

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOLOGIA
PÓS-GRADUAÇÃO EM BIOLOGIA COMPARADA

**Diversidade genética intraespecífica na espécie cactófila *Drosophila serido* Vilela
& Sene, 1977 do Grupo *Drosophila repleta*.**

Gabriel Simei Faganello

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para obtenção do título de Mestre em Ciências, obtido no Programa de Pós-Graduação em Biologia Comparada.

Ribeirão Preto – SP

2023

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOLOGIA
PÓS-GRADUAÇÃO EM BIOLOGIA COMPARADA

Diversidade genética intraespecífica na espécie cactófila *Drosophila serido* Vilela & Sene,
1977 do Grupo *Drosophila repleta*.

Gabriel Simei Faganello

VERSÃO ORIGINAL

Dissertação apresentada à Faculdade de Filosofia,
Ciências e Letras de Ribeirão Preto da Universidade de
São Paulo, como parte das exigências para obtenção do
título de Mestre em Ciências, obtido no Programa de Pós-
Graduação em Biologia Comparada.

Orientadora: Profa. Dra. Maura Helena Manfrin

Ribeirão Preto – SP

2023

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

FICHA CATALOGRÁFICA

Faganello, Gabriel Simei

Diversidade genética intraespecífica na espécie cactófila *Drosophila serido* Vilela & Sene, 1977 do Grupo *Drosophila repleta*. Ribeirão Preto, 2023.

80 p.

Dissertação de Mestrado apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – USP. Área de concentração: Biologia Comparada.

Orientadora: Manfrin, Maura Helena.

1. *Drosophila serido*. 2. Estrutura populacional. 3. Filogeografia.

Nome: FAGANELLO, Gabriel Simei

Título: Diversidade genética intraespecífica na espécie cactófila *Drosophila serido* Vilela & Sene, 1977 do Grupo *Drosophila repleta*.

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, como parte das exigências para obtenção do título de Mestre em Ciências, obtido no Programa de Pós-Graduação em Biologia Comparada.

Aprovado em: ____/____/_____

Banca Examinadora

Prof(a). Dr(a). _____

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof(a). Dr(a). _____

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof(a). Dr(a). _____

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof(a). Dr(a). _____

Instituição: _____

Julgamento: _____ Assinatura: _____

AGRADECIMENTOS

Agradeço imensamente à Profa. Dra. Maura Helena Manfrin pela oportunidade de orientação, paciência, compreensão, conversas e correções durante todos os processos que envolveram o desenvolvimento e finalização do presente trabalho.

À Dra. Dora Yovanna Barrios Leal, por todos os ensinamentos e disponibilidade de ajuda nas análises sempre que precisei.

Ao técnico Paulo Ricardo Epifânio, pelos auxílios e pelo trabalho que realiza no Laboratório de Genética Evolutiva.

A todos os pesquisadores que foram responsáveis pela coleta de material biológico, extração e sequenciamento dos dados genéticos dos quais pude utilizar para a realização do presente trabalho. Não existem palavras para agradecer o trabalho, o esforço e a contribuição de vocês.

Ao Prof. Dr. Milton Groppo Junior, coordenador do Programa de Pós-Graduação em Biologia Comparada do Departamento de Biologia da USP-RP e à Vera Cassia Cicilini de Lucca, secretária do mesmo programa, por todo o trabalho e suporte ao durante a pós-graduação.

À FAPESP, CAPES, CNPq e USP, cujo fomento mantém o Laboratório de Genética Evolutiva, da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto – USP.

À minha família. Aos meu pais, Marcelo e Glaucia e meu irmão Rafael. Obrigado por existirem, pelo amor e pelo apoio incondicional em todas as minhas decisões.

"It is to be stressed that all of these estimates are, at best, guesses."

- Ashburner *et al.*, 1984

RESUMO

Faganello, Gabriel Simei. “**Diversidade genética intraespecífica na espécie cactófila *Drosophila serido* Vilela & Sene, 1977 do Grupo *Drosophila repleta*.**” Dissertação – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, 2023, 80 folhas.

Ao longo de milhões de anos, a restrita associação ecológica de drosófilas neotropicais com cactáceas pode ter tido um papel fundamental na diversificação e irradiação desses animais. O presente trabalho teve como objetivo, por meio de análises moleculares utilizando os marcadores COI, *kl-5*, *GstD1* e *α -esterase5*, testar a hipótese de estruturação genética na espécie cactófila *Drosophila serido* em dois grupos populacionais distintos: um ocorrendo no bioma Caatinga e no litoral da região Nordeste (Grupo Nordeste) e o outro ocorrendo predominantemente no litoral do sul da Bahia e nos litorais Sudeste e Sul do Brasil (Grupo Litoral). Os resultados das análises multilocus são congruentes com a hipótese de estruturação em *D. serido*, mas em um padrão diferente do que foi estabelecido na hipótese inicial, sugerindo a existência de uniformidade genética do Grupo Nordeste e estruturação populacional dentro do Grupo Litoral, que parece ter sofrido múltiplos eventos assíncronos de dispersão populacional ao longo do litoral Atlântico brasileiro.

PALAVRAS-CHAVE: *Drosophila serido*; Estrutura populacional; Filogeografia.

ABSTRACT

Faganello, Gabriel Simei. “**Intraspecific genetic diversity in the cactophilic species *Drosophila serido* Vilela & Sene, 1977 from the *Drosophila repleta* group.**”

Dissertação – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo, 2023, 80 folhas.

Over millions of years, the narrow ecological association between neotropical drosophilids and cacti may have played a fundamental role in the diversification and radiation of these animals. The present study aimed, through molecular analyses using COI, *kl-5*, *GstD1*, and *α-esterase5* markers, to test the hypothesis of genetic structuring in the cactophilic species *Drosophila serido* into two distinct population groups: one occurring in the Caatinga biome and along the coast of the Northeast region (Northeast Group), and the other occurring predominantly along the southern coast of Bahia and the Southeast and South coasts of Brazil (Coastal Group). The results of the multilocus analyses are congruent with the hypothesis of structuring in *D. serido*, but in a different pattern than established in the initial hypothesis, suggesting the existence of genetic uniformity of the Northeast Group and population structuring within the Coastal Group, which seems to have undergone multiple asynchronous events of population dispersal along the Brazilian Atlantic coast.

KEYWORDS: *Drosophila serido*; Population structure; Phylogeography.

SUMÁRIO

1. INTRODUÇÃO	4
1.1. As espécies cactófilas do “cluster” <i>Drosophila buzzatii</i> (grupo <i>repleta</i>, gênero <i>Drosophila</i>)	4
1.2. <i>Drosophila serido</i> Vilela & Sene, 1977	4
1.3. Filogeografia de <i>Drosophila serido</i> e Marcadores Moleculares	6
1.3.1. Gene mitocondrial COI.....	8
1.3.2. Gene nuclear <i>kl-5</i> ligado ao cromossomo Y.....	9
1.3.3. Genes nucleares autossômicos <i>GstD1</i> e α - <i>esterase5</i>	9
2. OBJETIVOS	11
3. MATERIAL E MÉTODOS	12
3.1. Material Biológico	12
3.2. Análises Moleculares	18
3.2.1. Análises Filogenéticas.....	18
3.2.2. Estatística Descritiva.....	20
3.2.3. Estrutura Populacional.....	20
3.2.4. Análises Demográficas.....	22
4. RESULTADOS	24
4.1. Análises Moleculares	24
4.1.1. Análises Filogenéticas.....	24
4.1.2. Estatística Descritiva.....	31
4.1.3. Estrutura Populacional.....	40
4.1.4. Análises Demográficas.....	49
5. DISCUSSÃO	53
5.1. Estrutura Populacional e História Demográfica de <i>Drosophila serido</i>	53
6. CONCLUSÕES	62
7. REFERÊNCIAS	63

1. INTRODUÇÃO

1.1. As espécies cactófilas do “cluster” *Drosophila buzzatii* (grupo *repleta*, gênero *Drosophila*)

Na América do Sul, alguns grupos do gênero *Drosophila* possuem associação ecológica obrigatória com espécies da família Cactaceae, por utilizarem seus tecidos vegetais necrosados como sítio de desenvolvimento larval (Manfrin & Sene, 2006). Entre esses, no grupo *D. repleta*, uma radiação forma o “cluster” *Drosophila buzzatii*, que atualmente compreende sete espécies: *D. buzzatii* Patterson & Wheeler, 1942, que ocorre por toda a área de distribuição do “cluster”; *D. koepferae* Fontdevila & Wasserman, 1988, que ocorre no bioma Chaco; *D. antonietae* Tidon-Sklorz & Sene, 2001, que ocorre nas áreas da bacia dos rios Paraná-Paraguai; *D. serido* Vilela & Sene, 1977 que ocorre no bioma Caatinga e em restingas da Mata Atlântica e *D. seriema* Tidon-Sklorz & Sene, 1995a, *D. gouveai* Tidon-Sklorz & Sene, 2001 e *D. borborema* Vilela & Sene, 1977, que ocorrem em enclaves de vegetação aberta e campos rupestres no bioma Cerrado e pelos biomas Caatinga e do litoral do bioma Mata Atlântica (Manfrin & Sene, 2006).

O “cluster” *Drosophila buzzatii* é considerado um grupo monofilético, como foi sugerido por análises de marcadores cromossômicos (Ruiz & Wasserman, 1993), de genes nucleares (Franco *et al.*, 2010; Santos *et al.*, 2009) e citoplasmáticos (Manfrin *et al.*, 2001). Entretanto, as topologias das diferentes inferências filogenéticas realizadas apresentam incongruências, o que pode ser uma consequência das dificuldades intrínsecas associadas à própria reconstrução filogenética molecular de espécies com pouco tempo relativo de divergência em sua história evolutiva (Machado & Hey, 2003), como é o caso das populações desse “cluster”, cujos eventos cladogenéticos parecem ter ocorrido majoritariamente durante o Quaternário (Fransak, 2011; Franco & Manfrin, 2013; Barrios-Leal *et al.*, 2019).

1.2. *Drosophila serido* Vilela & Sene, 1977

A espécie *Drosophila serido* possui uma ampla distribuição geográfica no Neotrópico, ocorrendo no bioma Caatinga, em populações isoladas nos estados de Goiás, Minas Gerais e em enclaves de vegetação xerófila no litoral Atlântico. As populações dessa espécie possuem uma inversão cromossômica fixa $2x7$ (Wasserman & Richardson, 1987) e edeago

morfotipo “A” (Silva & Sene, 1991), embora exista variação morfológica ao longo de sua distribuição geográfica (Franco *et al.*, 2008).

Essa espécie também está associada a várias espécies de cactos, como *Pilosocereus piauhyensis* (Gürke) Byles & G. D. Rowley, 1957, *Cereus hildmannianus* K. Schumann, 1876, *C. fernambucensis* Lemaire, 1839, *Opuntia ficus-indica* (Linnaeus) P. Miller, 1768 e *O. monacantha* Haworth, 1819 (Manfrin & Sene, 2006). Análises de hidrocarbonetos da cutícula (Oliveira *et al.*, 2011), inversões cromossômicas polimórficas (Tosi & Sene, 1989; Ruiz *et al.*, 2000), diversidade haplotípica mitocondrial (Manfrin *et al.*, 2001; Morales, 2005; Franco & Manfrin, 2013) e análises cariotípicas (Baimai *et al.*, 1983) sugerem a existência de dois grupos populacionais em *D. serido*: as populações do Nordeste e as do litoral Sudeste e Sul brasileiros.

As populações do Nordeste de *D. serido*, predominantemente do bioma Caatinga, mas também no litoral dos estados da região Nordeste, possuem, sobre a inversão $2x^7$, as inversões cromossômicas polimórficas $2a^8$, $2b^8$, $2c^8$ e $2d^8$ (Tosi & Sene, 1989) e todos os marcadores moleculares utilizados até o momento indicam uma homogeneidade genética para esse grupo populacional (Baimai *et al.*, 1983; Silva & Sene, 1991; Ruiz *et al.*, 2000; Franco *et al.*, 2008).

Ao longo do litoral Sudeste e Sul, estudos com diversos marcadores sugerem que as populações dessa região são politípicas. As populações do litoral de São Paulo e Santa Catarina possuem, sobre a inversão $2x^7$, a inversão cromossômica fixa $2y^9$ e, sobre esta, as inversões $2x^8$ e $2w^8$. Essas duas últimas inversões foram observadas apenas nas populações de Florianópolis – SC e Guaratuba – SP (Ruiz *et al.*, 2000). Com relação as placas metafásicas, foram descritos cromossomos metafásicos do tipo III em indivíduos de Arraial do Cabo – RJ, do tipo IV em indivíduos de Peruíbe – SP e de um tipo diferente dos anteriores em Florianópolis – SC (Baimai *et al.*, 1983; Biffi *et al.*, 2001). Um exame da variação morfológica dos edeagos dessas populações também apontou a existência de dois grupos distintos no litoral Atlântico: um grupo contendo populações de Bertiooga – SP e Penha – SC e outro com populações de São Sebastião – SP e Arraial do Cabo – RJ (Franco *et al.*, 2008). Análises morfométricas das asas desses animais também corroboram a hipótese de estruturação nas populações dessa região (Esguícero, 2006).

1.3. Filogeografia de *Drosophila serido* e Marcadores Moleculares

Quando tratamos de táxons amplamente distribuídos, a abordagem filogeográfica pode ser utilizada para analisar a estruturação genética geográfica das populações que se deseja analisar (Avice, 2000; Zanella, 2011). A partir de métodos estatísticos, é possível fazer a distinção entre eventos vicariantes causados por barreiras físicas, que segregam uma população supostamente contínua e estável, e de fenômenos como aqueles influenciados pela dinâmica das flutuações climáticas do Quaternário por meio de testes de neutralidade e de relógios moleculares (Martins & Domingues, 2017).

A história evolutiva da América do Sul está intimamente conectada a uma sucessão de eventos geológicos que, nos últimos 250 milhões de anos, foram responsáveis por grandes modificações tanto nas configurações do continente e das bacias oceânicas quanto no clima e na evolução das espécies da região (Lavina & Fauth, 2011). A combinação desses eventos históricos possibilitou grande diversificação da atual cobertura florestal do continente, que compreende, em suas porções tropical e subtropical, extensos biomas que são delimitados majoritariamente por diferenças em seus níveis pluviométricos, englobando desde regiões semiáridas até florestas úmidas (Hueck, 1972; Zanella, 2011).

As regiões semiáridas são caracterizadas por florestas e formações abertas tropicais sazonalmente secas, também chamadas de florestas tropicais sazonalmente secas (FTSS). Possuem uma ampla faixa de clima com chuvas sazonais e de restrição hídrica, com ocorrência tanto de vegetação aberta quanto herbácea, do tipo savana, de vegetação xerófila e com florestas estacionais decíduas ou semidecíduas (Zanella, 2011). Essa vegetação heterogênea predomina por uma área que se estende em um eixo diagonal do nordeste ao sudoeste da América do Sul, denominada de “diagonal seca” (**Figura 1**), e que se distribui de modo descontínuo pelas províncias biogeográficas da Caatinga, do Cerrado e do Gran Chaco (Ab’Sáber, 1977; Prado & Gibbs, 1993).

Ao longo de milhões de anos, a restrita associação ecológica de drosófilas neotropicais com cactáceas pode ter tido um papel fundamental na diversificação e irradiação desses animais, tanto em aspectos ecológico-adaptativos quanto de distribuição histórica. Sob o ponto de vista ecológico-adaptativo, a diversificação das várias populações desses insetos pode ter sido influenciada pela composição química dos cactos utilizados e também pela expressão de genes relacionados aos processos de desintoxicação de compostos secundários dos cactos no organismo do animal (Matzkin *et al.*, 2006; Matzkin, 2008;

2012, Rane *et al.*, 2018; Carreira *et al.*, 2022; Moreyra *et al.*, 2022). Sob o ponto de vista de sua distribuição histórica, é possível que eventos subsequentes e recorrentes como os de expansão e retração das formações vegetais abertas durante paleoeventos climáticos possam ter tido influência significativa na distribuição das populações de cactos (Taylor & Zappi, 2004; Bonatelli *et al.*, 2014; Franco *et al.*, 2017; Lavor *et al.*, 2018; Silva *et al.*, 2018; Amaral *et al.*, 2021) e, por consequência, na distribuição das populações de *Drosophila* (Morales, 2005; Moraes *et al.*, 2009; Santos, 2011; Franco & Manfrin, 2013; Franco *et al.*, 2017; Barrios-Leal *et al.*, 2018; Barrios-Leal *et al.*, 2019; Rane *et al.*, 2019; Koser *et al.*, 2020; Barrios-Leal *et al.*, 2021; Barrios-Leal *et al.*, 2021).



Figura 1. A descontínua “diagonal seca” da América do Sul, com as delimitações das regiões da Caatinga, Cerrado e Gran Chaco. Modificado de Prado & Gibbs (1993).

A realização de testes estatísticos em análises filogenéticas só é possível por meio da obtenção de dados de sequências genéticas do modelo de estudo (Avice, 2000). A análise de diferentes *loci*, contudo, se faz necessária para melhor compreensão da história evolutiva e “status” taxonômico de espécies. Genealogias de genes inferidas a partir de um único gene podem resultar em conclusões equivocadas, pois cada reconstrução filogenética representa somente um ponto no espaço dentre todas as genealogias

possíveis. Assim, a depender da história natural da população e das perguntas que se busca responder em um estudo, a árvore do gene pode ser inferida de forma correta, mas não corresponderá necessariamente à árvore da espécie (Maddison, 1997; Martins & Domingues, 2017). Quando é possível unir as informações de mais genes ou até mesmo de genomas num estudo filogenético, aumentam-se as chances de se recuperar topologias que sejam mais próximas à árvore que representa a real história evolutiva da espécie (Futuyma & Kirkpatrick, 2017).

A partir dessas premissas, o presente estudo busca avaliar a estrutura genética populacional de elementos da espécie *Drosophila serido* ao longo de sua distribuição no bioma Caatinga e em algumas regiões do litoral Nordeste, Sudeste e Sul onde predomina o bioma Mata Atlântica. Desse modo, a análise filogeográfica de marcadores moleculares de um táxon cujas populações ocorrem nos dois biomas supracitados pode resultar em dados que permitam uma melhor resolução na reconstrução dos processos evolutivos e ecológico-adaptativos ocorridos durante a história evolutiva desses animais.

Para recuperar a história ecológica-evolutiva das populações de *D. serido* ao longo de sua distribuição nos diversos biomas do Brasil, utilizamos quatro marcadores moleculares. Os marcadores e suas características estão sumarizados nos itens a seguir.

1.3.1. Gene mitocondrial COI

O gene mitocondrial COI (citocromo C oxidase subunidade I) é frequentemente utilizado como marcador molecular em estudos filogenéticos e filogeográficos, pois apresenta algumas vantagens em relação a outros marcadores moleculares: ele é facilmente amplificado pela PCR (reação em cadeia da polimerase), é altamente conservado entre espécies relacionadas, mas suficientemente variável para permitir a diferenciação entre espécies e populações, e é encontrado em grande número de organismos (Hebert *et al.*, 2003).

O gene mitocondrial COI sofre forte seleção natural, tanto positiva (Castoe *et al.*, 2008) quanto purificadora (Ward & Holmes, 2007), em muitas espécies devido à sua função na cadeia de transporte de elétrons da mitocôndria, o que pode levar à sua evolução acelerada em diferentes linhagens (Avice, 2000).

Além disso, o gene mitocondrial COI tem sido amplamente utilizado em estudos de *barcoding* (identificação de espécies com base em sequências de DNA) e é uma ferramenta importante para a identificação rápida e precisa de espécies em diversas áreas da biologia, incluindo a biologia da conservação, a ecologia e a saúde pública (Hebert *et al.*, 2003).

1.3.2. Gene nuclear *kl-5* ligado ao cromossomo Y

Alguns autores mostraram que o cromossomo Y é capaz de adquirir novos genes, que podem ser recrutados via eventos de duplicação e/ou translocação (Skaletsky *et al.*, 2003; Koerich *et al.*, 2008). Em *Drosophila*, sabe-se que a maioria dos genes ligados ao cromossomo Y são produtos de aquisições autossômicas recentes e que, de maneira geral, o cromossomo Y sofre mais ganhos do que perdas genéticas. (Carvalho *et al.*, 2009). A ausência de recombinação também torna esse cromossomo essencialmente suscetível aos efeitos da deriva genética, que pode acabar por fixar alelos não-adaptativos a longo prazo (Gordo & Charlesworth, 2000; Dyer *et al.*, 2011).

O gene nuclear *kl-5* ligado ao cromossomo Y é um gene de cópia única que codifica uma proteína do tipo *dynein heavy chain*, um fator de fertilidade masculino essencial expresso durante a espermatogênese em *D. melanogaster* (Gepner & Hays, 1993). É sugerido que esse gene tenha sido adquirido independentemente ao menos três vezes pelo cromossomo Y no gênero *Drosophila*. Dyer e colaboradores (2010) identificaram uma possível terceira translocação histórica do gene *kl-5* do cromossomo Y para um gene autossômico no grupo testaceae em *Drosophila*.

Quando um gene como o *kl-5* se transloca de um cromossomo não-recombinante para um autossômico, ele se torna suscetível a outros mecanismos evolutivos além da seleção neutra. Dessa forma, é provável que os polimorfismos desse gene sejam recentes e que estejam relacionados a fatores reprodutivos que influenciam diretamente no *fitness* dos machos portadores do *kl-5* e, conseqüentemente, passível de seleção sexual (Chippindale & Rice, 2001; Koerich *et al.*, 2008; Carvalho, 2002). Nesse contexto, o gene *kl-5* pode se mostrar um marcador útil na compreensão dos possíveis processos de causaram a diferenciação das linhagens evolutivas observadas na espécie *D. serido*, já que também é provável que os polimorfismos desse gene sejam recentes e que tenham sua origem antes da aquisição do mesmo pelo cromossomo Y (Dyer *et al.*, 2011).

1.3.3. Genes nucleares autossômicos *GstD1* e α -esterase5

A composição bioquímica do tecido vegetal de cactáceas, que varia entre as espécies, é tóxica para muitos grupos de *Drosophila* que exploram esse vegetal como recurso (Frank & Fogleman, 1992; Matzkin, 2008). Assim, o entendimento das características bioquímicas desses vegetais e dos processos metabólicos de desintoxicação das moscas é essencial na determinação dos padrões de relação entre os cactos hospedeiros e as espécies que dele dependem (Frank & Fogleman, 1992).

Os processos de desintoxicação envolvem diferentes *loci* (Matzkin *et al.*, 2006). Estudos com *D. mojavensis* (Matzkin, 2008) e *D. mettleri* (Bono *et al.*, 2008) mostraram que as relações entre as linhagens evolutivas de moscas e os cactos hospedeiros podem ser inferidas por meio da análise de genes pertencentes às famílias glutatona S-transferase (*GstD1*) e citocromo P450, respectivamente. Santos *et al.* (2009) também observaram que o gene α -esterase5 (*$\alpha E5$*) é informativo a nível interespecífico em análises filogenéticas entre espécies do "cluster" *D. buzzatii*, obtendo topologias de genes com elevado suporte estatístico nos ramos. Ainda segundo Rane *et al.* (2018), esse gene pode sofrer influência de seleção positivas em populações de moscas cactófilas.

Como já mencionado, a espécie *D. serido* associa-se a diversas espécies de cactos ao utilizar seus tecidos vegetais necrosados como sítio de desenvolvimento larval (Manfrin & Sene, 2006). Sabe-se que o nível de endemismo de cactos na Caatinga é mais elevado do que nas regiões de restingas ao longo do litoral Atlântico (Taylor & Zappi, 2004). O uso de marcadores genéticos como os genes *GstD1* e *$\alpha E5$* em análises filogeográficas podem oferecer novos *insights* sobre os processos e padrões ecológico-adaptativos que moldaram a evolução das populações de *D. serido* da Caatinga e do litoral Atlântico, que pode ter sido influenciada pela estrita relação desses organismos com seus cactos hospedeiros.

2. OBJETIVOS

O presente trabalho tem como **objetivo principal** testar a hipótese de que a espécie cactófila *Drosophila serido* se estrutura em dois grupos populacionais distintos: um ocorrendo no bioma Caatinga (Grupo Nordeste) e o outro ocorrendo predominantemente nas regiões de restingas do litoral do sul da Bahia e dos litorais Sudeste e Sul do Brasil (Grupo Litoral).

Os **objetivos específicos** são:

- Analisar as sequências parciais do gene mitocondrial *COI* e dos genes nucleares *kl-5*, *GstD1* e *αE5* de indivíduos previamente coletados de *D. serido*;
- Avaliar a presença e, se for o caso, a quantidade de linhagens intraespecíficas em *D. serido* no Nordeste e no litoral Atlântico do Brasil por meio de análises genéticas e populacionais;
- Estabelecer hipóteses sobre a influência de eventos paleoclimáticos na diferenciação populacional de *D. serido*.

3. MATERIAL E MÉTODOS

3.1. Material Biológico

Sequências parciais dos genes nucleares *kl-5*, *GstD1*, *αE5* e mitocondrial COI foram obtidas a partir de amostras de DNA genômico de indivíduos de *Drosophila serido* em trabalhos anteriores (Morales, 2005; Franco *et al.*, 2010; Franco & Manfrin, 2013). Também foram para esse trabalho sequências disponíveis no banco de dados do *NCBI GenBank* (gene COI: JN124572.1 – JN124595.1 e JN124592.1 – JN124674.1). A **Tabela 1** sumariza todo o material biológico utilizado no presente estudo. Os *primers* senso e anti-senso para cada um dos cinco marcadores moleculares utilizados estão descritos na **Tabela 2**. A **Figura 2** exibe os pontos de coleta dos espécimes utilizados, seus respectivos códigos de localidade (assim como na **Tabela 1**), os biomas onde ocorriam e a hipótese inicial de dois agrupamentos populacionais.

Tabela 1. Populações amostradas de *Drosophila serido* em trabalhos anteriores.

Localidade	CL	CG (S, W)	Coletores	Data de Coleta	Nco1	Nkl-5	NGstD1	NαE5
Água Fria - GO	AF	-15.00, -47.87	Tidon, R.	-	06	04	04	05
Bertioga - SP	H49	-23.90, -46.10	Sene F.M., Monteiro S.G., Ruiz A.	10 a 18-V-1995	04	04	04	03
Ilha de Alcatrazes - SP	ILHA	-24.10, -45.69	-	-	04	06	06	06
Penha - SC	J52	-27.75, -48.66	-	-	07	05	01	05
Milagres - BA	J92	-11.20, -39.90	Morales A.C., Kuhn G.C.S., Franco F.F., Brisson J.A.	23-III-2002	06	06	04	05
Manoel Vitorino - BA	J93	-14.14, -40.24	Morales A.C., Kuhn G.C.S., Franco F.F., Brisson J.A.	23-III-2002	06	06	06	05
Itaberaba - BA	J95	-12.56, -40.31	Morales A.C., Kuhn G.C.S., Franco F.F., Brisson J.A.	23-III-2002	04	02	03	04
Mucuri - BA (1)	MUC1	-18.09, -33.55	-	-	01	-	-	-
Mucuri - BA (2)	MUC2	-17.99, -39.49	-	-	04	06	06	06
Camboriú - SC	N10	-27.05, -48.59	Morales A.C., Franco F.F., Silva Bernardi E.C.C.	24 a 31-III- 2003	02	03	-	03
Paúba - SP	N19	-23.80, -45.55	Morales A.C., Franco F.F.	05 a 09-X-2003	07	06	06	06
Guaratuba - SP	N2	-23.85, -46.13	Manfrin M.H., Prado N.	VIII-2002	06	03	04	06
Arraial do Cabo - RJ	N20	-23.00, -42.00	Morales A.C., Franco F.F., Silva Bernardi E.C.C.	14 a 20-III- 2004	06	06	06	06
Macaé - RJ	N22	-22.32, -41.82	-	-	05	03	03	05
Praia da Jureia - SP	N29	-23.76, -45.79	-	-	07	03	02	05

Itaúnas - ES	N31	-18.41, -39.70	Sene F.M., Franco F.F., Almeida J.G.	04 a 12-VII-2005	06*	02	04	06
Cabrália - BA	N35	-16.31, -39.01	Sene F.M., Franco F.F., Almeida J.G.	04 a 12-VII-2005	05*	05	05	05
Morro da Barrinha - BA	N37	-09.90, -40.30	Franco F.F., Oliveira C., Esgúicero A.	05 a 14-VIII-2008	10*	06	04	05
10km S de Morro do Chapéu - BA	N40	-11.65, -41.29	Franco F.F., Oliveira C., Esgúicero A.	05 a 14-VIII-2008	05*	04	05	04
Cachoeira Ferro Doido - BA	N42	-11.63, -41.00	Franco F.F., Oliveira C., Esgúicero A.	05 a 14-VIII-2008	10*	06	06	06
Irecê - BA	N43	-11.22, -41.95	Franco F.F., Oliveira C., Esgúicero A.	05 a 14-VIII-2008	09*	06	06	06
Mucugê - BA	N45	-13.00, -41.40	Franco F.F., Oliveira C., Esgúicero A.	05 a 14-VIII-2008	05*	04	03	04
Andaraí - BA	N46	-12.86, -41.31	Franco F.F., Oliveira C., Esgúicero A.	05 a 14-VIII-2008	06*	06	05	05
Serra do Cipó - MG	N56	-19.31, -43.61	Franco F.F., Esgúicero A.	19 a 24-XI-2006	08*	06	06	06
Salvador - BA	N65	-12.93, -39.32	Franco F.F.	07 a 09-X-2007	10*	05	06	06
Exu - PE	N67	-07.65, -39.79	Franco F.F., Santos M.H.S.	16 a 23-II-2008	02*	02	-	01
Pocinhos - PB	N68	-07.17, -36.05	Franco F.F., Santos M.H.S.	16 a 23-II-2008	04*	06	06	06
Junco do Serido - PB	N70	-07.00, -36.71	Franco F.F., Santos M.H.S.	16 a 23-II-2008	10*	06	06	06
Cabedelo - PB	N71	-07.03, -34.83	Franco F.F., Santos M.H.S.	16 a 23-II-2008	07*	06	06	05
Florianópolis - SC	N88	-27.56, -48.56	Fransak F., Mateus R.P., Bizzo L.	25 a 29-IX-2008	01	03	02	06
Gov. Celso Ramos - SC	N9	-27.37, -48.53	Morales A.C., Franco F.F., Silva Bernardi E.C.C.	24 a 31-III-2003	03	03	-	03

Riacho - BA	R27	-09.61, -38.22	Lavagnini T.C., Leal D.Y.B., Manfrin M.H., Mateus R.P.	15 a 22-IX-2012	04	05	05	05
Monte Alegre de Sergipe - SE	R28	-10.13, -37.54	Lavagnini T.C., Leal D.Y.B., Manfrin M.H., Mateus R.P.	15 a 22-IX-2012	06	05	03	05
Porto das Cabras - SE	R29	-10.80, -39.63	Lavagnini T.C., Leal D.Y.B., Manfrin M.H., Mateus R.P.	15 a 22-IX-2012	05	04	01	04
Penedo - AL	R30	-10.32, -36.52	Lavagnini T.C., Leal D.Y.B., Manfrin M.H., Mateus R.P.	15 a 22-IX-2012	06	06	06	06
Pacatuba - SE	R31	-10.46, -36.63	Lavagnini T.C., Leal D.Y.B., Manfrin M.H., Mateus R.P.	15 a 22-IX-2012	06	06	06	05
Costa Azul - BA	R32	-11.61, -37.46	Lavagnini T.C., Leal D.Y.B., Manfrin M.H., Mateus R.P.	15 a 22-IX-2012	06	04	03	06
Rodolfo Fernandes - RN	R33	-05.82, -37.99	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C., Lavagnini M.C.	23 a 30-VI-2012	01	01	01	01
Jucurutu - RN	R34	-05.95, -37.04	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C., Lavagnini M.C.	23 a 30-VI-2012	01	01	01	01
Tabuleiro do Norte - CE	R35	-05.28, -38.16	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C., Lavagnini M.C.	23 a 30-VI-2012	02	02	02	02
Russas - CE	R36	-04.85, -38.04	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C., Lavagnini M.C.	23 a 30-VI-2012	06	06	04	04

Lajes Pintadas - RN	R37	-06.17, -36.11	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C., Lavagnini M.C.	23 a 30-VI-2012	08	06	06	06
Bom Jesus - RN	R38	-06.02, -35.65	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C., Lavagnini M.C.	23 a 30-VI-2012	06	06	06	05
Maxaranguape - RN (1)	R39	-05.54, -35.27	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C., Lavagnini M.C.	23 a 30-VI-2012	05	06	06	06
Almenara-MG	R56	-16.00, -40.95	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C.	19 a 28-IX-2013	10	09	08	06
Jequitinhonha - MG	R59	-16.52, -41.30	Lavagnini T.C., Leal D.Y.B., Lavagnini A.C.	19 a 28-IX-2013	01	01	01	-

CL: códigos das localidades; CG: coordenadas geográficas (grau-decimal); NCOI: número de amostras sequenciadas para o gene mitocondrial COI; *N α E5*: número de amostras sequenciadas para o gene nuclear *α -esterase5*; *NGstD1*: número de amostras sequenciadas para o gene nuclear *GstD1*; *Nkl-5*: número de amostras sequenciadas para o gene nuclear *kl-5*; *: localidades em que foram utilizadas sequências disponíveis no *NCBI GenBank*; -: ausência de dados.

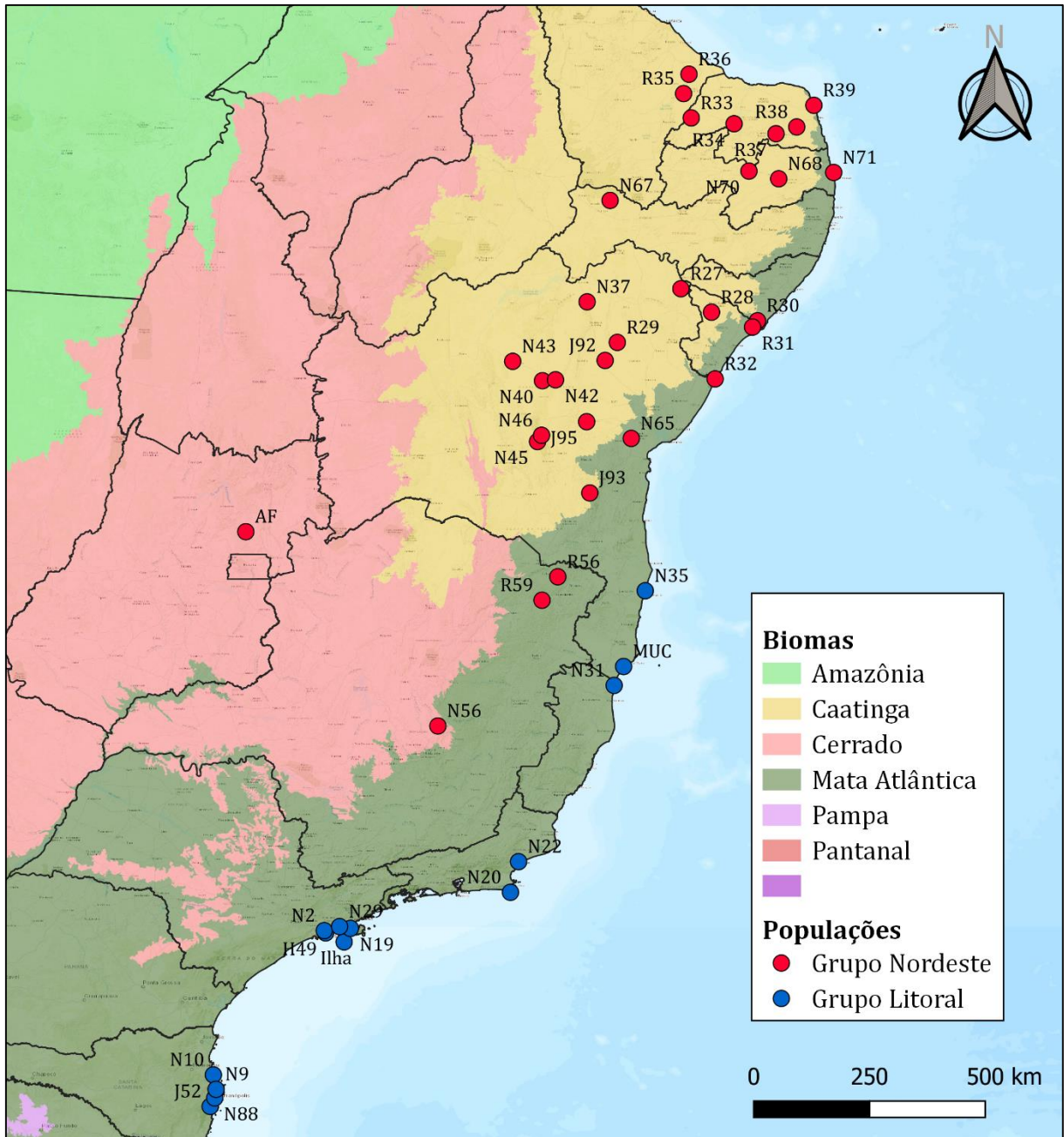


Figura 2. Mapa da costa leste do Brasil contendo as localidades amostradas de *Drosophila serido*, seus códigos de referência (assim como na **Tabela 1**), respectivos biomas (dados do IBGE de 2019) e a hipótese inicial de dois agrupamentos populacionais. Mapa construído no programa QGIS v3.22.6.

Tabela 2. Primers senso e anti-senso e condições da PCR para cada fragmento de gene.

Marcador (referência)	primer senso	primer anti-senso
COI (Simon <i>et al.</i> , 1994)	<i>TY-J-1460</i> 5' TACAATTTATCGCCTAAACTTCAGCC 3'	<i>C1-N-2191</i> 5' CCCGGTAAAATTTAAAATATAAACTTC 3'
<i>kl-5</i> (Dyer <i>et al.</i> , 2011)	<i>K15-F8</i> 5' KTTYGARYTRCAAGGKCCAGATCC 3'	<i>K15-R6</i> 5' GCSGGCCAYTCRAATATCCAMAC 3'
<i>GstD1</i> (desenhados em trabalho anterior*)	<i>GstD1deg-F</i> 5' CCCGAGTTCGTGAAGATCAAYCCNCA 3'	<i>GstD1deg-R</i> 5' TTCTCGTACCACCTGTTTCACGTTNKCR 3'
<i>αE5</i> (Santos <i>et al.</i> , 2009)	<i>Ester 2F</i> 5'TGGACTGAAGGACCAGGTTT 3'	<i>Ester CH1 R</i> 5' AGCCATGCCAGAAGATCCTA 3'

*: Trabalho não publicado.

3.2. Análises Moleculares

Para a realização das análises moleculares, foram utilizados alinhamentos previamente disponíveis, que resultaram em: 620pb de 249 indivíduos amostrados em 46 localidades para o gene COI; 1200pb de 207 indivíduos amostrados em 45 localidades para o gene *kl-5*; 380pb de 184 indivíduos amostrados em 44 localidades para o gene *GstD1* e, por fim, 690pb de 212 indivíduos amostrados em 43 localidades para o gene *αE5*.

3.2.1. Análises Filogenéticas

Modelos de Substituição Nucleotídica

Os **modelos de substituição nucleotídica** para cada alinhamento foram estimados com auxílio do programa jModelTest 2.1.10 (Darriba *et al.*, 2012), utilizando como parâmetro estatístico o *Akaike Information Criterion* (AIC), em que o modelo mais adequado é selecionado pelo menor valor de AIC. Esse critério indica o modelo de substituição nucleotídica que melhor se aproxima do processo real de evolução molecular dos fragmentos de genes com os quais se deseja trabalhar (Posada & Buckley, 2004).

Análise de Congruência Filogenética

A verificação da existência de informações congruentes entre os diferentes loci dos quatro marcadores utilizados foi realizada a partir de uma **análise de congruência filogenética** em toda a base de dados com o pacote “*ape*” no programa RStudio (R Core Team), que testa a congruência entre matrizes de distância (*Congruence among distance matrices* – CADM) utilizando a função `CADM.global` com 1.000 permutações e 95% de significância entre as sequências.

O coeficiente de concordância de Kendall (W) (Campbell *et al.*, 2011) que varia de 0 (incongruência completa) a 1 (congruência completa) foi levado em conta na verificação da congruência entre as quatro diferentes matrizes de distância obtidas pelo programa Arlequin v3.5.2.2 (Excoffier & Lischer, 2010). As diferentes matrizes foram analisadas par-a-par e em concatenação.

Reconstruções Filogenéticas e Estimativa do Tempo de Divergência Comum

Cinco hipóteses filogenéticas foram recuperadas: uma para cada marcador molecular e outra utilizando todas as quatro partições de cada marcador para reconstruir uma filogenia concatenada (*Species Tree* – Edwards *et al.*, 2007). As análises foram conduzidas no programa Beast2 v2.6.7 (Bouckaert *et al.*, 2019), que implementa o modelo de estatística Bayesiana (IB) para estimar filogenias. O pacote STACEY (Jones, 2016) foi utilizado para agrupar as sequências de indivíduos em suas respectivas populações (**Tabela 1**) e as análises levaram em conta os modelos de substituição nucleotídica estimados para cada marcador (**Tabela 3**). Foram utilizadas sequências da espécie *Drosophila buzzatii*, disponíveis no banco de dados do *NCBI GenBank* (gene *COI*: MH612367.1 – gene *GstD1*: MW805280.1 – gene α -*esterase5*: DQ204660.1) para a composição de um grupo-externo às sequências de *D. serido*.

Para a escolha do modelo de relógio molecular de cada partição, foram conduzidas análises utilizando todas as partições no programa Beast2 v2.6.7 (Bouckaert *et al.*, 2019), para conferência dos valores de seus respectivos parâmetros de *ucl.d.stdev* no programa Tracer v1.7.1 (Rambaut *et al.*, 2018) sob o modelo *Uncorrelated Relaxed Lognormal Clock* (Drummond *et al.*, 2006). De acordo com o manual do programa (Drummond & Rambaut, 2007), sob este modelo, se a estimativa de *ucl.d.stdev* possui um valor igual a zero considera-se que não há variação das taxas de substituição nos diferentes os ramos das topologias das sequências analisadas, fazendo com que os dados sejam adequados à hipótese de modelo de relógio estrito. Por outro lado, se esse parâmetro for maior do que 1.0, isso indica heterogeneidade das taxas de substituição dos ramos das linhagens analisadas, sendo recomendado o uso de um relógio molecular mais “relaxado” do que o relógio estrito para a execução das análises. O modelo de relógio estrito foi escolhido como parâmetro para as partições dos genes *COI*, *kl-5* e α -*esterase5*, ao passo em que o modelo *Relaxed Clock Lognormal* foi escolhido para o gene *GstD1*.

Na análise filogenética concatenada, o tempo de divergência comum para cada nó foi estimado utilizando as taxas mínima e máxima de substituições nucleotídicas do gene *COI*, estimadas para o gênero *Drosophila*, de 0.87% e 1.75% de substituições por milhão de anos, respectivamente (Barrios-Leal *et al.*, 2018). Esse intervalo de valores foi utilizado como um parâmetro de distribuição uniforme nas estimativas de tempo para o marcador *COI*, enquanto parâmetros com intervalos mais abrangentes (1.0×10^{-6} a 1.0) foram adotados para as estimativas de tempo das outras partições (tanto nessa reconstrução quanto nas individuais para os três marcadores nucleares), seguindo as premissas de Bouckaert (2015). O parâmetro do modelo coalescente de Yule (Yule, 1924) foi utilizado para estimativa de uma árvore filogenética única a partir de todas as partições. As análises foram conduzidas utilizando 200 milhões de iterações na Cadeia de Markov Monte Carlo (MCMC), com um *burn-in* de 25% e amostragem de árvores a cada 5.000 passos. As árvores consenso foram obtidas com o programa TreeAnnotator v2.6.7 (Bouckaert *et al.*, 2019) e foram visualizadas e editadas nos programas FigTree v1.4.4 (Rambaut, 2010) e Inkscape (Inkscape Project, 2020).

O poder estatístico das análises foi verificado por meio do valor das métricas de *Effective Sample Size* (ESS) e de probabilidades posteriores (*posterior*) exibidos no programa Tracer (Rambaut *et al.*, 2018). O índice ESS calculado pelo programa representa a estimativa do número médio de dados na distribuição posterior das genealogias amostradas a partir dos dados fornecidos e que garante que o número de réplicas na MCMC foi capaz de amostrar uma relativa diversidade dessas genealogias (Villanea *et al.*, 2020). Apesar de não haver um valor limite para a métrica ESS, valores menores do que 200 indicam baixa convergência de informação, ou seja, pouca diversidade de genealogias recuperadas, indicando que a análise não utilizou um número suficiente de réplicas na MCMC (Kuhner, 2009).

3.2.2. Estatística Descritiva

A composição da **estatística descritiva** dos grupos genéticos avaliados, contendo o número de sítios polimórficos (S), número de haplótipos (H/h), diversidade haplotípica (Hd), diversidade nucleotídica (π) e número médio de diferenças nucleotídicas (k), foi realizada com uso do programa DnaSP v6.12.03 (Rozas *et al.*, 2017).

3.2.3. Estrutura Populacional

Análise Discriminante dos Componentes Principais (ADCP)

A hipótese do presente estudo considera a existência de dois **grupos genéticos** distintos, com seus elementos definidos *a priori*, dentro das populações amostradas de *Drosophila serido*. Nessa etapa, contudo, optou-se por uma abordagem estatística onde o número de *clusters* (K) é inferido *a posteriori* para evitar qualquer tipo de viés atribuído à hipótese inicial. Desse modo, foi realizada uma Análise Discriminante dos Componentes Principais (ADCP) sobre a base de dados utilizando o pacote *adegenet* v2.1.3 (Jombart, 2008; Jombart & Collins, 2015) no programa RStudio. Nessa abordagem, os dados genotípicos são transformados por meio de uma análise de Componentes Principais (CPs) para obtenção de variáveis não-correlacionadas que, por sua vez, serão utilizadas como *input* para uma Análise Discriminante (AD). A AD busca maximizar o peso das variações entre grupos e a diminuição do peso das variações dentro dos grupos (Thia, 2022). A ADCP é comumente utilizada em estudos de diversidade genética e estrutura populacional, já que não necessita de suposições estabelecidas *a priori* e também pela relativa rapidez em processar uma grande quantidade de dados (Miller *et al.*, 2020).

Essa análise possui uma etapa interativa durante sua execução. Por meio da função `find.clusters()`, o usuário indica o número de CPs e de funções discriminantes a serem mantidos para o conjunto de dados. Em seguida, o algoritmo calcula o valor do *Bayesian Information Criterion* (BIC) para cada valor de K dentro de um intervalo pré-definido no arquivo de script da própria análise.

O valor de BIC é calculado automaticamente pelo programa *adegenet* e representa um equilíbrio entre a qualidade do ajuste do modelo e a complexidade do mesmo, sendo que um valor menor indica uma melhor escolha de K (Jombart & Collins, 2015). Assim, o número de *clusters* que minimiza o valor de BIC é geralmente escolhido como o número ótimo de *clusters* para representar a estruturação populacional nos dados. Esse valor pode ser frequentemente indicado por um “cotovelo” na curva de valores BIC em função de K (Jombart & Collins, 2015), ou seja, pelo um ponto de inflexão da curva.

Após o cálculo, um gráfico com os valores de BIC versus o número de possíveis *clusters* é gerado. O usuário pode, então, decidir a partir desse gráfico qual o valor de K que melhor representa os seus dados e, por fim, obter um novo gráfico de barras com a representação em cores das hierarquias inferidas. Para o conjunto de dados do presente trabalho, foi

definido no script da análise um intervalo de K de 1 a 8 para a inferência do número de *clusters* pelo pacote *adeigenet* v2.1.3 (Jombart, 2008; Jombart & Collins, 2015).

Redes Haplotípicas

Redes haplotípicas para todos os quatro alinhamentos de nucleotídeos individuais e concatenados foram produzidas com o programa Haplotype Viewer (<http://www.cibiv.at/~greg/haploviewer>) utilizando como *input* uma topologia inferida pelo método estatístico de Máxima Verossimilhança (*Maximum Likelihood*) gerada no programa MEGA X v10.2.4 (Kumar *et al.*, 2018) utilizando o método de *bootstrap* (Felsenstein, 1985) como teste de filogenia com 1.000 replicações.

Análise de Variância Molecular (AMOVA) e Teste de Mantel

A avaliação da estruturação genética obtida na abordagem anterior e escolha do valor de K também foram balizadas com o suporte estatístico da Análise de Variância Molecular (Excoffier *et al.*, 1992), implementada no programa Arlequin v3.5.2.2 (Excoffier & Lischer, 2010), com 10.000 permutações para cada população de cada marcador genético. Foram realizadas duas AMOVAs distintas, uma com a composição populacional definida *a priori* e outra com a composição definida estatisticamente pela ADCP no item anterior, a fim de se analisar comparativamente os índices obtidos em cada uma.

Para testar a hipótese de isolamento por distância (*Isolation by Distance* – Wright, 1943), foi utilizado o **Teste de Mantel** no programa Arlequin, que avalia a existência de correlação entre duas ou mais matrizes de dados. Nesse caso, entre os dados de distância genética e os de distância geográfica das populações de *D. serido* definidas de acordo com a hipótese inicial de estruturação (“Grupo Nordeste” e “Grupo Litoral”). As matrizes de distância genética, F_{st} par-a-par, foram calculadas pelo programa Arlequin. As matrizes de distância geográfica par-a-par entre as populações de cada grupo foram geradas pelo programa Geographic Distance Matrix Generator v1.2.3 (Ersts, 2006) e, posteriormente, linearizadas em escala logarítmica (Slatkin, 1993; Rousset, 1997).

Pelo fato de o Teste de Mantel ser altamente permissível ao erro Tipo 1, o coeficiente de correlação (r^2) foi considerado como significativo apenas quando o valor-p era muito menor ($p \leq 0,001$) do que 0,05 (Oden & Sokal, 1992; Diniz-Filho *et al.*, 2013).

3.2.4. Análises Demográficas

Testes de Neutralidade e *Mismatch Distribution*

Desvios do modelo de neutralidade nucleotídica (Kimura, 1968) foram verificados por meio dos testes de neutralidade D de Tajima (Tajima, 1989), F_s de Fu (Fu, 1997) e pela análise de *mismatch distribution* (Rogers & Harpending 1992; Harpending *et al.*, 1993), que mostra a relação da frequência das diferenças nucleotídicas entre pares de sequências de DNA. Esses testes avaliam se a distribuição de alelos em uma população é consistente com o que seria esperado sob um modelo de equilíbrio neutro, e são úteis para detectar desvios significativos desse modelo, que podem indicar eventos históricos como expansões, gargalos ou estabilidade populacionais, indícios de diferentes tipos de seleção natural ou migração (Nielsen, 2001).

Os índices D de Tajima e F_s de Fu foram estimados com auxílio do programa DnaSP v6 (Rozas *et al.*, 2017) utilizando 1.000 *bootstraps*. A análise de *mismatch distribution* foi conduzida no programa Arlequin v3.5.2.2 (Excoffier & Lischer, 2010) assumindo como premissa o modelo de expansão demográfica súbita e utilizando 1.000 *bootstraps*. Todos os testes levaram em consideração os agrupamentos inferidos anteriormente por meio da ADCP.

Extended Bayesian Skyline Plot (EBSP)

Os eventos de flutuação do tamanho populacional foram estimados e representados graficamente em função do tempo por meio da análise multilocus *Extended Bayesian Skyline Plot* (EBSP) implementada no programa Beast2 v2.6.7 (Bouckaert *et al.*, 2019), utilizando todas as quatro partições genéticas de *D. serido*. Os modelos de substituição nucleotídica e os demais parâmetros selecionados no programa BEAUti v2.6.7 (Bouckaert *et al.*, 2019) foram os mesmos utilizados na reconstrução filogenética (*species tree*).

Para cada agrupamento inferido, foi realizada uma análise única com 100 milhões de passos na Cadeia de Markov Monte Carlo (MCMC), com um *burn-in* de 25% e amostragem de árvores a cada 2.000 passos. O arquivo .log originado teve seus valores de *Effective Sample Size* (ESS) e de probabilidades posteriores (*posterior*) verificados no programa Tracer (Rambaut *et al.*, 2018). Os gráficos de tamanho populacional em função do tempo, para cada população analisada, foram gerados no programa RStudio a partir do script “plotEBSP.R” (Heled & Drummond, 2008).

4. RESULTADOS

4.1. Análises Moleculares

4.1.1. Análises Filogenéticas

Modelos de Substituição Nucleotídica

A **Tabela 3** exibe os modelos de substituição nucleotídica estimados pelo programa jModelTest 2.1.10 (Darriba *et al.*, 2012). Na ausência dos modelos estimados nos programas utilizados, modelos com o valor de AIC mais próximos do modelo original estimado foram escolhidos para as análises.

Tabela 3. Marcadores moleculares analisados e seus respectivos modelos de substituição nucleotídica estimados no programa jModelTest.

Marcador	Modelo (AIC)	Modelo utilizado no Beast2
<i>COI</i>	TVM+I	HKY+I
<i>kl-5</i>	TPM1uf+I	HKY+I
<i>Gstd1</i>	TIM2+G	GTR+G
<i>αE5</i>	SYM+I+G	GTR+I

Análise de Congruência Filogenética

Os coeficientes de concordância de Kendall (W) (Campbell *et al.*, 2011) para cada matriz de distância dos marcadores utilizados estão compilados na **Tabela 4**.

Tabela 4. Coeficientes de concordância de Kendall (W) entre cada matriz distância par-a-par de cada marcador molecular analisados de indivíduos de *Drosophila serido*.

W	<i>COI</i>	<i>kl-5</i>	<i>GstD1</i>	<i>αE5</i>
<i>COI</i>	-	0,6187061	0,7007075	0,6400465
<i>kl-5</i>	0,6187061	-	0,6084851	0,6898947
<i>GstD1</i>	0,7007075	0,6084851	-	0,7354295
<i>αE5</i>	0,6400465	0,6898947	0,7354295	-

A análise de congruência entre todas as matrizes concatenadas resultou num valor de $W = 0,5057118$, o que Landis & Koch (1977) definem em seu trabalho como uma “concordância moderada” (“ $0.40 \leq W < 0.60$ - *Moderate agreement*”).

Reconstruções Filogenéticas e Estimativa do Tempo de Divergência Comum

O sinal filogenético recuperado pela hipótese filogenética obtida a partir do marcador COI (**Figura 3**), com baixos valores de suporte nos ramos, mostra uma história onde as populações do litoral de Santa Catarina são evolutivamente mais próximas das populações da Caatinga, com as duas pertencendo a um clado único, que é irmão do clado formado pelas populações dos litorais sul da Bahia e do Sudeste. Além disso, esses dois cladros compõem um clado único, que é irmão da população amostrada na localidade da Serra do Cipó – MG (N56).

A hipótese filogenética obtida a partir do marcador *kl-5* (**Figura 4**), com baixos valores de suporte nos ramos, recupera dois grandes cladros irmãos. Um deles é formado por todas as populações dos litorais do sul da Bahia, Espírito Santo, Rio de Janeiro e algumas populações do litoral de São Paulo com exceção das populações de Bertioga – SP (H49), Ilha de Alcatrazes – SP (ILHA), Guaratuba – SP (N2) e Paúba – SP (N19). O outro é formado majoritariamente por populações da Caatinga e do litoral do Nordeste, mas com a adição das populações do litoral de São Paulo mencionadas como ausentes no clado anterior.

As hipóteses filogenéticas obtidas a partir dos marcadores *GstD1* (**Figura 5**) e *αE5* (**Figura 6**), ambas com valores elevados de suporte nos ramos, recuperam dois grandes cladros irmãos, um composto pelas populações associadas ao Grupo Nordeste e o outro composto pelas populações associadas ao Grupo Litoral, o que é consistente com a hipótese inicial de estruturação populacional em *D. serido*. A filogenia consenso (**Figura 7**), por fim, recupera um cladograma com as mesmas topologias dos marcadores *GstD1* e *αE5*, indicando que a árvore de espécies estimada com os sinais filogenéticos de todos os quatro marcadores utilizados apresenta uma diferenciação evolutiva entre dois conjuntos populacionais que é compatível com a hipótese inicial do presente trabalho.

a) Gene mitocondrial COI

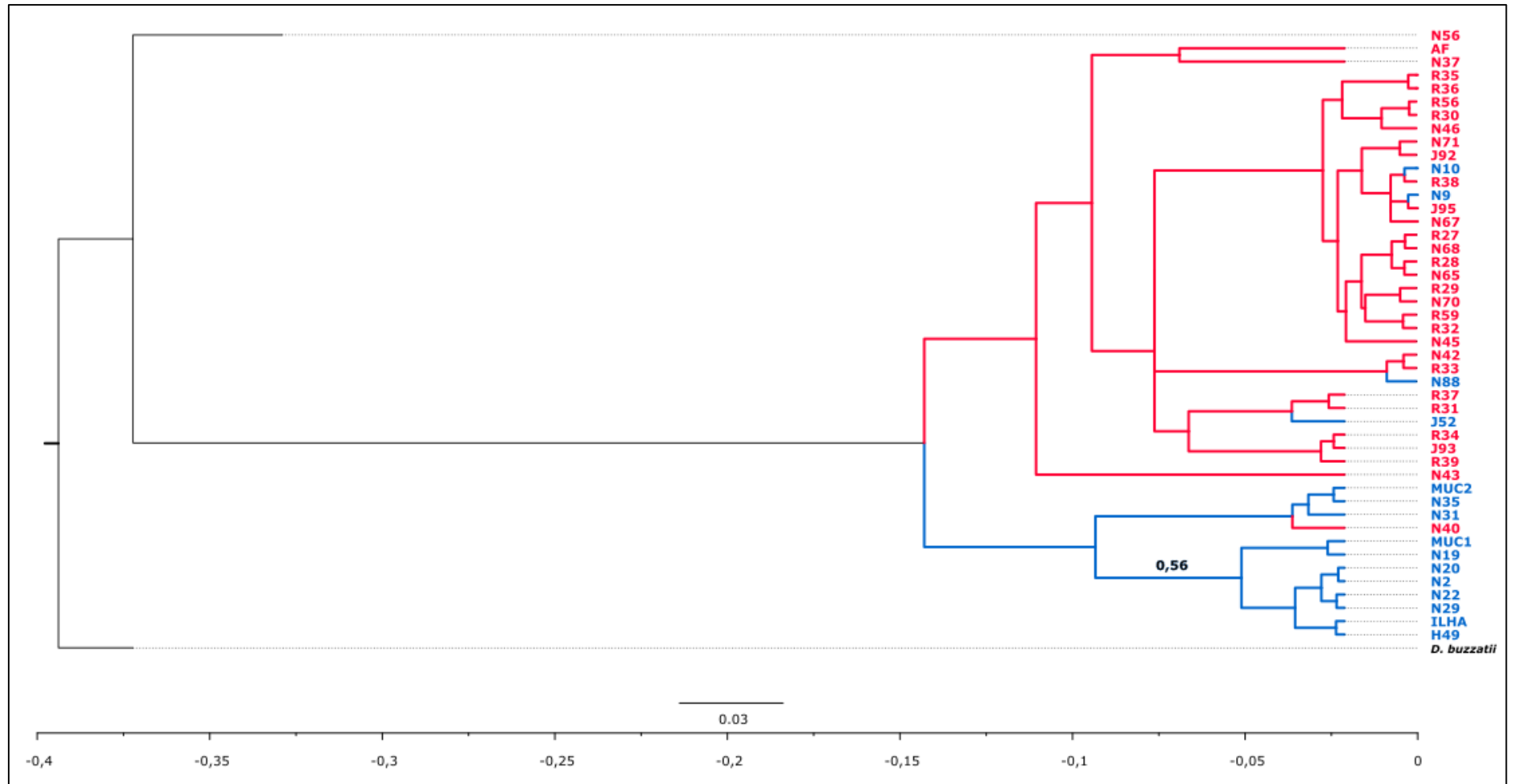


Figura 3. Reconstrução filogenética de *D. serido* baseada no gene COI. As probabilidades posteriores (>0,50) se encontram na parte superior dos ramos. As cores vermelho e azul representam, respectivamente, as populações do Grupo Nordeste e do Grupo Litoral definidos na hipótese inicial. Os valores negativos da barra inferior, lida da direita para a esquerda, representam o tempo em milhões de anos a partir do presente (zero).

b) Gene nuclear *kl-5*

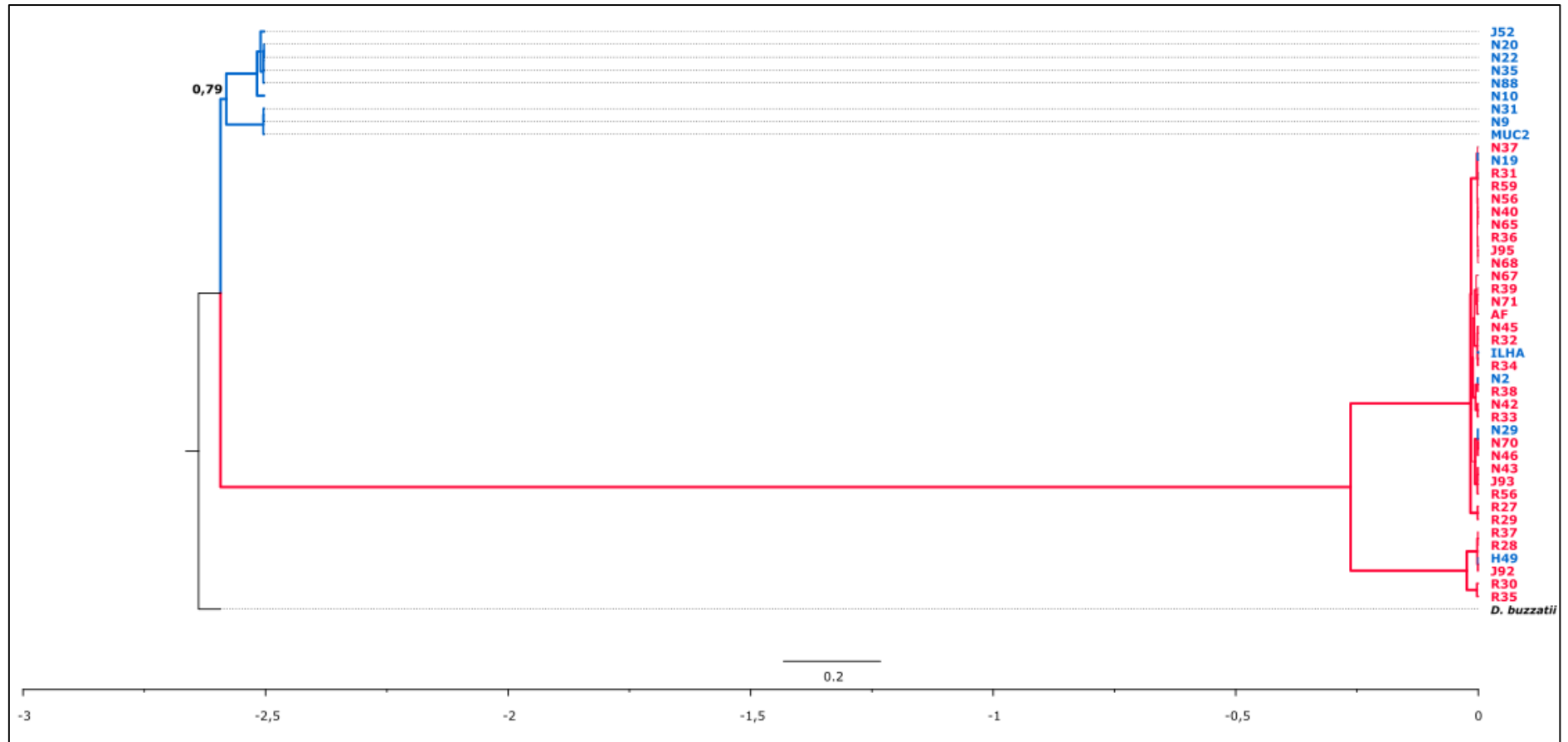


Figura 4. Reconstrução filogenética de *D. serido* baseada no gene *kl-5*. As probabilidades posteriores ($>0,50$) se encontram na parte superior dos ramos. As cores vermelho e azul representam, respectivamente, as populações do Grupo Nordeste e do Grupo Litoral definidos na hipótese inicial. Os valores negativos da barra inferior, lida da direita para a esquerda, representam o tempo em milhões de anos a partir do presente (zero).

c) Gene nuclear autossômico *GstD1*

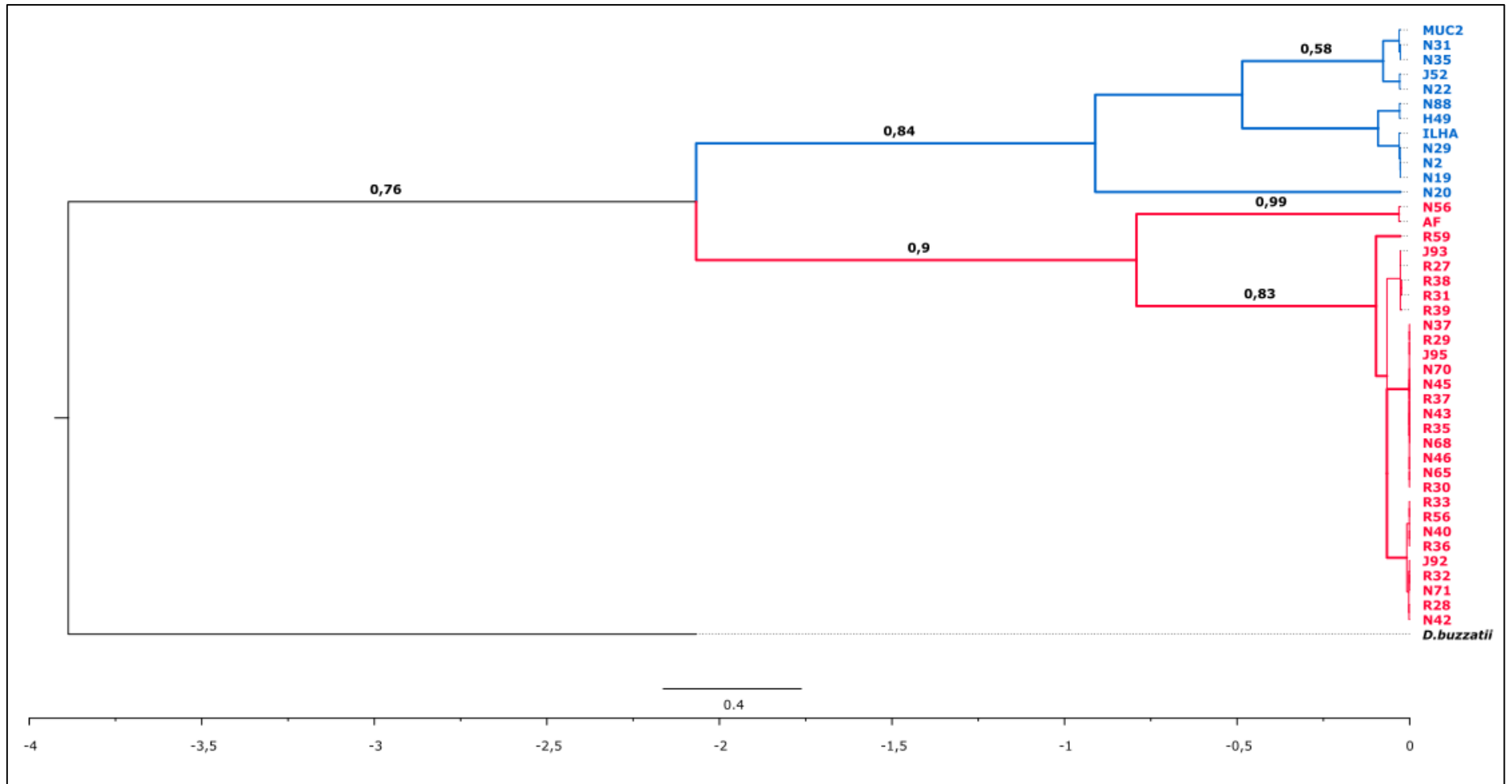


Figura 5. Reconstrução filogenética de *D. serido* baseada no gene *GstD1*. As probabilidades posteriores (>0,50) se encontram na parte superior dos ramos. As cores vermelho e azul representam, respectivamente, as populações do Grupo Nordeste e do Grupo Litoral definidos na hipótese inicial. Os valores negativos da barra inferior, lida da direita para a esquerda, representam o tempo em milhões de anos a partir do presente (zero).

d) Gene nuclear autossômico α -E5

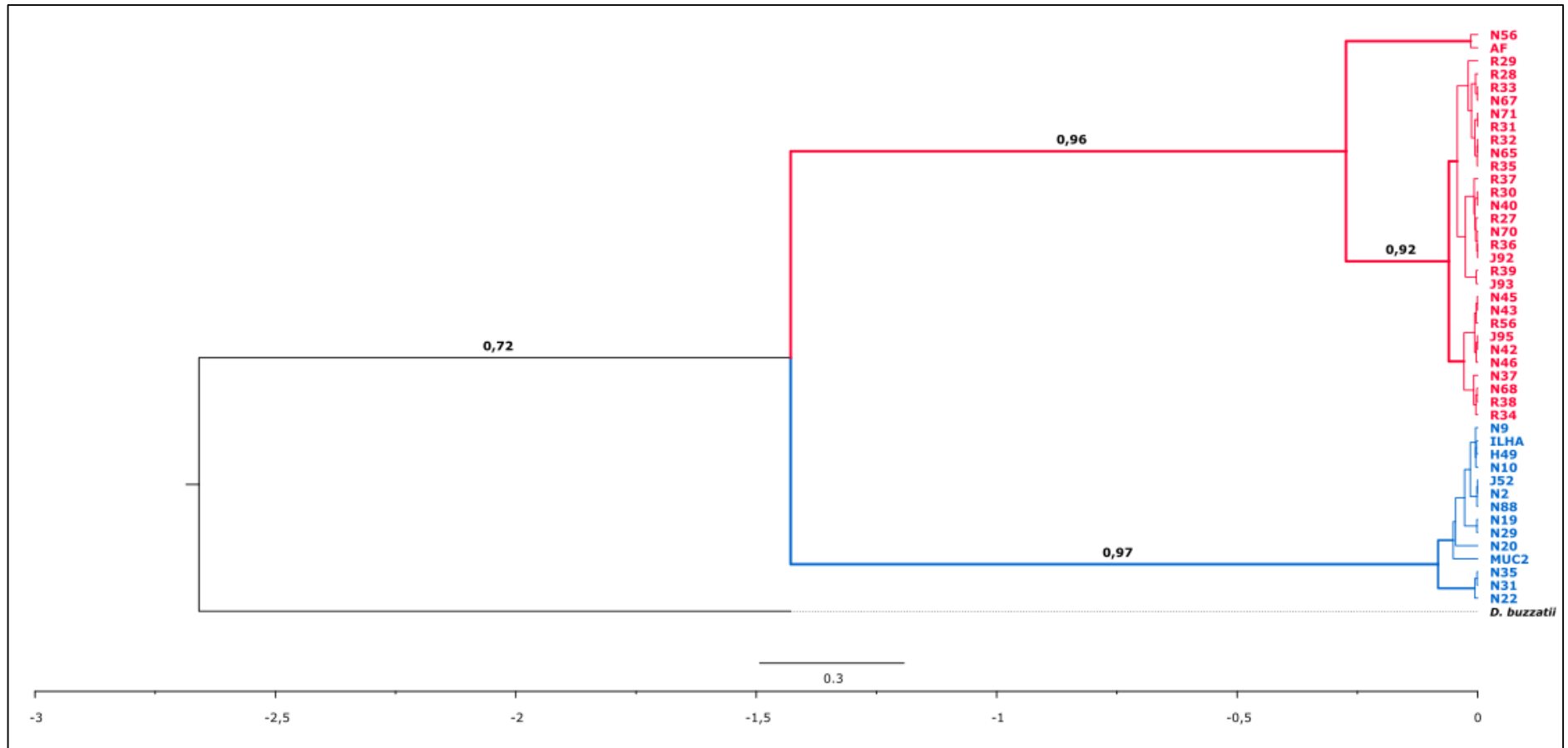


Figura 6. Reconstrução filogenética de *D. serido* baseada no gene α -E5. As probabilidades posteriores (>0,50) se encontram na parte superior dos ramos. As cores vermelho e azul representam, respectivamente, as populações do Grupo Nordeste e do Grupo Litoral definidos na hipótese inicial. Os valores negativos da barra inferior, lida da direita para a esquerda, representam o tempo em milhões de anos a partir do presente (zero).

e) Árvore Consenso

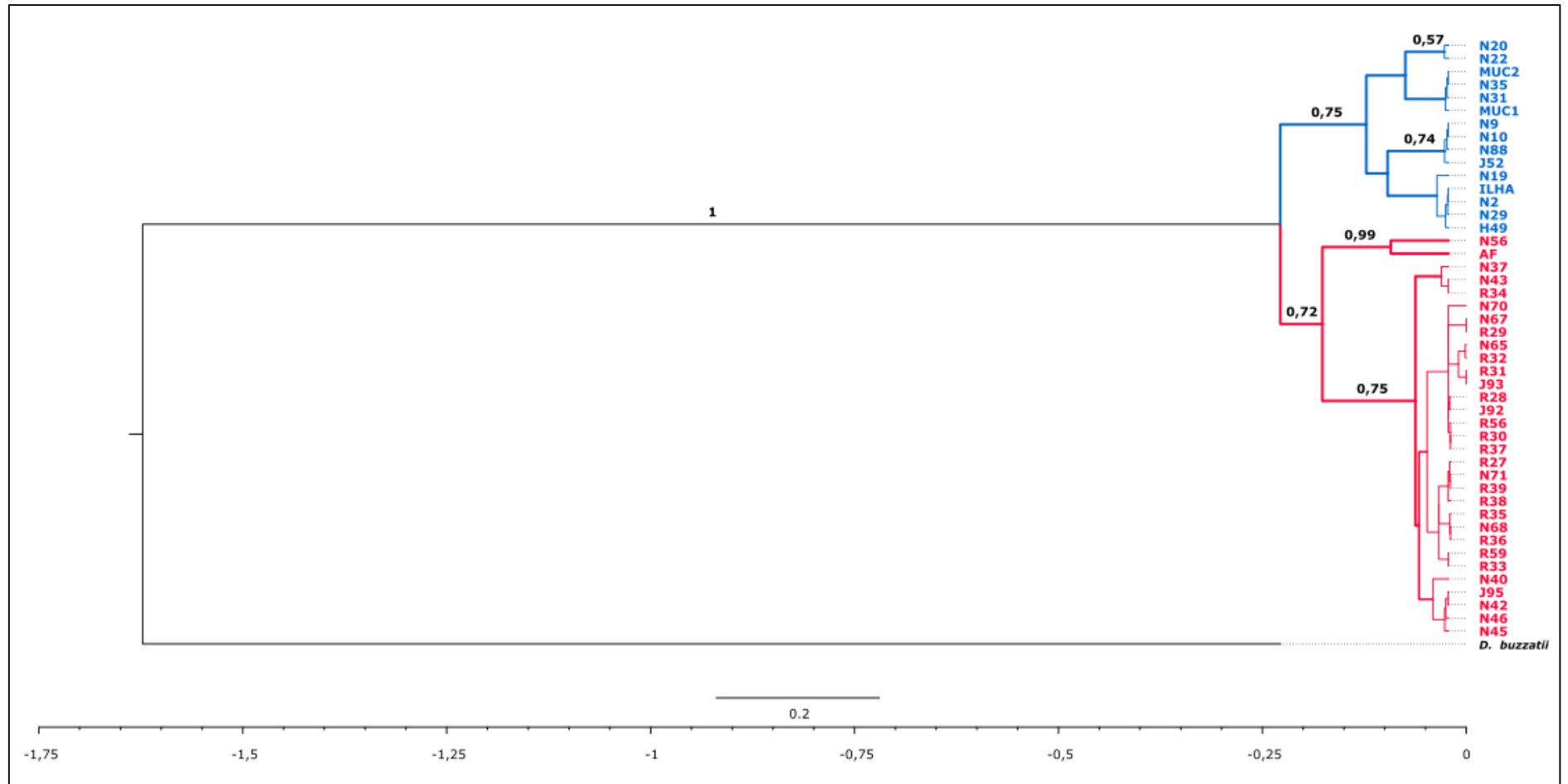


Figura 7. Reconstrução filogenética consenso de *D. serido* baseada nos genes *COI*, *kl-5*, *GstD1* e α -*E5*. As probabilidades posteriores (>0,50) se encontram na parte superior dos ramos. As cores vermelho e azul representam, respectivamente, as populações do Grupo Nordeste e do Grupo Litoral definidos na hipótese inicial. Os valores negativos da barra inferior, lida da direita para a esquerda, representam o tempo em milhões de anos a partir do presente (zero).

4.1.2. Estatística Descritiva

Os valores globais dos índices que compõem a estatística genética descritiva para cada um dos quatro marcadores moleculares de *Drosophila serido* utilizados estão compilados na **Tabela 5**.

Tabela 5. Estatística genética descritiva para todo o conjunto de dados de sequências de *D. serido*.

Marcador	Índices de Diversidade Genética					
	N	S	H	Hd (\pm SD)	π (\pm SD)	K
COI	249	62	59	0,787 (0,022)	0,00645 (0,00042)	3,997
<i>kl-5</i>	207	6	5	0,337 (0,037)	0,00102 (0,00012)	1,221
<i>GstD1</i>	184	1	2	0,011 (0,011)	0,00003 (0,00003)	0,011
<i>αE5</i>	212	17	13	0,569 (0,032)	0,00295 (0,00016)	1,719

N: número de sequências utilizadas; S: número de sítios segregantes; H: número de haplótipos; Hd: índice de diversidade haplotípica; π : índice de diversidade nucleotídica; k: número médio de diferenças nucleotídicas; SD: desvio padrão (*standard deviation*).

Os índices de diversidade genética das populações para cada marcador estão contidos nos itens abaixo (a – e).

a) Gene mitocondrial COI

A **Tabela 6** contém os valores dos índices de diversidade genética para cada população analisada a partir deste mesmo marcador.

Tabela 6. Estatística genética descritiva para cada população de *D. serido* a partir das sequências do gene mitocondrial COI.

População	Índices de Diversidade Genética					
	n	S	h	Hd (\pm SD)	π (\pm SD)	k
AF	6	0	0	0,000	0,00000	0,000
H49	4	0	0	0,000	0,00000	0,000
ILHA	4	1	2	0,500 (0,265)	0,00081 (0,00043)	0,500
J52	7	0	0	0,000	0,00000	0,000

J92	6	2	3	0,733 (0,155)	0,00151 (0,00042)	0,933
J93	6	2	3	0,733 (0,155)	0,0014 (0,00040)	0,867
J95	4	0	0	0,000	0,00000	0,000
J99*	1	-	-	-	-	-
MUC1*	1	-	-	-	-	-
MUC2	4	8	2	0,667 (0,204)	0,00860 (0,00263)	5,333
N10	2	2	2	1,000 (0,500)	0,00323 (0,00161)	2,000
N19	7	5	3	0,524 (0,209)	0,00230 (0,00106)	1,429
N2	6	0	0	0,000	0,00000	0,000
N20	6	0	0	0,000	0,00000	0,000
N22	5	1	2	0,400 (0,237)	0,00065 (0,00038)	0,400
N29	6	0	0	0,000	0,00000	0,000
N31	6	10	4	0,800 (0,172)	0,00806 (0,00218)	5,000
N35	5	8	3	0,700 (0,218)	0,00516 (0,00263)	3,200
N37	10	17	7	0,933 (0,062)	0,00993 (0,00255)	6,156
N40	5	3	3	0,800 (0,164)	0,00226 (0,00078)	1,400
N42	10	1	2	0,200 (0,154)	0,00032 (0,00025)	0,200
N43	9	18	7	0,917 (0,092)	0,01093 (0,00278)	6,778
N45	5	0	0	0,000	0,00000	0,000
N46	6	3	3	0,600 (0,215)	0,00194 (0,00069)	1,200

N56	8	8	4	0,643 (0,184)	0,00409 (0,00162)	2,536
N65	10	9	6	0,867 (0,085)	0,00333 (0,00098)	2,067
N67	2	1	2	1,000 (0,500)	0,00161 (0,00081)	1,000
N68	4	2	3	0,833 (0,222)	0,00188 (0,00059)	1,167
N70	10	7	8	0,956 (0,059)	0,00276 (0,00043)	1,711
N71	7	3	4	0,810 (0,130)	0,00169 (0,00042)	1,048
N88*	1	1	-	-	-	-
N9	3	0	0	0,000	0,00000	0,000
R27	4	1	2	0,667 (0,204)	0,00108 (0,00033)	0,667
R28	6	2	3	0,733 (0,155)	0,00140 (0,00040)	0,867
R29	5	5	5	1,000 (0,126)	0,00323 (0,00072)	2,000
R30	6	3	4	0,800 (0,172)	0,00161 (0,00049)	1,000
R31	6	3	4	0,800 (0,172)	0,00161 (0,00049)	1,000
R32	6	5	4	0,800 (0,172)	0,00269 (0,00102)	1,667
R33*	1	1	-	-	-	-
R34*	1	1	-	-	-	-
R35	2	1	2	1,000 (0,500)	0,00161 (0,00081)	1,000
R36	6	3	4	0,867 (0,129)	0,00194 (0,00045)	1,200

R37	8	3	4	0,643 (0,184)	0,00121 (0,00043)	0,750
R38	6	1	2	0,033 (0,215)	0,00054 (0,00035)	0,333
R39	5	1	2	0,600 (0,175)	0,00097 (0,00028)	0,600
R56	10	2	3	0,511 (0,164)	0,00090 (0,00033)	0,556
R59*	1	1	-	-	-	-

n: número de sequências por população; S: número de sítios segregantes; h: número de haplótipos; Hd: índice de diversidade haplotípica; π : índice de diversidade nucleotídica; k: número médio de diferenças nucleotídicas; SD: desvio padrão (*standard deviation*); *: populações com apenas uma sequência, não sendo possível a estimativa de seus índices de diversidade genética.

b) Gene nuclear *kl-5*

A **Tabela 7** contém os valores dos índices de diversidade genética para cada população analisada a partir deste mesmo marcador.

Tabela 7. Estatística genética descritiva de cada população de *D. serido* a partir das sequências do gene nuclear *kl-5*.

População	Índices de Diversidade Genética					
	n	S	h	Hd (\pm SD)	π (\pm SD)	k
AF	4	0	0	0,000	0,00000	0,000
H49	4	1	2	0,500 (0,265)	0,00042 (0,00022)	0,500
ILHA	6	0	0	0,000	0,00000	0,000
J52	5	0	0	0,000	0,00000	0,000
J92	6	1	2	0,533 (0,172)	0,00044 (0,00014)	0,533
J93	6	0	0	0,000	0,00000	0,000
J95	2	0	0	0,000	0,00000	0,000
MUC2	6	0	0	0,000	0,00000	0,000
N10	3	0	0	0,000	0,00000	0,000
N19	6	0	0	0,000	0,00000	0,000
N2	3	0	0	0,000	0,00000	0,000
N20	6	0	0	0,000	0,00000	0,000
N22	3	0	0	0,000	0,00000	0,000
N29	3	0	0	0,000	0,00000	0,000
N31	2	0	0	0,000	0,00000	0,000

N35	5	0	0	0,000	0,00000	0,000
N37	6	0	0	0,000	0,00000	0,000
N40	4	0	0	0,000	0,00000	0,000
N42	6	0	0	0,000	0,00000	0,000
N43	6	0	0	0,000	0,00000	0,000
N45	4	0	0	0,000	0,00000	0,000
N46	6	0	0	0,000	0,00000	0,000
N56	6	0	0	0,000	0,00000	0,000
N65	5	0	0	0,000	0,00000	0,000
N67	2	1	2	1,000 (0,500)	0,00083 (0,00042)	1,000
N68	6	0	0	0,000	0,00000	0,000
N70	6	0	0	0,000	0,00000	0,000
N71	6	0	0	0,000	0,00000	0,000
N88	3	0	0	0,000	0,00000	0,000
N9	3	1	2	0,667 (0,314)	0,00056 (0,00026)	0,667
R27	5	0	0	0,000	0,00000	0,000
R28	5	1	2	0,400 (0,237)	0,00033 (0,00020)	0,400
R29	4	0	0	0,000	0,00000	0,000
R30	6	0	0	0,000	0,00000	0,000
R31	6	0	0	0,000	0,00000	0,000
R32	4	0	0	0,000	0,00000	0,000
R33*	1	-	-	-	-	-
R34*	1	-	-	-	-	-
R35	2	0	0	0,000	0,00000	0,000
R36	6	0	0	0,000	0,00000	0,000
R37	6	1	2	0,333 (0,215)	0,00028 (0,00018)	0,333
R38	6	0	0	0,000	0,00000	0,000
R39	6	0	0	0,000	0,00000	0,000
R56	9	0	0	0,000	0,00000	0,000
R59*	1	-	-	-	-	-

n: número de sequências por população; S: número de sítios segregantes; h: número de haplótipos; Hd: índice de diversidade haplotípica; π : índice de diversidade nucleotídica; k: número médio de diferenças nucleotídicas; SD: desvio padrão (*standard deviation*); *: populações com apenas uma sequência, não sendo possível a estimativa de seus índices de diversidade genética.

c) **Gene nuclear autossômico *GstD1***

A **Tabela 8** contém os valores dos índices de diversidade genética para cada população analisada a partir deste mesmo marcador.

Tabela 8. Estatística genética descritiva de cada população de *D. serido* a partir das sequências do gene nuclear autossômico *GstD1*.

População	Índices de Diversidade Genética					
	n	S	h	Hd (\pm SD)	π (\pm SD)	k
AF	4	0	0	0,000	0,00000	0,000
H49	4	0	0	0,000	0,00000	0,000
ILHA	6	0	0	0,000	0,00000	0,000
J52*	1	-	-	-	-	-
J92	4	0	0	0,000	0,00000	0,000
J93	6	0	0	0,000	0,00000	0,000
J95	3	0	0	0,000	0,00000	0,000
MUC2	6	0	0	0,000	0,00000	0,000
N19	6	1	2	0,333 (0,215)	0,00090 (0,00058)	0,333
N2	4	2	2	0,500 (0,265)	0,00270 (0,00143)	1,000
N20	6	1	2	0,333 (0,215)	0,00090 (0,00058)	0,333
N22	3	0	0	0,000	0,00000	0,000
N29	2	0	0	0,000	0,00000	0,000
N31	4	0	0	0,000	0,00000	0,000
N35	5	0	0	0,000	0,00000	0,000
N37	4	0	0	0,000	0,00000	0,000
N40	5	0	0	0,000	0,00000	0,000
N42	6	0	0	0,000	0,00000	0,000
N43	6	0	0	0,000	0,00000	0,000
N45	3	0	0	0,000	0,00000	0,000
N46	5	0	0	0,000	0,00000	0,000
N56	6	0	0	0,000	0,00000	0,000
N65	6	0	0	0,000	0,00000	0,000
N68	6	0	0	0,000	0,00000	0,000
N70	6	0	0	0,000	0,00000	0,000
N71	6	0	0	0,000	0,00000	0,000
N88	2	6	2	1,000 (0,500)	0,01592 (0,00796)	6,000
R27	5	0	0	0,000	0,00000	0,000
R28	3	1	2	0,667 (0,314)	0,00177 (0,00083)	0,667
R29*	1	-	-	-	-	-
R30	6	0	0	0,000	0,00000	0,000

R31	6	0	0	0,000	0,00000	0,000
R32	3	0	0	0,000	0,00000	0,000
R33*	1	-	-	-	-	-
R34*	1	-	-	-	-	-
R35	2	0	0	0,000	0,00000	0,000
R36	4	0	0	0,000	0,00000	0,000
R37	6	0	0	0,000	0,00000	0,000
R38	6	0	0	0,000	0,00000	0,000
R39	6	0	0	0,000	0,00000	0,000
R56	8	0	0	0,000	0,00000	0,000
R59*	1	-	-	-	-	-

n: número de sequências por população; S: número de sítios segregantes; h: número de haplótipos; Hd: índice de diversidade haplotípica; π : índice de diversidade nucleotídica; k: número médio de diferenças nucleotídicas; SD: desvio padrão (*standard deviation*); *: populações com apenas uma sequência, não sendo possível a estimativa de seus índices de diversidade genética.

d) Gene nuclear autossômico α -esterase5

A **Tabela 9** contém os valores dos índices de diversidade genética para cada população analisada a partir deste mesmo marcador.

Tabela 9. Estatística genética descritiva de cada população de *D. serido* a partir das sequências do gene nuclear autossômico α -esterase5.

População	Índices de Diversidade Genética					
	n	S	h	Hd (\pm SD)	π (\pm SD)	k
AF	5	0	0	0,000	0,00000	0,000
H49	3	1	2	0,667 (0,314)	0,00097 (0,00046)	0,667
ILHA	6	0	0	0,000	0,00000	0,000
J52	5	0	0	0,000	0,00000	0,000
J92	5	5	2	0,400 (0,237)	0,00470 (0,00279)	3,200
J93	5	16	4	0,900 (0,161)	0,01168 (0,00238)	8,000
J95	4	5	4	1,000 (0,177)	0,00420 (0,00114)	2,833
MUC2	6	3	3	0,733 (0,155)	0,00250 (0,00061)	1,667
N10	3	0	0	0,000	0,00000	0,000
N19	6	1	2	0,533 (0,0172)	0,00078 (0,00025)	0,533
N2	6	0	0	0,000	0,00000	0,000
N20	6	0	0	0,000	0,00000	0,000

N22	5	1	2	0,400 (0,237)	0,00059 (0,00035)	0,400
N29	5	7	2	0,600 (0,175)	0,00612 (0,00179)	4,200
N31	6	0	0	0,000	0,00000	0,000
N35	5	0	0	0,000	0,00000	0,000
N37	5	6	3	0,800 (0,164)	0,00469 (0,00096)	3,200
N40	4	7	3	0,833 (0,222)	0,00638 (0,00182)	4,333
N42	6	15	6	1,000 (0,096)	0,01032 (0,00182)	7,000
N43	6	20	6	1,000 (0,096)	0,01122 (0,00238)	7,667
N45	4	18	3	0,833 (0,222)	0,01526 (0,00437)	10,500
N46	5	10	5	1,000 (0,126)	0,00709 (0,00157)	4,800
N56	6	0	0	0,000	0,00000	0,000
N65	6	0	0	0,000	0,00000	0,000
N67*	1	-	-	-	-	-
N68	6	3	4	0,867 (0,129)	0,00218 (0,00046)	1,467
N70	6	8	3	0,733 (0,155)	0,00597 (0,00120)	4,067
N71	5	0	0	0,000	0,00000	0,000
N88	6	0	0	0,000	0,00000	0,000
N9	3	0	0	0,000	0,00000	0,000
R27	5	6	2	0,600 (0,175)	0,00529 (0,00154)	3,600
R28	5	4	4	0,900 (0,161)	0,00238 (0,00071)	1,600
R29	4	5	3	0,833 (0,222)	0,00366 (0,00153)	2,500
R30	6	0	0	0,000	0,00000	0,000
R31	5	2	2	0,600 (0,175)	0,00178 (0,00052)	1,200

R32	6	1	2	0,333 (0,215)	0,00049 (0,00032)	0,333
R33*	1	-	-	-	-	-
R34*	1	-	-	-	-	-
R35	2	0	0	0,000	0,00000	0,000
R36	4	0	0	0,000	0,00000	0,000
R37	6	8	4	0,867 (0,129)	0,00454 (0,00153)	3,067
R38	5	0	0	0,000	0,00000	0,000
R39	6	10	4	0,800 (0,172)	0,00486 (0,00243)	3,333
R56	6	10	5	0,933 (0,122)	0,00584 (0,00164)	3,933

n: número de sequências por população; S: número de sítios segregantes; h: número de haplótipos; Hd: índice de diversidade haplotípica; π : índice de diversidade nucleotídica; k: número médio de diferenças nucleotídicas; SD: desvio padrão (*standard deviation*); *: populações com apenas uma sequência, não sendo possível a estimativa de seus índices de diversidade genética.

4.1.3. Estrutura Populacional

Análise Discriminante dos Componentes Principais (ADCP)

a) Gene mitocondrial COI

A ADCP realizada para o gene COI considerou a discriminação de dois grupos geneticamente distintos ($K = 2$), denominados Grupo 1 e Grupo 2, dentro das populações amostradas de *D. serido* (**Figura 8**). O Grupo 1 é composto majoritariamente por haplótipos de populações da Caatinga, enquanto que o Grupo 2 é composto majoritariamente por haplótipos de populações do litoral Atlântico.

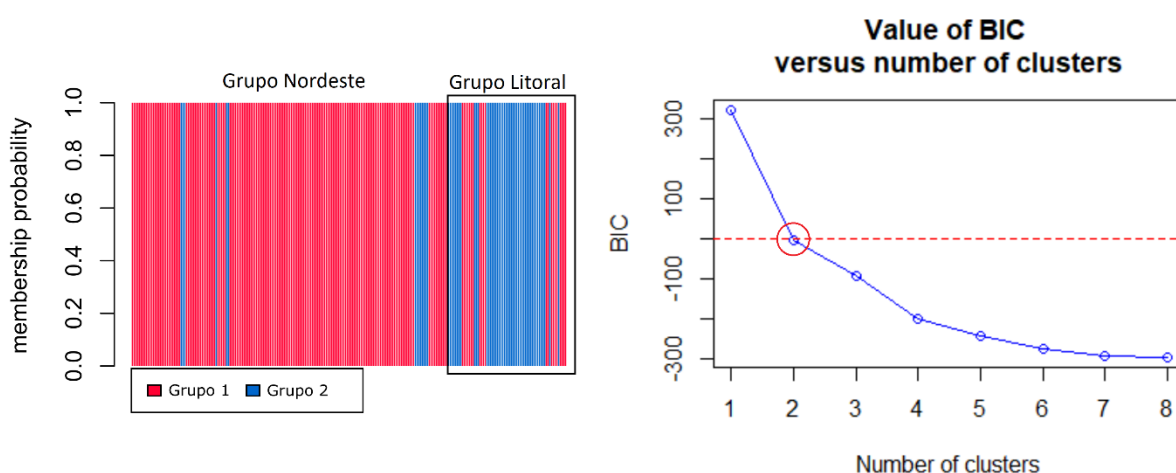


Figura 8. O gráfico à esquerda possui probabilidades de associação (*membership probability*) de cada haplótipo a um determinado grupo obtidas por meio da ADCP a partir do gene COI. O contorno em preto na figura destaca os indivíduos atribuídos ao Grupo Litoral da hipótese inicial. Os demais indivíduos, à esquerda do contorno, pertencem ao Grupo Nordeste. O gráfico à direita contém os valores de BIC versus o número de *clusters* do intervalo pré-definido na ADPC. O círculo vermelho destaca o valor de K escolhido. A linha vermelha tracejada representa o ponto onde o valor de BIC é igual a zero.

A composição haplotípica dos dois grupos genéticos inferidos pela ADPC para o gene COI, entretanto, difere da hipótese de estruturação estabelecida *a priori*, pois há intercâmbio de haplótipos dos grupos genéticos 1 e 2 com os intervalos de sequências definidos como Grupo Nordeste e Grupo Litoral. Esse compartilhamento é mais expressivo dentro do Grupo Litoral, que contém muitos haplótipos associados ao Grupo 1.

O Grupo Litoral abarca haplótipos das localidades de Penha – SC (J52-3, J52-4, J52-5, J52-6, J52-7, J52-8, J52-9 e J52-10), Mucuri – BA (MUC2-6 e MUC2-7), Governador Celso Ramos – SC (N9-1, N9-2 e N9-3), Camboriú – SC (N10-1 e N10-2), Itaúnas – ES (N31-6 e

N31-6), Cabrália – BA (N35-4, N35-5, N35-6 e N35-7), e Florianópolis – SC (N88-2) que, embora considerados como parte do Grupo Litoral na hipótese inicial, foram associados geneticamente ao Grupo 1, um “cluster” onde predominam haplótipos de populações da Caatinga.

O Grupo Nordeste abarca haplótipos das localidades de Morro da Barrinha – BA (N37-7 e N37-8), Irecê – BA (N43-9 e N43-10) e Serra do Cipó (N56-7, N56-8, N56-12, N56-13, N56-14, N56-15 e N56-16), que, apesar de pertencerem a esse mesmo grupo na hipótese inicial, foram associados geneticamente ao Grupo 2, um “cluster” onde predominam haplótipos de populações do litoral Atlântico (do sul da Bahia à Santa Catarina).

As populações de Mucuri – BA (MUC2), Itaúnas – ES (N31), Cabrália – BA (N35), Morro da Barrinha – BA (N37) e Irecê – BA (N43) compartilham haplótipos tanto com o Grupo Nordeste quanto com o Grupo Litoral da hipótese inicial.

b) Gene nuclear *kl-5*

A ADCP realizada para o gene *kl-5* considerou a discriminação de dois grupos geneticamente distintos ($K = 2$), denominados Grupo 1 e Grupo 2, dentro das populações amostradas de *D. serido* (Figura 9). O Grupo 1 é composto majoritariamente por haplótipos de populações da Caatinga, enquanto que o Grupo 2 é composto majoritariamente por haplótipos de populações do litoral Atlântico.

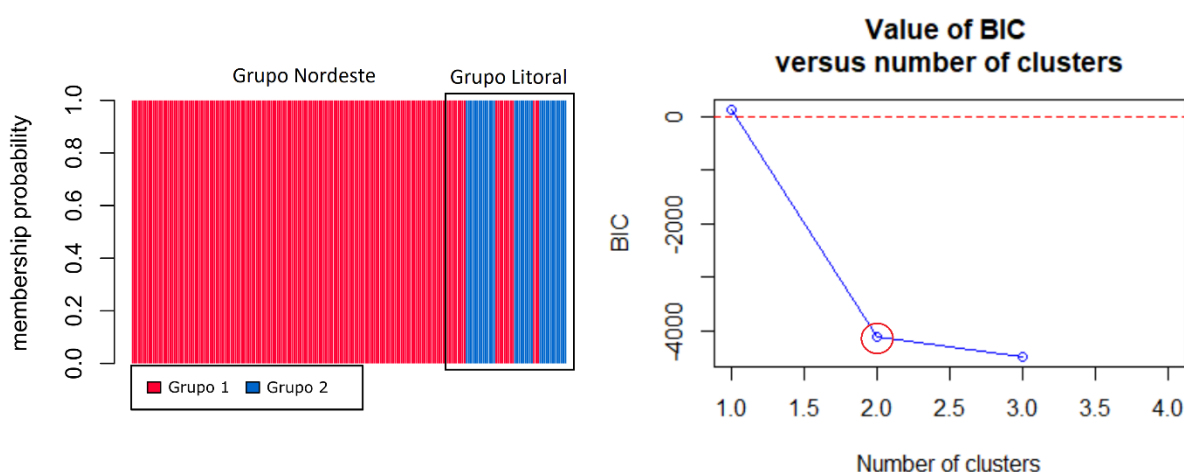


Figura 9. O gráfico à esquerda possui probabilidades de associação (*membership probability*) de cada haplótipo a um determinado grupo obtidas por meio da ADCP a partir do gene *kl-5*. O contorno em preto na figura destaca os indivíduos atribuídos ao Grupo Litoral da hipótese inicial. Os demais indivíduos, à esquerda do contorno, pertencem ao Grupo Nordeste. O gráfico à direita contém os valores de BIC versus o número de *clusters* do intervalo pré-definido na ADPC. O círculo

vermelho destaca o valor de K escolhido. A linha vermelha tracejada representa o ponto onde o valor de BIC é igual a zero.

A composição haplotípica dos dois grupos genéticos inferidos pela ADPC para o gene *kl-5* difere da hipótese de estruturação estabelecida *a priori*, pois há intercâmbio de haplótipos dos grupos genéticos 1 e 2 com os intervalos de sequências definidos como Grupo Nordeste e Grupo Litoral. Esse compartilhamento ocorre apenas dentro do Grupo Litoral, que contém muitos haplótipos associados ao Grupo 1, um “cluster” onde predominam haplótipos de populações da Caatinga. O Grupo Nordeste é composto apenas por haplótipos associados ao Grupo 1.

Dentro do Grupo Litoral, todos os haplótipos das localidades de Bertioga – SP (H49), Ilha de Alcatrazes – SP (ILHA), Guaratuba – SP (N2), Paúba – SP (N19) e Praia da Jureia – SP (N29) foram considerados como geneticamente pertencentes ao Grupo 1 pela ADCP.

c) Gene nuclear autossômico *GstD1*

A ADCP realizada para o gene *GstD1* considerou a discriminação de dois grupos geneticamente distintos, Grupo 1 e Grupo 2, dentro das populações amostradas de *D. serido* (**Figura 10**). O Grupo 1 é composto majoritariamente por haplótipos de populações da Caatinga, enquanto que o Grupo 2 é composto majoritariamente por haplótipos de populações do litoral Atlântico (do sul da Bahia à Santa Catarina).

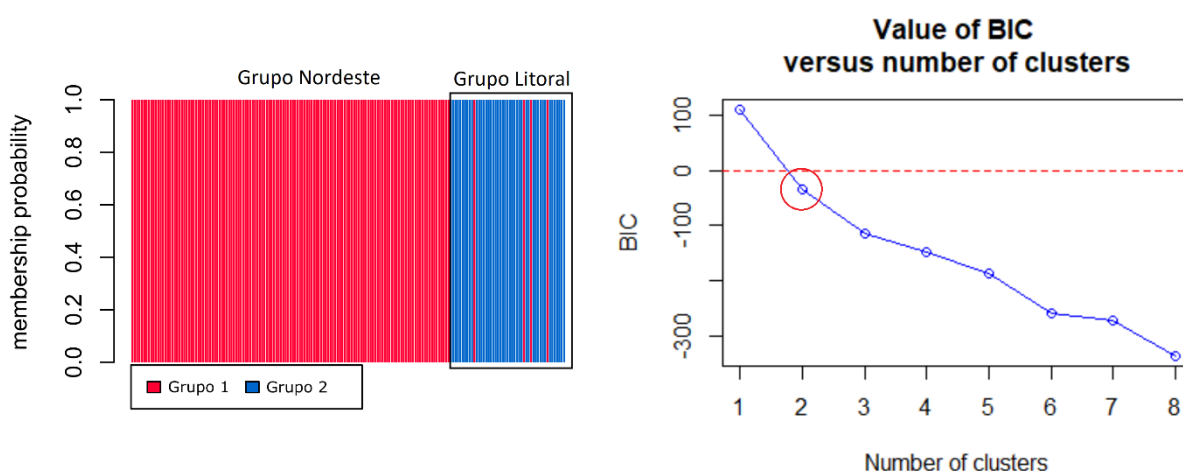


Figura 10. O gráfico à esquerda possui probabilidades de associação (*membership probability*) de cada haplótipo a um determinado grupo obtidas por meio da ADCP a partir do gene *GstD1*. O contorno em preto na figura destaca os indivíduos atribuídos ao Grupo Litoral da hipótese inicial. Os demais indivíduos, à esquerda do contorno, pertencem ao Grupo Nordeste. O gráfico à direita contém os valores de BIC versus o número de *clusters* do intervalo pré-definido na ADPC. O círculo

vermelho destaca o valor de K escolhido. A linha vermelha tracejada representa o ponto onde o valor de BIC é igual a zero.

A composição haplotípica dos dois grupos genéticos inferidos pela ADCP para o gene *GstD1* difere da hipótese de estruturação estabelecida *a priori*, pois há intercâmbio de haplótipos dos grupos genéticos 1 e 2 com os intervalos de sequências definidos como Grupo Nordeste e Grupo Litoral, ainda que em menor frequência do que nos marcadores anteriores. Os resultados mostram que o intervalo definido como Grupo Litoral possui haplótipos associados geneticamente ao Grupo 1, um “cluster” composto majoritariamente por populações da Caatinga inferido por meio da ADCP. O Grupo Nordeste é composto apenas por haplótipos definidos como pertencentes ao grupo genético Grupo 1.

Dentro do intervalo de sequências denominado Grupo Litoral, apenas um haplótipo das populações totais de Penha – SC (J52-2), Arraial do Cabo – RJ (N20-7), Macaé – RJ (N22-6) e Itaúnas – ES (N31-6) foram considerados como geneticamente pertencentes ao Grupo 1 pela ADCP.

d) **Gene nuclear autossômico α -E5**

A ADCP realizada para o gene α -E5 considerou a discriminação de dois grupos geneticamente distintos, Grupo 1 e Grupo 2, dentro das populações amostradas de *D. serido* (**Figura 11**). O Grupo 1 é composto majoritariamente por haplótipos de populações da Caatinga, enquanto que o Grupo 2 é composto majoritariamente por haplótipos de populações do litoral Atlântico (do sul da Bahia à Santa Catarina).

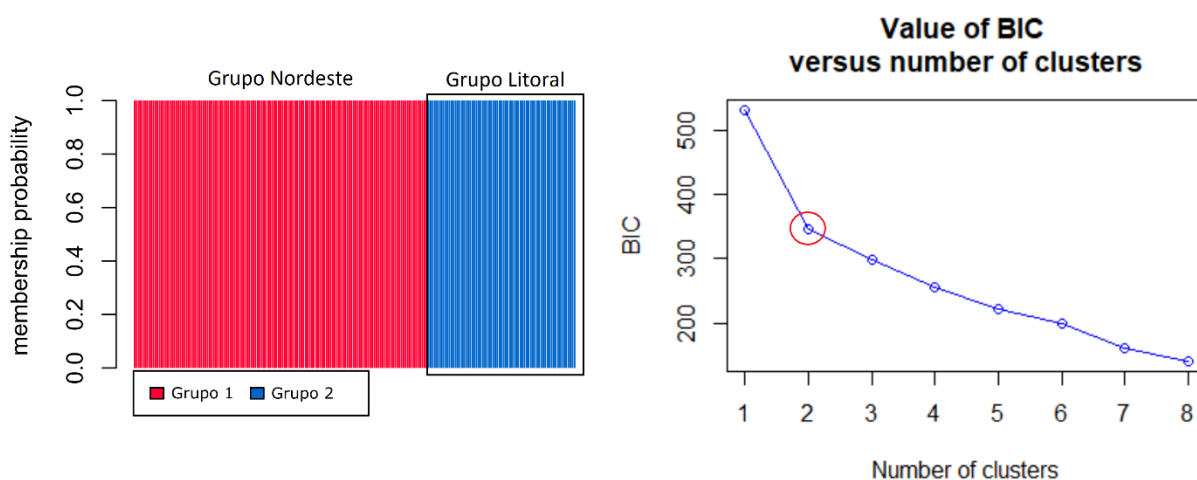


Figura 11. O gráfico à esquerda possui probabilidades de associação (*membership probability*) de cada haplótipo a um determinado grupo obtidas por meio da ADCP a partir do gene α -E5. O

proporcionais ao número de indivíduos que eles contêm, que, por sua vez, encontra-se descrito no centro de cada haplótipo. Cada ponto nas linhas representa um passo mutacional. A escala da rede pode ser visualizada no canto inferior direito da figura.

b) Gene nuclear *kl-5*

Foram estimados 5 haplótipos a partir das 207 sequências do marcador *kl-5* de *D. serido* (**Figura 13**). A configuração das distribuições haplotípicas da rede indica um compartilhamento de haplótipos entre indivíduos de populações do Grupo Litoral com os de haplótipos do Grupo Nordeste. Os haplótipos compartilhados entre os dois diferentes grupos são os mesmos descritos na seção da ADCP para o mesmo marcador.

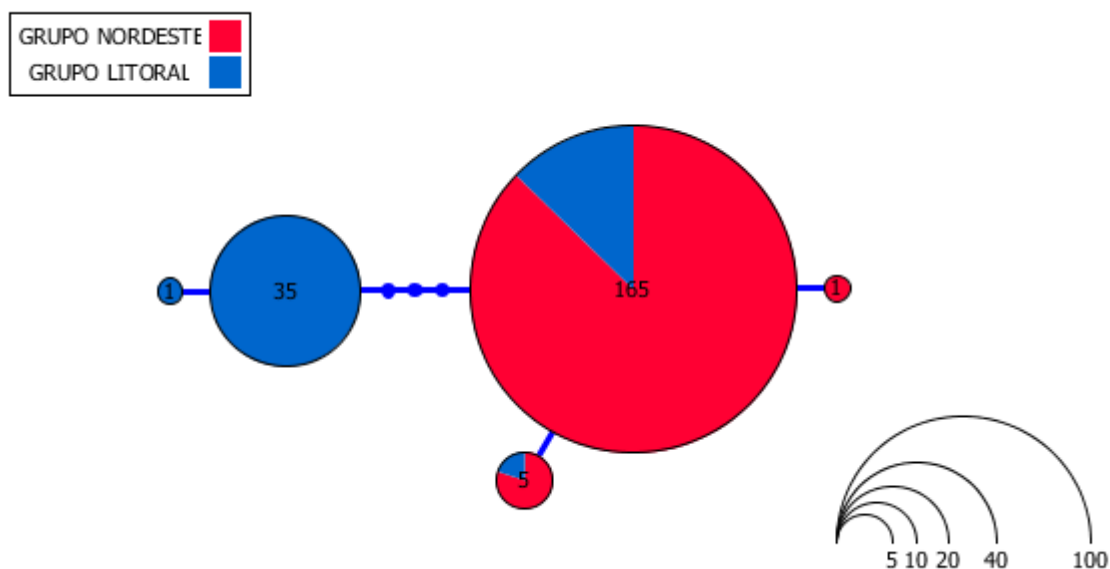


Figura 13. Rede haplotípica não enraizada obtida a partir do gene *kl-5* para *D. serido* por meio do programa Haplotype Viewer. Os haplótipos são representados por círculos, cujas dimensões são proporcionais ao número de indivíduos que eles contêm, que, por sua vez, encontra-se descrito no centro de cada haplótipo. Cada ponto nas linhas representa um passo mutacional. A escala da rede pode ser visualizada no canto inferior direito da figura.

c) Gene nuclear autossômico *GstD1*

Foram estimados 55 haplótipos a partir das 184 sequências do marcador *GstD1* de *D. serido* (**Figura 14**). A configuração das distribuições haplotípicas da rede indica um compartilhamento de haplótipos entre indivíduos de populações do Grupo Litoral com os de haplótipos do Grupo Nordeste. Os haplótipos compartilhados entre os dois diferentes grupos são os mesmos descritos na seção da ADCP para o mesmo marcador.

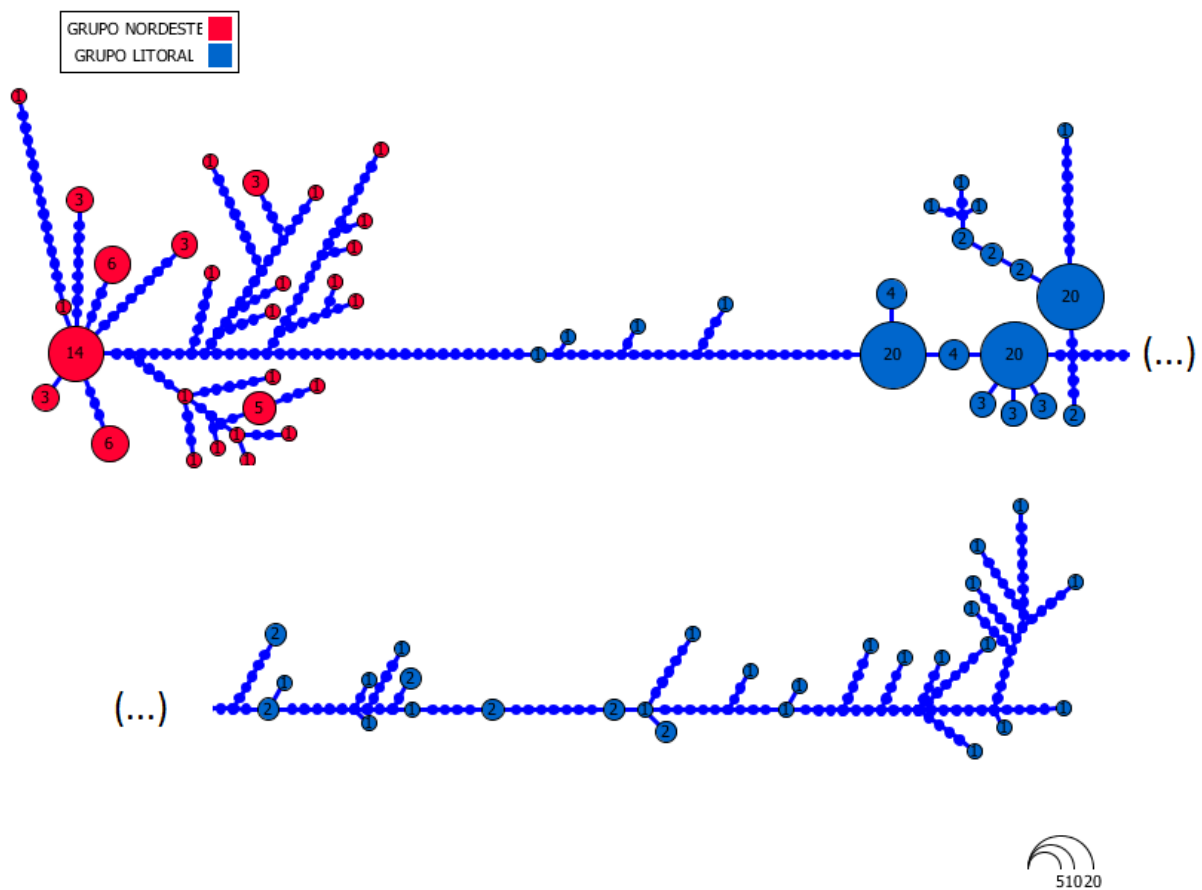


Figura 15. Rede haplotípica não enraizada obtida a partir do gene α -E5 para *D. serido* por meio do programa Haplotype Viewer. Os haplótipos são representados por círculos, cujas dimensões são proporcionais ao número de indivíduos que eles contêm, que, por sua vez, encontra-se descrito no centro de cada haplótipo. Cada ponto nas linhas representa um passo mutacional. A escala da rede pode ser visualizada no canto inferior direito da figura.

Análise de Variância Molecular (AMOVA) e Teste de Mantel

A **Tabela 10** contém os resultados das AMOVAs Globais e Hierárquicas, considerando a hipótese inicial de dois grupos distintos (Nordeste e Litoral) em *Drosophila serido*, realizadas separadamente para todos os marcadores moleculares.

Tabela 10. Resultados da Análises de Variância Molecular (AMOVAs) Global e Hierárquicas.

Fonte de Variação	d.f.	Componentes de Variância	Porcentagem de Variação	Índice de Fixação
COI				
AMOVA Global				
Entre as Populações	45	1,24829 Va	61,59	Fst = 0,61594*
Dentro das Populações	203	0,77836 Vb	38,41	
Total	248	2,02665	100	
AMOVA Hierárquica				
Entre os Grupos	1	1,54942 Va	52,63	Fct = 0,52627*

Entre Populações	44	0,61637 Vb	20,94	Fsc = 0,44193*
Dentro dos Grupos				
Dentro das Populações	203	0,77836 Vc	26,44	Fst = 0,73563*
Total	248	2,94414	100	

kl-5

AMOVA Global

Entre as Populações	44	0,59277 Va	95,16	
Dentro das Populações	162	0,03014 Vb	04,84	Fst = 0,95161*
Total	206	0,62292	100	

AMOVA Hierárquica

Entre os Grupos	1	0,75379 Va	70,80	Fct = 0,70800*
Entre Populações				
Dentro dos Grupos	43	0,28074 Vb	26,37	Fsc = 0,90304*
Dentro das Populações	162	0,03014 Vc	02,83	Fst = 0,97169*
Total	206	1,06468	100	

Gstd1

AMOVA Global

Entre as Populações	41	0,39952 Va	91,28	
Dentro das Populações	142	0,03815 Vb	08,72	Fst = 0,91284*
Total	183	0,43766		

AMOVA Hierárquica

Entre os Grupos	1	0,63085 Va	77,42	Fct = 0,77416*
Entre Populações				
Dentro dos Grupos	40	0,14588 Vb	17,90	Fsc = 0,79272*
Dentro das Populações	142	0,03815 Vc	04,68	Fst = 0,95319*
Total	183	0,81488	100	

αE5

AMOVA Global

Entre as Populações	42	2,73201 Va	83,25	
Dentro das Populações	168	0,54970 Vb	16,75	Fst = 0,8325*
Total	211	3,28171	100	

AMOVA Hierárquica

Entre os Grupos	1	5,26030 Va	85,03	Fct = 0,85026*
Entre Populações				
Dentro dos Grupos	42	0,32350 Vb	06,20	Fsc = 0,41371*
Dentro das Populações	168	0,54315 Vc	08,78	Fst = 0,91221*
Total	211	6,18673	100	

*: $p < 0,01$.

O Teste de Mantel, feito para cada grupo definido *a priori*, não mostrou correlação entre as distâncias genéticas das populações e suas distâncias geográficas. Para os dois grupos testados e, para cada um dos quatro marcadores, os resultados não obtiveram um valor- p significativo ($p \leq 0,001$ – Diniz-Filho *et al.*, 2013). Logo, não se pode rejeitar a hipótese nula da ausência de isolamento por distância nas populações analisadas.

4.1.4. Análises Demográficas

Testes de Neutralidade e Mismatch Distribution

Nos testes realizados no programa DnaSP (Rozas *et al.*, 2017), os valores da estatística D de Tajima se mostraram significativos apenas para o gene COI nos Grupos Nordeste e Litoral e, para o gene $\alpha E5$, somente no Grupo Nordeste. O índice Fs de Fu se mostrou significativo para os Grupos Nordeste e Litoral para o gene COI e apenas no Grupo Nordeste para o gene $\alpha E5$. A **Tabela 11** sumariza os resultados dos testes de neutralidade, obtidos para cada grupo em cada um dos marcadores. Do mesmo modo, os resultados da análise de *mismatch distribution* estão descritos na **Tabela 12** e suas respectivas curvas para cada marcador e grupo populacional estão contidas na **Figura 16**.

Tabela 11. Resultados dos Testes de Neutralidade realizados para todos os marcadores considerando os dois grupos inferidos na hipótese inicial (Nordeste e Litoral).

População	n	Testes de Neutralidade	
		D de Tajima	Fs de Fu
COI			
Grupo Nordeste	181	-2,55817***	-68,190**
Grupo Litoral	68	-1,83578*	-6,495**
kl-5			
Grupo Nordeste	149	-1,12435	-2,821
Grupo Litoral	58	-1,13321	-1,362
GstD1			
Grupo Nordeste	135	-0,99214	-2,491
Grupo Litoral	49	-1,47321	-2,918*
$\alpha E5$			
Grupo Nordeste	141	-2,41526**	-17,883**
Grupo Litoral	71	1,71751	2,115

n: número de indivíduos por grupo; *: $p < 0,05$; **: $p < 0,02$; ***: $p < 0,001$.

Os valores negativos e significativos observados nos índices Fs de Fu podem indicar uma maior diversidade haplotípica do que o esperado sob um modelo de equilíbrio neutro, o que pode ser explicado por uma expansão populacional recente que aumentou o tamanho efetivo da população, aumentando a diversidade genética (Fu, 1997).

No caso do teste D de Tajima, valores negativos podem sugerir um excesso de sítios segregantes na amostra em relação ao que seria esperado sob um modelo de equilíbrio neutro, o que pode ser explicado por uma expansão populacional recente (após um

gargalo populacional) que aumentou o tamanho efetivo da população e permitiu que variantes alélicas raras se tornassem mais comuns (Tajima, 1989).

Tabela 12. Resultados da análise de *mismatch distribution* realizada para todos os marcadores sobre os dois grupos inferidos na hipótese inicial. Os valores foram computados levando em conta as premissas contidas no *sudden expansion model*.

População	n	<i>mismatch distribution</i>	
		Curva	SSD (valor-p)
<i>COI</i>			
Grupo Nordeste	181	Unimodal	0,007 (0,005)
Grupo Litoral	68	Bimodal	0,072 (0,320)
<i>kl-5</i>			
Grupo Nordeste	149	a	-
Grupo Litoral	58	a	-
<i>GstD1</i>			
Grupo Nordeste	135	a	-
Grupo Litoral	49	a	-
<i>αE5</i>			
Grupo Nordeste	141	Unimodal	0,016 (0,633)
Grupo Litoral	71	Unimodal	0,036 (0,001)

n: número de indivíduos por grupo; SSD: *Sum of Squared Deviation* – soma das diferenças quadráticas, com respectivo valor-p entre parênteses; a: o procedimento de mínimos quadrados (*least-squares*) para ajustar a *mismatch distribution* do modelo e a distribuição observada não convergiu após 2.000 etapas; -: ausência de informação.

Numa análise de *mismatch distribution*, a presença de uma curva bimodal pode sugerir que a população em questão evoluiu dentro das premissas do modelo de equilíbrio neutro por um período prolongado de tempo, já que os dois picos refletem mutações ocorridas em diferentes momentos da história da população. Em contrapartida, se a curva for unimodal, (como no caso dos Grupos Nordeste e Litoral para o gene $\alpha E5$ e do Grupo Nordeste para o gene *COI*) isso pode sugerir que a população passou por um evento de gargalo populacional ou expansão populacional recente, uma vez que as diferenças nucleotídicas refletem mutações recentes e raras (Tajima, 1989; Rogers & Harpending, 1992; Nielsen, 2001).

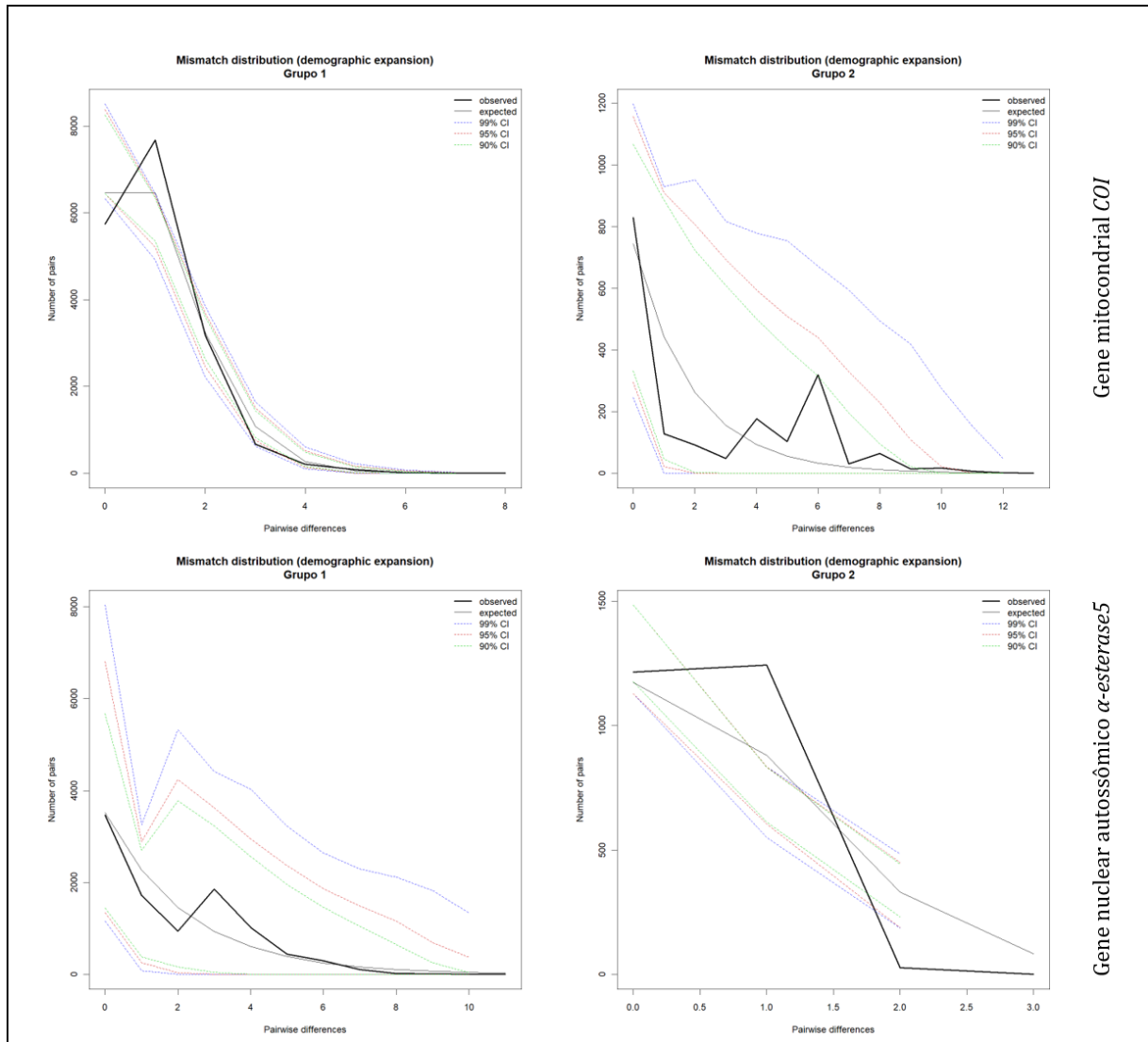


Figura 16. Resultados do teste de *mismatch distribution* assumindo o modelo de expansão repentina. As linhas mais grossas representam a relação do número de pares pelo número diferenças par-a-par observado. A linhas mais finas representam a relação esperada no caso de eventos de expansão populacional.

Extended Bayesian Skyline Plot (EBSP)

A análise coalescente EBSP se mostrou congruente com o cenário de expansão populacional recente sugerido pelos testes de neutralidade e *mismatch distribution* apenas para as populações do Grupo Nordeste, que exibiu um crescimento exponencial em seu tamanho populacional que se iniciou há aproximadamente 40 mil anos atrás, logo após uma pequena contração entre 70 e 50 mil anos atrás (**Figura 17**). Para o Grupo Litoral, entretanto, o EBSP não sugere evidência significativa de mudanças no tamanho efetivo dessa população ao longo do tempo, já que o parâmetro `sum(indicators.alltrees)` visualizado no programa Tracer incluía o valor de zero dentro do intervalo 95% HPD

(Heled & Drummond, 2008). A partir desses resultados, não se pode rejeitar a hipótese nula de equilíbrio demográfico para o Grupo Litoral.

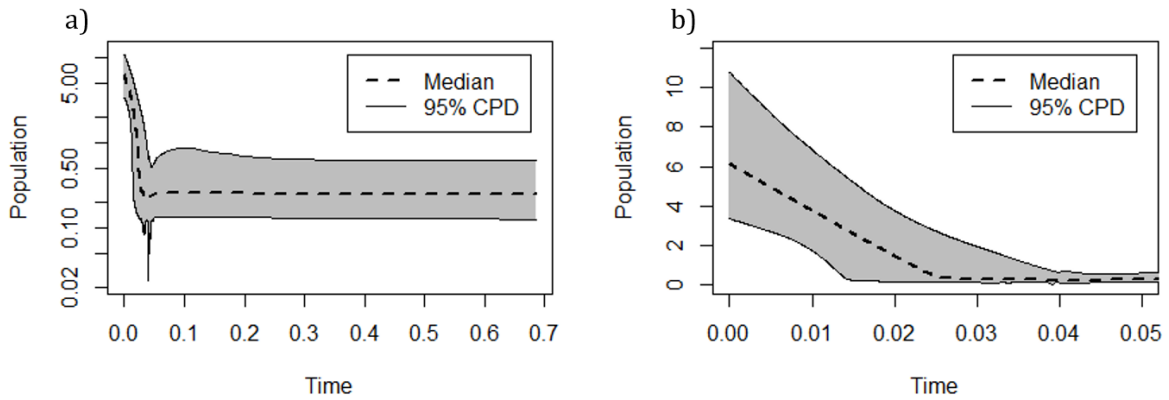


Figura 17. (a) Reconstrução dos eventos de expansão populacional ao longo do tempo, em milhões de anos, a partir do método de EBSP para sequências de DNA de *D. serido* do Grupo Nordeste. A linha tracejada representa a mediana das variações do tamanho populacional efetivo. As linhas contínuas representam o intervalo de 95% *Central Posterior Density* (CPD) dessas mesmas variações. (b) O mesmo gráfico em (a), mas com maior resolução da escala temporal para melhor visualização do evento da expansão populacional.

5. DISCUSSÃO

5.1. Estrutura Populacional e História Demográfica de *Drosophila serido*

A hipótese inicial deste trabalho considerou a existência de dois grupos populacionais distintos ocorrendo dentro da espécie cactófila *D. serido*, conforme sugerido por estudos anteriores, que apontaram a presença de estrutura populacional no grupo (Baimai *et al.*, 1983; Tosi & Sene, 1989; Ruiz *et al.*, 2000; Manfrin *et al.*, 2001; Morales, 2005; Franco; 2009; Franco & Manfrin, 2013). Uma dessas linhagens, levando a denominação de Grupo Nordeste, é constituída por populações que ocorrem majoritariamente na região Nordeste do Brasil, no bioma Caatinga, mas com algumas populações presentes nas restingas de Mata Atlântica do litoral Nordeste e no interior da região Sudeste. A outra, denominada Grupo Litoral, é composta por populações que ocorrem exclusivamente nas zonas de restingas do litoral Atlântico no sul da Bahia e nas regiões Sudeste e Sul do Brasil.

As análises multilocus realizadas levaram à confirmação da existência de estruturação genética em *D. serido*, com ao menos dois grupos ocorrendo dentro da espécie: um grupo formado essencialmente pelas populações que ocorrem na Caatinga (Grupo Nordeste), geneticamente mais homogêneo, e o outro formado pelas populações que ocorrem no litoral Atlântico a partir do sul da Bahia (Grupo Litoral). Entretanto, as análises também apontaram a existência de estruturação dentro do Grupo Litoral, chamando a atenção para a existência de mais subpopulações *D. serido* do que se considerou na hipótese inicial.

Estatística descritiva

Os índices de diversidade genética, quando analisados em conjunto, podem fornecer uma visão mais completa da história evolutiva das populações e das forças que as influenciaram. Considerando os índices obtidos para todo o conjunto de dados (índices globais – **Tabela 5**), o gene COI apresentou os maiores índices de diversidade genética em comparação a todos os outros marcadores analisados, o que indica uma elevada quantidade de variações genéticas e grande diversidade na distribuição dessas variações para esse gene em *D. serido*.

A elevada diversidade genética presente no gene COI também é observada, ainda que em menor intensidade, no gene $\alpha E5$. Uma alta diversidade haplotípica, ao menos no caso do gene COI, pode sugerir a existência de subpopulações na amostragem, tendo sua

possível origem em eventos históricos que modificaram a composição genética das populações amostradas (Tajima, 1989; Fu, 1997).

Os marcadores *kl-5* e *GstD1*, por outro lado, mostraram-se significativamente mais conservados, ou seja, com menos variações genéticas do que os dois anteriores. Isso pode ser explicado pela influência da própria seleção natural: a ação de pressões seletivas pode levar à fixação de alelos favoráveis em detrimento de outros alelos, resultando em uma diminuição da diversidade genética (Hedrick, 2005), como López-Olmos *et al.* (2017) observaram para o gene *GstD1* em *D. hexastigma*, que apresentou forte seleção positiva para esse gene, e que pode estar associada à adaptação desses animais aos seus cactos hospedeiros.

No caso do gene *kl-5*, como os polimorfismos desse gene também estão associados a fatores que influenciam diretamente no *fitness* dos machos portadores, a presença de seleção sexual positiva (Chippindale & Rice, 2001; Koerich *et al.*, 2008; Carvalho, 2002) também pode ser a explicação para os seus baixos índices de diversidade.

Estrutura populacional

Nas análises de estrutura populacional conduzidas no programa *adegenet* e com base na interpretação dos gráficos de BIC (**Figuras 8, 9, 10 e 11**), optou-se pela escolha de um valor de $K = 2$ para todos os marcadores utilizados. Na prática, a escolha do número ideal de *clusters* costuma ser mais difícil em estudos com dados empíricos (Jombart & Collins, 2015), podendo ser influenciada por outros fatores como o conhecimento prévio da biologia e da ecologia das populações estudadas ou os objetivos específicos da análise.

Desse modo, prezando pela padronização dessa escolha e levando em consideração os pontos de inflexão das quatro curvas analisadas (nesse caso, o ponto em que elas mudam sua orientação e começam a decrescer menos), assim como as hipóteses dos estudos anteriores de estruturação no modelo de estudo e também as considerações impostas pela hipótese do presente trabalho, um $K = 2$ foi definido como o número de *clusters* que melhor representa nossos dados disponíveis.

Os padrões de compartilhamento haplotípico observados nos resultados da ADCP entre a hipótese inicial e a hipótese genética *a posteriori* também ocorrem nas redes haplotípicas para todos os marcadores (**Figuras 12, 13, 14 e 15**). De modo geral, tanto as

análises de estrutura genética por meio do *adegenet* quanto as redes haplotípicas sugerem uma maior homogeneidade das populações da Caatinga para a maioria dos marcadores do que as do Grupo Litoral, que apresentam diferentes padrões de associações haplotípicas que variam muito para cada um dos quatro marcadores.

Dentre as AMOVAs hierárquicas realizadas para cada marcador (**Tabela 10**), aquelas cujos índices de diferenciação entre grupos (Fct) obtiveram valores significativos e elevados em comparação aos índices de diferenciação entre as populações dentro dos grupos (Fsc), indicando estruturação populacional significativa, foram aquelas realizadas para os genes COI e $\alpha E5$, levando em consideração a subdivisão em dois grupos populacionais, de acordo com a hipótese inicial.

Os genes *kl-5* e *GstD1* apresentaram valores de Fsc maiores do que de Fct (igualmente elevados e significativos nos dois casos) quando considerada a hipótese de dois grupos para as populações de *D. serido*, indicando que a variação genética está distribuída igualmente tanto dentro quanto entre as subpopulações, não havendo uma estruturação genética significativa entre as subpopulações (Excoffier *et al.*, 1992) a partir dessa hipótese.

No caso específico dos marcadores *kl-5* e *GstD1*, essa inconsistência com a hipótese inicial pode ser explicada pelo fato de que ambos os marcadores possuem índices de diversidade genética muito baixos para cada população amostrada, muitos deles com valores iguais a zero (**Tabela 5**). Contudo, mesmo com baixa diversidade genética global (**Tabelas 7 e 8**), algumas poucas populações exibem índices muito elevados em comparação à maioria, o que pode ser a explicação para os índices de Fst e Fct igualmente elevados para os dois marcadores na AMOVA.

No caso do marcador *kl-5*, as populações com índices elevados de diversidade genética são as de Guaratuba – SP (N2), Governador Celso Ramos – SC (N9), Arraial do Cabo – RJ (N20), Florianópolis – SC (N88) e Monte Alegre de Sergipe – SE (R28) e, para o marcador *GstD1*, as de Bertiooga – SP (H49), Milagres – BA (J92), Governador Celso Ramos – SC (N9), Exu – PE (N67), Monte Alegre de Sergipe – SE (R28) e Lajes Pintadas – RN (R37).

Essas populações mencionadas possuem representantes, para todos os marcadores, tanto no Grupo Nordeste quanto no Grupo Litoral, fazendo com que seja razoável supor que a quantidade de suas variações genéticas seja a explicação algébrica para os elevados

índices de diferenciação entre os dois grupos (Fct) e também entre as populações dentro dos grupos (Fsc) para os genes *kl-5* e *GstD1*, mesmo com os dois marcadores apresentando índices de diversidade genética globais muito baixos em comparação aos genes COI e *αE5*.

Análises Filogenéticas

A existência de duas linhagens para o conjunto de dados moleculares também se reforça na reconstrução filogenética total (**Figura 7**), que recupera dois cladogramas distintos com alto suporte de probabilidades posteriores nos ramos dentro da espécie *D. serido*. A reconstrução filogenética utilizando o gene COI (**Figura 3**), em especial, nos mostra a recuperação de três cladogramas. O primeiro clado, mais externo, é composto apenas por indivíduos da localidade de Serra do Cipó – MG (N56). O segundo, formado por todos os indivíduos do Grupo Nordeste, com a única exceção de Morro do Chapéu – BA (N40), e também por todos os indivíduos das populações do litoral de Santa Catarina. O terceiro clado, irmão do clado anterior, compreende todas as populações do litoral do sul da Bahia, Espírito Santo, Rio de Janeiro e São Paulo com a adição de Serra do Cipó – MG (N40).

De acordo com Franco (2009), a localidade de Serra do Cipó – MG (N56) foi a única que não foi agrupada com nenhuma outra população de *D. serido* analisada em seu trabalho, pois as mesmas apresentaram valores de *Fst* muito elevados e significativos nas comparações par-a-par com outras populações. Segundo o autor isso faz sentido, já que os indivíduos dessa população se encontram em relativo isolamento geográfico no topo de um morro a 860m de altitude acima do nível do mar, onde predomina uma grande quantidade de cactos do gênero *Cereus* (Franco, 2009).

A proximidade evolutiva entre populações de *D. serido* do litoral de Santa Catarina com as populações da Caatinga, no entanto, é curiosa e também pode ser verificada na rede haplotípica para o gene COI (**Figura 12**). Esses padrões podem sugerir um cenário em que duas linhagens ancestrais de *D. serido* divergiram de um ponto de irradiação comum, onde uma delas colonizou o bioma Caatinga e, a outra, as áreas de restinga do litoral Sul, o que explicaria o compartilhamento de haplótipos observado. Num segundo evento de expansão, as populações que permaneceram nesse mesmo ponto de irradiação hipotético, podem ter originado as populações do litoral sul da Bahia e da região Sudeste, o que explicaria as diferenças haplotípicas encontradas para esse gene.

Contudo, como demonstrado por Ruiz *et al.* (2000), as populações dos litorais de São Paulo e Santa Catarina possuem, sobre a inversão $2x^7$, a inversão cromossômica fixa $2y^9$ e, sobre esta, as inversões $2x^8$ e $2w^8$, o que as distingue das populações da Caatinga, que possuem, sobre mesma inversão $2x^7$, as inversões cromossômicas polimórficas $2a^8$, $2b^8$, $2c^8$ e $2d^8$ (Tosi & Sene, 1989) distintas do grupo anterior. Isso sugere que as populações do litoral de Santa Catarina estão evolutivamente mais próximas das populações do litoral de São Paulo do que as da Caatinga e do litoral Nordeste, ao contrário do que indica o sinal filogenético recuperado para o gene COI.

Testes de Neutralidade e Padrões de História Demográfica

Os testes de neutralidade, representados pelos índices F_s de F_u e D de Tajima (**Tabela 11**), apontam para um evento de expansão demográfica após um gargalo populacional recente, que aumentou o tamanho efetivo da população e que também é suportado pela curva bimodal de *mismatch distribution* (**Figura 16**) do Grupo Nordeste (Tajima, 1989; Fu, 1997) para os marcadores COI e $\alpha E5$.

Em contrapartida, também é possível que um valor negativo no teste D de Tajima se deva em razão de efeitos de seleção positiva (Tajima, 1989), como no caso do marcador $\alpha E5$, que pode ter aumentado a frequência de alelos favoráveis nas populações analisadas através de varredura seletiva recente (Tajima, 1989). A influência de seleção positiva em genes do “cluster” α -esterase foi observada por Rane *et al.* (2018) num estudo genômico com 14 espécies do gênero *Drosophila*.

Essa segunda possibilidade é especialmente interessante, pois o gene $\alpha E5$ tem papel fundamental na codificação de proteínas que auxiliam na desintoxicação de compostos metabólicos dos cactos pelas moscas (Matzkin *et al.*, 2006). Desse modo, concomitantemente a um evento de expansão populacional, é possível que os índices de neutralidade para o gene $\alpha E5$ também estejam indicando um evento de seleção positiva de polimorfismos associados à adaptação das novas populações de *D. serido* à elevada diversidade de cactos hospedeiros endêmicos no bioma Caatinga (Taylor & Zappi, 2004).

O uso de diferentes cactos por moscas do gênero *Drosophila* já foi associado a diferenças na composição genética desses animais. Um estudo genômico com *D. mojavensis* e *D. buzzatii* observou que essas duas espécies estavam sofrendo uma rápida seleção positiva em genes envolvidos na regulação e no catabolismo de algumas das toxinas heterocíclicas encontradas em seus cactos hospedeiros (Guillen *et al.*, 2015).

Comparações transcriptômicas entre populações de moscas cactófilas vivendo em diferentes hospedeiros tanto em *D. mojavensis* quanto em *D. mettleri*, destacaram mudanças transcricionais em vias metabólicas e sensoriais importantes que podem contribuir para a adaptação à dessecação e/ou ao hospedeiro (Matzkin & Markow, 2009; Matzkin *et al.*, 2011; Hoang *et al.*, 2015). No entanto, a interpretação dessas associações ainda é limitada pelo número reduzido de espécies estudadas e, em alguns casos, pela substancial distância filogenética envolvida na comparação (Rane *et al.*, 2019).

A hipótese de diferenciação genética por seleção adaptativa aos cactos hospedeiros também pode ser corroborada pela reconstrução filogenética para os genes *$\alpha E5$* (**Figura 6**) e *GstD1* (**Figura 5**), que mostram uma evidente separação entre as populações do Grupo Nordeste e as populações do Grupo Litoral da hipótese inicial. Como as duas regiões possuem diferentes composições de vegetação xerófitica, é razoável supor com base nos dados apresentados que as espécies cactofílicas que nelas ocorrem também poderiam possuir mecanismos bioquímicos distintos para desintoxicação dos metabólitos secundários de seus hospedeiros. Essa hipótese também é levantada por Ruiz *et al.* (2000).

A inferência de expansão populacional recente também se mostra provável para as populações do Grupo Litoral a partir dos resultados negativos e significativos de F_s de F_u e D de Tajima para o gene COI, apesar de que a análise de *mismatch distribution* para esse grupo denota um padrão de curva bimodal, indicando, a princípio, que as populações do Grupo Litoral evoluíram dentro das premissas do modelo de equilíbrio neutro (Tajima, 1989; Rogers & Harpending, 1992; Nielsen, 2001). Para o gene *$\alpha E5$* , os mesmos índices de neutralidade não possuem valores significativos, mas a respectiva curva de *mismatch distribution* apresenta um padrão unimodal, não sendo possível a rejeição da hipótese de expansão populacional para esse agrupamento (Rogers & Harpending, 1992).

Alternativamente, assim como no caso dos índices de neutralidade supracitados, outros fatores além da história evolutiva de uma dada população podem afetar a distribuição da curva de *mismatch distribution* como, por exemplo, a presença de subpopulações na amostragem, que pode produzir uma curva de padrão bimodal (Rogers & Harpending, 1994), como observado no Grupo Litoral para o gene COI, ou mesmo influências da própria seleção natural sobre as frequências alélicas (Tajima, 1989; Fu, 1997), como sugerido aqui para a história evolutiva do marcador *$\alpha E5$* .

O padrão bimodal observado no Grupo Litoral, nesse caso, pode ocorrer pelo simples fato de que cada subpopulação possui sua própria história demográfica e, portanto, sua própria curva de *mismatch distribution*. Se essas subpopulações entraram em contato secundário recentemente, ou se ocorreu de as mesmas estarem agrupadas em um único conjunto a critério do pesquisador, pode haver um aumento na diversidade genética da população resultante, o que pode levar a um segundo pico na curva de *mismatch distribution* (Rogers & Harpending, 1992; Nielsen, 2001).

Estudos populacionais anteriores utilizando o marcador COI em *D. serido* apontaram a existência de mais subpopulações do que as compreendidas pela hipótese inicial do presente estudo. Morales (2005) sugere a existência de três grupos populacionais em *D. serido*: um grupo formado por indivíduos do interior do Nordeste, no bioma Caatinga, e os outros dois formados por indivíduos ao longo do litoral Atlântico brasileiro, corroborando com os padrões mostrados pelas curvas de *mismatch distribution* (**Figura 16**).

A hipótese de dois eventos de expansão populacional independentes da espécie *D. serido* também foi levantada por Franco (2009), que recuperou um evento de colonização do litoral norte da Bahia (Salvador – BA) e da Chapada da Borborema por meio da *Nested Clade Phylogeographic Analysis* (NCPA) e um segundo evento de expansão das populações desses animais para o litoral de São Paulo, Rio de Janeiro, Espírito Santo e sul da Bahia. Segundo o autor, evidências oriundas tanto de datação baseada em *mismatch distribution* quanto de análises coalescentes evidenciam que esses dois eventos distintos provavelmente ocorreram de forma assíncrona, como também sugere a datação molecular contida na reconstrução filogenética total (**Figura 7**).

Posteriormente, em um estudo que relaciona a história demográfica de grupos cactófilos de *Drosophila* com as mudanças paleoclimáticas do Quaternário na América do Sul, Franco & Manfrin (2013) identificaram uma rota de expansão das populações de *D. serido* saindo da região da Chapada Diamantina para o norte do Brasil, em direção à Chapada de Borborema, e outra para o sul, em direção à região de Salvador – BA. O estudo recente de Barrios-Leal *et al.* (2019), baseado na reconstrução de eventos paleogeográficos para as espécies do “cluster” *D. buzzatii*, sugere múltiplas rotas de expansão populacional para *D. serido*. Contudo, todas essas rotas ocorrem ao redor da área da Chapada Diamantina e são congruentes com os resultados de Franco e Manfrin (2013).

A hipótese de que a região da Chapada Diamantina seja o centro de dispersão das populações de *D. serido* também é compatível com os resultados observados para *D. serido* no presente estudo. Os testes de neutralidade como um todo (**Tabelas 11 e 12; Figura 16**) apontam para uma expansão populacional recente tanto para o Grupo Nordeste quanto para o Grupo Litoral. As redes haplotípicas para os genes COI, *GstD1* e *αE5* (**Figuras, 3, 5 e 6**) evidenciam a mesma expansão para o Grupo Nordeste, em que os haplótipos ancestrais de uma mesma região próxima à Chapada Diamantina (nesse caso, Milagres – BA) se expandem para o norte do Nordeste e também para o sul da Bahia. Isso é indicado, nas redes dos três marcadores, pelos haplótipos ao redor do haplótipo maior, o mais antigo, que se organizam em uma conformação em forma de estrela que indica expansão populacional a partir do haplótipo maior, tal como já foi observado no trabalho de Lavagnini-Pizzo (2015) para a mesma espécie.

No caso das populações de *D. serido* do litoral Atlântico nos estados da Bahia, Rio de Janeiro e São Paulo, Morales (2005) sugere que os eventos de expansão populacional dessas populações podem estar intimamente relacionados aos eventos de expansão da vegetação xerofítica durante os períodos interglaciais do Pleistoceno. A autora também sugere que as populações fundadoras de *D. serido* podem ter migrado desde o interior do estado da Bahia até o litoral por meio de depressões de relevo existentes nos vales dos rios, que têm sua origem na Caatinga e que desaguam no Oceano Atlântico, como o rio Mucuri – BA (Morales, 2005). Os sinais de expansão populacional de *D. serido* sugeridos por Morales (2005) e Franco & Manfrin (2013) são compatíveis com a rota de expansão pelo Norte utilizada por Prado & Gibbs (1993) para explicar a expansão das vegetações abertas da Caatinga durante os períodos interglaciais do Quaternário (Franco, 2009).

O compartilhamento de haplótipos entre populações que foram alocadas em dois clados distintos nas localidades de Itaúnas – SP (N31), Cabrália – BA (N35) e Mucuri – BA (MUC1 e MUC2) conforme observado na rede haplotípica do gene COI (**Figura 12**), é consistente com as rotas inferidas por Morales (2005) e podem indicar uma possível região de contato secundário das duas linhagens evolutivas de *D. serido*, que divergiram por alopatria segundo Franco & Manfrin (2013). Segundo os autores, essa região também é considerada uma zona de contato secundário para outros organismos com diferentes capacidades de dispersão, como anfíbios (Carnaval *et al.*, 2009) e lagartos (Pellegrino *et al.*, 2011), sugerindo uma congruência em larga escala do padrão filogeográfico observado em *D. serido*.

Complementarmente a essas evidências cronológicas e geográficas descritas acima e, levando em conta um detalhado estudo palinológico conduzido na região de Salitre – MG (Mayle, 2006), Franco (2009) sugere que as expansões populacionais observadas em *D. gouveai* e *D. serido* em seu estudo tenham, de fato, acompanhado a expansão das FTSS durante o Pleistoceno, o que corrobora, também, com a hipótese do Arco Pleistocênico (Prado & Gibbs, 1993; Pennington *et al.*, 2000). De modo similar, a mesma conclusão pode ser extrapolada para explicar a história populacional de todas as *Drosophila* cactófilas do “cluster” *buzzatii* – com exceção de *D. seriema* (Franco & Manfrin, 2013; Barrios-Leal *et al.*, 2019).

A idade dos eventos de expansão demográfica, estimada por meio da análise coalescente EBSP, obtida apenas para o Grupo Nordeste, mostrou uma pequena contração do tamanho efetivo populacional entre 70 e 50 mil anos atrás, seguida de um crescimento exponencial iniciado há aproximadamente 40 mil anos e que perdura por todo o Holoceno, congruente com os resultados descritos por Franco & Manfrin (2013) e Barrios-Leal *et al.* (2019). Outras datações utilizando análises coalescentes mostraram que as populações analisadas das espécies cactófilas *D. serido*, *D. gouveai*, *D. borborema* e *D. antonietae* (com exceção apenas de *D. seriema*, que apresenta uma história evolutiva singular) possuem histórias demográficas semelhantes às inferidas para *D. serido* no presente estudo (Franco & Manfrin, 2013).

6. CONCLUSÕES

- Os resultados das análises multilocus são congruentes com a hipótese de estruturação em *D. serido*, mas em um padrão diferente do que foi estabelecido na hipótese inicial de dois grupos populacionais distintos;
- A reconstrução filogenética total recupera a existência de dois grandes clados distintos, um formado pelas populações do Grupo Nordeste e o outro formado pelas populações do Grupo Litoral, consistente com a hipótese inicial de estruturação em *D. serido*;
- É possível que os padrões de diversidade e estruturação genéticos observados para os genes *GstD1* e *αE5* estejam mais relacionados a fatores ecológico-adaptativos do que históricos, como no caso do gene COI;
- É possível que a baixa diversidade genética do *kl-5* esteja ligada à influência de seleção positiva sobre o *fitness* dos machos portadores do gene;
- As redes haplotípicas corroboram com a hipótese de expansão dos grupos da Caatinga a partir da região da Chapada Diamantina e posterior fragmentação numa possível zona de contato da qual faz parte a região de Mucuri – BA, onde ocorre compartilhamento de haplótipos dos dois grupos da hipótese inicial;
- As análises de estruturação populacional e as curvas de *mismatch distribution* sugerem uniformidade genética do Grupo Nordeste e estruturação populacional dentro do Grupo Litoral, que parece ter sofrido múltiplos eventos assíncronos de dispersão ao longo do litoral Atlântico brasileiro;
- A análise EBSP infere que os eventos de expansão demográfica do Grupo Nordeste ocorreram há aproximadamente 40 milhões de anos;
- A história da expansão populacional de *D. serido* parece ter sido influenciada pelos eventos de expansão das FTSS durante o Pleistoceno.

7. REFERÊNCIAS

- AB'SÁBER, A. N. (1977). Os Domínios morfoclimáticos na América do Sul. Primeira aproximação. **Geomorfologia**, 52: 1-24.
- AMARAL, D. T.; MINHÓS-YANO, I; OLIVEIRA, J. V. M. *et al.* (2021). Tracking the xeric biomes of South America: The spatiotemporal diversification of Mandacaru cactus. **Journal of Biogeography**, 48(12): 3085-3103.
- AVISE, J. C. (2000). Phylogeography: the history and formation of species. **Harvard University Press**, 1: 3-9.
- BAIMAI, V.; SENE, F. M.; PEREIRA, M. A. O. R. (1983). Heterochromatin and karyotypic differentiation of some neotropical cactus-breeding species of the *Drosophila repleta* species group. **Genetica**, 60(2): 81-92.
- BARRIOS-LEAL, D. Y.; FRANCO, F. F.; CARVALHO-SILVA, E. C. *et al.* (2018). Deep intraspecific divergence in *Drosophila meridionalis*, a cactophilic member of the New World *Drosophila repleta* group. **Biological Journal of the Linnean Society**, 123: 163-178.
- BARRIOS-LEAL, D. Y.; MATEUS, R. P.; SANTOS, C. G. *et al.* (2021). Plastic Variation in the Phyletic Lineages of Cactophilic *Drosophila meridionalis* and Relation to Hosts as Potential for Diversification. **Neotropical Entomology**, 50: 515-523.
- BARRIOS-LEAL, D. Y.; MENEZES, R. S. T.; RIBEIRO, J. V. *et al.* (2021). A holocenic and dynamic hybrid zone between two cactophilic *Drosophila* species in a coastal lowland plain of the Brazilian Atlantic Forest. **Journal of Evolutionary Biology**, 34(11): 1737-1751.
- BARRIOS-LEAL, D. Y.; NEVES-DA-ROCHA, J; MANFRIN, M. H. (2019). Genetics and Distribution Modeling: The Demographic History of the Cactophilic *Drosophila buzzatii* Species Cluster in Open Areas of South America. **Journal of Heredity**, 110(1): 22-33.
- 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. (2011). Zona de contato secundário: simpatria entre duas espécies de *Drosophila*. **Tese (Doutorado em Ciências) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto**, 134 f.
- BONATELLI, I. A. S.; PEREZ, M. F.; PETERSON, A. T. *et al.* (2014). Interglacial microrefugia and diversification of a cactus species complex: phylogeography and palaeodistributional reconstructions for *Pilosocereus aurisetus* and allies. **Molecular Ecology**, 23: 3044-3063.
- BONO, J. M.; MATZKIN, L. M.; CASTREZANA, S. *et al.* (2008). Molecular evolution and population genetics of two *Drosophila mettleri* cytochrome P450 genes involved in host plant utilization. **Molecular Ecology**, 17(13): 3211-3221.
- BOUCKAERT, R. (2015). Help, BEAST acts weird! (or how to set up rates). Disponível em: <https://beast2.blogs.auckland.ac.nz/tag/clock-rates/> <Acessado em 10/01/2023>.
- BOUCKAERT, R.; VAUGHAN, T. G.; BARIDO-SOTTANI, J. *et al.* (2019). BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis. **PLOS Computational Biology**, DOI: 10.1371/journal.pcbi.1006650.
- CAMPBELL, V.; LEGENDRE, P.; LAPOINTE, F.-J. (2011). The performance of the Congruence Among Distance Matrices (CADM) test in phylogenetic analysis. **BMC Evolutionary Biology**, 11: 64.
- CARREIRA, V.; BOUZAS, S.; PADRÓ, J. *et al.* (2022). P450 gene family mediates allelochemical detoxification and tolerance to alkaloids in cactophilic *Drosophila*. **Entomologia Experimentalis et Applicata**, 170(11): 984-956.
- CARNAVAL, A. C.; HICKERSON, M. J.; HADDAD, C. F. B. *et al.* (2009). Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. **Science**, 323: 785-789.
- CARVALHO, A. B. (2002). Origin and evolution of the *Drosophila* Y chromosome. **Current Opinion in Genetics and Development**, 12(6): 664-668.
- CARVALHO, A.; DOBO, B.; VIBRANOVSKI, M. *et al.* (2001). Identification of five new genes on the Y chromosome of *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences USA**, 98(23): 13225-13230.

- CARVALHO, A. B.; KOERICH, L. B.; CLARK, A. G. (2009). Origin and evolution of Y chromosomes: *Drosophila* tales. **Trends in Genetics**, 25(6): 270-277.
- CASTOE, T. A.; JIANG, Z. J.; GU, W. *et al.* (2008). Adaptive Evolution and Functional Redesign of Core Metabolic Proteins in Snakes. **PLoS One**, DOI: 10.1371/journal.pone.0002201.
- CHIPPINDALE, A. K.; RICE, W. R. Y. (2001). Y chromosome polymorphism is a strong determinant of male fitness in *Drosophila melanogaster*. **Proceedings of the National Academy of Sciences USA**, 98(10): 5677-5682.
- DARRIBA, D.; TABOADA, G. L.; DOALLO, R. *et al.* (2012). jModelTest 2: more models, new heuristics and parallel computing. **Nature Methods**, 9(8): 772.
- DASILVA, M. B.; PINTO-DA-ROCHA, R. (2011). História Biogeográfica da Mata Atlântica: Opiliões (Arachnida) como Modelo para sua Inferência. In: Carvalho, C. J. B. & Almeida, E. A. B. (Eds.). Biogeografia da América do Sul: Padrões e Processos. **Editora Roca, São Paulo**: 221-238.
- DINIZ-FILHO, J. A.; SOARES, T. N.; LIMA, J. S. *et al.* (2013). Mantel test in population genetics. **Genetics and Molecular Biology**, 36: 475-485.
- DE BRITO, R. A. D.; MANFRIN, M. H.; SENE, F. M. (2002b). Nested cladistic analysis of Brazilian populations of *Drosophila serido*. **Molecular Phylogenetics and Evolution**, 22(1): 131-143.
- DRUMMOND, A. J.; RAMBAUT, A.; SHAPIRO, B. *et al.* (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. **Molecular Biology and Evolution**, 22(5): 1185-1192.
- DRUMMOND, A. J.; HO, S. Y. W.; PHILLIPS, M. J. *et al.* (2006). Relaxed phylogenetics and dating with confidence. **PLoS Biology**, 4(5): e88.
- DRUMMOND, A. J.; SUCHARD, M. A.; XIE, D. *et al.* (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. **Molecular Biology and Evolution**, 29(8): 1969-1973.
- DRUMMOND, A. J.; RAMBAUT, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. **BMC Evolutionary Biology**, 7: 214.

- DYER, K. A.; WHITE, B.; BRAY, M. J. *et al.* (2011). Molecular evolution of a Y chromosome to autosome gene duplication in *Drosophila*. **Molecular Biology and Evolution**, 28(3): 1293-1306.
- EDWARDS, S. V.; LIU, L.; PEARL, D. K. (2007). High-resolution species trees without concatenation. **Proceedings of the National Academy of Sciences of the United States of America**, 104: 5841-5936.
- ERSTS, P. J. (Internet) Geographic Distance Matrix Generator (version 1.2.3). **American Museum of Natural History, Center of Biodiversity and Conservation**. Disponível em http://biodiversityinformatics.amnh.org/open_source/gdmg. <Acessado em 10/01/2023>.
- ESGUÍCERO, A. L. H. (2006). Estudo taxonômico e evolutivo em populações de *Drosophila serido* e *Drosophila antonietae* (Diptera; Drosophilidae). **Dissertação de Mestrado, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto**.
- EXCOFFIER, L.; LISCHER, H. E. L. (2010). Arlequin Suite ver 3.5, a New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. **Molecular Ecology Resources**, 10: 564-567.
- EXCOFFIER, L.; SMOUSE, P. E.; QUATTRO, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. **Genetics**, 131(2): 479-491.
- FELSENSTEIN, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, 39: 783-791.
- FONTDEVILA, A.; WASSERMAN, M. (1988). In: FONTDEVILA, A., PLA, C.; HASSON, E.; WASSERMAN, M.; SANCHEZ, A.; NAVEIRA, H. & RUIZ, A. (Eds.). *Drosophila koepferae*: A New Member of the *Drosophila serido* (Diptera, Drosophilidae) Superspecies Taxon. **Annals of the Entomological Society of America**, 81(3): 380-385.
- FRALEY, C. & RAFTERY, A. E. (2006). MCLUST Version3: An R Package for Normal Mixture Modeling and Model-Based Clustering.

- FRANCO, F. F.; JOJIMA, C. L.; PEREZ, M. F. *et al.* (2017). The xeric side of the Brazilian Atlantic Forest: The forces shaping phylogeographic structure of cacti. **Ecology and Evolution**, 7(22): 9281-9293.
- FRANCO, F. F.; MANFRIN, M. H. (2013). Recent demographic history of cactophilic *Drosophila* species can be related to Quaternary palaeoclimatic changes in South America. **Journal of Biogeography**, 40(1): 142-154.
- FRANCO, F. F.; SILVA, E. C. C.; BARRIOS-LEAL, D. Y. *et al.* (2017). The Calibrated Phylogeny of the *Drosophila fasciola* Subgroup (*D. repleta* Group Wasserman) Indicates Neogene Diversification of Its Internal Branches. **Neotropical Entomology**, 46: 537-545.
- FRANCO, F. F.; SILVA-BERNARDI, E. C. C.; SENE, F. M. *et al.* (2010). Intra- and interspecific divergence in the nuclear sequences of the clock gene *period* in species of the *Drosophila buzzatii* cluster. **Journal of Zoological Systematics and Evolutionary Research**, 48(4): 322-331.
- FRANCO, F. F.; SOTO, I. M.; SENE, F. M. *et al.* (2008). Phenotypic variation of the aedeagus of *Drosophila serido* Vilela & Sene (Diptera: Drosophilidae). **Neotropical Entomology**, 37(5): 558-563.
- FRANCO, F. F., MANFRIN, M. H. (2013). Recent demographic history of cactophilic *Drosophila* species can be related to the Quaternary paleoclimatic changes in South America. **Journal of Biogeography**, 40: 142-154.
- FRANK, M. R.; FOGLEMAN, J. C. (1992). Involvement of cytochrome P450 in host-plant utilization by Sonoran Desert *Drosophila*. **Proceedings of the National Academy of Sciences USA**, 89(24): 1198-2002.
- FRANSAK, R. F. (2011). Filogenia do complexo *Drosophila buzzatii* (Grupo *repleta*): inferências de análises multilocus mitocondriais e nucleares. **Dissertação (Mestrado em Genética) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto**, 100 f.
- FU, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. **Genetics**, 147: 915-925.
- FUTUYMA, D. J.; KIRKPATRICK, M. (2017). Evolution. **Sinauer Associates, Inc**: 428-430.

- GEPNER, J.; HAYS, T. S. (1993). A fertility region on the Y chromosome of *Drosophila melanogaster* encodes a dynein microtubule motor. **Proceedings of the National Academy of Sciences USA**, 90(23): 11132-11136.
- GORDO, I.; CHARLESWORTH, B. (2000). The degeneration of asexual haploid populations and the speed of Muller's ratchet. **Genetics**, 154(3): 1379-1387.
- GUILLEN, Y.; RIUS, N.; DELPRAT, A. *et al.* (2015). Genomics of ecological adaptation in cactophilic *Drosophila*. **Genome Biology and Evolution**, 7(1):349-366.
- HARPENDING, H.C.; SHERRY, S. T.; ROGERS, A. R. *et al.* (1993). Genetic structure of ancient human populations. **Current Anthropology**, 34: 483-496.
- HEBERT, D. N.; CYWINSKA, A.; BALL, S. L. *et al.* (2003). Biological identifications through DNA barcodes. **Proceedings: Biological Sciences**, 270(1512):313-321.
- HEDRICK, P. W. (2007). A standardized genetic differentiation measure. **Evolution**, 59(8): 1633-1638.
- HELED, J.; DRUMMOND, A. J. (2008). Bayesian inference of population size history from multiple loci. **BMC Evolutionary Biology**, 8: 289.
- HOANG, K.; MATZKIN, L. M.; BONO, J. M. (2015). Transcriptional variation associated with cactus host plant adaptation in *Drosophila mettleri* populations. **Molecular Ecology**, 24(20):5186-5199.
- HUECK, K. (1972). As Florestas da América do Sul. **Editora Polígono**.
- IWATA, H.; UKAI, Y. (2002). SHAPE: A Computer Program Package for Quantitative Evaluation of Biological Shapes Based on Elliptic Fourier Descriptors. **Journal of Heredity**, 93(5): 384-385.
- JOMBART, T. (2008). *adegenet*: a R package for the multivariate analysis of genetic markers. **Bioinformatics**, 24(11): 1403-1405.
- JOMBART, T.; COLLINS, C. (2015). A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0. Disponível em: <https://adegenet.r-forge.r-project.org/files/tutorial-dapc.pdf>. <Acessado em 10/01/2023>.

- JONES, G. (2017). Algorithmic improvements to species delimitation and phylogeny estimation under the multispecies coalescent. **Journal of Mathematical Biology**, 74: 447-467.
- KANESHIRO, K. Y. (1969). A study of the relationships of Hawaiian *Drosophila* species based on external male genitalia. **The University of Texas Publication**, 6918: 55-70.
- KIMURA, M. (1968). Evolutionary Rate at the Molecular Level. **Nature**, 217: 624-626.
- KOERICH, L.; WANG, X.; CLARK, A. *et al.* (2008). Low conservation of gene content in the *Drosophila* Y chromosome. **Nature**, 456(7224): 949-951.
- KOSER, J. R.; BARRIOS-LEAL, D. Y.; MENEZES, R. S. T. *et al.* (2020). Independent Evolutionary Responses to Quaternary Landscape and Climatic Changes of Two Codistributed Cactophilic *Drosophila* Species (*Drosophila repleta* group) in Open Areas of South America. **Annals of the Entomological Society of America**, 113(5): 389-397.
- KUHNER, M. K. (2009). Coalescent genealogy samplers: windows into population history. **Trends in Ecology & Evolution**, 24(2): 86-93.
- KUMAR, S.; STECHER, G.; LI, M. *et al.* (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. **Molecular Biology and Evolution**, 35(6): 1547-1549.
- LANDIS, J. R.; KOCH, G. G. (1977). The measurement of observer agreement for categorical data. **Biometrics**, 33: 159-174.
- LAVINA, E. L.; FAUTH, G. (2011). Evolução Geológica da América do Sul nos Últimos 250 Milhões de Anos. In: Carvalho, C. J. B. & Almeida, E. A. B. (Eds.). *Biogeografia da América do Sul: Padrões e Processos*. **Editores Roca, São Paulo**: 3-13.
- LAVOR, P.; CALVENTE, A.; VERSIEUX, L. M. *et al.* (2018). Bayesian spatio-temporal reconstruction reveals rapid diversification and Pleistocene range expansion in the widespread columnar cactus *Pilosocereus*. **Journal of Biogeography**, 46(1): 238-250.
- LEDO, R. M. D.; COLLI, G. R. (2017). The historical connections between the Amazon and the Atlantic Forest revisited. **Journal of Biogeography**, 44(11): 2551-2563.

- LÓPEZ-OLMOS K.; MARKOW T. A.; MACHADO C. A. (2017). Evolution of GSTD1 in Cactophilic *Drosophila*. **Journal of Molecular Evolution**, 84(5-6):285-294.
- MACHADO, C. A.; HEY, J. (2003). The causes of phylogenetic conflict in a classic *Drosophila* species group. **Proceedings of the Royal Society London B**, 270: 1193-1202.
- MADDISON, W. P. (1997). Gene Trees in Species Trees. **Systematic Biology**, 46(3): 523–536.
- MANFRIN, M. H.; DE BRITO, R. O. A.; SENE, F. M. (2001). Systematics and evolution of the *Drosophila buzzatii* (Diptera: Drosophilidae) cluster using mtDNA. **Annals of the Entomological Society of America**, 94(3): 333-346.
- MANFRIN, M. H.; SENE, F. M. (2006). Cactophilic **Drosophila** in South America: A model for evolutionary studies. **Genetica**, 126(1): 57-75.
- MARTINS, F. M.; DOMINGUES, M. V. (2011). Filogeografia. In: Carvalho, C. J. B. & Almeida, E. A. B. (Eds.). Biogeografia da América do Sul: Padrões e Processos. **Editora Roca, São Paulo**: 137-150.
- MATZKIN, L. M. (2012). Population transcriptomics of cactus host shifts in *Drosophila mojavensis*. **Molecular Ecology**, 21(10): 2428-2439.
- MATZKIN, L. M. (2008). The molecular basis of host adaptation in cactophilic *Drosophila*: molecular evolution of a Glutathione S-Transferase gene (GstD1) in *Drosophila mojavensis*. **Genetics**, 178(2): 1073-1083.
- MATZKIN, L. M.; JOHNSON, S.; PAIGHT, C. *et al.* (2011). Dietary protein and sugar differentially affect development and metabolic pools in ecologically diverse *Drosophila*. **The Journal of Nutrition**, 141(6):1127–1133.
- MATZKIN, L. M.; MARKOW, T. A. (2009). Transcriptional regulation of metabolism associated with the increased desiccation resistance of the cactophilic *Drosophila mojavensis*. **Genetics**, 182(4):1279–1288.
- MATZKIN, L. M.; WATTS, T. D.; BITLER, B. G. *et al.* (2006). Functional genomics of cactus host shifts in *Drosophila mojavensis*. **Molecular Ecology**, 15(14): 4635-4643.

- MILLER, J.M.; CULLINGHAM, C.I; PEERY, R.M. (2020). The influence of a priori grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. **Heredity (Edinb.)**, 125 269-280.
- MORAES, E. M.; YOTOKO, K.; MANFRIN, M. H. *et al.* (2009). Phylogeography of the cactophilic species *Drosophila gouveai*: demographic events and divergence timing in dry vegetation enclaves in eastern Brazil. **Journal of Biogeography**, 36(11): 2136-2147.
- MORALES, A. C. (2005). Análise histórico-evolutiva de populações das espécies cactofílicas *Drosophila serido* e *Drosophila antonietae* (Diptera, Drosophilidae). **Tese (Doutorado em Ciências) – Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto**, 110 f.
- MOREYRA, N. N.; ALMEIDA, F. C.; ALLAN, C. *et al.* (2022). Phylogenomics provides insights into the evolution of cactophily and host plant shifts in *Drosophila*. **Molecular Phylogenetics and Evolution**, 178(1): 107653.
- NIELSEN, R. (2001). Statistical tests of selective neutrality in the age of genomics. **Heredity**, 86: 641-647.
- ODEN, N. L.; SOKAL, R. R. (1992). An investigation of three-matrix permutation tests. **Journal of Classification**, 9: 275-290.
- OLIVEIRA, C. C.; MANFRIN, M. H.; SENE, F. M. *et al.* (2011). Variations on a theme: diversification of cuticular hydrocarbons in a clade of cactophilic *Drosophila*. **BMC Evolutionary Biology**, 11: 179.
- PATTERSON, J. T.; WHEELER, M. R. (1942). Description of new species of the subgenera *Hirtodrosophila* and *Drosophila*. **The University of Texas Publication**. 4213: 67-109.
- PELLEGRINO, K. C. M.; RODRIGUES, M. T.; HARRIS, D. J. *et al.* (2011). Molecular phylogeny, biogeography and insights into the origin of parthenogenesis in the Neotropical genus *Leposoma* (Squamata: Gymnophthalmidae): ancient links between the Atlantic Forest and Amazonia. **Molecular Phylogenetics and Evolution**, 61: 446-459.

- PENNINGTON, R. T.; PRADO, D. E.; PENDRY, C. A. (2000). Neotropical seasonally dry forests and Quaternary vegetation changes. **Journal of Biogeography**, 27(2): 261-273.
- POR, F. D. (1992). Sooretama – The Atlantic Rain Forest of Brazil. **SPB Academic Publishing**: 130.
- POSADA, D.; BUCKLEY, T. R. (2004). Model selection and model averaging in phylogenetics: advantages of Akaike Information Criterion and Bayesian Approaches over Likelihood Ratio Tests. **Systematic Biology**, 53(5): 793-808.
- PRADO, D. E.; GIBBS, P. E. (1993) Patterns of species distributions in the dry seasonal forests of South America. **Annals of the Missouri Botanical Garden**, 80: 902-927.
- RAMBAUT, A. (2010). FigTree v1.3.1. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh. Disponível em: <http://tree.bio.ed.ac.uk/software/figtree/> <Acessado em 30/01/2023>.
- RAMBAUT, A.; DRUMMOND, A. J.; XIE, D. *et al.* (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. **Systematic Biology**, 67(5): 901-904.
- RANE, R. V.; CLARKE, D. F.; PEARCE, S. L. *et al.* (2018). Detoxification Genes Differ Between Cactus-, Fruit- and Flower-Feeding *Drosophila*. **Journal of Heredity**, 110(1): 80-91.
- RANE, R. V.; PEARCE, S. L.; LI, F. *et al.* (2019). Genomic changes associated with adaptation to arid environments in cactophilic *Drosophila* species. **BMC Genomics**, 20(1):52.
- ROGERS, A. R.; HARPENDING, H. C. (1992). Population growth makes waves in the distribution of pairwise differences. **Molecular Biology and Evolution**. 9: 552-569.
- RONQUIST, F.; HUELSENBECK, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. **Bioinformatics**, 19(12): 1572-1574.
- ROUSSET, F. (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. **Genetics**, 145(4): 1219-1228.
- ROZAS, J.; FERRER-MATA, A.; SÁNCHEZ-DELBARRIO, J. C. *et al.* (2017). DnaSP v6: DNA Sequence Polymorphism Analysis of Large Datasets. **Molecular Biology and Evolution**, 34: 3299-3302.

- RUIZ, A.; CANSIAN, A. M.; KUHN, G. C. S. *et al.* (2000). The *Drosophila serido* speciation puzzle: putting new pieces together. **Genetica**, 108(3): 217-227.
- RUIZ, A.; WASSERMAN, M. (1993). Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. **Heredity**, 70(6): 582-596.
- SANTOS, M. H. (2011). As rotas de dispersão de *Drosophila buzzatii* na América do Sul. **Tese (Doutorado em Genética) - Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto**, 128f.
- SANTOS, M. H.; HASSON, E.; SENE, F. M. *et al.* (2009). Esterase-5 gene sequences from the *Drosophila buzzatii* cluster. **Drosophila Information Service**, 92: 44-56.
- SILVA, G. A. R.; ANTONELLI, A.; LENDEL, A. *et al.* (2018). The impact of early Quaternary climate change on the diversification and population dynamics of a South American cactus species. **Journal of Biogeography**, 45(1): 76-88.
- SILVA, A. F. G. & SENE, F. M. (1991). Morphological geographic variability in *Drosophila serido* (Diptera, Drosophilidae). **Revista Brasileira de Entomologia**, 35: 455-468.
- SKALETSKY, H.; KURODA-KAWAGUCHI, T; MINX, P. *et al.* (2003). The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. **Nature**, 423(6942): 825-837.
- SLATKIN, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. **Evolution**, 47(1): 264-279.
- TAJIMA, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. **Genetics**, 123: 584-595.
- TAYLOR, N.; ZAPPI, D. (2004). Cacti of Eastern Brazil. **The University of Chicago Press**.
- TIA, J. A. (2022). Guidelines for standardizing the applications of discriminant analysis of principal components to genotype data. **Molecular Ecology Resources**, 00: 1-16.
- TIDON-SKLORZ, R.; SENE, F. M. (1995a). *Drosophila seriema*: a new member of the *Drosophila serido* (Diptera, Drosophilidae) superspecies taxon. **Annals of the Entomological Society of America**, 88(1): 139-142.
- TIDON-SKLORZ, R.; SENE, F. M. (2001). Two New Species of the *Drosophila serido* Sibling Set (Diptera, Drosophilidae). **Iheringia (Série Zoologia)**, 90: 141-146.

- TOSI, D.; SENE, F. M. (1989). Further studies on chromosomal variability in *Drosophila serido* (Diptera, Drosophilidae). **Revista Brasileira de Genética**, 1: 729-746.
- VILELA, C.R.; SENE, F. M. (1977). Two new neotropical species of the *repleta* group of the genus *Drosophila* (Diptera, Drosophilidae). **Papéis Avulsos de Zoologia**, 30(20): 295-299.
- VILLANEA, F. A., KITCHEN, A; KEMP, B. M. (2020). Review: Applications of Bayesian Skyline Plots and Approximate Bayesian Computation for Human Demography. **Human Biology Open Access Pre-Prints**: 167.
- WARD, R. D.; HOLMES, B. H. (2007). An analysis of nucleotide and amino acid variability in the barcode region of cytochrome c oxidase I (cox1) in fishes. **Molecular Ecology Resources**, 7(6): 899-907.
- WASSERMAN, M.; RICHARDSON, R. H. (1987). Evolution of Brazilian *Drosophila mulleri* complex species. **Journal of Heredity**, 78(5): 282-286.
- WRIGHT, S. (1943). Isolation by Distance. **Genetics**, 28(2): 114-138.
- YULE, G. U. (1924). A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. **Philosophical Transactions of the Royal Society of London. Series B**, 213: 21-87.
- ZANELLA, F. C. V. (2011). Evolução da Biota da Diagonal de Formações Abertas Secas da América do Sul. In: Carvalho, C. J. B. & Almeida, E. A. B. (Eds.). Biogeografia da América do Sul: Padrões e Processos. **Editores Roca, São Paulo**: 198-220.