

UNIVERSIDADE ESTADUAL DO CENTRO-OESTE

**TAMANHO EFETIVO E FLUXO GÊNICO EM
POPULAÇÕES DE *Acrocomia aculeata* (Jacq.) Lodd.
ex Mart. USANDO TEORIA DA COALESCÊNCIA**

DISSERTAÇÃO

RENAN MARCELO PORTELA

IRATI-PR

2019

RENAN MARCELO PORTELA

TAMANHO EFETIVO E FLUXO GÊNICO EM POPULAÇÕES DE *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. USANDO TEORIA DA COALESCÊNCIA

Dissertação apresentada à Universidade Estadual do Centro-Oeste, como parte das exigências do Programa de Pós-Graduação em Ciências Florestais, área de concentração Manejo Sustentável de Recursos Florestais, área de atuação Silvicultura, para obtenção do título de Mestre em Ciências Florestais.

Prof. Dr. Evandro Vagner Tambarussi
Orientador

Prof. Dr. João Ricardo Bachega Feijó Rosa
Coorientador

Prof. Dra. Ananda Virgínia de Aguiar
Coorientadora

IRATI-PR

2019

Catálogo na Fonte
Biblioteca da UNICENTRO

P843t	<p>PORTELA, Renan Marcelo. Tamanho efetivo e fluxo gênico em populações de <i>Acrocomia aculeata</i> (Jacq.) Lodd. ex Mart. usando teoria da coalescência / Renan Marcelo Portela. – Irati, PR : [s.n.], 2019. 98f.</p> <p>Orientador: Prof. Dr. Evandro Vagner Tambarussi Coorientador: Prof. Dr. João Ricardo Bachega Feijó Rosa Coorientadora: Prof. Dra. Ananda Virgínia de Aguiar Dissertação (mestrado) – Programa de Pós-Graduação em Ciências Florestais. Área de concentração Manejo Sustentável de Recursos Florestais. Área de atuação Silvicultura. Universidade Estadual do Centro-Oeste, PR.</p> <p>1. Engenharia Florestal – dissertação. 2. Genética de populações. 3. Macaúba – palmeira. 4. Melhoramento – migração. I. Tambarussi, Evandro Vagner. II. Rosa, João Ricardo Bachega Feijó. III. Aguiar, Ananda Virgínia de. IV. UNICENTRO. V. Título.</p>
-------	--

CDD 634.92

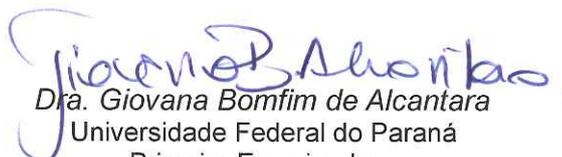
TERMO DE APROVAÇÃO

Defesa Nº 124

Renan Marcelo Portela

**“Tamanho efetivo e fluxo gênico em populações de *Acrocomia aculeata* (Jacq.)
Lodd. ex Mart. usando teoria da coalescência”**

Dissertação aprovada em 26/02/2019 como requisito parcial para obtenção do grau de Mestre, no Programa de Pós-Graduação em Ciências Florestais, área de concentração em Manejo Sustentável de Recursos Florestais, da Universidade Estadual do Centro-Oeste, pela seguinte Banca Examinadora:


Dra. Giovana Bomfim de Alcantara
Universidade Federal do Paraná
Primeira Examinadora


Dra. Viviane Fernandes de Souza
Universidade Estadual do Centro-Oeste
Segunda Examinadora


Dr. Evandro Vagner Tambarussi
Universidade Estadual do Centro-Oeste
Orientador e Presidente da Banca Examinadora

Irati - PR
2019

AGRADECIMENTOS

A toda minha família, em especial para minha querida avó Daluz, pela educação, ensinamentos e todos os momentos felizes que ela me proporcionou, do meu nascimento até agora.

Aos meus pais, Valquíria e Celso, por sempre me apoiarem e incentivarem, pelos sacrifícios feitos para viabilizar minha formação acadêmica.

A Bruna Strappazon, por estar ao meu lado todo esse tempo e me apoiar incondicionalmente.

Ao meu orientador, Prof. Dr. Evandro Vagner Tambarussi, por todas as oportunidades, orientações, aulas, incentivos, amizade, paciência e por sempre acreditar no meu trabalho e na minha capacidade.

A minha co-orientadora, Prof. Dra. Ananda Virgínia de Aguiar, agradeço imensamente pela amizade, orientações, conselhos e por todo conhecimento compartilhado, além das oportunidades de desenvolver e expandir meus horizontes acadêmicos.

Ao meu co-orientador, Prof. Dr. João Ricardo Bachega Feijó Rosa, pelas orientações, explicações, disponibilidade, amizade e por mostrar o quão fascinante é a teoria da Coalescência.

Aos meus amigos e colegas da Unicentro, Armando, Rafael, Rayssa, Richeliel, Emilio, Thiago, Silvio, Alexandre, Willian, Pedro, Fabio, João Gabriel, Matheus, Bell, Dani, Caio e demais não nominados. Pelos momentos de descontração, amizade e apoio.

As amizades e parcerias que fiz nesses incríveis anos de aprendizado.

Aos meus amigos de Rebouças pela força e pelos momentos de felicidade e companheirismo que me propiciaram.

Aos professores que passaram por minha formação, desde a base até agora, minha eterna gratidão.

A Unicentro, a Adriana do PPGF, além de todos os funcionários que tive oportunidade de conviver nesse período.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes), pela concessão da bolsa, e a todos que de alguma forma contribuíram para a realização deste trabalho.

"O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de financiamento 001".

SUMÁRIO

SUMÁRIO.....	4
RESUMO	7
ABSTRACT	8
1. INTRODUÇÃO	9
2. OBJETIVOS.....	12
3. REVISÃO BIBLIOGRÁFICA.....	13
3.1. <i>Acrocomia aculeata</i> (Jacq.) Lodd. ex Mart.	13
3.2. Marcadores microssatélites	15
3.2.1. Mutação em microssatélites e modelos teóricos para estimativas	16
3.3. Uso da genética de populações na conservação de recursos vegetais	18
3.4. Teoria da Coalescência e Genética de Populações	20
3.5. Tamanho efetivo populacional.....	23
3.6. Fluxo gênico	24
3.6.1. Estimação do fluxo gênico	25
4. MATERIAL E MÉTODOS	31
4.1. Local de Estudo e amostragem	31
4.2. Extração de DNA, reações de PCR e checagem de dados.....	32
4.3. Fluxo gênico e tamanho efetivo via coalescência.....	32
5. RESULTADOS	35
5.1. Ajuste dos dados para frequência de alelos nulos.....	35
5.2. Fluxo gênico e tamanho efetivo via coalescência.....	36
6. DISCUSSÃO	40
6.1. Ajuste dos dados para frequência de alelos nulos.....	40
6.2. Fluxo gênico e tamanho efetivo populacional pela coalescência	40
7. CONCLUSÃO	46
8. REFERÊNCIAS	47

LISTA DE FIGURAS

- Figura 1:** Características fenotípicas genéricas da palmeira *A. aculeata* (macaúba), folhas e estipe (A), inflorescência e espata (B); cachos de frutos (C)..... 14
- Figura 2:** Esquema referente a estatísticas F e as relações entre os níveis hierárquicos populacionais. 28
- Figura 3:** Representação gráfica da distribuição espacial dos genótipos coletados nas localidades de Teodoro Sampaio I (TSI), Teodoro Sampaio II (TSII), Amparo (AP) e Rosana (RA) em que *A. aculeata* foram amostradas no Estado de São Paulo. 31
- Figura 4:** Esquema dos modelos demográficos e de fluxo gênico que foram utilizados entre as quatro subpopulações de *A. aculeata* utilizadas nesse estudo, ambas no Estado de São Paulo, Brasil. 33
- Figura 5:** Estimativas do número efetivo de migrantes por geração (N_m), obtidas para espécie *A. aculeata* nas localidades Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana, no Estado de São Paulo, para cada loco utilizado no estudo. 38

LISTA DE TABELAS

Tabela 1: Matriz com as distâncias geográficas (Km) entre as localidades amostradas para espécie <i>A. aculeata</i> nas populações de Teodoro Sampaio I (TSI), Teodoro Sampaio II (TSII), Amparo (AP) e Rosana (RA).	32
Tabela 2: Estimativa das frequências de alelos nulos utilizando o estimador de van Oosterhout, considerando um intervalo de confiança de 95% para cada loco individualmente, em quatro subpopulações amostradas no Estado de São Paulo, Brasil.	35
Tabela 3: Estimativas das taxas de migração, escalonadas pela taxa de mutação (M_{ij}), obtidas para a espécie <i>A. aculeata</i> envolvendo as localidades TSI, TSII, RA e AP (Estado de São Paulo), considerando a média de todos os locos microssatélites utilizados.	36
Tabela 4: Estimativas do número efetivo de migrantes por geração (N_{mji}), para espécie <i>A. aculeata</i> nas localidades Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana, no Estado de São Paulo, Brasil, e considerando o intervalo de confiança de 95%.	37
Tabela 5: Estimativas do tamanho efetivo populacional, do intervalo de confiança (95%) e ($\theta_i = 4N_e\mu$) para espécie <i>A. aculeata</i> nas quatro localidades estudadas no Estado de São Paulo, considerando todos os locos utilizados.	39

RESUMO

Acrocomia aculeata é uma palmeira nativa da América Central e do Sul, que possui significativo potencial econômico, social e ambiental. Possui diversas aplicações, sendo usada principalmente como fonte de matéria-prima para biocombustíveis, bem como nas indústrias alimentícias, farmacêuticas e cosmética. Entretanto, os avanços territoriais do agronegócio, juntamente com a expansão urbana e outros fatores que causam a fragmentação e destruição de seu *habitat*, acabam levando à perda de genes e genótipos em populações locais de *A. aculeata*. Esses processos restringem a variabilidade genética, afetando a produtividade e a capacidade das populações de *A. aculeata* em responder às mudanças climáticas e outros estresses bióticos e abióticos. Dessa forma, estes processos podem comprometer a perpetuação não só da espécie, mas também da sua flora e fauna dependentes. Neste contexto, o objetivo deste estudo foi estimar o fluxo gênico e o tamanho efetivo populacional utilizando uma abordagem baseada na teoria da coalescência. Quatro subpopulações localizadas nos municípios de Teodoro Sampaio (TSI e TSII), Rosana (RA) e Amparo (AP), no Estado de São Paulo, foram avaliadas com nove locos microssatélites. Para estimar o tamanho efetivo e o fluxo gênico, utilizou-se o método de inferência bayesiana, baseado na teoria de coalescência implementada no *software* MIGRATE-N. O tamanho efetivo populacional (N_e) foi obtido assumindo uma taxa de mutação de 5×10^{-5} . O fluxo gênico (Nm) entre os pares de populações variou de 0,28 a 1,17, com níveis mais altos de migração entre os três locais geograficamente próximos (TSI, TSII e RA). As estimativas do tamanho efetivo populacional (N_e) foram 444, 835, 838 e 874 para AP, TSII, RA e TSI, respectivamente. Os resultados obtidos agregam conhecimento acerca da compreensão da genética de populações de *A. aculeata* e mostram que alguns métodos tradicionais podem ser insatisfatórios para fazer inferências sobre o fluxo gênico e processos evolutivos históricos.

PALAVRAS-CHAVE: macaúba; genética de populações; conservação; melhoramento; migração.

ABSTRACT

Acrocomia aculeata is a tropical palm tree native to Central and South America, which has significant economic, social, and environmental potential. It has several applications and is used mainly as a source of raw material for biofuels, as well as in the culinary, pharmaceutical, and cosmetics industries. However, land encroachment due to the expansion of agribusiness, together with urban growth and other factors have resulted in the fragmentation and destruction of its habitat, leading to the loss of genes and genotypes in local *A. aculeata* populations. These deleterious genetic processes restrict genetic variability, affecting productivity, and the capacity of *A. aculeata* populations to respond to climate change and other biotic and abiotic stresses. As such, these processes can compromise the continuation of not only the species, but also its dependent flora and fauna. In this context, the objective of this study was to estimate the gene flow and effective population size using an approach based on coalescence theory. Four subpopulations located in the municipalities of Teodoro Sampaio (TSI and TSII), Rosana (RA), and Amparo (AP), São Paulo State, Brazil, were evaluated with nine microsatellite loci. In order to evaluate the effective size and gene flow, the Bayesian inference method was used, based on the coalescence theory implemented in the MIGRATE-N software. The effective population size (N_e) was obtained considering a mutation rate of 5×10^{-5} . Gene flow (Nm) for pairs of populations ranged from 0.28 to 1.17, with higher levels of migration between the three geographically proximal locations (TSI, TSII, and RA). The estimates of effective population size (N_e) were 444, 835, 838, and 874 for AP, TSII, RA, and TSI, respectively. The results obtained add to our understanding of *A. aculeata* population genetics and show that some traditional methods may be unsatisfactory in making inferences about gene flow and historical evolutionary processes.

KEYWORDS: macaúba; population genetics; conservation; genetic improvement; migration.

1. INTRODUÇÃO

A espécie *Acrocomia aculeata* (Jacq.) Lodd. ex Mart. é uma palmeira arborescente nativa das Américas Central e do Sul (LORENZI, 2006; CARVALHO et al., 2011), sendo considerada a palmeira de maior dispersão geográfica no Brasil (LORENZI, 2006), tendo sua ocorrência confirmada nas regiões Centro-Oeste, Norte, Nordeste, Sudeste e Sul (SCARIOT et al., 1995; LEITMAN et al., 2015).

Como característica ecológica, *A. aculeata* possui hábito pioneiro, ocupando principalmente áreas abertas e com grande incidência solar, ocorrendo, em grande parte, de forma agrupada e em grandes maciços. Suas flores são unissexuais, com flores masculinas no topo e femininas na base, possuindo dicogamia protogínica, de maneira que as flores masculinas amadurecem após as femininas (BRITO, 2013). O sistema reprodutivo da *A. aculeata* é o misto (COELHO, et al 2018), preferencialmente alógama (SCARIOT et al., 1991; BRITO, 2013). A polinização da *A. aculeata* é realizada principalmente por besouros e pelo vento, e secundariamente por abelhas, sendo polinizadores de curta distância. Entretanto, a dispersão das sementes ocorre por grandes mamíferos e aves (POTT; POTT, 1994; SCARIOT et al., 1998).

A. aculeata apresenta grande potencial econômico, social e ambiental, possuindo várias aplicações e usos, principalmente como fonte de matéria-prima para biocombustíveis, alimentos, fármacos e cosméticos (COLOMBO, 2017). Além disso, considerando que *A. aculeata* é uma espécie nativa de alta rusticidade, o seu potencial se estende para ocupação de áreas marginais degradadas ou destinadas à preservação e possibilidade de integração com sistemas agroflorestais (COELHO, 2018).

Levando-se em consideração que *A. aculeata* é uma espécie tida como não domesticada sem programas de melhoramento genético bem desenvolvidos, além do cultivo ser feito com pouca ou nenhuma tecnologia, sua produtividade de óleo é elevada. A produção de óleo pode chegar a 5.000 kg/ha (MOTA, 2011). Isso se torna evidente principalmente em comparação com culturas anuais tradicionais, onde o melhoramento genético e o domínio tecnológico já estão avançados, como soja (450 kg/ha) e girassol (700 kg/ha). Há, ainda, a vantagem de *A. aculeata* ser uma espécie perene (COLOMBO, 2017).

As áreas de ocorrência natural da espécie no Estado de São Paulo eram amplas, porém, foram gradativamente substituídas por cultivos de café, cana-de-açúcar e pastagens (NOVAES, 1952; AZEVEDO FILHO, 2012). Dessa maneira, os avanços

territoriais do agronegócio, juntamente com a expansão urbana e outros fatores que podem gerar fragmentação e destruição de biomas, podem fazer com que *A. aculeata* fique sujeita à perda de genes e genótipos presentes em populações locais, restringindo a variabilidade genética e levando à deriva genética das populações.

As alterações nas frequências alélicas e genotípicas, além de afetarem a produtividade e, podem, ainda, alterar a capacidade de *A. aculeata* em responder às mudanças climáticas e aos demais estresses bióticos e abióticos, podendo comprometer a perpetuação da espécie e de sua fauna e flora dependentes da espécie (PRIMACK; RODRIGUES, 2001). Nesse contexto, pesquisas que possam gerar informações sobre adequadas estratégias de conservação genética para a espécie são de grande importância, principalmente por essa espécie possuir grande potencial socioeconômico e seus *habitats* estarem sendo reduzidos. Além disso, essas informações também podem ser importantes para o estabelecimento de programas de melhoramento genético, que ainda não estão disponíveis para esta espécie.

A diversidade genética e a estrutura populacional em espécies arbóreas têm sido normalmente investigadas com base em várias metodologias, tais como as estatísticas de Wright (1951,1978) e de Nei (1973), inferência bayesiana (PRITCHARD et al., 2000), componentes principais (CRUZ et al., 2011) e agrupamentos hierárquicos clássicos (ODONG et al., 2011). A utilização dessas abordagens serve principalmente para estimar a variabilidade genética existente entre e dentro de populações e como esta encontra-se estruturada no espaço.

De modo geral, a maioria destes métodos estatísticos utiliza as frequências alélicas estimadas para capturar níveis de diferenciação entre populações. Entretanto, esses níveis de diferenciação podem ser o resultado de eventos de fluxo gênico recentes, de modo a não considerar eventos migratórios ocorridos no passado (ROSA, 2016; WILSON; RANNALA, 2003). Consequentemente, acaba dificultando o entendimento das relações históricas e do passado evolutivo das populações. Portanto, no contexto deste trabalho, uma visão genealógica do fluxo gênico e da estrutura populacional de *A. aculeata* é de grande importância (ROSA, 2016). Possivelmente, sendo utilizada para nortear as amostragens de sementes, pólen e propágulos para conservação genética, bem como para definir as populações prioritárias para a conservação.

Uma visão genealógica de eventos populacionais pode ser alcançada com base na teoria da coalescência (KINGMAN, 1982 a, b, c). Esta abordagem faz uso de modelos retrospectivos baseados na genealogia dos alelos, de modo a rastrear os alelos de um gene

compartilhado por todos os membros amostrados da população atual até uma única cópia ancestral, conhecida como ancestral comum mais recente da amostra (ACMR). Assim, é possível considerar, por exemplo, eventos de mutação e fluxo gênico que podem ter ocorrido há muitas gerações entre populações de determinada espécie, fornecendo estimativas históricas e mais precisas de tais eventos. Diante da necessidade de estudos nessa área, a proposta desta pesquisa foi a de utilizar a abordagem baseada em coalescência para estimar as taxas de mutação e fluxo gênico de populações de *A. aculeata*, de modo a verificar as relações histórico-evolutivas de tais populações. Essa abordagem é tida por vários cientistas como a moderna (PALCZEWSKI; BEERLI, 2013) e mais popular (MARJORAM; TAVARÉ, 2006) abordagem em genética de populações, a qual vem revolucionando o entendimento acerca dos processos histórico-evolutivos envolvidos (HOLSINGER; WEIR, 2009).

Espera-se colaborar com o entendimento de processos como o fluxo gênico em populações, estudos de diversidade genética e tamanho efetivo populacional da *A. aculeata*. Desta forma, essa pesquisa tem como base a obtenção de subsídios para a utilização sustentável e correto manejo dos recursos gerados por *A. aculeata*, assim como para o melhoramento genético desta espécie.

Hipóteses e questões acerca do trabalho:

- Ocorre fluxo gênico entre as localidades geograficamente próximas?
- Qual é a magnitude desse fluxo? Caso haja, este fluxo é simétrico? O número de migrantes é suficiente para evitar a divergência entre as subpopulações?
- Quais as implicações para conservação?

2. OBJETIVOS

Objetivo Geral

Estimar parâmetros genéticos populacionais para *Acrocomia aculeata* (Jacq.) Lodd. ex Mart baseados na teoria da coalescência.

Objetivos Específicos

Estimar índices de diversidade em populações naturais de *Acrocomia aculeata* (Jacq.) Lodd. ex Mart;

Estimar o tamanho efetivo populacional (N_e) e fluxo gênico entre tais populações.

3. REVISÃO BIBLIOGRÁFICA

Nessa revisão são abordados tópicos sobre os aspectos relativos a importância econômica, social e ambiental da *A. aculeata*, bem como suas características botânicas, ecológicas e geográficas; Logo após é abordada a temática referente aos marcadores genéticos, breves aspectos históricos relacionados, principais tipos, dando ênfase aos microssatélites, apresentando de suas características gerais, dinâmica mutacional e modelos teóricos propostos para estimar mutação; Na sequência é discorrido acerca do uso da genética de populações na conservação de recursos vegetais, com foco em sua importância na preservação e manutenção dos recursos genéticos; Aborda-se também sobre a teoria da coalescência e genética de populações, mostrando o foco e o interesse da genética de populações, sua importância, métodos usuais empregados nas estimativas de parâmetros, mostrando noções importantes acerca da teoria da coalescência, bem como aspectos relativos ao desenvolvimento e utilização dessa teoria; Por fim, são abordados dois parâmetros importantíssimos para fins de conservação, sendo o tamanho efetivo populacional e o fluxo gênico. Em ambos são abordadas as bases conceituais, histórico e importância, além de subtópico onde aborda-se os métodos usuais de estimativa de fluxo gênico.

3.1. *Acrocomia aculeata* (Jacq.) Lodd. ex Mart.

A espécie *Acrocomia aculeata* (Jacq.) Lodd. ex Mart., popularmente conhecida como macaúba, é uma palmeira arborescente da família Arecaceae, nativa das Américas Central e do Sul (LORENZI, 2006; CARVALHO et al., 2011). Sua altura varia entre 10 e 20 metros de comprimento, tendo estipe reta e sem ramificações, com diâmetro variando entre 25 e 50 centímetros e espinhos escuros e finos revestindo a região de nós caulinar. Suas folhas são pinadas e compostas, tendo, em média, de 4 a 5 metros de comprimento, que dão um aspecto plumoso à copa, por estarem ordenadas em diferentes planos (CARVALHO, 2008).

No Brasil, a espécie ocorre nos domínios fitogeográficos da Mata Atlântica, Amazônia, Caatinga e Cerrado, tendo sua ocorrência registrada nas regiões Norte (Amazonas, Pará, Roraima e Tocantins), Nordeste (Bahia, Ceará, Maranhão, Pernambuco e Piauí), Centro-Oeste (Distrito Federal, Goiás, Mato Grosso do Sul e Mato Grosso),

Sudeste (Minas Gerais, Rio de Janeiro e São Paulo) e Sul (Paraná) (SCARIOT et al., 1995; LEITMAN et al., 2015), sendo considerada a palmeira de maior dispersão geográfica no Brasil (LORENZI, 2006). Tem hábito de crescimento pioneiro, ocupando principalmente áreas abertas e com grande incidência solar, ocorrendo em grande parte de forma agrupada e em grandes maciços. Apesar de ocorrer em solos arenosos e com pouca umidade, desenvolve-se melhor em solos férteis (LORENZI, 2006).

Figura 1: Características fenotípicas genéricas da palmeira *A. aculeata*, folhas e estipe (A), inflorescência e espata (B); cachos de frutos (C).



Fonte: Retirado de Coelho (2017)

Suas inflorescências são em forma de espádice, interfoliares e de coloração amarelo claro, que ficam agrupadas em cachos pendentes (SODRÉ, 2005). As flores são diclinas, estaminadas no topo e pistiladas na base, possuindo dicogamia protogínica, sendo que as flores estaminadas amadurecem após as pistiladas (BRITO, 2013). A polinização da *A. aculeata* é feita principalmente por besouros e pelo vento, e secundariamente por abelhas, sendo polinizadores de curta distância. Entretanto, a

dispersão da semente ocorre por grandes mamíferos e aves (POTT; POTT, 1994; SCARIOT et al., 1998).

O sistema reprodutivo da *A. aculeata* é do tipo misto, ocorrendo tanto por alogamia quanto por autogamia, sendo que a polinização cruzada corresponde a uma significativa porcentagem na produção de frutos, porém há autocompatibilidade (SCARIOT et al., 1991; BRITO, 2013). Os frutos são marrons amarelados, com o formato de drupa globosa e esféricos (SCARIOT et al., 1991).

A. aculeata tem grande potencial econômico, social e ambiental, possuindo várias aplicações e usos, principalmente como fonte de matéria-prima na produção de biocombustíveis, alimentos, fármacos e cosméticos. Além disso, considerando que *A. aculeata* é uma espécie nativa de alta rusticidade, o seu potencial se estende para ocupação de áreas marginais degradadas ou destinadas à preservação e possibilidade de integração com sistemas agroflorestais (COELHO et al., 2018).

Apesar de ser uma espécie considerada não domesticada, não haver programas de melhoramento bem desenvolvidos e seu cultivo ser feito com pouca ou nenhuma tecnologia, sua produtividade em óleo é elevada (média de 5.000 kg/ha) (MOTA, 2011), sendo maior que culturas anuais tradicionais, onde o melhoramento genético e o domínio tecnológico já estão avançados, tais como soja (450 kg/ha) e girassol (700 kg/ha), havendo ainda a vantagem de *A. aculeata* ser uma espécie perene (AZEVEDO FILHO et al., 2012), onde os gastos anuais com operações de plantio, preparo do solo, mão-de-obra e manutenção são reduzidos em relação a culturas anuais, além de produzir mesmo com chuvas irregulares ou atrasadas sem redução drástica da produtividade, permanecer de 30 a 40 anos sem precisar de replantio, evitando práticas mais intensivas de cultivo, e ainda provendo serviços ecossistêmicos (BIODIESELBR, 2011; COLOMBO, 2017; FAQUIN, 2001; GUREVITCH et al., 2009).

3.2. Marcadores microssatélites

Os marcadores microssatélites são sequências repetidas constituídas de 1 a 6 nucleotídeos que se repetem lado a lado em forma de tandem. Esse tipo de marcador é amplamente usado em análises genéticas, em uma vasta gama de seres vivos, sobretudo, devido ao seu elevado grau de informações e transferibilidade (BORÉM; CAIXETA, 2009).

Os microssatélites são usados em espécies florestais para diversos propósitos, tais como: estudos relacionados a análises de parentesco, sistema de reprodução, diversidade genética e estudos evolutivos, principalmente por serem sequências com alta taxa evolutiva, que, mesmo em comparações de genótipos com estreita base genética, geralmente é possível detectar um elevado número de alelos em um único loco (SELKOE; TOONEN, 2006).

Os polimorfismos são em decorrência das diferenças no número de vezes em que os nucleotídeos se repetem em determinado loco, podendo ser mono-, di-, tri-, tetra-, penta- e hexa-nucleotídeos. Os microssatélites podem ainda ser classificados de acordo com as sequências repetidas, podendo ser (i) perfeita (quando as sequências não são interrompidas por bases que não sejam do motivo), (ii) imperfeita (quando ocorre a presença de alguma base alheia ao motivo), (iii) interrompida (ocorre de forma intercalada uma pequena sequência alheia ao motivo), e (iv) composta (ocorrem duas sequências repetidas de forma adjacente) (BUSO, 2003). Essas variações geram diferenças nos comprimentos dos segmentos, as quais podem ser detectadas pela reação de polimerase em cadeia (PCR), visualizadas especificamente na etapa de separação de fragmentos amplificados no gel de eletroforese (TURCHETTO-ZOLET et al., 2017)

Os microssatélites têm como vantagem o fato de serem marcadores codominantes (diferenciam entre indivíduos homozigotos e heterozigotos), multialélicos, baseados em PCR, estarem distribuídos de forma ampla no genoma e necessitarem de pequenas quantidades de DNA coletado dos indivíduos a serem analisados (BUSO, 2003).

3.2.1. Mutação em microssatélites e modelos teóricos para estimativas

Diversos mecanismos mutacionais para microssatélites vêm sendo sugeridos e propostos, de modo a buscar explicar e elucidar a alta taxa de mutação, citando-se os erros durante a recombinação; *crossing-over* desigual; deslizamento da polimerase durante a replicação ou reparo do DNA (STRAND et al., 1993).

Em relação à inclusão de erros durante a recombinação, cepas de *E. coli*, com ou sem sistema de recombinação funcional, apresentaram uma taxa de mutação semelhante, sugerindo que a recombinação não é o mecanismo predominante na geração de variabilidade de microssatélites (LEVINSON; GUTMAN, 1987).

Sobre o *crossing-over* desigual, quando ocorre, pode acarretar mudanças acentuadas, como a perda ou o ganho de um grande número de repetições. Isso ocorre porque quando regiões repetitivas de microssatélites estão presentes, um grampo pode ser formado durante a sinapse, o que significa que apenas partes, geralmente de comprimento diferente, de cada cromossomo, serão trocadas, e um cromossomo receberá um maior fragmento por causa do maior número de repetições de microssatélites trocados, enquanto o cromossomo homólogo recebe um número menor de repetições (OLIVEIRA et al., 2006).

No caso da replicação ou reparo do DNA, pode ocorrer deslizamento (*slippage*) da DNA polimerase, onde uma fita de DNA dissocia-se temporariamente da outra e revincula-se rapidamente em uma posição diferente da anterior, ocasionando erros de emparelhamento de bases, podendo ocorrer adição ou perda de sequências (OLIVEIRA et al., 2006).

A escolha do modelo mutacional teórico que servirá de base para se obter corretamente os parâmetros genéticos populacionais a serem estimados a partir de dados de microssatélites é muito importante. Modelos mutacionais são utilizados para estimar o número esperado de alelos em uma população a partir da heterozigosidade observada e nas análises estatísticas da variabilidade genética. Entretanto, de forma geral, os modelos conhecidos apresentam desvantagens quando aplicados a dados de microssatélites (OLIVEIRA et al., 2006). Os mesmos autores argumentam que quatro modelos mutacionais podem ser utilizados, tais como o modelo de alelos infinitos (*Infinite alleles model*), mutação passo-a-passo (*Stepwise mutation model*), modelo de duas fases (*Two phase model*) e o modelo de K-alelos (*K-alleles model*).

O modelo de alelos infinitos (*Infinite alleles model*) considera que cada mutação gera aleatoriamente um novo alelo (KIMURA; CROW, 1964). Logo, ao considerar este modelo em marcadores microssatélites, é possível assumir que as mutações alteram o número de repetições que ocorrem em tandem (KIMURA; CROW, 1964). Este é o modelo clássico proposto por Wright (1931), com o qual são utilizadas as estatísticas-*F* (OLIVEIRA et al., 2006).

O modelo de mutação passo-a-passo (*Stepwise mutation model*) considera que quando um loco microssatélite sofre mutação, ele ganha ou perde uma repetição (KIMURA; OHTA, 1978). Consequentemente, implica que dois alelos que diferem por apenas um motivo são mais relacionados (isto é, compartilham um ancestral comum mais recente) do que os alelos que diferem por várias repetições (KIMURA; OHTA, 1978).

Slatkin (1993), propôs uma medida de diferenciação genética R_{ST} semelhante ao F_{ST} (WRIGHT, 1951) e ao G_{ST} (NEI, 1973), porém, com base no modelo de mutação passo-a-passo. Este modelo é geralmente preferido ao estimar relações genéticas entre indivíduos e estrutura populacional, exceto na presença de homoplasia (quando dois alelos são idênticos por estado, mas não por descendência) (OLIVEIRA et al., 2006). A homoplasia pode influenciar de forma acentuada os estudos populacionais que envolvem altas taxas de mutação e grandes tamanhos populacionais, juntamente com fortes restrições de tamanho dos alelos (ESTOUP et al., 2002).

O modelo denominado de *Two phase model* (DI RIENZO et al. 1994) foi proposto como uma extensão do modelo de mutação passo-a-passo para estudos em microssatélites, onde postula que a maioria dos eventos mutacionais resulta em aumento ou diminuição de uma unidade de repetição, embora também ocorram alterações infrequentes de um grande número de repetições (DI RIENZO et al. 1994).

Para o modelo *K-alleles model* todos os alelos são considerados como equivalentes, ou seja, a probabilidade de ocorrer mutação de um alelo para qualquer outro alelo é igualmente provável (CROW; KIMURA, 1970). Isso significa que se houver exatamente K alelos possíveis em um determinado loco, a probabilidade de um dado alelo “mutacionar” em qualquer outro é dado por $[\mu / (K-1)]$, onde μ é a taxa de mutação para qualquer um dos $(K-1)$ estados alélicos (ESTOUP et al., 2002; OLIVEIRA et al., 2006).

3.3. Uso da genética de populações na conservação de recursos vegetais

Ecossistemas florestais representam uma fonte imensurável de recursos genéticos potenciais a humanidade, seja isso de forma direta ou indireta. Entretanto, grande parte desses recursos vêm sendo destruídos de modo irreversível, exigindo medidas urgentes a serem tomadas para sua conservação (FRANKHAM, 2002; KAGEYAMA, 1987).

A fragmentação do ambiente inclui a erosão da variabilidade genética e o aumento da divergência entre as populações por meio da deriva genética, a elevação dos níveis de endogamia dentro das populações, a diminuição do fluxo gênico interpopulacional e uma maior probabilidade de extinção de fragmentos locais (YOUNG et al., 2001). Nesse sentido, conhecer e caracterizar os recursos genéticos, por meio de análises genéticas moleculares, são essenciais para elucidar aspectos relevantes da biologia e evolução das

espécies, promovendo o manejo e a conservação desses recursos (HODGKIN et al., 2001; RAO; HODGKIN, 2002).

Os recursos genéticos são definidos como parte da biodiversidade que tem previsão de uso atual e/ou potencial, sendo sua conservação uma medida estratégica para a segurança alimentar, econômica e ecológica da humanidade. Isso porque é a partir desses recursos genéticos que dependerá a sobrevivência das gerações futuras e interesses econômicos das nações (ALHO, 2012; GARCIA, 1995). Desta forma, há uma grande preocupação quanto à manutenção, avaliação e troca da diversidade genética em âmbito mundial (NASS, 2007).

A conservação genética de uma espécie é um processo técnico que deve estar fundamentado em bases científicas sólidas, que possibilitem a manutenção da variabilidade genética dos genótipos em condições especiais, seja para fins de preservação ou melhoramento genético. Baseando-se nisso, são descritas várias formas de conservação e estudos populacionais. Estas são a base para a definição de estratégias de conservação e escolha de áreas prioritárias para manejo (EIZIRIK, 1996).

A estratégia de conservação adotada dependerá de uma série de fatores, tais como a natureza do material, o objetivo e o alcance da conservação (NASS et al., 2007). A natureza do material diz respeito à duração do ciclo total, ao modo de reprodução, ao tamanho dos indivíduos e se o material é domesticado ou não (NASS et al., 2007). Deve-se ainda considerar o tempo (curto, médio e longo prazos) e os locais onde tal conservação será realizada (NASS et al., 2007).

Para definir a melhor estratégia de conservação, os princípios da genética de populações são fundamentais (FRANKHAM, 2002; 2014). A genética de populações baseia-se nos alelos e em suas frequências contidos em uma ou mais populações ao longo de gerações, levando-se em consideração forças evolutivas como seleção natural, deriva genética, mutação, recombinação e fluxo gênico, permitindo inferir sobre a distribuição da variabilidade genética no espaço e no tempo (nos diferentes níveis hierárquicos populacionais). Esta inferência está relacionada a processos de estimação de parâmetros como diversidade, fluxo gênico, taxas de cruzamento, tamanho efetivo, parentesco, endogamia, entre outras informações importantes sobre as populações em questão, que servirão de base para o entendimento das relações histórico-evolutivas envolvidas (HARTL, 2010).

Fazendo uso dos conceitos e das teorias da genética de populações, é possível, por exemplo, definir as melhores estratégias para minimizar a probabilidade de extinção de

uma determinada espécie, considerando possíveis problemas associados ao reduzido tamanho efetivo, aos efeitos deletérios da endogamia, à redução da diversidade e da capacidade de evolução em resposta a estresses bióticos e mudanças no ambiente, além de efeitos deletérios decorrentes do cruzamento entre indivíduos muito distintos geneticamente (depressão por exogamia - em alguns casos, devido ao aumento acentuado da heterozigosidade, ocorre a depressão de determinada característica) (FRANKHAM et al., 2002).

Por meio da genética de populações, é possível estudar o efeito da fragmentação populacional e o impacto no fluxo gênico entre populações estruturadas, a redução da diversidade genética e o efeito do acúmulo de mutações deletérias ou letais (FRANKHAM, 2002). Por outro lado, também é possível estudar os problemas relacionados às incertezas taxonômicas, estabelecer espécies e populações prioritárias para conservação e definir unidades evolutivamente significativas e unidades de manejo, idealizando preservar os recursos genéticos e a biodiversidade (MORITZ, 2002).

3.4. Teoria da Coalescência e Genética de Populações

Historicamente, métodos estatísticos baseados em frequências alélicas foram usados para estimar os parâmetros genéticos populacionais a partir de conjuntos de dados (ROSA, 2016). Esses métodos foram aprimorados e se expandiram em diversas variantes que foram aceitos e debatidos por muitos pesquisadores, citando-se as estatísticas F de Wright (1951, 1978), as de Nei (1973) e suas extensões (NATH; GRIFFITHS, 1996), inferência bayesiana STRUCTURE (PRITCHARD et al., 2000), componentes principais (CRUZ et al., 2011), agrupamentos hierárquicos clássicos (ODONG et al., 2011), entre outros. De modo geral, a maioria desses métodos utiliza as frequências alélicas para capturar níveis de diferenciação. Entretanto, isso pode ser o resultado de eventos de fluxo gênico recentes, de modo a não considerar os eventos migratórios ocorridos no passado (WILSON; RANNALA, 2003; ROSA, 2016), dificultando o entendimento das relações e do passado evolutivo das populações.

As estatísticas F de Wright ainda são um marco inicial de qualquer estudo genético populacional, especialmente em estudos visando à conservação. Entretanto, ao longo dos anos, as limitações têm se tornado evidentes (NEIGEL, 2002). Muitos desses métodos citados usam suposições restritivas, por exemplo, não permitindo considerar mutação. As

estatísticas F , assim como os métodos análogos ao F_{ST} , são frequentemente empregados em pares de populações, podendo levar a estimativas tendenciosas dos parâmetros (BEERLI, 2004; SLATKIN, 2005) e a reutilização de dados nesses métodos de pares é indesejável de um ponto de vista estatístico (BEERLI, 2009).

Em 1982, Sir John Frank Charles Kingman desenvolveu a teoria da coalescência (Kingman, 1982 a, b, c). Os trabalhos de Kingman passaram a ter enorme representatividade para os estudos que envolvem a genética de populações e evolução, abrindo portas para novos métodos em genética de populações, levando-se em consideração que as inferências baseadas na teoria da coalescência não eram praticáveis até meados dos anos 1990, devido a restrições computacionais (BEERLI, 2009).

A teoria da coalescência é uma abordagem retrospectiva que tem como objetivo modelar a genealogia dos alelos amostrados, tendo como base o modelo populacional proposto por Fisher (1930) e Wright (1931). A hipótese é de que os alelos amostrados no presente tenham surgido de um único alelo ancestral, dito alelo comum mais recente (NORDBORG, 2001).

Como exemplo, considere um determinado loco amostrado em uma única população, que segue o modelo de Fisher-Wright (FW; Fisher, 1930; Wright, 1931). A ancestralidade dessa amostra é investigada a partir do presente em direção ao passado, até o ponto em que os alelos do referido loco se coalescem em seu ancestral comum mais recente (ACMR). A ocorrência de um evento coalescente é a fusão de dois alelos do tempo (t) em um único alelo qualquer presente no tempo ($t-1$). Em visão prospectiva da história evolutiva das populações, os eventos coalescentes correspondem aos eventos de divergência (FU; LI, 1999).

Em uma amostra com tamanho n , têm-se $n-1$ eventos de coalescência caminhando-se em direção ao passado, de forma que o último evento coalescente corresponde ao ancestral comum mais recente da amostra (NORDBORG, 2001; WAKELEY, 2009). Ao considerar uma amostra de tamanho n , tem-se um total de $\frac{n(n-1)}{2}$ possíveis combinações para o primeiro evento de coalescência, sendo o raciocínio análogo aos demais eventos coalescentes (FU; LI, 1999).

A probabilidade de duas amostras quaisquer coalescerem em alguma geração no passado é de $\frac{1}{2N_e}$. Tem-se ainda que a probabilidade de duas amostras não coalescerem é de $1 - \frac{1}{2N_e}$, sendo N_e o tamanho efetivo populacional. Logo, a probabilidade de ocorrer

o primeiro evento coalescente é de $\frac{n(n-1)}{2} * \frac{1}{2N_e}$ considerando um total de n amostras. Isso se dá pelo fato de que em uma população com um tamanho efetivo populacional constante com $2N_e$ cópias de cada locus, existem $2N_e$ "pais em potencial" na geração anterior (ROSA, 2016). Nota-se ainda que, a princípio, o modelo de FW descreve um processo de amostragem binomial de alelos em cada geração.

Quando os eventos coalescentes são modelados de acordo com a distribuição binomial, o tempo de coalescência, que diz respeito aos intervalos entre os eventos coalescentes sucessivos, em cada geração anterior, é modelado de acordo com a distribuição geométrica, de forma que a probabilidade de não ocorrer coalescência nas gerações precedentes de $t-1$ é multiplicada pela probabilidade de coalescência na geração de interesse, sendo $P_c(t) = \left(1 - \frac{1}{2N_e}\right)^{t-1} * \frac{1}{2N_e}$ (HALLIBURTON et al., 2004).

Quando uma amostra de tamanho n é muito grande, a probabilidade de coalescer é muito pequena, e os eventos coalescentes são estatisticamente modelados seguindo a distribuição de Poisson, tendo a taxa de um evento de coalescência ($\lambda = 1$) para cada amostra em um intervalo de tempo (t_1, t_2, t_3 , até t_n). Isso ocorre devido a distribuição binomial se aproximar de uma distribuição de Poisson, quando a probabilidade de sucesso é pequena e n muito grande. Para esse caso, os tempos de coalescência são modelados usando a distribuição exponencial, onde o tempo esperado de coalescência é dado por $P_c(t) = \frac{1}{2N_e} e^{-\frac{t-1}{2N_e}}$ (HALLIBURTON et al., 2004).

As diferentes combinações possíveis para se obter eventos de coalescência determinam o número de configurações possíveis (também chamado de número de genealogias) envolvendo as amostras das populações atuais. Dessa maneira, o processo de inferência de certa genealogia, segundo a teoria da coalescência, baseia-se em diferentes modelos probabilísticos (ROSA, 2016).

O tempo necessário para ocorrer um evento coalescente é proporcional ao tamanho efetivo populacional e amostral, de modo em que os primeiros eventos coalescentes devem acontecer, em média, mais rapidamente no passado, enquanto que os coalescentes mais antigos irão demorar cada vez mais à medida em que o ACMR se aproxima (WAKELEY, 2009).

Kingman (1982 a, b, c) ampliou a observação de Wright levando dois cromossomos escolhidos aleatoriamente em uma população de tamanho N sobre $2N$ gerações até se encontrarem no seu ancestral comum mais recente. Suas descobertas

permitiram o uso de uma amostra aleatória de indivíduos para inferir parâmetros para toda a população, demonstrando, teoricamente, que é possível calcular a probabilidade de certa genealogia para qualquer número de indivíduos (NORDBORG, 2001).

Nos últimos anos, programas computacionais que possibilitam estimar parâmetros usando dados genéticos sob vários modelos de coalescência foram desenvolvidos (BEERLI; FELSENSTEIN, 1999, 2001; BAHLO; GRIFFITHS, 2000; WILSON; RANNALLA, 2003; De IORIO; GRIFFITHS, 2004; HEY; NIELSEN 2004; BEERLI, 2006; EWING; RODRIGO, 2006; KUHNER, 2006). A utilização dessas abordagens possibilita estimar parâmetros relacionados à genética de populações por meio de uma abordagem que considera o passado evolutivo e as relações genealógicas dos alelos.

3.5. Tamanho efetivo populacional

O tamanho efetivo populacional (N_e) é um dos vários “conceito-chave” em genética de populações, sendo inicialmente introduzido por Wright (1930) em seu trabalho “*Evolution in Mendelian Populations*”. Sua finalidade é fornecer uma maneira de calcular a taxa de mudança evolutiva causada pela amostragem aleatória de frequências alélicas em uma população finita (deriva genética).

O conceito de tamanho efetivo populacional refere-se ao número de indivíduos que se reproduzem e deixam descendentes, desta maneira, transmitindo os genes para a próxima geração. Esse conceito é fundamental para fins evolutivos, conservação genética e melhoramento de espécies (CABALLERO, 1994; CHARLESWORTH, 2009) pois é um dos principais fatores determinantes na perda de variabilidade em populações ameaçadas de extinção (SOLÉ-CAVA, 2001). Quanto menor o tamanho efetivo de uma população, maiores são os efeitos da deriva genética, como aumento da variância entre populações e diminuição da diversidade dentro de populações, colaborando para fixação ou perda de alelos (RIDLEY, 2006).

O tamanho efetivo conceitualmente equivale-se ao de uma população idealizada em relação ao modelo populacional de Fisher (1930) e Wright (1931), considerando que reduz a variabilidade genética na mesma taxa que a população do estudo (CABALLERO, 1994; CHARLESWORTH, 2009; SEBBENN; SEOANE, 2005). Dessa forma, o tamanho efetivo populacional teórico pode diferir do tamanho senso de uma população, sendo, na maioria das vezes, menor (WRIGHT, 1938).

Na literatura, são descritos diferentes tipos de estimadores para o tamanho efetivo, podendo eles serem baseados em dados fenotípicos ou em marcadores moleculares, sendo usado como medida de representatividade genética a partir determinada amostra (VENCOVSKY; CROSSA, 2003).

3.6. Fluxo gênico

O fluxo gênico é um termo genérico que diz respeito à migração de genes de uma população para outra, onde inclui-se todos os mecanismos possíveis que resultam na movimentação de alelos entre populações (NEIGEL, 1997). É juntamente com as medidas de diversidade genética e o grau de estruturação genética espacial das populações, um dos parâmetros genéticos populacionais mais importantes para fins de conservação genética e manejo, pois o fluxo pode restringir o processo evolutivo (SLATKIN, 1989; TAMBARUSSI et al., 2015).

Considerando o processo evolutivo como uma mudança nas frequências alélicas em uma população, ao longo do tempo (WILSON; BOSSERT, 1971), e que essas mudanças evolucionárias ocorrem por meio de quatro fenômenos diferentes, frequentemente chamados de “forças evolutivas” ou “mecanismos evolutivos”, os quais são a mutação, a seleção, a deriva e o fluxo gênico, onde esses mecanismos, agindo separadamente ou em conjunto, afetam a mudança de frequências alélicas em uma população de geração em geração.

Fatores como tamanho populacional pequeno e a presença de endogamia reduzem a variabilidade genética dentro de populações e acentuam os efeitos da deriva genética (TAMBARUSSI et al., 2015), contribuindo no sentido de aumentar a diferenciação, divergência e heterogeneidade genética entre as populações. O fluxo gênico age justamente no sentido de se contrapor aos efeitos da deriva genética, permitindo o aumento da variabilidade existente nas populações, espalhando novos genes e combinações de genes em populações (SLATKIN, 1989), e, ao mesmo tempo, a redução da variabilidade entre populações devido ao número de indivíduos que migram, por geração, de uma população para outra (WRIGHT, 1951). No entanto, caso o fluxo gênico não ocorra, é possível que as populações acumulem diferenças genéticas em uma escala temporal, podendo se tornarem espécies distintas (FUTUYAMA, 1992; RIDLEY, 2006; HARTL, 2010).

Há descrito, genericamente, quatro modelos corriqueiramente utilizados para explicar como o fluxo gênico atua nas metapopulações (FUTUYAMA, 1992). Os modelos são:

- I. Modelo de n-ilhas: a migração ocorre ao acaso dentre um grupo de pequenas subpopulações;
- II. Modelo continente-ilha: existe, efetivamente, um movimento unidirecional de uma subpopulação grande para uma menor, quase isolada;
- III. Modelo de “alpondras” (*stepping-stone*): cada subpopulação recebe migrantes somente das subpopulações vizinhas;
- IV. Modelo de isolamento por distância: o fluxo gênico ocorre preferencialmente entre indivíduos mais próximos espacialmente, em subpopulações de distribuição contínua.

3.6.1. Estimação do fluxo gênico

O processo de estimação das taxas de fluxo gênico é uma atividade bastante complexa, pois estas taxas não dependem apenas do processo de dispersão, mas, também, do sucesso na reprodução do indivíduo migrante e do tamanho da população que recebe o migrante (ENNOS, 1994), além de fatores como sistema reprodutivo da espécie e as relações ecológicas entre as plantas e seus polinizadores e dispersores (DICK et al., 2003; ELLSTRAND, 2014; RIDLEY, 2006).

O estudo do fluxo gênico pode ser realizado com base em duas estratégias: (i) métodos diretos, e (ii) métodos indiretos (SLATKIN, 1985; 1987; BEERLI, 1998). Os métodos diretos utilizam observações diretas ou estimativas de distância de dispersão e de sucesso reprodutivo, as quais são baseadas em observações do movimento dos vetores de pólen e sementes (LATTA et al., 1998) e na análise da paternidade (JONES; ARDREN, 2003). Já os métodos indiretos utilizam dados genéticos de amostras de indivíduos, oriundos de várias subpopulações, para realizar inferência das taxas de migração (BEERLI, 1998).

Os métodos diretos podem ajudar a determinar o padrão de migração dos indivíduos durante o estudo e podem fornecer informações sobre um histórico muito recente. Pressupõe-se que os poucos indivíduos rastreados com tais métodos sejam escolhidos aleatoriamente e que seus movimentos não sofram viés, e os dados coletados

para este caso podem fornecer informações interessantes sobre o padrão de migração de uma população específica (BEERLI, 1998).

Entretanto, os métodos diretos possuem várias limitações, sendo pouco viável em populações arbóreas naturais, isso porque taxas de migração pequenas podem ser indetectáveis e a precisão e a confiabilidade das estimativas dos parâmetros são baixas quando o estudo é baseado em poucos indivíduos (BEERLI, 1998). Se o estudo for de curta duração e não for repetido, não é possível saber se o padrão de migração observado foi acidental ou geral, tendo em vista que a observação da dispersão está necessariamente limitada no tempo e no espaço. Dessa forma, as estimativas indicam o fluxo gênico de um tempo particular, não sendo capaz de captar os eventos ocorridos no passado histórico das populações (BEERLI, 1998).

Os métodos indiretos, utilizando dados genéticos de amostras de indivíduos, são muito utilizados para populações vegetais, fazendo uso de modelos probabilísticos. Os dados genéticos podem vir de diferentes tipos de marcadores moleculares, sendo esses dados a base para inferência nos padrões de fluxo gênico (GRIFFITHS et al., 2004).

Como vantagem do método indireto, não há necessidade de rastrear os indivíduos ao longo do tempo. Para este método, as estimativas são médias ao longo do tempo evolutivo e revelam padrões de migração gerais (não individuais) (BEERLI, 1998). Esse método pode ser aplicado a qualquer organismo; contudo, apresenta como desvantagem a necessidade de assumir que os marcadores são seletivamente neutros, de modo que a similaridade genética entre diferentes subpopulações seja resultado da migração e não de pressões de seleção (BEERLI, 1998). Ainda, os marcadores precisam mostrar elevado polimorfismo para que seja possível avaliar as diferenças entre subpopulações. Um marcador com uma taxa de mutação muito lenta não revelará eventos de migração recentes, porém, pode ter algumas informações sobre eventos de migração no passado, quando comparado com populações geograficamente mais distantes (BEERLI, 1998).

Várias abordagens usando dados de marcadores genéticos para a inferência de taxas de migração são descritas, citando-se os estimadores baseados em frequências alélicas e estatística F de Wright, análogos como descrito por Michalakis e Excoffier, (1996); estimadores de máxima verossimilhança, baseados em frequências alélicas (RANNALA; HARTIGAN, 1996; TUFTO et al. 1996); estimadores baseados nas genealogias dos indivíduos amostrados (KINGMAN, 1982), com taxas de migração estimadas usando procedimentos de Bahlo e Griffiths (1998); Beerli e Felsenstein, (1998); Wakeley, (1998).

Alguns estimadores são misturas dos grupos citados (BEERLI, 1998), como no caso de Slatkin e Maddison (1989), que desenvolveram um método que utiliza resultados da teoria da coalescência e logo em seguida apresenta uma tabela de interpolação produzida pela simulação da coalescência com migração, na qual o número mínimo de eventos de migração encontrados na melhor genealogia está relacionado a uma taxa de migração $4N_e m$.

A maioria desses estimadores foram desenvolvidos partindo de pressupostos simplificados, assumindo que a migração e as taxas de mutação são constantes ao longo do tempo (BEERLI, 1998). Além disso, quase todos os métodos usam uma população de tamanho constante. Os métodos que permitem supor que as subpopulações não são constantes estão pouco desenvolvidos e são escassos, citando-se Bahlo e Griffiths (1998). Entretanto, a disponibilidade de métodos para estimar parâmetros populacionais sob condições não constantes aumentará o desenvolvimento de novas ferramentas estatísticas e computacionais, e a implementação de um novo método geralmente são lentas. Além disso, métodos mais complexos normalmente possuem maior quantidade de parâmetros, que demandarão mais dados para serem estimados, tornando muitas vezes impraticável o processo de estimação (BEERLI, 1998).

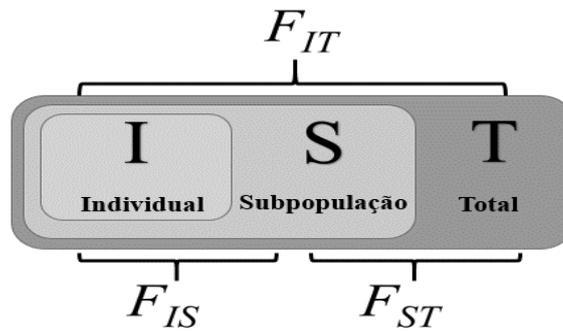
Escolher uma abordagem para a estimação de fluxo gênico não é uma tarefa simples, pois, cada um desses métodos tem seu próprio conjunto de suposições, as quais podem ser incompatíveis com o estudo. Além disso, os resultados obtidos com dados reais fornecem uma orientação pouco confiável para a escolha de um método, porque normalmente não sabemos os valores reais dos parâmetros subjacentes (BEERLI, 2009).

Grande parte dos modelos de estimativa de fluxo gênico baseia-se em populações de Fisher-Wright (FISHER, 1930; WRIGHT, 1931). O modelo populacional de Fisher-Wright com migração é relativamente simples e de fácil entendimento, tendo propriedades que facilitam as manipulações matemáticas (BEERLI, 2009). O modelo consiste em uma subpopulação, com um número constante de indivíduos, onde em cada geração os indivíduos produzem um grande número de gametas os quais podem permanecer em uma subpopulação ou migrarem para outra subpopulação. Dessa forma, os novos indivíduos são formados pela “escolha” aleatória de dois gametas em uma subpopulação, onde os novos indivíduos substituem seus genitores (HARTL, 2010).

Wright (1951) descreveu uma estrutura que utiliza as estatísticas F , para uma população subdividida, propondo três coeficientes: F_{IS} , F_{IT} e F_{ST} . Usando o modelo de alelos infinitos, F pode ser entendido como uma probabilidade de que dois alelos

escolhidos aleatoriamente sejam idênticos por descendência, e os F 's são proporções onde as letras subscritas são os níveis hierárquicos, como no caso o ' I ' é indivíduo, ' S ' é subpopulação e ' T ' é o total de subpopulações.

Figura 2: Esquema referente a estatísticas F e as relações entre os níveis hierárquicos populacionais.



Fonte: Adaptado de acordo com Hartl e Clark (2010)

Beerli (1998), utilizando conceitos de Michalakis e Excoffier (1996), explica que o F_{ST} pode ser descrito de forma genérica como sendo:

$$F_{ST} = \frac{g(\sigma_w) - g(\sigma_b)}{g(\sigma_b)} \quad (3.1)$$

Onde, σ_w é a variação dentro da população, σ_b é a variação entre populações e $g(x)$ são funções com correção para os diferentes estimadores análogos ao F_{ST} com variância dentro de uma população (σ_w) ou entre populações (σ_b). Dessa maneira, o F_{ST} pode ser substituído por diferentes estimadores específicos, citando-se alguns como o θ (WEIR, 1996), N_{ST} (LYNCH; CREASE, 1990), $\langle F_{ST} \rangle$ (HUDSON et al., 1992), Φ_{ST} (EXCOFFIER; SMOUSE 1994), ρ_{ST} (ROUSSET, 1996), G_{ST} (NEI, 1973) e R_{ST} (SLATKIN, 1993).

Essas variâncias são proporcionais aos tempos médios de coalescência em uma subpopulação em relação a toda a população. Sendo o F_{ST} interpretado como uma medida da diferenciação entre subpopulações, onde valores próximos de zero indicam que a população não está estruturada (FRANKHAM et al., 2002).

O F_{ST} é de forma corriqueira transformado em uma medida mais direta para migração. Wright (1951) mostrou que, para um modelo de população em ilhas, com um número infinito de subpopulações e nenhuma mutação, o F_{ST} pode ser escrito como sendo:

$$F_{ST} = \frac{1}{1 + 4N_e m} = N_e m = \frac{1}{4} \left(\frac{1}{F_{ST}} - 1 \right) \quad (3.2)$$

em que, N_e é o tamanho efetivo de uma população Fisher-Wright e m é a taxa de migrantes por geração. Beerli (1998) ressalta que, mesmo a população não se comportando exatamente como uma população de Fisher-Wright, ainda é possível usar N_e para fazer comparações com outras populações. Relaxar a suposição de ter um número infinito de subpopulações é relativamente simples e foi descrito anteriormente, citando-se Li (1976), onde demonstra:

$$F_{ST} = \frac{1}{1 + 4N_e m \frac{d^2}{(d-1)^2}} \quad (3.3)$$

em que d é o número de subpopulações, d é certamente diferente do número de populações amostradas. Ressalta-se que os resultados para $N_e m$ podem ser muito diferentes dependendo de qual versão do análogo ao F_{ST} que foi utilizada (CHARLESWORTH, 1998).

O modelo de população de n-ilhas faz uso de apenas dois parâmetros, sendo o tamanho efetivo da população (N_e) e a taxa de imigração por geração m . Assume-se que os tamanhos das subpopulações são os mesmos e que a taxa de migração é a mesma entre todas as subpopulações. Entretanto, essas suposições são frequentemente violadas em estudos de populações naturais, para as quais não conhecemos nem os padrões reais de migração nem os tamanhos populacionais (BEERLI, 2009).

Beerli (1998), fazendo uso dos conceitos propostos na teoria da coalescência, expõe que é possível construir a genealogia dos indivíduos amostrados e que, em uma única população, este processo depende apenas do tamanho efetivo populacional.

Kingman (1982) mostrou que a probabilidade de ocorrer um evento de coalescência no passado, entre duas cópias genéticas escolhidas aleatoriamente a partir de uma amostra de tamanho n , no intervalo de tempo (t), que é medido em gerações, escalado pela taxa de mutação (μ) é:

$$\text{Prob}(t | N_e, \mu) = \exp\left(-t \frac{n(n-1)}{4N_e}\right) \frac{2}{4N_e} \quad (3.4)$$

Pode-se ainda estimar $\text{Prob}(g | N_e, \mu)$, sendo a probabilidade de toda uma genealogia (g) iniciando com n amostras de alelos ou sequências e, voltando no tempo, multiplicando as probabilidades para cada intervalo de tempo entre os “nós” nesta genealogia. Dessa forma, seria possível examinar todas as possíveis genealogias e encontrar a genealogia ou grupo de genealogias para as quais a probabilidade dada aos parâmetros da população é a maior.

Esta estrutura proposta por Kingman (1982) pode ser expandida de modo a incluir todos os possíveis parâmetros da população por meio dessa abordagem, de modo que a estrutura original da coalescência pode ser expandida incorporando-se outros parâmetros populacionais (HUDSON, 1990), tais como crescimento populacional (GRIFFITHS; TAVARÉ 1994; KUHNER et al. 1998), taxas de migração (NATH; GRIFFITHS 1996; BAHLO; GRIFFITHS 1998; BEERLI; FELSENSTEIN 1998), taxas de recombinação (GRIFFITHS; MARJORAM 1996), e seleção (KRONE; NEUHAUSER 1997; NEUHAUSER; KRONE 1997). Inclui-se ainda modelos com migração e fluxo gênico assimétrico, podendo ser descrito na forma um sistema de duas populações (BEERLI, 1998) da seguinte maneira:

$$\text{Prob}(t | N_e^{(1)}, N_e^{(2)}, m_{21}, m_{12}, \mu) = \exp\left(-t \left[\frac{n_1(n_1-1)}{4N_e^{(1)}} + t \frac{n_2(n_2-1)}{4N_e^{(2)}} + n_1m_1 + n_2m_2 \right] \right) \beta \quad (3.5)$$

em que t é o intervalo de tempo; n é o tamanho da amostra; μ é a taxa de mutação; N_e é o tamanho efetivo populacional; β é a probabilidade de um coalescente na população i , com probabilidade $\frac{2}{4N_e^{(i)}}$, ou um evento de migração da população j para i , com probabilidade m_{ji} ;

As metodologias mais recentes de fluxo gênico via coalescência, a partir de marcadores microssatélites, baseiam-se, principalmente, nos métodos de máxima verossimilhança e inferência bayesiana (KUHNER; SMITH, 2007).

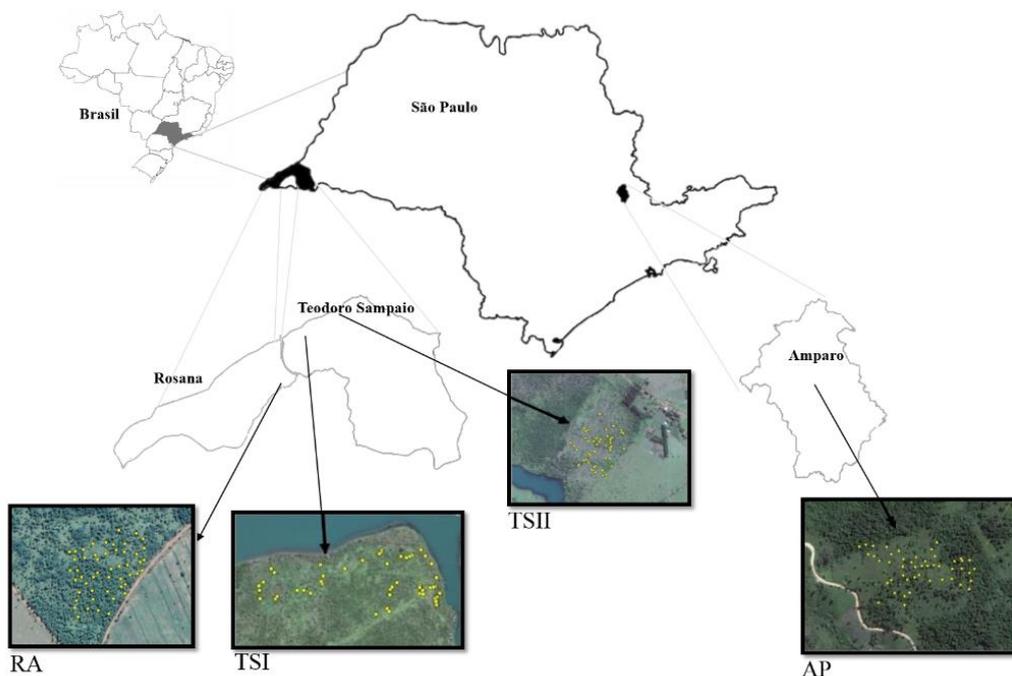
4. MATERIAL E MÉTODOS

4.1. Local de Estudo e amostragem

Foram utilizadas quatro subpopulações de *A. aculeata*, localizadas nos municípios de Teodoro Sampaio (duas populações, TSI e TSII), Rosana (RA) e Amparo (AP), municípios do Estado de São Paulo, Brasil (Figura 2). As três subpopulações na região do Pontal do Paranapanema (TSI, TSII e RA) fazem parte das reservas legais de assentamentos rurais, e a subpopulação de AP corresponde a uma propriedade particular.

Essa amostragem e genotipagem é a mesma utilizada por Coelho (2017), cujo dados são oriundos do projeto FAPESP 2013/23168-4, gentilmente cedidos pelos proponentes para a realização desta pesquisa.

Figura 3: Representação gráfica da distribuição espacial dos genótipos coletados nas localidades de Teodoro Sampaio I (TSI), Teodoro Sampaio II (TSII), Amparo (AP) e Rosana (RA) em que *A. aculeata* foram amostradas no Estado de São Paulo.



Fonte: Autor

Foram amostrados 50 indivíduos de *A. aculeata* em cada população, a partir da coleta de amostras das folhas para extração de DNA e coordenadas geográficas de cada indivíduo usando GPS (Garmin: GPSMAP 76CXS), totalizando 200 amostras (Metodologia realizada por Coelho, 2017).

Tabela 1: Matriz com as distâncias geográficas (Km) entre as localidades amostradas para espécie *A. aculeata* nas populações de Teodoro Sampaio I (TSI), Teodoro Sampaio II (TSII), Amparo (AP) e Rosana (RA).

Localidades	TSI	TSII	AP	RA
TSI	-			
TSII	12	-		
AP	586	596	-	
RA	29	19	602	-

4.2. Extração de DNA, reações de PCR e checagem de dados

A extração de DNA foi realizada no Laboratório de Reprodução e Genética de Espécies Arbóreas- LARGEA (Esalq-USP), as reações de PCR e genotipagem foram feitas pela empresa Hereditas Tecnologia em Análise de DNA[®], seguindo método CTAB (DOYLE; DOYLE, 1990) com PVP 40 a 2% e mercapto-ethanol a 3%. Após a extração, eletroforese, reações de amplificação e análise de fragmentos foram realizadas conforme descrito em Silva et al. (2016), com os nove locos microssatélites, sendo quatro desenvolvidos e descritos por Nucci (2007), Aacu30, Aacu26, Aacu07, Aacu12 e cinco não publicados, MAC245, MAC115, MAC201, MAC240, MAC23 (GRATTAPAGLIA, não publicado, 2016; COELHO et al., 2017).

As frequências de alelos nulos foram estimadas utilizando o programa Micro-Checker 2.2.3 (VAN OOSTERHOUT et al., 2006), baseado em simulações pelo método de Monte Carlo via procedimento *bootstrap*, para gerar frequências esperadas dos alelos em homozigose e heterozigose, utilizando um intervalo de confiança de 95% para cada loco individualmente. Os genótipos foram ajustados baseando-se no estimador de Oosterhout et al. (2004).

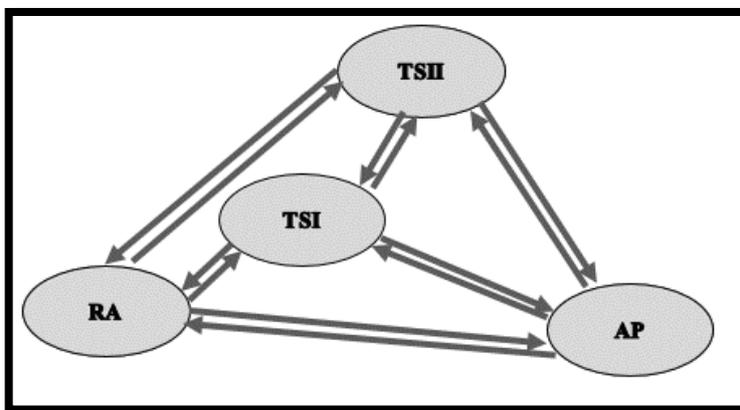
4.3. Fluxo gênico e tamanho efetivo via coalescência

Diferentes modelos demográficos e de conectividade entre as populações foram utilizados (Figura 3). Diferentes cenários de fluxo gênico foram testados, visando elucidar se populações estão em panmixia ou se estas se mostram estruturadas e se há indícios de

isolamento por distância. A comparação dos modelos foi realizada estimando a probabilidade do modelo (equação 4.1), onde os modelos responsáveis pela maior proporção de verossimilhança são os mais prováveis, estimada a partir probabilidades marginais obtidas pela integração termodinâmica no *software* Migrate-N 3.6.11 (BEERLI; PALCZEWSKI, 2010).

Figura 4: Esquema dos modelos demográficos e de fluxo gênico que foram utilizados entre as quatro subpopulações de *A. aculeata* utilizadas nesse estudo, ambas no Estado de São Paulo, Brasil.

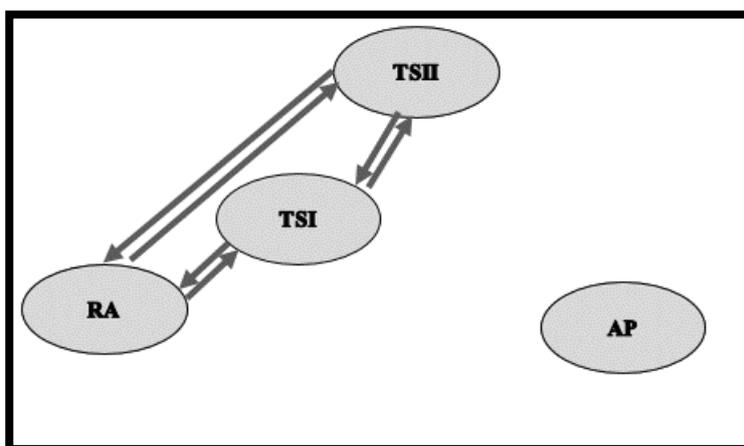
Modelo 1:



Pop	TSI	TSII	AP	RA
TSI	*	*	*	*
TSII	*	*	*	*
AP	*	*	*	*
RA	*	*	*	*

em que TSI= População I de Teodoro Sampaio; TSII= População II de Teodoro Sampaio; RA= População de Rosana; AP= População de Amparo; “*” significa que este parâmetro é estimado; “0” significa que este parâmetro não é estimado; “s” significa que o parâmetro de migração é simétrico ($M_{ij} = M_{ji}$).

Modelo 2:



Pop	TSI	TSII	AP	RA
TSI	*	*	0	*
TSII	*	*	0	*
AP	0	0	*	0
RA	*	*	0	*

em que TSI= População I de Teodoro Sampaio; TSII= População II de Teodoro Sampaio; RA= População de Rosana; AP= População de Amparo; “*” significa que este parâmetro

é estimado; “0” significa que este parâmetro não é estimado; “s” significa que o parâmetro de migração é simétrico ($M_{ij} = M_{ji}$).

Modelo 3:

A probabilidade do modelo é obtida dividindo cada probabilidade marginal pela soma das probabilidades marginais de todos os modelos utilizados.

$$Prob(\text{modelo}_i) = \frac{mL\text{modelo}(i)}{\sum mL\text{modelos}(todos)} \quad (4.1)$$

Em que, $Prob(\text{modelo}_i)$ é a probabilidade do modelo relação aos demais; $mL\text{modelo}(i)$ é a probabilidade marginal do modelo (i); $\sum mL\text{modelos}(todos)$ é a somatória das probabilidades de todos os modelos utilizados.

Para avaliar o tamanho efetivo e fluxo gênico assimétrico entre as populações foi utilizado o método de inferência bayesiana, baseado na teoria da coalescência implementado no *software* MIGRATE-N (BEERLI; PALCZEWSKI, 2010), onde dois parâmetros, o (Θ_i), e a taxa de migração escalonada pela taxa de mutação (M_{ij}) são estimados simultaneamente:

$$\Theta_i = xN_e\mu \quad (4.2)$$

Em que x é um multiplicador que depende da ploidia e herança dos dados ($x=4$ para diploides nucleares); N_e é o tamanho efetivo da população; e μ é a taxa de mutação.

$$M_{ij} = \frac{m_{ij}}{\mu} \quad (4.3)$$

Em que m_{ij} = taxa de imigração entre grupo i para grupo j ; μ = taxa de mutação.

A partir desses valores, foi obtido o número efetivo de migrantes por geração ($xN_i m_{ji}$), o qual pode ser estimado por:

$$xN_i m_{ji} = \Theta_i * M_{ij} \quad (4.4)$$

A distribuição a *priori* dos parâmetros escolhida foi a uniforme, com os valores de mínimo, máximo e delta respectivamente de 0, 100 e 10. O processo de estimação teve 5.000.000 iterações, com um período de aquecimento (*burnin*) de 10.000. Empregou-se o esquema de aquecimento estático com quatro cadeias de diferentes temperaturas (1,0;

1,5; 3,0 e 100.000,0) (BEERLI; PALCZEWSKI, 2010). Após o término de cada análise, a área abaixo da curva de verossimilhança, realizada por meio de integração termodinâmica, foi calculada utilizando a aproximação pela curva de Bezier, a qual possibilita uma melhor aproximação com um pequeno número de cadeias aquecidas (BEERLI; PALCZEWSKI, 2010). A taxa de mutação variou apenas entre locos, e assumiu-se uma taxa de mutação de 5×10^{-5} , próxima as taxas descritas no trabalho de Marriage et al. (2009).

5. RESULTADOS

5.1. Ajuste dos dados para frequência de alelos nulos

Desvios estatisticamente significativos do equilíbrio de Hardy-Weinberg, devido à deficiência de heterozigotos, foram observados para os locos MAC115 e MAC201 nas populações TSI e RA, e para os locos Aacu12 e MAC115 na população de AP. Não houve evidências de alelos nulos observadas na população de TSII (Tabela 2). Os genótipos foram ajustados baseando-se no estimador de Oosterhout et al. (2004), para as frequências $>0,05$).

Tabela 2: Estimativa das frequências de alelos nulos utilizando o estimador de van Oosterhout, para cada loco individualmente, em quatro subpopulações amostradas no Estado de São Paulo, Brasil.

Locos	TSI	TSII	AP	RA
	<i>P</i> nulo	<i>P</i> nulo	<i>P</i> nulo	<i>P</i> nulo
<i>Aacu30</i>	0,039	-0,133	0,095	0,086
<i>Aacu26</i>	-0,020	0,009	0,099	0,069
<i>Aacu07</i>	-0,062	-0,062	-0,010	-0,010
<i>Aacu12</i>	-0,025	-0,028	0,225*	0,056
<i>MAC245</i>	0,021	-0,001	-0,049	0,041
<i>MAC115</i>	0.132*	0,098	0,433*	0,208*
<i>MAC201</i>	0.225*	0,035	0,040	0,233*
<i>MAC240</i>	0,011	-0,014	0,064	0,000
<i>MAC23</i>	0,075	0,060	-0,044	-0,008

P nulo = Estimador de van Oosterhout para a frequência de alelos nulos em cada loco.

5.2. Fluxo gênico e tamanho efetivo via coalescência

O modelo de fluxo gênico adotado para as inferências foi o “modelo 1”, baseando-se como critério de escolha na probabilidade do modelo (“modelo 1” com probabilidade $>0,99$; o outro modelo apresentou valores $< 0,01$ de probabilidade). Foram obtidas as estimativas do parâmetro M_{ij} , bem como o intervalo de confiança de 95% (Tabela 3) para tal parâmetro, sendo M_{ij} a taxa de migração dimensionada pela taxa de mutação $M_{ij} = m_{ij}/\mu$.

Tabela 3: Estimativas das taxas de migração, escalonadas pela taxa de mutação (M_{ij}), obtidas para a espécie *A. aculeata* envolvendo as localidades TSI, TSII, RA e AP (Estado de São Paulo), considerando a média de todos os locos microsatélites utilizados.

Valores de M_{ij}		Recebem			
		TSI	TSII	AP	RA
Doam	TSI	*	1.97 (1.20-2.73)	1.23 (0.53-1.93)	1.7 (0.93-2.47)
	TSII	1.63 (0.87-2.33)	*	0.9 (0.27-1.53)	1.63 (0.8-2.33)
	AP	0.43 (0.00-0.93)	0.83 (0.20-1.40)	*	1.3 (0.53-1.93)
	RA	1.63 (0.80-2.33)	2.77 (1.93-3.60)	1.23 (0.47-1.87)	*

Onde $M_{ij} = (m_{ij}/\mu)$ sendo m_{ij} a taxa de migração entre localidade i para j , e μ a taxa de mutação; TSI, TSII, RA e AP são respectivamente Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana.

As estimativas médias para as taxas de imigração (M_{ij}) foram de 1,63; 0,43 e 1,63 recebidos pela localidade de TSI, oriundos de TSII, AP e RA, respectivamente. A localidade de TSII recebeu imigrantes com taxas de 1,97; 0,83 e 2,77 das localidades TSI, AP e RA, respectivamente. A localidade de AP recebeu taxas de 1,23; 0,9 e 1,23, provenientes das localidades de TSI, TSII e RA. Na localidade de RA, as taxas de imigração recebidas foram de 1,7; 1,63 e 1,3 oriundas de TSI, TSII e AP (Tabela 3).

Estimou-se o número efetivo de migrantes por geração ($N_i m_{ji}$), obtida pela expressão $xN_i m_{ji} = \Theta_i * M_{ij}$, e o intervalo de confiança (95%) para a estimativa (Tabela 4).

Tabela 4: Estimativas do número efetivo de migrantes por geração ($N_i m_{ji}$), para espécie *A. aculeata* nas localidades Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana, no Estado de São Paulo, Brasil, e considerando o intervalo de confiança de 95%.

Valores de $N_i m_{ij}$		Recebem			
		TSI	TSII	AP	RA
Doam	TSI	*	0.83 0.50-1.59	0.28 0.12-0.67	0.72 0.39-1.45
	TSII	0.72 0.38-1.43	*	0.20 0.06-0.59	0.68 0.34-1.42
	AP	0.19 0.00-0.91	0.35 0.08-1.09	*	0.54 0.22-1.23
	RA	0.71 0.35-1.46	1.17 0.81-2.03	0.27 0.10-0.67	*

Onde os valores são número efetivo de migrantes por geração ($N_i m_{ji}$), obtida por $xN_i m_{ji} = \Theta_i * M_{ij}$ e abaixo os intervalos de confiança de 95%, as siglas TSI, TSII, AP e RA correspondem, respectivamente, às localidades de Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana.

As estimativas do número efetivo de migrantes (Nm) em cada geração, recebidos pela localidade de Teodoro Sampaio I, foram de 0,72; 0,19 e 0,71, oriundos respectivamente de Teodoro Sampaio II, Amparo e Rosana. Para a localidade de Teodoro Sampaio II, os migrantes recebidos foram de 0,83; 0,35 e 1,17, respectivamente de Teodoro Sampaio I, Amparo e Rosana. A localidade de Amparo recebeu 0,28; 0,2 e 0,27, vindo das localidades de Teodoro Sampaio I, Teodoro Sampaio II e Rosana. Na localidade de Rosana, os migrantes recebidos foram de 0,72; 0,68 e 0,54, oriundas de Teodoro Sampaio I, Teodoro Sampaio II e Amparo (Tabela 4).

Foi estimado o número de migrantes por geração para cada loco (Figura 5) e o tamanho efetivo populacional para as quatro localidades (Tabela 5).

Figura 5: Estimativas do número efetivo de migrantes por geração ($N_i m_{ji}$), obtidas para espécie *A. aculeata* nas localidades Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana, no Estado de São Paulo, para cada loco.

		<i>Locus Aacu30</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	37.96	2.09	16.86
	TSII	10.55	*	1.62	12.01
	AP	1.44	10.26	*	7.28
	RA	11.2	16.78	2.79	*
		Recebe			

		<i>Locus Aacu26</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	3.39	0.18	2.44
	TSII	4.71	*	0.12	2.54
	AP	0.66	0.39	*	0.71
	RA	3.48	1.87	0.14	*
		Recebe			

		<i>Locus Aacu07</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	2.46	0.15	0.96
	TSII	1.42	*	0.15	1.54
	AP	0.38	0.43	*	0.47
	RA	1.48	1.74	0.3	*
		Recebe			

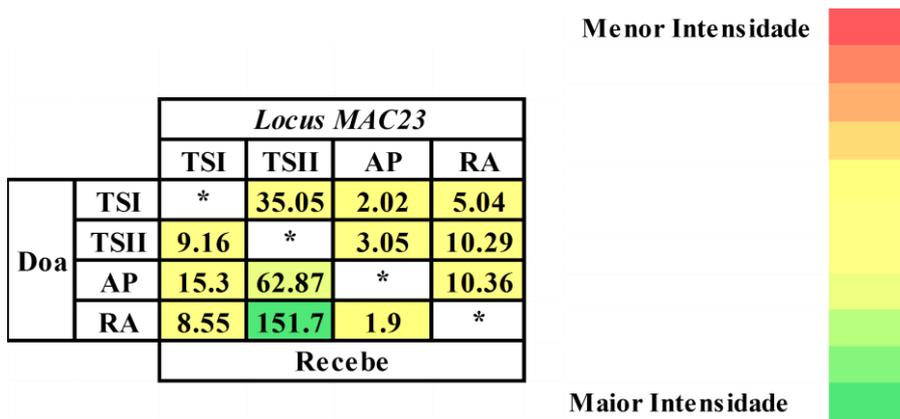
		<i>Locus Aacu12</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	40.1	1.65	35.06
	TSII	43.05	*	2.61	35.2
	AP	40.18	57.52	*	5.07
	RA	19.48	44.94	0.33	*
		Recebe			

		<i>Locus MAC245</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	2.12	1.57	2.6
	TSII	1.52	*	1.79	3.83
	AP	0.72	0.46	*	4.64
	RA	1.19	2.32	3.26	*
		Recebe			

		<i>Locus MAC115</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	4.85	0.83	23.92
	TSII	1.3	*	1.82	15.42
	AP	0.36	0.75	*	21.92
	RA	2	3.57	1.16	*
		Recebe			

		<i>Locus MAC201</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	2.83	0.84	3.58
	TSII	5.77	*	0.56	3.41
	AP	4.27	2.06	*	3.37
	RA	2.58	3.13	0.49	*
		Recebe			

		<i>Locus MAC240</i>			
		TSI	TSII	AP	RA
Doa	TSI	*	2.92	1.82	8.59
	TSII	8.18	*	0.5	9.32
	AP	1.85	0.48	*	10.73
	RA	2.84	2.49	2.3	*
		Recebe			



Onde os valores são número efetivo de migrantes por geração; TSI, TSII, AP e RA são respectivamente as localidades de Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana.

Considerando cada loco, houve uma grande amplitude nos valores estimados do número efetivo de migrantes (N_m), que variaram de 0,12 para o loco *Aacu26*, no sentido de Teodoro Sampaio II para Amparo, até 151,7 para o loco *MAC23*, indo no sentido de Rosana para Teodoro Sampaio II (Figura 5).

As estimativas do tamanho efetivo populacional (N_e) foram de 874, 835, 444 e 838 para TSI, TSII AP e RA respectivamente (Tabela 5).

Tabela 5: Estimativas do tamanho efetivo populacional, do intervalo de confiança (95%) e ($\theta_i = 4N_e\mu$) para espécie *A. aculeata* nas quatro localidades estudadas no Estado de São Paulo, considerando todos os locos utilizados.

Localidade	Tamanho efetivo populacional	$\theta_i = 4N_e\mu$
TSI	874 500-1700	1,749
TSII	835 467-1667	1,669
AP	444 5-1267	0,887
RA	838 433-1667	1,676

Onde N_e é o tamanho efetivo populacional; μ é a taxa de mutação por geração; TSI, TSII, AP e RA são respectivamente as localidades de Teodoro Sampaio I, Teodoro Sampaio II, Amparo e Rosana.

6. DISCUSSÃO

6.1. Ajuste dos dados para frequência de alelos nulos

Os alelos nulos são alelos que não amplificam e quando estão em heterozigose, resultam em um falso e aparente homozigoto. Dessa forma, ressalta-se importante a verificação de alelos nulos e respectivos ajustes para as análises estatísticas. A presença dos mesmos em marcadores microssatélites é frequente (BRUFORD; WAYNE, 1993; LI et al., 2003; KARLSSON; MORRIS, 2005). A presença de alelos nulos pode produzir um erro importante nas proporções genóticas, havendo um claro e explícito desvio do equilíbrio de Hardy-Weinberg devido a um excesso de homozigotos.

Para os locos *MAC115* e *MAC201* (populações TSI e RA), e os locos *Aacu12* e *MAC115*, (população de AP), o *software* gerou genótipos ajustados, baseado no estimador de Oosterhout et al. (2004), em que o número de genótipos homozigotos para cada classe de comprimento é ajustado de modo a refletir o número “real” estimado de homozigotos. Esse método assume que as amostras não amplificadas são de causas como DNA degradado, entre outros fatores. O ajuste das frequências dos alelos nulos é estimado a partir do cálculo prévio das frequências alélicas dos locos visíveis.

6.2. Fluxo gênico e tamanho efetivo populacional pela coalescência

As estimativas do tamanho efetivo e do fluxo gênico, obtidas pelo método proposto por Beerli e Palczewski (2010), mostraram evidências importantes em relação às populações. Dentre os diferentes cenários de fluxo gênico que foram propostos, o escolhido como mais provável foi o do modelo 1, baseando-se como critério de escolha na probabilidade do modelo, onde a comparação das verossimilhanças para decisão foi feita de acordo com o sugerido por Beerli (2012). Assim, os diferentes cenários dos modelos são somados e depois os modelos são proporcionalizados a essa soma para comparação, onde os modelos responsáveis pela maior proporção de verossimilhança são os mais prováveis.

Ressalta-se que ainda não se encontram descritos na literatura trabalhos envolvendo palmeiras, ou espécies filogeneticamente e ecologicamente próximas usando marcadores *SSR* e que abordem a metodologia semelhante para estimativa de fluxo gênico ou tamanho efetivo baseado na teoria da coalescência.

As maiores taxas de migração escalonada pela taxa de mutação (M_{ij}) foram entre as três localidades geograficamente próximas (TSI, TSII e RA) enquanto as menores foram relacionadas a população de Amparo, a qual encontra-se, respectivamente a cerca de 586, 596 e 602 quilômetros de distância das demais populações, resultado que já era esperado devido grande distância geográfica.

Entretanto, os valores de M_{ij} devem ser interpretados com cautela, pois são oriundos da relação (m_{ij}/μ), onde m_{ij} é a taxa de imigração da população a qual é dependente da taxa de mutação, de modo que representam a razão do quão importante é o processo de migração, em relação a taxa de mutação, no sentido de introduzir a variabilidade nas populações estudadas.

O parâmetro estimado mais usual para migração é o número efetivo de migrantes por geração ($N_i m_{ji}$), o qual pode ser facilmente obtido por meio da expressão $x N_i m_{ji} = \Theta_i * M_{ij}$. Esse parâmetro é preferido, pois quando se obtêm valores elevados para M_{ij} , isso pode indicar, de modo ambíguo, um fluxo gênico elevado, ou baixa taxa mutação, dificultando uma real interpretação sem conhecimento prévio das taxas de mutação (Beerli, 1998). Resumidamente, as taxas de migração diferentes podem ter o mesmo impacto na estrutura populacional, e as taxas de migração semelhantes podem ter impacto diferente na estrutura populacional.

Os menores valores referentes ao número efetivo de migrantes por geração (Nm ou $N_i m_{ji}$) ocorreram na localidade de Amparo (Tabela 4) que está geograficamente a mais de 580 km da subpopulação mais próxima utilizada no estudo, de maneira que a diferença entre os valores obtidos para as regiões geograficamente próximas foram mais altos, e menos contrastantes entre-si, em relação a Amparo. O resultado obtido era esperado, visto a grande distância de Amparo para as demais localidades.

Além disso, considerando os resultados de Nm para locos específicos nas subpopulações de *A. aculeata* pode haver indícios de uma história evolutiva distinta, visto grande variação entre o número de migrantes por geração.

Entretanto, deve-se tomar cuidado ao se inferir sobre processos associados ao passado evolutivo e ao fluxo gênico baseando-se em um ou poucos locos, pois a utilização de vários locos melhora as estimativas dos parâmetros, e quando há incorporação de mais locos, eles são considerados réplicas independentes (BEERLI, 1998). Outro ponto importante em relação aos locos individuais, é que a teoria de coalescência sugere que para uma melhor estimativa, deve-se aumentar a quantidade de dados não por amostragem

de mais indivíduos, mas por amostragem de mais locos (KUHNER et al. 1998), principalmente para a estimativa das taxas de migração (WAKELEY, 1998; BEERLI, 1998).

Hartl e Clark (2010), discutem acerca do número de migrantes por geração (Nm), e demonstram que um único migrante por geração é suficiente para anular parcialmente o efeito divergente da deriva genética, fazendo com que as populações se mantenham como uma única unidade evolutiva, de modo que, no presente estudo, apenas na localidade de Rosana para Teodoro Sampaio II o Nm foi acima de uma unidade.

As estimativas de fluxo gênico relativamente altas para as localidades geograficamente próximas (TSI, TSII e RA) podem indicar uma maior homogeneidade genética devido ao fluxo. Essa possível homogeneidade é corroborada por Coelho et al. (2018), que estudando os mesmos indivíduos e as mesmas populações, porém com abordagens metodológicas distintas, obtiveram dois agrupamentos bem definidos (1=TSI, TSII e RA; 2=AP) correspondentes às duas regiões de estudo (Pontal do Paranapanema e Amparo), usando o *software* desenvolvido por Pritchard et al. (2000). Esse agrupamento pode ter sido causado pela menor distância geográfica entre as populações, permitindo o fluxo e prevenindo a heterogeneidade genética entre as três localidades. De maneira antagônica, o menor fluxo gênico observado para a localidade de Amparo, em relação às demais, contribui de maneira acentuada para a diferenciação genética, evidenciando a formação de um segundo grupo.

Outro fator a ser considerado, além da grande distância geográfica para as demais populações, que pode ter contribuído para essa diferenciação de AP, é o tamanho da área onde a população está inserida. Enquanto as populações de TSI, TSII e RA estão inseridas em áreas muito maiores, ou seja, 115 ha, 155 ha e 40 ha, respectivamente, a população de AP está inserida em apenas 4 ha.

Lanes et al. (2015) utilizando seis locos microssatélites estimou para um agrupamento de quatro subpopulações de *A. aculeata*, com genótipos oriundos dos Estados da Paraíba, Pernambuco, Pará, Mato Grosso do Sul, São Paulo e Minas Gerais valores de F_{ST} que variaram entre 0,152 a 0,546, e quando utilizados na equação (3.2) correspondem respectivamente a um Nm de 1,39 e 0,54. Apesar dos autores empregarem métodos distintos para as estimativas, os valores obtidos assemelham-se aos obtidos nesse trabalho (que variou entre $Nm= 0,20$ e $Nm= 1,17$).

Oliveira et al. (2012) estudando populações de *A. aculeata* nos municípios de Montes Claros, Itacambira, Brasília de Minas, Mirabela e Grão Mogol no norte do Estado

de Minas Gerais, utilizando microssatélites, relata um Nm médio 2,4 valor expressivamente mais alto que os obtidos na presente pesquisa. Entretanto, não relata qual estimador foi utilizado para obtenção do mesmo. Baseando-se no valor relatado do $G_{ST}=0,172$ pelo mesmo autor, quando utilizado na equação (3.2) obtêm-se $Nm=1,2$ para a média das populações, valor próximo aos obtidos nesse trabalho para as populações geograficamente próximas.

O fluxo gênico tem grande interação com seleção natural, pois esta pode manter duas populações distintas mesmo se entre elas houver grande fluxo gênico (RIDLEY, 2006). Quando a magnitude da seleção for reduzida, o fluxo gênico pode homogeneizar as frequências alélicas rapidamente em populações distintas, porém, na teoria, uma grande força seletiva poderá manter a heterogeneidade gênica em duas populações, mesmo ocorrendo fluxo gênico (HARTL; CLARCK, 2010).

A capacidade que a seleção natural tem em manter populações distintas, mesmo com a presença de fluxo gênico, pode ser entendida pelo trabalho de Bradshaw (1971). Neste trabalho, o autor estudou populações da gramínea *Agrostis tenuis* em montes de rejeitos com concentrações tóxicas de metais pesados e as áreas adjacentes. Alguns genótipos de *A. tenuis* foram capazes de colonizar o ambiente contaminado, enquanto outros, que ocorriam ao redor da área, não foram capazes. Assim, pode-se dizer que a seleção natural atuou contra as sementes dos genótipos não resistentes, pois eram inviabilizadas quando em contato com o monte de rejeito. Curiosamente, fora do ambiente contaminado, a seleção natural atuou contra os genótipos resistentes, provavelmente por causa do mecanismo que expressa a tolerância requerer um “alto custo”, e quando ele não é necessário, os genótipos sem a resistência se sobressaem. As populações de *A. tenuis* mostraram divergência genética entre os indivíduos resistentes e não resistentes, de modo que o padrão observado foi favorecido devido à seleção natural. Hipoteticamente, se os montes de rejeitos fossem limpos, a tendência era de que as populações ficassem homogêneas em poucas gerações devido ao fluxo gênico, havendo menor atuação da seleção natural.

Outra força evolutiva que tem interação com o fluxo gênico é a deriva genética. Quando ocorre equilíbrio entre fluxo gênico e deriva, os valores de F_{ST} diminuem rapidamente quando o Nm aumenta, ficando claro que o fluxo gênico contribui para redução da divergência genética entre as subpopulações, citando-se os exemplos de Hartl e Clark (2010), onde para cada Nm apresenta um respectivo F_{ST} , de modo que com (Nm de 0,25, obtêm-se um $F_{ST} = 0,50$); (Nm de 0,50, obtêm-se um $F_{ST} = 0,33$); (Nm de 1,00,

obtêm-se um $F_{ST} = 0,20$); (Nm de 2,00 obtêm-se um $F_{ST} = 0,11$), de modo que quando o número de migrantes aumenta, a divergência entre as populações diminuem.

Em relação aos mecanismos que ocasionam o fluxo gênico, *A. aculeata*, é um vegetal e tem mobilidade limitada, de modo que o fluxo gênico poderá ocorrer por meio da dispersão de sementes ou dispersão do pólen, geralmente ocorrendo a combinação desses mecanismos, de maneira que o fluxo gênico existente entre as subpopulações de *A. aculeata* é dependente do sucesso migratório dos agentes de dispersão de polens e semente. Entretanto, esses eventos ocorrem de maneiras distintas, e deve-se levar em consideração a biologia reprodutiva da espécie.

O reconhecimento desses estágios de dispersão gênica é fundamental para se compreender a distribuição da diversidade (LEVIN, 1981), além de fornecer subsídios para compreensão dos padrões demográficos e formação das comunidades vegetais.

Para o fluxo gênico a partir do pólen, são relatados como os principais agentes polinizadores os insetos, sendo os coleópteros, com destaque para as famílias Nitidulidae, Curculionidae e Escarabaeidae; e as abelhas, destacando-se o grupo das Trigonas (HENDERSON et al., 1995; SCARIOT, 1998, citados por LORENZI, 2006).

Nesse sentido, a espécie *A. aculeata* apresenta características ecológicas pioneiras que facilitam sua colonização inicial no ambiente, além de produzir grande quantidade de frutos que fazem parte da dieta de animais silvestres, que são grandes dispersores das sementes, tais como como araras, periquitos, cuícas, emas, capivaras, antas, entre outros, além das áreas abertas de pastagem, servindo de alimento para o gado (POTT; POTT, 1994; SCARIOT, 1998), além das atividades extrativistas intervenções antrópicas, que ocorrem de forma corriqueira (COELHO, 2017) nas áreas de estudo.

A situação mais preocupante em relação ao tamanho efetivo estimado foi na localidade de Amparo (Tabela 5), devido ao tamanho efetivo reduzido, os efeitos causados pela deriva genética são mais acentuados, como aumento da variância entre populações, diminuição da diversidade dentro de populações, colaborando para fixação ou perda de alelos. Deve-se considerar ainda que a área tem apenas 4 ha, e está localizada praticamente em zona urbana em um local antropizado, de modo que pressões ambientais, aliadas as forças evolutivas, poderiam acarretar em uma possível extinção de genótipos locais em um prazo relativamente curto.

Apesar de não usar o mesmo estimador de tamanho efetivo via coalescência, Mori et al. (2015) com objetivo de elucidar as implicações para a coleta de sementes com fins de conservação e melhoramento genético, analisaram, utilizando locos microssatélites,

192 progênies de polinização aberta coletadas de 30 árvores matrizes de *Handroanthus heptaphyllus* (Ipê-rosa) oriundas de duas populações de Botucatu. Os autores obtiveram um tamanho efetivo de variância menor do que o esperado em progênies de meios-irmãos ($N_e=4$), o qual variou entre 2,15 a 3,81. Implicando conseqüentemente em um número maior de árvores matrizes para a coleta de sementes, devido a autofecundações e cruzamentos entre aparentados aumentarem o parentesco dentro das progênies e reduzirem o tamanho efetivo, logo, sementes devem ser coletadas de um número maior de árvores para conservação.

Tambarussi et al. (2015), estudaram três populações naturais de *Cariniana legalis* (jequitibá-rosa) localizadas uma em Ibicatu e duas em Mogi Guaçu, ambas no Estado de São Paulo/Brasil, utilizando microssatélites. No estudo, foi estimado o tamanho efetivo populacional de variância (N_e) de acordo com $N_e=0,5/\theta$, sendo θ neste caso o coeficiente de coancestria. O tamanho do censo foi de 65 árvores com ($N_e=32,5$), 22 árvores com ($N_e=10,9$) e 4 árvores com ($N_e= 2,3$), respectivamente para as populações Ibicatu, Mogi Guaçu I e Mogi Guaçu II. Os autores ressaltam que devido à endogamia e à presença de um elevado número de indivíduos aparentados nas populações, e que dessa forma, caso não ocorra fluxo de pólen, as populações terão uma redução no potencial evolutivo, e mesmo com altos níveis de diversidade genética, os altos níveis de coancestria dentro das populações podem levar à endogamia devido ao acasalamento entre os parentes, resultando em depressão por endogamia.

A comparação dos resultados do tamanho do censo (N) com tamanhos de população (N_e) usando a razão de N_e/N pode mostrar informações importantes, principalmente nos aspectos ligados a conservação e evolução da espécie, entretanto, na presente pesquisa foi realizada somente a amostragem dos indivíduos, o que impede a comparação.

7. CONCLUSÃO

A inferência de parâmetros relativos ao tamanho efetivo e fluxo gênico foi de grande importância para compreensão da genética populacional nas regiões onde *A. aculeata* foi estudada.

Foi possível estimar o fluxo gênico entre as subpopulações estudadas, o qual mostrou ser assimétrico entre os pares de populações, indicando que as relações históricas, evolutivas e demográficas, bem como os fatores associados ao fluxo gênico, são heterogêneas. O resultado evidencia que alguns métodos clássicos, que partem de pressupostos pouco realistas, podem ser insatisfatórios para fazer inferências acerca do fluxo gênico e dos processos histórico-evolutivos envolvidos.

As informações obtidas podem servir de base e auxiliar em projetos de manejo, conservação e melhoramento genético.

A situação mais preocupante em relação ao fluxo gênico e tamanho efetivo ocorre na população de Amparo, onde além de haver indícios de ocorrer barreiras geográficas e ecológicas que teoricamente impediriam o fluxo gênico no panorama atual. A área conta com apenas 4 ha localizada praticamente na zona urbana, de modo que pressões naturais ou antrópicas que alterem a dinâmica e a paisagem, poderiam acarretar em uma possível extinção de genótipos locais em um prazo relativamente curto.

8. REFERÊNCIAS

ALHO, C. J. R. Importância da biodiversidade para a saúde humana: uma perspectiva ecológica. **Estudos Avançados**, v. 26, p. 151-166, 2012.

AZEVEDO FILHO, J. A. de; COLOMBO, C. A.; BERTON, L. H. C. MACAÚBA: PALMEIRA NATIVA COMO OPÇÃO BIOENERGÉTICA. **Pesquisa & Tecnologia**, Campinas, SP, p. 1-10, 18. 2012.

BAHLO M, GRIFFITHS R. C. **Genetree**. Program and documentation distributed by the authors. Department of Mathematics, Monash University, Sydney, Australia., 1998.

BAHLO, M.; GRIFFITHS, R. Inference from gene trees in a subdivided population. **Theoretical population biology**, v. 57, p. 79-95, 2000.

BEERLI. P., FELSENSTEIN. J. Maximum likelihood estimation of migration rates and population numbers of two populations using a coalescent approach. **Genetics** v. 152, p.763-773, 1999.

BEERLI, P. Estimation of migration rates and population sizes in geographically structured populations. In **Advances in Molecular Ecology**, edited by G. Carvalho, of NATO Science Series A: Life Sciences. Amsterdam: IOS Press. v. 306, p. 39-53, 1998.

BEERLI, P. Effect of unsampled populations on the estimation of population sizes and migration rates between sampled populations. **Molecular Ecology**, v.13, p. 827-836, 2004.

BEERLI, P. How to use MIGRATE or why are Markov chain Monte Carlo programs difficult to use? **Population genetics for animal conservation**. Cambridge University Press, Cambridge, United Kingdom, p.39-77, 2009.

BEERLI, P. **MIGRATE documentation** (version 3.6). Technical Report. 2013 Disponível em: <http://popgen.sc.fsu.edu/Migrate/Migrate-n.html>. Acesso em: 14 de junho de 2018.

BEERLI, P., FELSENSTEIN, J. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. **Proceedings of the National Academy of Sciences**. n.98, p. 4563-4568, 2001.

BEERLI P., PALCZEWSKI M. Unified framework to evaluate panmixia and migration direction among multiple sampling locations. **Genetics**. v.185, p. 313-326, 2010.

BIODIESELBR. [Análise] **Plantas perenes e anuais: vantagem e desvantagem**. <<https://www.biodieselbr.com/biodiesel/analises/plantas-perenes-anuais-vantagem-desvantagem-10-03-081>> Acesso em: 07/01/2019.

BORÉM, A.; CAIXETA, E.T. **Marcadores moleculares**. 2. ed. Viçosa: Universidade Federal de Viçosa, 2009. 532 p.

BORÉM, A.; MIRANDA, G.V. **Melhoramento de plantas**. 7. edição. Viçosa: UFV, 2017.

BRITO, A. C. **Biologia Reprodutiva de Macaúba: floração, polinizadores, frutificação e conservação de pólen**. Tese (Doutorado), Universidade Federal de Viçosa-UFV, Viçosa-MG, 2013.

BRUFORD, M. W.; WAYNE, R. K. Microsatellites and their application to population genetic studies. **Current Opinion in Genetics and Development**. v.3, p.939-943, 1993.

BUSO, G. S. C. Marcadores microssatélites em espécies vegetais. **Biotecnologia, Ciência e Desenvolvimento**, v. 6, p. 46-50, 2003.

CABALLERO, A. Developments in the prediction of effective population size. *Heredity*, v. 73, p. 657-679, 1994.

CARVALHO, P.E.R. **Espécies arbóreas brasileiras**. Coleção Espécies Arbóreas Brasileiras. Vol. 3. Brasília, DF; Colombo, PR; Embrapa Florestas 2008.593 p.

CARVALHO, K.J. de; SOUZA, A.L. de; MACHADO, C.C. Macaúba-Acrocomia aculeata (Jacq.) Lood. ex Mart. Viçosa:UFV/Pólo de Excelência em Florestas/Convênio de Cooperação Técnica SECTES/FAPEMIG. 2011. 38p.

COLOMBO, C. A.; BERTON, LUIZ HENRIQUE CHORFI ; DIAZ, B. G. ; FERRARI, R. A. . Macauba: a promising tropical palm for the production of vegetable oil. **OCL. Oléagineux Corps Gras Lipides**, v. 24, p. 1, 2017.

COELHO, N. H. P. Estudos genéticos em populações naturais de Macaúba em Reservas Legais de assentamentos rurais no Pontal do Paranapanema. Dissertação de mestrado, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba. p. 69, 2017.

COELHO, N. H. P., TAMBARUSSI, E. V., AGUIAR, B. I., ROQUE, R. H., PORTELA, R. M., BRAGA, R. C., GANDARA, F. B. Understanding genetic diversity, spatial genetic structure, and mating system through microsatellite markers for the conservation and sustainable use of *Acrocomia aculeata* (Jacq.) Lodd. Ex Mart.. **Conservation genetics**, v. 19, p. 879-891, 2018.

CHARLESWORTH, B. Measures of divergence between populations and the effect of forces that reduce variability. **Molecular and Biological Evolution**, v. 15, p.538-543, 1998.

CHARLESWORTH, B. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, v. 10, n. 3, p. 195-205, 2009.

CROW, J.F; KIMURA, M.A. **An introduction to population genetics theory**. London: Harper Row, 591 p., 1970.

CRUZ CD; FERREIRA FM; PESSONI LA. **Biometria aplicada ao estudo da diversidade genética**. Visconde do Rio Branco: Suprema, 2011. 620p

DICK, C. W.; ETCHELECU, G.; AUSTERLITZ, F. Pollen dispersal of tropical trees (*Dinizia excelsa* : Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest. **Molecular Ecology** v.12, p.753-764, 2003.

DI RIENZO, A. et al. Mutational processes of simple sequences repeat loci in human populations. **Proceedings of the National Academy of Sciences of the United States of America**, v. 91, n. 3166-3170, 1994.

DE IORIO, M.; GRIFFITHS, R. C. Importance sampling on coalescent histories. I. **Adv. Appl. Prob.** v.36, p.417-433, 2004.

DOYLE, J. J.; DOYLE, J. L. Isolation of plant DNA from fresh tissue. **Focus**, v.12 p. 13-15. 1990.

EIZIRIK, E. Ecologia molecular, genética da conservação e o conceito de unidades evolutivamente significativas. **Brazilian Journal of Genetics**, Ribeirão Preto, v.19, n.4, p.23-29, 1996.

ELLSTRAND, N. C. Is gene flow the most important evolutionary force in plants? **American Journal of Botany**, v. 101, p. 737-753, 2014.

ENNOS, R. A. Estimating the relative rates of pollen and seed migration among plant populations. **Heredity**, London, v.72, p.250-259, 1994.

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

EWING, G.; RODRIGO, A. Coalescent-based estimation of population parameters when the number of demes changes over time. **Mol. Biol. Evol.** V.23, p. 988-96, 2006.

EXCOFFIER. L.; SMOUSE, P. Using Allele Frequencies and Geographic Subdivision to Reconstruct Gene Trees Within Species: Molecular Variance Parsimony. **Genetics** v.136, p. 343-359, 1994.

FAQUIN, V. **Nutrição mineral de plantas**. Lavras, Universidade Federal de Lavras, 2001. 182p.

FERREIRA, M. E.; GRATAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 2 ed. Brasília: Embrapa-Cenargen, 1998.

FISHER, R. A. **The Genetical Theory of Natural Selection**. Dover; New York 2ªed. 291 p. 1930.

FU, Y. X; LI, W. H. Coalescing into the 21st century: an overview and prospects of coalescent theory. **Theoretical Population Biology**, v.56, p. 1-10, 1999.

FUTUYMA, D. J. **Biologia evolutiva**. Ribeirão Preto: Sociedade Brasileira de Genética, p. 631, 1992.

FRANKHAM, R.; BALLOU, J.D.; BRISCOE, D.A. Introduction to Conservation Genetics. Cambridge University Press, Cambridge, p. 617, 2002.

FRANKHAM, R.; BRADSHAW, C.J.A.; BROOK, B.W. Genetics in conservation management: Revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. **Biological Conservation**. v. 170, p. 56-63, 2014.

FRANKHAM, R. How closely does genetic diversity in finite populations conform to predictions of neutral theory: Large deficits in regions of low recombination. **Heredity**, v.168, p.167-178, 2012.

GARCIA, E. S. Biodiversidade, Biotecnologia e Saúde. **Cad. Saúde Públ.**, Rio de Janeiro, v. 11, n.3, p. 495-500, 1995.

GRIFFITHS, R. C.; P. MARJORAM. Ancestral inference from samples of DNA sequences with recombination. **Journal of Computational Biology** v. 3 p. 479-502, 1996.

GRIFFITHS, R. C.; TAVARÉ, S. Sampling theory for neutral alleles in a varying environment. Philosophical Transactions of the Royal Society of London. **Series B: Biological Sciences** v. 344, p. 403-410, 1994.

GRIFFITHS, A. J. F; S.R. WESSLER; R.C. LEWONTIN; W. M. GELBART, D. T. SUZUKI, J. H. MILLER. An Introduction to Genetic Analysis. Editora W. H. Freeman; 8 ed. 2004. 800 p.

GROVER, A.; SHARMA. P. C. Development and use of molecular markers: past and presente. **Critical Reviews in Biotechnology** v. 36, p. 290-302, 2016.

GUREVITCH, J.; SCHEINER, S. M; FOX, G. A. **Ecologia Vegetal**. 2. ed. Porto Alegre: Artmed, 2009. 592 p.

HALLIBURTON, R. **Introduction to population genetics**. Pearson Education International, New Jersey.651p, 2004.

HARDY, G. H. Mendelian proportions in a mixed population. **Science** v.28, p. 49-50. 1908.

HARTL, D.L.; CLARK, A. G. **Princípios de Genética de populações**. 4^a ed. Porto Alegre: Artmed, 660 p. 2010.

HEIN, J., SCHIERUP, M. H. & WIUF, C. **Gene Genealogies, Variation and Evolution**. Oxford Univ. Press, Oxford, 2005.

HENDERSON, A.; GALEANO, G.; BERNAL, R. **Field guide to the palms of the Americas**. Princeton Princeton Univ Press. 1995. 352p.

HEY, J.; R. NIELSEN, Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. **Genetics** v.167, p. 747-760, 2004.

HODGKIN, T.; ROVIGLIONI, R.; DE VICENTE, M. C.; DUDNIK, N. Molecular methods in the conservation and use of plant genetic resources. **Acta Horticulturae**, v. 546, p. 107-118, 2001.

HOLSINGER, K. E.; WEIR, B. S. Genetics in geographically structured populations: defining, estimating and interpreting F_{ST} . **Nature Reviews Genetics** v.10, p. 639-650, 2009.

HUDSON, R.R. Gene genealogies and the coalescent process. **Oxford Surveys in Evolutionary Biology** v. 7, p. 1-44, 1990.

HUDSON, R.R.; SLATKIN, M.; MADDISON, W.P. Estimation of levels of gene flow from DNA sequence data. **Genetics**, v.132, p.583-589, 1992.

JONES, A. G.; W. R. ARDREN. Methods of parentage analysis Extension to the more general case of polygamy for in natural populations. **Mol. Ecol.** v.12, p. 2511-2523, 2003.

KAGEYAMA, P. Y. **Conservação “in situ” de recursos genéticos de plantas**. IPEF, Piracicaba, v.35, p.7-37, 1987.

KARLSSON, S.; MORK, J. Deviation from Hardy-Weinberg equilibrium, and temporal instability in allele frequencies at microsatellite loci in a local population of Atlantic cod. **ICES J. Mar. Sci.** v. 62, p. 1588-96, 2005.

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

KIMURA, M.; CROW, J.F. The Number of Alleles That Can Be Maintained in a Finite Population. **Genetics**, v.49, p. 725-738, 1964.

KIMURA, M.; OHTA, T. On some principles governing molecular evolution. **Proceedings of the National Academy of Sciences of the United States of America**, v. 71, p. 2848-2852, 1974.

KINGMAN, J. F. **Exchangeability and the evolution of large populations**. 1982A.

KINGMAN, J. F. The coalescent. **Stochastic processes and their applications** v.13, p. 235-248, 1982B.

KINGMAN, J. F., On the genealogy of large populations. **Journal of Applied Probability**. p. 27-43. 1982C.

KORDROSTAMI. M.; RAHIMIM, M. Molecular markers in plants: concepts and applications. **G3M**, v.13, p. 4024-4031, 2015

KRONE, S. M.; NEUHAUSER, C. Ancestral processes with selection. *Theor. Pop. Biol.* v.51, p.210-237, 1997.

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

KUHNER, M. K.; J. YAMATO; J. FELSENSTEIN. Maximum likelihood estimation of population growth rates based on the coalescent. **Genetics** v.149, p. 429-434, 1998.

KUHNER, M. K; SMITH, L. P. Comparing likelihood and Bayesian coalescent estimation of population parameters. **Genetics** , v.175, p. 1-47, 2007.

LANES, E. C., MOTOIKE, S. Y., KUKI, K. N., NICK, C., FREITAS, R. D. Molecular characterization and population structure of the macaw palm, *Acrocomia aculeata*(Arecaceae), ex situ germplasm collection using microsatellites markers. **Journal of Heredity**, v. 106, p.102-112,2015.

LATTA, R.G.; LINHART, Y.; FLECK, D.; ELLIOT, M. Direct and indirect estimates of seed versus pollen movement within a population of ponderosa pine. **Evolution** v.52, p.61-67, 1998.

- LEITMAN, P.; SOARES, K.; HENDERSON, A.; NOBLICK, L.; MARTINS, R.C. 2015. Arecacehamae in **Lista de Espécies da Flora do Brasil**. Jardim Botânico do Rio de Janeiro. Disponível em: <<http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB15663>>.
- LEVIN, D. A. Dispersal versus gene flow in plants. *Annals of the Missouri Botanical Garden*, p. 233-253, 1981.
- LEVINSON, G.; GUTMAN, G.A. Slipped-strand mispairing: A major mechanism for DNA sequence evolution. **Mol Biol Evol.** v.4, p.203-221, 1987.
- LI, W. H. Effect of migration on genetic distance. **American Naturalist** 110: 841-847, 1976.
- LORENZI, H. **Palmeiras do Brasil: exóticas e nativas**. Nova Odesa: Editora Plantarum, p.1-20, 1992.
- LORENZI, G. M. A. C. **Acrocomia aculeata (Jacq.) Lodd. ex Mart. - Areaceae: bases para o extrativismo sustentável**. Curitiba, 2006, 156f. Tese (Doutorado em Ciências Agrárias) Universidade Federal do Paraná, Curitiba, 2006.
- LYNCH, M; CREASE, T. The analysis of population survey data on DNA sequence variation. **Molecular Biology and Evolution** v.7, p.377-394, 1990.
- MARJORAM, P.; TAVARÉ, S. Modern computational approaches for analysing molecular genetic variation data. **Nature Reviews Genetics** v. 7, p.759-770, 2006.
- MARJORAM, P.; P. JOYCE, Practical implications of coalescent theory. In Lenwood, L. S. and Ramakrishnan, N. (eds.), *Problem Solving Handbook in Computational Biology and Bioinformatics*. New York: Springer, p. 359, 2010.
- MARJORAM, P.; S. TAVARE. Modern computational approaches for analysing molecular genetic variation data. **Nature Reviews Genetics** v.7, p. 759-770, 2006.
- MARRIAGE, T. N.; HUDMAN, S.; MORT, M. E.; ORIVE, M. E.; SHAW, R. G.; KELLY, J. K. Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). **Heredity**, v.103, n. 4, p. 310-317, 2009.

- MICHALAKIS, Y.; EXCOFFIER, L. A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. **Genetics** v. 142, p. 1061-1064, 1996.
- MORI, N. T.; MORI, E. S.; TAMBARUSSI, E. V.; MORAES, M. L. T.; SEBBENN, A. M. Sistema de cruzamento em populações de *Handroanthus heptaphyllus* (Veli.) Mattos e suas implicações para a coleta de sementes para fins de conservação e Melhoramento genético. **Scientia Forestalis**, v. 43, n. 107, p. 675-681, 2015.
- MORITZ, C. Strategies to protect biological diversity and the evolutionary processes that sustain it. **Systematic Biology**. v. 51, p. 238-254, 2002.
- MOTA, C. S.; CORREA, T. R.; GROSSI, J. A. S.; CASTRICINI, A.; RIBEIRO, A. S. Exploração sustentável da macaúba para produção de biodiesel: Colheita, pós-colheita e qualidade dos frutos. **Informe Agropecuário** (Belo Horizonte), v. 32, p. 1-11, 2011.
- NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. **Recursos genéticos e melhoramento- Plantas**, Rondonópolis: Fundação MT, 1183p. 2001.
- NASS, L. L. **Recursos genéticos vegetais. Brasília, DF: Embrapa Recursos Genéticos e Biotecnologia**, 858 p. 2007.
- NATH, H.; GRIFFITHS, R Estimation in an island model using simulation. **Theoretical population biology**. v.50, p.227-253, 1996.
- NEI, M. Analysis of gene diversity in subdivided populations. **Proceedings of the National Academy of Sciences of the United States of America**. Washington, v. 70, p. 3321-3323, 1973.
- NEIGEL, J.E. A comparison of alternative strategies for estimating gene flow from genetic markers. **Annu Rev Ecol Syst**, v. 28, p. 105-128, 1997
- NEIGEL, J. E. Is FST obsolete? **Conservation Genetics**, v.3, p. 167-173, 2002.
- NEUHAUSER, C.; KRONE, S. M. The genealogy of samples in models with selection. **Genetics** v. 145, p. 519-534, 1997.

- NORDBORG, M. Coalescent theory. In: Handbook of Statistical Genetics (eds Balding DJ, Bishop M, Cannings C) John Wiley and Sons Ltd., Chichester, UK, p. 179-212, 2001.
- NOVAES, R. F. **Contribuição para o estudo do coco macaúba**. Piracicaba, 1952. 85f. Tese- Escola Superior de Agricultura “Luiz de Queiroz” da Universidade de São Paulo, Piracicaba.
- ODONG, T.; VAN HEERWAARDEN, J.; JANSEN, J.; VAN HINTUM, T. J; VAN EEUWIJK, F. Determination of genetic structure of germplasm collections: are traditional hierarchical clustering methods appropriate for molecular marker data **Theoretical and Applied Genetics**. v. 123, p. 195-205, 2011.
- OHTA, T.; GILLESPIE, J. H. Development of neutral and nearly neutral theories. **Theoretical Population Biology**, v. 49, n. 2, p. 128-42, 1996.
- PALCZEWSKI, M.; BEERLI, P. A continuous method for gene flow. **Genetics** v.194, p. 687-696, 2013.
- PATERSON, A. H., TANKSLEY, S. D.; SORRELLS, M. E. DNA markers in plant improvement. **Advances in Agronomy**, v. 46, p. 39-90, 1991.
- POTT, A.; POTT, V. J. **Plantas do pantanal**. Corumbá: Embrapa Pantanal.1994.320p.
- PRIMACK, R. B; RODRIGUES, E. **Biologia da conservação**. Londrina: Planta, 327 p. 2001.
- PRITCHARD, J. K.; STEPHENS, M.; ROSENBERG, N. A.; Donnelly. P. Association mapping in structured populations. **The American Journal of Human Genetics**. v.67, p.170-181, 2000.
- RANNALA, B.; HARTIGAN, J. A. Estimating gene flow in island populations. **Genetical Research**, v. 67, n.2, p. 147-158, 1996.
- RAO, V.; HODGKIN, T. Genetic diversity and conservation and utilization of plant genetic resources. **Plant Cell Tissue Organ Cult**, v.68, p. 1-19, 2002.
- RIDLEY, M.. **Evolução**. 3a. ed. Porto Alegre:Artmed, 2006.

ROCHA, O.J.; AGUILAR, G. Reproductive history of the dry forest tree *Enterolobium cyclocarpum* (guanacaste) in Costa Rica: a comparison between trees left in pastures and trees in continuous forest. **Am. J. Bot.** v. 88, p. 1607-1614, 2001.

ROSA, J. R. B. F. Análise da estrutura populacional e do desequilíbrio de ligação de um painel de acessos de sorgo: uma abordagem usando teoria da coalescência. Tese de Doutorado, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba. p. 75, 2016.

ROUSSET, F. Equilibrium values of measures of population subdivision for stepwise mutation processes. **Genetics** v.142, p. 1357-1362, 1996.

SCARIOT, A. O.; LLERAS, E.; HAY, J. D. Flowering and fruiting phenologies of the palm *Acrocomia aculeata*: patterns and consequences. **Biotropica**, Washington, v.27, n. 2, p. 168-173, 1995.

SELKOE, K. A.; TOONEN, R. J. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. **Ecology Letters**, v. 9, n. 5, p. 615-629, 2006.

SEBBENN, A. M. SEOANE, C. E. S. Estimativa de tamanho efetivo de endogamia por marcadores genéticos. **R. Árvore**, Viçosa-MG, v.29, n.1, p.1-7, 2005.

SILVA, P.H.M.; SHEPHERD, M; GRATTAPAGLIA, D; SEBBENN, A.M. Use of genetic markers to build a new generation of *Eucalyptus pilularis* breeding population. **Silvae Genetica** v. 64, p. 170-181, 2016

SLATKIN, M. Gene flow in natural populations. **Annual Review of Ecology and Systematics**, Palo Alto, v. 16, p. 393-430, 1985.

SLATKIN, M.; MADDISON, W. A cladistic measure of gene flow inferred from the phylogenies of alleles. **Genetics** v. 123, p. 603-613, 1989.

SLATKIN, M.; BARTON, N. H. A comparison of three indirect methods for estimating average levels of gene flow. **Evolution**, Washington, v. 43, n. 7, p. 1349-1368, 1989.

SLATKIN, M. A measure of population subdivision based on microsatellite allele frequencies. **Genetics** v.139, p.457-462, 1993.

SLATKIN, M. Seeing ghosts: the effect of unsampled populations on migration rates estimated for sampled populations. **Molecular Ecology** v.14, p. 67-73, 2005.

SODRÉ, L.B. **Morfologia das palmeiras como meio de identificação e uso paisagístico**. Monografia- Universidade Federal de Lavras, Lavras, 2005. 62f.

SOLÉ-CAVA, A. M. Biodiversidade molecular e genética da conservação. In: MATIOLI, S. R. **Biologia molecular e evolução**. Ribeirão Preto: Holos, 2001. p. 193-198.

SOUZA, E., SORRELLS, M.E. Relationship among 70 North American oat germplasms: II. Cluster analysis using qualitative characters. **Crop Science**, Madison, v. 31, p. 605-612, 1991.

STRAND, M. et al. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. **Nature**, v. 365, p. 274-276, 1993.

TAMBARUSSI, E. V.; BOSHIER, D.; VENCOVSKY, R.; FREITAS, M. L. M.; SEBBENN, A. M. Paternity analysis reveals significant isolation and near neighbour pollen dispersal in small *Cariniana legalis* Mart. Kuntze populations in the Brazilian Atlantic Forest. **Ecology and Evolution**, Oxford, v. 5, p. 4735-5147, 2015.

TUFTO J.; ENGEN, S.; HINDAR, K. Inferring patterns of migration from gene frequencies under equilibrium conditions. **Genetics** v.144, p.1911-1921, 1996.

TURCHETTO-ZOLET, A. C. et al. **Marcadores moleculares na era genômica: metodologias e aplicações**. Ribeirão Preto: Sociedade Brasileira de Genética, 2017.

WAKELEY, J. Segregating sites in Wright's island model. **Theoretical Population Biology** v.53, p. 166-174, 1998.

WAKELEY, J. **Coalescent Theory. An Introduction**. Greenwood Village Roberts and Company Publishers, p. 326, 2009.

WEIR, B. S. **Genetic Data Analysis II**. Sinauer Associates, Sunder. Land. MA. 1996.

WILSON, E. O.; BOSSERT, W. H. **A primer of population biology**. Sinauer, Stamford, Connecticut, USA, 1971.

WILSON, G. A.; RANNALA, B. Bayesian inference of recent migration rates using multilocus genotypes. **Genetics**. v.163, p.1177-1191, 2003.

WRIGHT, S. **Evolution in mendelian populations**. *Genetics*, v.16, p.97-159, 1931.

WRIGHT, S. Size of population and breeding structure in relation to evolution. **Science**, v. 87, n. 2263, p. 430-2264, 1938.

WRIGHT, S. The genetical structure of populations. **Annual of Eugenics**. v. 15, p. 313-354, 1951.

WRIGHT, S. **Evolution and genetics of populations**. University of Chicago Press, Chicago, 511 p. 1978.

YOUNG, A.; BOYLE, T.; BROWN, T. The populations genetic consequences of habitat fragmentation for plants. **Trends in Ecology and Evolution**, London, v.11, n.10, p.413-418, 2001.

VAN OOSTERHOUT, C.; HUTCHINSON W. F.; WILLS, D. P. M. Micro Checker: software for identifying and correcting genotyping errors in microsatellite data. **Mol. Ecol. Notes**, v.4, p. 535 - 538, 2004.

VAN OOSTERHOUT, C.; WEETMAN, D.; HUTCHINSON, W.F. Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. **Mol Ecol Notes**. v.4, p.535-538, 2006.

VENCOVSKY, R.; CROSSA, J. Measurements of representativeness used in genetic resources conservation and plant breeding. **Crop Science**, v. 43, p. 1912-1921, 2003.

ZHONG-HU, H. An investigation of the relationship between the F1 potential and the measures of genetic distance among wheat lines. **Euphytica**, v. 58, p. 65-170, 1991.

ZHU, C.; GORE, M.; BUCKLER, E. S; YU, J. Status and prospects of association mapping in plants. **The plant genome**. v.1, p. 5-20, 2008.