



**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

**FELIPE LISS ZCHONSKI**

**Determinação do tamanho mínimo amostral para estudos genéticos populacionais em plantas, utilizando marcadores microssatélites**

**Guarapuava**

**2021**

**FELIPE LISS ZCHONSKI**

**Determinação do tamanho mínimo amostral para estudos genéticos populacionais em plantas, utilizando marcadores microssatélites**

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).

**Orientador**

Prof. Dr. Paulo Roberto Da Silva

**Guarapuava**

**2021**

Catálogo na Publicação

Rede de Bibliotecas da Unicentro

Z39d Zchonski, Felipe Liss  
Determinação do tamanho mínimo amostral para estudos genéticos populacionais em plantas utilizando marcadores microssatélites / Felipe Liss Zchonski. -- Guarapuava, 2021.  
viii, 59 f. : il. ; 28 cm

Dissertação (mestrado) - Universidade Estadual do Centro-Oeste em Associação Ampla com a Universidade Estadual de Ponta Grossa, Programa de Pós-Graduação em Biologia Evolutiva. Área de concentração: Biologia Evolutiva, 2021.

Orientador: Paulo Roberto Da Silva  
Banca examinadora: Claudete Aparecida Mangolim, Luciana Paes de Barros Machado, André Luiz Gaglioti

Bibliografia

1. Tamanho amostral. 2. SSR. 3. Microssatélites. 4. Genética de populações. I. Título. II. Programa de Pós-Graduação em Biologia Evolutiva.

CDD 576.2

*Dedico à Ivete Liss, minha amiga, minha mãe,  
pelo apoio incondicional nas minhas decisões.*

## AGRADECIMENTOS

Agradeço primeiramente a Deus por me conceder saúde, força e disposição para a realização deste trabalho.

Agradeço a minha mãe, Ivete Liss pelo incentivo, apoio e amor incondicional.

Agradeço ao professor Dr. Paulo Roberto Da Silva pela orientação, paciência, dedicação e por ter dado as condições necessárias para a realização deste trabalho. Agradeço também pelo convívio diário que, sem dúvida, foi importante no meu aprendizado. Obrigado por ser um excelente professor orientador e um grande amigo.

À UNICENTRO, ao Programa de Pós-Graduação em Biologia Evolutiva e ao Laboratório de Genética Molecular. Agradeço a oportunidade dada, desde minha graduação.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelas bolsas que recebi no decorrer do curso de mestrado em Biologia Evolutiva.

Agradeço a minhas colegas, Barbara Mainardes Araujo, Bruna Santos Silva e Pamela Estyse Campos Bini pela ajuda nas atividades do projeto.

Agradeço aos ex-alunos de Mestrado do programa, Jesus Alberto Diaz Cruz, Laura Pilati e Ricardo Micolino que produziram parte dos dados utilizados neste trabalho.

A todos que colaboraram direta ou indiretamente com este trabalho, meu muito obrigado.

*"Uma das regras básicas do universo é que nada é perfeito. A perfeição simplesmente não existe. Sem imperfeição, nem você nem eu existiríamos."*

*Stephen Hawking*

## RESUMO

Ao iniciar um estudo genético populacional, o pesquisador primeiramente deve decidir o número de indivíduos a ser usado para cada população. Para animais, essa questão parece estar esclarecida. No entanto, para plantas, estudos com dados reais para determinar a influência do tamanho amostral nos índices genéticos são limitados, estando quase que restritos a gimnosperma. Para auxiliar na resposta a esta questão, neste estudo foram avaliadas quatro espécies de angiosperma, sendo uma árvore de grande porte (*Campomanesia xanthocarpa* O. Berg), uma árvore de médio porte (*Cecropia pachystachya* Trécul), uma árvore de pequeno porte (*Psidium guajava* L.) e um subarbusto (*Baccharis crispa* Spreng.). A partir de um conjunto de dados reais de 100 indivíduos de cada espécie, obtidos com marcadores microssatélites, foram realizadas por sorteio 50 repetições para os tamanhos amostrais de 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 e 95 indivíduos, totalizando 2400 subpopulações (600 para cada espécie). Para cada subpopulação foram calculados o número de alelos ( $N_a$ ), a heterozigosidade esperada ( $H_E$ ), a frequência alélica e coeficiente de diferenciação genética ( $F_{ST}$ ). Com as análises realizadas, foi determinada a amplitude dos dados, a mediana, e os quartis. Os resultados obtidos, evidenciaram que, para o  $N_a$ , o tamanho amostral de 45 foi capaz de detectar 90% dos alelos em *C. xanthocarpa* e *B. crispa*, e 35 indivíduos em *C. pachystachya* e *P. guajava*. Para a  $H_E$ , o tamanho amostral de 30 indivíduos é suficiente para detecção das estimativas genéticas para três das quatro espécies estudadas. O tamanho amostral de 40 indivíduos é suficiente para a detecção em todas as espécies dos alelos que apresentaram frequências acima de 0,05 na população original. Já, os alelos raros não são passíveis de detecção em todas as subpopulações em nenhum tamanho amostral. Para o  $F_{ST}$  a partir do tamanho amostral de 10 indivíduos, a diferença entre a população original e as subpopulações foi menor que 0,05 em *C. xanthocarpa* e *C. pachystachya*; e 15 para *P. guajava* e *B. crispa*. Na análise de alelos raros, não foram encontradas diferenças nos índices genéticos e na estruturação entre populações com e sem alelos raros. A análise geral dos dados evidencia que, a partir do tamanho amostral de 30 indivíduos, os dados podem ser considerados confiáveis, pois a maioria dos índices já apresenta valores próximos aos observados na população original. Estudos com número de indivíduos menores que 30, somente devem ser realizados quando o tamanho da população não permite atingir a amostragem ideal.

**Palavras-chave:** Tamanho amostral, SSR, microssatélites, genética de populações.

## ABSTRACT

When starting a population genetic study, the researcher must first decide the number of individuals to be used for each population. For animals, this question seems to be clear. However, for plants, studies with real data to determine the influence of sample size on genetic indices are limited, being almost limited to Gymnosperm. To help answer this question, four species of Angiosperm were evaluated in this study: a large tree (*Campomanesia xanthocarpa* O. Berg), a medium tree (*Cecropia pachystachya* Trécul), a small tree (*Psidium guajava* L.) and an subshrub (*Baccharis crispa* Spreng.). From a set of real data of 100 individuals of each species, obtained with microsatellite markers, 50 repetitions were obtained by random for the sample sizes of 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 and 95 individuals, totaling 2400 subpopulations (600 for each species). For each subpopulation, the number of alleles ( $N_a$ ), expected heterozygosity ( $H_E$ ), allelic frequency and coefficient of genetic differentiation ( $F_{ST}$ ) were calculated. With the data obtained, the range of data, the median, and the quartiles were determined. The results showed that, for  $N_a$ , the sample size of 45 was able to detect 90% of alleles in *C. xanthocarpa* and *B. crispa*, and 35 individuals in *C. pachystachya* and *P. guajava*. For  $H_E$ , the sample size of 30 individuals is sufficient to detect genetic estimates for three of the four species studied. The sample size of 40 individuals is sufficient for detection in all species of alleles that presented frequencies above 0.05 in the original population. On the other hand, rare alleles are not detectable in all subpopulations in any sample size. For the  $F_{ST}$  from the sample size of 10 individuals, the difference between the original population and the subpopulations was less than 0.05 in *C. xanthocarpa* and *C. pachystachya*; and 15 for *P. guajava* and *B. crispa*. In the analysis of rare alleles, no differences were found in genetic indices and structure between populations of 100 and 30 individuals with and without rare alleles. The overall analysis of the data shows that, from the sample size of 30 individuals, the data can already be considered reliable, as most indices already show values close to those observed in the original population. Studies with a number of individuals smaller than 30 should only be carried out when the population size does not allow reaching the ideal sample size.

**Keywords:** Sample size, SSR, microsatellites, population genetics



## SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>10</b>
<b>2. REFERENCIAL TEÓRICO .....</b>	<b>12</b>
2.1. Genética de populações .....	12
2.2. Marcadores microssatélites .....	14
2.3. Tamanho amostral .....	15
2.4. Alelos raros na genética de populações.....	18
<b>3. OBJETIVOS .....</b>	<b>20</b>
3.1. Objetivo geral .....	20
3.2. Objetivos específicos .....	20
<b>4. MATERIAL E MÉTODOS .....</b>	<b>21</b>
4.1. Tamanho amostral .....	21
4.1.1. Obtenção dos dados .....	21
4.1.2. Análise dos dados .....	24
4.2. Alelos raros.....	25
4.2.1. Obtenção dos dados .....	25
4.2.2. Análise dos dados .....	26
<b>5. RESULTADOS .....</b>	<b>27</b>
<b>5.1. Tamanho amostral .....</b>	<b>27</b>
5.1.1. <i>Campomanesia xanthocarpa</i> .....	27
5.1.2. <i>Cecropia pachystachya</i> .....	30
5.1.3. <i>Psidium guajava</i> .....	34
5.1.4. <i>Baccharis crispa</i> .....	38
<b>5.2. Alelos raros.....</b>	<b>44</b>
5.2.1. <i>Campomanesia xanthocarpa</i> .....	44
5.2.2. <i>Psidium guajava</i> .....	46
<b>6. DISCUSSÃO .....</b>	<b>50</b>
<b>7. REFERÊNCIAS .....</b>	<b>53</b>

## 1. INTRODUÇÃO

Determinar o número amostral mínimo a ser coletado em uma população é uma questão crítica para a obtenção de dados genéticos populacionais confiáveis. Até o momento, não há um consenso no número mínimo amostral para ser utilizado em estudos de genética populacional de plantas. Neste sentido, diferentes estudos utilizaram tamanhos amostrais para a obtenção das estimativas genéticas (MANGARAVITE *et al.*, 2016; WANG *et al.*, 2016; GUO *et al.*, 2019; ISLAM *et al.*, 2019). Portanto, uma das tarefas mais difíceis ao iniciar em estudo genético populacional é definir quantos indivíduos são necessários, por população, para que as estimativas genéticas tenham suas frequências com maior precisão em relação a população total.

Estabelecer o número amostral mínimo ideal para estudos em genética de populações em plantas com marcadores microssatélites é uma tarefa difícil. No entanto, alguns estudos tentaram responder a esta questão, porém estes apresentam problemas metodológicos e/ou de perguntas a serem respondidas. Por exemplo, o trabalho de Danusevicius; Kavaliauskas e Fussi (2016), utilizando pinheiros silvestres [*Pinus sylvestris* L. (Pinaceae)], os autores concluem que de 20 a 25 indivíduos é uma margem segura para amostrar alelos microssatélites com frequência acima de 0,05. Este estudo deve ser avaliado com cautela, uma vez que os autores amostraram 400 indivíduos em uma área de apenas um hectare, o que pode ter levado a coleta de organismos muito próximos fisicamente. A coleta de indivíduos próximos tende a levar a erros de precisão nas estimativas genéticas devido ao elevado grau de parentesco entre as plantas. Segundo Gandara *et al.* (2019), para evitar o problema de parentesco elevado em espécies arbóreas, uma distância mínima de 50 metros entre os indivíduos deve ser respeitada. Em outro estudo, Gapare; Yanchuk e Aitken (2008), trabalhando com *Picea sitchensis* (Bong.) Carrière (Pinaceae), afirmaram que tamanhos amostrais mínimos de 150 indivíduos são suficientes para capturar 95% da diversidade genética de populações não estruturadas. Esta conclusão é inviável financeiramente, pois o tamanho amostral é muito elevado. Já, com valores intermediários, Bashalkhanov; Pandey e Rajora (2009) utilizando dados simulados, concluíram que o tamanho amostral mínimo, para detectar todos os alelos de uma população, deve ser maior que 60 indivíduos. Hale; Burg e Steeves (2012) pontuaram que em vez de perguntar quantos indivíduos são necessários para detectar todos os alelos presentes em uma população, a estratégia mais útil em estudos

genéticos populacionais é considerar quantos indivíduos são necessários para detectar os alelos informativos e em frequências representativas da população original.

Os trabalhos citados têm pouca aplicação na genética de populações, seja por possíveis tendenciosidade de amostragem que podem subestimar ou superestimar o tamanho amostral, situação provavelmente presente no trabalho de Danusevicius; Kavaliauskas e Fussi (2016) e Gapare; Yanchuk e Aitken (2008), respectivamente, ou por possíveis equívocos na pergunta que deve ser feita quando se pretende realizar estudos de genética de populações, esta situação pode ser observada no trabalho de Bashalkhanov; Pandey e Rajora (2009). Neste sentido, para auxiliar na determinação do tamanho mínimo amostral em plantas, devem ser realizados trabalhos utilizando dados reais, com rigor metodológico e com análises de índices informativos para genética de populações.

Para algumas espécies animais, o tamanho amostral mínimo já foi determinado. Hale; Burg e Steeves (2012), analisaram dados reais obtidos com marcadores microssatélites em quatro espécies de animais, duas aves, um mamífero e um inseto. Nesse estudo, dados de frequências alélicas, da heterozigosidade esperada e das distâncias genéticas entre as populações das espécies avaliadas foram obtidos por subamostragem aleatória de populações de 5 a 100 indivíduos a partir de dados reais de uma população original. Os resultados foram semelhantes nas quatro espécies estudadas, e os autores concluíram que de 25 a 30 indivíduos são suficientes para realizar estudos de genética de populações baseados em marcadores microssatélites. Para plantas, conforme apontado anteriormente, os estudos realizados chegaram a conclusões muito discrepantes, com tamanhos amostrais variando de 25 a 150 indivíduos. Nesse sentido, a realização de análises utilizando dados reais, obtidos com marcadores microssatélites, é fundamental para definir o número amostral mínimo confiável para estudos de genética de populações em plantas.

## **2. REFERENCIAL TEÓRICO**

### **2.1. Genética de populações**

A genética de populações é o estudo da composição genotípica das populações, incluindo distribuições e mudanças no genótipo e na frequência genotípica, em resposta a processos evolutivos como seleção natural, deriva genética, mutação e fluxo gênico (GILLESPIE, 1998; JOHNSON, *et al.*, 2018; NOSIL, *et al.*, 2018). As aplicações da genética de populações são extensas, com interesses voltados para medicina, agricultura e conservação de espécies (HARTL; CLARK, 1997). Para espécies nativas, os estudos genéticos populacionais buscam compreender os processos evolutivos que envolvem as populações para ao entendimento de sua história evolutiva bem como para inferir estratégias de conservação das mesmas (CRUZ *et al.*, 2011; KIANI; MEMARIANI; ZARGHAMI, 2012; HUANG *et al.*, 2014).

A genética de populações aborda a base genética da evolução, especialmente as frequências e valores adaptativos dos genótipos ao longo do tempo. A escala de mudança para a maioria dessas variantes genéticas que ocorrem naturalmente é muito longa, variando de dezenas de milhares a milhões de anos. Essas mudanças são impossíveis de observar diretamente (GILLESPIE, 1998). Portanto, entender a base genética das populações e como os fatores de seleção podem influenciar na evolução, pode ser útil em estratégias de conservação das populações estudadas (BOSSÉ *et al.*, 2017).

As mutações são as principais fontes de variabilidade genética na formação de novos alelos, é a variação sobre a qual a seleção natural pode atuar e são as causas de muitas novidades que ocorrem na evolução. Uma pequena porcentagem das mutações é benéfica, algumas são neutras, entretanto, a maioria das mutações diminui o valor adaptativo dos indivíduos e são eliminadas por ação da seleção natural (KIMURA, 1968; KING; JUKES, 1969). A capacidade de produzir variações no DNA faz das mutações grandes precursoras da variabilidade genética, o que pode resultar em diferentes características nos indivíduos. Essas características podem surgir em uma população e estar ausente em outra, resultando na diferenciação das mesmas (MCDERMOTT; MCDONALD, 1993). Da mesma maneira que as mutações, a deriva genética ocorre aleatoriamente, mas pode levar ao aumento ou diminuição na frequência dos alelos, ou mesmo a perda de alguns deles em uma população.

A deriva genética é uma mudança aleatória nas frequências dos alelos ao longo das gerações. O processo de deriva genética torna-se cada vez mais importante quando o tamanho da população diminui, podendo levar a diferentes frequências alélicas ou à fixação de diferentes alelos em populações isoladas e, assim, reduzir a variabilidade genética (GILLESPIE, 1998). As chances de um alelo se tornar mais ou menos comum dentro das populações são igualmente prováveis, e seu efeito é maior para aqueles alelos presentes em baixa frequência (MCDERMOTT; MCDONALD, 1993; WAHL, 2011). A frequência dos alelos também pode ser alterada por meio do fluxo gênico, com a inserção de novos alelos em uma população.

O fluxo gênico é um mecanismo de movimento de genes e/ou alelos entre populações. Esse mecanismo resulta na homogeneidade genética das populações, tornando determinados genes comuns entre populações de diferentes locais (MAYR, 1963; SLARKIN, 1985; SLATKIN, 1987). Devido a tendência à homogeneização, o fluxo gênico é frequentemente visto como uma força restritiva na diferenciação genética. A seleção natural tende a adaptar a população a condições ambientais locais, porém o fluxo gênico pode impedi-la de se diferenciar e, conseqüentemente, evoluir para diferentes espécies (SLATKIN, 1987). Quando as populações estão muito distantes geograficamente, o isolamento por distância pode afetar o fluxo gênico e, conseqüentemente, as frequências gênicas entre as populações. O isolamento por distância geográfica e fluxo gênico restrito gera estruturação genética, devido à deriva genética ser aleatória e ocorrer localmente (HARTL; CLARK, 1997). Ainda, além da deriva genética, o fluxo gênico pode interferir na diferenciação promovida pela seleção natural.

A seleção natural é um fator evolutivo que pode levar diferenciação das populações, principalmente pela seleção dos genótipos mais adaptados a uma determinada condição ecológica e eliminar os menos aptos, moldando-a de acordo com as condições locais (LEWONTIN, 1970; HURST, 2009). Este processo pode levar a diferenciação genotípica da população. Quase sempre, isso se deve a recursos limitados, levando organismos com características que favorecem a sobrevivência e reprodução a deixarem mais descendentes. Neste sentido, as características presentes nestes organismos se tornam mais comuns nas próximas gerações. As atividades antrópicas têm alterado os habitats naturais diminuindo o tamanho das populações e, conseqüentemente têm influenciando na sobrevivência e na reprodução das espécies. Populações pequenas e altamente homozigotas, característica comum em áreas fragmentadas, limitam a ação da seleção natural, podendo levar a extinção

da população (TOWNSEND; BEGON; HARPER, 2010; BOSSE *et al.*, 2017; JHONSON *et al.*, 2018).

Atualmente, os pesquisadores da área de genética de populações não só se preocupam com o estudo da história evolutiva das espécies, mas também em obter dados comparativos entre populações em áreas conservadas e degradadas. Estes dados são úteis para o entendimento dos efeitos das ações antrópicas nas populações e na tomada de decisões para conservação. A técnica mais utilizada para obtenção dos dados genéticos destas populações é o uso de marcadores moleculares, dentre esses os microssatélites.

## 2.2. Marcadores microssatélites

Os marcadores microssatélites ou SSR (*Simple Sequence Repeats*), são sequências de DNA repetidos *in tandem* (de 2 a 5 nucleotídeos), localizados ao longo dos genomas dos organismos e com número de repetições altamente variável (LITT; LUTY, 1989; TAUTZ, 1989; WEBER; MAY, 1989). O alto nível de polimorfismo nesses marcadores os torna ferramentas poderosas para avaliar a similaridade genética entre indivíduos (GUICHOUX *et al.*, 2011). Esses marcadores são usados em variadas aplicações. Por exemplo, o mapeamento de genomas, análise forense, análise de parentesco, genética de conservação, identificação de parentesco de poliploides, filogeografia e genética de populações (ESSELINK; NYBOM; VOSMAN, 2004; FALEIRO, 2007; KALIA *et al.*, 2011).

As principais vantagens dos marcadores microssatélites são a codominância, que possibilita a visualização dos *loci* em heterozigose, a alta reprodutibilidade e o elevado polimorfismo (FALEIRO, 2007). Comparado aos marcadores dominantes, a principal desvantagem desses marcadores está relacionada ao custo para o desenvolvimento de *primers* específicos, quando eles não estão disponíveis para a espécie a ser estudada. Neste caso, é necessária a construção de bibliotecas enriquecidas com sequências microssatélites e sequenciamento para o desenho dos *primers* (LITT; LUTY, 1989; FALEIRO, 2007). No entanto, muitos *primers* microssatélites podem ser usados em espécies geneticamente relacionadas, permitindo estudos multiespécies (FALEIRO, 2007; DA-SILVA; MILACH; TISIAN, 2011; FAGUNDES *et al.*, 2016; HODEL *et al.*, 2016). Ainda, mesmo com custo maior que os marcadores dominantes, o desenvolvimento dos *primers* microssatélites, quando comparado ao alto custo das técnicas de sequenciamento de última geração (SNP), é vantajoso. Isso faz com que os microssatélites sejam, atualmente, o marcador mais utilizado

em estudos de genética de populações que requerem a identificação dos heterozigotos e altos níveis de polimorfismo (HODEL *et al.*, 2016).

O uso de marcadores microssatélites está consolidado como altamente confiável em genética de populações (KALIA *et al.*, 2011; VIEIRA *et al.*, 2016). Porém, o simples uso dos marcadores microssatélites não garante a determinação de estimativas genéticas confiáveis em genética de populações. Isso é devido ao tamanho amostral ser um fator crucial que pode comprometer os resultados obtidos. Para a obtenção de estimativas genéticas confiáveis, além da escolha do marcador robusto, é necessário tamanhos amostrais que, apesar de serem uma fração da população em estudo, traduzam o mais próximo possível o *status* genético da população como um todo.

Mesmo com o advento de técnicas mais avançadas de biologia molecular para detecção de polimorfismos, os marcadores microssatélites têm sido amplamente utilizados (DAVEY *et al.*, 2011; HODEL *et al.*, 2016). Até 2016, foram publicados aproximadamente 225 mil artigos científicos utilizando marcadores microssatélites (HODEL *et al.*, 2016). Ainda, grande parte destes estudos foram realizados em espécies vegetais. Neste sentido, determinar um número confiável de indivíduos a ser amostrado capaz de representar as frequências genéticas da população total utilizando marcadores microssatélites, é um conhecimento de grande valor no desenvolvimento de estudos genéticos populacionais em plantas.

### **2.3. Tamanho amostral**

Em genética de populações, o tamanho amostral é um subconjunto de um determinado grupo de indivíduos de uma população, coletado para fazer inferências ou extrapolações da amostra para a população (FALEIRO, 2007). A maioria dos pesquisadores acreditam que estender o esforço amostral ao máximo é a melhor estratégia para que as estimativas cheguem o mais próximo possível da população real. No entanto, quando uma espécie possui ampla distribuição e com populações grandes, coletar muitos indivíduos de uma população é uma tarefa pouco prática e onerosa economicamente (SPIEGEL *et al.*, 2004). Por outro lado, pequenas amostragens podem levar a erros significativos nas estimativas de diversidade genética. Esta situação não se aplica quando a população é muito pequena e todos os indivíduos são amostrados (BASHALKHANOV; PANDEY; RAJORA, 2009). Neste sentido,

estabelecer um número amostral mínimo que estime dados confiáveis de uma população é de grande valia para estudos genéticos populacionais.

A determinação do número mínimo amostral para estudos genéticos populacionais é uma tarefa difícil e não está disponível para todos os grupos de organismos. Porém, para algumas espécies foram realizados trabalhos com o objetivo de resolver essa questão. Merece destaque o trabalho de Hale; Burg e Steeves (2012) que, com base em dados reais, determinaram que o tamanho amostral de 25 a 30 indivíduos é suficiente para determinação de dados genéticos populacionais em algumas espécies animais utilizando marcadores microssatélites. Para dados reais de marcadores microssatélites em plantas, os poucos trabalhos realizados foram majoritariamente com gimnospermas (GAPARE; YANCHUK; AITKEN, 2008; BASHALKHANOV; PANDEY; RAJORA, 2009; DANUSEVICIUS; KAVALIAUSKAS; FUSSI, 2016). Nesses estudos, ao depender do índice genético avaliado, o tamanho amostral mínimo variou de 25 a 150 indivíduos.

Para angiospermas, os dois trabalhos encontrados são de espécies que apresentam características biológicas bem diferentes, uma árvore de grande porte e uma erva. Miyamoto *et al.* (2008), trabalhando com dados reais de marcadores microssatélites obtidos de populações de *Fraxinus excelsior* L. (Oleaceae) evidenciaram que, dependendo da estimativa genética, o tamanho amostral mínimo pode variar de 3 a 386 indivíduos para confiabilidade de 95%. Ainda, esses autores são reticentes em extrapolar estes dados para outras espécies. Segundo os autores, os dados podem ser aplicados somente para espécies com polinização exclusiva pelo vento, com capacidade de dispersão de pólen e sementes a longas distâncias, ausência total de autopolinização (características que favorecem alta variabilidade genética). Já, no estudo de Herrmann; Flajoulot e Julier (2010), com a leguminosa *Medicago sativa* L. (Fabaceae), a avaliação das populações de três tamanhos amostrais (10, 20, 40), evidenciou que o tamanho amostral de 40 indivíduos é suficiente para estudos de diversidade genética em *M. Sativa*, com marcadores microssatélites. No entanto, para estudos em grande escala, o uso de 20 indivíduos pode ser uma alternativa econômica, e ainda assim gerar dados confiáveis de diversidade genética nesta espécie de alfafa, cuja polinização é entomófila, realizada principalmente por abelhas. A discrepância nos dados apresentados nestes dois trabalhos, e considerando que aproximadamente 78% da flora temperada e 94% da flora tropical (incluindo espécies arbóreas) é polinizada por animais (OLLERTON; WINFREE;



TARRANT, 2011), torna-se necessário estudos de tamanho mínimo amostral neste dominante grupo de angiospermas.

Para marcadores ISSR, dados produzidos em nosso grupo de estudo, e que estão sendo submetidos para publicação, foi demonstrado que tamanhos amostrais de 30 indivíduos são suficientes para estimar de forma confiável, o índice de Shannon (I), a diversidade genética de Nei (h), a distância genética e o índice de diferenciação genética ( $G_{ST}$ ) em angiospermas com predomínio de polinização por animais. Já, para marcadores moleculares SNPs (*Single Nucleotide Polymorphism*), também com espécie de polinização zoocórica [*Amphirrhox longifolia* (A. St.-Hil.) Spreng. (Violaceae)], um tamanho amostral maior que oito indivíduos têm pouco impacto nas estimativas de diversidade genética das populações (NAZARENO *et al.*, 2017).

É importante pontuar que alguns estudos, que não definem espécies, usam técnicas de rarefação, abordagens de regressão e amostragem aleatória repetida para determinar tamanhos amostrais representativos nas estimativas dos dados genéticos (KALINOWSKI, 2005; (BASHALKHANOV; PANDEY; RAJORA, 2009; MARANDEL *et al.*, 2020). Esses trabalhos além de utilizar dados simulados, que dificilmente representam os dados reais de uma população, utilizam técnicas que estimam a necessidade de amostrar todos, ou quase todos, os alelos de uma população, incluindo os considerados raros. Esta abordagem é inviável no dia-dia, pois, necessitam de números exageradamente grandes de tamanhos amostrais (acima de 150), enquanto que os recursos para pesquisa são limitados. Ainda, segundo Hale; Burg e Steeves (2012) em vez de amostrar todos os alelos, uma estratégia mais útil na genética de populações é amostrar os alelos mais representativos da população, desconsiderando os alelos raros. Isso pode ser feito sem necessidade de amostrar centenas de indivíduos, diminuindo o tempo e os custos do estudo.

Frequentemente, em estudos de genética de populações, o aumento do tamanho amostral resulta no aumento da precisão das estimativas genéticas. No entanto, a taxa na qual a precisão aumenta deve se nivelar conforme o aumento do esforço de amostragem, enquanto o custo para genotipar mais indivíduos não aumenta (HALE; BURG; STEEVES, 2012).

O desequilíbrio desta equação, principalmente relacionado ao aumento do custo ou esforço e não aumento da precisão, não justifica o acréscimo no tamanho amostral. Esta situação ficou clara no trabalho de Hale; Burg e Steeves (2012) com tamanhos amostrais acima de 30 indivíduos. Segundo estes autores, com este tamanho amostral, os alelos

representativos das populações são adequadamente amostrados. Neste sentido, ao definir o tamanho amostral de um estudo de genética de populações deve estar claro os objetivos, pois no caso da necessidade de amostrar os alelos raros em sua totalidade, o número amostral deve ser elevado. Porém, os alelos raros parecem não serem importantes quando o interesse é o entendimento da estrutura e diversidade genética de espécies, mas podem ser importantes quando o objetivo é selecionar populações para conservação.

#### **2.4. Alelos raros na genética de populações**

Os alelos podem ser classificados de acordo com suas frequências relativas, que podem ser comuns ou raros. Para Kimura (1983), alelos raros são aqueles com frequências menores que 0,01. Porém, em outros estudos, são considerados raros, aqueles alelos que são encontrados em frequências inferiores a 0,05 (EANES; KOEHN, 1977; HALE; BURG; STEEVES, 2012).

No genoma das plantas, os alelos raros são distribuídos aleatoriamente em grupos, podendo ocorrer em qualquer *locus* (MEMON *et al.*, 2016, BROWN; KELLY, 2020). Sua origem pode ser por mutações aleatórias ou pela migração de indivíduos de outras populações. Ainda, os alelos raros que surgem por meio de mutações ou por fluxo gênico podem desaparecer de uma população devido a mutações deletérias, que diminuam o valor adaptativo de alguns indivíduos (MEMON *et al.*, 2016).

Os alelos raros são úteis em análises e cálculos de parentesco (HALE; BURG; STEEVES, 2012); além de poderem fornecer informações sobre os processos evolutivos que envolvem a relação das populações com o meio ambiente e também, possibilitar às populações uma adaptação a futuras mudanças em seu habitat (MEYERS; BULL, 2002). Um exemplo da vantagem evolutiva fornecida por alelos raros é a chamada ‘seleção dependente da frequência’.

A hipótese da ‘seleção dependente de frequência’ também é descrita como hipótese da vantagem do alelo raro ou a hipótese da Rainha Vermelha (VAN VALEN, 1973; LADLE, 1992; EBERT; HAMILTON, 1996; SOMMER, 2005). Ocorre quando um alelo ou genótipo é favorecido, em função de outros, ocorrendo alteração na frequência destes alelos, porém mantendo a variabilidade genética (AYALA; CAMPBELL, 1974; LEVIN, 1988; HEINO; METZ; KAITALA, 1998). Em alguns casos, a seleção dependente da frequência pode ser negativa, onde os genótipos raros são favorecidos em relação aos comuns (NOVIČIĆ *et al.*,

2020). Por exemplo, os alelos raros do Complexo Principal de Histocompatibilidade (MHC) que dão ao hospedeiro uma vantagem quando parasitado, são selecionados positivamente e tornam-se comuns nas populações. Por outro lado, isso aumenta a pressão de seleção do parasita para evitar o reconhecimento por esses alelos que se tornaram comuns. No entanto, à medida que a antigenicidade do parasita muda, o valor adaptativo dos genótipos comuns do hospedeiro diminui e fornece uma vantagem seletiva para outros alelos raros (SOMMER, 2005). A seleção dependente de frequência negativa também foi proposta como o mecanismo pelo qual as frequências dos *loci* podem se restaurar, mesmo após a aplicação de vacinas (CORANDER *et al.*, 2017).

A identificação de alelos raros pode ser um importante indicador de fluxo gênico em populações naturais (SLATKIN, 1985). Uma vez que esses alelos são identificados, eles podem ser vantajosos, principalmente na genética de conservação, a fim de garantir a manutenção da variabilidade genética das populações (SCHROEDER, *et al.*, 2007; SZPIECH *et al.*, 2008; SZPIECH; ROSENBERG, 2011). No entanto, de acordo com Hale; Burg e Steeves (2012), os alelos raros contribuem pouco para estudos em genética de populações, quando o objetivo é a descrição dos índices de diversidade e o estabelecimento da estruturação das populações. Para esse tipo de estudo, em animais são necessários alelos informativos, ou seja, aqueles encontrados em frequências maiores que 0,05 (HALE; BURG; STEEVES, 2012). Diante disso, são necessárias análises da contribuição de alelos raros nas estimativas genéticas em plantas. Para mais, é preciso comparar os índices genéticos obtidos em populações com a presença de alelos raros com aqueles em que estes alelos estão ausentes.

### **3. OBJETIVOS**

#### **3.1. Objetivo geral**

Determinar o número amostral mínimo para obtenção de dados genéticos populacionais confiáveis em plantas utilizando marcadores microsatélites.

#### **3.2 Objetivos específicos**

- Determinar o número de alelos ( $N_a$ ), a Heterozigosidade esperada ( $H_E$ ), as frequências alélicas e o coeficiente de diferenciação genética ( $F_{ST}$ ), em populações de tamanho amostral variando de 5 a 95 indivíduos para as espécies estudadas;
- Estimar o tamanho amostral ideal para obtenção de dados confiáveis para número de alelos ( $N_a$ ), a Heterozigosidade esperada ( $H_E$ ), as frequências alélicas e o coeficiente de diferenciação genética ( $F_{ST}$ ) nas espécies estudadas;
- Analisar os efeitos da presença dos alelos raros em parâmetros genéticos populacionais de *C. xanthocarpa* e *P. guajava*.

## 4. MATERIAL E MÉTODOS

### 4.1. Tamanho amostral

#### 4.1.1. Obtenção dos dados

Neste estudo foram avaliadas quatro espécies de angiospermas: Uma árvore monoica de grande porte, conhecida como guabirobeira (*Campomanesia xanthocarpa* O. Berg) (LIMA; GOLDENBERG; SOBRAL, 2011), uma árvore dióica de médio porte, conhecida como embaúba, (*Cecropia pachystachya* Trécul) (ROMANIUC NETO *et al.*, 2009), uma árvore monoica de pequeno porte, conhecida como goiabeira (*Psidium guajava* L.) (SURESH *et al.*, 2008), e um subarbusto dióico, conhecido como carqueja (*Baccharis crispa* Spreng.) (CORTADI *et al.*, 1999). Os 100 indivíduos das espécies *C. xanthocarpa* e *B. crispa* foram amostrados ao longo da Mata Atlântica, da região Sul do Brasil. Já, os indivíduos das espécies *C. pachystachya* e *P. guajava*, foram amostrados ao longo de toda a Mata Atlântica.

Os dados genotípicos de *C. xanthocarpa* foram obtidos neste estudo. Já, os dados das demais espécies foram obtidos em trabalhos anteriores realizados pelo grupo de pesquisa do Laboratório de Genética Molecular (DNA Lab) da UNICENTRO.

Para obtenção dos dados da espécie *C. xanthocarpa*, foram coletadas folhas jovens que foram mantidas em sílica gel até a extração do DNA. O DNA de cada indivíduo foi extraído de acordo com o protocolo proposto por Sharma; Mishra e Misra (2008). Para a genotipagem, o DNA dos indivíduos foram amplificados individualmente por PCR, utilizando os pares de *primers* microssatélites desenvolvidos para outras Myrtaceae e transferidos para a espécie por Fagundes *et al.* (2016) (Tabela 1). Os produtos das ampliações foram separados por eletroforese em gel de poliacrilamida a 12%, com migração inicial de 40 min a 500 V 55 MA 5 W para empilhamento das amostras, seguido de migração por 4 horas a 2000 V 55 MA 55 W. A revelação das amostras no gel foi realizada por coloração com nitrato de prata de acordo com o protocolo descrito por Sanguinetti; Dias Neto e Simpson, 1994.

**Tabela 1.** Relação dos sete pares de *primers* microssatélites utilizados para a obtenção de dados para a avaliação do tamanho amostral em *Campomanesia xanthocarpa* O. Berg. TA: temperatura de pareamento em °C; TPB: tamanho do fragmento esperado em pares de bases.

<i>Locus</i>	Sequência (5'-3') dos pares de <i>primers</i>	TA	TPB
Embra14 <sup>1</sup>	F: GCCTCAAACCAATTCAAAT R: CATGATTCTCCCCTCCTC	49	100-105
Embra69 <sup>1</sup>	F: TGTGTTCTCGGTTTCAAAC	50	140-300

	R: TGTGAAGTGATGCGAAGC		
Embra148 <sup>1</sup>	F: TGGATGCTGTTCTCATCC R: GGGTTTCTTTGTGAAACGA	50	180-300
Embra166 <sup>1</sup>	F: CTGGTCAACGTCGAAAC R: ATGCTGCAGAGGGCATAA	50	200-320
Embra210 <sup>1</sup>	F: CGTGTGGTTATGTGAACT R: CCTAACAATGCATAAGCTC 1	49	195-210
Eun5 <sup>2</sup>	F: GCTGCAAAGGAAAATAAATGA R: TCGTGCAGCCCTTAATCAG	54	140-162
Eun11 <sup>2</sup>	F: CGCAACATCTTCCCCTATT R: CCATCTCACATCCCTTTTCC	52	162-182

Marcadores desenvolvidos por <sup>1</sup>Brondani *et al.* (1998); <sup>2</sup>Ferreira-Ramos (2008).

Para *C. pachystachya*, foi utilizado um conjunto de dados de 100 indivíduos disponibilizados por Pilati (2017) obtidos utilizando os 10 pares de *primers* descritos na Tabela 2. Os produtos das ampliações foram separados por eletroforese em gel de poliacrilamida a 12%.

**Tabela 2.** Relação dos 10 pares de *primers* microssatélites utilizados para obtenção de dados para a avaliação do tamanho amostral mínimo para *Cecropia pachystachya* Trécul. TA: temperatura de pareamento em °C; TPB: tamanho do fragmento esperado em pares de bases.

<b>Locus</b>	<b>Sequência (5'-3') dos pares de primers</b>	<b>TA</b>	<b>TPB</b>
CEC_08	F: CTGCAATTGACTTGCCACAC R: GGTGTGAAATGAAAGTGACCC	59*	149-206
CEC_12	F: TTCCAATCCGGAGATAAACG R: AAGCAAGAATCTCAAAGCCG	59*	110-128
CEC_17	F: TTCTTGATCGTGTGTTGCTGC R: AAATGTTCAAGGCATTGGTTC	59* a 53	115-127
CEC_31	F: GGGTGTATGCTCTCACACTTG R: TCCATGATATGGTTTGGGTG	59* a 53	129-138
CEC_34	F: TTAGGACTACTGCCTTCGCAC R: TATTGAGGCATGGAGGCTTG	59* a 53	153-163
CEC_37	F: CAAGAGATGCGTCGAGAGTG R: GGCAATCAATTTGCGTAACC	59* a 53	151-157
CEC_43	F: TTCGTGTATGAGGACAACGAG R: AATCCACGAGGAAGCAGAG	59* a 53	293-317
CEC_45	F: TTTACCAAACCCAATTCCC R: ATTCTCAGCAAGTTCCCAGC	59* a 53	118-152
CEC_46	F: AGTACAACACCCGGATCGAC R: TCGAATATAACGCCTCTCGC	59* a 53	112-136
CEC_56	F: TGGCCTTCTTGAGTTGTTTG R: TCAGCCACTCTCACTTTCG	59* a 53	145-155

Marcadores desenvolvidos por Wei e Dick (2014). \* touchdown.

Para *P. guajava* e *B. crispera*, os dados foram obtidos dos trabalhos de Diaz-Cruz (2016) e Micolino (2016), respectivamente. Os pares de *primers* microssatélites utilizados para cada espécie são descritos nas Tabelas 3 e 4. A extração do DNA destas espécies também foi realizada conforme descrita para *C. xanthocarpa*. Os produtos da amplificação do DNA foram resolvidos em gel de agarose de alta resolução a 3% com corrida de aproximadamente 6 horas a 110 V e corados com brometo de etídio.

**Tabela 3.** Relação dos 12 pares de *primers* microssatélites utilizados para obtenção dos dados para a avaliação do tamanho amostral mínimo para *Psidium guajava* L. TA: temperatura de pareamento em °C; TPB: tamanho do fragmento esperado em pares de bases.

<i>Locus</i>	Sequência (5'-3') dos pares de <i>primers</i>	TA	TPB
mPgCIR03	F: TTGTGGCTTGATTCC R: TCGTTTAGAGGACATTTCT	55	118-250
mPgCIR07	F: ATGGAGGTAGGTTGATG R: CGTAGTAATCGAAGAAATG	55	148-160
mPgCIR09	F: GCGTGTCGTATTGTTTC R: ATTTTCTTCTGCCTTGTC	55	156-176
mPgCIR10	F: GTTGGCTCTTATTTTGGT R: GCCCATATCTAGGAAG	55	262-320
mPgCIR11	F: TGAAAGACAACAAACGAG R: TTACACCCACCTAAATAAGA	55	298-214
mPgCIR13	F: CCTTTTTCCCGACCATTACA R: TCGCACTGAGATTTTGTGCT	55	240-260
mPgCIR15	F: TCTAATCCCCTGAGTTTC R: CCGATCATCTCTTTCTTT	55	144-172
mPgCIR16	F: AATACCAGCAACACCAA R: CATCCGTCTCTAACCTC	55	268-296
mPgCIR19	F: AAAATCCTGAAGACGAAC R: TATCAGAGGCTTGCATTA	55	258-280
mPgCIR22	F: CATAAGGACATTTGAGGAA R: AATAAGAAAGCGAGCAGA	55	236-252
mPgCIR25	F: GACAATCCAATCTCACTTT R: TGTGTCAAGCATACTTC	55	104-130
mPgCIR26	F: CTACCAAGGAGATAGCAAG R: GAAATGGAGACTTTGGAG	55	180-198

Marcadores desenvolvidos por Risterucci *et al.* (2005).

**Tabela 4.** Relação dos seis pares de *primers* microssatélites utilizados para obtenção dos dados para a avaliação do tamanho amostral mínimo para *Baccharis crispa* Spreng. TA: temperatura de pareamento em °C; TPB: tamanho do fragmento esperado em pares de bases.

<i>Locus</i>	Sequência (5'-3') dos pares de <i>primers</i>	TA (°C)	TPB
SS20E <sup>1</sup>	F: CACACAGACACTCAAAGCTTCA R: ACCCGCCCTAAAATAAAGA	50	273-299
SS24F <sup>1</sup>	F: AGCTTTTCTTCGCCATTTCCCTTCC R: AATTTGGTTACTGGGTTTTCTTGA	59	156-222
EARI4-6 <sup>2</sup>	F: GCGGTTTGTGTAGAAAGTCC R: ATCTCACTGGTGAATTTTCAGAG	57	207-295
SG_2 <sup>3</sup>	F: TCTAAACTGTAAGTCTTTGATGAAACC R: GCCGTC AATCCTTACAATCC	55*	167-248
SG_8 <sup>3</sup>	F: TCCCTCTTTATTCTTTCAACAAACC R: GTTT AACACCAACATTGCAATCCC	55*	126-172
ER-HAJZC <sup>4</sup>	F: GGATATCGGTTTGGCTTGA R: GGAATCCCTTCTCTTCTGA	55*	179-195

Marcadores desenvolvidos por <sup>1</sup>Wieczorek e Geber (2002); <sup>2</sup>Lindsay *et al.* (2012); <sup>3</sup>Beck *et al.* (2014); <sup>4</sup>Takayama *et al.* (2012). \* touchdown.

#### 4.1.2. Análise dos dados

Para cada espécie, a partir da população original de 100 indivíduos foram construídos subconjuntos de dados consistindo de tamanhos amostrais de 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75 e 95 indivíduos. Cada subconjunto de dados foi produzido de modo independente 50 vezes, totalizando 50 repetições. As repetições foram obtidas utilizando um algoritmo elaborado na plataforma Excel, via linguagem de programação *Visual Basic for Applications* (VBA) onde números foram sorteados aleatoriamente sendo, o número inferior 1 e o número superior 100. Respeitando o que ocorre em uma coleta a campo, cada repetição de cada tamanho amostral foi feita sem reposição, de modo que nenhum indivíduo estivesse presente mais de uma vez em um subconjunto de dados (uma repetição). Porém, como cada subconjunto (repetição) de dados é independente, o mesmo indivíduo poderia estar presente em mais de um subconjunto.

Ao final de todos os sorteios, foram obtidas 600 subpopulações para cada espécie (total de 2400 subpopulações). Após a obtenção das subpopulações, estas foram avaliadas utilizando o programa PopGene (versão 1.32) (YEH *et al.*, 1999) para o cálculo das seguintes medidas de diversidade genética; número de alelos ( $N_a$ ), heterozigosidade esperada ( $H_E$ ), frequência alélica e coeficiente de diferenciação genética ( $F_{ST}$ ). Esses parâmetros foram calculados individualmente para cada uma das 600 subpopulações das quatro espécies. Para o



cálculo de  $F_{ST}$ , as 600 subpopulações de cada espécie foram comparadas com os dados da matriz original de 100 indivíduos. Para visualização dos dados obtidos para o número observado de alelos ( $N_a$ ), a heterozigosidade esperada ( $H_E$ ) e o índice de variação genética ( $F_{ST}$ ) com o auxílio do software PAST (versão 3.16) (HAMMER; HARPER; RYAN, 2001), foram determinadas a amplitude dos dados, a mediana e os quartis. Com base nestes dados, foram desenhados a representação gráfica dos valores em um gráfico *Boxplot*. Para desenho do gráfico de frequências alélicas, foram selecionados, a partir da população original, o alelo de maior frequência, um alelo de frequência intermediária, e um alelo de baixa frequência (frequência entre 0,1 e 0,05) e um alelo raro (frequência entre 0,04 e 0,001). A porcentagem de detecção destes alelos nos diferentes tamanhos amostrais foi demonstrada graficamente.

Para a interpretação dos dados, o número observado de alelos ( $N_a$ ), foi considerado suficiente quando em todas as repetições de um determinado tamanho amostral a detecção dos alelos presentes na população original de 100 indivíduos foi >90%. A heterozigosidade esperada ( $H_E$ ), foi considerada suficiente no tamanho amostral em que em nenhuma das repetições o valor obtido foi >10% acima ou abaixo do valor obtido na população original de 100 indivíduos. Para o coeficiente de diferenciação genética ( $F_{ST}$ ), o tamanho amostral foi considerado suficiente quando em todas as repetições a diferença com a população original foi menor que 0,05, que segundo Wright (1978), indica baixa diferenciação genética.

Todos os índices pontuados acima também foram calculados para a população original de 100 indivíduos e os resultados utilizados para avaliação dos resultados dos diferentes tamanhos amostrais.

## **4.2. Alelos raros**

### **4.2.1. Obtenção dos dados**

Para o entendimento dos efeitos dos alelos raros nos parâmetros número de alelos ( $N_a$ ), diversidade de Nei ( $h$ ), na heterozigosidade esperada ( $H_E$ ), na heterozigosidade observada ( $H_O$ ), no coeficiente de diferenciação genética ( $F_{ST}$ ), na distância genética de Nei ( $D$ ), na análise de coordenadas Principais (PCoA) e na análise Bayesiana a partir da população total de 100 indivíduos de *C. xanthocarpa* e *P. guajava*, foi sorteada novamente uma população de 30 indivíduos para cada espécie. Após a obtenção destas populações foi mantido um arquivo com todos os alelos e em outro arquivo os alelos raros (com frequência inferior a 0,05 na população de 100 indivíduos) foram excluídos. Neste sentido, foram obtidas

quatro populações para cada espécie, sendo uma de 100 indivíduos com todos alelos, uma de 100 indivíduos sem os alelos raros, uma de 30 indivíduos com todos os alelos, e uma de 30 indivíduos sem os alelos raros.

#### 4.2.2. Análise dos dados

O número de alelos ( $N_a$ ), a diversidade genética de Nei ( $h$ ), a heterozigosidade esperada ( $H_E$ ), a heterozigosidade observada ( $H_o$ ), o coeficiente de diferenciação genética ( $F_{ST}$ ) e a distância genética de Nei ( $D$ ) foram calculadas para todas as populações utilizando o software POPGENE (YEH *et al.*, 1999). Para o  $F_{ST}$  e a  $D$ , os dados foram comparados par-a-par.

A Análise de Coordenadas Principais (PCoA), para verificar a distribuição espacial dos indivíduos nos quadrantes e suas conseqüentes contribuições para diversidade genética da espécie foi realizada individualmente para cada população no *software* GenAlEx versão 6.501 (PEAKALL; SMOUSE, 2012). A análise Bayesiana, para obtenção dos *clusters* (grupos genéticos), foi feita também individualmente para cada população utilizando o programa STRUCTURE (PRITCHARD; WEN; FALUSH, 2010). Para determinar o número ideal de *clusters* ( $K$ ), foram feitas simulações partindo do princípio que é possível a obtenção de qualquer número de *clusters* entre de 1 e 5, sendo que para cada simulação foram feitas 10 repetições. Para esta análise foi utilizado o modelo de ascendência sem mistura, e as frequências alélicas foram correlacionados por 500.000 *burn-in* e 1.000.000 MCMC (*Markov Chain Monte Carlo*) repetições após *burn-in*. Para definição do número ideal de  $K$  (grupos genéticos), foi utilizado o *software* STRUCTURE HARVESTER (EARL; VONHOLD, 2012).

A análise dos dados obtidos para os parâmetros cujos os resultados são numéricos ( $h$ ,  $H_E$ ,  $H_o$ ,  $F_{ST}$ ,  $D$ ) foi por comparação entre os valores obtidos nas populações dos diferentes tamanhos amostrais com e sem os alelos raros. Para o número de alelos, os valores foram comparados somente entre os pares de populações com e sem alelos raros. Para a PCoA foi avaliada a distribuição dos indivíduos nos quadrantes e na análise Bayesiana, o número de grupos genéticos obtidos.

## 5. RESULTADOS

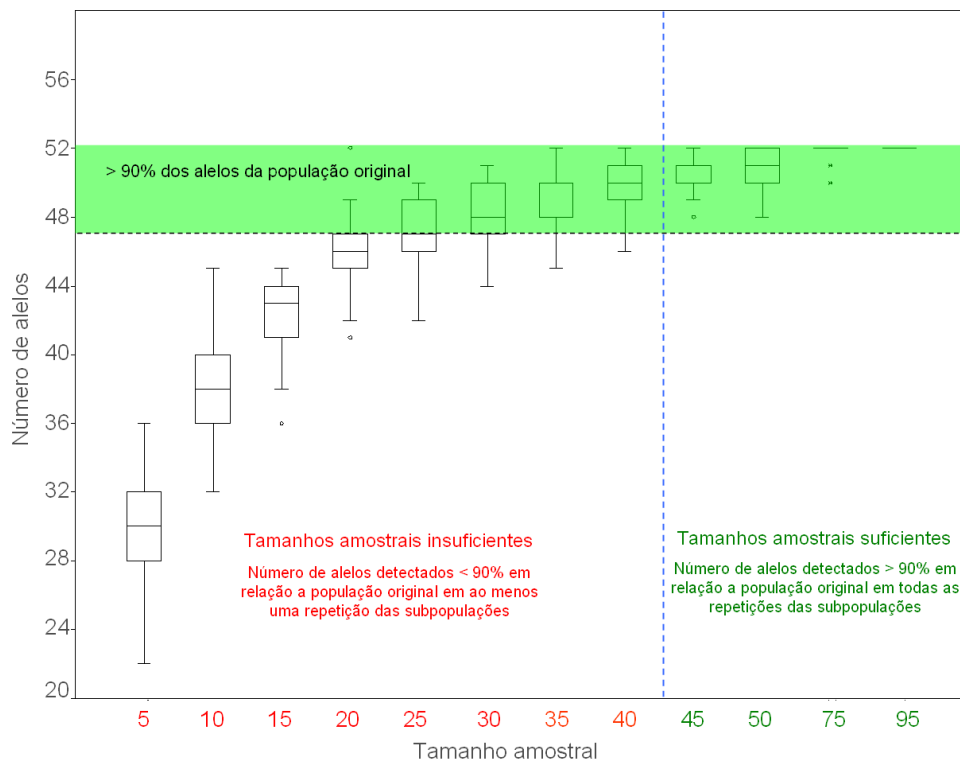
### 5.1. Tamanho amostral

A avaliação do tamanho mínimo amostral, considerando cada índice e cada espécie foi variável. Neste sentido, para facilitar o entendimento, os resultados serão apresentados separadamente por espécie.

#### 5.1.1. *Campomanesia xanthocarpa*

##### Número de alelos (Na)

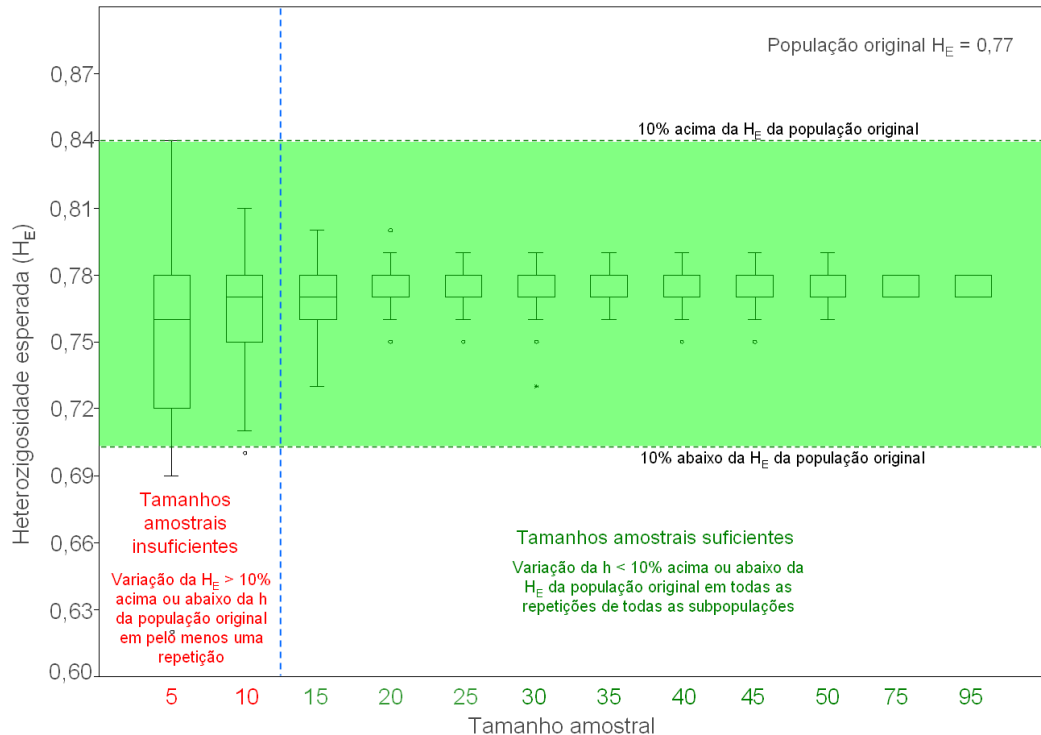
Na população original foram obtidos 52 alelos. A avaliação dos diferentes tamanhos amostrais evidenciou que a partir do tamanho amostral de 45, foi possível detectar 90% ou mais dos alelos presentes na população original de 100 indivíduos (Figura 1).



**Figura 1.** Variação no número de alelos em subpopulações de guabirobeira (*Campomanesia xanthocarpa* O. Berg) de diferentes tamanhos amostrais. Os dados de cada tamanho amostral são oriundos de 50 repetições (subpopulações) obtidas a partir da população original de 100 indivíduos.

## Heterozigosidade esperada ( $H_E$ )

Para a  $H_E$  média, a população original apresentou valor de 0,77. A avaliação das subpopulações dos diferentes tamanhos amostrais evidenciou que, a partir do tamanho amostral de 15 indivíduos, a variação foi inferior a 10% abaixo ou acima da diversidade da população original (Figura 2).

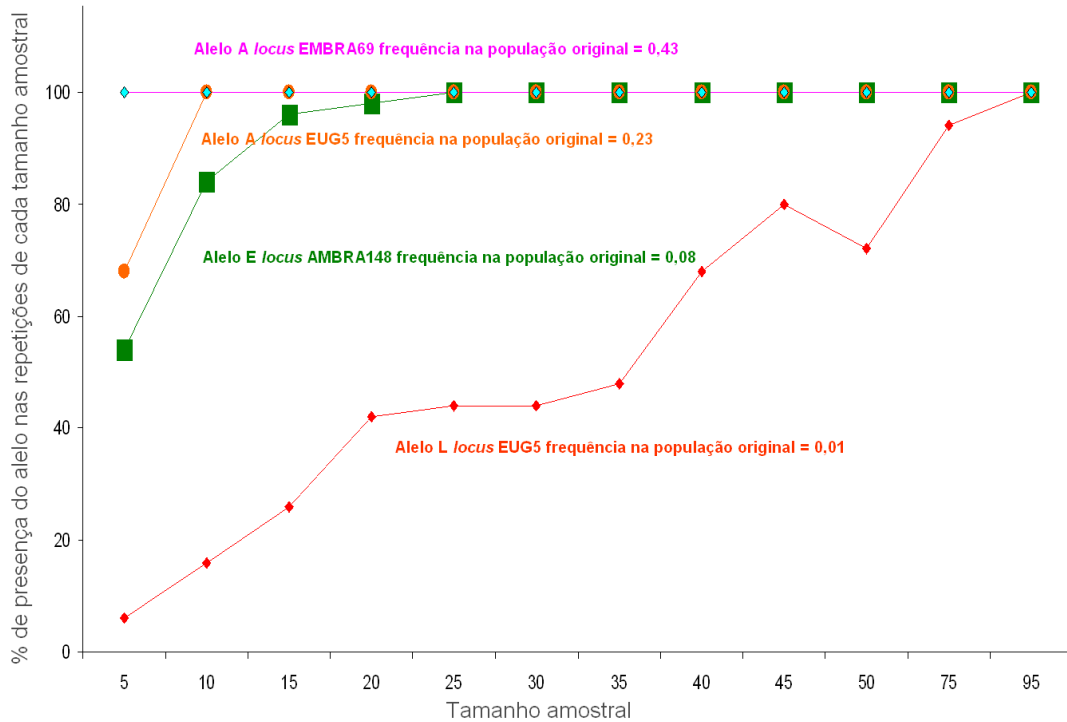


**Figura 2.** Variação na Heterozigosidade esperada ( $H_E$ ) em subpopulações de guabirobeira (*Campomanesia xanthocarpa* O. Berg) de diferentes tamanhos amostrais. Os dados de cada tamanho amostral são oriundos de 50 repetições (subpopulações) obtidas a partir da população original de 100 indivíduos.

## Frequência alélica

Na população original, o alelo A do *locus* EMBRA69 teve a maior frequência (0,43). Quando avaliado em diferentes tamanhos amostrais, foi possível detectar esse alelo em todas as subpopulações de todos os tamanhos amostrais. Portanto, com um tamanho amostral de 5 indivíduos, já é possível detectar o alelo em todas as populações (Figura 3). O alelo A do *locus* Eug5 teve frequência intermediária (0,23) na população original. A avaliação desse alelo mostrou que a partir do tamanho amostral 10 é possível detectar este alelo em todas as subpopulações (Figura 3). Porém, para o alelo E do *locus* EMBRA148, observado em baixa frequência (0,08) na população original, sua detecção em todas as subpopulações ocorre a

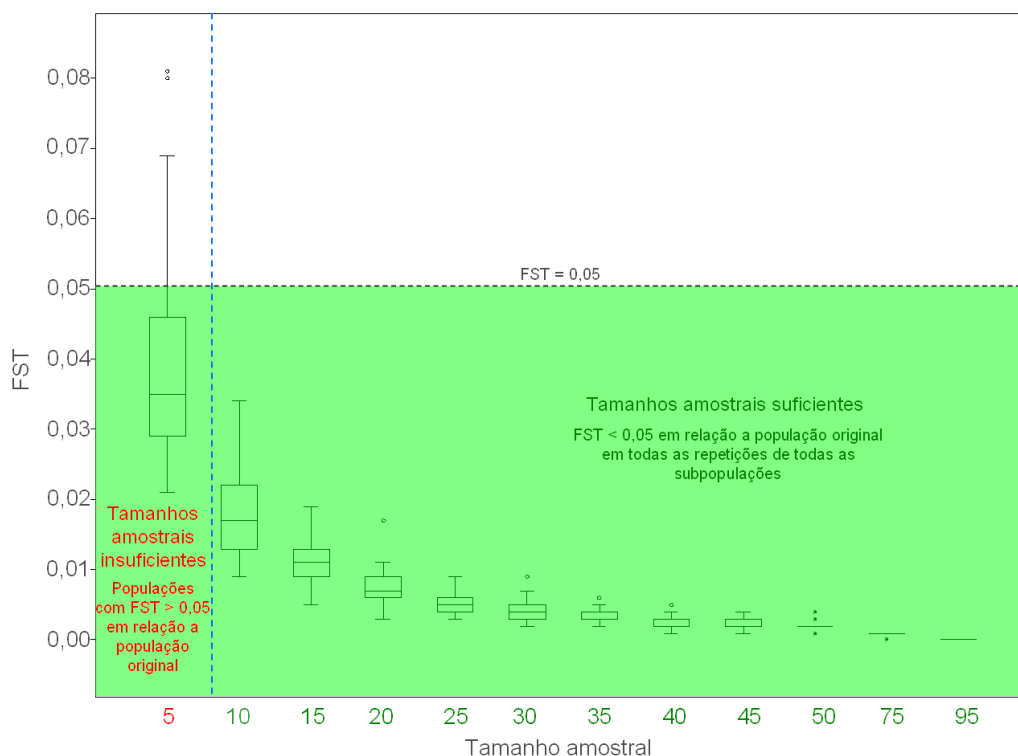
partir do tamanho amostral 25. Para o alelo L do *locus* EUG 5, considerado como de frequência rara na população original (0,01), a detecção em todas as subpopulações só é possível no tamanho amostral de 95 indivíduos (Figura 3).



**Figura 3.** Impacto do tamanho da amostra na detecção de alelos com diferentes frequências na população original de guabiobeira (*Campomanesia xanthocarpa* O. Berg). Os dados de cada tamanho amostral são oriundos de 50 repetições (subpopulações) obtidas a partir da população original de 100 indivíduos.

### Coefficiente de diferenciação genética ( $F_{ST}$ )

A comparação de cada subpopulação de cada tamanho amostral com a população original mostrou que, a partir do tamanho amostral de 10 indivíduos, o valor do  $F_{ST}$  foi inferior a 0,05 em todas as repetições (Figura 4).

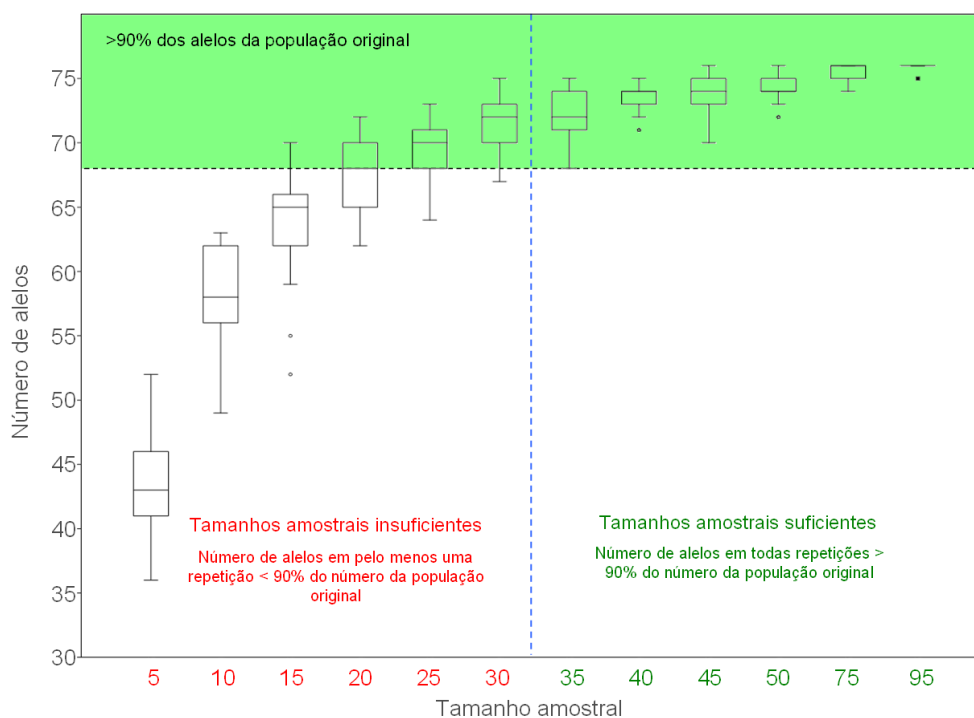


**Figura 4.** Variação no valor do  $F_{ST}$  entre as subpopulações de guabirobeira (*Campomanesia xanthocarpa* O. Berg) de diferentes tamanhos amostrais e a população original contendo 100 indivíduos. Os dados de cada tamanho amostral são oriundos de 50 repetições (subpopulações) obtidas a partir da população original.

### 5.1.2. *Cecropia pachystachya*

#### Número de alelos ( $N_a$ )

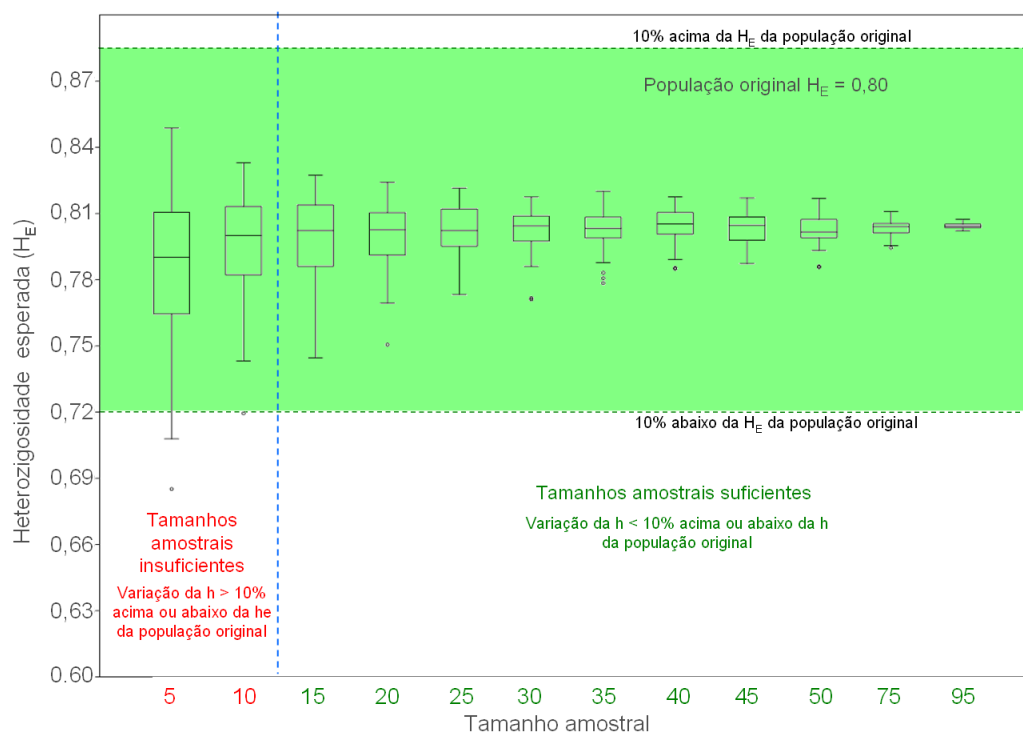
A população original de 100 indivíduos apresentou 76 alelos. A avaliação dos diferentes tamanhos amostrais evidenciou que, a partir do tamanho amostral 35, é possível detectar em todas as repetições 90% ou mais dos alelos presentes na população original de 100 indivíduos (Figura 5).



**Figura 5.** Variação no número de alelos detectados em subpopulações de Embaúba (*Cecropia pachystachya* Trécul) de diferentes tamanhos amostrais. Os dados de cada tamanho amostral são oriundos de 50 repetições (subpopulações) obtidas a partir da população original de 100 indivíduos.

### Heterozigosidade esperada ( $H_E$ )

A  $H_E$  da população original foi de 0,83. Para a  $H_E$  média nas subpopulações, a partir do tamanho amostral 15, a variação que ocorre é inferior a 10% acima ou abaixo do valor da população original (Figura 6).

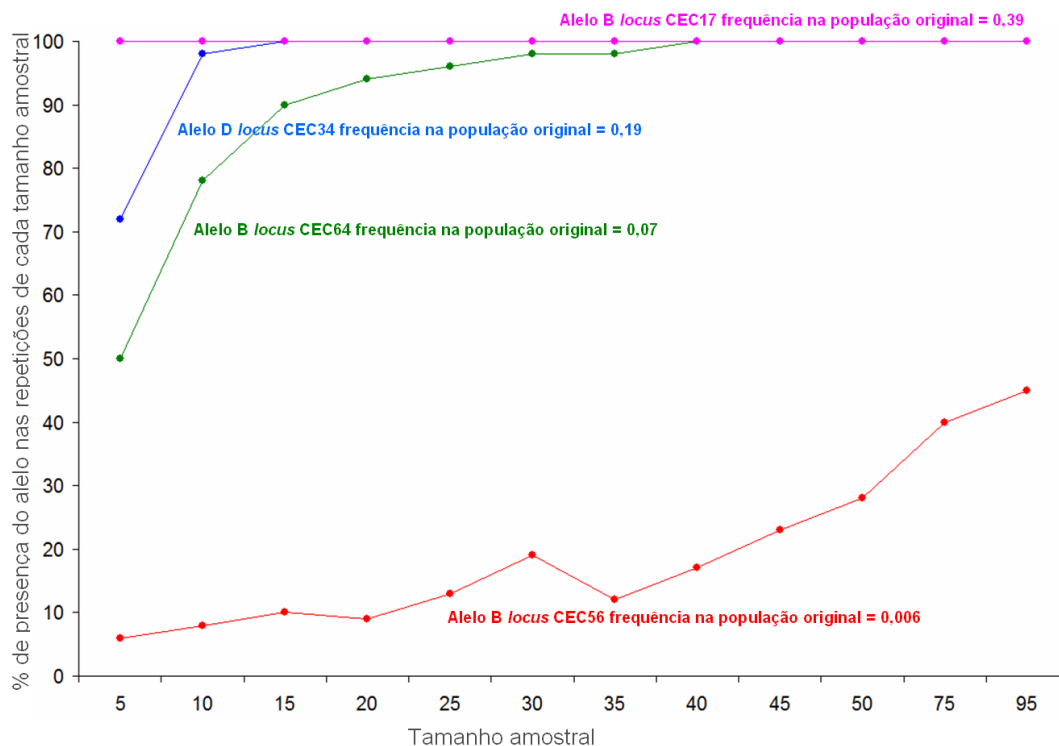


**Figura 6.** Variação na Heterozigidade esperada ( $H_E$ ) em subpopulações de Embaúba (*Cecropia pachystachya* Trécul) de diferentes tamanhos amostrais. Os dados de cada tamanho amostral são oriundos de 50 repetições (subpopulações) obtidas a partir da população original de 100 indivíduos.

### Frequência alélica

Na população original, o alelo B do *locus* CEC\_17 teve a maior frequência (0,39), sendo possível detectar alelos de alta frequência em todos os tamanhos amostrais. Portanto, com um tamanho amostral de 5 indivíduos já é possível detectar o alelo em todas as populações (Figura 7). Para o alelo D do *locus* CEC\_34 de frequência intermediária (0,19) na população original, houve estabilidade a partir do tamanho amostral 15 (Figura 7). A avaliação desse alelo mostrou que a partir do tamanho amostral 15, já é possível detectar este alelo em todas as subpopulações da população original. No entanto, para o alelo B do *locus* CEC64 que ocorre em baixa frequência na população original (0,07), houve estabilidade a partir do tamanho amostral de 40 indivíduos. O alelo B do *locus* CEC\_56, raro na população original (frequência de 0,006), não esteve presente em todas as subpopulações em nenhum tamanho amostral (Figura 7).

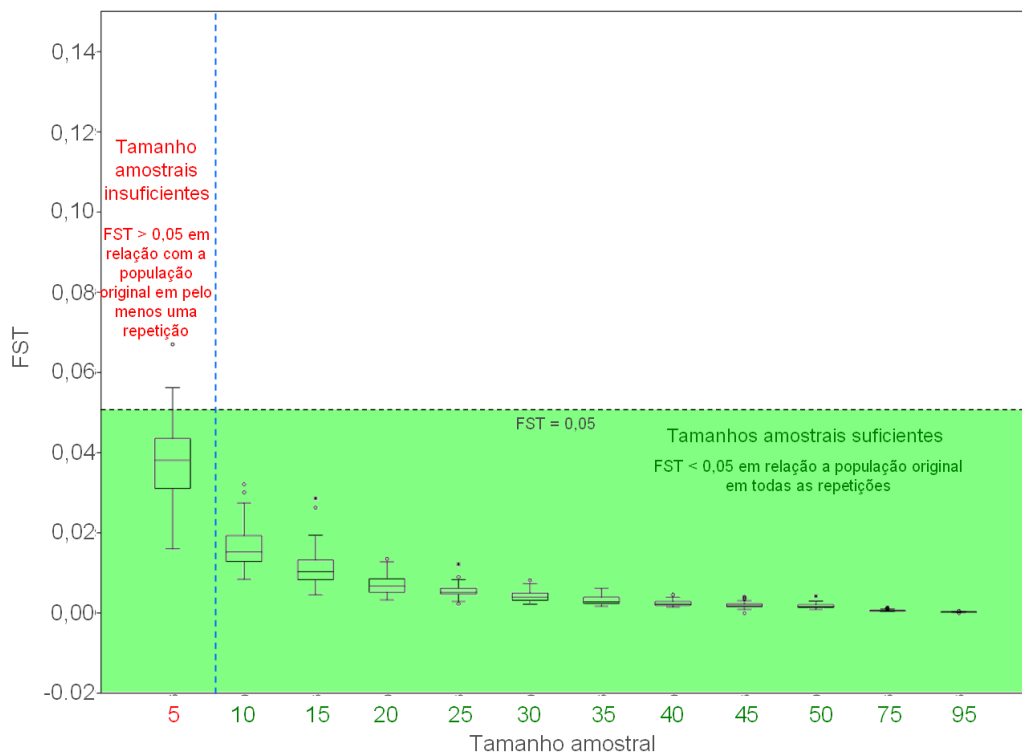




**Figura 7.** Porcentagem das subpopulações de Embaúba (*Cecropia pachystachya* Trécul) de cada tamanho amostral que foram capazes de detectar o alelo de maior frequência (alelo B locus CEC\_17) e menor frequência (alelo B locus CEC\_56), de frequência média (alelo D CEC\_34) e frequência entre 0,05 e 0,1 (alelo B CEC\_64) na população original contendo 100 indivíduos. Cada tamanho amostral, obtido a partir da população original, foi composto de 50 repetições.

### **Coefficiente de diferenciação genética ( $F_{ST}$ )**

No coeficiente de diferenciação genética ( $F_{ST}$ ) foi observado que a partir do tamanho amostral 10, a diferença entre a população original e as repetições foi menor que 0,05 (Figura 8).

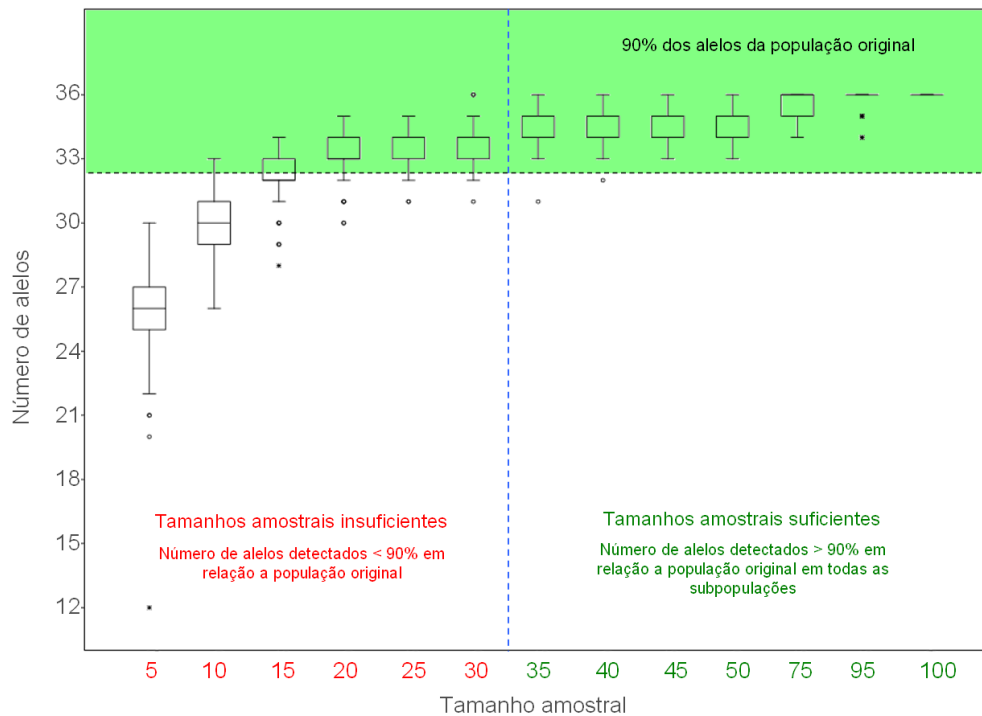


**Figura 8.** Variação no valor do  $F_{ST}$  entre subpopulações de Embaúba (*Cecropia pachystachya* Trécul) e a população original contendo 100 indivíduos. Cada tamanho amostral, obtido a partir da população original, foi composto de 50 repetições.

### 5.1.3. *Psidium guajava*

#### Número de alelos ( $N_a$ )

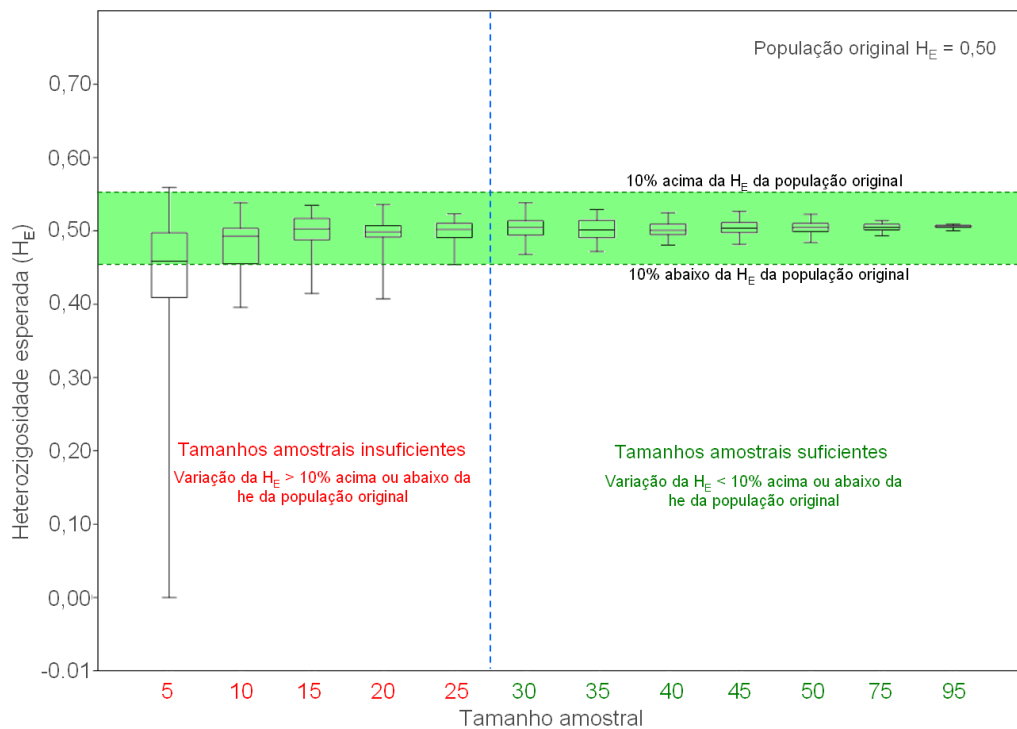
O número de alelos na população original foi de 36. A avaliação dos diferentes tamanhos amostrais evidenciou que, a partir do tamanho amostral 35 foi possível detectar em todas as subpopulações 90% ou mais dos alelos presentes na população original de 100 indivíduos (Figura 9).



**Figura 9.** Variação no número de alelos em subpopulações de goiabeiras (*Psidium guajava* L.) de diferentes tamanhos amostrais. Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições. A população original foi composta de 100 indivíduos e apresentou 36 alelos.

### Heterozigosidade esperada ( $H_E$ )

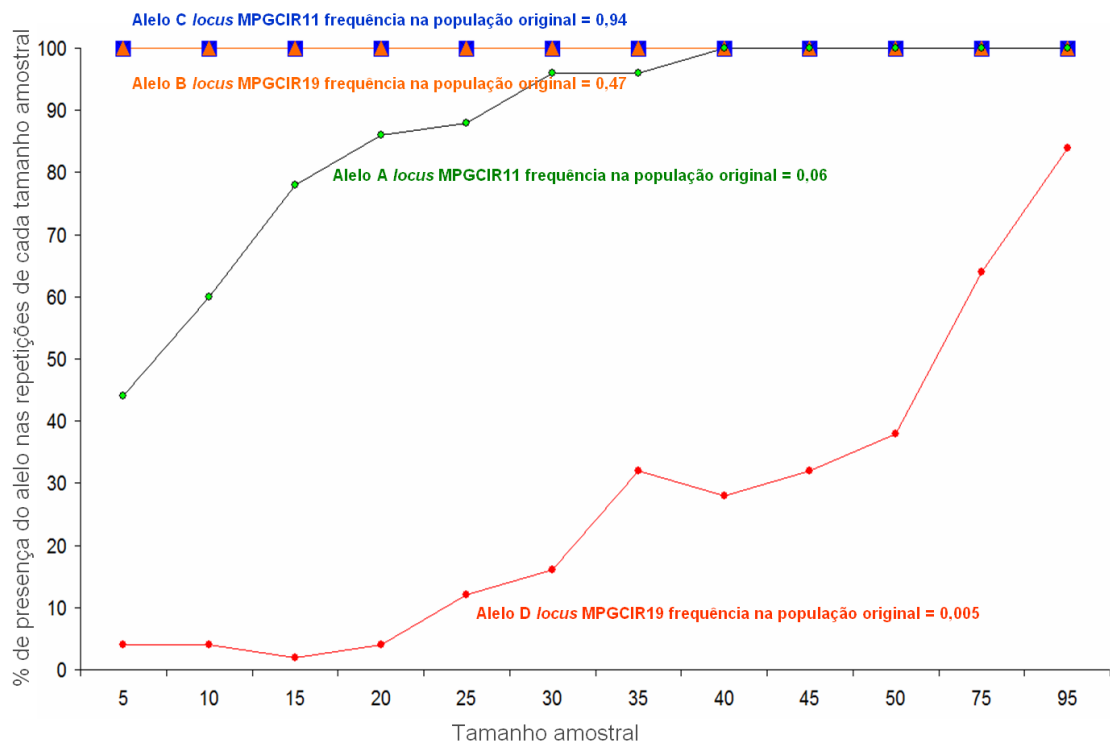
Para a  $H_E$  média, a população original apresentou valor de 0,50. A avaliação das subpopulações dos diferentes tamanhos amostrais evidenciou que a partir do tamanho amostral de 30, a variação foi menor que 10% abaixo ou acima da diversidade da população original (Figura 10).



**Figura 10.** Heterozigosidade esperada ( $H_E$ ) em subpopulações de goiabeiras (*Psidium guajava* L.) de diferentes tamanhos amostrais. Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições. A população original foi composta de 100 indivíduos.

### Frequência alélica

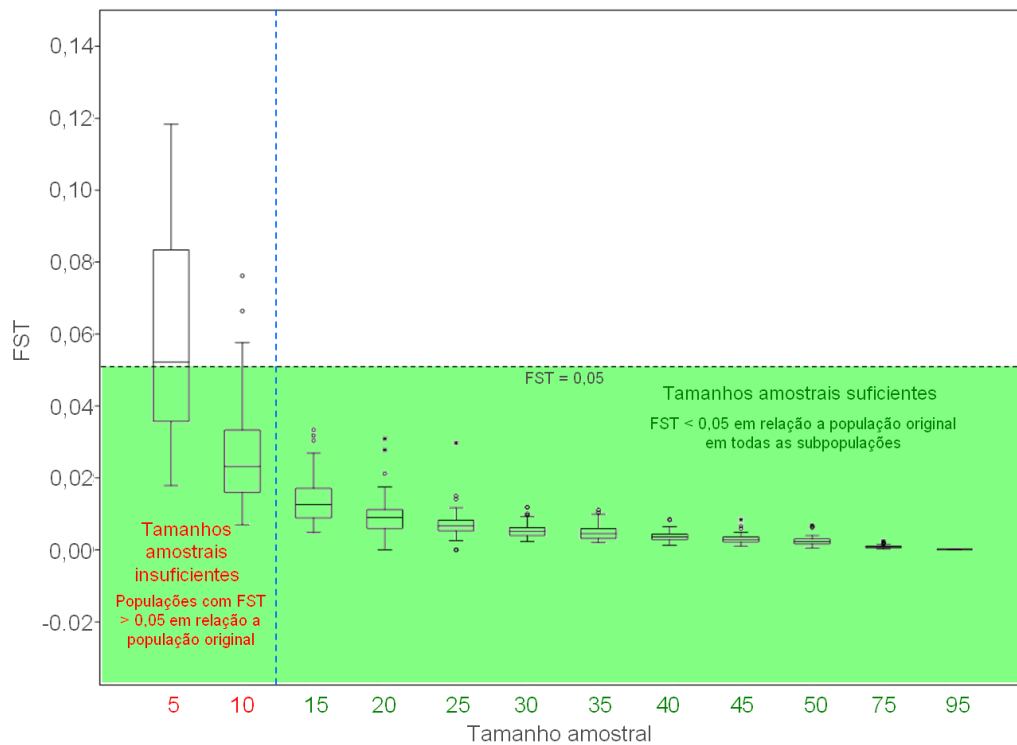
Na população original, alelo C do *locus* MPGCIR11 teve a maior frequência alélica (0,94). Já o alelo B do *locus* MPGCIR19 apresentou frequência intermediária (0,47). Para esses alelos, o tamanho amostral de 5 indivíduos já é o suficiente para suas detecções em todas as subpopulações (Figura 11). O alelo A do *locus* MPGCIR11, foi detectado em baixa frequência (0,06). A avaliação desse alelo mostrou que a partir do tamanho amostral 40, é possível detectar esse alelo em todas as subpopulações da população original. Todavia, o alelo D do *locus* MPGCIR19, ocorreu em frequências raras (0,005) e não se estabilizou nas subpopulações (Figura 11).



**Figura 11.** Impacto do tamanho da amostra na detecção de alelos nas subpopulações de goiabeiras (*Psidium guajava* L.) da frequência alélica na população original contendo 100 indivíduos, das frequências do alelo de maior (MPGCIR11 alelo C) e menor (MPGCIR19 alelo D) valor, alelo intermediário (MPGCIR19 alelo B) sendo alelos em uma frequência real ( $\geq 0,05$ ). E alelo com frequência variável entre 0,005 a 0,1 (MPGCIR11 alelo A). Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições.

### **Coefficiente de diferenciação genética ( $F_{ST}$ )**

No coeficiente de diferenciação genética ( $F_{ST}$ ) foi observado que a partir do tamanho amostral 15, a diferença entre a população original e as subpopulações foi menor que 0,05 (Figura 12).

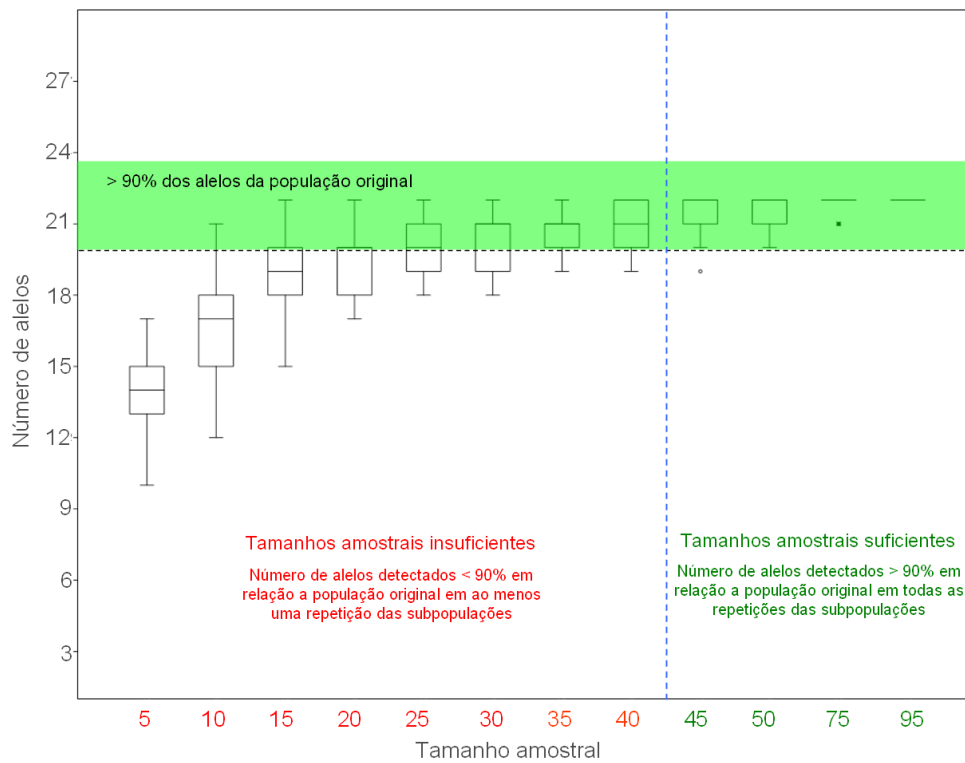


**Figura 12.** Variação no valor do  $F_{ST}$  entre subpopulações de goiabeiras (*Psidium guajava* L.) e a população original contendo 100 indivíduos. Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições. A população original foi composta de 100 indivíduos.

#### 5.1.4. *Bacharis crispa*

##### Número de alelos ( $N_a$ )

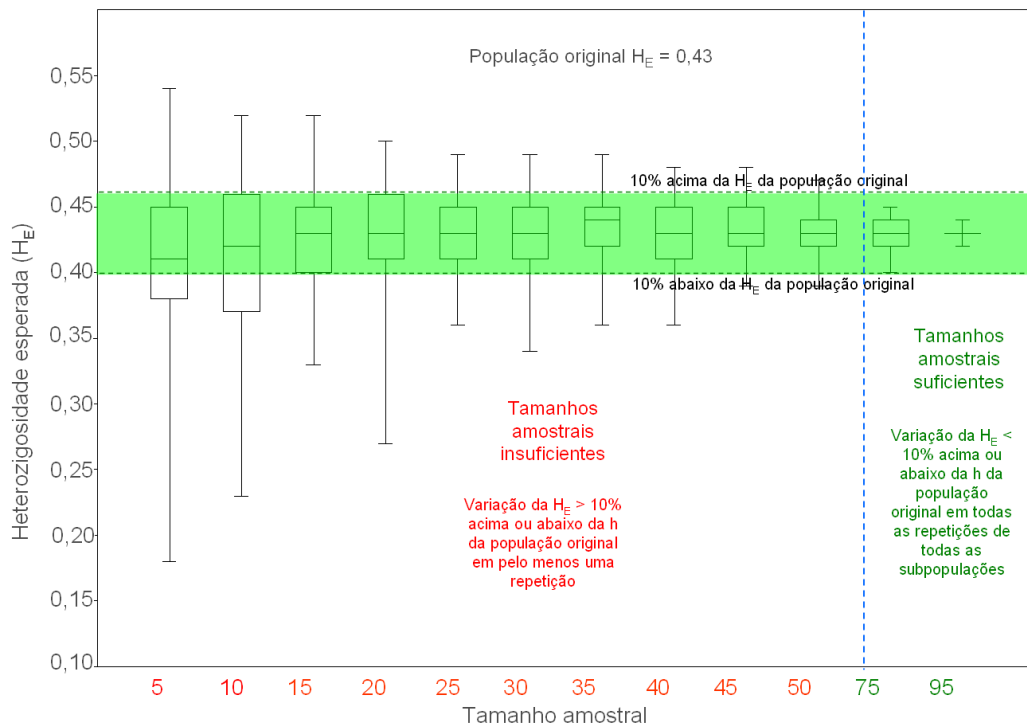
A avaliação dos diferentes tamanhos amostrais evidenciou que a partir do tamanho amostral 45, foi possível detectar 90% ou mais dos alelos dos 22 presentes na população original de 100 indivíduos (Figura 13).



**Figura 13.** Variação no número de alelos em subpopulações de Carqueja (*Baccharis crispa* Spreng.) de diferentes tamanhos amostrais. Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições. A população original foi composta de 100 indivíduos e apresentou 22 alelos.

### Heterozigosidade esperada ( $H_E$ )

Para a  $H_E$  média, a população original apresentou valor de 0,43. A avaliação das subpopulações dos diferentes tamanhos amostrais evidenciou que a partir do tamanho amostral de 75, a variação foi inferior a 10% abaixo ou acima da  $H_E$  da população original (Figura 14).

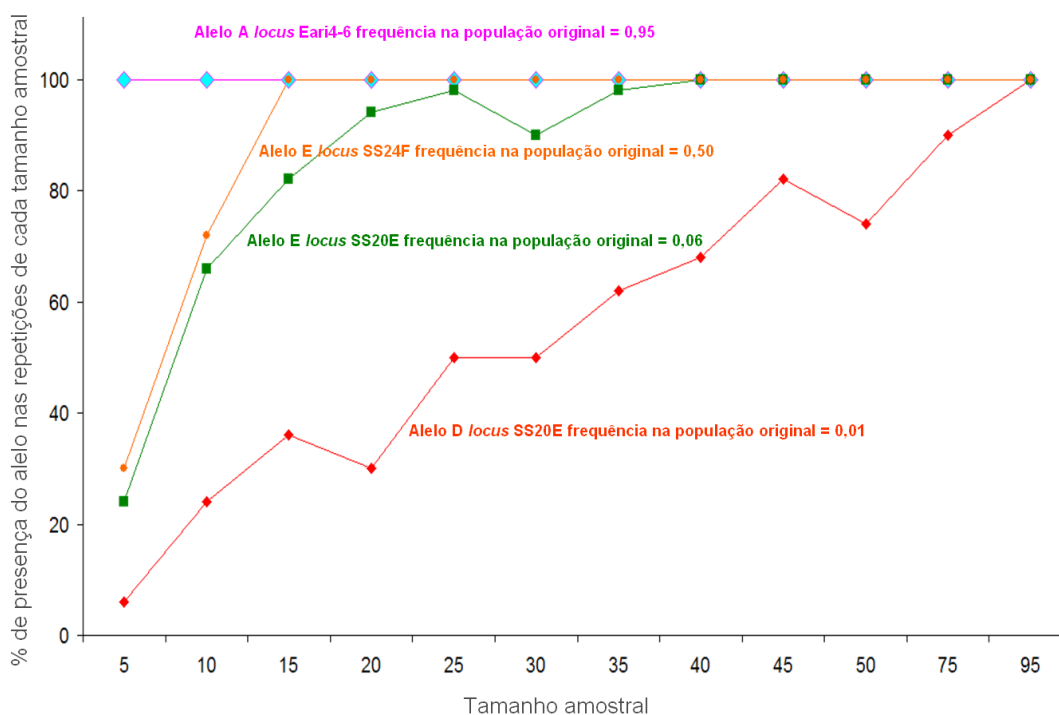


**Figura 14.** Heterozigiosidade esperada ( $H_E$ ) em subpopulações de Carqueja (*Baccharis crispa* Spreng.) de diferentes tamanhos amostrais. Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições. A população original foi composta de 100 indivíduos.

### Frequência alélica

Na população original, o alelo A do *locus* Eari4-6 teve a maior frequência (0,95). Quando avaliado nos diferentes tamanhos amostrais, foi possível detectar esse alelo em todas as subpopulações de todos tamanhos amostrais. Portanto, um tamanho amostral de 5 indivíduos é o suficiente para detectar o alelo em todas as subpopulações (Figura 15). O alelo E do *locus* SS24F apresentou frequência intermediária (0,50) na população original. A avaliação deste alelo mostrou que a partir do tamanho amostral 15 já é possível detectar esse alelo em todas as subpopulações (Figura 15). O alelo E do *locus* SS20E, foi detectado em baixa frequência na população original (0,06). A detecção desse em todas as subpopulações ocorre a partir do tamanho amostral 40. Para o alelo D do *locus* SS20E, considerado com frequência rara na população original (0,01), a detecção em todas as subpopulações só foi possível no tamanho amostral de 95 indivíduos (Figura 15).

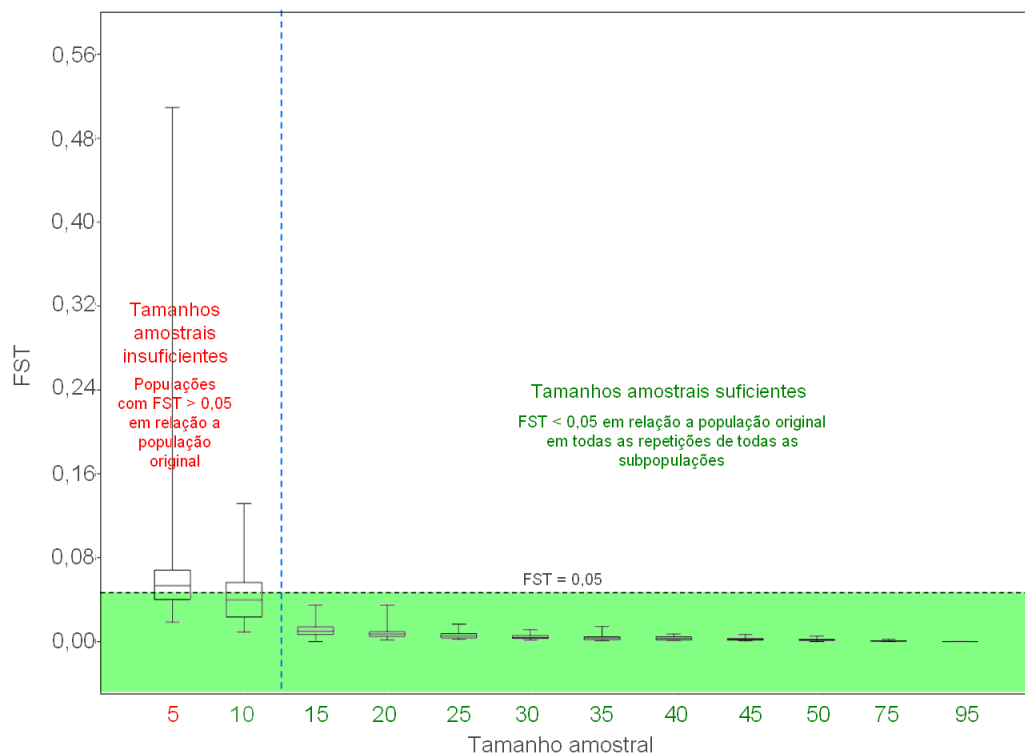




**Figura 15.** Impacto do tamanho da amostra na detecção nas subpopulações de Carqueja (*Baccharis crispa* Spreng.) de alelos com diferentes frequências na população original contendo 100 indivíduos. Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições.

### **Coefficiente de diferenciação genética ( $F_{ST}$ )**

O confronto de cada subpopulação de cada tamanho amostral com a população original evidenciou que a partir do tamanho amostral de 15 indivíduos, o valor de  $F_{ST}$  foi menor que 0,05 (Figura 16).



**Figura 16.** Variação no valor do  $F_{ST}$  entre subpopulações de Carqueja (*Baccharis crispa* Spreng.) e a população original contendo 100 indivíduos. Cada tamanho amostral, obtido a partir da população original, foi composta de 50 repetições. A população original foi composta de 100 indivíduos.

A compilação dos dados para cada índice avaliado e em cada espécie estão apresentados na Tabela 5.

**Tabela 5.** Relação dos tamanhos amostrais mínimos para a obtenção de dados genéticos confiáveis utilizando marcadores microssatélites nas espécies *Campomanesia xanthocarpa* O. Berg. *Cecropia pachystachya* Trécul, *Psidium guajava* L. e *Baccharis crispa* Spreng.

<b>Índices genéticos</b>	<i>C. xanthocarpa</i>	<i>C. pachystachya</i>	<i>P. guajava</i>	<i>B. crispa</i>
Detecção de 90% dos alelos presentes na população original	45	35	35	45
Heterozigosidade esperada ( $H_E$ ) variando menos de 10% abaixo ou acima da população original	15	15	30	75
Detecção de 95% dos alelos de...				
$F_{ST}$ menor que 0,05 em relação a população original	10	10	15	15

## 5.2. Alelos raros

A avaliação de populações com ausência e presença dos alelos raros evidenciou que estes alelos não influenciam os parâmetros número de alelos ( $N_a$ ), diversidade de Nei ( $h$ ), na heterozigosidade esperada ( $H_E$ ), na heterozigosidade observada ( $H_O$ ), no coeficiente de diferenciação genética ( $F_{ST}$ ), na distância genética de Nei ( $D$ ), na análise de coordenadas Principais (PCoA) e na análise Bayesiana nas espécies *C. xanthocarpa* e *P. guajava*. Para facilitar o entendimento, os resultados são apresentados separadamente por espécie.

### 5.2.1. *Campomanesia xanthocarpa*

O  $N_a$  total na população original foi de 52, sendo que destes, 11 foram considerados raros, com frequência menor que 0,05. A diferença do  $N_a$  entre os pares de populações foi de apenas 1 alelo (Tabela 6). Também, para todos parâmetros  $h$ ,  $H_E$ ,  $H_O$  não houve diferença relevante nos valores obtidos (Tabela 6). A distância genética entre qualquer população foi menor que 0,006 (Tabela 7). O  $F_{ST}$  dos pares de populações foram todos menores que 0,05 (Tabela 7).

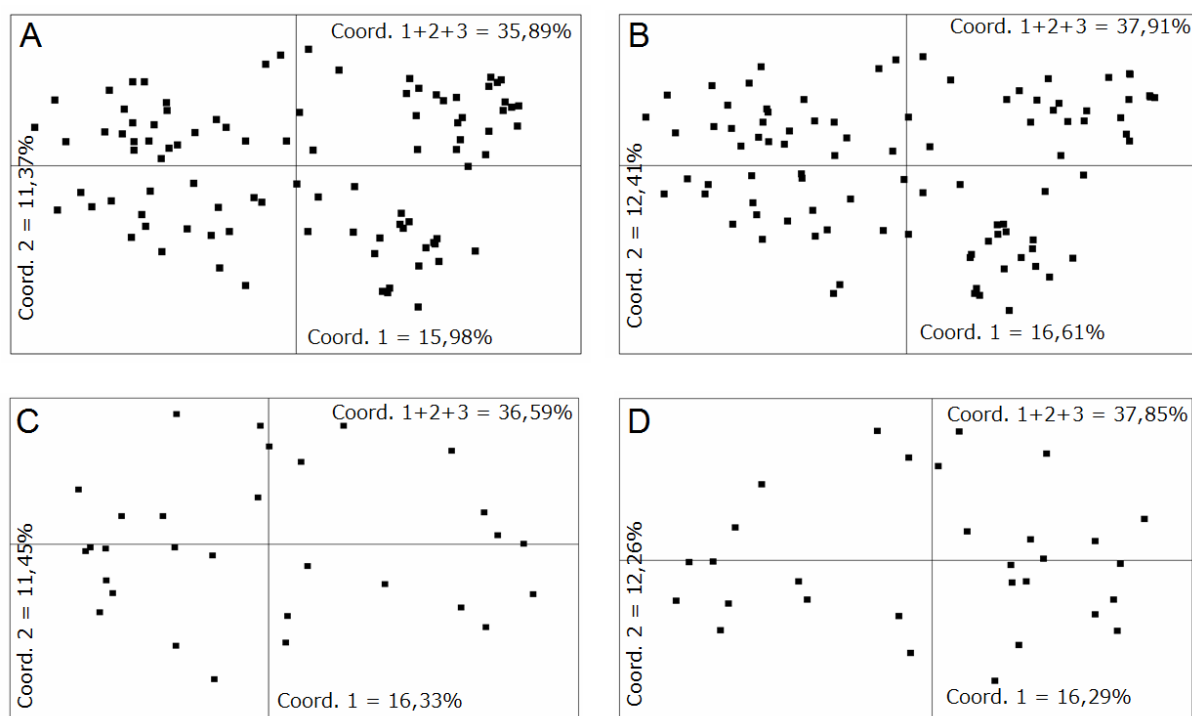
**Tabela 6.** Número observado de alelos ( $N_a$ ), diversidade de Nei ( $h$ ), heterozigosidade esperada ( $H_e$ ) e heterozigosidade observada ( $H_o$ ) das subpopulações de *Campomanesia xanthocarpa* com e sem alelos raros.

	$N_a$	$h$	$H_e$	$H_o$
População de 100 indivíduos com alelos raros	52	0,77	0,78	0,58
População de 30 indivíduos com alelos raros	51	0,77	0,78	0,62
População de 100 indivíduos sem alelos raros	41	0,76	0,77	0,57
População de 30 indivíduos sem alelos raros	40	0,76	0,77	0,60

**Tabela 7.** Estimativas das distâncias genéticas de Nei (1978) (abaixo da diagonal) e o coeficiente de diferenciação genética ( $F_{ST}$ ) (acima da diagonal) entre as quatro subpopulações de *Campomanesia xanthocarpa* O. Berg. avaliadas. 100: populações com 100 indivíduos; 30: populações com 30 indivíduos; CR: com alelos raros; SR: sem alelos raros

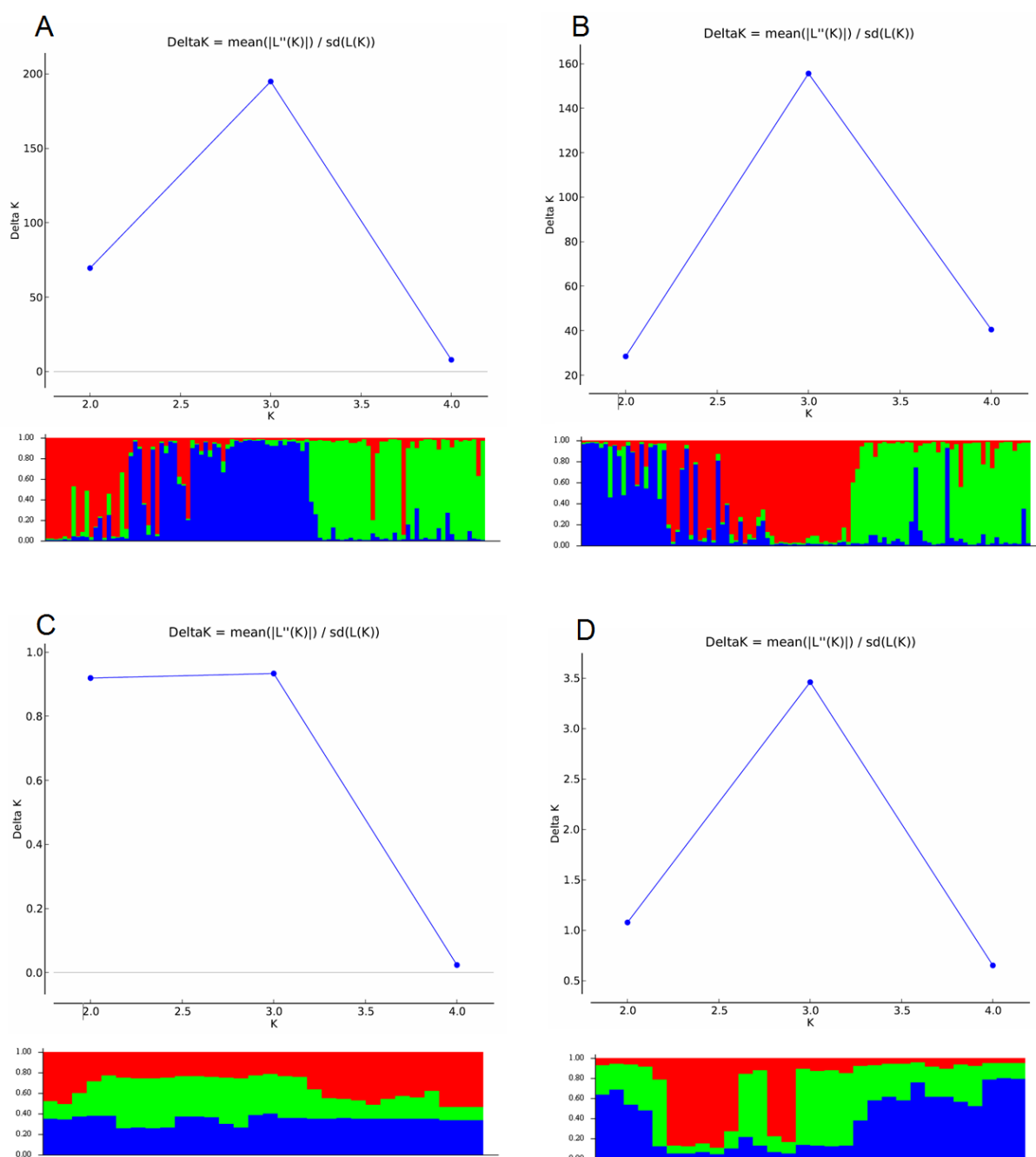
Populações	100CR	100SR	30CR	30SR
100CR		0,001	0,004	0,006
100SR	-0,002		0,005	0,004
30CR	0,004	0,011		0,002
30SR	0,016	0,016	-0,020	

A análise de coordenadas principais (PCoA) reduziu a dimensionalidade dos dados e permitiu a observação da estrutura espacial dos indivíduos e cada uma das populações com e sem alelos raros. A distribuição espacial (área ocupado nos quadrantes pelas populações) e a contribuição das coordenadas para variabilidade genética da espécie foi semelhante em todas as populações independente da presença ou ausência dos alelos raros (Figura 17).



**Figura 17.** Análise de Coordenadas Principais (PCoA) das subpopulações de *Campomanesia xanthocarpa* O. Berg. com e sem alelos raros. **A)** população de 100 indivíduos com os alelos raros; **B)** população de 100 indivíduos sem os alelos raros; **C)** população de 30 indivíduos com os alelos raros; **D)** população de 30 indivíduos sem os alelos raros.

Nas simulações feitas pelo programa STRUCTURE com abordagem Bayesiana, o número de K (Clusters, grupos genéticos) foi definido como três para as quatro subpopulações (Figura 18).



**Figura 18.** Número de grupos genéticos (k) e distribuição destes grupos nas subpopulações de *Campomanesia xanthocarpa* O. Berg. com e sem alelos raros. **A)** população de 100 indivíduos com os alelos raros; **B)** população de 100 indivíduos sem os alelos raros; **C)** população de 30 indivíduos com os alelos raros; **D)** população de 30 indivíduos sem os alelos raros.

### 5.2.2. *Psidium guajava*

O Na total na população original foi de 36, e destes, 6 com frequência menor que 0,05, foram considerados raros. A diferença entre os pares de populações (com e sem alelos raros) foi de 6 alelos (Tabela 8). Para os parâmetros  $h$ ,  $H_E$ ,  $H_o$  não houve diferença relevante entre

os valores obtidos tanto quando comparado os pares de populações com e sem alelos raros de mesmo tamanho amostral, como de diferentes tamanhos (Tabela 8). A distância genética entre qualquer par de populações foi menor que 0,006 (Tabela 8). O  $F_{ST}$  dos pares de populações foram todos menores que 0,05 (Tabela 9).

**Tabela 8.** Número observado de alelos ( $N_a$ ), diversidade de Nei ( $h$ ), heterozigosidade esperada ( $H_E$ ) e heterozigosidade observada ( $H_o$ ) das subpopulações de *Psidium guajava*.

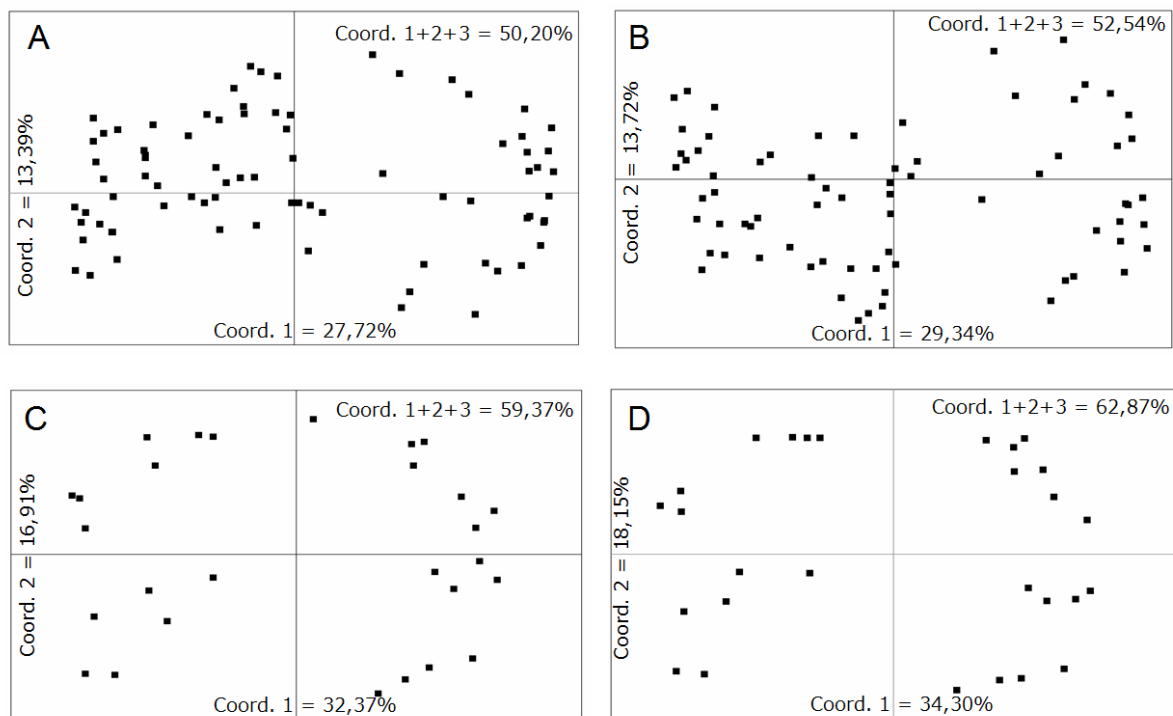
	$N_a$	$h$	$H_E$	$H_o$
População de 100 indivíduos com alelos raros	36	0,50	0,51	0,10
População de 100 indivíduos sem alelos raros	30	0,49	0,49	0,09
População de 30 indivíduos com alelos raros	36	0,49	0,50	0,09
População de 30 indivíduos sem alelos raros	30	0,47	0,48	0,08

As distâncias genéticas de Nei (1978) variaram de -0,0069 a 0,0140. No entanto, o coeficiente de diferenciação genética ( $F_{ST}$ ) variou de 0.0005 a 0.0103 (Tabela 8). Tanto nas distâncias genéticas quanto em  $F_{ST}$ , o aumento do tamanho amostral e a análise com e sem os alelos raros mostraram que as diferenças nos índices genéticos foram mínimas.

**Tabela 9.** Estimativas das distâncias genéticas de Nei (1978) (abaixo da diagonal) e o coeficiente de diferenciação genética ( $F_{ST}$ ) (acima da diagonal) entre as quatro subpopulações de *Psidium guajava* L. avaliadas neste trabalho.

	<b>100CR</b>	<b>100SR</b>	<b>30CR</b>	<b>30SR</b>
<b>100CR</b>		0,001	0,009	0,010
<b>100SR</b>	-0,002		0,0098	0,001
<b>30CR</b>	0,013	0,014		0,001
<b>30SR</b>	0,014	0,013	-0,007	

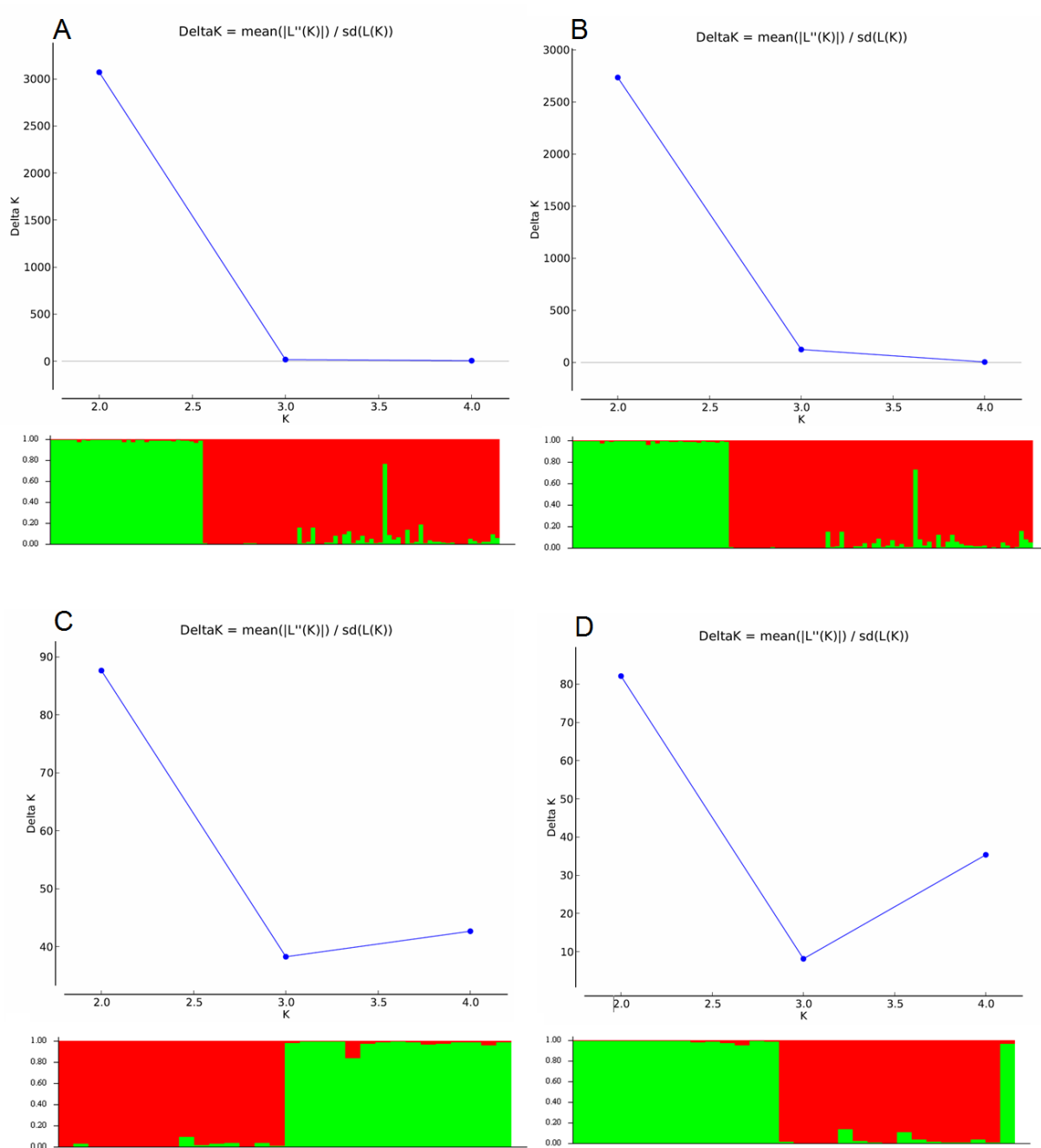
Na PCoA, a redução da dimensionalidade dos dados permitiu observar que a distribuição espacial (área ocupado nos quadrantes pelas populações) e a contribuição das coordenadas para variabilidade genética da espécie foi semelhante em todas as populações independente da presença ou ausência dos alelos raros (Figura 19).



**Figura 19.** Análise de Coordenadas Principais (PCoA) das subpopulações de *Psidium guajava* L. com e sem alelos raros. **A)** população de 100 indivíduos com os alelos raros; **B)** população de 100 indivíduos sem os alelos raros; **C)** população de 30 indivíduos com os alelos raros; **D)** população de 30 indivíduos sem os alelos raros.

A análise Bayesiana evidenciou que o número de K (Clusters, grupos genéticos) não diferiu em função da presença de alelos raros e nem em função do tamanho amostral (Figura 20).





**Figura 20.** Número de grupos genéticos (k) e distribuição destes grupos nas subpopulações de *Psidium guajava* L. com e sem alelos raros. **A)** população de 100 indivíduos com os alelos raros; **B)** população de 100 indivíduos sem os alelos raros; **C)** população de 30 indivíduos com os alelos raros; **D)** população de 30 indivíduos sem os alelos raros.

## 6. DISCUSSÃO

O tamanho amostral ideal considerando todos os índices e todas as espécies avaliadas é de 45 indivíduos, porém, com 30 indivíduos já é possível quantificar a diversidade e estruturação genética das populações de forma confiável.

Nas angiospermas aqui estudadas, para detectar 95% dos alelos ( $N_a$ ) da população original foram necessários 95 indivíduos (considerando uma população de 100 indivíduos). Em gimnospermas, utilizando dados reais, o tamanho amostral ideal para detectar todos os alelos de uma população parece ser dependente do tamanho real da população (informação quase sempre indisponível) (BASHALKHANOV; PANDEY; RAJORA, 2009). Porém, esses autores afirmam que para populações com tamanhos finitos, amostras de 150 indivíduos podem ser suficientes para detectar 95% dos alelos, corroborando com os valores obtidos por Gapare; Yanchuk e Aitken (2008). Já, Danusevicius; Kavaliauskas e Fussi (2016), com uma análise mais refinada, concluem que para detecção com 95% de confiança de alelos com frequência acima de 0,05, são necessários de 25 a 30 indivíduos. No entanto, para alelos com frequência abaixo de 0,01, o tamanho amostral deve ser de 300 a 350 indivíduos. Avaliando uma população de uma espécie de angiosperma, Miyamoto *et al.* (2008) conclui que para detectar todos os alelos são necessários em torno de 300 indivíduos. No estudo aqui realizado com angiospermas, os achados corroboram com o pontuado por Miyamoto *et al.* (2008) e Danusevicius; Kavaliauskas e Fussi (2016), quanto menor a frequência do alelo na população original, maior o tamanho amostral para conseguir detectar esse alelo (Figuras 3, 7, 11 e 15). O mesmo ocorreu com o trabalho de Herrmann; Flajoulot e Julier (2010); o tamanho amostral de 40 indivíduos foi suficiente para detectar os alelos representativos, porém, neste tamanho amostral, os alelos que não estavam presentes foram em sua maioria raros. Em nosso estudo, em todas as espécies avaliadas, 95% dos alelos com frequência na população original acima de 0,05, são detectados com o tamanho amostral de 30 indivíduos.

A importância dos alelos raros (aqui considerados aqueles com frequência menor que 0,05) deve ser relativada de acordo com o objetivo do estudo. Esses alelos podem ser importantes na genética da conservação, pois são justamente eles que podem possibilitar a adaptação a uma nova condição ambiental que possa surgir (PRIMACK, 2006; BASHALKHANOV; PANDEY; RAJORA, 2009). Porém, o esforço amostral, considerando trabalho e recursos financeiros, para detectar esses alelos, torna trabalhos com 150 a 300 indivíduos por população inviáveis. Hale; Burg e Steeves (2012), afirmam que em estudos genéticos populacionais o que deve ser priorizado não é a detecção de todos os alelos, mas

sim daqueles que são informativos e em frequências representativas da população original. Segundo esses autores, os alelos raros teriam pouco efeito nos índices genéticos em animais e, assim, mesmo que não seja possível detectá-los em tamanhos amostrais menores, os índices podem continuar robustos. A análise do efeito dos alelos raros nos índices comumente utilizados em estudos genéticos populacionais em plantas corroborou com o observado por Hale; Burg e Steeves (2012) em animais. Nesse sentido, não observamos diferença nos índices genéticos populacionais de  $N_a$ ,  $h$ ,  $H_E$ ,  $H_o$ ,  $F_{ST}$ ,  $D$ , na PCoA e no número de grupos genéticos quando comparadas populações com 100 e 30 indivíduos com e sem alelos raros (Tabelas 5 a 8 e Figuras 17 a 20). Por fim, a análise geral dos dados evidencia que os alelos raros não influenciam os parâmetros de  $h$ ,  $H_E$ ,  $H_o$ ,  $F_{ST}$ ,  $D$ , a PCoA e a determinação do número de grupos genéticos, possibilitando a utilização de tamanhos amostrais de 30 indivíduos.

O tamanho amostral mínimo quando considerado a  $H_E$  foi variável, com destaque para *B. crispera* que ficou muito acima das demais espécies (Tabela 7). Em gimnospermas, tamanhos amostrais de 3 a 15 indivíduos já são suficientes para alcançar valores de  $H_E$  muito próximos da população original, fazendo com que esse índice, nesse grupo de plantas, seja considerado insensível ao tamanho amostral (BUCHERT *et al.*, 1997; RAJORA *et al.*, 2000; BASHALKHANOV; PANDEY; RAJORA, 2009). Hale; Burg e Steeves (2012) demonstraram que a precisão das heterozigosidades esperadas ( $H_E$ ) é influenciada pelo nível de polimorfismo de cada *locus* nas espécies estudadas de animais. No estudo desses autores, o tamanho amostral representativo ideal (com valores de  $H_E$  mais próximos da população real) foi menor para os *loci* com altas  $H_E$ , e maior para os *loci* com baixas  $H_E$ . Com isso, nos *loci* com alta  $H_E$  (acima de 0,77) e baixa (0,50), as heterozigosidades podem ser estimadas com precisão com tamanhos amostrais a partir de 20 a 100 indivíduos, respectivamente. Essas observações podem justificar a insensibilidade da  $H_E$  das gimnospermas ao tamanho amostral, pois esse grupo apresenta polinização cruzada, com conseqüente elevada heterozigose. Para três das angiospermas aqui estudadas, os sistemas reprodutivos são bem conhecidos, *Cecropia pachystachya* ( $H_E = 0,83$ ) e *C. xanthocarpa* ( $H_E = 0,77$ ) são de polinização cruzada (RIBAS-SANTOS; SANTOS; GUERRA, 2019; PEDERSOLI; GAGLIOTI; TEIXEIRA 2020), e *P. guajava* ( $H_E = 0,50$ ) possui de 60 a 75% de autopolinização (SALAZAR *et al.*, 2006; ALVEZ e FREITAS, 2007). Porém, para *B. crispera* ( $H_E = 0,50$ ), as plantas são dióicas, mas o modo de reprodução ainda é pouco conhecido (HEIDEN; IGANCI; MACIAS, 2009). Como esperado, as  $H_{ES}$  dessas espécies, os valores parecem decair de acordo com a facilidade de ocorrer

autopolinização. *Cecropia pachystachya*, com maior  $H_E$  apresenta plantas dióicas, *C. xanthocarpa*, também com alta  $H_E$ , é monóica, porém autoimcompatível, e *P. guajava*, com baixa  $H_E$ , apresenta elevado grau de autopolinização. Ainda, é provável que apesar de *B. crispera* ser dióica, a reprodução da espécie a partir dos rizomas (HEIDEN; IGANCI; MACIAS, 2009) possa contribuir para os valores de  $H_E$  observados. Mesmo considerando que o comportamento reprodutivo influencia diretamente os valores de  $H_E$ , e esses no tamanho mínimo amostral, uma amostragem de 30 indivíduos foi ideal para estimativas precisas de  $H_E$  para três das quatro espécies estudadas.

A análise de diferenciação genética ( $F_{ST}$ ) evidenciou que o aumento amostral acima de 15 indivíduos não tem impacto nos valores obtidos para as quatro espécies estudadas. Com base em dados simulados de microssatélites, Kalinowski (2005) sugere que 20 indivíduos seria um tamanho amostral razoável, quando o valor do  $F_{ST}$  é 0,05. Entretanto, em marcadores *SNPs*, em que o tamanho populacional ideal é oito indivíduos, estimativas confiáveis de diferenciação genética podem ser obtidas com apenas dois indivíduos (NAZARENO *et al.*, 2017). Esses resultados mostram que tamanhos amostrais com poucos indivíduos, geram estimativas precisas de  $F_{ST}$ . Porém, vale ressaltar que o aumento no número de *locus* no estudo resulta em um aumento na precisão da detecção do  $F_{ST}$ . Se houver *loci* suficientes, as estimativas de diferenciação genética podem ser obtidas com poucos indivíduos (NEI, 1987; KALINOWSKI, 2005). No presente estudo, também foi possível obter informações sobre o grau de diferenciação genética, isto é, à medida que aumenta o tamanho amostral, diminui a diferenciação entre as subpopulações. No entanto, essa diferença não é biologicamente significativa e o uso de 15 indivíduos pode ser considerado adequado para gerar estimativas de  $F_{ST}$ . Porém, a utilização de poucos indivíduos pode afetar a precisão de outros índices genéticos.

Para estudos de genética de populações o tamanho amostral apropriado é aquele que permite que as estimativas genéticas sejam representativas da população total. Neste estudo, os índices genéticos analisados nas quatro espécies mostraram que o tamanho amostral de 30 indivíduos é ideal para estimar dados genéticos populacionais em angiospermas quando o marcador utilizado é o microssatélite. Amostrando mais de 30 indivíduos por população, a variação nos índices (exceto  $N_a$ ) é tão pequena que não leva a uma interpretação diferente dos dados. Além disso, o aumento além de 30 indivíduos apenas elevaria os custos e a duração da pesquisa e não seria capaz de interferir na interpretação dos dados genéticos populacionais da espécie.

## 7. REFERÊNCIAS

- ALVES, J.E.; FREITAS, B.M. Requerimentos de polinização da goiabeira. **Ciência Rural** v.37, n.5, p.1281-1286, set-out, 2007.
- AYALA, F.J.; CAMPBELL, C.A. Frequency-dependent selection. **Annual review of Ecology and systematics**, v. 5, n. 1, p. 115-138, 1974.
- BASHALKHANOV, S.; PANDEY, M.; RAJORA, O. P. A simple method for estimating genetic diversity in large population from finite sample sizes. **BMC Genetics**, v. 10, n. 84, 2009.
- BECK, J.B.; SEMPLE, J.C.; BRULL, J.M.; LANCE, S.L.; PHILLIPS, M.M.; HOOT, S.B.; MEYER, G.A. Genus-wide microsatellite primers for the goldenrods (*Solidago*; Asteraceae). **Applications in Plant Sciences**, v. 2, n. 4, 2014.
- BOSSE, M., SPURGIN, L.G., LAINE, V.N., COLE, E.F., FIRTH, J.A., GIENAPP, P.; GOSLER, A.G.; MCMAHON, K.; POISSANT, J.; VERHAGEN, I.; GROENEN, M.A.M.; OERS, K.V.; SHELDON, B.C.; VISSER, M.E.; SLATE, J. Recent natural selection causes adaptive evolution of an avian polygenic trait. **Science**, v. 358, n. 6361, p. 365-368, 2017.
- BRONDANI, R.P.V.; BRONDANI, C.; TARCHINI, R.; GRATTAPAGLIA, D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics**, Berlin, v. 97, p. 816–827, 1998.
- BROWN, K.E.; KELLY, J.K. Severe inbreeding depression is predicted by the “rare allele load” in *Mimulus guttatus*. **Evolution**, v.74, n.3, p. 587-596, 2020.
- BUCHERT, G.P.; RAJORA, O.P.; HOOD, J.V.; DANCİK, B.P. Effects of harvesting on genetic diversity in old-growth Eastern white pine in Ontario, Canada. **Conservation Biology**, v. 11, n. 3, p. 747-758, 1997.
- CORANDER, J.; FRASER, C.; GUTMANN, M.U.; ARNOLD, B.; HANAGE, W.P.; BENTLEY, S.D.; LIPSITCH, M.; CROUCHER, N.J. Frequency-dependent selection in vaccine-associated pneumococcal population dynamics. **Nature ecology & evolution**, v.1, n.12, p.1950-1960, 2017.
- CORTADI, A.; DI SAPIO, O.; MC CARGO, J.; SCANDIZZI, A.; GATTUSO, S.; GATTUSO, M. Anatomical Studies of *Baccharis Artriculata*, *Baccharis Crispa* and *Baccharis Trimeria*, “Carquejas” Used in Folk Medicine. **Pharmaceutical Biology**, v. 37, n. 5, 357-365, 1999.
- CRUZ, D.T.; SELBACH-SCHNADELBACH, A.; LAMBERT, S.M.; RIBEIRO, P.L.; BORBA, E.L. Genetic and morphological variability in *Cattleya elongata* Barb. Rodr. (Orchidaceae), endemic to the campo rupestre vegetation in northeastern Brazil. **Plants Systematics and Evolution**, n.294, p.87-98, 2011.

DANUSEVIČIUS, D.; KAVALIAUSKAS, D.; FUSSI, B. Optimum sample size for SSR-based estimation of representative allele frequencies and genetic diversity in Scots pine populations. **Baltic Forestry**, v. 22, n. 2, p. 194-202, 2016.

DA-SILVA, P.R.; MILACH, S.C.K.; TISIAN, L.M. Transferability and utility of white oat (*Avena sativa*) microsatellite markers for genetic studies in black oat (*Avena strigosa*). **Genetics and Molecular Research**, v. 10, p. 2916-2923, 2011.

DAVEY, J.W.; HOHENLOHE, P.A.; ETTER, P.D.; BOONE, J.Q.; CATCHEN J.M.; BLAXTER, M.L. Descoberta de marcadores genéticos em todo o genoma e genotipagem usando sequenciamento de última geração . **Nature Reviews. Genetics**, v.12, p. 499-510, 2011.

DIAZ-CRUZ, J.A. **Diversidade e estrutura genética de populações de *Psidium guajava* L. (Myrtaceae) oriundas do Brasil e do México**. 2016. **Dissertação (Mestrado em Biologia Evolutiva)**. Universidade Estadual do Centro-Oeste, Guarapuava, 2016.

EANES, W.F.; KOEHN, R.K. The correlation of rare alleles with heterozygosity: determination of the correlation for the neutral models. **Genetics Research**, v. 29, n. 3, 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. **Conservation Genetics Resources**, v.4, n.2, p.359-361, 2012.

EBERT, D.; HAMILTON, W.D. Sex against virulence: the coevolution of parasitic diseases. **Trends in Ecology & Evolution**, v. 11, n. 2, p. 79-82, 1996.

ESSELINK, G.D.; NYBOM, H.; VOSMAN, B. Assignment of allelic configuration in polyploids using the MAC-PR (Microsatellite DNA Allele Counting-Peak Ratios) method. **Theoretical and Applied Genetics**, v. 109, p. 402- 408, 2004.

FAGUNDES, B S.; SILVA, L F.; GIACOMIN, R.M.; SECCO, D.; DIAZ-CRUZ, J.A.; Da-Silva, P.R. Transferability of microsatellite markers among Myrtaceae species and their use to obtain population genetics data to help the conservation of the Brazilian Atlantic Forest. **Tropical Conservation Science**, v. 9, p. 408-422, 2016.

FALEIRO, F.G. **Marcadores Genético-moleculares aplicados a programas de conservação e uso de recursos genéticos**. Planaltina, Embrapa Cerrados, 2007.

FERREIRA-RAMOS, R.; LABORDA, P.R.; SANTOS, M.D.; MAYOR, M.S.; MESTRINER, M.A.; DE SOUZA, A.P.; ALZATE-MARIN, A.L. Genetic analysis of forest species *Eugenia uniflora* L. through of newly developed SSR markers. **Conservation Genetics** v. 9: p.1281–1285, 2008.

GANDARA, F.B.; DA-SILVA, P.R.; DE MOURA, T.M.; PEREIRA, F.B.; GOBATTO, C.R.; FERRAZ, E.M.; KAGEYAMA, P.Y.; TAMBARUSSI, E.V. The Effects of Habitat Loss on Genetic Diversity and Population Structure of *Cedrela fissilis* Vell.. **Tropical Plant Biology**, v. 12, p. 1-11, 2019.

GAPARE, W.J.; YANCHUK, A.D.; AITKEN, S.N. Optimal sampling strategies for capture of genetic diversity differ between core and peripheral populations of *Picea sitchensis* (Bong.) Carr. **Conservation Genetics**, v. 9(2), p. 411-418, 2008.

GILLESPIE, J.H. **Population Genetics. A Concise Guide**. Baltimore/London: The Johns Hopkins University Press. p.19-48, 1998.

GUICHOUX, E.; LAGACHE, L.; WAGNER, S.; CHAUMEIL, P.; LÉGER, P.; LEPAIS, O.; LEPOITTEVIN, C.; MALAUSA, T.; REVARDEL, E.; SALIN, F.; PETIT, R.J. Current trends in microsatellite genotyping. **Molecular Ecology Resources**, v. 11, p. 591-611, 2011.

GUO, H.-Y.; WANG, Z.-L.; HUANG, Z.; CHEN, Z.; YANG, H.-B.; KANG, X.-Y. Genetic Diversity and Population Structure of *Alnus cremastogyne* as Revealed by Microsatellite Markers. **Forests**, v. 10, 2019.

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**, v. 7, n. 9, p. 1-10, 2012.

HAMMER, Ø.; HARPER, D.A.T.; RYAN, P.D. PAST: Paleontological statistics software package for education and data analysis. **Palaeontologia Electronica**, v. 4 (1), 9pp. 2001.

HARTL, D.L.; CLARK, A.G. **Principles of Population Genetics**, 3<sup>o</sup> ed. Sunderland USA, Sinauer Associates Inc. p. 542, 1997.

HEIDEN, G.; IGANCI, J.R.V.; MACIAS, L. *Baccharis* sect. *Caulopterae* (Asteraceae, Astereae) in Rio Grande do Sul state, Brazil. **Rodriguésia**, v. 60, n. 4, p. 943-983, 2009.

HEINO, M.; METZ, J.A.J.; KAITALA, V. The enigma of frequency-dependent selection. **Trends in Ecology & Evolution**, v. 13, n. 9, p. 367-370, 1998.

HERRMANN, D.; FLAJOULOT, S.; JULIER, B. Sample size for diversity studies in tetraploid alfalfa (*Medicago sativa*) based on codominantly coded SSR markers. **Euphytica**, v. 171, n. 3, p. 441-446, 2010.

HODEL, R.G.; SEGOVIA-SALCEDO, M.C.; LANDIS, J.B.; CROWL, A.A.; SUN, M.; LIU, X.; GITZENDANNER, M.A.; DOUGLAS, N.A.; GERMAIN-AUBREY, C.C.; CHEN, S.; SOLTIS, D.E.; SOLTIS, P.S. The report of my death was an exaggeration: a review for researchers using microsatellites in the 21st century. **Applications in plant sciences**, v. 4, n. 6, 2016.

HUANG, W.; ZHAO, X.; ZHAO, X.; LI, Y.; LIAN, J.; YUN, J. Relationship between the genetic diversity of *Artemisia halodendron* and climatic factors. **Acta Oecologica**, n.55, p.97-103, 2014.

HURST, L.D. Fundamental concepts in genetics: genetics and the understanding of selection. **Nature Reviews Genetics**, v. 10 (2), p. 83-93, 2009.

ISLAM, M.R.; LI, Z.Z.; GICHIRA, A.W.; ALAM, M.N.; FU, P.C.; HU, G.W.; WANG, Q.F.; CHEN, L.Y. Population genetics of *Calotropis gigantea*, a medicinal and fiber resource plant,

as inferred from microsatellite marker variation in two native countries. **Biochemical genetics**, v. 57, n. 4, p. 522-539, 2019.

JOHNSON, M.T.; PRASHAD, C.M.; LAVOIGNAT, M.; SAINI, H.S. Contrasting the effects of natural selection, genetic drift and gene flow on urban evolution in white clover (*Trifolium repens*). **Proceedings of the Royal Society B: Biological Science**, v. 285, n. 1883, p. 20181019, 2018

KALIA, R.K.; RAI, M.K.; KALIA, S.; SINGH, R.; DHAWAN, A. Microsatellite markers: An overview of the recent progress in plants. **Euphytica**, v. 111, p. 309– 334, 2011.

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

KIANI, M.; MEMARIANI, F.; ZARGHAMI, H. Molecular analysis of species of *Tulipa* L. from Iran based on ISSR markers. **Plant Systematics and Evolution**, v.298, p.1515-1522, 2012.

KIMURA, M. Evolutionary rate at the molecular level. **Nature**, v. 217, n. 5129, p. 624-626, 1968.

KIMURA, M. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, 1983.

KING, J.L.; JUKES, T.H. Non-Darwinian evolution. **University of California, Berkeley**, v. 94720, p. 788, 1969.

LADLE, R.J. Parasites and sex: catching the Red Queen. **Trends in ecology & evolution**, v. 7, n. 12, p. 405-408, 1992.

LEVIN, B.R. Frequency-dependent selection in bacterial populations. **Philosophical Transactions of the Royal Society of London. B, Biological Sciences**, v. 319, n. 1196, p. 459-472, 1988.

LEWONTIN, R.C. The Units of Selection. **Annual Review of Ecology and Systematics**, v. 1, p. 1–18, 1970.

LIMA, D.F.; GOLDENBERG, R.; SOBRAL, M. O gênero *Campomanesia* (Myrtaceae) no estado do Paraná, Brasil. **Rodriguésia**, v. 62, n. 3, p. 683-693, 2011.

LINDSAY, D.L.; BAILEY, P.P.; ANDERSON, J.L.; JUNG, M.G.; EDWARDS, C.; LANCE, R.; Isolation and characterization of microsatellite loci for a hyper-rare cliff endemic, *Erigeron lemmonii*, and a more widespread congener, *Erigeron arisolius* (Asteraceae). **Conservation Genetics Resources**, v. 4, p. 849-852, 2012.

LITT, M.; LUTY, J.A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle action gene. **American Journal of Human Genetics**, v. 44, n. 3, p.397-401, 1989.



MANGARAVITE, É.; VINSON, C.C.; RODY, H.V.; GARCIA, M.G.; CARNIELLO, M.A.; SILVA, R.S.; OLIVEIRA, L.O. Contemporary patterns of genetic diversity of *Cedrela fissilis* offer insight into the shaping of seasonal forests in eastern South America. **American Journal of Botany**, v. 103(2), p. 307-16, 2016.

MARANDEL, F.; CHARRIER, G.; LAMY, J.B.; LE CAM, S.; LORANCE, P.; TRENKEL, V.M. Estimating effective population size using RADseq: Effects of SNP selection and sample size. **Ecology and evolution**, v.10, n. 4, p. 1929-1937, 2020.

MAYR, E. *Animal Species and Evolution*. Cambridge, Mass: Harvard Univ. Press, 1963.

MCDERMOTT, J.M.; MCDONALD, B.A. Gene flow in plant pathosystems. **Journal Annual Review of Phytopathology**, v. 31, n.1, p.353-373, 1993.

MEMON, S.; JIA, X.; GU, L.; ZHANG, X. Genomic variations and distinct evolutionary rate of rare alleles in *Arabidopsis thaliana*. **BMC Evolutionary Biology**, v.16, n.1, p. 1-12, 2016.

MEYERS, L.A.; BULL, J.J. Fighting change with change: adaptive variation in an uncertain world. **Trends in Ecology & Evolution**, v.17, n.12, p. 551-557, 2002.

MICOLINO, R. **Uso de marcadores microssatélites da família Asteraceae para estudos genéticos em Carqueja (*Baccharis crispa* Spreng., Asteraceae). Dissertação (Mestrado em Biologia Evolutiva)**. Universidade Estadual do Centro-Oeste, Guarapuava, 2016.

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)? **Annals of forest science**, v. 65, n. 4, p. 1, 2008.

NAZARENO, A.G.; BEMMELS, J.B.; DICK, C.W.; LOHMANN, L.G. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. **Molecular Ecology Resources**, v. 17(6), p. 1136-1147, 2017.

NEI, M. Estimation of average heterozygosity and genetic distance from a small number of individuals. **Genetics**, Madison. v. 89, n. 3, p. 586-590, 1978.

NEI, M. *Molecular Evolutionary Genetics*, Columbia University Press: New York, NY, 1987.

NOSIL, P.; VILLOUTREIX, R.; DE CARVALHO, C.F.; FARKAS, T.E.; SORIA-CARRASCO, V.; FEDER, J.L.; CRESPI, B.J.; GOMPERT, Z. Natural selection and the predictability of evolution in *Timema* stick insects. **Science**, v. 359, n. 6377, p. 765-770, 2018.

NOVIČIĆ, Z.K.; SAYADI, A.; JELIĆ, M.; ARNQVIST, G. Negative frequency dependent selection contributes to the maintenance of a global polymorphism in mitochondrial DNA. **BMC evolutionary biology**, v. 20, n. 1, p. 20, 2020.

OLLERTON, J.; WINFREE, R.; TARRANT, S. How many flowering plants are pollinated by animals?. **Oikos**, v. 120, n. 3, p. 321-326, 2011.

PEDERSOLI, G.D.; GAGLIOTI, A.L.; TEIXEIRA, S.P. Floral development elucidates the formation of a tubular calyx and connate stamens in Cecropieae species (Urticaceae). *Plant Systematics and Evolution*, v. 306, n. 4, p.1-18, 2020.

PEAKALL, R.; SMOUSE, P.E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics*, v. 28, p. 2537-2539, 2012.

PILATI, L. **Estrutura Genética-Populacional de Embaúba dos Estados do Paraná, São Paulo e Rio de Janeiro. Trabalho de conclusão de curso.** Universidade Estadual do Centro-Oeste, Guarapuava, 2017.

PRIMACK, Richard. **Essentials Of Conservation Biology.** 4º Edição, Sinauer Associates. ISBN 978-0-87893-720-2, 2006.

PRITCHARD, J.K.; WEN, X.; FALUSH, D. Documentation for structure software. Version 2.3. **University of Chicago**, Chicago, USA. 2010.

RAJORA, O.P.; RAHMAN, M.H.; BUCHERT, G.P.; DANCİK, B.P. Microsatellite DNA analysis of genetic effects of harvesting in old- rowth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Molecular Ecology*, v. 9, n. 3, p. 339-348, 2000.

RIBAS-SANTOS, C.M.; SANTOS, M.; GUERRA, M.P. Analysis of S-incompatibility in *Campomanesia xanthocarpa* (Myrtaceae). *Acta Horticulturae*, 1231, p. 91-98, 2019.

RISTERUCCI, A.M.; DUVAL, M.F.; ROHDE, W.; BILLOTTE, N. Isolation and characterization of microsatellite loci from *Psidium guajava* L. *Molecular Ecology Notes*, v. 5, p. 745-748, 2005.

ROMANIUC NETO, S.; GAGLIOTI, A.L.; GUIDO, B.M.O. Urticaceae Juss. do Parque Estadual das Fontes do Ipiranga, São Paulo, SP, Brasil. *Hoehnea*, v.36, n.1, p. 193-205, 2009.

SALAZAR, D.M.; MELGAREJO, P.; MARTÍNEZ, R.; MARTÍNEZ, J.J.; HERNÁNDEZ, F.; BURGUERA, M. Phenological stages of the guava tree (*Psidium guajava* L.). *Scientia Horticulturae*, v. 108, n. 2, p. 157–161, 2006.

SANGUINETTI, C.J.; DIAS NETO, E; SIMPSON, A.J. Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques*, v. 17, n. 5, p. 914-921, 1994.

SCHROEDER, K.B.; SCHURR, T.G.; LONG, J.C.; ROSENBERG, N.A.; CRAWFORD, M.H.; TARSKAIA, L.A.; OSIPOVA, L.P.; ZHADANOV, S.I.; SMITH, D.G. A private allele ubiquitous in the Americas. *Biology Letters*, v. 3, n. 2, p. 218-223, 2007.

SHARMA, K.; MISHRA, A.K.; MISRA, R.S. A simple and efficient method for extraction of genomic DNA from tropical tuber crops. *African Journal of Biotechnology*, v. 7, p.1018-1022, 2008.

SLARKIN, M. Gene flow in natural populations. *Annual review of ecology and systematics*, v. 16, n. 1, p. 393-430, 1985.

SLATKIN, M. Gene flow and the Geographic Structure of Natural Populations. **Science**, v.236, p.787-92, 1987.

SLATKIN, M. Rare alleles as indicators of gene flow. **Evolution**, v. 39, n. 1, p. 53-65, 1985.

SOMMER, S. The importance of immune gene variability (MHC) in evolutionary ecology and conservation, **Frontiers in Zoology** v.2, n.16, 2005.

SURESH, K.; DEEPA, P.; HARISARANRAJ, R.; VAIRA A.V. Antimicrobial and Phytochemical Investigation of the Leaves of *Carica papaya* L., *Cynodon dactylon* (L.) Pers., *Euphorbia hirta* L., *Melia azedarach* L. and *Psidium guajava* L. **Ethnobotanical Leaflets**, v. 2008, n. 1, p. 157, 2008.

SPIEGEL, M.R.; SCHILLER, J.J.; SRINIVASAN, R.A; CARMONA, S.I.C. **Teoria e problemas de probabilidade e estatística**. 2º edição. Porto Alegre: Bookman, 2004.

SZPIECH, Z.A.; JAKOBSSON, M.; ROSENBERG, N.A. ADZE: a rarefaction approach for counting alleles private to combinations of populations. **Bioinformatics**, v.24, n.21, p. 2498-2504, 2008.

SZPIECH, Z.A.; ROSENBERG, N.A. On the size distribution of private microsatellite alleles. **Theoretical Population Biology**, v. 80, n. 2, p. 100-113, 2011.

TAKAYAMA, K.; LÓPEZ-SEPÚLVEDA, P.; KOHL, G.; NOVAK, J.; STUESSY, T.F. Development of microsatellite markers in species of *Erigeron* (Asteraceae) endemic to the Juan Fernández Archipelago, Chile. **American Journal of Botany**, p. 487-489, 2012.

TAUTZ D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, v. 17, p. 6463-6471, 1989.

TOWNSEND, C. R.; BEGON, M.; HARPER, J.L. **Fundamentos em Ecologia**. 3º edição. Porto Alegre: Artmed. p. 53-76, 2010.

VAN VALEN, L. A new evolutionary law. **Evolutionary Theory**, v.1, p. 1-30, 1973.

VIEIRA, M.L.C.; SANTINI, L.; DINIZ, A.L.; MUNHOZ, C.D.F. Microsatellite markers: what they mean and why they are so useful. **Genetics and molecular biology**, v. 39, n. 3, p. 312-328, 2016.

WAHL, L.M. Fixation when N and s vary: classic approaches give elegant new results. **Genetics**. v.188 (4), p.783-785, 2011.

WANG, A.; ZHANG, P.; LIU, X.; LIANG, J.; LI, W. Genetic structure and diversity of *Glehnia littoralis*, an endangered medicinal plant in China. **Biochemical Systematics and Ecology**, v. 66, p. 265-271, 2016.

WEBER, J. L.; MAY, P.E. Abundant class of human DNA polymorphisms which can be typed using the Polymerase Chain Reaction. **American Journal of Human Genetics**, v. 44, p. 388-396, 1989.

WEI, N.; DICK, C.W. Characterization of twenty-six microsatellite markers for the tropical pioneer tree species *Cecropia insignis* Liebm (Urticaceae). **Conservation Genetic Resources**, v. 6, n. 2, 2014.

WIECZOREK, A.M.; GEBER, M.A. Microsatellite loci for studies of population differentiation and range expansion in *Solidago sempervirens* L. (Asteraceae). **Molecular Ecology Notes**, n. 2, p. 554-556, 2002.

WRIGHT, S. **Evolution and the Genetics of Populations: Variability within and among natural populations**. University of Chicago Press: Chicago, USA, vol. 4, 1978.

YEH, F.C.; YANG, R.C.; BOYLE, T. POPGENE: Microsoft Windows-based freeware for population genetic analysis. Version 1.31. **University of Alberta, Canada and Centre for International Forestry Research**, Canada, 1999.