de haplótipos compartilhados pelas populações de *D. serido* na Rede A e B encontra-se apresentada nas Tabelas 35 e 36, respectivamente.

Para a Rede A (Figura 26), valores significativos nas análises filogeográficas foram observados em oito clados e interpretados por meio da Chave de Inferência Filogeográfica (Tabela 37).

O evento de fluxo gênico restrito por distância foi inferido para o clado 1-11 (Tabela 37), neste clado estão agrupados haplótipos compartilhados por localidades no sul da Bahia (Mucuri e Cabralia) e Itaúnas-ES. Estas são localidades relativamente próximas e localizadas na costa Atlântica, além disso, a AMOVA hierárquica sugere que as populações do Espírito Santo e Rio de Janeiro formem um agrupamento com Mucuri-BA (Tabela 30) sendo assim, este o evento parece plausível e os resultados para ambas as análises são concordantes.

Para o clado 2-9 foi inferido o evento de fluxo gênico restrito por distância entre os clados 1-16 e 1-17 (Figura 26, Tabela 37). Neste clado estão agrupados haplótipos compartilhados entre populações de SP e SC, concordando com o compartilhamento de diversidade genética sugerido pela AMOVA hierárquica (Tabela 30) para estes grupos.

Para o clado 3-1 foi inferido fluxo gênico restrito com isolamento por distância (Tabela 37), envolvendo populações do sul da Bahia (Mucuri e Cabralia), Espírito Santo, Rio de Janeiro e Ilha de Alcatrazes-SP. Este evento histórico evidencia uma possível e importante conexão desta região da costa Atlântica com uma ilha, que pode estar relacionada à hipótese da presença de cactos ao longo da costa Atlântica conectando São Paulo e Rio de Janeiro.

O evento histórico de expansão de área contígua foi inferido para os clados 3-2 e 3-3 (Tabela 37) evidenciando que a colonização de novas áreas e o fluxo gênico pode ter contribuído para o compartilhamento de haplótipos/variação genética entre as localidades aninhadas nestes clados, que ocorrem desde o sul da Bahia até São Paulo.

Para o clado 3-4, em que estão agrupadas populações de Santa Catarina, São Paulo e Mucuri-BA, foi inferido fluxo gênico restrito por distância (Tabela 37). A disposição dos haplótipos na rede sugere que o movimento tenha sido no sentido sul-norte, uma

vez que o haplótipo mais antigo da amostra ocorre em Santa Catarina e São Paulo, e o haplótipo de Mucuri-BA possui uma posição derivada com relação aos demais (Figura 26). Esta inferência concorda com a direção da rota migratória sugerida pelo parâmetro $M'32$ (Tabela 34).

Os eventos históricos inferidos para os clados de nível 3 sugerem a ocorrência de diversos eventos de expansão/fluxo gênico, que parecem ter sido assincrônicos, uma vez que ocorrem em várias direções.

O evento de expansão de área contígua foi inferido no clado 4-2 (Tabela 37), que agrupa os clados 3-2 e 3-4, reforçando que o sentido da expansão teria sido sul-norte, uma vez que os haplótipos do clado 3-2 são mais derivados e incluem populações do sul da Bahia, Espírito Santo e Rio de Janeiro. Para o clado 5-1 o resultado da Chave de Inferência Filogeográfica foi inconclusivo (Tabela 37), pois não foi possível definir a condição ponta ou interior para os clados aninhados (Figura 26).

Para a Rede B (Figura 27), valores significativos nas análises filogeográficas foram observados em 14 clados e interpretados por meio da Chave de Inferência Filogeográfica (Tabela 38). Contudo, em três clados (1-43, 1-45 e 3-3) o resultado da análise foi inconclusivo.

O evento de fragmentação alopátrica foi inferido para os clados 1-46 e 2-19 (Tabela 38). Dois haplótipos estão agrupados no clado 1-46, 71 e 72, sendo que o primeiro é compartilhado pelas populações com ampla distribuição na região Nordeste, sendo elas: Cachoeira do Ferro Doido-BA, Irecê-BA, Salvador-BA, Riacho-BA, Costa Azul-BA, Penedo-AL, Lajes Pintadas-RN, Almenara-MG. Enquanto que o haplótipo 72 apresenta distribuição geográfica mais restrita, ocorrendo apenas na população Morro da Barrinha-BA, interior da Bahia na divisa com Pernambuco. Para o clado 2-19 o evento histórico inferido sugere fragmentação alopátrica entre populações do interior da Bahia e litoral da Bahia e Sergipe.

O evento de fluxo gênico restrito por distância foi inferido para cinco clados (1-49, 2-18, 2-26, 4-1 e 4-4) (Tabela 38) na rede haplotípica B (Figura 27). Este evento parece plausível, uma vez que os clados incluem populações da região Nordeste, interior e litoral, e do vale do Rio Jequitinhonha e, portanto, o fluxo gênico entre localidades geograficamente próximas parece mais provável do que entre localidades geograficamente mais distantes.

Para os clados 3-6, 3-8, 4-2 e 5-1 foi inferido o evento filogeográfico de expansão de área contígua (Tabela 38). No clado 3-6 encontram-se aninhadas apenas populações de

Minas Gerais (Serra do Cipó e vale do Rio Jequitinhonha) e Goiás (Água Fria) e este evento auxilia no entendimento da forma como ocorreu a ocupação desta área. No clado 3-8 estão agrupadas populações desde o vale do Rio Jequitinhonha, passando pela Chapada Diamantina até a Paraíba, próximo à Chapada da Borborema, sugerindo que o sentido da expansão tenha sido sul-norte. No clado 4-2, além do clado 3-6, estão aninhados os clados 3-7 e 3-10, sendo que estes últimos incluem populações do Nordeste (interior e litoral, incluindo a Chapada Diamantina) e Ilha de Alcatrazes-SP. A partir da observação deste clado, é possível inferir que a expansão discutida para o clado 3-6 tenha se iniciado com populações da Chapada Diamantina, confirmando a hipótese de que este pode ser o centro de dispersão para a espécie (FRANCO; MANFRIN, 2013).

Para o clado 5-1, que envolve todos os clados da rede haplotípica B (Figura 27), foi inferido expansão de área contígua (Tabela 38) e nesta rede estão todas as populações do Nordeste e vale do Rio Jequitinhonha.

Figura 26. Rede haplotípica não-enraizada, Rede A, contendo haplótipos do gene nuclear $\alpha E5$ para *Drosophila serido*. Cada linha da rede representa um passo mutacional. O tamanho dos círculos/quadrados que representam os haplótipos é proporcional ao número de indivíduos portando o haplótipo. Os círculos vazios representam haplótipos hipotéticos, não amostrados. A localidade geográfica de cada haplótipo pode ser visualizada na Tabela 35.

Figura 27. Rede haplotípica não-enraizada, Rede B, contendo haplótipos do gene nuclear $\alpha E5$ para *Drosophila serido*. Cada linha da rede representa um passo mutacional. O tamanho dos círculos/quadrados que representam os haplótipos é proporcional ao número de indivíduos portando o haplótipo. Os círculos vazios representam haplótipos hipotéticos, não amostrados. A localidade geográfica de cada haplótipo pode ser visualizada na Tabela 36.

Tabela 35. Lista de haplótipos da Rede A do gene nuclear $\alpha E5$ compartilhados por populações de *Drosophila serido* em cada localidade geográfica analisada. Os códigos das localidades são os mesmos das Tabelas 1 e 2.

CL	Haplótipos
J52	10xHap32
H49	2xHap27, 4xHap32
Ilha	7xHap27, 1xHap32, 1xHap03
N32	1xHap06, 1xHap24, 1xHap12, 1xHap18, 4xHap19, 1xHap11, 1xHap14, 1xHap13
N17	1xHap27, 3xHap32, 4xHap18, 1xHap28, 1xHap31
N2	3xHap27, 6xHap32, 3xHap28
N88	11xHap32, 1xHap29
N31	7xHap12, 1xHap09, 1xHap04, 1xHap05, 2xHap20
N35	4xHap12, 2xHap14, 1xHap08, 2xHap20, 1xHap21
R29	
N19	1xHap27, 3xHap32, 4xHap18, 1xHap26, 3xHap25
N10	6xHap32
N9	5xHap32, 1xHap30
N22	2xHap12, 3xHap18, 1xHap02, 1xHap01, 1xHap22, 1xHap23, 1xHap15
N20	5xHap18, 1xHap16, 1xHap17, 1xHap08, 1xHap09, 2xHap10, 1xHap07

CL: código da localidade.

Tabela 36. Lista de haplótipos da Rede B do gene nuclear *αE5* compartilhados por populações de *Drosophila serido* em cada localidade geográfica analisada. Os códigos das localidades são os mesmos das Tabelas 1 e 2.

CL	Haplótipos	CL	Haplótipos
AF	2xHap33, 8xHap35	R27	3xHap25, 3xHap85, 2xHap67, 1xHap71, 1xHap16
Ilha	2xHap42, 1xHap43	R30	5xHap85, 1xHap41, 2xHap67, 1xHap71, 1xHap75, 1xHap64, 1xHap83
N32	1xHap44	R31	4xHap85, 1xHap62, 1xHap45, 4xHap49
J92	1xHap25, 1xHap13, 1xHap68, 3xHap85, 1xHap62, 1xHap41, 1xHap40, 1xHap67	R32	7xHap85, 1xHap62, 2xHap71, 1xHap49, 1xHap75
J95	1xHap85, 2xHap62, 1xHap73, 1xHap18, 1xHap12, 1xHap54, 1xHap53	R39	2xHap25, 5xHap85, 1xHap79, 2xHap84, 2xHap80
R29	2xHap85, 1xHap53, 2xHap39, 1xHap57, 2xHap74	R38	2xHap85, 3xHap62, 1xHap67, 1xHap61, 2xHap70, 1xHap14
N67	2xHap81	R56	1xHap85, 4xHap71, 2xHap45, 2xHap37, 1xHap52, 1xHap32, 1xHap31
N65	4xHap85, 1xHap62, 1xHap71, 1xHap79, 3xHap49, 1xHap76, 1xHap56	N71	5xHap85, 1xHap67, 4xHap79
J93	4xHap25, 2xHap02, 2xHap50, 1xHap58, 1xHap46	R28	1xHap85, 1xHap67, 2xHap73, 1xHap23, 1xHap78, 1xHap03, 1xHap38, 2xHap82
N46	1xHap18, 2xHap20, 2xHap06, 1xHap69, 1xHap11, 1xHap24, 2xHap63	R33	1xHap85, 1xHap51
N68	4xHap62, 1xHap67, 1xHap23, 2xHap27, 1xHap48, 1xHap75, 2xHap78	R36	1xHap85, 2xHap62, 5xHap67
N40	3xHap25, 2xHap67, 1xHap05, 1xHap55, 1xHap19	R37	2xHap25, 1xHap85, 2xHap67, 1xHap71, 1xHap16, 2xHap26, 1xHap22, 1xHap7, 1xHap04
N42	2xHap25, 1xHap62, 1xHap67, 1xHap10, 1xHap09, 2xHap60, 2xHap01, 1xHap71, 1xHap61	N56	2xHap35, 3xHap28, 3xHap37, 1xHap36, 1xHap29, 1xHap34, 1xHap30
N70	2xHap25, 5xHap85, 1xHap66, 1xHap17, 1xHap47, 1xHap21, 1xHap16	N37	1xHap85, 4xHap62, 4xHap72, 1xHap23
R34	1xHap62, 1xHap51	N43	2xHap67, 1xHap10, 2xHap71, 2xHap59, 2xHap62, 1xHap15, 2xHap45
R35	3xHap85, 1xHap62	N45	2xHap85, 1xHap10, 2xHap45, 1xHap08, 2xHap07

CL: código da localidade.

Tabela 37. Resultados da Chave de Inferência Filogeográfica para os clados aninhados da rede haplotípica A (Figura 26), obtida com sequências do gene nuclear *αE5* para populações de *Drosophila serido*, que apresentaram resultados significativos na análise.

Clado	Etapas na Chave de Inferência Filogeográfica
1-11	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
2-9	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
3-1	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
3-2	1 – 2 – 11 – 12: Não (Expansão de área contígua)
3-3	1 – 2 – 11 – 12: Não (Expansão de área contígua)
3-4	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
4-2	1 – 2 – 11 – 12: Não (Expansão de área contígua)
5-1	1 – 2: não é possível definir a condição ponta/interior (Resultado inconclusivo)

Tabela 38. Resultados da Chave de Inferência Filogeográfica para os clados aninhados da rede haplotípica B (Figura 27), obtida com sequências do gene nuclear *αE5* para populações de *Drosophila serido*, que apresentaram resultados significativos na análise.

Clado	Etapas na Chave de Inferência Filogeográfica
1-43	1 – 2 – 11 – 17: Não (Resultado inconclusivo)
1-45	1 – 2 – 11 – 17: Não (Resultado inconclusivo)
1-46	1 – 19: Não (Fragmentação alopátrica)
1-49	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
2-18	1 – 19 – 20 – 2 – 11 – 17 – 4: Não (Fluxo gênico restrito com isolamento por distância)
2-19	1 – 19: Não (Fragmentação alopátrica)
2-26	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
3-3	1 – 19 – 20 – 2 – 11 – 17: Não (Resultado inconclusivo)
3-6	1 – 2 – 11 – 12: Não (Expansão de área contígua)
3-8	1 – 2 – 11 – 12: Não (Expansão de área contígua)
4-1	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
4-2	1 – 19 – 20 – 2 – 11 – 12: Não (Expansão de área contígua)
4-4	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
5-1	1 – 2 – 11 – 12: Não (Expansão de área contígua)

4.6 – Gene mitochondrial COI

4.6.1 – Estatística descritiva, modelo de substituição nucleotídica e taxa de substituição nucleotídica

Foram sequenciados 629 pb em 240 indivíduos amostrados em 42 localidades. A estatística descritiva mostrou que há 60 haplótipos (h) com diversidade haplotípica (Hd) de 0,771, 60 sítios polimórficos (S), diversidade nucleotídica (π) de 0,00708 e número médio de diferenças nucleotídicas (k) de 4,154. A estatística descritiva para as populações de *D. serido* analisadas por meio de sequências do gene COI está apresentada na Tabela 39.

Tabela 39. Resultados obtidos da estatística descritiva para cada população de *Drosophila serido* a partir das sequências do gene mitocondrial COI.

Localidade	N	S	h	Hd	π	k
N71	07	04	05	0,90	0,00212	1,333
N70	10	07	08	0,95	0,00272	1,711
N68	04	02	03	0,83	0,00185	1,167
N67	02	01	02	1,00	0,00159	1,000
J92	06	02	03	0,73	0,00148	0,933
N32	05	09	02	0,60	0,00859	5,400
N37	10	18	07	0,93	0,01035	6,511
N40	05	03	03	0,80	0,00223	1,400
N43	09	19	07	0,92	0,01157	7,278
N42	10	01	02	0,20	0,00032	0,200
N45	05	00	01	0,00	0,00000	0,000
N65	10	11	06	0,87	0,00392	2,467
N35	05	09	03	0,70	0,00572	3,600
N46	06	03	03	0,60	0,00191	1,200
N39*	01	-	-	-	-	-
J95	04	00	01	0,00	0,00000	0,000
N31	06	11	04	0,80	0,00880	5,533
N20	06	00	01	0,00	0,00000	0,000
N22	05	02	03	0,70	0,00127	0,800
N29	07	01	02	0,29	0,00045	0,286
H49	04	01	02	0,67	0,00106	0,667
N2	06	01	02	0,53	0,00085	0,533
N88	07	23	04	0,81	0,02142	12,571
J52	07	00	01	0,00	0,00000	0,000
N56	08	09	04	0,64	0,00443	2,786
AF	06	00	01	0,00	0,00000	0,000
R27	04	01	02	0,67	0,00106	0,667
R28	06	02	03	0,73	0,00138	0,867
R29	05	05	05	1,00	0,00318	2,000
R30	06	03	04	0,80	0,00159	1,000
R31	06	03	04	0,80	0,00159	1,000
R32	06	05	04	0,80	0,00265	1,667
R33*	01	-	-	-	-	-
R34*	01	-	-	-	-	-
R35	02	01	02	1,00	0,00159	1,000
R36	06	03	04	0,87	0,00191	1,200
R37	08	03	04	0,64	0,00119	0,750
R38	06	01	02	0,33	0,00053	0,333
R39	05	02	03	0,70	0,00159	1,000
R40	06	01	02	0,60	0,00100	0,600
R56	10	02	03	0,51	0,00088	0,556
R59*	01	-	-	-	-	-

N = número de indivíduos, S = número de sítios polimórficos, h = número de haplótipos, Hd = diversidade haplotípica, π = diversidade nucleotídica e k = número médio de diferenças nucleotídicas. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

*Não foi possível calcular os parâmetros da estatística descritiva para populações com tamanho amostral menor que dois.

O modelo de substituição nucleotídica que melhor explica o processo de evolução dos fragmentos obtidos para o gene COI é TPM3uf +I +G (manual programa jModelTest), contudo, na ausência deste modelo, dentre as opções dos programas utilizados, foi escolhido o modelo HKY +I +G (HASEGAWA; KISHINO; YANO, 1985). A taxa de substituição nucleotídica para o gene COI, estimada a partir de calibração secundária, é de 1,084 substituições/sítio/Myr.

4.6.2 – Análise de Variância Molecular, teste de Mantel e estrutura populacional

O resultado da AMOVA global mostrou estruturação populacional entre as amostras de *D. serido*, sendo que o Φ_{st} foi de 0,54 ($p=0,00$) (Tabela 40). A AMOVA hierárquica detectou cinco agrupamentos populacionais para as sequências do gene mitocondrial COI analisadas, sendo eles: 1) AF: localidade de Água Fria-GO; 2) N56: localidade Serra do Cipó-MG; 3) Nordeste: incluindo populações do interior e litoral da região Nordeste, e populações do Norte do Estado de Minas Gerais situadas no vale do Rio Jequitinhonha; 4) SP-RJ-ES: populações dos Estados de São Paulo, Rio de Janeiro e Espírito Santo; e 5) SC: populações do Estado de Santa Catarina (Tabela 40).

Tabela 40. Análise de Variância Molecular (AMOVA) utilizando sequências do gene mitocondrial COI obtidas a partir de populações de *Drosophila serido*.

	Fonte de variação	Componentes de Variância	% de variação	Fixation Index
AMOVA global	Entre populações	1,24 (Va)	54,05	$\Phi_{st} = 0,54^*$
	Dentro das populações	1,05 (Vb)	45,95	($p=0,00$)
AMOVA hierárquica: 05 grupos	Entre grupos	2,42 (Va)	66,12	$\Phi_{ct} = 0,66^*$ ($p=0,00$)
	Entre populações dentro dos grupos	0,18 (Vb)	5,04	$\Phi_{sc} = 0,15^*$ ($p=0,00$)
	Dentro das populações	1,05 (Vc)	28,84	$\Phi_{st} = 0,71^*$ ($p=0,00$)

*valores estatisticamente significativos ($p<0,05$).

O teste de Mantel não apresentou correlação significativa entre a distância genética e a distância geográfica ($p = 1,00$) para as populações analisadas, ou seja, a distância geográfica não explica a distância genética observada.

Com relação à análise de estruturação populacional realizada pelo programa Structure foram inferidos quatro grupos populacionais ($K = 4$), resultado relativamente próximo ao encontrado com a AMOVA hierárquica (Tabela 40). As populações do Nordeste e SC foram classificadas no grupo definido pela cor preta, com elevada probabilidade (Figura 28). Entretanto, a população de Mucuri-BA (N32) apresentou probabilidades semelhantes de classificação tanto no Nordeste quanto no agrupamento que inclui populações do litoral Sudeste, denominadas pela cor cinza médio (Figura 28).

A população de Serra do Cipó-MG (N56) foi classificada em um terceiro grupo, denominado pela cor cinza claro, e a população Água Fria-GO (AF) apresentou probabilidades semelhantes de classificação nos grupos representados pelas cores cinza escuro e cinza claro (Figura 28).

Figura 28. Estrutura populacional em *Drosophila serido*, por meio da análise de classificação bayesiana para um modelo populacional com $K = 4$, utilizando sequências do gene mitocondrial COI. Os códigos das localidades são os mesmos apresentados nas Tabelas 1 e 2.

4.6.3 – Testes de neutralidade

O teste de neutralidade D de Tajima mostrou-se negativo e significativo apenas para o agrupamento Nordeste (Tabela 41), o que sugere seleção purificadora, efeito carona ou expansão populacional. Entretanto, podemos desconsiderar a hipótese de efeito carona, pois o teste de neutralidade H de Fay & Wu não foi significativo (Tabela 41), para nenhum dos agrupamentos populacionais avaliados.

Com exceção do agrupamento AF, para todos os demais o teste de neutralidade Fs de Fu foi negativo e significativo sugerindo evento de expansão populacional (Tabela 41).

Tabela 41. Testes de neutralidade realizados a partir de sequências do gene mitocondrial COI em agrupamentos populacionais de *Drosophila serido*.

Agrupamento populacional	D de Tajima	Fs de Fu	H de Fay & Wu
AF	0,00 (p=1,00)	>100 (p=1,00)	0,66 (p=0,33)
N56	-0,97 (p=0,21)	-5,90 (p=0,00)*	0,58 (p=0,34)
Nordeste	-2,29 (p=0,00)*	-27,24 (p=0,00)*	0,46 (p=0,34)
SP-RJ-ES	-1,48 (p=0,05)	-27,61 (p=0,00)*	0,60 (p=0,33)
SC	1,12 (p=0,89)	-6,62 (p=0,00)*	0,41 (p=0,36)

*valores estatisticamente significativos.

4.6.4 – Testes de seleção

Os testes de seleção aplicados em sequências do gene mitocondrial COI mostraram, por meio do LRT, evidência de pressão seletiva variável entre os sítios e que alguns deles poderiam estar sob seleção positiva (Tabela 42). Contudo, o NEB obtido a partir do modelo M3 não encontrou nenhuma trinca de base com evidência estatisticamente significativa para seleção positiva. Em concordância, neste modelo 99% dos sítios foram classificados sob seleção purificadora ($\omega = 0,012$), nenhum sob evolução neutra e apenas 1% sob seleção positiva ($\omega = 999,00$) (Tabela 42).

Tabela 42. Estimativa dos parâmetros de cada modelo, valores de ω , valores de *log likelihood* (lnL) e testes estatísticos para os modelos de sítios (LRT – $2\Delta l$), disponíveis no programa PAML, utilizando sequências do gene mitocondrial COI obtidas em *Drosophila serido*.

Modelo	Parâmetros	Valores ω	#p	lnL	$2\Delta l$
M0 (one ratio)	-	0,00	1	-4374,84	M0 vs. M3
M3 (discrete)	$\omega_0 = 0,01, \omega_1 = 60,88, \omega_2 = 999,00$ $p_0 = 0,99, p_1 = 0,00, p_2 = 0,01$	5,12	5	-1689,97	5369,73***
M1a (nearly neutral)	$\omega_0 = 0,00, \omega_1 = 1,00$ $p_0 = 0,26, p_1 = 0,74$	0,74	2	-2339,48	M1a vs. M2a
M2a (selection)	$\omega_0 = 0,01, \omega_1 = 1,00, \omega_2 = 999,00$ $p_0 = 0,99, p_1 = 0,00, p_2 = 0,01$	5,12	4	-1689,97	1299,02***
M7 (beta)	$p = 0,02, q = 0,01$	0,80	2	-2348,47	M7 vs. M8
M8 (beta, ω)	$p_0 = 0,99, p = 0,07, q = 3,90$ $p_1 = 0,01, \omega_s = 999,00$	5,13	4	-1687,53	1321,87***

O LRT entre M0 vs. M3 é estatisticamente significativo quando grau de liberdade = 4 e valores críticos de qui-quadrado são 9,49, 13,28 e 18,47 em 5%*, 1%** e 0,1%*** de significância, respectivamente. O LRT entre M1a vs. M2a e entre M7 vs. M8 é estatisticamente significativo quando grau de liberdade = 2 e valores críticos de qui-quadrado são 5,99, 9,21 e 13,82 em 5%*, 1%** and 0,1%*** de significância, respectivamente.

Figura 29. Sítios de aminoácidos (trinca de bases) e classes de ω , obtidos a partir do modelo M3 dos testes de seleção, em sequências do gene mitocondrial COI para *Drosophila serido*.

A ausência de trincas de bases que não apresentam significância estatística para seleção positiva é de certa forma esperado, uma vez que o gene COI está relacionado à respiração celular (NELSON; COX, 2006) e, portanto, mudanças não-sinônimas na

sequência de aminoácidos podem acarretar em características deletérias para o indivíduo.

4.6.5 – Análises demográficas

Para a estimativa dos parâmetros populacionais realizada por meio do programa Lamarc as populações AF e N56, que a AMOVA hierárquica (Tabela 40) definiu como grupos independentes, tiveram que ser incluídas no agrupamento Nordeste, devido ao pequeno tamanho amostral.

Foi observado que o Nordeste possui o maior valor de θ , indicando maior tamanho populacional que os demais, que apresentaram valores de θ idênticos (Tabela 43). Em concordância com este parâmetro, a taxa de crescimento exponencial (g) mostrou evidência de crescimento populacional para o Nordeste, e de redução populacional para os demais grupos, pois os intervalos de confiança incluem zero (Tabela 43).

Tabela 43. Valores de theta (θ) e taxas de crescimento exponencial (g) para as populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 40), utilizando sequências do gene mitocondrial COI.

Grupo	Theta (θ)	Taxa de crescimento exponencial (g)
Nordeste (AF + N56)	0,058 (0,040; 0,089)	976,64 (471,87; 1001,12)
SP-RJ-ES	0,003 (0,001; 0,009)	58,57 (-431,71; 950,33)
SC	0,003 (0,001; 0,012)	-395,83 (-482,79; 892,71)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

As taxas de migração (M) foram convertidas para um parâmetro (M') que mede o balanço entre fluxo gênico e deriva genética como forças que podem promover divergência entre populações/grupos (Tabela 44 e Figura 30). Para o parâmetro M' o valor estimado é maior que 1, indicativo de que a força diversificadora da migração consegue sobrepor-se à deriva genética (Tabela 44). Este parâmetro auxilia na explicação do maior tamanho populacional observado para o Nordeste e também a redução populacional em SP-RJ-ES mostrado pelos valores de θ (Tabela 43).

Para os demais parâmetros estimados (Tabela 44 e Figura 30), os valores foram menores que 1, evidenciando que os eventos de migração não devem ser frequentes,

possivelmente ausentes, ou seja, a variação genética observada é resultado de deriva genética e não de eventos de migração.

Tabela 44. Taxas de migração para populações de *Drosophila serido* geneticamente estruturadas de acordo com os agrupamentos detectados pela AMOVA hierárquica (Tabela 40), utilizando sequências do gene mitocondrial COI.

	Taxa de migração – M	Taxa de migração – M'
M21	3,29 (-0,24; 80,22)	0,19 (-0,01; 7,14)
M31	58,95 (13,89; 96,85)	3,42 (0,55; 8,62)
M12	94,28 (13,86; 99,22)	0,30 (0,01; 1,22)
M32	7,77 (1,47; 96,89)	0,02 (0,00; 1,19)
M13	95,04 (17,15; 99,61)	0,31 (0,02; 0,89)
M23	7,75 (1,35; 97,05)	0,02 (0,00; 0,87)

Valores entre parênteses correspondem aos limites mínimo e máximo, respectivamente, do intervalo de confiança de 95%.

1) Nordeste; 2) SC; 3) SP-RJ-ES.

M21: taxa de migração de SC para o Nordeste; M31: taxa de migração de SP-RJ-ES para o Nordeste; M12: taxa de migração do Nordeste para SC; M32: taxa de migração de SP-RJ-ES para SC; M13: taxa de migração do Nordeste para SP-RJ-ES; M23: taxa de migração de SC para SP-RJ-ES.

M=m/mu: em que m é a chance de uma linhagem imigrar por geração, e μ é a taxa de mutação por sítio por geração.

M'=3Nm: em que N é o número efetivo de migrantes e m é a chance de uma linhagem imigrar por geração.

Figura 30. Taxas de migração estimadas (Tabela 44), por meio de sequências do gene nuclear COI, entre os agrupamentos populacionais de *Drosophila serido*. Seta contínua: $1 < M' < 5$; Seta tracejada: $M' > 5$; Seta pontilhada: $M' \leq 1$.

Simulações das histórias demográficas por meio da BSP (Figura 31) foram realizadas para todos os agrupamentos observados na AMOVA hierárquica (Tabela 40) para o gene COI, com exceção de AF e N56, devido ao pequeno tamanho amostral.

O T_{MRCA} para o agrupamento Nordeste foi estimado em 1,05 kyr (HPD 95%: 3,95 até 12,60 kyr), e a curva de simulação demográfica (Figura 31A) indica crescimento populacional desde o T_{MRCA} até os dias atuais, concordando com os resultados obtidos para θ e g (Tabelas 43).

A reconstrução demográfica para o agrupamento SP-RJ-ES apresentou redução populacional desde o T_{MRCA} até próximo dos dias atuais quando há um rápido crescimento populacional (Figura 31B). Seu T_{MRCA} foi estimado em 6,91 kyr (HPD 95%: 2,44 até 12,30 kyr). Já para o agrupamento SC foi observado o T_{MRCA} mais antigo, 15,03 kyr (HPD 95%: 9,21 até 25,60 kyr), e a BSP mostrou redução populacional (Figura 31C), desde seu T_{MRCA} até os dias atuais, sugerindo que as populações desta região passaram por gargalo populacional.

Figura 31. Simulações de história demográfica por meio da *Bayesian Skyline Plot* (BSP) para as populações de *Drosophila serido* estruturadas em (A) Nordeste; (B) RJ-ES-N32 e (C) SP-SC, utilizando sequências do gene mitocondrial COI. A curva negra corresponde à variação média na população ao longo do tempo. As curvas cinza correspondem ao HPD 95%. O tamanho populacional médio (N) é medido em unidades de $\log N_e \cdot \tau$ e o tempo em milhões de anos (Myr). A linha tracejada indica o tempo do ancestral comum mais recente da amostra (T_{MRCA}).

4.6.6 – Análises filogeográficas

As sequências do gene mitocondrial COI utilizadas para obtenção da rede haplotípica não-enraizada geraram duas redes individuais. A primeira delas, com maior número de haplótipos agrupados (Figura 32), foi utilizada para inferências filogeográficas por meio da NCPA e foi denominada neste trabalho como Rede A. A segunda rede haplotípica obtida (Figura 33) apresentou uma quantidade de haplótipos agrupados

consideravelmente menor que a primeira e, portanto, não foi possível realizar a NCPA. No presente trabalho, esta rede foi denominada como Rede B e possui haplótipos apenas de Florianópolis-SC.

Na Rede A, o haplótipo 57 foi considerado, pela análise, como o mais antigo, com o maior valor de *outgroup weight* (0,146), o que corresponde à probabilidade de ser a raiz ou grupo externo da rede haplotípica (CASTELLOE; TEMPLETON, 1994). Além disso, ele é o mais frequente na amostra e está localizado no interior da rede haplotípica (Figura 32). Na Tabela 45 encontra-se a lista de haplótipos compartilhados pelas populações de *D. serido*. Valores significativos nas análises filogeográficas foram observados em cinco clados e interpretados por meio da Chave de Inferência Filogeográfica (Tabela 46).

Para o clado 1-18 foi inferido o evento histórico de fluxo gênico restrito com isolamento por distância (Tabela 46). Neste clado estão agrupados haplótipos que ocorrem em populações dos Estados de São Paulo, Rio de Janeiro, Espírito Santo e uma população do sul da Bahia (Cabrália – BA), sugerindo que as populações mais próximas geograficamente tem maior chance de fluxo gênico do que populações mais distantes geograficamente. Este resultado corrobora a hipótese, obtida por meio da abordagem bayesiana, de fluxo gênico entre populações do litoral Sudeste e Nordeste (item 4.6.5).

No clado 1-20 foi inferido o evento de fragmentação alopátrica entre o haplótipo 48 e os demais haplótipos aninhados neste clado (Tabela 46). A razão desta inferência filogeográfica está no fato de que o haplótipo 48 é exclusivo da população de Água Fria – GO (AF).

As inferências filogeográficas seguintes referem-se aos níveis hierárquicos mais elevados da rede haplotípica (Figura 32). O evento de expansão de área contígua foi inferido para o clado 4-1 no qual estão aninhadas populações do clado 1-18 (descritas acima) e mais duas populações, Morro da Barrinha e Irecê, que estão localizadas nas pontas da rede (Figura 32), sugerindo que elas são mais recentes do que os haplótipos do litoral Sudeste. Este fato oferece suporte à hipótese de que teria ocorrido fluxo gênico entre as populações do litoral Sudeste e Nordeste, sendo que a direção do movimento migratório teria sido sul-norte, e que tal evento teria sido recente.

Para o clado 4-2 foi inferido o evento de colonização de longa distância e/ou fragmentação passada, eventos não mutuamente exclusivos (Tabela 46). O clado 4-2 é formado pelo aninhamento dos clados 3-3 e 3-4, sendo que o primeiro é constituído apenas por duas populações Morro da Barrinha – BA e Serra do Cipó-MG, e o último

possui haplótipos com ampla distribuição geográfica. Neste cenário, é possível que tenha ocorrido inicialmente a colonização de longa distância, pois o haplótipo de Morro da Barrinha – BA é mais antigo do que aqueles da Serra do Cipó-MG, seguido de fragmentação, pois esta última população está localizada em um morro com 860m de altitude (FRANCO, 2009), o que pode ter favorecido seu isolamento.

Para o clado 5-1 foi inferido o evento filogeográfico de expansão de área gradual passada seguida por fragmentação ou grande expansão de área passada seguida de extinção dos intermediários.

Figura 32. Rede haplotípica não-enraizada, Rede A, contendo haplótipos do gene mitocondrial COI para *Drosophila serido*. Cada linha da rede representa um passo mutacional. O tamanho dos círculos/quadrados que representam os haplótipos é proporcional ao número de indivíduos portando o haplótipo. Os círculos vazios representam haplótipos hipotéticos, não amostrados. A localidade geográfica de cada haplótipo pode ser visualizada na Tabela 45.

A Rede B (Figura 33) é composta por haplótipos exclusivos de Florianópolis-SC. A separação em duas redes independentes está relacionada ao critério de parcimônia assumido pela análise, pois seriam necessários muitos passos mutacionais para conectar as Redes A e B.

Figura 33. Rede haplotípica não-enraizada, Rede B, contendo haplótipos do gene mitocondrial COI para *Drosophila serido*. Cada linha da rede representa um passo mutacional. O tamanho dos círculos/quadrados que representam os haplótipos é proporcional ao número de indivíduos portando o haplótipo. Os círculos vazios representam haplótipos hipotéticos, não amostrados. A localidade geográfica de cada haplótipo pode ser visualizada na Tabela 45.

Tabela 45. Lista de haplótipos, do gene mitocondrial COI, compartilhados pelas redes A e B por populações de *Drosophila serido* em cada localidade geográfica analisada. Os códigos das localidades são os mesmos das Tabelas 1 e 2.

CL	Haplótipos	CL	Haplótipos
AF	6xHap48	R30	3xHap57, 1xHap20, 1xHap47, 1xHap46
J52	7xHap57	R31	3xHap57, 1xHap20, 1xHap04, 1xHap45
N32	2xHap57, 3xHap31	R32	3xHap57, 1xHap50, 1xHap07, 1xHap37
J92	3xHap57, 2xHap20, 1xHap18	R39	1xHap57, 3xHap20, 1xHap51
J95	4xHap57	R56	7xHap57, 1xHap20, 2xHap46
N29	6xHap31, 1xHap28	N71	2xHap57, 2xHap20, 1xHap36, 1xHap34, 1xHap37
N2	2xHap31, 4xHap28	R28	3xHap57, 2xHap20, 1xHap38
N88	3xHap57, 1xHap59, 2xHap58, 1xHap60	R33	1xHap57
N31	1xHap57, 3xHap31, 1xHap49, 1xHap16	R36	2xHap57, 2xHap20, 1xHap54, 1xHap06
N35	3xHap57, 1xHap49, 1xHap29	N20	6xHap31
N46	4xHap57, 1xHap09, 1xHap10	R37	5xHap57, 1xHap20, 1xHap44, 1xHap56
N68	2xHap57, 1xHap20, 1xHap19	N56	1xHap25, 1xHap23, 5xHap21, 1xHap22
N40	2xHap57, 2xHap49, 1xHap17	N37	2xHap57, 2xHap20, 1xHap44, 1xHap24, 1xHap13, 1xHap12, 2xHap35
N42	9xHap57, 1xHap43	R38	5xHap57, 1xHap20
N70	2xHap57, 2xHap20, 1xHap32, 1xHap39, 1xHap42, 1xHap33, 1xHap52, 1xHap40	R59	1xHap57
N65	3xHap57, 3xHap20, 1xHap11, 1xHap26, 1xHap55, 1xHap05	N67	1xHap57, 1xHap53
R29	1xHap57, 1xHap20, 1xHap08, 1xHap03, 1xHap37	N45	5xHap57
H49	2xHap31, 2xHap28	R34	1xHap20
R40	3xHap57, 3xHap20	R35	1xHap57, 1xHap54
N22	3xHap31, 1xHap28, 1xHap33	N43	3xHap57, 1xHap20, 1xHap11, 1xHap15, 1xHap14, 1xHap02, 1xHap01
R27	2xHap57, 2xHap20	N39	1xHap57

CL: código da localidade

Tabela 46. Resultados da Chave de Inferência Filogeográfica para os clados aninhados da rede haplotípica (Figura 32), obtida com sequências do gene mitocondrial COI para populações de *Drosophila serido*, que apresentaram resultados significativos na análise.

Clado	Etapa na Chave de Inferência Filogeográfica
1-18	1 – 2 – 3 – 4: Não (Fluxo gênico restrito com isolamento por distância)
1-20	1 – 2 – 3 – 4 – 9: Não (Fragmentação alopátrica)
4-1	1 – 2 – 11 – 12: Não (Expansão de área contígua)
4-2	1 – 2 – 11 – 12 – 13 – 14: Não (colonização de longa distância e/ou fragmentação passada – eventos não mutuamente exclusivos)
5-1	1 – 2 – 11 – 12 – 13 – 21: Não (Expansão de área gradual passada seguida por fragmentação ou grande expansão de área passada seguida de extinção dos intermediários)

5 – Discussão

Drosophila serido é uma espécie cactófila, ocorrendo em formações florestais abertas na porção sudeste da América do Sul. Suas populações, por meio da análise de distintos marcadores moleculares, são divididas em dois grupos: populações do Nordeste e do litoral.

Neste trabalho, a partir da análise de sequências nucleotídicas dos genes *$\alpha E5$* e *GstD1*, ambos autossômicos, *kl-5* e *period*, ligados aos cromossomos sexuais Y e X, respectivamente, e do mitocondrial COI, foi corroborada, de forma geral, a divisão citada acima das populações de *D. serido*. Por outro lado, dependendo do marcador, a estruturação das populações na costa Atlântica foi alterada e diferentes eventos demográficos foram descritos para cada um deles. Além disso, evidências de seleção, como um fator causal da variação, foram detectados nas análises das sequências dos gene autossômicos.

Inferimos que estes processos demográficos populacionais na evolução de *D. serido* são pleistocênicos, pois análises filogenéticas e demográficas com sequências do gene mitocondrial COI estimam a separação entre *D. serido* e a espécie irmã *D. antonietae* há aproximadamente 1,27 (FRANSACK, 2011) e 1,73 Myr (FRANCO; MANFRIN, 2013), respectivamente, sendo que um evento causal seria isolamento de áreas devido a influências paleoclimáticas durante o Pleistoceno (FRANCO; MANFRIN, 2013).

5.1 – Populações do Nordeste

Populações de *D. serido* da região Nordeste do Brasil têm sido investigadas ao longo dos anos por meio de marcadores citogenéticos (BAIMAI; SENE; PEREIRA, 1983; RUIZ et al., 2000; TOSI; SENE, 1989) e morfológicos (FRANCO et al., 2008; SILVA; SENE, 1991). Embora haja diversidade de ambientes nesta região, que inclui: 1) Mata Atlântica, com enclaves de vegetação xérica; 2) Agreste, que consiste em uma faixa estreita que se caracteriza como área de transição entre as florestas úmidas a leste e a Caatinga (sertões) a oeste (PRADO, 2003); e 3) Caatinga; estes trabalhos sugerem ausência de estruturação populacional. Da mesma forma, tal inferência foi obtida no presente trabalho, por meio de testes de AMOVA hierárquica (Tabelas 7, 15, 22, 30, 40), para os cinco marcadores analisados.

Há duas hipóteses, não mutuamente exclusivas, que podem auxiliar na compreensão da manutenção da homogeneidade genética entre estas populações, sendo elas: 1) a

evidência de áreas de estabilidade climática na Caatinga, mesmo durante as flutuações climáticas do Quaternário, e 2) a geomorfologia da região Nordeste.

Mapas com projeções paleoclimáticas, realizadas para táxons de ocorrência na Caatinga, apresentam grandes áreas de estabilidade climática na região Nordeste, ao longo das flutuações climáticas do Quaternário, sendo que apenas as ecorregiões Raso da Catarina, Planalto da Borborema e porção sul da Depressão Sertaneja, apresentam-se nos mapas de predição como áreas de instabilidade climática (WERNECK et al., 2011).

Com relação à geomorfologia da região Nordeste, ela apresenta complexidade topográfica relativamente reduzida, possuindo poucas formações como as Chapadas do Araripe, Apodi, Borborema e a Serra do Espinhaço, favorecendo dessa forma o fluxo gênico entre as populações, uma vez que a quantidade de barreiras é reduzida (WERNECK et al., 2012b). Para o lagarto *Phyllorhynchus pollicaris*, as análises genéticas apresentaram ausência de estruturação entre as populações da região Nordeste (WERNECK et al., 2012b), assemelhando-se ao padrão observado para *D. serido*.

Da mesma forma, o sistema hídrico desta região parece não constituir barreira ao fluxo gênico entre populações de *D. serido*. As análises não mostraram evidência de estruturação populacional ou eventos filogeográficos concordantes com a localização dos rios São Francisco e Jequitinhonha. No entanto, há evidências de que a importância dos rios atuando como barreira ao fluxo gênico pode variar de acordo com o grupo taxonômico analisado. Para a lagartixa *Gymnodactylus darwinii* (PELLEGRINO et al., 2005) análises genéticas mostraram estruturação populacional de acordo com os rios São Francisco e Jequitinhonha, e para o jacarandá-da-Bahia, *Dalbergia nigra*, a análise de estruturação genética apresentou agrupamentos ao norte e ao sul do Rio Jequitinhonha (RIBEIRO et al., 2011).

Localidades próximas à Chapada Diamantina compartilham o haplótipo mais antigo observado nas redes haplotípicas para os genes *period*, *kl-5*, *αE5* e COI (Figuras 11, 16, 26, 32), e possuem vários haplótipos em menor frequência ao redor deste haplótipo, caracterizando um formato em estrela, indicativo de expansão. Este fato observado no presente trabalho corrobora a hipótese de que a Chapada Diamantina seja o centro de dispersão de populações de *D. serido* (FRANCO; MANFRIN, 2013).

A única exceção para esta constatação consistiu na rede haplotípica obtida por meio de sequências do gene nuclear *GstD1* (Figura 21), em que o haplótipo mais antigo, segundo a análise de parcimônia, ocorre em Arraial do Cabo-RJ (haplótipo 30). No entanto, derivado do haplótipo 30 encontra-se o haplótipo 29, observado em elevada

frequência e compartilhado por populações da região Nordeste, inclusive em localidades próximas à Chapada Diamantina. É possível que a discordância observada tenha sido favorecida por eventos de recombinação (item 4.4.1) e seleção (item 4.4.4; Tabela 24) detectados nas análises deste gene.

Ao norte e ao sul da Chapada Diamantina, três rotas foram inferidas por Franco e Manfrin (2013), sendo elas: 1) nordeste, em direção ao Planalto da Borborema; 2) norte-centro, em direção à Serra dos Pirineus na região do Planalto Central; e 3) sudeste, em direção à Serra da Canastra e Serra do Cipó, ao longo da Serra do Espinhaço. Estas rotas também foram observadas nas análises filogeográficas do presente trabalho, no entanto, novas rotas também puderam ser inferidas.

Além das rotas sugeridas para moscas do “cluster” *D. buzzatii*, rotas semelhantes já haviam sido propostas anteriormente por Prado e Gibbs (1993) para plantas, com objetivo de explicar as conexões entre os núcleos da FTSS, formando o Arco Pleistocênico (PRADO; GIBBS, 1993). A rota norte ligou a Caatinga até o Estado de Goiás, passando por áreas do bioma Cerrado, provavelmente ao longo dos vales do Rio Grande, até o Distrito Federal. Neste ponto a rota desvia ao norte, em direção à Ilha do Bananal-TO, e ao sul, em direção ao Pantanal mato-grossense. A rota sul estendeu-se primeiramente pelos vales dos rios São Francisco e Jequitinhonha, e ao sul pelo Rio das Velhas, até a região da cidade de Belo Horizonte, chegando até o Rio de Janeiro (PRADO, 2000; PRADO; GIBBS, 1993).

Para o gene *period*, os eventos filogeográficos inferidos para os clados 1-14 e 2-5 (Tabela 13) sugerem diversos movimentos populacionais a partir da Chapada Diamantina. No clado 1-14 estão agrupados haplótipos compartilhados por localidades da Chapada Diamantina e do Planalto da Borborema, sendo que o formato em estrela corrobora o evento de fluxo gênico restrito/dispersão, mas com alguma dispersão de longa distância, concordando com a rota nordeste, que é compartilhada por *D. serido* e *D. borborema* (FRANCO; MANFRIN, 2013). Esta rota também foi observada por meio da análise filogeográfica utilizando sequências do gene nuclear *αE5* (Rede B, clado 3-8 – Figura 27, Tabela 38).

No clado 1-14 (gene *period*) também estão agrupados haplótipos das populações de São Paulo, que podem ser indicativo de uma rota de expansão. Não há haplótipos de populações intermediárias que indiquem se a dispersão ocorreu pelo interior ou pelo litoral, no entanto, as taxas de migração (Tabela 11, Figura 9) estimadas entre os agrupamentos da AMOVA hierárquica (Tabela 7) sugerem uma rota pelo interior, em

que houve fluxo gênico entre o Nordeste e São Paulo, sentido norte-sul. Sendo assim, é possível inferir que populações da Chapada Diamantina dispersaram-se até São Paulo por meio da rota sudeste (FRANCO; MANFRIN, 2013), em direção à Serra da Canastra, ao longo da Serra do Espinhaço.

No clado 2-5, além dos haplótipos do clado 1-14, estão agrupados haplótipos das populações Serra do Cipó (N56) (Figura 11, Tabela 12). Com relação à população N56, o evento filogeográfico inferido (Tabela 13) concorda com a rota sudeste (FRANCO; MANFRIN, 2013) de dispersão em direção à Serra do Cipó, ao longo da Serra do Espinhaço. Esta rota de dispersão é compartilhada por populações de *D. gouveai* e *D. seriema* (FRANCO; MANFRIN, 2013), sendo que Moraes et al (2009) datou este movimento para populações de *D. gouveai* em 170 kyr. Embora as estimativas do T_{MRCA} para os agrupamentos populacionais de *D. serido* (Figura 10) no presente trabalho não concordem com a datação obtida por Moraes et al (2009), é importante ressaltar que esta discordância pode estar relacionada 1) aos métodos estatísticos utilizados; 2) ao marcador molecular utilizado, Moraes et al. (2009) usaram sequências do gene mitocondrial COII; e 3) os movimentos de dispersão para *D. serido* e *D. gouveai* podem tratar-se de eventos assíncronicos.

As rotas norte-centro e sudeste (FRANCO; MANFRIN, 2013) foram identificadas no presente trabalho por meio das análises filogeográficas com sequências do gene *GstD1*. Nos clados 1-11 e 1-15 estão agrupados haplótipos compartilhados por populações do Nordeste, vale do Rio Jequitinhonha, Água Fria-GO (AF) e N56, sendo que os eventos filogeográficos inferidos (Tabela 28) sugerem o seguinte cenário: eventos de dispersão a partir da Chapada Diamantina em direção à AF e N56 e, posteriormente, um evento de fragmentação alopátrica teria favorecido o isolamento das populações do Nordeste e AF.

A população AF encontra-se inserida em um fragmento de FTSS no vale do Rio Paranã, no bioma Cerrado (WERNECK, 2011), evidência de que eventos de expansão e retração da vegetação xerofítica durante as flutuações climáticas do Quaternário (WERNECK et al., 2011; 2012a) podem ter influenciado também a dinâmica das populações de *D. serido*. O evento de fragmentação alopátrica pode ter sido favorecido pela área de estabilidade climática identificada no Cerrado, denominado Refúgio Serra Geral de Goiás (WERNECK et al., 2012a), a oeste da bacia do Rio Paranã (CARVALHO JÚNIOR; HERMUCHE; GUIMARÃES, 2006). Este evento histórico

também foi detectado por meio da análise filogeográfica utilizando sequências do gene mitocondrial COI (clado 1-20 – Figura 32, Tabela 46).

Esta rota norte-centro é compartilhada por populações de *D. serido* e *D. gouveai* (FRANCO; MANFRIN, 2013), sendo que para *D. gouveai* ela foi datada em 450 kyr (MORAES et al., 2009). Novamente, a discordância entre as estimativas do T_{MRCA} para os agrupamentos populacionais de *D. serido* (Figura 20) no presente trabalho e a datação obtida por Moraes et al (2009), pode estar relacionada 1) aos métodos estatísticos utilizados; 2) ao marcador molecular utilizado; e 3) os movimentos de dispersão podem ter sido assíncronos.

Para o gene *$\alpha E5$* os eventos filogeográficos inferidos para os clados 3-6, 4-2 e 5-1 da Rede B (Figura 27, Tabela 36) concordam com as rotas norte-centro e sudeste (FRANCO; MANFRIN, 2013) discutidas acima para os genes *period* e *GstD1*, uma vez que nestes clados estão aninhados haplótipos compartilhados por populações do Nordeste, vale do Rio Jequitinhonha, AF e N56.

Com relação ao litoral da região Nordeste, a análise filogeográfica realizada para o gene *period*, permitiu pela primeira vez a inferência de uma rota de expansão em direção ao Planalto da Borborema pelo litoral. O evento histórico de expansão de área contígua inferido para o clado 2-6 (Figura 11), confirma nossa hipótese de que o Rio São Francisco não constitui barreira ao fluxo gênico, uma vez que a expansão ocorre ao norte e ao sul da distribuição geográfica do haplótipo 48. Este haplótipo é compartilhado por Penedo-AL (foz do Rio São Francisco), Salvador-BA, Itaúnas-ES e Macaé-RJ e o evento de expansão ocorreu ao norte em direção ao Planalto da Borborema e ao sul em direção à Paúba-SP.

Dois eventos de fragmentação alopátrica, clados 1-46 e 2-19, da Rede haplotípica B, obtida para o gene *$\alpha E5$* (Figura 27, Tabela 38), foram inferidos. No clado 1-46 este evento ocorreu entre a população Morro da Barrinha-BA (N37) e as populações Cachoeira do Ferro Doido-BA (N42), Irecê-BA (N43), Salvador-BA (N65), Riacho-BA (R27), Costa Azul-BA (R32), Penedo-AL (R30), Lajes Pintadas-RN (R37), Almenara-MG (R56). Projeções paleoclimáticas com táxons da Caatinga mostraram que a ecorregião Raso da Catarina constituiu uma área de instabilidade climática, ao longo do Quaternário (WERNECK et al., 2011). Portanto, esta região pode ter favorecido a ocorrência do evento de fragmentação alopátrica entre N37 e as demais populações, uma vez que N37 e R27 estão localizadas próximo ao Raso da Catarina.

No clado 2-19, o evento histórico inferido pode estar relacionado à diferença na diversidade de cactáceas entre o interior e litoral da região Nordeste, uma vez que o evento ocorreu entre Milagres-BA (clado 1-36) e populações de Salvador-BA, Costa Azul-BA e Pacatuba-SE (clado 1-35). A região Nordeste possui elevada diversidade de cactos (TAYLOR; ZAPPI, 2004), no entanto, no litoral há ocorrência de *Cereus fernambucensis* (JOJIMA, 2014). Sendo assim, é importante considerar que a distribuição da diversidade genética atual em sequências do gene $\alpha E5$ pode ter sido influenciada por fatores ecológicos, uma vez que este gene pode estar envolvido em processos de desintoxicação dos metabólitos secundários presentes nos tecidos de cactos necrosados (FRANK; FOGLEMAN, 1992; MATZKIN, 2008). Além disso, testes de seleção mostraram evidência de seleção positiva em diversas trincas de bases (item 4.5.5, Tabela 32), fato que será discutido adiante.

5.2 – Populações do Litoral

Embora *D. serido* não esteja relacionada diretamente ao bioma Mata Atlântica, sua ocorrência no litoral se dá em áreas de restinga, definidas como planícies costeiras com faixas de vegetação entre as encostas das montanhas e o mar (SCARANO, 2002), que estão sujeitas às influências da seca, salinidade e reduzida quantidade de nutrientes (LACERDA; ARAUJO; MACIEL, 1993; SCARANO, 2002). Desta forma, a discussão sobre o processo de diversificação populacional em *D. serido* nesta região será conduzida de forma a extrapolar os eventos históricos relatados para Mata Atlântica, considerando que eles possam ter influenciado também a dinâmica evolutiva em áreas de restinga.

As populações com distribuição na costa Atlântica apresentam, de forma geral, uma divisão norte-sul, sendo que a posição geográfica desta divisão pode variar de acordo com o grupo taxonômico e a barreira considerada (CABANNE; SANTOS; MIYAKI, 2007; CARNAVAL et al., 2009; COSTA, 2003; DOLMAN; MORITZ, 2006; FITZPATRICK et al., 2009; GRAZZIOTIN et al., 2006; LARA-RUIZ; CHIARELLO; SANTOS, 2008; MARTINS et al., 2007; MORAES-BARROS et al., 2006; PELLEGRINO et al., 2005; RESENDE et al., 2010; THOMÉ et al., 2010; TONINI et al., 2013). No caso das populações de *D. serido*, a divisão norte-sul também foi observada, nos cinco marcadores analisados, sendo que ela coincide com a divisão já proposta: populações do Nordeste e do Litoral (Tabelas 7, 15, 22, 30, 40). No caso de

D. serido, a descontinuidade genética (divisão norte-sul) ocorre entre as localidades de Mucuri-BA e Itaúnas-ES. Este fato já foi sugerido anteriormente por análises de hidrocarbonetos de cutícula (OLIVEIRA et al., 2011), inversões cromossômicas polimórficas (RUIZ et al., 2000; TOSI; SENE, 1989), diversidade haplotípica mitocondrial (FRANCO; MANFRIN, 2013; MANFRIN; DE BRITO; SENE, 2001; MORALES, 2005) e análises cariotípicas (BAIMAI; SENE; PEREIRA, 1983).

Esta inferência recorrente nas análises populacionais de *D. serido* pode estar relacionada à presença de um manguezal no estuário do Rio Mucuri – Bahia (CUZZUOL; CAMPOS, 2001). Estudos têm mostrado que as flutuações climáticas do Quaternário, associadas às flutuações no nível do mar, também influenciaram as áreas de ocorrência dos manguezais, favorecendo eventos de expansão e retração de sua vegetação (COHEN et al., 2008, 2009; FRANÇA et al., 2013; PESSENDA et al., 2012). Desta forma, considerando a possibilidade de mudanças nas áreas de ocorrência dos manguezais, devido aos paleoclimas do Quaternário, e as incompatibilidades entre as características do solo para estabelecimento de manguezais e populações de Cactaceae, é possível inferir que o manguezal do estuário do Rio Mucuri pode ter atuado como barreira à dispersão e fluxo gênico dos cactos, assim como das populações de moscas cactófilas. Embora não haja datações ou dados que relatem estes eventos no manguezal do estuário do Rio Mucuri.

A relação entre o Rio Mucuri e a diversificação de linhagens foi relatada na espécie de formiga *Dinoponera lucida* (RESENDE et al., 2010) e para o anfíbio *Chiasmocleis carvalhoi* (TONINI et al., 2013), que apresentaram padrão semelhante ao observado em *D. serido*, podendo auxiliar na elucidação de outra hipótese quanto a descontinuidade genética observada. No caso de *D. lucida*, a estruturação de acordo com o Rio Mucuri foi datada em aproximadamente 1,22 Myr. Além disso, a elevada diversidade genética apresentada pelas populações de *D. lucida* que ocorrem ao norte das margens do Rio Mucuri, corrobora a hipótese de que tenha existido uma área de estabilidade climática para a Mata Atlântica, durante os períodos mais frios e secos do Quaternário, denominado de Refúgio Bahia (CARNAVAL et al., 2009; CARNAVAL; MORITZ, 2008). Considerando que frequentemente as idades de formação de rios não coincidem com estimativas de tempo de divergência, uma vez que constituem eventos históricos muito antigos, não é indicada sua associação direta com eventos de diversificação de linhagens/espécies (THOMÉ et al., 2014). Sendo assim, é possível que o Refúgio Bahia tenha atuado como barreira ao fluxo gênico entre populações de *D. serido* do Nordeste e

costa Atlântica, especialmente pelo fato de que o Rio Mucuri está na área *core* deste refúgio (RESENDE et al., 2010).

As análises filogeográficas e de estrutura populacional mostram que os rios Doce e Paraíba do Sul parecem não ter constituído barreira ao fluxo gênico entre as populações de *D. serido* ao longo do tempo. No entanto, a importância dos rios que compõem o sistema hídrico da costa Atlântica como barreira que tenha favorecido a diversificação de linhagens é controversa, e tem sido amplamente discutida (CABANNE; SANTOS; MIYAKI, 2007; COLOMBI; LOPES; FAGUNDES, 2010; PELLEGRINO et al., 2005; RESENDE et al., 2010; RIBEIRO et al., 2011; THOMÉ et al., 2010; 2014; TONINI et al., 2013).

Análises filogeográficas em *D. serido* (presente trabalho), *Rhinella crucifer* (anfíbio) (THOMÉ et al., 2010; 2014) e *Akodon cursor* (mamífero roedor) (COLOMBI; LOPES; FAGUNDES, 2010) sugerem que o Rio Doce não constituiu barreira geográfica relevante no processo de diferenciação de linhagens nestas espécies. Entretanto, o mesmo não foi observado em *Xiphorhynchus fuscus* (Arapaçu-rajado – ave; CABANNE; SANTOS; MIYAKI, 2007), *D. lucida* (formiga; RESENDE et al., 2010), *G. darwinii* (lagartixa; PELLEGRINO et al., 2005) e em *Dendropsophus elegans* e *C. carvalhoi* (anfíbios; TONINI et al., 2013), em que o Rio Doce coincidiu com as descontinuidades genéticas observadas. Além do Rio Doce, *G. darwinii* (PELLEGRINO et al., 2005) também apresentou estruturação populacional de acordo com o Rio Paraíba do Sul. Nestes casos, as possíveis explicações para a relação entre o sistema hídrico da costa Atlântica e a estruturação de linhagens estão em: 1) áreas de estabilidade climática na Mata Atlântica (CABANNE; SANTOS; MIYAKI, 2007; RESENDE et al., 2010); 2) características biológicas da espécie, como baixa capacidade de dispersão (RESENDE et al., 2010); 3) associação entre distâncias geográficas e a presença de barreiras (rios) (PELLEGRINO et al., 2005); e 4) estocasticidade do processo coalescente (TONINI et al., 2013).

Ao longo da costa Atlântica, as populações de *D. serido* têm apresentado padrões distintos quanto à sua estruturação, considerando marcadores citogenéticos (BAIMAI; SENE; PEREIRA, 1983; BIFFI et al., 2001; RUIZ et al., 2000) e morfológicos (FRANCO et al., 2008).

A maior estruturação populacional ao longo da costa Atlântica pode estar relacionada ao grau de influência que as flutuações climáticas do Quaternário apresentaram sobre a distribuição geográfica da vegetação. Projeções paleoclimáticas realizadas para táxons

de ocorrência na Mata Atlântica apresentam as florestas úmidas como áreas fragmentadas, concordando com hipóteses que sugerem que o impacto das fases mais secas do Pleistoceno teve maior influência na porção sul da Mata Atlântica do que ao norte (CARNAVAL; MORITZ, 2008). Além disso, por estar no limite leste do escudo brasileiro, esta área é mais complexa do ponto de vista topográfico, provavelmente devido à neotectônica e às mudanças no nível do mar durante o Quaternário (LIMA, 2000; SUGUIO et al., 2005).

Para os cinco marcadores moleculares analisados no presente trabalho, a politipia das populações de *D. serido* nesta região foi observada, tornando-a ainda mais evidente. Embora eles tenham apresentado três padrões de estruturação populacional.

Os genes *period* e *kl-5*, ligados aos cromossomos sexuais X e Y, respectivamente, foram concordantes quanto à estruturação genética das populações de *D. serido* na costa Atlântica. As divisões propostas para estas populações, RJ-ES, SP e SC, concordam com os três tipos de placas metafásicas inferidos para Arraial do Cabo-RJ (tipo III), Peruíbe-SP (tipo IV) e Florianópolis-SC (outro tipo, não nomeado) (BAIMAI; SENE; PEREIRA, 1983; BIFFI et al., 2001). No entanto, esta estruturação populacional não concorda com as hipóteses sugeridas a partir de dados de diversidade morfológica em edeagos, que sugerem agrupamento de populações de SP e SC (FRANCO et al., 2008), e concorda parcialmente com as divisões em setores geomorfológicos da costa brasileira proposta por Ab'Saber (2003a), pois ES e RJ pertencem à diferentes macrossetores da costa Atlântica, Litoral Leste e Sudeste, respectivamente.

A estruturação das populações de SP com as demais pode ter sido influenciada pela área de estabilidade climática, denominada Refúgio São Paulo (CARNAVAL et al., 2009; CARNAVAL; MORITZ, 2008), inferida por meio de projeções paleoclimáticas com táxons de ocorrência na Mata Atlântica. Sendo assim, esta área pode ter contribuído para o processo de diversificação destas linhagens.

Com relação à estruturação das populações de SC, a presença de manguezais no Paraná, como na Baía de Guaratuba (AB'SABER, 2003a), pode ter tido papel relevante no processo de isolamento e diversificação destas populações com relação às demais. Além disso, diversas coletas foram realizadas pelo nosso grupo de pesquisa e não há registros de ocorrência de cactos/moscas ao longo do litoral do Paraná (M. H. MANFRIN, comunicação pessoal). Na região entre SP e SC também foi observada uma quebra genética dentro da linhagem C (C1 e C2) do complexo de espécies *R. crucifer* (THOMÉ et al., 2010; 2014), cuja datação é de 600 kyr, sendo que para C1 o limite de

sua distribuição na costa Atlântica ocorre em Peruíbe-SP e a área de ocorrência de C2 se inicia em Guaratuba-PR (THOMÉ et al., 2010). Thomé et al. (2014) discutem que esta quebra genética é coincidente com a localização do lineamento Guapiara e com a zona de cisalhamento de Cubatão, ambas consideradas falhas tectônicas (SAADI et al., 2002), contudo, defini-las como elementos causais do processo de diferenciação entre linhagens é uma questão delicada, pois não há informação disponível com relação à datação e extensão destes eventos.

Com relação aos genes autossômicos, *GstDI* e *αE5*, os resultados quanto à estruturação populacional foram congruentes, sugerindo dois agrupamentos na costa Atlântica, RJ-ES e SP-SC. Esta divisão 1) não concorda com os três tipos de placas metafásicas (BAIMAI; SENE; PEREIRA, 1983; BIFFI et al., 2001), uma vez que o agrupamento SP-SC possui dois tipos; 2) não concorda com os setores geomorfológicos propostos por Ab'Saber (2003a), pois SC pertence ao macrossetor Litoral Sul; 3) concorda parcialmente com os dados de diversidade morfológica em edeagos (FRANCO et al., 2008), e 4) concorda com os dados de inversões polimórficas, pois populações de Guaratuba-SP e Florianópolis-SC compartilham as inversões $2x^7y^9x^8w^8$ (RUIZ et al., 2000).

Experimentos com indivíduos de *D. buzzatii*, com a inversão cromossômica *2st*, utilizando meio de cultura contendo o cacto *Trichocereus terscheckii*, apresentaram viabilidade significativamente maior do que nos outros tratamentos. Já indivíduos de *D. buzzatii* com a inversão *2jq⁷* tiveram maior viabilidade em meios de cultura contendo os cactos *Opuntia ficus-indica* e *Opuntia vulgaris*. Estes dados sugerem que a interação cariótipo/cacto pode ser determinante no valor adaptativo da espécie (IRIARTE; HASSON, 2000). A concordância com os dados de inversão cromossômica e a estruturação genética do agrupamento SP-SC, observado no presente trabalho, é um fato interessante, pois as inversões cromossômicas têm sido relacionadas à seleção (HASSON et al., 1995; RUIZ et al., 1986) e, os testes de seleção para os genes *GstDI* e *αE5* mostraram evidência de seleção positiva (itens 4.4.4 e 4.5.4).

Para estes marcadores nucleares autossômicos não há evidência de que a presença de manguezais no Paraná (AB'SABER, 2003a) possa ter sido uma barreira ao fluxo gênico entre SP e SC. Na verdade, o compartilhamento de diversidade genética entre estas populações também pode estar relacionado ao tempo de coalescência dos marcadores moleculares, uma vez que genes autossômicos possuem o maior tempo de coalescência, devido ao seu modo de herança (TEMPLETON, 2011). Assim, o compartilhamento

observado pode remeter a um cenário mais antigo na história evolutiva de *D. serido*. Por outro lado, a evidência de recombinação (itens 4.4.1 e 4.5.1) e seleção positiva (itens 4.4.4 e 4.5.4) podem gerar incongruências com relação à história mostrada em diferentes marcadores (HARE, 2001).

Análises filogeográficas utilizando sequências do gene *GstD1* sugerem para o clado 2-4 o evento histórico de fluxo gênico restrito com isolamento por distância (Figura 21, Tabela 28). Neste clado estão agrupadas populações com distribuição ao longo da costa Atlântica, localizadas no sul da Bahia (Mucuri e Cabralia), Itaúnas-ES, São Paulo (Bertioga, Paúba e Guaratuba) e Florianópolis-SC. Há uma conexão entre os haplótipos aninhados do sul da Bahia-ES e Santa Catarina, sugerindo que este pode ser um polimorfismo ancestral compartilhado, uma vez que para todos os marcadores SC é sugerida como o agrupamento mais antigo (Figuras 10, 15, 20, 25, 31).

Para o gene *αE5* foi inferido o evento de fluxo gênico restrito com isolamento por distância para os clados 3-1 e 3-4 (Figura 26, Tabela 37). Para o primeiro, o evento histórico envolve populações do sul da Bahia (Mucuri e Cabralia), Espírito Santo, Rio de Janeiro e Ilha de Alcatrazes-SP, sugerindo a possibilidade de conexão com uma ilha da costa Atlântica, que pode constituir uma rota de dispersão entre populações do RJ e SP. Para o clado 3-4, a inferência filogeográfica está relacionada às populações de Santa Catarina, São Paulo e Mucuri-BA. No entanto, a disposição dos haplótipos na rede sugere que o movimento tenha sido no sentido sul-norte, uma vez que o haplótipo mais antigo da amostra ocorre em Santa Catarina e São Paulo, e o haplótipo de Mucuri-BA possui uma posição derivada com relação aos demais (Figura 26). Esta inferência concorda com as rotas migratórias estimadas (Tabela 34), e reforça a hipótese de ancestralidade para as populações de SC (Figuras 10, 15, 20, 25, 31).

A ocorrência e direção dos movimentos populacionais ao longo da costa Atlântica sugerem que eles devem ter sido recorrentes e assíncronos (Tabelas 34 e 44; Figuras 24 e 30). As estimativas do T_{MRCA} para os agrupamentos populacionais utilizando sequências dos genes *GstD1* e *αE5* indicam que o efeito causal destes eventos históricos está nas flutuações climáticas do Quaternário (Figuras 20 e 25).

O gene mitocondrial COI apresentou um terceiro padrão na estruturação populacional ao longo da costa Atlântica, com dois agrupamentos, ES-RJ-SP e SC. Esta divisão não concorda com aquela sugerida pelos tipos de placas metafásicas (BAIMAI; SENE; PEREIRA, 1983; BIFFI et al., 2001) e pela diversidade morfológica dos edeagos (FRANCO et al., 2008), que agrupa populações de SP e SC. No entanto, concorda

parcialmente com os setores geomorfológicos propostos por Ab'Saber (2003a), pois Itaúnas-ES faz parte do macrossetor Litoral Leste e as populações do RJ e SP encontram-se na área abrangente do setor Litoral Sudeste, considerado o macrossetor mais diversificado e acidentado de todo o país (AB'SABER, 2003a).

A estruturação das populações de SC em um agrupamento independente também corresponde ao setor geomorfológico descrito por Ab'Saber (2003a), denominado Litoral Sul. A presença de manguezais no limite sul do Estado de São Paulo (PESSENDA et al., 2012) e no Paraná (Baía de Guaratuba, por exemplo) (AB'SABER, 2003a) podem ter influenciado e favorecido o isolamento entre as populações de SC e as demais da costa Atlântica, como já discutido para os genes nucleares *period* e *kl-5*. Da mesma forma, a descontinuidade genética observada entre o litoral sudeste e sul para o gene mitocondrial COI concorda com a diversificação dentro da linhagem C do complexo de espécies *R. crucifer* (THOMÉ et al., 2010; 2014) citada anteriormente.

5.3 – Populações de Santa Catarina

Diversas hipóteses têm sido sugeridas, com base em análises de sequências do gene mitocondrial COI, com relação ao processo de diversificação populacional de *D. serido* (FRANCO, 2009; FRANCO; MANFRIN, 2013; MORALES, 2005). Dentre elas, a conexão sugerida por redes haplotípicas entre populações da Caatinga e Santa Catarina, sem aparente conexão com populações da costa Atlântica, tem sido discutida (FRANCO, 2009; FRANCO; MANFRIN, 2013; MORALES, 2005). Por meio das análises filogeográficas, com abordagem multi-*locus*, realizadas no presente trabalho, foi possível realizar inferências para explicar o compartilhamento de haplótipos entre estas duas regiões.

Estimativas do T_{MRCA} obtidas no presente trabalho, para todos os genes exceto *aE5*, sugerem que as populações de *D. serido* de Santa Catarina sejam mais antigas que as demais (Figuras 10, 15, 20, 25 e 31). É possível que alguns dos eventos de diversificação da espécie *D. serido* tenham ocorrido na região sul do Brasil, uma vez que a área de ocorrência de sua espécie irmã, *D. antonietae*, abrange a região da bacia dos rios Paraná-Paraguai (MANFRIN; SENE, 2006). Além disso, ambas as espécies compartilham a inversão cromossômica fixa $2x^7$ (RUIZ et al., 2000; WASSERMAN; RICHARDSON, 1987) e, portanto, é mais plausível inferir que esta inversão

cromossômica seja uma condição ancestral mantida, do que inferir que ela tenha surgido independentemente em *D. serido* e *D. antonietae*.

Eventos de migração entre Santa Catarina e Nordeste, em que o primeiro doa migrantes para o segundo, foram inferidos por meio de sequências do gene nuclear *period* (Tabela 11 e Figura 9). Desta forma, considerando que as populações mais antigas de *D. serido* encontram-se em Santa Catarina, a hipótese filogeográfica que sumariza as evidências é a de que indivíduos de *D. serido* teriam migrado pelo interior do continente até a região Nordeste, onde teriam se estabelecido. Devido à grande diversidade de cactos nesta região (TAYLOR; ZAPPI, 2004), as populações se estabeleceram, diversificaram, sendo que, posteriormente, teriam expandido sua área de ocorrência colonizando novas áreas a partir da Chapada Diamantina, considerado o centro de dispersão para diversas espécies do “cluster” *D. buzzatii* (FRANCO; MANFRIN, 2013; presente trabalho).

A conexão entre Santa Catarina e Nordeste, por meio do interior do continente, provavelmente foi favorecida pela expansão da vegetação xerofítica durante as flutuações climáticas do Quaternário, concordando com a Hipótese do Arco Pleistocênico (PRADO; GIBBS, 1993), que sugere a expansão dos núcleos da FTSS durante períodos climáticos mais secos e frios do que o atual, formando um contínuo na distribuição da vegetação, há aproximadamente 18 – 12 kyr (PRADO; GIBBS, 1993). Esta datação coincide com as estimativas do T_{MRCA} para o Nordeste e SC a partir de sequências dos genes *period* (Figura 10), *kl-5* (Figura 15), *GstD1* (Figura 20) e COI (Figura 31).

Embora haja evidências quanto à ancestralidade das populações de SC, é possível que após o estabelecimento das populações de *D. serido* no Nordeste, os movimentos de dispersão a partir da Chapada Diamantina tenham permitido um novo contato entre estas populações e as de SC. Esta hipótese baseia-se no evento filogeográfico inferido para o clado 2-5 (Tabela 13), por meio da análise de sequências do gene *period*, que sugere expansão de área gradual e passada seguida de fragmentação ou grande expansão de área no passado seguida por extinção dos intermediários. Neste clado estão agrupados haplótipos que ocorrem no Nordeste, N56 e SC, assim a evidência da rota sudeste, que sugere dispersão a partir da Chapada Diamantina em direção à Serra do Cipó (N56), ao longo da Serra do Espinhaço, dá suporte para uma conexão mais ao sul chegando até SC pelo interior do continente.

5.4 – Seleção

Embora seleção positiva tenha sido detectada em todos os marcadores (Figuras 8, 13, 18, 23; Tabelas 9, 17, 24, 32), com exceção do gene mitocondrial COI (Figura 29; Tabela 42), ela foi mais evidente no gene nuclear *αE5*. Contudo, considerando que os produtos dos genes nucleares *GstD1* e *αE5* podem estar relacionados ao processos de desintoxicação dos metabólitos secundários, presentes nos tecidos de cactos necrosados (MATZKIN et al., 2006; ROBIN et al., 2000), e à concordância da estruturação populacional (Tabelas 22 e 30), a influencia da seleção será discutida para ambos os genes.

É possível que a divisão entre as populações Nordeste e Litoral (divisão norte-sul), ao longo da distribuição geográfica de *D. serido*, tenha sido favorecida pela seleção exercida por compostos secundários encontrados nas diferentes espécies de cactos, uma vez que o Nordeste apresenta elevada diversidade de cactáceas (TAYLOR; ZAPPI, 2004), enquanto que na costa Atlântica há registros de ocorrência de 1) *C. fernambucensis*, em matas de restinga ao longo do litoral brasileiro, desde o Rio Grande do Norte até o sul de São Paulo (JOJIMA, 2014); 2) *Pilosocereus arrabidaei* em áreas de restinga e afloramentos rochosos no litoral de Rio de Janeiro, Espírito Santo e sudeste da Bahia (ZAPPI, 1994); 3) *Pilosocereus brasiliensis brasiliensis* em áreas de restinga e afloramentos rochosos no litoral de Rio de Janeiro e Espírito Santo (ZAPPI, 1994); 4) *Pilosocereus ulei* em afloramentos rochosos no litoral, próximo à região de Cabo Frio-RJ (ZAPPI, 1994); e 5) *C. hildmannianus* e *O. monacantha* no litoral de Santa Catarina (BIZZO, 2011; RUIZ et al., 2000; SILVA, 2013).

Dentre estes registros, há informações de *D. serido* que emergiram de cladódios em decomposição de *C. fernambucensis* amostrados em Peruíbe-SP (PEREIRA; VILELA; SENE, 1993) e Guaratuba-SP (RUIZ et al., 2000), e de cladódios em decomposição de *C. hildmannianus* e *O. monacantha* amostrados no litoral de Santa Catarina (BIZZO, 2011; RUIZ et al., 2000). Embora não haja informação de que *D. serido* utilize os cactos do gênero *Pilosocereus* citados acima, é possível que ela o faça, pois Pereira; Vilela; Sene (1993) relataram que *D. serido* emergiu de cladódios em decomposição de *Ce. piauhyensis*, que segundo Zappi (1994), é *Pilosocereus piauhyensis*.

A diversidade de cactos na costa Atlântica parece ser condizente com os agrupamentos genéticos sugeridos pelos genes *GstD1* e *αE5*, sendo eles: RJ-ES e SP-SC. Apesar da ocorrência de *C. fernambucensis* ser comum aos dois agrupamentos, a

diferença entre eles ocorre quanto à presença de cactos *Pilosocereus* no primeiro, e *C. hildmannianus* e *O. monacantha* no segundo. É importante considerar que cactos *Opuntia* são quimicamente menos complexos que cactos colunares (OLIVEIRA et al., 2012), sendo que estes possuem alcalóides, ácidos graxos de cadeia média, esteróis e triterpenos (CORIO et al., 2013). Desta forma, é possível inferir que as diferentes espécies de cactos ao longo da costa Atlântica tenham favorecido a diversificação destas populações por meio de seleção.

Ainda com relação às populações na costa Atlântica, vale ressaltar a relação entre o agrupamento genético SP-SC e o compartilhamento da inversão cromossômica $2x^7y^9x^8w^8$, observado entre populações de Guaratuba-SP e Florianópolis-SC (RUIZ et al., 2000). Hasson et al. (1995) sugerem que, em diferentes cenários ecológicos, o uso de cactos hospedeiros pode constituir o fator causal da observação de diferentes inversões cromossômicas. Desta forma, como relatado por Iriarte e Hasson (2000) em experimentos de laboratório, indivíduos de *D. buzzatii* com diferentes inversões cromossômicas apresentaram diferenças na viabilidade com relação ao cacto utilizado no meio de cultura, sugerindo que a interação cariótipo/cacto pode ser determinante no valor adaptativo da espécie. Sendo assim, é possível que em populações de *D. serido* também haja uma relação entre o hospedeiro utilizado e a inversão cromossômica encontrada.

Com relação às populações da região Nordeste, um dado relevante foi obtido por meio da análise filogeográfica utilizando sequências do gene *αE5*. Para o clado 2-19 (Rede B, Figura 27, Tabela 38), foi inferido o evento de fragmentação alopátrica entre uma população do interior, Milagres-BA, e outras populações do litoral, todas localizadas na região Nordeste. Este evento histórico pode ter tido como causa primária a diversidade de cactos, fator ecológico, em ambas regiões, dando suporte à nossa hipótese de que para os genes *GstD1* e *αE5* fatores ecológicos, por meio de seleção, podem ter sido mais relevantes no processo de diversificação de *D. serido* do que fatores históricos, como as flutuações climáticas do Quaternário.

6 – Conclusões

- A região da Chapada Diamantina parece ser o centro de dispersão das populações de *D. serido*, tendo sido colonizada inicialmente por populações de Santa Catarina;
- Não há estruturação populacional entre as populações do Nordeste;
- Há uma divisão norte e sul ao longo da costa Atlântica, que coincide com a denominação populações do Nordeste e Litoral, e que ocorre em Mucuri-BA;
- Houve discordância entre os marcadores moleculares quanto à estrutura populacional na porção sudeste-sul da costa Atlântica:
 - Genes nucleares autossômicos *GstD1* e *αE5* → dois grupos (RJ-ES e SP-SC);
 - Genes nucleares *period* (cromossomo X) e *kl-5* (cromossomo Y) → três grupos (RJ-ES, SP e SC);
 - Gene mitocondrial COI → dois grupos, porém não coincidem com a mesma divisão sugerida pelos genes nucleares autossômicos (RJ-ES-SP e SC);
- Com exceção do gene mitocondrial COI, todos os demais marcadores moleculares apresentaram evidência de seleção positiva;
- A estruturação populacional sugerida pelos genes autossômicos *GstD1* e *αE5* parecem estar mais relacionada à fatores ecológicos (seleção) do que históricos;
- A influência das flutuações climáticas do Quaternário parece ter contribuído para o processo de diversificação populacional de *D. serido*;
- A colonização das localidades de Água Fria-GO e Serra do Cipó-MG parece ter ocorrido por eventos de dispersão/fluxo gênico, a partir de localidades próximas à Chapada Diamantina, com posterior fragmentação;
- Múltiplos movimentos de dispersão/fluxo gênico assíncronicos parecem ter ocorrido ao longo da costa Atlântica entre as populações de *D. serido*.

7 – Referências

AB'SABER, A. N. Espaços ocupados pela expansão dos climas secos da América do Sul por ocasião dos períodos glaciais quaternários. **Paleoclimas**, v. 3, p. 1-19, 1977.

AB'SÁBER, A.N. **Litoral do Brasil - Brazilian Coast**. Metalivros, 2003a.

AB'SÁBER, A.N. **Os Domínios de natureza no Brasil**. Ed. Ateliê, 2003b.

ANDREWS, K. R. et al. The evolving male: spinner dolphin (*Stenella longirostris*) ecotypes are divergent at Y chromosome but not mtDNA or autosomal markers. **Molecular Ecology**, v. 22, p. 2408 – 2423, 2013.

AULER, A. S. et al. Quaternary ecological and geomorphic changes associated with rainfall events in presently semi-arid northeastern Brazil. **Journal of Quaternary Science**, v. 19, p. 693–701, 2004.

AVISE, J. C. et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. **Annual Review of Ecology, Evolution and Systematics**, v. 18, p. 489-522, 1987.

AVISE, J. C. Phylogeography: retrospect and prospect. **Journal of Biogeography**, v. 36, p. 3-15, 2009.

AVISE, J. C. **Phylogeography: The history and formation of species**. Cambridge: Harvard University Press, 2000.

BAIMAI, V.; SENE, F. M.; PEREIRA, M. A. O. R. Heterochromatin and karyotypic differentiation of some neotropical cactus-breeding species of the *Drosophila repleta* species group. **Genetica**, v. 60, n. 2, p. 81-92, 1983.

BARKER, J. S.; STARMER, W. T. **Ecological Genetics and Evolution. The cactus-yeast-Drosophila Model System**. Austrália: Academic Press, 1982.

- BARR, N. B.; CUI, L.; MCPHERON, B. A. Molecular systematics of nuclear gene *period* in genus *Anastrepha* (Tephritidae). **Annals of the Entomological Society of America**, v. 98, n. 2, p. 173-180, 2005.
- BAUZER, L. G. S. R. et al. The *period* gene and genetic differentiation between three Brazilian populations of *Lutzomyia longipalpis*. **Insect Molecular Biology**, v. 11, n. 4, p. 315-323, 2002.
- BEAUMONT, MA; et al. In defence of model-based inference in phylogeography. **Molecular Ecology**, v. 19, p. 436-446, 2010.
- BIFFI, F. et al. 2001. **Padrão cariotípico de uma área de contato entre as espécies cactófilas *Drosophila serido* e *Drosophila antonietae* (Grupo *repleta*)**. Resumos do 47º Congresso Brasileiro de Genética.
- BIZZO, L. E. M. **Zona de contato secundário: simpatria entre duas espécies de *Drosophila***. 2011. 134 f. Tese (Doutorado em Ciências) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2011.
- BONATELLI, I. A. S. et al. Interglacial microrefugia and diversification of a cactus species complex: phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. **Molecular Ecology**, v. 23, p. 3044-3063, 2014.
- BONO, J. M. et al. Molecular evolution and population genetics of two *Drosophila mettleri* cytochrome P450 genes involved in host plant utilization. **Molecular Ecology**, v. 17, p.3211-3221, 2008.
- BRITO, P. H.; EDWARDS, S. V. Multilocus phylogeography and phylogenetics using sequence-based markers. **Genetica**, v. 135, p. 439-455, 2009.
- BRUEN, T. C.; PHILIPPE, H.; BRYANT, D. A simple and robust statistical test for detecting the presence of recombination. **Genetics**, v. 172, p. 2665 – 2681, 2006.

CABANNE, G. S.; SANTOS, F. R.; MIYAKI, C. Y. Phylogeography of *Xiphorhynchus fuscus* (Passeriformes, Dendrocolaptidae): vicariance and recent demographic expansion in southern Atlantic forest. **Biological Journal of the Linnean Society**, v. 91, p. 73 – 84, 2007.

CARLING, M. D.; BRUMFIELD, R. T. Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). **PLos One**, v. 2, n. 1, p.e160, 2007.

CARNAVAL, A. C. et al. Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. **Science**, v. 323, p. 785-789, 2009.

CARNAVAL, A. C.; BATES, J. M. Amphibian DNA shows marked genetic structure and tracks Pleistocene climate change in northeastern Brazil. **Evolution**, v. 61, p. 2942-2957, 2007.

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

CARVALHO JÚNIOR, O. A.; HERMUCHE, P. M.; GUIMARÃES, R. F. Identificação regional da floresta estacional decidual na bacia do Rio Paranã a partir da análise multitemporal de imagens modis. **Revista Brasileira de Geofísica**, v. 24, n. 3, p. 319 – 332, 2006.

CARVALHO, A. B. Origin and evolution of the *Drosophila* Y chromosome. **Current Opinion in Genetics and Development**, v. 12, p. 664-668, 2002.

CASTELLOE, J.; TEMPLETON, A. R. Root probabilities for intraspecific gene trees under neutral coalescent theory. **Molecular Phylogenetics and Evolution**, v. 3, n. 2, p. 102-113, 1994.

CHIPPINDALE, A. K.; RICE, W. R. Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. **PNAS**, v. 98, n. 10, p. 5677-5682, 2001.

CLEMENT, M.; POSADA, D.; CRANDALL, K. A. TCS: a computer program to estimate gene genealogies. **Molecular Ecology**, v. 9, p. 1657-1659, 2000.

COHEN, M. C. L. et al. Impact of sea-level and climatic changes on the Amazon coastal wetlands during the late Holocene. **Vegetation History and Archaeobotany**, v. 18, p. 425 – 439, 2009.

COHEN, M. C. L. et al. Wetland dynamics of Marajó Island, northern Brazil, during the last 1000 years. **Catena**, v. 76, p. 70 – 77, 2008.

COLOMBI, V. H.; LOPES, S. R.; FAGUNDES, V. Testing the Rio Doce as a riverine barrier in shaping the Atlantic rainforest population divergence in the rodent *Akodon cursor*. **Genetics and Molecular Biology**, v. 33, n. 4, p. 785 – 789, 2010.

COLOT, H. V.; HALL J. C; ROSBASH, M. Interspecific comparison of the *period* gene of *Drosophila* reveals large blocks of non-conserved coding DNA. **The EMBO Journal**, v. 7, n. 12, p. 3929 – 3937, 1988.

CORIO, C. et al. An alkaloid fraction extracted from the cactus *Trichocereus terscheckii* affects fitness in the cactophilic fly *Drosophila buzzatii* (Diptera: Drosophilidae). **Biological Journal of the Linnean Society**, v. 109, p. 342-353, 2013.

COSTA, L. P. The historical bridge between the Amazon and the Atlantic Forest of Brazil: a study of molecular phylogeography with small mammals. **Journal of Biogeography**, v. 30, p. 71 – 86, 2003.

CUZZUOL, G. R. F.; CAMPOS, A. Aspectos nutricionais na vegetação de manguezal do estuário do Rio Mucuri. **Revista Brasileira de Botânica**, v. 24, n. 2, p. 227 – 234, 2001.

DE BRITO, R. A. D.; MANFRIN, M. H.; SENE, F. M. Mitochondrial DNA phylogeography of Brazilian populations of *Drosophila buzzatii*. **Genetics and Molecular Biology**, v. 25, n. 2, p. 161-171, 2002a.

DE BRITO, R. A. D.; MANFRIN, M. H.; SENE, F. M. Nested cladistic analysis of Brazilian populations of *Drosophila serido*. **Molecular Phylogenetics and Evolution**, v. 22, n. 1, p. 131-143, 2002b.

DE OLIVEIRA, P. E.; BARRETO, A. M. F.; SUGUIO, K. Late Pleistocene/Holocene climatic and vegetational history of the Brazilian caatinga: the fossil dunes of the middle São Francisco River. **Palaeogeography, Palaeoclimatology, Palaeoecology**, v. 152, p. 319-337, 1999.

DOLMAN, G.; MORITZ, C. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). **Evolution**, v. 60, n. 3, p. 573 – 582, 2006.

DRUMMOND, A. J. et al. Bayesian coalescent inference of past population dynamics from molecular sequences. **Molecular Biology and Evolution**, v. 22, n. 5, p. 1185-1192, 2005.

DRUMMOND, A. J. et al. Bayesian phylogenetics with BEAUti and the BEAST 1.7. **Molecular Biology and Evolution**, v. 29, n. 8, p. 1969-1973, 2012.

DYER, K. A. et al. Molecular evolution of a Y chromosome to autosome gene duplication in *Drosophila*. **Molecular Biology and Evolution**, v. 28, n. 3, p. 1293-1306, 2011.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792 – 1797, 2004.

ERSTS, P. J. Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Disponível em <http://biodiversityinformatics.amnh.org/open_source/gdmg>. Acesso em: 08 nov. 2014.

EVANNO, G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software Structure: a simulation study. **Molecular Ecology**, v. 14, p. 2611 – 2620, 2005.

EXCOFFIER, L.; LISCHER, H. E. L. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Molecular Ecology Resources**, v. 10, n. 3, p. 564-567, 2010.

EXCOFFIER, L.; SMOUSE, P. E.; QUATTRO, J. M. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**, v. 131, n. 2, p. 479-491, 1992.

FAY, J. C.; W, C. I. Hitchhiking under positive Darwinian selection. **Genetics**, v. 155, p. 1405-1413, 2000.

FELSENSTEIN, J. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? **Molecular Biology and Evolution**, v. 23, n. 3, p. 691-700, 2006.

FIASCHI, P.; PIRANI, J. R. Review of plant biogeographic studies in Brazil. **Journal of Systematics and Evolution**, v. 47, n. 5, p. 477-496, 2009.

FITZPATRICK, S. W. et al. Geographical variation in genetic structure of an Atlantic Coastal Forest frog reveals regional differences in habitat stability. **Molecular Ecology**, v. 18, p. 2877 – 2896, 2009.

FLOT, J-F. SeqPhase: a web tool for interconverting PHASE input/output files and FASTA sequence alignments. **Molecular Ecology Resources**, v. 10, p. 162 – 166, 2010.

FRANÇA, M. C. et al. Mangrove vegetation changes on Holocene terraces of the Doce River, southeastern Brasil. **Catena**, v. 110, p. 59 – 69, 2013.

FRANCO, F. F. et al. 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**, v. 48, n. 4, p. 322-331, 2010.

FRANCO, F. F. et al. Phenotypic variation of the aedeagus of *Drosophila serido* Vilela & Sene (Diptera: Drosophilidae). **Neotropical Entomology**, v. 37, n. 5, p. 558-563, 2008.

FRANCO, F. F. **História evolutiva do “cluster” *Drosophila buzzatii* (grupo *D. repleta*): eventos históricos e diversificação de espécies no Brasil**. 2009. 197 f. Tese (Doutorado em Genética) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009.

FRANCO, F. F.; MANFRIN, M. H. Recent demographic history of cactophilic *Drosophila* species can be related to the Quaternary paleoclimatic changes in South America. **Journal of Biogeography**, v. 40, p. 142-154, 2013.

FRANK, M. R.; FOGLEMAN J. C. Involvement of cytochrome P450 in host-plant utilization by Sonoran Desert *Drosophila*. **PNAS**, v. 89, p. 11998-12002, 1992.

FRANSAK, R. F. **Filogenia do complexo *Drosophila buzzatii* (Grupo *repleta*): inferências de análises multilocus mitocondriais e nucleares**. 2011. 100 f. Dissertação (Mestrado em Genética) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2011.

FU, Y. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. **Genetics**, v. 147, p. 915-925, 1997.

GEPNER, J.; HAYS, T. S. A fertility region on the Y chromosome of *Drosophila melanogaster* encodes a dynein microtubule motor. **PNAS**, v. 90, p. 11132-11136, 1993.

GRAZZIOTIN, F. G. et al. Phylogeography of the *Bothrops jararaca* complex (Serpentes: Viperidae): past fragmentation and island colonization in the Brazilian Atlantic Forest. **Molecular Ecology**, v. 15, p. 3969 – 3982, 2006.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, v. 41, p. 95-98, 1999.

HARE, M. P. Prospects for nuclear gene phylogeography. **TRENDS in Ecology & Evolution**, v. 16, n. 12, p. 700-706, 2001.

HASEGAWA, M.; KISHINO, H.; YANO, T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. **Journal of Molecular Evolution**, v. 22, p. 160-174, 1985.

HASSON, E. C. et al. The evolutionary history of *Drosophila buzzatii*. XXVI. Macrogeographic patterns in the inversion polymorphisms in New World populations. **Journal of Evolutionary Biology**, v. 8, p. 369–384, 1995.

HEIN, J.; SCHIERUP, M. H.; WIUF, C. **Gene genealogies, variation and evolution**. Oxford, 2005.

HICKERSON, M. J. et al. Phylogeography's past, present, and future: 10 years after Avise, 2000. **Molecular Phylogenetics and Evolution**, v. 54, p. 291-301, 2010.

HO, S. Y. W. Calibrating molecular estimates of substitution rates and divergence times in birds. **Journal of Avian Biology**, v. 38, p. 409-414, 2007.

HUECK, K. **As Florestas da América do Sul**. Editora Polígono, 1972.

HUSON, D. H.; BRYANT, D. Application of phylogenetic networks in evolutionary studies. **Molecular Biology and Evolution**, v. 23, n. 2, p. 254 – 267, 2006.

IRIARTE, P. F.; HASSON, E. The role of the use of different host plants in the maintenance of the inversion polymorphism in the cactophilic *Drosophila buzzatii*. **Evolution**, v. 54, n. 4, p. 1295-1302, 2000.

JAKOBSSON, M.; ROSENBERG, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. **Bioinformatics**, v. 23, n. 14, p. 1801 – 1806, 2007.

JOJIMA, C. L. **Filogeografia de *Cereus fernambucensis* (Cactaceae, Cereeae) inferida por marcadores moleculares**. 2014. 73 f. Dissertação (Mestrado em Diversidade Biológica e Conservação) – Universidade Federal de São Carlos, Sorocaba, 2014.

KINGMAN, J. F. C. The coalescent. **Stochastic Processes and their Applications**, v. 13, p. 235-248, 1982.

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

KOERICH, L. B. et al. Low conservation of gene content in the *Drosophila* Y chromosome. **Nature**, v. 456, p. 949-951, 2008.

KOKUDAI, C. B. S.; SENE, F. M.; MANFRIN, M. H. Sympatry and asymmetric introgression between the cactophilic species *Drosophila serido* and *Drosophila antonietae* (Diptera: Drosophilidae). **Annals of the Entomological Society of America**, v. 104, n. 3, p. 434-442, 2011.

KONOPKA, R. J.; BENZER, S. Clock mutants of *Drosophila melanogaster*. **PNAS**, v. 68, n. 9, p. 2112-2116, 1971.

KUHNER, M. K. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. **Bioinformatics**, v. 22, n. 6, p. 768 – 770, 2006.

KYRIACOU, C. P.; HALL, J. C. Circadian rhythm mutations in *Drosophila melanogaster* affect short-term fluctuations in the male's courtship song. **PNAS**, v. 77, n. 11, p. 6729-6733, 1980.

LACERDA, L. D.; ARAUJO, D. S. D. ; MACIEL, N. C. Dry coastal ecosystem of the tropical Brazilian coast. In: VAN DER MAAREL, E. (Ed.). **Dry coastal ecosystems: Africa, America, Asia and Oceania**. Amsterdam: Elsevier, 1993, p. 477-493.

LARA-RUIZ, P.; CHIARELLO, A. G.; SANTOS, F. R. Extreme population divergence and conservation implications for the rare endangered Atlantic Forest sloth, *Bradypus torquatus* (Pilosa, Bradypodidae). **Biological Conservation**, v. 141, p. 1332 – 1342, 2008.

LEDRU, M. P. Late Quaternary environment and climatic changes in central Brazil. **Quaternary Research**, v. 39, p. 90-98, 1993.

LIBRADO, P.; ROZAS, J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. **Bioinformatics**, v. 25, p. 1451-1452, 2009.

LIMA, C. C. U. O neotectonismo na costa do sudeste e do nordeste brasileiro. **Revista de Ciência e Tecnologia**, v. 15, p. 91-102, 2000.

LOW, W. Y.; et al. Molecular evolution of glutathione S-transferases in the genus *Drosophila*. **Genetics**, v. 177, p. 1363-1375, 2007.

MACHADO, C. A.; HEY, J. The causes of phylogenetic conflict in a classic *Drosophila* species group. **Proceedings of the Royal Society London B**, v. 270, p. 1193-1202, 2003.

MAGALHÃES, I. L. F. et al. Strong spatial structure, Pliocene diversification and cryptic diversity in the Neotropical dry forest spider *Sicarius cariri*. **Molecular Ecology**, v. 23, p. 5323 – 5336, 2014.

MANFRIN, M. H.; DE BRITO, R. O. A.; SENE, F. M. Systematics and evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) cluster using mtDNA. **Annals of the Entomological Society of America**, v. 94, n. 3, p. 333-346, 2001.

MANFRIN, M. H.; SENE, F. M. Cactophilic *Drosophila* in South America: A model for evolutionary studies. **Genetica**, v. 126, n. 1, p. 57-75, 2006.

MARTINS, F. M. et al. Mitochondrial DNA phylogeography reveals marked population structure in the common vampire bat, *Desmodus rotundus* (Phyllostomidae). **Journal of Zoological Systematics and Evolutionary Research**, v. 45, n. 4, p. 372 – 378, 2007.

MARTINS, F. M. Historical biogeography of the Brazilian Atlantic forest and the Carnaval–Moritz model of Pleistocene refugia: what do phylogeographical studies tell us? **Biological Journal of the Linnean Society**, v. 104, n. 3, p. 499-509, 2011.

MATZKIN, L. M. et al. Functional genomics of cactus host shifts in *Drosophila mojavensis*. **Molecular Ecology**, v. 15, n. 14, p. 4635-4643, 2006.

MATZKIN, L. M. Population transcriptomics of cactus host shifts in *Drosophila mojavensis*. **Molecular Ecology**, v. 21, n. 10, p. 2428-39, 2012.

MATZKIN, L. M. The molecular basis of host adaptation in cactophilic *Drosophila*: molecular evolution of a Glutathione S-Transferase gene (GstD1) in *Drosophila mojavensis*. **Genetics**, v. 178, n. 2, p. 1073-1083, 2008.

MAYLE, F. E. Assessment of the Neotropical dry forest refugia hypothesis in the light of palaeoecological data and vegetation model simulations. **Journal of Quaternary Science**, v. 19, n. 7, p. 713-720, 2004.

MCDONALD, J. H.; KREITMAN, M. Adaptive protein evolution at the Adh locus in *Drosophila*. **Nature**, v. 351, p. 652-654, 1991.

MOONEY, H.A.; BULLOCK, S.H.; MEDINA, E. **Seasonally dry tropical forests**. Cambridge University Press, 1995.

MORAES-BARROS, N. et al. Comparative phylogeography of the Atlantic forest endemic sloth (*Bradypus torquatus*) and the widespread three-toed sloth (*Bradypus variegatus*) (Bradypodidae, Xenarthra). **Genetica**, v. 126, p. 189-198, 2006.

MORAES, E. M. et al. Phylogeography of the cactophilic species *Drosophila gouveai*: demographic events and divergence timing in dry vegetation enclaves in eastern Brazil. **Journal of Biogeography**, v. 36, n. 11, p.2136-2147, 2009.

MORALES, A. C. **Análise histórico-evolutiva de populações das espécies cactofílicas *Drosophila serido* e *Drosophila antonietae* (Diptera, Drosophilidae)**. 2005. 110 f. Tese (Doutorado em Ciências) - Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2005.

MURPHY, P. G.; LUGO, A. E. Ecology of tropical dry forest. **Annual Review of Ecology and Systematics**, v. 17, p. 67-88, 1986.

NELSON, D. L.; COX, M. M. **Princípios de bioquímica**. Sarvier, 2006.

NIELSEN, R.; YANG, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. **Genetics**, v. 148, p. 929 – 936, 1998.

OLIVEIRA, C. C. et al. Variations on a theme: diversification of cuticular hydrocarbons in a clade of cactophilic *Drosophila*. **BMC Evolutionary Biology**, v. 11, p. 179, 2011.

OLIVEIRA, D. C. S. G. et al. Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. **Molecular Phylogenetics and Evolution**, v. 64, n. 3 ,p. 533-544, 2012.

OLIVEIRA, M. T.; GARESSE, R.; KAGUNI, L. S. Animal models of mitochondrial DNA transactions in disease and ageing. **Experimental Gerontology**, v. 45, n. 7-8, p. 489-502, 2010.

PANCHAL, M. The automation of Nested Clade Phylogeographic Analysis. **Bioinformatics**, v. 23, n. 4, p. 509-510, 2007.

PELLEGRINO, K. C. M. et al. Phylogeography and species limits in the *Gymnodactylus darwinii* complex (Gekkonidae, Squamata): genetic structure coincides with river systems in the Brazilian Atlantic Forest. **Biological Journal of the Linnean Society**, v. 85, n. 1, p. 13-26, 2005.

PENNINGTON, R. T. et al. Historical climate changes and speciation: Neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. **Philosophical Transactions of The Royal Society B**, v. 359, p. 515-538, 2004.

PENNINGTON, R. T.; PRADO, D. E.; PENDRY, C. A. Neotropical seasonally dry forests and Quaternary vegetation changes. **Journal of Biogeography**, v.27, n.2, p.261-273. 2000.

PEREIRA, M. A. Q. R.; VILELA, C. R.; SENE, F. M. Notes of breeding and feeding sites of some species of the repleta group of the genus *Drosophila* (Diptera, Drosophilidae). **Ciência y Cultura**, v. 35, p. 1313-1319, 1993.

PESSEDA, L. C. R. et al. Late Quaternary vegetation and coastal environmental changes at Ilha do Cardoso mangrove, southeastern Brazil. **Palaeogeography, Palaeoclimatology, Palaeoecology**, p. 57 – 68, 2012.

PLUZHNIKOV, A.; DONNELLY, P. Optimal sequencing strategies for surveying molecular genetic diversity. **Genetics**, v. 144, n. 3, p. 1247-1262, 1996.

POELCHAU, M. F.; HAMRICK, J. L. Palaeodistribution modelling does not support disjunct Pleistocene refugia in several Central American plant taxa. **Journal of Biogeography**, v. 40, p. 662 – 675, 2013.

PÔRTO, K. C.; CABRAL, J. J. P.; TABARELLI, M. **Brejos de altitude em Pernambuco e Paraíba: História natural, ecologia e conservação**. Ministério do Meio Ambiente, 2004.

POSADA, D. jModelTest: phylogenetic model averaging. **Molecular Biology and Evolution**, v. 25, n. 7, p. 1253 – 1256, 2008.

POSADA, D.; BUCKLEY, T. R. Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian Approaches over Likelihood Ratio Tests. **Systematic Biology**, v. 53, n. 5, p. 793-808, 2004.

POSADA, D.; CRANDALL, K. A. MODELTEST: testing the model of DNA substitution. **Bioinformatics**, v. 14, n. 9, p. 817-818, 1998.

POSADA, D.; CRANDALL, K. A.; TEMPLETON, A. R. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. **Molecular Ecology**, v. 9, n. 4, p. 487-488, 2000.

PRADO, D. E. **As caatingas da América do Sul**. In: Leal, I. R.; Tabarelli, M.; Silva, J. M. C. (Eds) *Ecologia e Conservação da Caatinga*. Ed. Universitária da UFPE, Recife, p. 3 – 74, 2003.

PRADO, D. E. Seasonally dry forests of tropical South America: from forgotten ecosystems to a new phytogeographic unit. **Edinburgh Journal of Botany**, v. 57, p. 437–461, 2000.

PRADO, D. E.; GIBBS, P. E. Patterns of species distributions in the dry seasonal forests of South America. **Annals of the Missouri Botanical Garden**, v. 80, p. 902-927, 1993.

PRITCHARD, J. K.; STEPHENS, M.; DONNELLY, P. Inference of population structure using multilocus genotype data. **Genetics**, v. 155, p. 945 – 959, 2000.

QUIJADA-MASCAREÑAS, A. et al. Phylogeographic patterns of trans-Amazonian vicariants and Amazonian biogeography: the Neotropical rattlesnake (*Crotalus durissus* complex) as an example. **Journal of Biogeography**, v. 34, p. 1296-1312, 2007.

RAMBAUT, A. et al. Tracer version 1.6, Disponível em: <<http://beast.bio.ed.ac.uk/Tracer>>. Acesso em: 08 dez. 2014.

RAMBAUT, A. FigTree: Tree figure drawing tool. Disponível em: <<http://tree.bio.ed.ac.uk/software/figtree/>>. Acesso em: 09 dez. 2014.

RAMÍREZ-SORIANO, A. et al. Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. **Genetics**, v. 179, p. 555-567, 2008.

RAMOS-ONSINS, S. E.; ROZAS, J. Statistical properties of new neutrality tests against population growth. **Molecular Biology and Evolution**, v. 19, n. 12, p. 2092-2100, 2002.

REGIER, J. C. et al. Phylogenetic relationships of wild silkmoths (Lepidoptera : Saturniidae) inferred from four protein-coding nuclear genes. **Systematic Entomology**, v. 33, p. 219-228, 2008.

RESENDE, H. C. et al. Pliocene and Pleistocene events shaping the genetic diversity within the central corridor of the Brazilian Atlantic Forest. **Biological Journal of the Linnean Society**, v. 101, p. 949 – 960, 2010.

RIBEIRO, R. A. et al. Phylogeography of the endangered rosewood *Dalbergia nigra* (Fabaceae): insights into the evolutionary history and conservation of the Brazilian Atlantic Forest. **Heredity**, v. 106, n. 1, p. 46-57, 2011.

ROBIN, G. C. Q. et al. Reconstructing the diversification of α -Esterases: comparing the gene clusters of *Drosophila buzzatii* and *D. melanogaster*. **Journal of Molecular Evolution**, v. 51, p. 149-160, 2000.

RONQUIST F. et al. MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. **Systematic Biology**, v. 61, p. 539-542, 2012.

RUIZ, A. et al. The *Drosophila serido* speciation puzzle: putting new pieces together. **Genetica**, v. 108, n. 3, p. 217-227, 2000.

RUIZ, A. et al. The evolutionary history of *Drosophila buzzatii*. VIII. Evidence for endocyclic selection acting on the inversion polymorphism in a natural population. **Evolution**, v. 40, p. 740-755, 1986.

RUIZ, A.; WASSERMAN, M. Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. **Heredity**, v. 70, n. 6, p. 582-596, 1993.

SAADI, A. et al. Map and database of Quaternary faults and lineaments in Brazil. U.S. Geological Survey. OpenFile Report 02-230, Version 1.0. Disponível em: <<http://pubs.usgs.gov/of/2002/ofr-02-230/>>. Acesso em: 09 jan. 2002.

SANTOS, M. H. **As rotas de dispersão de *Drosophila buzzatii* na América do Sul**. 128 f. Tese (Doutorado em Genética) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2011.

SANTOS, M. H. et al. *Esterase-5* gene sequences from the *Drosophila buzzatii* cluster. **DIS**, v. 92, p. 44-56, 2009.

SCARANO, F. R. Structure, function and floristic relationships of plant communities in stressful habitats marginal to the Brazilian Atlantic Rainforest. **Annals of Botany**, v. 90, p. 1 – 8, 2002.

SILVA, G. A. R. **Evolução de *Cereus hildmannianus* (Cactaceae) no sul do Brasil**. 2013. 129 f. Tese (Doutorado em Genética) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2013.

SILVA, A. F. G.; SENE, F. M. Morphological geographic variability in *Drosophila serido* (Diptera, Drosophilidae). **Revista Brasileira de Entomologia**, v. 35, p. 455-468, 1991.

SIMON, C. et al. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved Polymerase Chain Reaction primers. **Annals of the Entomological Society of America**, v. 87, n. 6, p. 651-701, 1994.

STEPHENS, M.; SCHEET, P. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. **The American Journal of Human Genetics**, v. 76, p. 449 – 462, 2005.

STEPHENS, M.; SMITH, N. J.; DONNELLY, P. A new statistical method for haplotype reconstruction from population data. **The American Journal of Human Genetics**, v. 68, p. 978 – 989, 2001.

SUGUIO, D. et al. **Paleoníveis do Mar e Paleolinhas de Costa**. In: Souza, C. R. G.; Suguio, K.; Oliveira, A. M. S.; Oliveira, P. E. (Eds) Quaternário do Brasil. Holos Editora, Ribeirão Preto, p. 114 – 129, 2005.

TAJIMA, F. Statistical analysis of DNA polymorphism. **The Japanese Journal of Genetics**, v. 68, p. 567 – 595, 1993.

TAJIMA, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, v. 123, n. 3, p. 585-595, 1989.

TAMURA, K.; NEI, M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. **Molecular Biology and Evolution**, v. 10, n. 3, p. 512-526, 1993.

TAVARÉ, S. Some probabilistic and statistical problems in the analysis of dna sequences. **Lectures on Mathematics in the Life Sciences**, v. 17, p. 57 – 86, 1986.

TAYLOR, N.; ZAPPI, D. **Cacti of Eastern Brazil**. The University of Chicago Press, 2004.

TEMPLETON, A. R. **Genética de Populações e Teoria Microevolutiva**. Sociedade Brasileira de Genética, 2011.

TEMPLETON, A. R. Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographical analysis vs. approximate Bayesian computation. **Molecular Ecology**, v. 18, p. 319 – 331, 2009.

TEMPLETON, A. R. Statistical phylogeography: methods of evaluating and minimizing inference errors. **Molecular Ecology**, v. 13, p. 789-809, 2004.

TEMPLETON, A. R.; BOERWINKLE, E.; SING, C. F. 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**, v. 117, n. 2, p. 343-351, 1987.

TEMPLETON, A. R.; ROUTMAN, E.; PHILLIPS, C. A. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial dna haplotypes in the tiger salamander, *Ambystoma tigrinum*. **Genetics**, v. 140, n. 2, p. 767-782, 1995.

THOMÉ, M. T. C. et al. Barriers, rather than refugia, underlie the origin of diversity in toads endemic to the Brazilian Atlantic Forest. **Molecular Ecology**, v. 23, p. 6152 – 6164, 2014.

THOMÉ, M. T. C. et al. Phylogeography of endemic toads and post-Pliocene persistence of the Brazilian Atlantic Forest. **Molecular Phylogenetics and Evolution**, v. 55, n. 3, p. 1018-1031, 2010.

THOMPSON, J.; HIGGINS, D.; GIBSON, T. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, v. 22, n. 22, p. 4673-4680, 1994.

TONINI, J. F. R.; COSTA, L. P.; CARNAVAL, A. C. Phylogeographic structure is Strong in the Atlantic Forest; predictive power or correlative palaeodistribution models, not always. **Journal of Zoological Systematics and Evolutionary Research**, v. 51, n. 2, p. 114 – 121, 2013.

TOSI, D.; SENE, F. M. Further studies on chromosomal variability in *Drosophila serido* (Diptera, Drosophilidae). **Revista Brasileira de Genética**, v. 1, p. 729-746, 1989.

VILELA, C. R. A revision of the *Drosophila repleta* species group (Diptera; Drosophilidae). **Revista Brasileira de Entomologia**, v. 27, p. 1-114, 1983.

WANG, R. L.; HEY, J. The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the *period* locus. **Genetics**, v. 144, n. 3, p. 1113-1126, 1996.

WANG, X. F. et al. Wet periods in northeastern Brazil over the past 210 kyr linked to distant climate anomalies. **Nature**, v. 432, p. 740–743, 2004.

WASSERMAN, M.; RICHARDSON, R. H. Evolution of Brazilian *Drosophila mulleri* complex species. **Journal of Heredity**, v. 78, n. 5, p. 282-286, 1987.

WERNECK, F. P. et al. Climatic stability in the Brazilian Cerrado: implications for biogeographical connections of South American savannas, species richness and conservation in a biodiversity hotspot. **Journal of Biogeography**, v. 39, p.1695-1706, 2012a.

WERNECK, F. P. et al. Deep diversification and long-term persistence in the South American ‘dry diagonal’: integrating continent-wide phylogeography and distribution modeling of geckos. **Evolution**, v. 66, n. 10, p. 3014 – 3034, 2012b.

WERNECK, F. P. et al. Revisiting the historical distribution of Seasonally Dry Tropical Forests: new insights based on palaeodistribution modelling and palynological evidence. **Global Ecology and Biogeography**, v. 20, n. 2, p. 272-288, 2011.

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

YANG, Z. **Computational Molecular Evolution**. Oxford University Press Inc, 2006.

YANG, Z. et al. Codon-substitution models for heterogeneous selection pressure at amino acid sites. **Genetics**, v. 155, p.431 – 449, 2000.

YANG, Z. PAML 4: phylogenetic analysis by maximum likelihood. **Molecular Biology and Evolution**, v. 24, n. 8, p. 1586 – 1591, 2007.

YANG, Z.; WONG, W. S. W.; NIELSEN, R. Bayes empirical bayes inference of amino acid sites under positive selection. **Molecular Biology and Evolution**, v. 22, n. 4, p. 1107 – 1118, 2005.