



UNIVERSIDADE ESTADUAL DO CENTRO-OESTE
SETOR DE CIÊNCIAS AGRÁRIAS E AMBIENTAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA EVOLUTIVA, PPGBIOEVOL
Associação Ampla entre a UEPG e a UNICENTRO

BRUNA CAROLINE FABIANI

Otimização da amostragem para estudos populacionais com locos de microssatélites em espécies Neotropicais de *Drosophila* (Diptera, Drosophilidae)

Guarapuava

2022

BRUNA CAROLINE FABIANI

**Otimização da amostragem para estudos populacionais com locos de microssatélites em espécies
Neotropicais de *Drosophila* (Diptera, Drosophilidae)**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Biologia Evolutiva da Universidade Estadual do Centro-Oeste em associação com a Universidade Estadual de Ponta Grossa, como parte dos requisitos para a obtenção do título de mestre em Ciências Biológicas (Área de Concentração em Biologia Evolutiva).

Orientadora

Profa. Dra. Luciana Paes de Barros Machado

Coorientadora

Profa. Dra. Emanuele Cristina Gustani Buss

Guarapuava

Catálogo na Publicação
Rede de Bibliotecas da Unicentro

F118o Fabiani, Bruna Caroline
Otimização da amostragem para estudos populacionais com locos de microssatélites em espécies Neotropicais de *Drosophila* (Diptera, Drosophilidae) / Bruna Caroline Fabiani. -- Guarapuava, 2022.
xi, 85 f. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual do Centro-Oeste, em Associação com a UEPG, Programa de Pós-Graduação em Biologia Evolutiva. Área de Concentração: Biologia Evolutiva, 2022.

Orientadora: Luciana Paes de Barros Machado
Coorientadora: Emanuele Cristina Gustani Buss
Banca examinadora: Maura Helena Manfrin, Maria Cláudia Colla Ruvolo Takasusuki

Bibliografia

1. *D. ornatifrons*. 2. *D. antonietae*. 3. *D. prosaltans*. 4. *D. sturtevanti*. 5. Genética de populações. 6. Marcador genético SSR. I. Título. II. Programa de Pós-Graduação em Biologia Evolutiva.

CDD 574

Aos meus pais, Marli e Carlinhos, por serem os pais de todas minhas conquistas.

“Por vezes sentimos que aquilo que fazemos
não é senão uma gota de água no mar. Mas o
mar seria menor se lhe faltasse uma gota”.
(Madre Teresa de Calcuta)

AGRADECIMENTOS

Gostaria de agradecer, primeiramente, à minha família. Especialmente aos meus pais, Marli e José Carlos, que sempre me apoiaram em minhas decisões. Meus maiores exemplos de vida, me sinto orgulhosa e sobretudo, privilegiada, por ter vocês na minha vida.

À minha irmã Mariana por todo amor, carinho, compreensão e apoio em tantos momentos difíceis desta caminhada que se iniciou com duas adolescentes saindo de casa sozinhas e finalizadas com duas mulheres profissionais prontas para alcançar o mundo.

À minha orientadora, Prof.^a Dr.^a Luciana Paes de Barros Machado, pelo papel fundamental na minha formação acadêmica, pelo grande exemplo de mulher pesquisadora, por todo o apoio e incentivo em inúmeros dias. Obrigada por toda paciência nesse período e por ter me aceitado de maneira tão afetuosa, mesmo eu não tendo nenhum vínculo anterior com o laboratório.

Ao Prof.^o Dr. Rogério Pincela Mateus, pela orientação fundamental para o andamento do projeto, principalmente nos cálculos estatísticos.

À minha co-orientadora Prof.^a Dr.^a Emanuele Cristina Gustani Buss pelas sugestões e disponibilidade para tirar dúvidas em todas as questões relacionadas ao projeto.

Ao Prof.^o Dr. Paulo Roberto da Silva, pelo exemplo de dedicação aos assuntos relacionados ao Programa de Pós-Graduação em Biologia Evolutiva da UNICENTRO.

Aos membros da banca examinadora, Prof.^a Dr.^a Maura Helena Manfrin e Prof.^a Dr.^a Maria Cláudia Colla Ruvolo Takasusuki por aceitarem participar e colaborar com esta dissertação.

Ao Programa de Pós-Graduação em Biologia Evolutiva da UNICENTRO e a todos os professores que proporcionam um ambiente de qualidade para a formação acadêmica.

A todos meus colegas de graduação e pós-graduação, por todos os momentos vividos e as histórias trocadas. Obrigada por contribuírem no meu amadurecimento.

A toda equipe da LL Ambiental Ltda., especialmente a minha coordenadora Jéssica Sieklicki, pela oportunidade de amadurecimento profissional e pessoal, como bióloga e como pessoa.

RESUMO

Definir o tamanho da amostra de indivíduos de populações naturais que possa representar de maneira fidedigna a variabilidade genética é um importante fator para o estudo da estrutura genética populacional. O baixo número de indivíduos amostrados pode trazer conclusões errôneas, enquanto um número muito alto, além de ser muito laborioso, pode ser um desafio à coleta para alguns grupos de animais, e também pode representar desperdício de recursos financeiros aplicados na análise. Desta maneira, torna-se relevante a determinação do tamanho amostral adequado que seja representativo da variação natural, com o menor custo e esforço laboratorial possível. O objetivo deste trabalho foi determinar o tamanho amostral que representa a variabilidade genética de locos de microssatélites para estudos populacionais de espécies de *Drosophila* da região Neotropical, com otimização dos recursos e esforços de coleta e laboratorial, o N ideal. Foram utilizados 100 genótipos de locos de microssatélites de indivíduos de populações naturais para cada uma de quatro espécies de *Drosophila*: *D. ornatifrons*, *D. antonietae*, pertencentes ao subgênero *Drosophila*, e *D. prosaltans* e *D. sturtevantii*, do subgênero *Sophophora*. Das matrizes originais, foram sorteadas 50 subamostras para os seguintes tamanhos populacionais (N): 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 e 95 indivíduos. Para todas as 50 subamostras para cada N, foi avaliada a porcentagem da detecção de todos os alelos representativos (frequência $\geq 5\%$) da amostra original, heterozigosidade média esperada (H_e), e o índice de diferenciação genética (F_{st}) com a amostra original. O N ideal para cada espécie foi estabelecido a partir da análise não paramétrica dos valores não significativos de H_e e F_{st} pareados entre os diferentes tamanhos populacionais, e para a detecção dos alelos representativos, foi considerada a presença destes em ao menos 95% das 50 subamostras para no mínimo 70% dos locos de microssatélites. Após a determinação do N ideal, os alelos raros (frequência $< 5\%$) da matriz original foram substituídos por outros genótipos com alelos não raros (também provenientes de populações naturais), e desta matriz sem alelos raros foram sorteadas 50 subamostras com o N ideal. As composições genéticas das 50 subamostras do N ideal com os alelos raros foi comparada com as do N ideal sem alelos raros, por meio do programa STRUCTURE. De acordo com os parâmetros estabelecidos, os tamanhos amostrais para *D. ornatifrons*, N = 25, para *D. antonietae*, N = 30, *D. prosaltans* e *D. sturtevantii*, N = 35, foram considerados ideais para a análise da estrutura genética. Não foi observado efeito acentuado dos alelos raros na distribuição das proporções dos grupos genéticos dos indivíduos, determinados pelo STRUCTURE, dentro das classes de agrupamento propostas. As espécies do subgênero *Drosophila* provavelmente apresentaram menor N ideal devido a menor diversidade genética, resultado da especificidade de suas dietas, proteica no caso de *D. ornatifrons* e cactofila para *D. antonietae*. Por outro lado, o generalismo das espécies do subgênero *Sophophora*, *D. prosaltans* e *D. sturtevantii*, resulta em maior diversidade genética e, conseqüentemente, maior N ideal para análises populacionais com locos de microssatélites. Os resultados deste trabalho também mostraram que o mesmo N ideal pode ser empregado em caso de locos de microssatélites transferidos entre espécies próximas filogeneticamente, como foi o caso dos locos descritos para *D. sturtevantii* e transferidos para *D. prosaltans*.

Palavras-chave: *D. ornatifrons*, *D. antonietae*, *D. prosaltans*, *D. sturtevantii*, genética de populações, marcador genético SSR.

ABSTRACT

The definition of the sample size of individuals from natural populations that can reliably represent genetic variability is an important factor for the study of population genetic structure. The low number of specimens sampled can lead to inaccurate conclusions, while a very high number can be a challenge to collect some groups of animals, in addition of being very laborious and could also represent a waste of financial resources. In this way, to determine the appropriate sample size that is representative of the natural variation, with the lowest possible cost and laboratory effort, becomes relevant. The objective of this work was to determine the sample size that represents the genetic variability of microsatellite loci for population studies of *Drosophila* species from the Neotropical region, with optimization of resources and collection and laboratory efforts. We used 100 individuals from natural populations genotyped for microsatellite loci for each of four *Drosophila* species: *D. ornatifrons*, *D. antonietae*, belonging to the *Drosophila* subgenus, and *D. prosaltans* and *D. sturtevantii*, of the *Sophophora* subgenus. From the original matrices, 50 subsamples were randomly obtained for the following population sizes (N): 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 and 95 individuals. For all 50 replicates of each N, the percentage of detection of all non-rare alleles (frequency $\geq 5\%$) from the original sample, expected mean heterozygosity (H_e), and the genetic differentiation (F_{st}) to the original 100 individuals matrix were evaluated. The ideal N for each species was established using a non-parametric analysis of the non-significant values of H_e and F_{st} paired between the different population sizes, and for the detection of the non-rare alleles, the presence of these in at least 95% of the 50 subsamples in at least 70% of the loci was considered. After determining the ideal N, the rare alleles (frequency $< 5\%$) of the original matrix were replaced by other genotypes with non-rare alleles (also from natural populations), and from this matrix without rare alleles, 50 subsamples with the ideal N were randomly selected. The genetic compositions of the 50 replicates of ideal N with rare alleles were compared with those of ideal N without rare alleles, using the STRUCTURE software. According to the established parameters, the sample sizes for *D. ornatifrons*, $N = 25$, *D. antonietae*, $N = 30$, *D. prosaltans* and *D. sturtevantii*, $N = 35$, were considered ideal for the analysis of the genetic structure. The presence of rare alleles produced inconspicuous differences in the distribution of the genetic clusters of individuals, obtained in the STRUCTURE software, within the grouping classes proposed. Species of the *Drosophila* subgenus probably presented lower ideal N due to lower genetic diversity, as a result of the specificity of their diets, more protein in the case of *D. ornatifrons* and cactophily for *D. antonietae*. On the other hand, the generalism of the *Sophophora* subgenus species, *D. prosaltans* and *D. sturtevantii*, results in greater genetic diversity and, consequently, greater ideal N for population analyzes using microsatellite loci. The results of this work also showed that the same ideal N can be used in case of microsatellite loci transferred between phylogenetically close species, as was the case of the loci described for *D. sturtevantii* and transferred to *D. prosaltans*.

Keywords: *D. ornatifrons*, *D. antonietae*, *D. prosaltans*, *D. sturtevantii*, Population Genetics, SSR genetic marker.

SUMÁRIO

1. INTRODUÇÃO	7
2. REVISÃO BIBLIOGRÁFICA	8
2.1 A caracterização da variabilidade genética das populações naturais	8
2.2 Espécies de <i>Drosophila</i> utilizadas como modelos na determinação do tamanho amostral ideal para estudos populacionais com locos de microssatélites	12
3. OBJETIVOS	15
3.1 Objetivo geral	15
3.2 Objetivos específicos	15
4. MATERIAL E MÉTODOS	15
4.1. Determinação do tamanho amostral ideal	15
4.2. Influência dos alelos raros na distribuição da composição genética das subamostras do N ideal	17
4.3. Análises estatísticas	18
5. RESULTADOS	18
6. CONSIDERAÇÕES FINAIS	42
7. REFERÊNCIAS BIBLIOGRÁFICAS	42

1. INTRODUÇÃO

O gênero *Drosophila* da família Drosophilidae (Diptera) é altamente diverso, com cerca de 2.000 espécies descritas e muitas ainda a serem identificadas, especialmente na região Neotropical. Algumas espécies deste gênero são muito utilizadas como modelo biológico para diferentes áreas das Ciências Biológicas há mais de um século, principalmente na Genética, por suas características favoráveis a experimentação, como rápido ciclo de vida, alta fecundidade, facilidade de coleta, manuseio e manutenção em laboratório, entre outros. No entanto, o potencial destes organismos em estudos da composição da fauna, interações ecológicas, e da relação da distribuição da variabilidade genética em seus ambientes naturais começou a ser explorado a partir de meados do século XX, e há muito ainda a ser investigado nas espécies da região Neotropical, onde existe a maior diversidade de espécies.

Para um entendimento amplo sobre da distribuição da variação encontrada, é necessário relacionar os dados genéticos com o contexto ecológico de cada espécie, e compreender os padrões e mecanismos evolutivos que concatenam a genética do organismo, a estrutura populacional, e o ambiente no qual ele vive, tema da Genética de Populações. Além disso, o conhecimento da variabilidade genética das populações naturais é um importante aspecto para políticas de conservação da natureza, visto que a adaptação para diferentes condições ambientais é possível somente quando existe variação, desse modo, os processos evolutivos dependem da variação pré-existente surgida por mutação e mantida ao longo das gerações pelos cruzamentos entre os indivíduos de uma população.

A variabilidade genética das populações naturais pode ser avaliada por uma gama de marcadores moleculares. Os marcadores neutros polimórficos são mais indicados, pois se um loco está ligado a uma região genômica a qual é alvo de seleção natural, ele deverá mostrar um desvio em relação à neutralidade esperada. O marcador de DNA microsatélite é hipervariável e, na maioria das vezes, neutro (por ser detectado preferencialmente nas regiões intergênicas), o que o torna adequado para aplicação nesses estudos de Genética de Populações.

Delimitar o tamanho da amostra para estudos de Genética de Populações, de modo que seja uma representação fiel da população natural, mas ao mesmo tempo seja exequível em termos de esforços laboratoriais e de recursos financeiros, é muito importante, especialmente para algumas espécies animais que apresentam dificuldades para coleta de espécimes e/ou estão ameaçadas de extinção. O objetivo deste trabalho foi

determinar o tamanho amostral ideal para estudos populacionais da variabilidade de locos de microssatélites de espécies de *Drosophila*, por meio de genótipos amostrados de populações naturais, para, assim, poder contribuir com a otimização dos recursos empregados nestas análises para insetos da Ordem Diptera, ou ainda de organismos com características ecológicas e demográficas, e de distribuição geográfica semelhantes a das espécies utilizadas neste trabalho.

2. REVISÃO BIBLIOGRÁFICA

2.1 A caracterização da variabilidade genética das populações naturais

Darwin já destacava a importância da variação na modificação das populações ao longo do tempo, pois seu primeiro postulado para a alteração das populações por meio da seleção natural é a existência de variação individual, seguida de que essa variação seja, ao menos parcialmente, transmitida para a prole, que alguns indivíduos apresentem maior probabilidade de sobrevivência que outros, e, finalmente, que essa sobrevivência diferencial não seja aleatória (FREEMAN; HERRON, 2009; DARWIN, 1859, tradução DUARTE; DUARTE, 2014).

Assim, desde a formulação da Síntese Moderna da Evolução entre os anos 30 e 40 do século XX, o estudo das mudanças genóticas, constituem a base das teorias evolutivas e das hipóteses para a especiação, e o conhecimento desta variabilidade genética é fundamental para a compreensão destas teorias, além de também ser uma medida das condições das populações naturais em suas respostas a alterações no ambiente, importante para estudos de conservação.

A variação genética das populações naturais foi inicialmente avaliada por meio de caracteres morfológicos discretos, e o termo “polimorfismo genético” foi definido nesta fase por Ford (1940) como a existência concomitante em uma população de duas formas descontínuas, em proporção tal que a frequência da forma mais rara não pode ser explicada por mutações recorrentes. A partir da década de 60 do século passado, a técnica de eletroforese permitiu avaliar a variabilidade genética por meio da expressão direta dos produtos gênicos, as proteínas, e foi verificado que a variabilidade genética é muito maior do que se supunha com a análise dos caracteres morfológicos, e sua manutenção não poderia ser explicada tão somente por seleção natural. Destas constatações, foi proposta a Teoria Neutralista para a Evolução Molecular que ressalta a importância da deriva genética aleatória na fixação das variantes moleculares ao longo do tempo (KIMURA, 1968; KING; JUKES, 1969; KIMURA; OHTA, 1971).

A descoberta da estrutura do DNA possibilitou o desenvolvimento de técnicas de sequenciamento desta molécula, começando pelo método de Sanger até as técnicas de sequenciamento de nova geração (*Next Generation Sequencing*) a partir de 2005 (WATSON; CRICK, 1953; SANGER; NICKLEN; COULSON, 1977; SANTOS et al., 2013). Estas tecnologias vêm demonstrando que a variabilidade nas sequências de nucleotídeos do DNA entre indivíduos de uma população é ainda maior que a verificada pela eletroforese de proteínas, e diferentes marcadores moleculares foram desenvolvidos por meio dos dados de sequenciamentos de diferentes “Projetos Genomas”, e podem ser aplicados sem que haja a necessidade de sequenciamento de todo genoma, contribuindo para os avanços dos estudos de Genética de Populações.

Os dados de sequenciamento de DNA aliado à técnica da reação em cadeia da polimerase (PCR – *Polymerase Chain Reaction*) (MULLIS et al., 1986) permitiu o desenvolvimento dos marcadores codominantes microssatélites, inicialmente aplicados na análise de paternidade (MOXON; WILLS, 1999), e sua utilização para estudos genético populacionais (ESTOUP et al, 1995; CHAMBERS; MACAVOY, 2000). Microssatélites, também conhecidos pelas siglas STR (*Short Tandem Repeats*), ou SSR (*Simple Sequence Repeats*). podem ser definidos como repetições em tandem de unidades curtas (até seis bases), denominadas motes, de sequências no DNA (HANCOCK, 1999). O comprimento mínimo de um microssatélite seria de oito nucleotídeos, pois sequências de tamanho inferiores ocorrem na mesma frequência esperada se elas fossem encontradas ao acaso no genoma (ROSE; FALUSH, 1998). Ou seja, um microssatélite composto por unidade de repetição com dois nucleotídeos, chamada unidade de repetição dinucleotídica, deve estar repetido ao menos quatro vezes em tandem para ser considerada uma sequência microssatélite, uma trinucleotídica no mínimo três, uma tetra, penta ou hexanucleotídica ao menos duas.

Os microssatélites apresentam distribuição ubíqua nos genomas procariotos e eucariotos, e apesar de algumas vezes estarem localizados em regiões codificadoras, ocorrem preferencialmente nas regiões intergênicas (HANCOCK, 1999; SREENU et al., 2003; LI et al., 2004), sendo, portanto, menos suscetíveis a ação da seleção natural e, assim, considerados seletivamente neutros, característica vantajosa para estudo populacionais, como pontuado anteriormente. Outra propriedade importante dos locos de microssatélites com benefícios na aplicação em Genética de populações, é seu polimorfismo e hipervariabilidade. Os microssatélites são separados por tamanho em gel por eletroforese após amplificação dos locos via PCR, e a variação no tamanho entre

alelos de um indivíduo, ou entre indivíduos de uma população é, em geral, devido ao número de repetições dos motes. A alta variação no número de repetições dos locos de microssatélites se deve a dois mecanismos principais: *crossing-over* desigual, que é o resultado da recombinação entre homólogos não alinhados corretamente em regiões de DNA repetitivo; e, o evento mais provável, DNA *slippage* durante a replicação, em que alças podem ser formadas na fita molde ou na fita que está sendo sintetizada como resultado da perda do pareamento nas sequências curtas repetidas em tandem. Quando essas alças não são corrigidas pelo mecanismo de reparo específico (reparo *mismatch*), a duplicação do DNA nesta condição pode resultar em uma unidade de repetição a mais, se a alça ocorre na fita recém-sintetizada, ou uma repetição a menos, caso a formação da alça tenha ocorrido na fita molde (LEVINSON; GUTMAN, 1987a; WRIGHT, 1994; GRAUR; LI, 2000).

A aplicação de locos microssatélites em estudos populacionais requer uma amostra de indivíduos que represente a variação natural, sem acarretar em um esforço muito grande de coleta, o que pode ser problemático para espécies em risco ou de difícil captura, especialmente no caso de animais; e em desperdício de recursos financeiros destinados a material laboratorial de consumo. É fundamental encontrar o equilíbrio entre o custo e a representatividade da amostra, pois o tamanho da amostra afeta a precisão das estimativas das frequências alélicas e diversidade genética, parâmetros utilizadas nas diferentes medidas da estrutura e diferenciação populacional (LUNA; DEXING, 2004; KALINOWSKI, 2005; MIYAMOTO et al., 2008; PRUETT; WINKER, 2008; HALE; BURG; STEEVES, 2012; DUBOIS et al., 2018; ZCHONSKI, 2021).

Os estudos que delimitam o tamanho da amostral ideal, aqui também denominado N ideal, no sentido menor custo (coleta e material laboratorial) e maior benefício (precisão na representação da variabilidade), para locos de microssatélites são realizados por meio da comparação de dados empíricos (“população real”) com simulações, e por análises de rarefação e regressão (KALINOWSKI, 2005; BASHALKHANOV; PANDEY; RAJORA, 2009; HALE; BURG; STEEVES, 2012). Diferentes fatores demográficos e ecológicos das populações naturais podem afetar o tamanho efetivo de uma população, sua dispersão na área amostrada, os quais têm consequências na distribuição da variabilidade genética intra e interpopulacional e podem, portanto, afetar as estimativas do tamanho amostral ideal (KALINOWSKI, 2005; PRUETT; WINKER, 2008; DUBOIS et al., 2018). Além disso, o nível de polimorfismo e o número de locos de microssatélites aplicados nos estudos populacionais também influenciam na

determinação do tamanho amostral ideal (KALINOWSKI, 2005; HALE; BURG; STEEVES, 2012; ARTHOFER et al., 2018; WANG et al., 2021).

As taxas de evolução das sequências microssatélites são significativamente mais altas que das substituições de bases, acarretando no alto polimorfismo destes locos e originando alelos raros (com frequência inferior a 5%) por mutação recorrente em diferentes populações (EANES; KOEHN, 1977; LEVINSON; GUTMAN, 1987b; DALLAS, 1992; STRAND et al., 1993; WEBER; WONG, 1993; HALE; BURG; STEEVES, 2012). As estimativas do N ideal para a detecção desses alelos são de centenas de indivíduos (KALINOWSKI, 2005; MIYAMOTO et al., 2008; ARTHOFER et al., 2018), contudo, apesar desses alelos serem importantes para medidas de conservação ou inferências de fluxo gênico, as frequências dos alelos não raros, das quais são baseados os parâmetros de diversidade genética, distância genética e índices de diferenciação, são mais representativas e importantes para a determinação da estrutura das populações (SLATKIN, 1985; HALE; BURG; STEEVES, 2012).

Em animais, a determinação do tamanho ideal para locos de microssatélites foi realizada para espécies de aves, pardal (PRUETT; WINKER, 2008), albatroz e kakī (HALE; BURG; STEEVES, 2012); mamífero, esquilo (HALE; BURG; STEEVES, 2012); e insetos, gafanhoto (LUNA; DEXING, 2004), formiga (HALE; BURG; STEEVES, 2012), e borboleta (DUBOIS et al., 2018). Arthofer e colaboradores (2018) determinaram o número de locos de microssatélites adequados para análises da estrutura de populações de *Drosophila nigrosarsa*, mas não o N ideal. Desta maneira, existe uma lacuna deste tipo de dado para espécies do gênero *Drosophila* da família Drosophilidae (Diptera, Insecta), um grupo muito diverso que compreende dois terços, do total de espécies presente nesta família (TIDON, 2006; GOTTSCHALK; HOFMANN; VALENTE, 2008; BÄCHLI, 2022).

As espécies endêmicas de *Drosophila* da região Neotropical apresentam sensibilidade às alterações ambientais, diferentes sistemas de acasalamento, podem ser generalistas ou apresentarem especificidade de substrato para alimentação e reprodução, e ocupam uma variedade de habitats distribuídos em biomas distintos, sendo, portanto um organismo modelo muito importante para a avaliação da qualidade das áreas naturais e das mudanças climáticas em relação tanto à composição da fauna quanto à distribuição da variabilidade genética, assim como para determinar os padrões biogeográficos e eventos evolutivos responsáveis pela dispersão e especiação nesta região (PEREIRA; VILELA; SENE, 1983; MARKOW, 1996; MATEUS; SENE, 2003, KLACZKO, 2006;

MANFRIN; SENE, 2006; TIDON, 2006; da MATA; MCGEOCH; TIDON, 2008, 2010; GOTTSCHALK et al., 2009; BATISTA; ANANINA; KLACZKO, 2012; GOÑI et al., 2012; CENZI DE RÉ et al., 2014; GUSTANI et al., 2015; SANTA-BRÍGIDA et al., 2019; VALADÃO et al., 2019). As populações de algumas espécies de *Drosophila* da região Neotropical foram avaliadas quanto à variabilidade de locos de microssatélites, e os genótipos provenientes destes estudos foram utilizados neste trabalho para determinar o tamanho amostral ideal (MACHADO et al., 2010.; ZORZATO, 2015; PAIXÃO, 2020; TRAVA et al., 2021).

2.2 *Espécies de Drosophila utilizadas como modelos na determinação do tamanho amostral ideal para estudos populacionais com locos de microssatélites*

O gênero *Drosophila* é constituído por mais de 1.100 espécies organizadas em 14 subgêneros, entre estes, os subgêneros *Sophophora* e *Drosophila* se destacam pela alta diversidade, com cerca de 336 e 734 espécies, respectivamente (GOTTSCHALK; HOFMANN; VALENTE, 2008; O'GRADY; DESALLE, 2018; BÄCHLI, 2022), e ainda existem muitas espécies a serem descritas na região Neotropical (MEDEIROS; KLACZKO, 2004). Estes subgêneros divergiram há aproximadamente 63 milhões de anos, e a alta diversidade deles deve-se a eventos de especiação resultantes de isolamento provocado pelos períodos de resfriamento do planeta, que foram também, em parte, responsáveis pela distribuição de espécies endêmicas e cosmopolitas (POWELL, 1997; TIDON-SKLORZ; SENE, 1999; TAMURA; NEI; KUMAR, 2004).

A análise da distribuição da variabilidade genética de espécies de *Drosophila* endêmicas da região Neotropical tem revelado o efeito de alterações ambientais históricas e em decorrência da fragmentação por ação antrópica na expansão/retração de suas populações e no processo de especiação nesta região com alta diversidade de espécies, que podem ser coincidentes para outros grupos animais que ocupam os mesmos habitats (KLACZKO, 2006; MANFRIN; SENE, 2006; MATEUS; SENE, 2007; MATEUS et al., 2010; CENZI DE RÉ et al., 2014; GUSTANI et al., 2015; SILVA; MACHADO; MATEUS, 2015; BARRIOS-LEAL et al., 2021; TRAVA et al., 2021; ZORZATO; YASSIN; MADI-RAVAZZI, 2022). Estes estudos envolvem tanto espécies que ocupam as matas, como espécies associadas à vegetação xerofítica, com populações do litoral ao interior do continente, da região tropical e subtropical, caracterizando o padrão de

distribuição da variação em uma amplitude de biomas e condições abióticas da região Neotropical.

Duas das espécies utilizadas neste trabalho pertencem ao grupo *saltans* do subgênero *Sophophora*, *D. prosaltans* (subgrupo *saltans*); e *D. sturtevanti* (subgrupo *sturtevanti*); e duas ao subgênero *Drosophila*, *D. antonietae*, do grupo *repleta*, e *D. ornatifrons*, do grupo *guarani*. Além das relações filogenéticas, essas espécies também divergem com relação à especificidade por substratos para alimentação e reprodução, *D. antonietae* é cactófila, enquanto as outras três são generalistas; e ao tipo de habitat, áreas de mata para *D. prosaltans*, *D. sturtevanti* e *D. ornatifrons*, clima ameno e alta umidade para *D. ornatifrons*, e áreas abertas com vegetação xerofítica para *D. antonietae*. Ainda, os locos utilizados nas análises de *D. ornatifrons* foram transferidos de *D. mediopunctata* (grupo *tripunctata*) (LABORDA; KLACZKO; SOUZA, 2009; TRACTZ et al., 2012) e das demais espécies, os *primers* utilizados para a amplificação dos locos são espécie específicos (MACHADO et al., 2003; PAIXÃO, 2020; ROMAN; TRAVA; MADIRAVAZZI, 2020). Essas divergências podem ter consequências na detecção da estrutura das populações que afetam a distribuição da variabilidade genética e, por conseguinte, podem ter influência na determinação do N ideal para estudos genéticos evolutivos com locos de microsatélites.

As espécies do subgênero *Sophophora* amostradas neste trabalho, *Drosophila prosaltans* e *D. sturtevanti*, apresentam ampla distribuição geográfica desde o México e Caribe, para *D. sturtevanti*, e desde a Costa Rica, para *D. prosaltans*, até o Sul do Brasil (principalmente na Mata Atlântica, mas também no Cerrado) e Paraguai (MAGALHÃES, 1962). Paixão (2020) e Trava e colaboradores (2021) analisaram mais de 10 locos polimórficos de microsatélites em oito populações de *D. prosaltans* e nove de *D. sturtevanti*, de indivíduos coletados do Sul ao Nordeste do Brasil, e encontraram resultados semelhantes: diferenciação moderada, e distribuição das frequências gênicas sem correlação com a localização geográfica. *Drosophila ornatifrons*, do subgênero *Drosophila*, também tem ampla distribuição na América do Sul, porém, apesar de ter sido capturada no cerrado, é encontrada em maior frequências nas áreas de mata do Sul do Brasil, as quais apresentam temperatura amena e alta umidade relativa (ver GUSTANI et al., 2015 para uma revisão das áreas em que esta espécie já foi coletada). Zorzato (2015), analisando 10 locos polimórficos de microsatélites em cinco populações da região Sul do Brasil, observaram o mesmo padrão das espécies do grupo *saltans* do subgênero *Sophophora*.

A outra espécie do subgênero *Drosophila* analisada aqui, *D. antonietae*, apresenta distribuição mais restrita em relação às demais, pois ocorre nas áreas abertas com vegetação xerofítica das regiões Sul e Sudeste do Brasil, até a costa leste do Chaco argentino (TIDON-SKLORZ; SENE, 2001; MANFRIN, 2006). Enquanto a diferenciação genética das populações de *D. ornatifrons*, *D. prosaltans* e *D. sturtevanti* é moderada, *D. antonietae* mostrou baixa diferenciação genética entre populações geograficamente isoladas (MACHADO et al., 2010). Outra diferença de *D. antonietae* em relação as outras espécies está na sua especificidade de alimentação e oviposição em tecidos de cactos em decomposição, sendo, portanto, uma espécie cactófila, enquanto as outras três são consideradas generalistas (TIDON-SKLORZ; SENE, 1995; TIDON, 2006; VALADÃO et al., 2019). Contudo, apesar de *D. ornatifrons* ser considerada generalista, ela parece ter preferência por fontes de recursos mais ricas em proteínas, pois além de poder ser coletada em carcaças animais (GOÑI et al., 2012), apresenta baixa viabilidade em meio de cultura padrão de banana (SANTOS et al. submetido), e em meio de cultura com menor concentração de proteína em relação a açúcar (MATZKIN et al., 2011) é totalmente inviável (SANTOS, 2020).

Considerando a eficiência dos marcadores microssatélites para a Genética de Populações, os vários fatores que influenciam a determinação do tamanho ideal para análises genéticas com esses marcadores, e a importância da qualidade e do desenho amostral para as inferências estatísticas das análises populacionais, torna-se importante definir o N ideal para o maior número de táxons e diferentes espécies de um mesmo táxon, que apresentem distintas especificidades ecológicas, sistemas de reprodução e habitats, para que possam ser confiavelmente generalizadas para grupos com configurações biológicas e espaciais semelhantes.

3. OBJETIVOS

3.1 Objetivo geral

Determinar o tamanho amostral ideal, definido pelo menor tamanho amostral representativo da variação natural, para análise de locos de microssatélites em espécies de *Drosophila* do Neotrópico.

3.2 *Objetivos específicos*

- Verificar a porcentagem das subamostras de cada N com todos alelos representativos (frequência acima de 5%) da matriz original para no mínimo 70 % dos locos de microsatélite analisados;
- Avaliar as diferenças estatísticas na heterozigosidade média esperada (H_e) das subamostras entre os diferentes N simulados;
- Avaliar as diferenças estatísticas da média do índice de diferenciação genética (F_{st}) das subamostras de cada N com a matriz original;
- Determinar o tamanho amostral ideal com base nos parâmetros acima para as quatro espécies;
- Comparar a distribuição da composição genética das subamostras do N ideal com e sem alelos raros;
- Comparar as estimativas do N ideal entre as espécies e inferir hipóteses para as semelhanças e/ou diferenças entre elas.

4. MATERIAL E MÉTODOS

4.1. Determinação do tamanho amostral ideal

Uma matriz com 100 genótipos aleatórios de locos autossômicos e polimórficos de microsatélites provenientes de populações naturais de *Drosophila antonietae*, *D. ornatifrons*, *D. prosaltans* e *D. sturtevantii* foi gerada para cada uma das espécies. Os locos aplicados e as populações analisadas para cada espécie, a metodologia para extração de DNA e as condições de amplificação estão descritas em Machado e colaboradores (2010) para *D. antonietae*, Zorzato (2015) para *D. ornatifrons*, Paixão (2020) para *D. prosaltans* e em Trava e colaboradores (2021) para *D. sturtevantii*. A metodologia para determinação e separação dos alelos em gel de poliacrilamida 6%, corado com nitrato de prata 15%, seguiu o protocolo descrito em Machado e colaboradores (2003) para todas as espécies. Apenas os locos com amplificações positivas para mais de 100 indivíduos foram utilizados na matriz original, desta maneira, foram amostrados nove locos para *D. ornatifrons*, e sete locos para *D. prosaltans*, *D. sturtevantii* e *D. antonietae* (Tabela 1). Os locos aplicados em *D. ornatifrons* e *D. prosaltans* foram transferidos de *D. mediopunctata* (LABORDA; KLACZKO; SOUZA, 2009; LABORDA; MORI; SOUZA, 2009; TRACTZ et al., 2012) e *D. sturtevantii* (ROMAN; TRAVA; MADI-RVAZZI, 2020), respectivamente. Os locos aplicados em *D. antonietae* e *D. sturtevantii* foram

descritos para estas espécies (MACHADO et al. 2003 e ROMAN; TRAVA; MADIRVAZZI, 2020, respectivamente).

Os critérios para determinação do tamanho amostral ideal (N ideal) adotados neste trabalho seguiram os parâmetros discutidos no artigo de Hale, Burg e Steeves (2012) de determinar o número amostral mínimo capaz de fornecer estimativas precisas das frequências alélicas representativas (frequência $\geq 5\%$) por loco e, como consequência, da heterozigosidade esperada (H_e), de uma população. Segundo as autoras, no geral, a detecção da frequência de alelos raros ($< 5\%$) não aumenta substancialmente com o incremento do tamanho amostral, além das frequências alélicas serem superestimadas quando os alelos raros são amostrados. Assim, neste trabalho, o N ideal foi selecionado entre três parâmetros: o tamanho amostral a partir do qual ao menos 95% de 50 sorteios de N (variando entre 5 e 95) detecte todos os alelos não raros em ao menos 70% dos locos; o menor tamanho amostral que não revele diferenças significativas nas análises par-a-par da variância de H_e com o tamanho amostral superior subsequente; o menor tamanho amostral que não revele diferenças significativas nas estimativas médias de F_{st} em relação a matriz original. A seguir, a descrição das análises que permitiram avaliar esses três critérios.

A partir da matriz original, foi realizado o sorteio, por meio do comando aleatório do Excel (Microsoft®), para 50 simulações (chamadas daqui em diante de subamostras), de diferentes tamanhos amostrais: 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 e 95. A frequência dos alelos por loco foi determinada na matriz original, utilizando o programa GenAlEx (PEAKALL; SMOUSE, 2012), para a detecção dos alelos raros. A partir desse resultado, foram construídos gráficos relacionando a frequência da presença de todos os alelos representativos nas subamostras dos diferentes N.

As heterozigosidades médias esperadas e o índice de diferenciação genética F_{st} das subamostras com a matriz original foram estimados com o auxílio do programa GenAlEx (PEAKALL; SMOUSE, 2012). A significância da variância entre as médias de H_e e F_{st} das subamostras foi avaliada pela comparação par-a-par entre os diferentes tamanhos amostrais.

Tabela 1. Locos de microssatélites e os respectivos número de alelos (N_a) provenientes de dados empíricos de espécies de *Drosophila*. *D. ornatifrons* – Zorzato (2015); *D. prosaltans* – Paixão (2020); *D. sturtevantii* – Trava et al. (2021); *D. antonietae* – Machado et al. (2010).

Espécies	Locos	Na	Espécies	Locos	Na	
<i>Drosophila ornatifrons</i>	34	12	<i>Drosophila sturtevantii</i>	e	10	
	53	2		g	12	
	57	5		j	11	
	79	2		k	8	
	87	8		l	12	
	96	6		m	9	
	107	7		o	13	
	118	6		<i>Drosophila antonietae</i>	1	6
	126	5			2	4
<i>Drosophila prosaltans</i>	A	16	3		6	
	B	15	4	4		
	L	14	5	4		
	D	14	6	6		
	F	15	7	6		
	G	13				
	J	13				

4.2. Influência dos alelos raros na distribuição da composição genética das subamostras do N ideal

A influência dos alelos raros na distribuição da composição genética do N ideal foi avaliada com a geração de uma nova matriz com 100 genótipos, na qual houve substituição dos genótipos com alelos raros por genótipos de dados empíricos sem estes alelos. A partir desta matriz sem alelos raros, foram sorteadas 50 subamostras do N ideal para cada espécie de *Drosophila*. A distribuição da composição genética das subamostras com alelos raros foi comparada com as das subamostras sem alelos raros pela análise Bayesiana com base no número de *clusters*, agrupamentos genéticos (K), utilizando o programa *Structure* (PRITCHARD; WEN; FALUSH, 2010).

Para esta análise assumiu-se a hipótese de mistura, com a existência de frequências correlacionadas, nas quais cada amostra é composta parcialmente pelo genoma de cada população ancestral. O número de agrupamentos foi determinado como sendo entre 1 e 10, utilizando 10.000 *burnin*, 500.000 repetições de cadeias de Markov Monte Carlo após o *burnin*, e 25 iterações. Por meio do programa STRUCTURE HARVESTER (EARL; VON HOLDT, 2012), utilizando o teste de Evanno (EVANNO; REGNAUT; GOUDET, 2005; EARL; VON HOLDT, 2012), foi determinado o melhor número de agrupamentos (K). A ferramenta CLUMPAK (KOPELMAN et al., 2015) foi utilizada para obtenção dos dados das frequências dos grupos genéticos para cada indivíduo. Posteriormente, cada grupo genético da análise com e sem alelos raros foi avaliado separadamente quanto à proporção de ocorrência das frequências individuais em três categorias: 10% inferior, 80% intermediário e 10% superior. A partir da amplitude

da distribuição das frequências individuais dentro de cada grupo genético, foram estimados os valores limites determinantes das três categorias. No Excel (Microsoft®) foram elaborados os gráficos da distribuição das frequências desses agrupamentos genéticos.

4.3. Análises estatísticas

O teste Shapiro Wilk da homocedasticidade dos dados de H_e e F_{st} revelou dados não paramétricos. Assim, as significâncias das comparações par-a-par da mediana de H_e entre os diferentes N, e da mediana de F_{st} nos N simulados com a matriz original, com os dados do arco seno da raiz quadrada, foram obtidas pela análise de Mann-Whitney e correção de Bonferroni no programa PAST (HAMMER; HARPER; RYAN, 2001).

5. RESULTADOS

Os resultados deste trabalho serão apresentados na forma de artigo redigido seguindo as normas do periódico “Journal of Zoological Systematics and Evolutionary Research”.

1 **Journal of Zoological Systematics and Evolutionary Research**

2 **Influência da Diversidade Genética sobre o Tamanho**

3 **Amostragem de Espécies Neotropicais de *Drosophila***

4 **Resumo**

5 O tamanho amostral (N) é um importante aspecto nos estudos da estrutura das
6 populações naturais, pois ao mesmo tempo que deve ser representativo da variação
7 natural, não deve desperdiçar recursos com número de indivíduos que não promovam
8 alteração na qualidade dos dados populacionais. O objetivo deste trabalho foi
9 determinar o N que otimiza recursos, mas mantém a representatividade da variação de
10 locos de microssatélites de espécies de *Drosophila* da região Neotropical: *D.*
11 *ornatifrons*, *D. antonietae*, subgênero *Drosophila*, e *D. prosaltans* e *D. sturtevanti*,
12 subgênero *Sophophora*. O N ideal foi determinado por meio de 50 simulações
13 (subamostras) de 12 tamanhos amostrais diferentes (variando de 5 a 95 indivíduos), com
14 a análise estatística de três parâmetros: 1) a detecção de no mínimo 95% dos alelos
15 representativos (frequência $\geq 5\%$) nas 50 subamostras em ao menos 70% dos locos; 2)
16 heterozigosidade média esperada (H_e); e o índice de diferenciação genética (F_{st}). O
17 efeito de alelos raros na distribuição dos indivíduos nos grupos genéticos foi analisado
18 pela comparação de 50 subamostras do N ideal, a partir de matrizes com e sem estes
19 alelos. O N ideal determinado foi semelhante para as espécies do subgênero *Drosophila*,
20 $N = 25$ para *D. ornatifrons* e $N = 30$ para *D. antonietae*, e o mesmo para as espécies do
21 subgênero *Sophophora*, $N = 35$. Não foi observado efeito acentuado dos alelos raros na
22 distribuição dos indivíduos nos grupos genéticos. Neste trabalho é discutida a influência
23 da diversidade genética na determinação do N ideal, a qual, por sua vez, pode ter
24 relação com os nichos ecológicos explorados por estas espécies. Os resultados
25 encontrados aqui podem ser aplicados para outros grupos animais que ocupam os
26 mesmos tipos de habitats, e ainda revelou que o mesmo N pode ser empregado em caso
27 de locos de microssatélites transferidos entre espécies próximas filogeneticamente.

28 **Keywords**

29 diferenciação genética, heterozigosidade, microssatélites, subgênero *Drosophila*,
30 subgênero *Sophophora*, tamanho amostral

31 **Introdução**

32 O estudo da variabilidade genética das populações naturais é fundamental para a
33 compreensão dos processos contemporâneos, históricos e demográficos responsáveis
34 pelo padrão observado, além de ser uma medida da capacidade de reposta das
35 populações frente à mudança ambiental, ou ainda, para fins de conservação das
36 espécies. Marcadores moleculares neutros e polimórficos, como os microssatélites, são
37 os mais adequados para estas análises, visto que a suposta ausência de seleção natural
38 não limitaria a variação, ou causaria desvio da neutralidade (Hedrick, 1986; Estoup et
39 al, 1995; Frankham, 1995; Wan et al., 2004; Machado et al. 2010; Trava et al., 2021).
40 Como essas análises populacionais por meio de locos de microssatélites envolvem
41 esforços e recursos substanciais para a coleta de indivíduos da natureza, e posterior

42 amplificação e análise desses marcadores, torna-se importante a otimização de um
43 tamanho amostral adequado que ainda seja representativo da variação natural.

44 Espécies de *Drosophila* da região Neotropical dos subgêneros *Drosophila* e
45 *Sophophora* tiveram a diversidade genética de algumas de suas populações amostradas
46 por meio de locos de microssatélites. As espécies do subgênero *Sophophora* apresentam
47 uma tendência de se alimentarem de frutos em decomposição, enquanto as espécies do
48 subgênero *Drosophila* se alimentam das leveduras que se desenvolvem não apenas em
49 frutos, mas também em outras partes das plantas, de fungos e podem se alimentar até
50 mesmo de microrganismos presentes em guano e carcaças de outros animais (Heed
51 & Wheele, 1957; Fellows & Heed, 1972; Ruiz & Heed, 1988; Tosi et al., 1990; Matzkin
52 et al., 2009; Goñi et al., 2012; Valadão et al., 2019). O grupo *repleta* do gênero
53 *Drosophila* possui representantes tanto de uma dieta generalista, como espécimes com
54 especificidade ecológica para alimentação e sítio de reprodução em tecidos em
55 decomposição de cactáceas (Pereira Vilela, & Sene, 1983). Originalmente, os espécimes
56 cactófilos do grupo *repleta* usavam principalmente as frutas como hospedeiros para
57 alimentação e reprodução e depois mudaram para cactáceas (Oliveira et al., 2012).

58 A utilização de recursos para alimentação e reprodução, entre outros fatores
59 demográficos, podem afetar as estimativas do tamanho otimizado para análises
60 populacionais, no sentido de que a heterogeneidade de recursos alimentar e de
61 reprodução das espécies pode manter o polimorfismo genético, enquanto que a
62 especificidade de nicho resulta na limitação da dispersão pela restrição e escassez de
63 recursos, e, conseqüentemente, menor diversidade genética (Brown & Kodric-Brown,
64 1977; Templeton et al. 1990, 2001; Frankham, 1995; Hartl & Clark, 1997; Mateus &
65 Sene, 2003; Dubois et al., 2018). Desta forma, o objetivo deste trabalho foi determinar o
66 menor tamanho amostral representativo da variação natural, N ideal, para análise de
67 locos de microssatélites em espécies de *Drosophila* do Neotrópico com diferentes
68 preferências ecológicas: *D. ornatifrons* e *D. antonietae*, representantes generalista e
69 cactófila, respectivamente, do subgênero *Drosophila*, e *D. prosaltans* e *D. sturtevantii*,
70 espécies próximas filogeneticamente e generalistas do subgênero *Sophophora*.

71 Material e Métodos

72 Determinação do tamanho amostral ideal

73 A partir de uma matriz original de 100 genótipos de locos de microssatélites
74 polimórficos, provenientes de populações naturais, foram sorteadas 50 subamostras para
75 cada um de 12 tamanhos populacionais (N): 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 e
76 95, com auxílio do comando aleatório do Excel (Microsoft®). Essas simulações foram
77 realizadas para determinar o N ideal, aqui definido como o menor tamanho amostral
78 sem diferença estatística significativa do N superior subsequente em relação a parâmetros
79 da estrutura populacional, das espécies: *Drosophila ornatifrons* (Zorzato, 2015) e *D.*
80 *antonietae* (Machado et al., 2010), pertencentes ao subgênero *Drosophila*, e *D.*
81 *prosaltans* (Paixão, 2020) e *D. sturtevantii* (Trava et al., 2021), do subgênero
82 *Sophophora*. Os locos de microssatélites aplicados nas espécies *D. antonietae*
83 (Machado et al., 2003) e *D. sturtevantii* (Roman, Trava, & Madi-Ravazzi, 2020) são

84 espécie-específicos, enquanto os locos de *D. ornatifrons* e *D. prosaltans* foram
85 transferidos de *D. mediopunctata* (Laborda, Klaczko, & Souza, 2009; Laborda, Mori, &
86 Souza, 2009; Tractz et al., 2012) e *D. sturtevantii* (Roman, Trava, & Madi-Ravazzi,
87 2020), respectivamente.

88 Os critérios para determinação do N ideal foram estabelecidos considerando três
89 parâmetros: o tamanho amostral a partir do qual ao menos 95% das 50 subamostras de
90 cada N detecte todos os alelos representativos (frequência $\geq 5\%$) em ao menos 70% dos
91 locos; o menor tamanho amostral que não revele diferenças significativas nas análises
92 par-a-par da variância de H_e com o tamanho amostral superior subsequente; o menor
93 tamanho amostral que não revele diferenças significativas nas estimativas médias de F_{st}
94 das subamostras em relação a matriz original. As heterozigosidades médias esperadas e
95 o índice de diferenciação genética F_{st} das subamostras com a matriz original foram
96 estimados com o auxílio do programa GenAIEx (Peakall & Smouse, 2012). O teste
97 Shapiro Wilk da homocedasticidade dos dados de H_e e F_{st} revelou dados não
98 paramétricos. Assim, as significâncias das comparações par-a-par da mediana de H_e
99 entre os diferentes N, e da mediana de F_{st} das subamostras nos N simulados com a
100 matriz original, com os dados do arco-seno da raiz quadrada, foram obtidas pela análise
101 de Mann-Whitney e correção de Bonferroni no programa PAST (Hammer, Harper, &
102 Ryan, 2001).

103 Influência dos alelos raros na distribuição da composição genética de subamostras do N
104 ideal

105 O número de agrupamentos genéticos, assumindo hipótese de mistura, foi determinado
106 como sendo entre 1 e 10, utilizando 10.000 *burnin*, 500.000 repetições de cadeias de
107 Markow Monte Carlo após o *burnin*, e 25 iterações. Por meio do programa
108 STRUCTURE HARVESTER (Earl & Von Holdt, 2012), utilizando o teste de Evanno
109 (Evanno, Regnaut, & Goudet, 2005; Earl & Von Holdt, 2012), foi determinado o
110 melhor número de agrupamentos (K) em 50 subamostras do N ideal determinado para
111 cada espécie, a partir de uma matriz de 100 genótipos de microssatélites contendo alelos
112 raros (frequência $< 5\%$) e outra matriz de 100 genótipos sem alelos raros. A ferramenta
113 CLUMPAK (Kopelman et al., 2015) foi utilizada para obtenção dos dados das
114 frequências dos grupos genéticos para cada indivíduo. Posteriormente, cada grupo
115 genético da análise com e sem alelos raros foi avaliado separadamente quanto à
116 proporção de ocorrência das frequências individuais em três categorias: 10% inferior,
117 80% intermediário e 10% superior. A partir da amplitude da distribuição das
118 frequências individuais dentro de cada grupo genético, foram estimados os valores
119 limites determinantes das três categorias. No Excel (Microsoft®) foram elaborados os
120 gráficos da distribuição das frequências desses agrupamentos genéticos.

121 Resultados

122 Determinação do tamanho amostral ideal

123 As frequências alélicas dos locos de microssatélites da matriz com 100 genótipos das
124 quatro espécies de *Drosophila*, a partir das quais foram geradas as 50 subamostras para
125 os diferentes N, são apresentadas na Tabela 1 MS (Material Suplementar).

126 O tamanho amostral para a detecção de todos os alelos representativos dos locos
127 de microssatélites de *D. ornatifrons* em ao menos 95% das subamostras para no mínimo
128 70% dos locos foi $N = 25$ (Figura 1A). As médias de H_e variaram entre 0,5468 para $N =$
129 5 e 0,6345 para $N = 95$ (Figura 2A). A análise par-a-par da H_e das subamostras revelou
130 valores significativos em todas as comparações com $N = 5$ ($p < 0,01$); $N = 10$ com $N \geq 20$
131 ($p < 0,01$); $N = 15$ com $N \geq 50$ ($p < 0,01$); $N = 25$ com $N \geq 50$ ($p < 0,05$); e $N = 30$ com $N =$
132 95 ($p < 0,01$). O tamanho amostral $N = 10$ foi o menor N que não apresentou diferenças
133 significativas nas comparações par-a-par para a mediana de H_e com os tamanhos
134 amostrais superiores subsequentes. A média do índice de diferenciação genética F_{st} das
135 subamostras variou de próximo de 0 para $N = 95$ até 0,035 para $N = 5$ (Figura 3A). A
136 análise par-a-par das medianas do F_{st} não revelou valores significativos apenas nas
137 comparações entre $N = 20$ e $N = 25$, $N = 30$ e $N = 35$, $N = 35$ e $N = 40$, $N = 40$ e $N =$
138 45, e $N = 45$ e $N = 50$. Portanto, o menor tamanho amostral que não apresentou
139 diferença estatística do F_{st} em relação a matriz original nas comparações par-a-par foi N
140 $= 20$ com $N = 25$. Avaliando os três parâmetros, $N = 25$ pode ser considerado o
141 tamanho amostral ideal para aplicação em estudos genéticos populacionais de *D.*
142 *ornatifrons*.

143 O tamanho amostral para a detecção de todos os alelos representativos dos locos
144 de microssatélites de *D. antonietae* em ao menos 95% das subamostras para no mínimo
145 70% dos locos foi $N = 25$ (Figura 1B). As médias de H_e variaram entre 0,4622 para $N =$
146 5 e 0,5361 para $N = 95$ (Figura 2B). A análise par-a-par da mediana da H_e das
147 subamostras revelou valores significativos nas comparações com $N = 5$ com $N \geq 15$
148 ($p < 0,01$); $N = 10$ com $N = 30$ e $N \geq 40$ ($p < 0,05$); $N = 15$ com $N = 30$ e $N \geq 40$ ($p < 0,05$); N
149 $= 25$ com $N = 95$ ($p < 0,05$); $N = 30$ com $N \geq 75$ ($p < 0,05$); e $N = 40$ com $N = 95$ ($p < 0,05$).
150 O tamanho amostral $N = 5$ foi o menor N que não apresentou diferenças significativas
151 nas comparações par-a-par para a mediana de H_e com os tamanhos amostrais superiores
152 subsequentes. A média do índice de diferenciação genética F_{st} das subamostras variou
153 de próximo de 0 para $N = 95$ até 0,038 para $N = 5$ (Figura 3B). A análise par-a-par das
154 medianas do F_{st} não revelou valores significativos apenas nas comparações entre $N =$
155 30 e $N = 35$, $N = 35$ e $N = 40$, e $N = 45$ e $N = 50$. Portanto, o menor tamanho amostral
156 que não apresentou diferença estatística do F_{st} em relação a matriz original nas
157 comparações par-a-par foi $N = 30$ com $N = 35$. Avaliando os três parâmetros, $N = 30$
158 pode ser considerado o tamanho amostral ideal para aplicação em estudos genéticos
159 populacionais de *D. antonietae*.

160 O tamanho amostral para a detecção de todos os alelos representativos dos locos
161 de microssatélites de *D. prosaltans* em ao menos 95% das subamostras para no mínimo
162 70% dos locos foi $N = 35$ (Figura 1C). As médias de H_e variaram entre 0,7609 para $N =$
163 5 e 0,8763 para $N = 95$ (Figura 2C). A análise par-a-par da mediana da H_e das
164 subamostras revelou valores significativos nas comparações de $N = 5$, $N = 10$ e $N = 15$
165 com todos os demais N ($p < 0,01$); $N = 20$ com $N \geq 30$ ($p < 0,001$); $N = 25$ com $N \geq 40$
166 ($p < 0,05$); $N = 30$, $N = 35$ e $N = 40$ com $N \geq 50$ ($p < 0,05$); e $N = 45$ e $N = 50$ com $N \geq 75$
167 ($p < 0,001$). O tamanho amostral $N = 20$ foi o menor N que não apresentou diferenças
168 significativas nas comparações par-a-par para a mediana de H_e com os tamanhos
169 amostrais superiores subsequentes. A média do índice de diferenciação genética F_{st} das
170 subamostras variou de próximo de 0 para $N = 95$ até 0,035 para $N = 5$ (Figura 3C). A

171 análise par-a-par das medianas do Fst não revelou valores significativos apenas na
172 comparação entre N = 30 e N = 35, portanto este foi o menor tamanho amostral que não
173 apresentou diferença estatística do Fst em relação a matriz original. Avaliando os três
174 parâmetros, N = 35 pode ser considerado o tamanho amostral ideal para aplicação em
175 estudos genéticos populacionais de *D. prosaltans*.

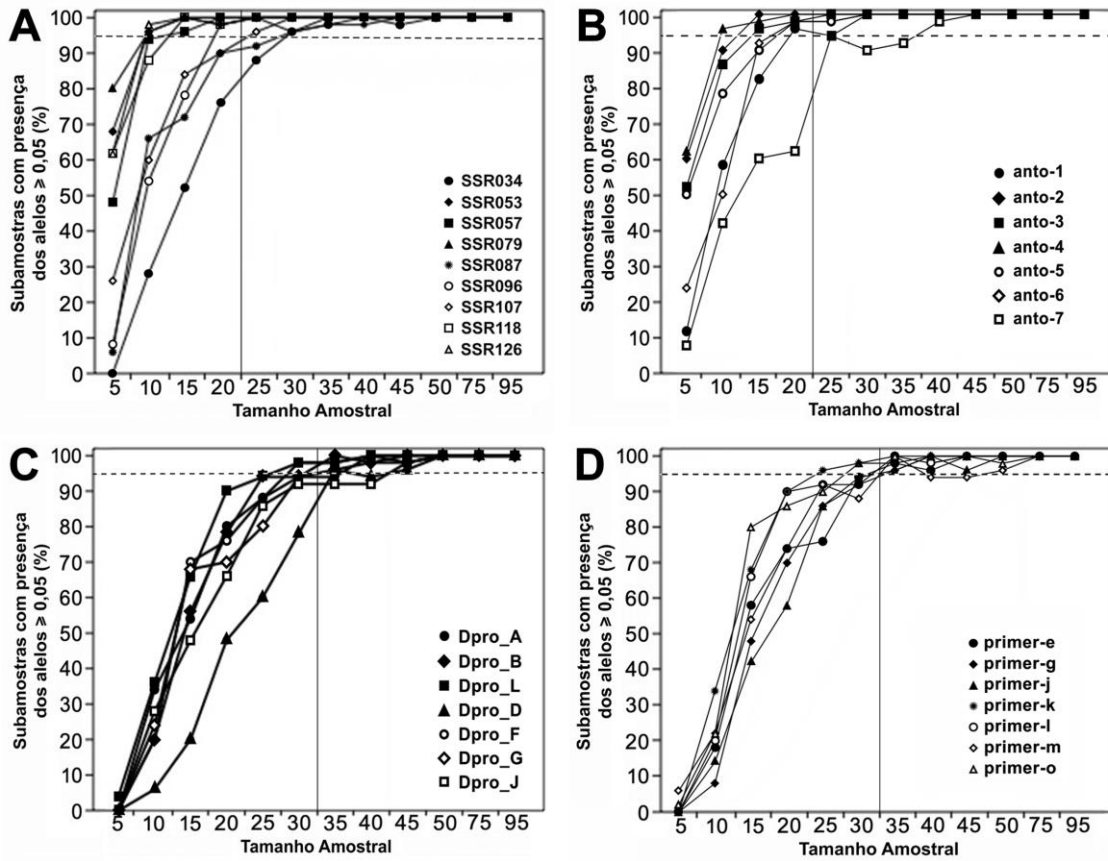
176 O tamanho amostral para a detecção de todos os alelos representativos dos locos
177 de microssatélites de *D. sturtevantii* em ao menos 95% das subamostras para no mínimo
178 70% dos locos foi N = 35 (Figura 1D). As médias de He variaram entre 0,7289 para N =
179 5 e 0,8488 para N = 95 (Figura 2D). A análise par-a-par da mediana da He das
180 subamostras não revelou valores significativos ($p > 0,05$) apenas nas comparações de N =
181 25 com N = 30; N = 30 com N = 35; N = 35 com N = 40 e N = 45; N = 40 com N = 45
182 e; N = 45 com N = 50. O tamanho amostral N = 25 foi o menor N que não apresentou
183 diferenças significativas nas comparações par-a-par para a mediana de He com os
184 tamanhos amostrais superiores subsequentes. A média do índice de diferenciação
185 genética Fst das subamostras variou de próximo de 0 para N = 95 até 0,038 para N = 5
186 (Figura 3D). A análise par-a-par das medianas do Fst não revelou valores significativos.
187 Avaliando os três parâmetros, N = 35 pode ser considerado o tamanho amostral ideal
188 para aplicação em estudos genéticos populacionais de *D. sturtevantii*.

189

190

191

192



193

194

195 Figura 1. Porcentagem das subamostras em que foram detectados todos os alelos com
 196 frequência $\geq 0,05$ para cada loco de microssatélite nos diferentes tamanhos amostrais. A
 197 linha vertical mostra o tamanho amostral a partir do qual todos os locos apresentam ao
 198 menos 95% (linha tracejada horizontal) das subamostras com os alelos representativos.

199 A. *Drosophila ornatifrons*, B. *D. antonietae*, C. *D. prosaltans*, D. *D. sturtevantii*.

200

201

202

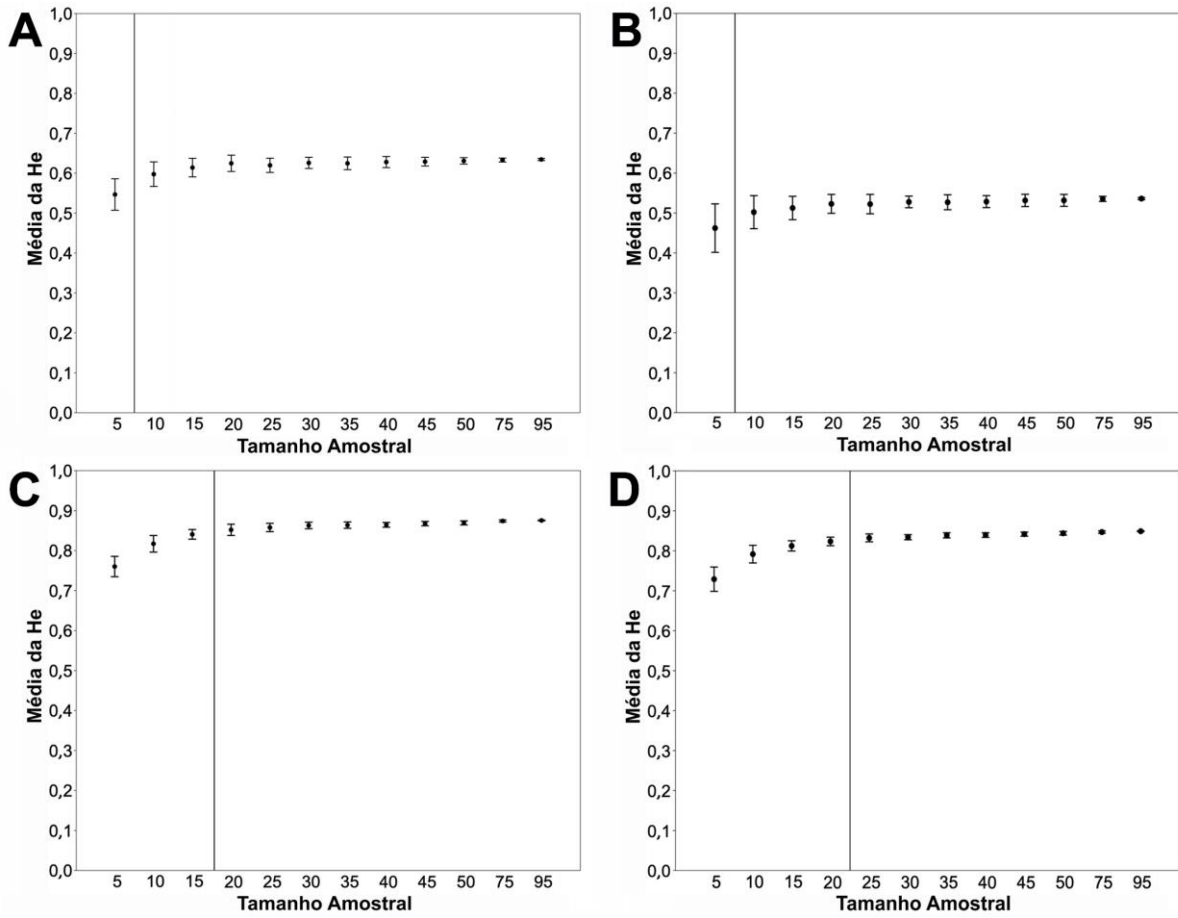
203

204

205

206

207



209

210

211 Figura 2. Média \pm desvio padrão da heterozigosidade média esperada (He) de 50
 212 subamostras para os diferentes tamanhos amostrais. A partir da linha vertical, não foram
 213 encontradas diferenças significativas entre pares de tamanhos amostrais
 214 subsequentemente superiores. A. *Drosophila ornatifrons*, B. *D. antonietae*, C. *D.*
 215 *prosaltans*, D. *D. sturtevantii*.

216

217

218

219

220

221

222

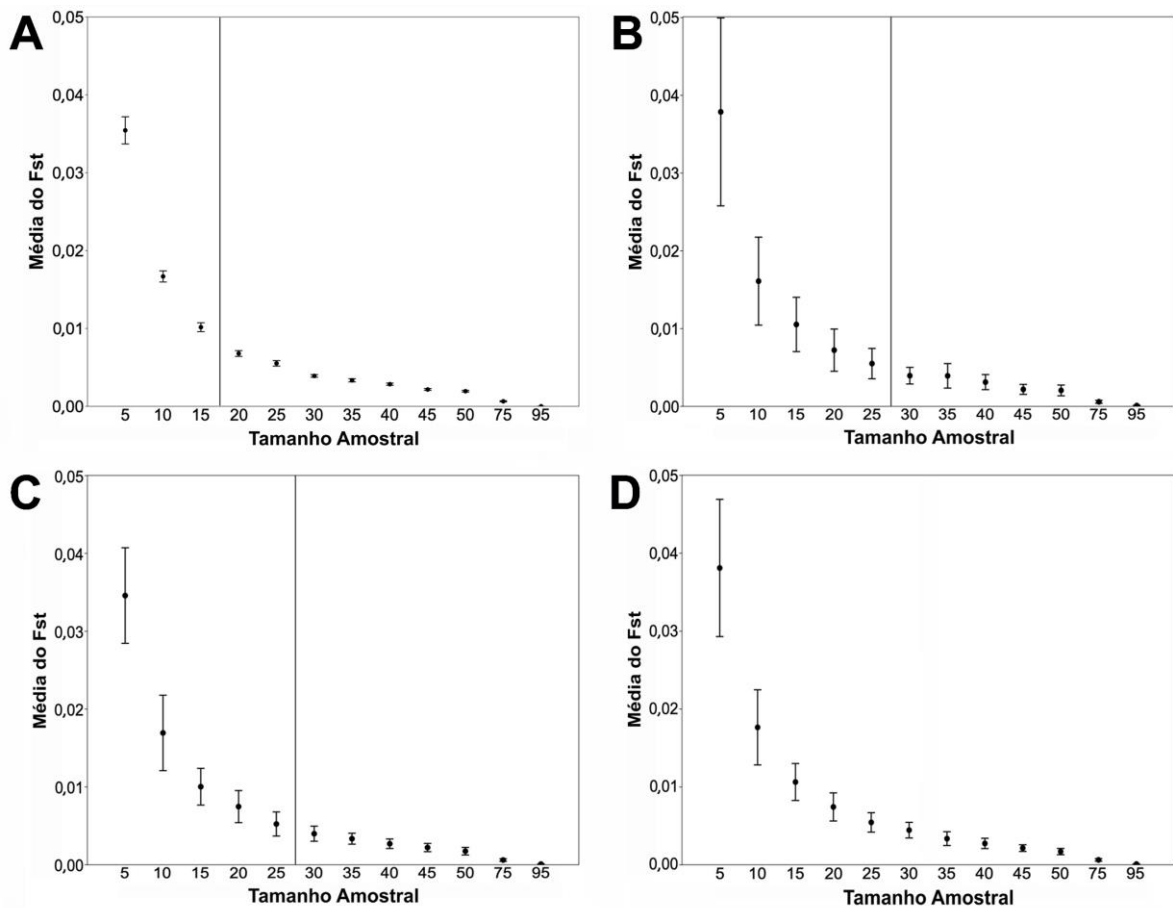
223

224

225

226

227



228

229

230

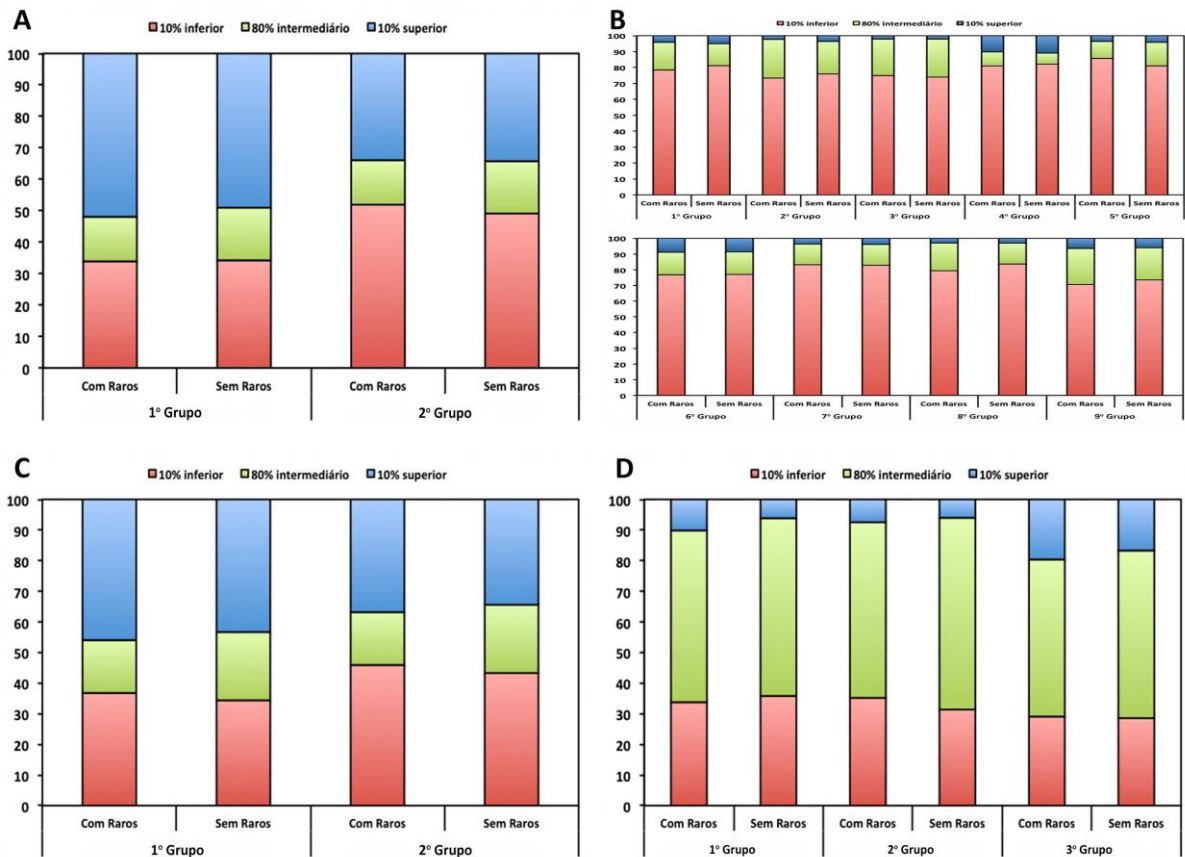
231

232 Figura 3. Média \pm desvio padrão do índice de diferenciação genética F_{st} entre 50
233 subamostras e a matriz original para os diferentes tamanhos amostrais. A partir da linha
234 vertical, não foram encontradas diferenças significativas entre pares de tamanhos
235 amostrais subsequentemente superiores. A. *Drosophila ornatifrons*, B. *D. antonietae*, C.
236 *D. prosaltans*, D. *D. sturtevanti*.

237 **Influência dos alelos raros na distribuição da composição genética de subamostras**
 238 **do N ideal**

239 A análise Bayesiana da estruturação populacional das subamostras com e sem alelos
 240 raros, realizada no programa STRUCTURE, mostrou a presença de dois grupos
 241 genéticos ($K = 2$) para *D. ornatifrons* e *D. prosaltans*, três ($K = 3$) para *D. sturtevantii* e
 242 nove ($K = 9$) para *D. antonietae*. As porcentagens de indivíduos que apresentaram
 243 proporções dentro de cada classe de agrupamento (10% inferior, 80% intermediário e
 244 10% superior) para cada grupo genético foram semelhantes com e sem a presença de
 245 alelos raros (Figura 4). Uma análise pormenorizada da distribuição das proporções de
 246 cada grupo em cada indivíduo, a partir dos dados com e sem alelos raros, mostrou que
 247 ocorreram diferenças pequenas e pouco pronunciadas nestas distribuições (Figuras 5 à
 248 8). Assim, pode ser concluído que a presença de alelos raros parece não afetar
 249 substancialmente os resultados produzidos pela análise Bayesiana.

250
 251



252
 253

254 Figura 4. Porcentagens de ocorrência em cada classe de agrupamento (10% inferior,
 255 80% intermediário e 10% superior - ver Material e Métodos) das proporções de cada
 256 grupo genético (K) nos indivíduos das subamostras com e sem alelos raros. A - *D.*
 257 *ornatifrons*: 1250 indivíduos (50 subamostras com 25 indivíduos) em cada simulação
 258 (com e sem alelos raros); B - *D. antonietae*: 1500 indivíduos (50 subamostras com 30
 259 indivíduos) em cada simulação (com e sem alelos raros); C - *D. prosaltans* e D - *D.*

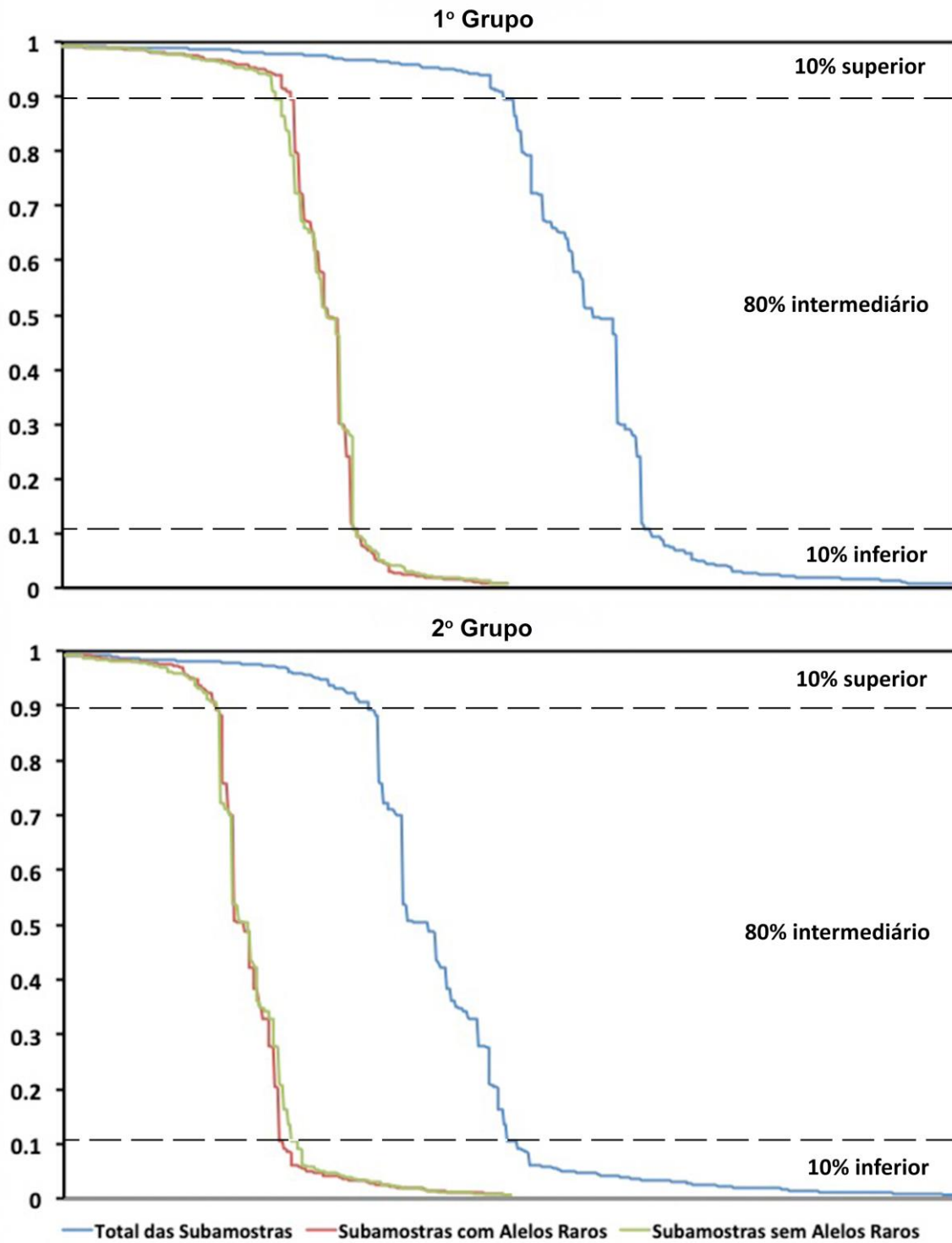
260 *sturtevantii*: 1750 indivíduos (50 subamostras com 35 indivíduos) em cada simulação
261 (com e sem alelos raros).

262

263

264

265



266

267

268

269

270

271

272 Figura 5. Distribuição das proporções dos grupos genéticos ($K = 2$) entre o total de
273 indivíduos de *D. ornatifrons* amostrados (2500), entre os com alelos raros (1250
274 indivíduos) e entre os sem alelos raros (1250 indivíduos). As linhas pontilhadas
275 delimitam as classes de agrupamentos (10% inferior, 80% intermediário e 10% superior
276 - ver Material e Métodos).

277

278

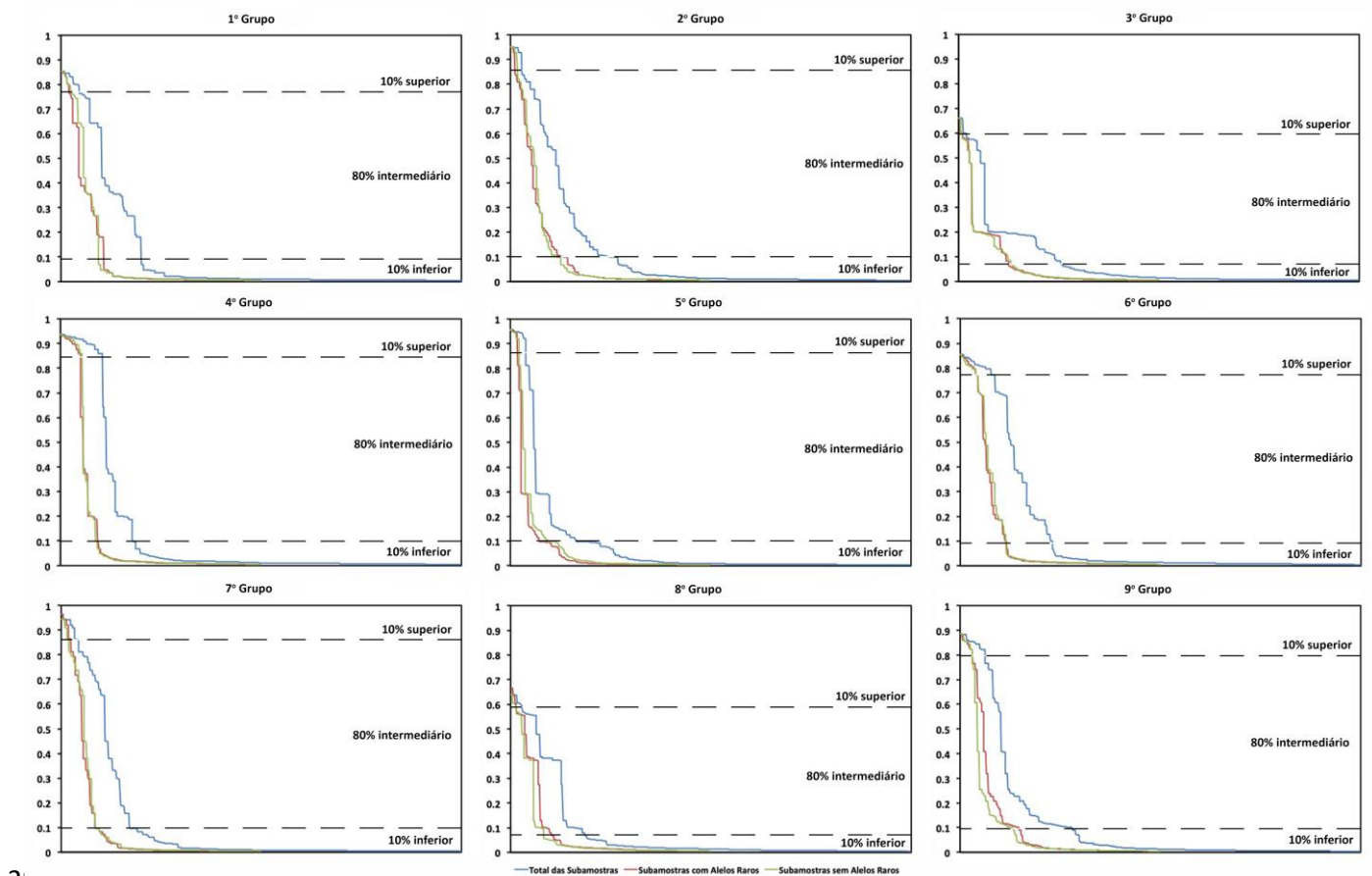
279

280

281

282

283

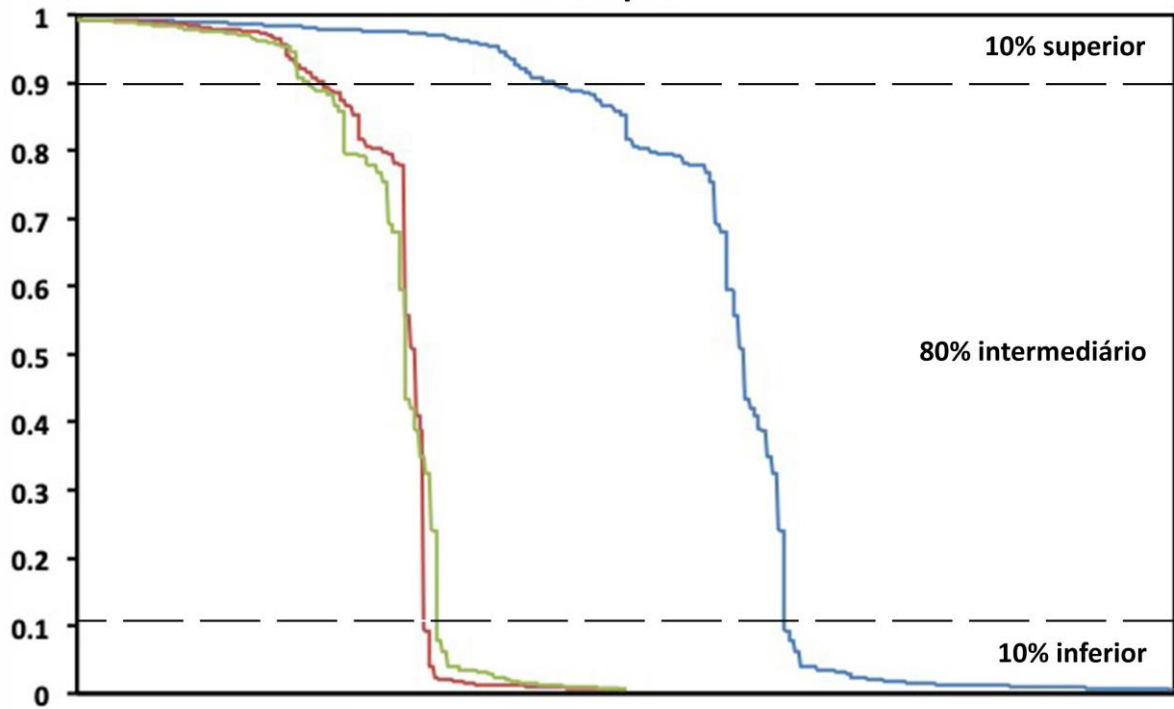


284

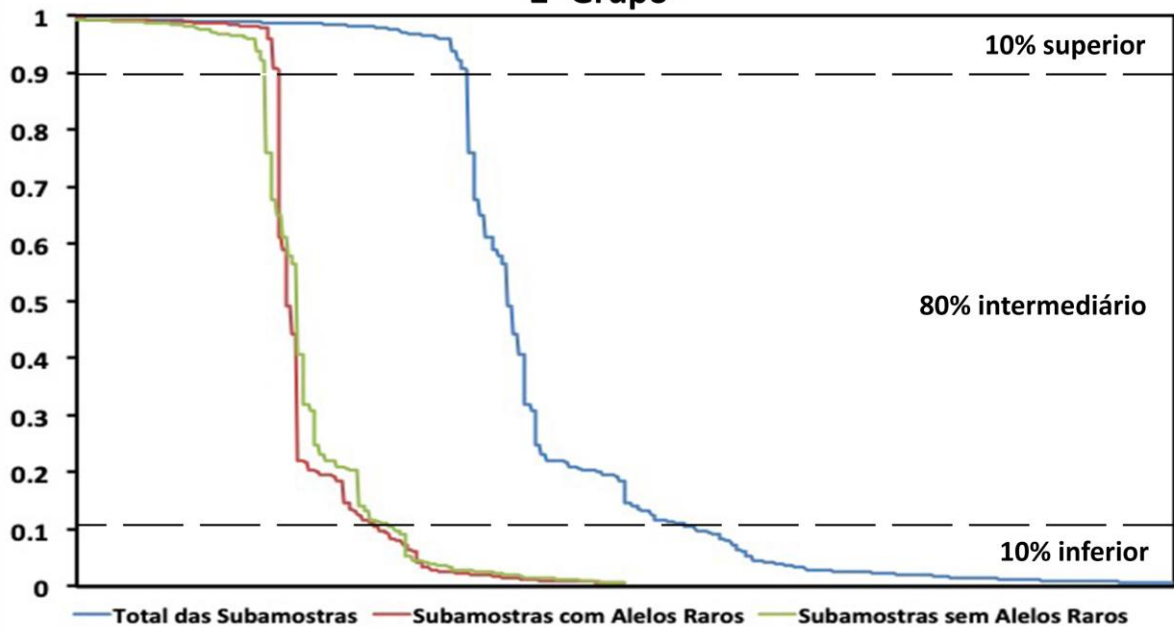
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311

Figura 6. Distribuição das proporções dos grupos genéticos ($K = 9$) entre o total de indivíduos de *D. antonietae* amostrados (3000), entre os com alelos raros (1500 indivíduos) e entre os sem alelos raros (1500 indivíduos). As linhas pontilhadas delimitam as classes de agrupamentos (10% inferior, 80% intermediário e 10% superior - ver Material e Métodos).

1º Grupo



2º Grupo



312

313

314

315

316

317

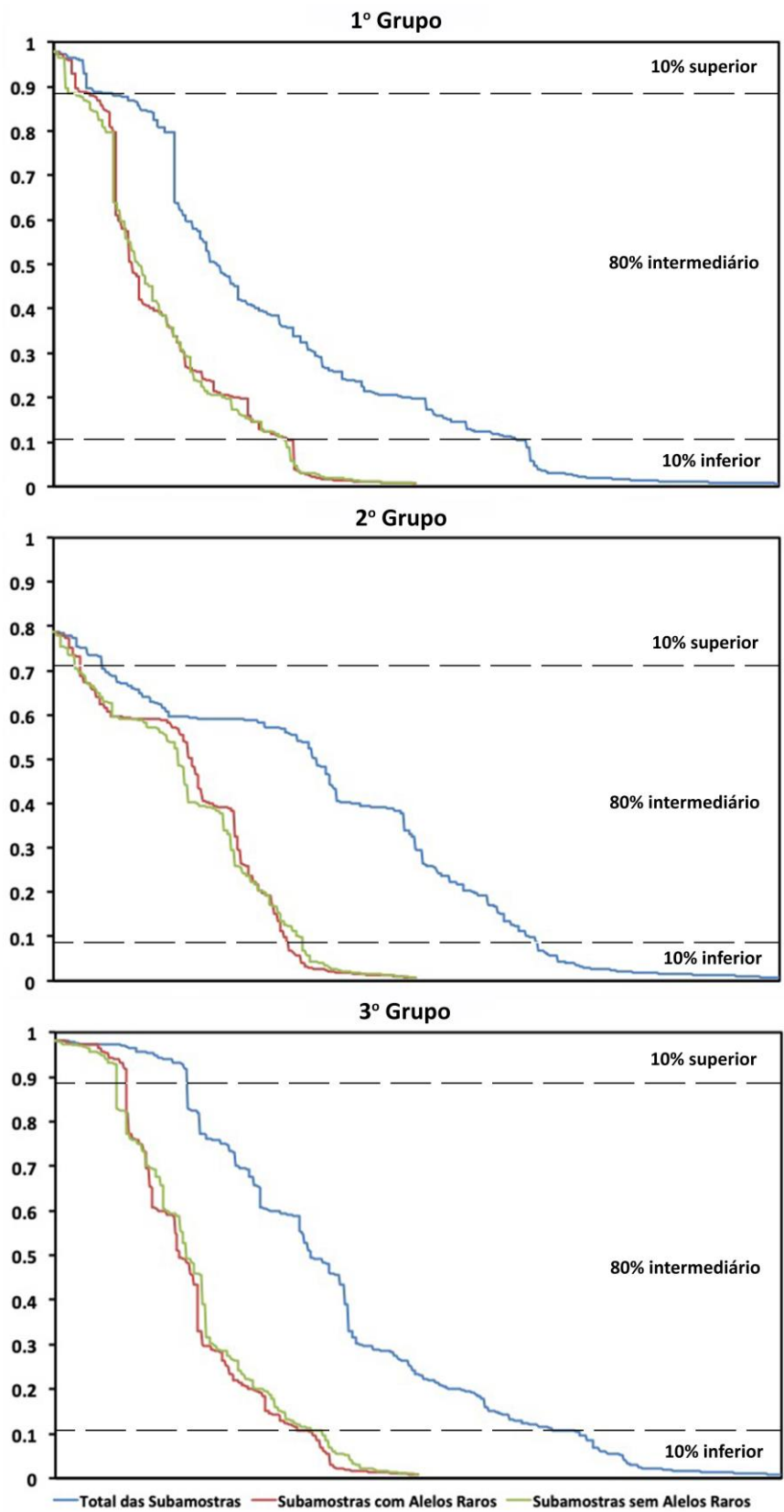
318

319 Figura 7. Distribuição das proporções dos grupos genéticos ($K = 2$) entre o total de
320 indivíduos de *D. prosaltans* amostrados (3500), entre os com alelos raros (1750
321 indivíduos) e entre os sem alelos raros (1750 indivíduos). As linhas pontilhadas
322 delimitam as classes de agrupamentos (10% inferior, 80% intermediário e 10% superior
323 - ver Material e Métodos).

324

325

326



327

328

329

330 Figura 8. Distribuição das proporções dos grupos genéticos ($K = 3$) entre o total de
331 indivíduos de *D. sturtevanti* amostrados (3500), entre os com alelos raros (1750
332 indivíduos) e entre os sem alelos raros (1750 indivíduos). As linhas pontilhadas
333 delimitam as classes de agrupamentos (10% inferior, 80% intermediário e 10% superior
334 - ver Material e Métodos).

335 Discussão

336 Um pequeno tamanho amostral (N) para análise da variação genética das populações
337 certamente resulta em desvio significativo dos parâmetros “reais”. No entanto, na
338 maioria das vezes, a partir de um determinado tamanho, o aumento no número de
339 indivíduos não mais representa um incremento significativo na qualidade dos dados
340 populacionais (Nei & Roychoudhury, 1974; Nei, 1978; Gorman & Jr-Renzi, 1979;
341 Pruett & Winker, 2008; Hale, Burg, & Steeves, 2012; Dubois et al., 2018; Zchonski,
342 2021). Considerando que as análises populacionais exigem coleta de material da
343 natureza, preferencialmente na abrangência da distribuição da espécie, que essas
344 análises são realizadas empregando diferentes números de locos disponíveis aplicados
345 separadamente por indivíduo, torna-se fundamental para a exequibilidade dos estudos, a
346 determinação de um tamanho amostral mínimo, tamanho amostral ideal, a partir do qual
347 a variação natural seja satisfatoriamente refletida, ao mesmo tempo em que reduza o
348 tempo, esforço e custo das análises de Genética de Populações. Neste trabalho, os
349 parâmetros aplicados para determinar o N ideal para espécies de *Drosophila* foram a
350 detecção de alelos representativos (> 5%), a heterozigosidade esperada (He), e os
351 índices de diferenciação genética (Fst) (Hale, Burg, & Steeves, 2012; Dubois et al.,
352 2018; Zchonski, 2021). O direcionamento da análise na detecção dos alelos
353 representativos se justifica pelos fatos da precisão na determinação de alelos raros
354 pouco se alterar com o aumento do tamanho amostral, e da importância dos alelos
355 representativos nos índices de He, os quais, apresentam maior relevância nas análises
356 populacionais (Hale, Burg, & Steeves, 2012; Dubois et al., 2018).

357 Apesar das espécies analisadas pertecerem a um mesmo táxon, suas associações com
358 hospedeiros (recursos alimentares) afetam diretamente a dispersão na área, o que,
359 consequentemente, influencia nas possibilidades de reprodução, que, por sua vez,
360 afetam o tamanho efetivo populacional (Ne) e o nível de variação das populações
361 (Brown & Kodric-Brown, 1977; Templeton et al. 1990, 2001; Frankham, 1995; Hartl &
362 Clark, 1997; Mateus & Sene, 2003; Dubois et al., 2018; Valadão et al., 2019). Esses, e
363 outros fatores demográficos, afetam a determinação do tamanho amostral ideal. Dentre
364 os diferentes critérios analisados para a determinação do N ideal, a diversidade genética,
365 representada pelo número de alelos por locos e He, foi o fator de maior variação entre as
366 espécies, e que mais influenciou nas diferenças observadas entre elas. Estudos da
367 variabilidade genética utilizando locos de microssatélites em populações naturais de
368 espécies de *Drosophila* da região Neotropical, revelaram maior diversidade para
369 espécies consideradas generalistas, como *D. maculifrons* (He = 0,745 – Silva, Machado,
370 & Mateus, 2015), *D. mediopunctata* (He = 0,864 – Heinz, 2012), do subgênero
371 *Drosophila*, e *D. sturtevantii* (He = 0,624 – Trava et al., 2021), subgênero *Sophophora*,
372 do que as especialistas ou de nichos mais restritos do subgênero *Drosophila*, como as
373 espécies cactófilas *D. gouveai* (He = 0,540 – Moraes & Sene, 2007) e *D. antonietae* (He
374 = 0,511 – Machado et al., 2010), e *D. ornatifrons* (He = 0,573 – Zorzato, 2015).

375 Os resultados obtidos neste trabalho para as espécies Neotropicais de *Drosophila*
376 indicam que diferentes espécies de um mesmo táxon mostraram distintos N ideais,
377 variando de 25 a 35 indivíduos, com os menores N (25 e 30) para as espécies com
378 preferências de recursos para alimentação e reprodução mais restritas, como foi o caso
379 de *D. ornatifrons* e a espécie cactófila *D. antonietae*, e o maior, N = 35, para as espécies
380 mais generalistas do subgênero *Sophophora*, *D. prosaltans* e *D. sturtevanti*, que se
381 alimentam preferencialmente de frutos em decomposição, recurso abundante nas áreas
382 de mata da região Neotropical (Matzkin et al., 2009; Goñi et al., 2012; Valadão et al.,
383 2019).

384 Nas subamostras deste trabalho, *Drosophila antonietae* apresentou os menores valores
385 de He e o segundo menor N ideal (N = 30). A menor diversidade genética desta espécie
386 se deve ao fato desta ser cactófila, que tende a não se dispersar na área, alimentando-se
387 e reproduzindo-se no mesmo sítio em que se desenvolveu (Moraes & Sene, 2002;
388 Mateus & Sene, 2003). *Drosophila ornatifrons* apresentou He maior apenas que *D.*
389 *antonietae* e o menor N ideal (N = 25). Apesar desta espécie ser considerada
390 generalista, por ter sido coletada em diferentes tipos de tecidos vegetais e fungos em
391 decomposição (Valadão et al., 2019), ela parece ter preferência por fontes de recursos
392 mais ricas em proteínas, visto que foi coletada em carcaças animais (Goñi et al., 2012).
393 Outras observações que reforçam a dieta com maior concentração de proteína desta
394 espécie é sua baixa viabilidade em meio de cultura padrão de banana (Santos et al.
395 submetido), e total inviabilidade em meio de cultura com menor concentração de
396 proteína em relação a açúcar (Santos, 2020), meio LPS (*Low Protein Sugar*)
397 desenvolvido por Matzkin et al. (2011). Além disso, *D. ornatifrons* é uma espécie pouco
398 abundante, mesmo em áreas com maior umidade e temperatura amena, condições
399 favoráveis a sua ocorrência (Gustani et al., 2015). A adaptação de *D. ornatifrons* a uma
400 dieta proteica, talvez menos disponível em relação a abundância de frutos e outros
401 tecidos vegetais, poderia resultar em menor Ne e diversidade genética e consequente no
402 menor N ideal revelado para esta espécie.

403 As espécies do subgênero *Sophophora*, *D. prosaltans* e *D. sturtevanti*, por outro
404 lado, apresentaram os maiores valores de He e coincidente N ideal igual a 35
405 indivíduos. Essas espécies, assim como as demais representantes do subgênero
406 *Sophophora*, utilizam uma variedade de frutos como hospedeiros, abundante recurso
407 nas áreas de mata da região Neotropical, conforme mencionado anteriormente (Matzkin
408 et al., 2009; Goñi et al., 2012; Valadão et al., 2019). Outro achado para estas espécies é
409 que a transferibilidade dos locos descritos em *D. sturtevanti* para *D. prosaltans* (Roman,
410 Trava, Madi-Ravazzi, 2020) não afetou a diversidade alélica dos três locos de
411 microssatélites analisados em comum (G, J e L, ver Tabela 1 MS em Material
412 Suplementar), e ainda resultou no mesmo N ideal, confirmando o sucesso da
413 amplificação cruzada de locos de microssatélites em espécies filogeneticamente
414 próximas (Laborda, Klaczko, & Souza, 2009; Tractz et al., 2012; Roman, Trava, &
415 Madi-Ravazzi, 2020).

416 A determinação do tamanho amostral ideal para análise de Genética de
417 Populações por meio de locos de microssatélites foi realizada para alguns grupos de
418 insetos, gafanhoto (*Locusta migratoria*) (Luna & Dexing, 2004), formiga (*Formica*
419 *lugubris*) (Hale, Burg, & Steeves, 2012), e borboleta (*Boloria aquilonaris*) (Dubois et

420 al., 2018). Para a espécie de borboleta, foram determinados tamanhos ideais de $N = 30$
421 (simulações) e $N = 40$ (dados empíricos), porém o ganho na precisão da determinação
422 das frequências gênicas é limitado em amostras acima de 20 indivíduos neste inseto.
423 Assim como observado para as espécies de *Drosophila* analisadas aqui, o trabalho com
424 borboletas também aponta para um aumento do N ideal quanto maior a diversidade das
425 populações estudadas. Hale, Burg, & Steeves (2012) sugeriram que as espécies de
426 formiga e de albatroz, *Thalassarche melanophris*, analisadas seriam modelos para
427 organismos com grande tamanho populacional. No entanto, ambas são espécies que
428 vivem em colônias, sendo as formigas insetos eussociais sob forte efeito de seleção de
429 parentesco (Hamilton, 1972), o que diminuiria a diversidade genética e a estimativa para
430 o N ideal. Também os albatrozes apresentam características de história de vida
431 peculiares e extremas, como alta taxa de sobrevivência dos adultos, baixa taxa
432 reprodutiva anual e maturação sexual tardia (Prince et al., 1994), que seriam fatores da
433 estrutura populacional a serem considerados na determinação do N ideal, antes que
434 generalizar a espécie como modelo para grandes populações.

435 A influência do número de locos de microssatélites na detecção da riqueza
436 alélica e de alelos raros em populações de *Drosophila nigrosparsa* foi avaliada por
437 Arthofer et al. (2018), mas não do tamanho amostral ideal para análises populacionais.
438 Além dos fatores demográficos discutidos acima, as mudanças nas frequências alélicas,
439 que influenciam nas estimativas dos parâmetros populacionais, também dependem do
440 número de locos e do nível de polimorfismo destes, visto que a diversidade de um loco
441 é proporcional a taxa de mutação e ao N_e (Eanes & Koehn, 1977; Kalinowski, 2005;
442 Arthofer et al., 2018; Wang et al., 2021). Wang et al. (2021) concluíram que a
443 confiabilidade de amostragens populacionais de espécies do gênero *Rhododendron*, da
444 família Ericaceae das plantas, é afetada com menos de 12 ou 11 locos de microssatélites
445 avaliados. Para a espécie *D. nigrosparsa*, a riqueza alélica e os alelos raros de locos de
446 microssatélites são corretamente determinados pela análise de oito locos, no entanto a
447 estrutura genética pode ser ainda verificada com a utilização de apenas dois locos
448 altamente polimórficos (Arthofer et al., 2018). Locos com altas taxas de mutação
449 também produzem estimativas mais confiáveis de distância genética, sem a necessidade
450 de grande tamanho amostral por população, do que locos com taxas baixas de mutação
451 (Kalinowski, 2005).

452 Estudos que estabeleceram a relação dos índices de F_{st} das populações com a
453 determinação do tamanho amostral revelaram que espécies com valores baixos de F_{st}
454 ($<0,01$) requerem tamanhos amostrais de 100 ou mais indivíduos por população
455 (Kalinowski, 2005; Pruett & Winker, 2008). Este não parece ser o caso para as espécies
456 de *Drosophila* analisadas neste trabalho, pois, segundo a classificação qualitativa para a
457 estatística F , proposta por Wright (1978), *D. ornatifrons*, *D. prosaltans* e *D. sturtevantii*
458 apresentam diferenciação moderada entre suas populações ($F_{st} = 0,0796$; $0,173$; e
459 $0,0674$, respectivamente) e *D. antonietae* apesar da baixa diferenciação, esta é maior
460 que $0,01$ ($F_{st} = 0,0457$) (Machado et al., 2010; Zorzato, 2015; Paixão, 2020; Trava et
461 al., 2021). Nas subamostras, a variação dos valores de F_{st} médios foi praticamente a
462 mesma entre as espécies, de próximo de zero para $N = 95$ à $0,035$ (*D. ornatifrons* e *D.*
463 *prosaltans*) e $0,038$ (*D. antonietae* e *D. sturtevantii*) para $N = 5$.

464 A detecção de alelos raros é relevante para estimativas de fluxo gênico,
465 importantes para estudos de conservação (Slatkin, 1985; Russello et al., 2007, Coelho et
466 al., 2018), contudo um tamanho amostral muito grande, da ordem de centenas de
467 indivíduos, é necessário para medidas precisas dessa riqueza alélica, enquanto que
468 estimativas da H_e , conforme descrito anteriormente, requer menor tamanho amostral ($<$
469 30). Além disso, quando alelos raros são amostrados, a frequência alélica é
470 superestimada, especialmente em pequeno tamanho amostral (Miyamoto et al., 2008;
471 Hale, Burg, & Steeves, 2012; Arthofer et al., 2018). Ainda, como esperado, alelos raros
472 não modificariam a determinação de grupos genéticos das populações (Zchonski, 2021),
473 pois a análise Bayesiana realizada pelo programa *Structure* é baseada na composição
474 genotípica dos indivíduos e nas frequências alélicas das populações, não sendo afetada,
475 portanto, por alelos com frequências muito baixas. Segundo os resultados deste trabalho
476 com espécies de *Drosophila*, os alelos raros também não interferiram na distribuição da
477 proporção dos grupos genéticos dos indivíduos dentro das classes de agrupamentos
478 propostas aqui (10% inferior, 80% intermediário e 10% superior, ver Figuras 5 à 8).
479 Esses dados reforçam a eficácia de estudos de tamanho amostral ideal baseados na
480 determinação das frequências alélicas representativas, as quais são resultado de
481 diferentes características demográficas das populações (Dubois et al., 2018).

482 Conclusões

483 O N ideal para análise de locos de microssatélites determinado para as espécies
484 Neotropicais de *Drosophila* avaliadas neste trabalho, variou em decorrência da
485 diversidade genética, que é afetada pela preferência por hospedeiros para alimentação e
486 reprodução: espécies generalistas do subgênero *Sophophora*, *D. prosaltans* e *D.*
487 *sturtevanti*, apresentaram maior H_e e maior N ideal, e as espécies com especificidades
488 ecológicas do subgênero *Drosophila*, *D. ornatifrons* com dieta proteica e *D. antonietae*
489 cactófila, com menor H_e e menor N ideal. Estas espécies podem ser modelos para outras
490 destes subgêneros, que são os mais diversos em termos de número de espécies entre os
491 oito subgêneros do gênero *Drosophila*, cerca de 300 espécies para *Sophophora* e o
492 dobro para *Drosophila*. Estes subgêneros apresentam longo tempo de divergência,
493 aproximadamente 60 milhões de anos (Tamura, Nei, & Kumar, 2004), e desde então as
494 espécies pertencentes a eles passaram por diferentes histórias adaptativas que podem ter
495 afetado a capacidade de ocupação de nichos. Neste sentido, os N ideais obtidos para as
496 espécies de *Drosophila* Neotropicais analisadas neste trabalho também podem ser
497 aplicados para outros grupos animais que ocupam tipos de habitats e fatores
498 demográficos semelhantes a elas.

499

500

501

502

503

504

505
506
507
508
509
510
511
512
513
514

515 Supplementary Materials

516 Tabela 1 MS. Frequências alélicas de 100 genótipos de microssatélites provenientes de
517 dados empíricos de espécies de *Drosophila*. *D. ornatifrons* – Zorzato (2015); *D.*
518 *prosaltans* – Paixão (2020); *D. sturtevantii* – Trava et al. (2021); *D. antonietae* –
519 Machado et al. (2010).

<i>Drosophila ornatifrons</i>			<i>Drosophila prosaltans</i>			<i>Drosophila sturtevantii</i>			<i>Drosophila antonietae</i>				
Loco	Alelo	Freq.	Loco	Alelo	Freq.	Loco	Alelo	Freq.	Loco	Alelo	Freq.		
34	1	0,005	A	1	0,010	E	1	0,010	1	1	0,030		
	2	0,010		2	0,010		2	0,025		2	0,095		
	3	0,025		3	0,010		3	0,055		3	0,105		
	4	0,165		4	0,030		4	0,150		4	0,135		
	5	0,125		5	0,030		5	0,280		5	0,545		
	6	0,185		6	0,105		6	0,120		6	0,090		
	7	0,230		7	0,180		7	0,175		2	1	0,090	
	8	0,115		8	0,075		8	0,045			2	0,865	
	9	0,065		9	0,150		9	0,090		9	0,020		
	10	0,005		10	0,070		10	0,050		10	0,025		
	11	0,060		11	0,160		G	1		0,025	3	1	0,025
	12	0,010		12	0,045			2		0,050		2	0,130
53	1	0,810	13	0,055	3	0,090	3	0,270					
	2	0,190	14	0,025	4	0,090	4	0,375					
57	1	0,430	15	0,040	5	0,080	5	0,190					

	2	0,150		16	0,005		6	0,145		6	0,010
	3	0,230	B	1	0,035		7	0,160	4	1	0,045
	4	0,175		2	0,015		8	0,155		2	0,260
	5	0,015		3	0,065		9	0,150		3	0,285
79	1	0,750		4	0,055		10	0,015		6	0,410
	2	0,250		5	0,155		11	0,025	5	1	0,095
87	1	0,040		6	0,070		12	0,015		2	0,705
	2	0,175		7	0,095	J	1	0,030		3	0,195
	3	0,205		8	0,085		2	0,070		4	0,005
	4	0,240		9	0,120		3	0,195	6	1	0,160
	5	0,185		10	0,090		4	0,260		2	0,010
	6	0,095		11	0,040		5	0,165		3	0,240
	7	0,055		12	0,105		6	0,130		4	0,020
	8	0,005		13	0,025		7	0,060		5	0,180
96	1	0,035		14	0,025		8	0,060		6	0,390
	2	0,450		15	0,020		9	0,020	7	1	0,010
	3	0,090	L	1	0,010		10	0,005		2	0,860
	4	0,195		2	0,015		11	0,005		3	0,005
	5	0,140		3	0,090	K	1	0,015		4	0,010
	6	0,090		4	0,135		2	0,075		5	0,065
107	1	0,005		5	0,205		3	0,155		6	0,050
	2	0,315		6	0,260		4	0,220			
	3	0,215		7	0,095		5	0,200			
	4	0,330		8	0,035		6	0,130			
	5	0,085		9	0,010		7	0,120			
	6	0,025		10	0,075		8	0,085			
	7	0,025		11	0,030	L	1	0,005			

520

Tabela 1 MS. Continuação.

<i>Drosophila ornatifrons</i>			<i>Drosophila prosaltans</i>			<i>Drosophila sturtevanti</i>			<i>Drosophila antonietae</i>		
Loco	Alelo	Freq.	Loco	Alelo	Freq.	Loco	Alelo	Freq.	Loco	Alelo	Freq.

118	1	0,015		12	0,030		2	0,015
	2	0,375		13	0,005		3	0,080
	3	0,450		14	0,005		4	0,185
	4	0,110	D	1	0,015		5	0,135
	5	0,045		2	0,030		6	0,140
	6	0,005		3	0,015		7	0,160
126	1	0,615		4	0,015		8	0,135
	2	0,180		5	0,055		9	0,075
	3	0,185		6	0,100		10	0,045
	4	0,010		7	0,245		11	0,015
	5	0,010		8	0,075		13	0,010
				9	0,250	M	1	0,025
				10	0,050		2	0,060
				11	0,075		3	0,265
				12	0,040		4	0,145
				13	0,020		5	0,235
				14	0,015		6	0,135
			F	1	0,005		7	0,100
				2	0,035		8	0,030
				4	0,005		9	0,005
			5	0,120	O	1	0,005	
			6	0,060		2	0,015	
			7	0,055		3	0,080	
			8	0,175		4	0,035	
			9	0,085		5	0,020	
			10	0,125		6	0,015	
			11	0,100		7	0,110	
			12	0,115		8	0,105	
			13	0,095		9	0,195	
			14	0,020		10	0,245	

	15	0,005		11	0,100
G	1	0,015		12	0,060
	2	0,020		13	0,015
	3	0,045			
	4	0,045			
	5	0,125			
	6	0,190			
	7	0,100			
	8	0,170			
	9	0,160			
	10	0,060			
	11	0,015			
	12	0,045			
	13	0,010			
J	1	0,005			
	2	0,040			
	3	0,045			

521

Tabela 1 MS. Continuação.

<i>Drosophila ornatifrons</i>			<i>Drosophila prosaltans</i>			<i>Drosophila sturtevanti</i>			<i>Drosophila antonietae</i>		
Loco	Alelo	Freq.	Loco	Alelo	Freq.	Loco	Alelo	Freq.	Loco	Alelo	Freq.
				4	0,010						
				5	0,050						
				6	0,140						
				7	0,090						
				8	0,160						
				9	0,120						
				10	0,130						
				11	0,160						
				12	0,045						
				13	0,005						

522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537

538 Referências

- 539 Arthofer, W., Heussler, C., Krapf, P., Schlick-Steiner, B. C., & Steiner, F. M. (2018). Identifying the
540 minimum number of microsatellite loci needed to assess population genetic structure: a case
541 study in fly culturing. *Fly*, 12(1), 13-22.
- 542 Brown, J. H., & Kodric-Brown, A. (1977). Turnover rates in insular biogeography: effect of immigration
543 on extinction. *Ecology*, 58(2), 445-449.
- 544 Coelho, N. H. P., Tambarussi, E. V., Aguiar, B. I., et al. (2018). Understanding genetic diversity, spatial
545 genetic structure, and mating system through microsatellite markers for the conservation and
546 sustainable use of *Acrocomia aculeata* (Jacq.) Lodd. *Ex Mart. Conservation Genetics*, 19(4),
547 879-891.
- 548 Dubois, Q., Lebigre, C., Schtickzelle, N., & Turlure, C. (2018). Sex, size and timing: Sampling design for
549 reliable population genetics analyses using microsatellite data. *Methods in Ecology and*
550 *Evolution*, 9(4), 1036-1048.
- 551 Eanes, W. F., & Koehn, R. K. (1977). The correlation of rare alleles with heterozygosity: determination
552 of the correlation for the neutral models. *Genetics Research*, 29(3), 223-230.
- 553 Earl, D. A., & VonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for
554 visualizing STRUCTURE output and implementing the Evanno method. *Conservation genetics*
555 *resources*, 4(2), 359-361.
- 556 Estoup, A., Garnery, L., Solignac, M., & Cornuet, J. M. (1995). Microsatellite variation in honey bee
557 (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and
558 stepwise mutation models. *Genetics*, 140(2), 679-695.
- 559 Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the
560 software STRUCTURE: a simulation study. *Molecular ecology*, 14(8), 2611-2620.

- 561 Fellows, D. P., & Heed, W. B. (1972). Factors affecting host plant selection in desert-adapted cactiphilic
562 *Drosophila*. *Ecology*, 53(5), 850-858.
- 563 Frankham, R. (1995). Conservation genetics. *Annual review of genetics*, 29(1), 305-327.
- 564 Goñi, B., Remedios, M., González-Vainer, P., Martínez, M., & Vilela, C. R. (2012). Species of
565 *Drosophila* (Diptera: Drosophilidae) attracted to dung and carrion baited pitfall traps in the
566 Uruguayan Eastern Serranías. *Zoologia (Curitiba)*, 29, 308-317.
- 567 Gorman, G. C., & Renzi Jr, J. (1979). Genetic distance and heterozygosity estimates in electrophoretic
568 studies: effects of sample size. *Copeia*, 242-249.
- 569 Gustani, E. C., Oliveira, A. P. F., Santos, M. H., Machado, L. P., & Mateus, R. P. (2015). Demographic
570 structure and evolutionary history of *Drosophila ornatifrons* (Diptera, Drosophilidae) from
571 Atlantic Forest of Southern Brazil. *Zoological science*, 32(2), 141-150.
- 572 Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic
573 studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies.
- 574 Hamilton, W. D. (1972). Altruism and related phenomena, mainly in social insects. *Annual Review of*
575 *Ecology and systematics*, 193-232.
- 576 Hammer, Ø., Harper, D. A., & Ryan, P. D. (2001). PAST: Paleontological statistics software package for
577 education and data analysis. *Palaeontologia electronica*, 4(1), 9.
- 578 Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). *Principles of population genetics* (Vol. 116).
579 Sunderland: Sinauer associates.
- 580 Hedrick, P. W. (1986). Genetic polymorphism in heterogeneous environments: a decade later. *Annual*
581 *review of ecology and systematics*, 535-566.
- 582 Heed, W. B., & Wheeler, M. R. (1957). Thirteen new species in the genus *Drosophila* from the
583 Neotropical region. *The University of Texas Publications*, 5721, 17-38.
- 584 Heinz, N. P. (2012). Variabilidade molecular sazonal de *Drosophila mediopunctata* (Diptera:
585 Drosophilidae). (Master Dissertation. Universidade Estadual Centro-Oeste, Brazil).
- 586 Kalinowski, S. T. (2005). Do polymorphic loci require large sample sizes to estimate genetic distances?.
587 *Heredity*, 94(1), 33-36.
- 588 Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: a
589 program for identifying clustering modes and packaging population structure inferences across
590 K. *Molecular ecology resources*, 15(5), 1179-1191.
- 591 Laborda, P. R., Klaczko, L. B., & de Souza, A. P. (2009). *Drosophila mediopunctata* microsatellites II:
592 cross-species amplification in the *tripunctata* group and other *Drosophila* species. *Conservation*
593 *Genetics Resources*, 1(1), 281-296.
- 594 Laborda, P. R., Mori, G. M., & de Souza, A. P. (2009). *Drosophila mediopunctata* microsatellites I: more
595 than a hundred polymorphic loci available for genetic studies. *Conservation Genetics Resources*,
596 1(1), 297-307.
- 597 Luna, Y., & Dexing, Z. (2004). Effects of sample size on various genetic diversity measures in population
598 genetic study with microsatellite DNA markers. *Dong wu xue bao.[Acta zoologica Sinica]*,
599 50(2), 279-290.

- 600 Machado, L. P. B., Manfrin, M. H., Silva-Junior, W. A., & Sene, F. M. (2003). Microsatellite loci in the
601 cactophilic species *Drosophila antonietae* (Diptera; Drosophilidae). *Molecular Ecology Notes*,
602 3(1), 159-161.
- 603 Machado, L. P., Mateus, R. P., Sene, F. M., & Manfrin, M. H. (2010). Microsatellite allele sequencing in
604 population analyses of the South American cactophilic species *Drosophila antonietae* (Diptera:
605 Drosophilidae). *Biological Journal of the Linnean Society*, 100(3), 573-584.
- 606 Mateus, R. P., & Sene, F. M. (2003). Temporal and spatial allozyme variation in the South American
607 cactophilic *Drosophila antonietae* (Diptera; Drosophilidae). *Biochemical Genetics*, 41(7), 219-
608 233.
- 609 Matzkin, L. M., Mutsaka, K., Johnson, S., & Markow, T. A. (2009). Metabolic pools differ among
610 ecologically diverse *Drosophila* species. *Journal of Insect Physiology*, 55(12), 1145-1150.
- 611 Matzkin, L. M., Johnson, S., Paight, C., Bozinovic, G., & Markow, T. A. (2011). Dietary protein and
612 sugar differentially affect development and metabolic pools in ecologically diverse *Drosophila*.
613 *The Journal of nutrition*, 141(6), 1127-1133.
- 614 Miyamoto, N., Fernández-Manjarrés, J. F., Morand-Prieur, M. E., Bertolino, P., & Frascaria-Lacoste, N.
615 (2008). What sampling is needed for reliable estimations of genetic diversity in *Fraxinus*
616 *excelsior* L. (Oleaceae)? *Annals of forest science*, 65(4), 1.
- 617 Moraes, E. M., & Sene, F. M. (2002). Breeding structure of an isolated cactophilic *Drosophila* population
618 on a sandstone table hill. *Journal of Zoological Systematics and Evolutionary Research*, 40(3),
619 123-128.
- 620 Moraes, E. M., & Sene, F. M. (2007). Microsatellite and morphometric variation in *Drosophila gouveai*:
621 the relative importance of historical and current factors in shaping the genetic population structure.
622 *Journal of Zoological Systematics and Evolutionary Research*, 45(4), 336-344.
- 623 Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of
624 individuals. *Genetics*, 89(3), 583-590.
- 625 Nei, M., & Roychoudhury, A. (1974). Sampling variances of heterozygosity and genetic distance.
626 *Genetics*, 76(2), 379-390.
- 627 Oliveira, D. C., Almeida, F. C., O'Grady, P. M., Armella, M. A., DeSalle, R., & Etges, W. J. (2012).
628 Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny
629 of the *Drosophila repleta* species group. *Molecular Phylogenetics and Evolution*, 64(3), 533-
630 544.
- 631 Paixão, J. F. (2020). Diversidade genética de *Drosophila prosaltans* e *D. austrosaltans* (Diptera,
632 Drosophilidae) de regiões de Mata Atlântica avaliada por marcadores microssatélites. (Master
633 Dissertation. Universidade Estadual Paulista, Brazil).
- 634 Peakall, R. & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software
635 for teaching and research—an update. *Bioinformatics*, 28, 2537e2539.
- 636 Pereira, M. R., Vilela, C. R., & Sene, F. D. M. (1983). Notes on breeding and feeding sites of some
637 species of the *repleta* group of the genus *Drosophila* (Diptera, Drosophilidae). *Ciência e Cultura*,
638 35(9), 1313-1319.
- 639 Prince, P. A., Rothery, P., Croxall, J. P., & Wood, A. G. (1994). Population dynamics of Black-browed
640 and Grey-headed Albatrosses *Diomedea melanophris* and *D. chrysostoma* at Bird Island, South
641 Georgia. *Ibis*, 136(1), 50-71.

- 642 Pruetz, C., & Winker, K. (2008). The effects of sample size on population genetic diversity estimates in
643 song sparrows *Melospiza melodia*. *Journal of Avian Biology*, 39(2), 252-256.
- 644 Roman, B. E., Trava, B. M., & Madi-Ravazzi, L. (2020). Development and transferability of
645 microsatellite markers for species of the *saltans* group of *Drosophila* (Diptera: Drosophilidae).
646 *Genet. Mol. Res*, 19(4).
- 647 Ruiz, A., & Heed, W. B. (1988). Host-plant specificity in the cactophilic *Drosophila mulleri* species
648 complex. *The Journal of Animal Ecology*, 237-249.
- 649 Russello, M. A., Brazaitis, P., Gratten, J., Watkins-Colwell, G. J., & Caccone, A. (2007). Molecular
650 assessment of the genetic integrity, distinctiveness and phylogeographic context of the Saltwater
651 crocodile (*Crocodylus porosus*) on Palau. *Conservation Genetics*, 8(4), 777-787.
- 652 Santos, C. H. (2020). Variação Bionômica e Metabólica em três espécies de *Drosophila* submetidas a
653 diferentes dietas (Master Dissertation. Universidade Estadual Centro-Oeste, Brazil).
- 654 Santos, C. H., Santos K. A. V., Machado, L. P. B., & Mateus, R. P. (submitted). Life history traits and
655 metabolic pool variation in Neotropical *Drosophila* species (Diptera, Drosophilidae). *Zoological
656 Studies*.
- 657 Silva, D. C., Machado, L. P. B., & Mateus, R. P. (2015). Migration rate and genetic diversity of two
658 *Drosophila maculifrons* (Duda, 1927) populations from Highland Araucaria Forest Fragments in
659 Southern Brazil. *Brazilian Journal of Biology*, 75, 254-255.
- 660 Slatkin, M. (1985). Rare alleles as indicators of gene flow. *Evolution*, 39(1), 53-65.
- 661 Tamura, K., Nei, M., & Kumar, S. (2004). Prospects for inferring very large phylogenies by using the
662 neighbor-joining method. *Proceedings of the National Academy of Sciences*, 101(30), 11030-
663 11035.
- 664 Templeton, A. R., Shaw, K., Routman, E., & Davis, S. K. (1990). The genetic consequences of habitat
665 fragmentation. *Annals of the Missouri Botanical Garden*, 13-27.
- 666 Templeton, A. R., Robertson, R. J., Brisson, J., & Strasburg, J. (2001). Disrupting evolutionary processes:
667 the effect of habitat fragmentation on collared lizards in the Missouri Ozarks. *Proceedings of the
668 National Academy of Sciences*, 98(10), 5426-5432.
- 669 Tosi, D., Martins, M., Vilela, C. R., & Pereira, M. A. Q. R. (1990). On a new cave-dwelling species of
670 bat-guano-breeding *Drosophila* closely related to *D. repleta* Wollaston (Diptera, Drosophilidae).
671 *Revta bras. Genet*, 13(1), 19-31.
- 672 Tractz, C. C., Salomon, G. R., Zorzato, S. V., Machado, L. P. B., Mateus, R. P. (2012). Allele diversity of
673 cross-species microsatellite amplification on populations of *Drosophila guarani* species group
674 from Araucaria Forest in Brazil. *Drosophila Information Service*, 95, 76-79.
- 675 Trava, B. M., Mateus, R. P., Machado, L. P. B., & Madi-Ravazzi, L. (2021). Moderate Population
676 Structure in *Drosophila sturtevantii* from the South American Atlantic Forest Biome. *Zoological
677 studies*, 60.
- 678 Valadão, H., Proença, C. E., Kuhlmann, M. P., Harris, S. A., & Tidon, R. (2019). Fruit-breeding
679 drosophilids (Diptera) in the Neotropics: playing the field and specialising in generalism?.
680 *Ecological Entomology*, 44(6), 721-737.
- 681 Wan, Q. H., Wu, H., Fujihara, T., & Fang, S. G. (2004). Which genetic marker for which conservation
682 genetics issue?. *Electrophoresis*, 25(14), 2165-2176.

- 683 Wang, H., Yang, B., Wang, H., & Xiao, H. (2021). Impact of different numbers of microsatellite markers
684 on population genetic results using SLAF-seq data for *Rhododendron* species. Scientific reports,
685 11(1), 1-9.
- 686 Wright, S. (1978). Variability within and among natural populations. Evolution and the Genetics of
687 Populations.
- 688 Zchonski, F. L. (2021). Determinação do tamanho mínimo amostral para estudos genéticos populacionais
689 em plantas utilizando marcadores microssatélites (Master Dissertation. Universidade Estadual
690 Centro-Oeste, Brazil).
- 691 Zorzato, S. V. (2015). Análise da Estrutura Genética Populacional de *Drosophila ornatifrons* (Diptera;
692 Drosophilidae) (Master Dissertation. Universidade Estadual Centro-Oeste, Brazil).

6. CONSIDERAÇÕES FINAIS

Diferentes grupos de organismos, diferentes espécies de um mesmo grupo, para diferentes conjuntos de locos de microssatélites polimórficos devem ser estudados com o objetivo de auxiliar na determinação do tamanho amostral ideal para estudos da estrutura das populações naturais. As metodologias para a determinação do tamanho amostral ideal para análises populacionais com locos de microssatélites devem se ajustar ao objetivo do trabalho, e quando o número de indivíduos sugerido não for atingido, por exemplo pela dificuldade de captura, ou para espécies ameaçadas, as medidas de diversidade genética realizadas devem considerar e comparar os resultados, com as devidas justificativas para as lacunas na amostragem e suas consequências nos resultados obtidos.

7. REFERÊNCIAS BIBLIOGRÁFICAS

ARTHOFER, W.; HEUSSLER, C.; KRAPF, P.; SCHLICK-STEINER, B. C.; STEINER, F. M. Identifying the minimum number of microsatellite loci needed to assess population genetic structure: a case study in fly culturing. **Fly**, 12: 13-22, 2018.

BÄCHLI, G. TaxoDros, the database on taxonomy of Drosophilidae. <http://taxodros.unizh.ch>. Acessado em 15 de julho de 2022.

BARRIOS-LEAL, D. Y.; MENEZES, R. S.; RIBEIRO, J. V.; BIZZO, L.; SENE, F. M.; NEVES-DA-ROCHA, J.; MANFRIN, M. H. A holocenic and dynamic hybrid zone between two cactophilic *Drosophila* species in a coastal lowland plain of the Brazilian Atlantic Forest. **J. Evol. Biol.**, 34: 1737-1751, 2021.

BATISTA, M. R. D.; ANANINA, G.; KLACZKO, L. B. Unexpected long-term changes in chromosome inversion frequencies in a Neotropical *Drosophila* species. **Clim. Res.**, 53: 131-140, 2012.

BASHALKHANOV, S.; PANDEY, M.; RAJORA, O. P. A simple method for estimating genetic diversity in large populations from finite sample sizes. **BMC Genetics**, 10: 1-10, 2009.

CENZI DE RÉ, F.; GUSTANI, E. C.; OLIVEIRA, A. P. F.; MACHADO, L. P. B.; MATEUS, R. P.; LORETO, E. L.; ROBE, L. J. Brazilian populations of *Drosophila maculifrons* (Diptera: Drosophilidae): low diversity levels and signals of a population expansion after the Last Glacial Maximum. **Biol. J. Linn. Soc.**, 112: 55-66, 2014.

CHAMBERS, G. K.; MACAVOY, E. S. Microsatellites: consensus and controversy. **Comp. Biochem. Physiol. B, Biochem. Mol. Biol.**, 126: 455-476, 2000.

DALLAS, J. F. Estimation of microsatellite mutation rates in recombinant inbred strains of mouse. **Mamm. Gen.**, 3: 452-456, 1992.

da MATA, R. A., MCGEOCH, M., TIDON, R. Drosophilid assemblages as a bioindicator system of human disturbance in the Brazilian Savanna. **Biodivers. Conserv.**, 17: 2899-2916, 2008.

da MATA, R. A.; MCGEOCH, M.; TIDON, R. Drosophilids (Insecta, Diptera) as tools for conservation biology. **Nat. & Conserv.**, 8: 60-65, 2010.

DARWIN, C. R. A origem das espécies por meio da seleção natural ou a preservação das raças favorecidas na luta pela vida. 1. ed. São Paulo: Martin Claret, 2014.

DUBOIS, Q.; LEBIGRE, C.; SHTICKZELLE, N.; TURLURE, C. Sex, size and timing: Sampling design for reliable population genetics analyses using microsatellite data. **Methods Ecol Evol.**, 9: 1036–1048, 2018.

EANES, W. F.; KOEHN, R. K. The correlation of rare alleles with heterozygosity: determination of the correlation for the neutral models. **Genet. Res.**, 29: p. 223-230, 1977.

EARL, D. A.; VONHOLD, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. **Conserv. Genet. Resour.**, v.4, n.2, p.359-361, 2012.

ESTOUP, A.; GARNERY, L.; SOLIGNAC, M.; CORNUET, J. M. Microsatellite variation in honey bee (*Apis mellifera* L.) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. **Genetics**, 140: 679-695, 1995.

ESTOUP, A.; JARNE, P.; CORNUET, J. M. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. **Mol. Ecol.**, 11: 1591-1604, 2002.

EVANNO G.; REGNAUT, S.; GOUDET, J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. **Mol. Ecol.**, 14: 2611–2620, 2005.

FORD, E. B. Polimorphism and taxonomy. In: Huxley, J. (ed.). **The new systematics**. Oxford: Clarendon Press, p. 493-523, 1940.

FREEMAN, S.; HERRON, J. C. Análise Evolutiva. 4. ed. Porto Alegre: Artmed, 2009.

GOÑI, B.; REMEDIOS, M.; GONZÁLEZ-VAINER, P.; MARTÍNEZ, M.; VILELA, C. R. Species of *Drosophila* (Diptera: Drosophilidae) attracted to dung and carrion baited pitfall traps in the Uruguayan Eastern Serranías. **Zoologia**, 29: 308-317, 2012.

GOTTSCHALK, M. S.; HOFMANN, P. R. P.; VALENTE, V. L. S. Diptera, Drosophilidae: historical occurrence in Brazil. **Check List**, 4: p. 485-518, 2008.

GOTTSCHALK, M. S.; BIZZO, L.; DÖGE, J. S.; PROFES, M. S.; HOFMANN, P. R. P.; VALENTE, V. L. S. *Drosophilidae* (Diptera) associated to fungi: differential use of resources in anthropic and Atlantic Rain Forest areas. **Iheringia Ser. Zool.**, 99: 442-448, 2009.

GRAUR, D.; LI, W-H. Genes, genetic codes, and mutation. In: (2nd ed.) **Fundamentals of molecular evolution**. Sunderland: Sinauer Associates, Inc., Publishers, p. 5-38, 2000.

GUSTANI, E. C.; OLIVEIRA, A. P. F.; SANTOS, M. H.; MACHADO, L. P. B.; MATEUS, R. P. Demographic structure and evolutionary history of *Drosophila ornatifrons* (Diptera, Drosophilidae) from Atlantic Forest of Southern Brazil. **Zool. Sci.**, 32: 141-150, 2015.

HALE, M. L.; BURG, T. M.; STEEVES, T. E. Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. **PLOS One**, 7: e45170, 2012.

HAMMER, O.; HARPER, D. A. T.; RYAN, P. D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. **Paleontol. Electron.**, 4: 2001.

HANCOCK, J. M. Microsatellites and other simple sequences: genomic context and mutational mechanisms. **Microsatellites: Evolution and applications**, 1999.

KALINOWSKI, S. T. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity*, 94: 33–36, 2005.

KIMURA, M. Evolutionary rate at the molecular level. **Nature**, 217: 624-626, 1968.

KIMURA, M.; OHTA, T. Protein polymorphism as a phase of molecular evolution. **Nature**, 229: 467-469, 1971.

KING, J. L.; JUKES, T. H. Non-darwinian evolution. **Science**, 164: 788-798, 1969.

KLACZKO, L. B. Evolutionary genetics of *Drosophila mediopunctata*. **Genetica**, 126: 43–55, 2006.

KOPELMAN, N. M.; MAYZEL, J.; JAKOBSSON M.; ROSENBERG, N. N.; MAYROSE, I. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. **Mol. Ecol. Resour.**, 15: 1179–1191, 2015.

LABORDA, P. R.; KLACZKO, L. B.; SOUZA, A. P. *Drosophila mediopunctata* microsatellites II: cross-species amplification in the *tripunctata* group and other *Drosophila* species. **Conserv. Genet. Resour.**, 1: 281-296, 2009.

LABORDA, P. R.; MORI, G. M.; SOUZA, A. P. *Drosophila mediopunctata* microsatellites I: more than a hundred polymorphic loci available for genetic studies. **Conserv. Genet. Resour.**, 1: 297-307, 2009.

- LEVINSON, G.; GUTMAN, G. A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. **Mol. Biol. Evol.**, 4: 203-221, 1987a.
- LEVINSON, G.; GUTMAN, G. A. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. **Nucl. Acids Res.**, 15: 5323-5338, 1987b.
- LI, Y. C.; KOROL, A. B.; FAHIMA, T.; NEVO, E. Microsatellites within genes: structure, function, and evolution. **Mol. Biol. Evol.**, 21: 991-1007, 2004.
- LUNA, Y.; DEXING, Z. Effects of sample size on various genetic diversity measures in population genetic study with microsatellite DNA markers. *Dong wu xue bao*. [Acta zoologica Sinica], 50: 279-290, 2004.
- MACHADO, L. P. B.; MANFRIN, M. H.; SILVA-JUNIOR, W. A.; SENE, F. M. Microsatellite loci in the cactophilic species *Drosophila antonietae* (Diptera; Drosophilidae). **Mol. Ecol. Notes**, 3: 159–161, 2003.
- MACHADO, L. P. B.; MATEUS, R. P.; SENE, F. M.; MANFRIN, M. H. Microsatellite allele sequencing in population analyses of the South American cactophilic species *Drosophila antonietae* (Diptera: Drosophilidae). **Biol. J. Linn. Soc.**, 100: 573–584, 2010.
- MAGALHÃES, L. E. Notes on the taxonomy, morphology and distribution of *saltans* group of *Drosophila*, with description of four new species. **The Univ. Tex. Publi.**, 6205: 135-154, 1962.
- MANFRIN, M. H.; SENE, F. M. Cactophilic *Drosophila* in South America: a model for evolutionary studies. **Genetica**, 126: 57-75, 2006.
- MARKOW, T. A. Evolution of *Drosophila* mating systems. **Evol. Biol.**, 29: 73-106, 1996.
- MATEUS, R. P.; SENE, F. M. Temporal and spatial allozyme variation in the South American cactophilic *Drosophila antonietae* (Diptera, Drosophilidae). **Biochem. Genet.**, 41: 219-233, 2003.
- MATEUS, R. P.; SENE, F. M. Population genetic study of allozyme variation in natural populations of *Drosophila antonietae* (Insecta, Diptera). **J. Zoolog. Syst. Evol. Res**, 45: 136-143, 2007.
- MATEUS, R. P.; MACHADO, L. P. B.; MORAES, E. M.; SENE, F. M. Allozymatic divergence between border populations of two cryptic species of the *Drosophila buzzatii* cluster species (Diptera: Drosophilidae). **Biochem. Syst. Ecol.**, 38: 410-415, 2010.

- MATZKIN, L. M.; JOHNSON, S.; PAIGHT, C.; BOZINOVIC, G.; MARKOW, T. A. Dietary protein and sugar differentially affect development and metabolic pools in ecologically diverse *Drosophila*. **J. Nutr.**, 141: 1127-1133, 2011.
- MEDEIROS, H. F.; KLACZKO, L. B. How many species of *Drosophila* (Diptera, Drosophilidae) remain to be described in the forests of São Paulo, Brazil? Species lists of three forest remnants. **Biota Neotrop.**, 4: 1–12, 2004.
- MIYAMOTO, N.; FERNÁNDEZ-MANJARRÉS, J. F.; MORAND-PRIEUR, M.-E.; BERTOLINO, P.; FRASCARIA-LACOSTE, N. What sampling is needed for reliable estimations of genetic diversity in *Fraxinus excelsior* L. (Oleaceae)? **Ann. For. Sci.**, 65, 403, 2008.
- MOXON, E. R.; WILLS, C. DNA microsatellites: agents of evolution?. **Sci. Am.**, 280: 94-99, 1999.
- MULLIS, K.; FALOONA, F.; SCHARF, S.; SAIKI, R.; HORN, G.; ERLICH, H. (1986, January). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In: **Cold Spring Harbor symposia on quantitative biology**. Cold Spring Harbor Laboratory Press, 51: 263-273, 1986.
- O'GRADY, P. M.; DESALLE, R. Phylogeny of genus *Drosophila*. **Genetics**, 209: p. 1-25, 2018.
- PAIXÃO, J. F. **Diversidade genética de *Drosophila prosaltans* e *D. austrosaltans* (Diptera, Drosophilidae) de regiões de Mata Atlântica avaliada por marcadores microssatélites**. 2020. 88 p. Dissertação (Mestrado em Biociências) - Universidade Estadual Paulista, São José do Rio Preto, 2020.
- PEAKALL, R.; SMOUSE, P. E. GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. **Bioinformatics**, 28: 2537e2539, 2012.
- PEREIRA, M. A. Q. R.; VILELA, C. R.; SENE, F. M. Notes on breeding and feeding sites on some species of the *repleta* group of the genus *Drosophila* (Diptera, Drosophilidae). **Cienc. Cult.**, 35: 1313–1319, 1983.
- POWELL, J. R. **Progress and prospects in evolutionary biology: the *Drosophila* model**. Oxford: Oxford University Press, 1997.
- PRITCHARD, J. K.; WEN, X.; FALUSH, D. **Documentation for structure software. Version 2.3**. University of Chicago, Chicago, USA. 2010.
- PRUETT, C. L.; WINKER, K. The effects of sample size on population genetic diversity estimates in song sparrows *Melospiza melodia*. **J. Avian Biol.**, 39: 252–256, 2008.

ROMAN, B. E.; TRAVA, B. M.; MADI-RAVAZZI, L. Development and transferability of microsatellites markers for species of the *sultans* group of *Drosophila* (Diptera: Drosophilidae). **Gen. Mol. Res.**, 19: gmr18700, 2020.

ROSE, O.; FALUSH, D. A threshold size for microsatellite expansion. **Mol. Biol. Evol.**, 15: 613-615, 1998.

SANGER, F.; NICKLEN, S.; COULSON, A. R. DNA sequencing with chain-terminating inhibitors (DNA polymerase/nucleotide sequences/bacteriophage 4X174). **Proc. Natl. Acad. Sci.**, 74: 5463-5467, 1977.

SANTA-BRIGIDA, R.; WARTCHOW, F.; MEDEIROS, P. S.; GOTTSCHALK, M. S.; MARTINS, M. B.; CARVALHO, C. J. B. D. Mycophagous Drosophilidae (Diptera) guild and their hosts in the Brazilian Amazon. **Pap. Avulsos Zool.**, 59: e20195920, 2019.

SANTOS, C. H. **Variação Bionômica e Metabólica em três espécies de *Drosophila* submetidas a diferentes dietas** 2020. 112 p. Dissertação (Mestrado em Biologia Evolutiva) - Universidade Estadual do Centro-Oeste, Guarapuava, 2020.

SANTOS, C. H.; SANTOS K. A. V.; MACHADO, L. P. B.; MATEUS, R. P. (submitted). Life history traits and metabolic pool variation in Neotropical *Drosophila* species (Diptera, Drosophilidae). **Zool. Stud.**

SANTOS, W. F.; DE OLIVEIRA, M. S.; ROMEIRO, L. C. T. L.; ALISSON, M.; DA CUNHA, A. (2013, October). Sequenciamento de DNA: métodos e aplicações. In Proceedings of Safety, Health and Environment World Congress (Vol. 13, pp. 139-141).

SANTOS, K.A.V.; ZORZATO, S. V.; SALOMON, G. R.; MACHADO, L. P. B.; MATEUS, R. P. Microsatellites null alleles detection in species of *guarani* and *guaramuru* groups collected in Highland Araucaria Forest of Brazil. **Dros. Inf. Serv.**, 103: 43-47, 2020.

SILVA, D. C.; MACHADO, L. P. B.; MATEUS, R. P. Migration rate and genetic diversity of two *Drosophila maculifrons* (Duda, 1927) populations from Highland Araucaria Forest Fragments in Southern Brazil. **Braz. J. Biol.**, 75, 254-255, 2015.

SLATKIN, M. Rare alleles as indicators of gene flow. **Evolution**, 39: 53-65, 1985.

SREENU, V. B.; ALEVOOR, V.; NAGARAJU, J.; NAGARAJARAM, H. A. MICdb: database of prokaryotic microsatellites. **Nucleic Acids Res.**, 31: 106-108, 2003.

STRAND, M.; PROLLA, T. A.; LISKAY, R. M.; PETES, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. **Nature**, 365: 274-276, 1993.

- TRACTZ, C. C.; SALOMON, G. R.; ZORZATO, S. V.; MACHADO, L. P. B.; MATEUS, R. P. Allele diversity of cross-species microsatellite amplification on populations of *Drosophila guarani* species group from Araucaria Forest in Brazil. **Dros. Inf. Serv.**, 95: 76-79, 2012.
- TAMURA, K.; NEI, M.; KUMAR, S. Prospects for inferring very large phylogenies by using the neighbor-joining method. **Proc. Natl. Acad. Sci. U.S.A.**, 101: p. 11030-11035, 2004.
- TIDON, R. Relationships between drosophilids (Diptera, Drosophilidae) and the environment in two contrasting tropical vegetations. **Biol. J. Linn. Soc.**, 87: 233–247, 2006.
- TIDON-SKLORZ, R.; SENE, F. M. Evolution of the *buzzatii* cluster (*Drosophila repleta* group) in the Northeastern South America. **Evolución Biológica**, 8/9: 71-85, 1995.
- TIDON-SKLORZ, R.; SENE, F. M. *Drosophila*. In **Biodiversidade do Estado de São Paulo, Brasil: síntese do conhecimento ao final do século XX** (Brandão C.R.F. & E.M. Canello, eds). 5 ed. Invertebrados terrestres, São Paulo: FAPESP, 1999.
- TIDON-SKLORZ, R.; SENE, F. M. Two new species of the *Drosophila serido* sibling set (Diptera, Drosophilidae). **Iheringia. Ser. Zool.**, 90: 141-146, 2001.
- TRACTZ, C. C.; SALOMON, G. R.; ZORZATO, S. V.; MACHADO, L. P. B.; MATEUS, R. P. Allele diversity of cross-species microsatellite amplification on populations of *Drosophila guarani* species group from Araucaria Forest in Brazil. **Dros. Inf. Serv.**, 95: 76-79, 2012.
- TRAVA, B. M.; MATEUS, R. P.; MACHADO, L. P. B.; MADI-RAVAZZI, L. Moderate Population Structure in *Drosophila sturtevantii* from the South American Atlantic Forest Biome. **Zool. Stud.**, 60: 46, 2021.
- VALADÃO, H.; PROENCA, C. E.; KUHLMANN, M. P.; HARRIS, S. A.;
- WANG, H.; YANG, B.; WANG, H.; XIAO, H. Impact of different numbers of microsatellite markers on population genetic results using SLAF-seq data for *Rhododendron* species. **Sci. Rep.**, 11: 1-9, 2021.
- WATSON, J. D.; CRICK, F. H. C. Molecular structure of Nucleic Acids. **Nature**, 171: 737-738, 1953.
- WEBER, J. L.; WONG, C. Mutation of human short tandem repeats. **Hum. Mol. Genet.**, 2: 1123-1128, 1993.
- WRIGHT, J. M. Mutation at VNTRs: are minisatellites the evolutionary progeny of microsatellites? **Genome**, 37: 345-347, 1994.

ZCHONSKI, F. L. **Determinação do tamanho mínimo amostral para estudos genéticos populacionais em plantas utilizando marcadores microssatélites.** 2021. 59 p. Dissertação (Mestrado em Biologia Evolutiva) - Universidade Estadual do Centro-Oeste, Guarapuava, 2021.

ZORZATO, S. V. **Análise da estrutura genética populacional de *Drosophila ornatifrons* (Diptera; Drosophilidae).** 2015. 60 p. Dissertação (Mestrado em Biologia Evolutiva) - Universidade Estadual do Centro-Oeste, Guarapuava, 2015.

ZORZATO, S. V.; YASSIN, A.; MADI-RAVAZZI, L. Signature of climatic differentiation on mitochondrial DNA of *Drosophila sturtevantii*. **Mitochondrial DNA Part A**, 1-9, 2022.