



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Campus de Botucatu



ANTONIO HIGO MOREIRA DE SOUSA

COEFICIENTES DE PARENTESCO EM ESPÉCIES FLORESTAIS

Botucatu

2018

ANTONIO HIGO MOREIRA DE SOUSA

COEFICIENTES DE PARENTESCO EM ESPÉCIES FLORESTAIS

Dissertação apresentada à Faculdade de Ciências Agronômicas da Unesp Campus de Botucatu, para obtenção do título de Mestre em Ciência Florestal.

Orientador: Prof. Dr. Evandro Vagner Tambarussi

Coorientador: Prof. Dr. Cristiano Bueno de Moraes

Botucatu

2018

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉCNICA DE AQUISIÇÃO E TRATAMENTO DA INFORMAÇÃO - DIRETORIA TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - UNESP - FCA - LAGEADO - BOTUCATU (SP)

S725c Sousa, Antonio Higo Moreira de, 1993-
Coeficientes de parentesco em espécies florestais /
Antonio Higo Moreira de Sousa. - Botucatu: [s.n.], 2018
65 p.: ils., grafs., tabs.

Dissertação (Mestrado) - Universidade Estadual Paulista, Faculdade de Ciências Agrônomicas, Botucatu, 2018
Orientador: Evandro Vagner Tambarussi
Coorientador: Cristiano Bueno de Moraes
Inclui bibliografia

1. Espécies florestais. 2. Seleção de plantas - Melhoramento genético. 3. Florestas - Reprodução. 4. Modelagem de dados. I. Tambarussi, Evandro Vagner. II. Moraes, Cristiano Bueno de. III. Universidade Estadual Paulista "Júlio de Mesquita Filho" (Câmpus de Botucatu). Faculdade de Ciências Agrônomicas. IV. Título.

Elaborada por Ana Lucia G. Kempinas - CRB-8:7310

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte"

CERTIFICADO DE APROVAÇÃO

TÍTULO DA DISSERTAÇÃO: COEFICIENTES DE PARENTESCO EM ESPÉCIES FLORESTAIS

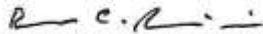
AUTOR: ANTONIO HIGO MOREIRA DE SOUSA

ORIENTADOR: EVANDRO VAGNER TAMBARUSSI

Aprovado como parte das exigências para obtenção do Título de Mestre em CIÊNCIA FLORESTAL,
pela Comissão Examinadora:


Prof. Dr. EVANDRO VAGNER TAMBARUSSI
Engenharia Florestal / Universidade Estadual do Centro-Oeste


Prof. Dr. CELSO LUÍS MARINO
Genética / Instituto de Biociências de Botucatu - UNESP


Dr. BRUNO CESAR ROSSINI
IBTEC - Instituto de Biotecnologia / Instituto de Biociências de Botucatu

Botucatu, 06 de setembro de 2018

*A Deus, aos meus amados pais, Iris (in
memory) e Antonio e minha noiva Lunara*

dedico

AGRADECIMENTOS

Agradeço a Deus pela força para percorrer este longo caminho até esta conquista. “Bendiga o Senhor a minha alma! Não esqueça nenhuma de suas bênçãos! Salmos 103:2.”

Aos meus pais, que são responsáveis pela minha educação e moral, na pessoa de Iris Moreira de Sousa, minha mãe, a quem sempre me disciplinou e inspirou a buscar por conhecimento e na pessoa de Antonio Borges de Sousa, meu pai, num exemplo de dedicação total a família, entrega a seus filhos, apostando sempre no meu futuro e mostrando um amor que me motiva a sonhar.

Ao meu irmãos Hugo e Lívia, pela confiança e apoio sempre depositados.

Aos demais familiares, pela união e fé depositada em mim, sendo um alicerce forte no qual está fundamentado meus valores.

A minha noiva e companheira Lunara de Alencar, que me acompanhou em cada momento deste capítulo da minha vida.

A amada Igreja de Cristo, que serviu de conforto espiritual e físico, acolhendo nos momentos que mais precisei e sendo alicerce para que a palavra de Deus não fosse apagada pelas lutas diárias e encorajando a seguir em frente.

Aos meus queridos amigos e pastores Gustavo Machado e Eduardo, ao qual como um filho trataram-me e sempre serão pessoas a quem pude e posso recorrer.

Ao projeto Testemunhar, pela cobertura de suas orações e a força nos momentos de fraqueza.

Ao Laboratório de Melhoramento Florestal, pelos momentos que me proporcionaram e por dividirem seu conhecimento comigo.

A republica Zona Azul pelo acolhimento e aos laços de amizades criados.

A Doutoranda, coach de apresentação e revisora do trabalho Fernanda Bortolanza pela paciência, apoio e disponibilidade em ajudar, ao Mestrando Renan Portela e a Doutoranda Isabel Homczinski pelos toques na dissertação e amizade criada.

Agradecimento pelo presente trabalho que foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Ao meu orientador e professor Evandro Vagner Tambarussi, pela confiança depositada e pela ajuda ao meu crescimento como profissional, disponibilizando ferramentas e direcionamento para a produção deste documento.

RESUMO

Parentesco entre indivíduos em populações naturais é uma informação com múltiplos usos e pode ser acessada por meio de estimadores que usam dados moleculares. Todavia, cada estimador possui pressuposições e, muitas vezes, rigorosas para espécies florestais. Assim, a modelagem dos dados é uma etapa importante e, quase em todos os casos, negligenciada. Este trabalho teve como objetivo avaliar a eficácia de diferentes estimadores de parentesco em *Acrocomia aculeata* (Jacq.) Lodd. ex Mart., *Hymenaea stigonocarpa* Mart. ex Hayne e *Dipteryx alata* Vogel., bem como em quatro diferentes populações simuladas. A partir das coancestrias médias ($\bar{\theta}$) estimadas, foi calculado o erro dos estimadores pressupondo que os indivíduos analisados eram meios-irmãos ($\bar{\theta}=0,125$). As estimativas de parentesco oscilaram conforme a espécie e o método utilizado, gerando diferentes valores de erro para cada estimador. A correlação foi observada apenas entre estimadores que possuíam o mesmo método de estimativa ou com pressupostos similares. As populações simuladas tiveram melhores valores estimados e menores erros em comparação com dados reais. Os valores de erro dos estimadores encontrados, demonstram que somente aplicação dos estimadores para a inferência de determinado grau de parentesco, pode gerar resultados viesados e corroborar para ineficácia da tomada de decisão, sendo necessário o uso de informações complementares associadas ao parentesco, como análise do sistema de reprodução, estrutura genética da população, permitindo inferir com maior precisão o parentesco dos indivíduos avaliados.

Palavras-chave: Estimadores de parentesco. Espécies florestais. Seleção Genômica. Simulações. sistema reprodutivo misto.

ABSTRACT

Understanding kinship between individuals in natural populations offers useful information that can be assessed based on estimators of molecular data. However, each estimation method is based on assumptions and is often restricted for forest species. Thus, modeling the data is an important step that is almost always neglected. This study aims to evaluate the efficacy of different kinship estimators for *Acrocomia aculeata* (Jacq.) Lodd, ex Mart., *Hymenaea stigonocarpa* Mart. Ex Hayne, and *Dipteryx alata* Vogel., and four different simulated populations. From the estimated mean coancestry ($\bar{\theta}$), the error estimates were calculated assuming that analyzed individuals were perfect half-siblings ($\bar{\theta}=0,125$). Estimates of kinship ranged according to the species and method used, generating different error values for each estimator. A correlation was observed only between estimators that used the same estimation method or similar assumptions. Simulated populations showed more accurate estimates and lower error values compared to actual data. The error values of the estimators demonstrate that the application of estimators to infer a certain degree of kinship can generate biased results and lead to inefficient decision making. Thus, the use of complementary information associated with kinship is necessary, such as analysis of the reproduction system and genetic structure of the population, enabling more precise inferences of the kinship between evaluated individuals.

Key words: Kinship estimators. Forest species. Genomic Selection. Simulations. mixed reproductive system.

LISTA DE FIGURAS

Figura 1 - Exemplo de arquivo de saída gerado pelo pacote <i>stringr</i> para a simulações de populações para o estudo de estimativas de parentescos.	32
Figura 2 - Distribuição das coancestrias estimadas pelos métodos de momentos em <i>Acrocomia aculeata</i>	38
Figura 3 - Distribuição das coancestrias estimadas pelos métodos de momentos em <i>Dipteryx alata</i>	39
Figura 4 - Distribuição das coancestrias estimadas pelos métodos de momentos em <i>Hymenaea stigonocarpa</i>	40
Figura 5 - Distribuição das coancestrias estimadas pelo método da verossimilhança para <i>Dipteryx alata</i> , <i>Hymenaea stigonocarpa</i> e <i>Acrocomia aculeata</i> respectivamente.	41
Figura 6 - Correlações entre os estimadores de coancestria (θ) para <i>Acrocomia aculeata</i>	46
Figura 7 - Correlações entre os estimadores de coancestria (θ) em <i>Hymenaea stigonocarpa</i>	47
Figura 8 - Correlações entre os estimadores de coancestria (θ) em <i>Dipteryx alata</i>	48

LISTA DE TABELAS

Tabela 1 - Locus microssatélites utilizados para a avaliação de diferentes modelos de estimativa de parentesco para as espécies.....	31
Tabela 2 - Descrição de população simulada de meios irmãos (<i>half-sibs</i>) para estimativa da coancestria por diferentes estimadores.....	32
Tabela 3 - Estados de <i>IBD</i> e suas probabilidades considerando quatro genes (Adaptado de Wang, 2007).	35
Tabela 4 - Relação do número de locos utilizado para estimativa da coancestria de <i>Acrocomia aculeata</i> e <i>Dipteryx alata</i>	35
Tabela 5 - Média e variância dos estimadores de coancestria para as espécies <i>Acrocomia aculeata</i> , <i>Hymenaea stigonocarpa</i> e <i>Dipteryx alata</i>	42
Tabela 6 - Erros dos quadrados médios (<i>MBE</i> em %), raiz dos erros quadrados médios (<i>RMSE</i> em %) para coancestrias dos diferentes estimadores para <i>Acrocomia aculeata</i> , <i>Hymenaea stigonocarpa</i> e <i>Dipteryx alata</i>	45
Tabela 7 - Erros dos quadrados médios (<i>MBE</i> em %), raiz dos erros quadrados médios (<i>RMSE</i> em %) para as estimativas de coancestrias dos diferentes modelos para as populações simuladas.....	51

SUMÁRIO

1	INTRODUÇÃO	19
2	REVISÃO BIBLIOGRÁFICA	22
2.1	Marcadores Microssatélites	22
2.2	Sistemas de Reprodução	23
2.3	Parentesco	24
2.4	Marcadores Moleculares <i>versus</i> Relações de Parentesco	26
2.5	Utilizações de Estimativa de Parentesco em Análises Genéticas	29
3	MATERIAL E MÉTODOS	31
3.1	Espécies e Marcadores Moleculares	31
3.2	POPULAÇÕES SIMULADAS	31
3.3	Coeficientes de Parentesco	33
3.4	Análise Comparativa dos Estimadores	35
4	RESULTADOS E DISCUSSÃO	37
4.1	Análise dos Estimadores para as Espécies	37
4.2	Análise dos Estimadores nas Populações Simuladas	51
4.3	Implicações do Uso de Estimadores de Coancestria na Conservação e Melhoramento De Espécies	53
5	CONCLUSÃO	55
	REFERÊNCIAS	57

1 INTRODUÇÃO

O parentesco tem importante aplicação na genética quantitativa, genética da conservação, genética forense, evolução e ecologia (WEIR et al., 2006). Isto porque as relações genéticas entre os indivíduos de uma população são informações importantes sob estes vários aspectos (CHARPENTIER et al., 2012; MATTILA et al., 2012; LIU et al., 2013).

Do ponto de vista genético, essa informação pode ser utilizada no planejamento de programas de melhoramento, por exemplo, para estimar parâmetros genéticos (FALCONER e MACKAY, 1996; RITLAND, 2012) de caracteres relacionados à produção (GYNCH e WALSH, 1998), uma vez que as relações de parentesco afetam o desempenho das progênies (NICHOLAS e SMITH, 1983).

Em relação ao aspecto ecológico ou conservacionista, o conhecimento das relações entre os indivíduos auxilia no gerenciamento de programas de melhoramento para conservação de espécies (JONES et al., 2002; KOZFKAY et al., 2008; GONCALVES DA SILVA et al., 2010; BERGNER et al., 2014), além de ser usado em estudos de depressão endogâmica (GRUEBER et al., 2010; NIELSEN et al., 2012; TAMBARUSSI et al., 2017a) e de tamanho efetivo das populações sob conservação e melhoramento (VENCOVSKY e CROSSA, 2003).

O parentesco pode ser entendido por meio do conceito de similaridade genética, definido por Malecot (1969) como a probabilidade de que dois alelos do mesmo loco, de dois indivíduos (i e j), serem idênticos por descenderem de um ou mais ancestrais comuns. Esta relação entre indivíduos pode ser explicada ou estimada por meio do coeficiente de parentesco ou do coeficiente de coancestria.

O coeficiente de coancestria (θ_{xy}) é a probabilidade de que dois alelos homólogos, retirados ao acaso de um par de indivíduos (x e y), sejam idênticos por descendência, enquanto que o coeficiente de parentesco (r_{xy}) mede a proporção de alelos entre dois indivíduos que são idênticos por descendência (LYNCH e WALSH, 1998; SEBBEN, 2005).

A coancestria é uma probabilidade que pode assumir valores entre 0 e 1, dependendo das relações dos indivíduos, variando desde não aparentado com $\theta_{xy}=0$ ou $r_{xy} = 0$ até irmãos completos com $\theta_{xy}= 0,25$ ou $r_{xy}= 0,5$; entre outras relações intermediárias. Como esta medida corresponde as relações biológicas que os

indivíduos podem ter na natureza, valores fora do intervalo entre 0 a 0,25 ou 0 a 0,5 não são considerados biologicamente corretos, incorporando assim, erros aos estudos de populações (MILLIGAN, 2003).

Desta forma, o conhecimento do sistema de reprodução das espécies é de fundamental importância, uma vez que determina a estrutura genética na geração descendente. A taxa de cruzamento, presença ou não de endogamia, além da coancestria dentro de progênies afetam o tamanho efetivo de amostras a serem coletadas, tornando essa informação prévia a chave para amostrar corretamente as populações naturais de espécies de interesse (DE MORAES et al., 2007).

Em relação ao sistema de reprodução, o estudo de espécies com sistema de reprodução misto tem ganhado enfoque, visto que estas espécies possuem heterogeneidade de indivíduos quanto ao grau de endogamia e parentesco que compõe a população (VENCOVSKY et al., 2001; TAMBARUSSI et al., 2017b). Dentre várias espécies nativas de relativa importância, podemos citar a *Acrocomia aculeata* (Jacq.) Lodd. ex Mart., *Hymenaea stigonocarpa* Mart. ex Hayne e a *Dipteryx alata* Vogel.

A. aculeata, também conhecida como macaúba, pertence à família *Arecaceae*, tem ampla distribuição no Brasil, ocorrendo em todas regiões do país desde o Amazonas até o Paraná, sendo abundante no bioma do cerrado (LORENZI, et al., 2004; FLORA DO BRASIL, 2018). A espécie possui sistema reprodutivo alógamo com algumas populações mistas (ABREU et al., 2012; BERTON et al., 2013; COELHO et al., 2018) e grande variabilidade genética, tanto dentro como entre populações (NUCCI et al., 2008; OLIVEIRA et al., 2012). Além disso, *A. aculeata* tem grande potencial energético pelo fato de que os óleos extraídos da polpa e da semente podem ser usados em indústrias de biocombustível e farmacêutica (COIMBRA e JORGE, 2012; BERTON et al., 2013)

H. stigonocarpa, conhecida por jatobá, ocorre nas regiões, Norte, Nordeste, Centro-Oeste e Sudeste (FLORA DO BRASIL, 2018), e também foi classificada como espécie com sistema reprodutivo misto (MORAES et al., 2007; MORAES e SEBBENN, 2011; MORENO, 2009). Uma característica importante da espécie é seu potencial como planta medicinal e uso na indústria farmacêutica (ORSI et al., 2012; 2014).

Por fim, *D. alata* não ocorre apenas na Região Sul (FLORA DO BRASIL, 2018) e primeiramente foi classificada por Oliveira e Sigris (2003) como alógama, com autoincompatibilidade de ação tardia e taxa de aborto alta. Todavia, Tambarussi et al. (2017b) constataram que a espécie tem sistema de reprodução misto, além de forte variação individual na taxa de cruzamento. A espécie pode ser utilizada para alimentação e com potencial para extração de óleo (FETZER et al., 2018; LEMOS et al., 2012; MARTINS et al., 2013).

Comumente, estimativas como o índice de fixação (F) ou do coeficiente de coancestria (θ), são obtidas e usadas para realizar inferência sobre populações naturais, com base nas relações dos indivíduos. Para tais inferências, o uso de simulações pode ser um bom recurso tanto para selecionar o melhor estimador, bem como verificar a confiabilidade da estimativa gerada, possibilitando uma maior robustez de possíveis conclusões (TAILOR, 2015). Vale ressaltar também que o uso de testes de progênies de polinização aberta e consequente estimativa do sistema de reprodução, juntamente com essas simulações, proporcionam uma maior confiabilidade para quaisquer estimativas de parentescos, pois a partir dessas técnicas e estimativas são geradas informações que auxiliam no planejamento e manejo dos indivíduos avaliados.

Desta forma, o presente estudo objetivou avaliar a eficiência de diferentes estimadores de parentesco em indivíduos de três espécies florestais de sistema reprodutivo misto para elucidar as seguintes hipóteses:

- i. o sistema de reprodução misto influencia no desempenho dos estimadores de parentesco e
- ii. há um estimador de parentesco adequado para cada espécie em estudo.

2 REVISÃO BIBLIOGRÁFICA

2.1 Marcadores microssatélites

Atualmente, há muitas tecnologias em genética molecular disponíveis para serem aplicadas em estudos de genética de populações, ecologia e no melhoramento genético, sendo que uma das técnicas mais eficientes é o uso de marcadores moleculares, devido ao seu poder de detectar diferenças genéticas sem o efeito de fatores ambientais (SANTOS, 2016).

Marcadores moleculares são definidos como um segmento específico de DNA que permite fazer inferências diretas sobre a diversidade genética e inter-relações entre os organismos (TURCHETTO-ZOLET et al., 2017). A aplicação de marcadores moleculares a um tipo específico de estudo partirá da escolha daquele que melhor se adapte às necessidades do pesquisador, tais como o objetivo do estudo, infraestrutura disponível, recursos financeiros, disponibilidade de mão de obra especializada, entres outros (FALEIRO, 2007; MIAH et al., 2013).

Há vários tipos de marcadores moleculares, sendo os microssatélites um dos mais utilizados, definidos como sequências simples de um a seis pares de bases, denominadas de regiões microssatélites, repetidas em *tandem* ao longo do DNA, (LITT e LUTY, 1989; MIAH et al., 2013). Os marcadores moleculares do tipo microssatélites também são conhecidos como STR (*Short Tandem Repeats*), SSLP (*Simple Sequence Length Polymorphism*) e SSR (*Simple Sequence Repeat*) (BOREM, 2009; FALEIRO, 2007).

Os SSRs são altamente polimórficos, em consequência de suas taxas de mutação elevadas, exibindo variações que se manifestam como diferenças no número de repetições da sequência do DNA entre indivíduos de uma amostra (GUICHOUX et al., 2011; MIAH et al., 2013). Soma-se às várias vantagens deste tipo de marcador, a relativa facilidade de transferência entre espécies intimamente relacionadas.

2.2 Sistemas de reprodução

O sistema reprodutivo pode ser entendido como a maneira que uma espécie transmite seus genes em cada geração, a fim de gerar descendentes, podendo gerar maior ou menor variabilidade genética (BOSHIER, 1996; GIUSTINA et al., 2018).

As espécies podem ser classificadas quanto ao sistema de reprodução, como alógamas, autógamas ou mistas. As espécies são consideradas alógamas quando a sua forma de reprodução se dá por acasalamentos entre plantas, apresentando uma taxa de cruzamento naturais superior a 95% (ALLARD, 1999; BESPALHOK, 2014; SOUZA JR, 2001), enquanto em espécies autógamas, esta taxa é inferior a 5%, se reproduzindo, portanto, predominantemente por autofecundações (95%) (ALLARD, 1999; RAMALHO et al., 2001; BESPALHOK, 2014). Espécies cujo a taxa de cruzamentos se encontra no intervalo entre 5% e 95% são consideradas como espécies com sistema misto de reprodução (BESPALHOK, 2014; GOODWILLIE et al., 2005; VENCOVSKY et al., 2001;).

Embora o sistema misto, para muitos autores, não seja considerado um sistema de reprodução evolutivamente estável, sendo tomado como apenas um estágio intermediário entre a alogamia e autogamia, vários são os trabalhos que enquadram espécies neste tipo de sistema e ressaltam a sua influência para o entendimento do comportamento das populações (ARISTA et al., 2017; GOODWILLIE et al., 2005; TAMBARUSSI et al. 2017a, 2017b). As espécies arbóreas tropicais geralmente apresentam sistema misto de reprodução com predomínio de cruzamentos, gerando progênes compostas por misturas de irmãos de autofecundação e meios-irmãos (SOUZA et al., 2003).

Um dos fatores que pode explicar o desenvolvimento dos sistemas mistos de reprodução é a forma de dispersão do pólen. Espécies que são polinizadas por vetores bióticos possuem maior tendência a apresentar sistema misto de reprodução, isto porque na falta de polinizadores para realizar cruzamentos, a espécie utiliza da autofecundação como um mecanismo de garantia reprodutiva (KALISZ e VOGLER, 2003; WHITEHEAD et al., 2018). Todavia, para espécies polinizadas por vetores

abióticos há uma baixa ocorrência de sistema misto que se atribui à maior estabilidade que os fatores abióticos oferecem (FRIEDMAN e BARRETT, 2009).

Conhecer o sistema reprodutivo de uma espécie permite obter maior acurácia ao se fazer inferências sobre a composição da população e as relações genéticas entre os indivíduos, além de contribuir para o desenvolvimento de estratégias mais eficientes para a conservação de seus recursos genéticos (SEBBENN, 2002; BREIER, 2008). O sistema reprodutivo tem importância ecológica uma vez que determina o nível e a distribuição genética dos indivíduos, além da variabilidade dentro e entre populações, por controlar o grau de parentesco e tamanho efetivo destas populações (GOODWILLIE et al., 2005; SEBBENN et al., 2011).

A variabilidade genética possibilita a sobrevivência e reprodução de uma espécie frente as adversidades ambientais, por corresponder a um maior número de alelos distintos na população que geram uma diversidade de genótipos, que podem sobreviver aos estresses e propagar seus genes (SILVA et al., 2014; VINSON et al., 2015), sendo a conservação desta variabilidade uma ferramenta importante para a conservação das espécies (SERROTE et al., 2016).

Da mesma forma que o aumento da variabilidade genética promove efeitos significativos numa população, a perda desta variabilidade pode gerar danos consideráveis a uma espécie. A curto prazo, a perda de variabilidade pode causar suscetibilidade a doenças, pragas ou eventos climáticos (geadas, estiagens), além do aumento na taxa de cruzamentos biparentais (endogâmicos), levando a depressão por endogamia (SERROTE et al., 2016; TAMBARUSSI et al, 2017a). A longo prazo, uma menor variabilidade genética leva a uma diminuição da resposta à pressão de seleção do ambiente (SERROTE et al., 2016).

2.3 Parentesco

De acordo com o sistema de reprodução de uma espécie, ela pode transmitir seus genes gerando descendentes com distintos graus de parentesco entre si (meios-irmãos, irmãos completos, irmãos de autofecundação, etc.) (GIUSTINA et al., 2018). As estimativas de parentesco, geralmente são feitas com base no conceito de alelos *identical by descent* (IBD) ou idênticos por descendência, propondo que, para

indivíduos relacionados, em uma população infinita, a probabilidade de um único loco ser *IBD* entre os gametas é igual à correlação no tipo alélico (THOMPSON, 2013).

O *IBD* é um conceito que pode ser entendido como um fragmento do DNA atual de um indivíduo que é a cópia idêntica do DNA de seu ancestral, gerando semelhanças genéticas nos indivíduos de uma população (THOMPSON, 2013). O *IBD* pode ser estimado por meio de genealogia, que determina a probabilidade de que alelos amostrados ao acaso, em um par de gametas segregantes, sejam idênticos por descendência (WRIGHT, 1922). Quando os alelos analisados em dois indivíduos são idênticos, porém não provêm de um ancestral comum, são considerados *identical in state* (IIS) ou idêntico por estado (MIRANDA FILHO, 2001).

Outro conceito ligado a *IBD* é a coancestria (θ_{xy}). Este conceito foi proposto pela primeira vez por Cockerham e Weir (1983) juntamente com outras medidas de *IBD*, como parâmetros importantes, capazes de informar sobre homozigosidade, deriva e endogamia em uma população. Posteriormente, Weir e Cockerham (1984) desenvolveram as seguintes definições e notações:

- F : a correlação entre genes dentro de indivíduos ou coeficiente de endogamia;
- θ : a correlação entre genes de diferentes indivíduos da mesma população ou coeficiente de coancestria.

Uma das formas de obter a estimativa de parentesco entre indivíduos é pelo uso de estimadores baseados em marcadores moleculares. Estes estimadores nada mais são que modelos estatísticos que estimam as relações entre indivíduos (coeficiente de parentesco ou coeficiente de coancestria) por meio de informações genéticas (frequência alélica, número de locos) fornecida pelos marcadores (BLOUIN, 2003).

Os estimadores com base em marcadores moleculares podem estimar o coeficiente de coancestria (θ_{xy}) ou coeficiente de parentesco (r_{xy}). Coeficiente de coancestria é a probabilidade de que dois alelos homólogos, selecionados ao acaso de um par de indivíduos x e y , sejam idênticos por descendência (LYNCH e WALSH, 1998; SEBBEN, 2005), enquanto que o coeficiente de parentesco (r_{xy}) mede a

proporção esperada de alelos compartilhados entre pares de indivíduos que são idênticos por descendência (BLOUIN, 2003).

Vários estimadores podem ser empregados para análise de parentesco, cada um tendo suas estimativas influenciadas por diferentes aspectos, como, por exemplo, as características dos marcadores moleculares (número de locos avaliados, polimorfismo), o tamanho da amostra, os erros de genotipagem, a mutação e a endogamia (CAPPA et al., 2016). Entre os marcadores moleculares, os microssatélites se destacam como uma ferramenta eficiente para estimar e/ou elaborar matrizes de parentesco, devido à natureza altamente informativa (TELFER et al., 2015), além de inclusive superar os marcadores SNP's (*Single Nucleotide Polymorphism*), em virtude do multialelismo (CAPPA et al., 2016).

Entre as aplicações da informação de parentesco pode-se listar a facilidade de estimativa do fluxo gênico contemporâneo (BURCZYK et al., 2002), o controle da endogamia e manutenção da diversidade genética adequada em programas de melhoramento, praticando-se seleção a partir de informações de parentesco (TELFER et al., 2015) e a estimativa dos valores genéticos das progênes e dos genitores (EL-KASSABY et al., 2011).

O parentesco ainda pode afetar o desempenho da seleção genômica em populações naturais, considerando a aplicação de seleção em uma amostra contendo subpopulações com diferentes relações de parentesco, afetando a estrutura da população (ASORO et al. 2011; DESTA e ORTIZ, 2014).

2.4 Marcadores moleculares *versus* relações de parentesco

O conhecimento do grau de parentesco é fundamental em muitas áreas de pesquisa como genética quantitativa, genética da conservação, estudos forenses, evolução e ecologia, onde a informação pode ser obtida de várias formas (RITLAND, 1996; LYNCH e RITLAND, 1999; BONATELLI). Uma delas é o uso de *pedigree* (informações dos genitores), em que, a partir deste, os coeficientes de coancestria ou parentesco podem ser estimados e posteriormente utilizados para estimativa de parâmetros genéticos (LYNCH e WALSH, 1998).

Todavia, o uso de *pedigree* pode ser dificultado em decorrência da baixa quantidade ou inexistência de informações para a sua construção, principalmente para espécies não domesticadas, além do problema do *pedigree* pressupor a inexistência de indivíduos aparentados, o que leva subestimação das relações de parentesco e aumento do viés nas estimativas (PEMBERTON, 2008; TAYLOR, 2015).

O uso de dados de marcadores moleculares pode ser uma saída para melhoria deste cenário uma vez que resulta em maior precisão na predição de graus de parentescos, produzindo “*pedigrees*” com maior confiabilidade e estimativas mais robustas (PEMBERTON, 2008). Estes marcadores também podem ser utilizados diretamente para estimativa de parentesco ou endogamia, por meio de estimadores de parentesco, prática que tem crescido com o surgimento de marcadores para diferentes espécies (MILLIGAN, 2003).

Os estimadores de parentesco podem ser agrupados em dois grupos: os com base no método dos momentos e os baseados em métodos de verossimilhança. Os estimadores fundamentados nos métodos dos momentos buscam estimar a relação entre os indivíduos em termos de probabilidades de identidade por descendência (*IBD*). Já os baseados em métodos de verossimilhança, estimam a probabilidade de indivíduos se encaixarem em um dado parentesco (TAYLOR, 2015).

Cada tipo de estimador possui suas vantagens, desde que suas pressuposições sejam atendidas. Os estimadores que se baseiam nos métodos dos momentos são imparciais (WANG, 2002), mas podem assumir valores negativos devido a erros de amostragem. Os estimadores de probabilidade possuem um intervalo que varia entre 0 a 1, com base no conceito de *IBD* e, como resultado, são viesados e superestimam as estimativas quando as pressuposições são violadas.

Dentre os estimadores baseados em métodos de momentos pode-se listar os mais usuais que são:

- Queller e Goodnight (1989): é um estimador desenvolvido para loco único, mas com várias formas de aplicação, por isso bastante popular e que pode ser facilmente aplicado pra amostras multilocos (LYNCH e RITLAND, 1999);

- Ritland (1996): é um modelo de correlação com base nas distribuições conjuntas de ambos os genótipos, condicionada às frequências alélicas (p_i), de maneira simétrica, para estimar o parentesco (r). Entretanto, uma desvantagem deste modelo, é que um dos pesos (S_i) assume o valor de zero quando um dos indivíduos analisados não possuem alelos semelhantes, gerando estimativas de parentesco nulas ou negativas;
- Loiselle et al. (1995): é um estimador de “coancestria” (ρ_{xy}) que analisa a autocorrelação espacial dos indivíduos, para entender a estrutura genética da sua população de estudo. Todavia, para uma amostra mista (níveis variados de parentesco ou sua ausência), este pode acarretar em erros na estimativa da coancestria. Na presença de indivíduos sem grau de parentesco, o método superestima o parâmetro enquanto que na ocasião de indivíduos aparentados, o subestima (WANG, 2017);
- Lynch e Ritland (1999): é um estimador de regressão bastante simples e de fácil aplicação que usa o princípio da similaridade (LYNCH, 1988; LI et al., 1993). Os pesos utilizados pelo estimador diferem de acordo com o genótipo de referência para um mesmo loco, influenciando a variância das estimativas de coancestria ($\hat{\theta}$) e, assim como o estimador de Ritland (1996), assumem parentescos nulos ($r=0$);
- Wang (2002): é um estimador criado a partir do índice de similaridade de Lynch (1988) e Li et al. (1993), mas que pode estimar o parentesco por meio de dois e quatro genes, seguindo à definição de parentesco do *IBD*.

Como exemplo de estimados que utilizando métodos de máxima verossimilhança tem-se os seguintes estimadores:

- Thompson (1975): é um modelo que assume que os indivíduos em questão são oriundos de uma única população em equilíbrio de Hardy-Weinberg, utilizando equações de probabilidade. Este método constrói razões de verossimilhança que comparam as probabilidades de os dados de marcadores assumirem determinado tipo de relação entre os indivíduos dentro de suas hipóteses (ANDERSON e WEIR, 2007);

- Milligan (2003): é um estimador que usa o mesmo princípio de Thompson (1975), utilizando os modelos de identidade (*IBD*) propostos por Jacquard (1974), estimando o parentesco;
- Wang (2007), utiliza os nove modelos de identidade (*IBD*) propostos por Jacquard (1974), diferente de Milligan (2003), que utiliza apenas três identidades.

Além disso, os estimadores podem ser afetados pela quantidade de informação genética ou pelas características da população estudada. A composição de uma população (VAN DE CASTEELE et al., 2001; CSILLÉRY et al., 2006), número de marcadores (ANDERSON e WEIR, 2007; BLOUIN, 2003; VAN DE CASTEELE et al., 2001;), polimorfismos dos marcadores (BLOUIN, 2003; WANG, 2014), alterações no nível de informação destes fatores podem causar mudanças no viés dos valores estimados, sendo que o nível de polimorfismo possui maior efeito sobre a performance dos estimadores.

Em virtude das especificidades de cada estimador, cada um deles possui um cenário adequado que maximiza seu desempenho, ou seja, não há um estimador universal indicado para qualquer situação. Isto porque a extensão do viés varia em favor do grau de parentesco real, o que torna impossível o uso de uma única correção para as diferentes situações ou cenários que os estimadores podem ser aplicados (WANG, 2017). Para contornar estes problemas, não se pode abrir mão de simulações para selecionar o estimador mais apropriado para um determinado cenário, uma prática que agrega maior segurança e confiabilidade na escolha do estimador (VAN DE CASTEELE et al., 2001; WANG, 2011).

2.5 Utilizações de estimativa de parentesco em análises genéticas

Estimar o parentesco a partir da informação de marcadores moleculares permitiu muitos estudos genéticos em populações naturais de várias espécies de plantas e animais, além de proporcionar maior robustez a metodologias que usam informações de parentesco (DEWOODY, 2005; GARANT e KRUIK, 2005).

Uma das várias aplicações de estimadores de parentesco com base em marcadores genéticos é a estimativa da diversidade genética de populações. Hansen et al. (2017), utilizaram estimadores de parentesco para estimar endogamia e parentesco em populações de *Tectona grandis* L. F. de três regiões diferentes, selecionando os estimadores com base em simulações, o que indicou que a variação nas populações estudadas provavelmente reflete suas regiões de origem genética.

Browne e Karubian (2018), estimaram o parentesco de matrizes da palmeira *Oenocarpus bataua* Mart., investigando a influência do efeito de alelos raros na preservação da espécie, o que mostrou que a seleção dependente de frequência negativa que oferece vantagem na permanência de genótipos maternos raros nas progênes gerando a diversidade genética a nível populacional.

Embora exista uma grande quantidade de estudos utilizando estimadores de parentesco por meio de marcadores moleculares, na maioria dos trabalhos não há uma comparação entre os estimadores disponíveis, o que pode levar a estimativas viesadas, uma vez que uma análise prévia comparativa dos estimadores disponíveis corrobora para uma maior robustez na inferência dos resultados gerados (VAN DE CASTEELE et al., 2001; WANG, 2011).

3 MATERIAL E MÉTODOS

3.1 Espécies e marcadores moleculares

Três espécies de sistemas reprodutivos misto, *A. aculeata*, *D. alata* e *H. stigonocarpa*, foram utilizadas para a avaliação dos modelos de predição de parentescos. As espécies são consideradas mistas, mas com predominância de alogamia.

Os dados moleculares utilizados são resultantes de projetos de auxílio a pesquisa aprovados pela Fundação de Amparo à Pesquisa do Estado de São Paulo sob números 2012/18747-2, 2013/23168-4 e 2007/59386-1 para *D. alata*, *A. aculeata* e *H. stigonocarpa* respectivamente. Para *A. aculeata* foram utilizados nove locos microssatélites, 10 locos para *D. alata* e seis locos para *H. stigonocarpa* (Tabela 1).

Tabela 1 - Locos microssatélites utilizados para a avaliação de diferentes modelos de estimativa de parentesco para as espécies.

Espécie	Primers ¹	Progênes	Autores
<i>Acrocomia aculeata</i>	Aacu30, Aacu26, Aacu07, Aacu12, MAC245, MAC115*, MAC201*, MAC240, MAC23	20	Nucci (2007)
<i>Dipteryx alata</i>	DaE20*, DaE41*, DaE67, BM164, DaE46*, DaE34*, DaE63, Dal14, Dod08*, Dal15*	12	Grattapaglia unpublished (2016)
<i>Hymenaea stigonocarpa</i>	Hc17, Hc33, Hc34, Hc35, Hc40, Hc42, Hc48, Hc49	6	Moreno (2009)

* Locos que não estão em Equilíbrio de Hardy-Weinberg.

3.2 Populações simuladas

Para estudar de maneira idealizada o comportamento dos estimadores, foram feitas simulações utilizando o pacote *stringr* (WICKHAM, 2017) do programa R versão 3.4.0 (*R Development Core Team* 2015), seguindo os seguintes passos:

1. Foram criadas populações simuladas de meias-irmãs multialélicas, completamente alógamas e em equilíbrio de Hardy-Weinberg;

- Para cada população foram criados por meio de simulações alelos e locos. Os locos foram codificados pela letra "X" acompanhada por um número (ex: loco: X1; X2) da mesma forma os alelos foram codificados por uma letra acompanhada por um número (Ex: Alelos: A1; B2), conforme a Figura 1:

Figura 1 - Exemplo de arquivo de saída gerado pelo pacote *stringr* para a simulações de populações para o estudo de estimativas de parentescos.

	X1	X2	X3
Alelos ←	A1A1	B1B3	C3C3
	A2A1	B2B4	C2C6

Loco ↗

- Afim de estudar o efeito do polimorfismo e o número de locos, foram criadas sete populações, combinando número de locos e de alelos por locos (Tabela 2);
- Em todos os cenários, fixou-se o número de progênie e de indivíduos por progênie (Tabela 2). Esses valores foram utilizados para se atingir o valor mínimo de número de indivíduos por população proposto por Ritland (2002).
- A 1ª, 2ª, 3ª e 4ª população foram usadas para comparação entre os estimadores de momento e os de verossimilhança, enquanto que para comparação entre os estimadores de momento foram usadas todas as populações.

Tabela 2 - Descrição de população simulada de meios-irmãos para estimativa da coancestria por diferentes estimadores.

Parâmetros	1ª Pop.	2ª Pop.	3ª Pop.	4ª Pop.	5ª Pop.	6ª Pop.	7ª Pop.
Nº de locos	8	8	20	20	15	20	25
Alelos/loco	2	10	2	10	2	20	2
Nº de prog.	15	15	15	15	15	15	15
Ind./prog.	20	20	20	20	20	20	20

f/alelo: frequência de cada alelo; Nº de prog: número de progênie; Ind/prog.: número de indivíduos por progênie.

3.3 Coeficientes de parentesco

A fim de se conhecer o melhor estimador do coeficiente de parentesco, alguns métodos de estimação foram testados, dentro dos dois grupos de estimadores. Utilizando o programa SPAGeDi (HARDY e VEKEMANS, 2002) foi estimado coeficiente médio de coancestria ($\bar{\theta}$) pelos estimadores de momentos de Ritland (1996), Lynch e Ritland (1999), Queller e Goodnight (1989), Loiselle et al. (1995) e Wang (2002). Também foram utilizados estimadores de verossimilhança, sendo estes Milligan (2003), Wang (2007) por meio do programa Coancestry v. 1.0.1.2 (WANG, 2011).

Assumindo que $A_i A_j$ e $A_k A_l$ são os genótipos dos indivíduos, δ_{ij} são variáveis indicadoras, sendo $\delta_{ij}=1$ se os alelos comparados entre os indivíduos i e j forem idênticos ou $\delta_{ij}=0$ se os valores forem diferentes. Entre dois indivíduos diploides há quatro valores de δ_{ij} , visto que existem quatro alelos, um para cada comparação A_k e A_l (δ_{ik}), A_i e A_l (δ_{il}), A_j e A_k (δ_{jk}) e A_j e A_l (δ_{jl}). O

Para os estimadores de momentos tem-se:

- Ritland (1996): $r_{ij} = \frac{[(\delta_{ik} + \delta_{il})/p_i] + [(\delta_{jk} + \delta_{jl})/p_j] - 1}{4(n-1)}$, em que n é o número

de alelos por loco em análise, p_i e p_j são as frequências dos alelos A_i e A_j no loco em análise na população. A estimativa multiloco é obtida pela soma das estimativas uniloco de r_{ij} , ponderadas por $(n-1)$;

- Lynch e Ritland (1999): $r_{ij} = \frac{p_i(\delta_{ik} + \delta_{il}) + p_j(\delta_{jk} + \delta_{jl}) - 4p_i p_j}{(1 + \delta_{ij})(p_i + p_j) - 4p_i p_j}$, em que o

primeiro genótipo, por exemplo $A_i A_j$ é o genótipo de referência (δ_{ij}). O estimador multiloco é obtido ponderando cada loco pelo inverso de $2p_i p_j / [(1 + \delta_{ij})(p_i + p_j) - 4p_i p_j]$;

- Queller e Goodnight (1989): $r_{ij} = \frac{0,5(\delta_{ik} + \delta_{il} + \delta_{jk} + \delta_{jl}) - p_i - p_j}{1 + \delta_{ij} - p_i - p_j}$. Este estimador é

indefinido quando o genótipo de referência é heterozigoto ($\delta_{ij} = 0$), para um loco dialélico (possua apenas dois alelos), visto que o denominador é zero, $1 + \delta_{ij} - p_i - p_j = 1 - 0 - 1 = 0$ (RITLAND e TRAVIS, 1994);

- Loiselle et al. (1995): $\hat{\rho}_{ij} = \frac{\sum_j (p_i - \bar{p})(p_j - \bar{p})}{k\bar{p}(1 - \bar{p})} + \frac{2}{(8k + 1)^{0,5} - 1}$ ($i < j$), onde

ρ_{ij} mede a correlação de frequência de alelos homólogos, p_i e p_j , em um mesmo loco em pares de indivíduos i e j . $k = n(n-1)/2$ é o número total de possíveis conexões entre os pares dos n indivíduos.

- Wang (2002): $r_{ij} = 1 - \frac{c_1 - c_1 P_1 - c_2 P_2 - c_3 P_3}{c_1 - c_1 e_1 - c_2 e_2 - c_3 e_3}$, onde, o genótipo dos indivíduos é

resumido em um conjunto de quatro indicadores, P_i e duas classes de constantes c_i e e_j .

Para os estimadores de verossimilhança, tem-se:

- Milligan (2003): $\theta_{xy} = \Delta_1 + \frac{1}{2}(\Delta_3 + \Delta_5 + \Delta_7) + \frac{1}{4}\Delta_8$, sendo $\theta_{xy} = 2r_{xy}$;
- Wang (2007): é um método de verossimilhança que utiliza os genótipos de uma tríade de indivíduos para estimar parentesco (r_{xy}), usando um terceiro indivíduo (Z) como referência na estimativa de r entre dois indivíduos (X e Y), reduzindo assim a chance de genes idênticos por estado (IIS) serem erroneamente inferidos como idênticos por descendentes (IBD), tendo assim:

$$r_{xy} = 2\Delta_1 + \Delta_3 + \Delta_5 + \Delta_7 + \frac{1}{2}\Delta_8;$$

Os estimadores de verossimilhança estimam a relação por meio da probabilidade que dois indivíduos (X e Y), sejam idênticos por descendência (IBD) considerando os nove estados de identidade condensados proposto por Jacquard (1972), sendo que cada configuração de IBD possui uma probabilidade correspondente (HARRIS, 1964 ; LYNCH e WALSH, 1998; WEIR, 1996) (Tabela 3).

Tabela 3 - Estados de *IBD* e suas probabilidades considerando quatro genes (Adaptado de Wang, 2007).

Estado da IBD	Genes IBD	Probabilidade
S ₁	(abcd)	Δ_1
S ₂	(ab,cd)	Δ_2
S ₃	(abc), (abd)	Δ_3
S ₄	(ab)	Δ_4
S ₅	(acd), (bcd)	Δ_5
S ₆	(cd)	Δ_6
S ₇	(ac,bd), (ad,bc)	Δ_7
S ₈	(ac), (ad), (bc), (bd)	Δ_8
S ₉	Nenhum	Δ_9

a e **b**: são genes homólogos no indivíduo X; **c** e **d** são genes homólogos no indivíduo X; S: configurações de IBD Δ : probabilidade das S_i.

3.4 Análise comparativa dos estimadores

A coancestria média ($\bar{\theta}$) foi estimada para cada progênie e para a população por cada um dos estimadores. Dos estimadores escolhidos, apenas Lynch e Ritland (1999) e Wang (2002) assumem equilíbrio de Hardy-Weinberg. Desta maneira, para *A. aculeata* e *D. alata*, na avaliação destes dois estimadores, utilizaram-se apenas os locos em equilíbrio, enquanto que para os demais estimadores todos os locos foram usados (Tabela 4). Para *Hymenaea stigonocarpa* utilizou-se a mesma quantidade de locos para todos os estimadores.

Tabela 4 - Relação do número de locos utilizados para estimativa da coancestria de *Acrocomia aculeata* e *Dipteryx alata*.

Estimadores	<i>Acrocomia Aculeata</i> (NL)	<i>Dipteryx alata</i> (NL)
Ritland (1996)	9	10
Lynch e Ritland (1999)	4	7
Queller e Goodnight (1989)	9	10
Loiselle et al. (1995)	9	10
Wang (2002)	4	7

NL: número de locos.

Para investigar o comportamento dos estimadores dentro das espécies, foi utilizada a Correlação de Pearson entre os estimadores dentro da população de cada espécie, afim de identificar grau de associação entre os parentescos estimados ($\hat{\theta}$), além de histogramas para observar a distribuição das estimativas.

Para estimar o erro dos estimadores das coancestria em cada uma das populações (*A. aculeata*, *D. alata*, *H. stigonocarpa* e simulações), foram utilizadas duas estatísticas. A primeira, *RMSE* (*Root Mean Square Error*), a qual mensura a diferença entre os valores do modelo de previsão e os valores reais observados ou esperados, agregando a magnitude dos erros de previsão em uma única medida de poder preditivo (no caso, $\hat{\theta}= 0,125$). A segunda estatística é o *MBE* (*Mean Bias Error*), também conhecido como erro médio, que mostra o desvio médio (ou tendência) do modelo em relação à uma variável. Ela fornece informações sobre a performance do modelo, erro sistemático, podendo ser positivo (quando superestima) ou negativo (quando subestima). Quanto mais próximo de zero o valor de *MBE*, mais precisas são as estimativas. Estas estatísticas são obtidas pelas expressões:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (\hat{y}_j - y_j)^2}; MBE = \frac{1}{n} \sum_{j=1}^n (\hat{y}_j - y_j);$$

Em que y_i é o valor real ou esperado do parâmetro avaliado; \hat{y}_j é o valor estimado pelo modelo e n é o número de comparações total.

Tanto o *RMSE* como o *MBE*, podem ser obtidos em porcentagem, multiplicando-se o valor encontrado por 100.

4 RESULTADOS E DISCUSSÃO

4.1 Análise dos estimadores para as espécies

As coancestrias apresentaram diferentes variações para cada estimador e em cada espécie (Figura 2, 3, 4 e 5). Isso pode ser explicado, pela forma que cada estimador responde à quantidade de informação genética disponível, por exemplo, número de locos e alelos segregantes por loco, que influencia fortemente o erro padrão da estimativa (MILLIGAN,2003).

Figura 2 - Distribuição das coancestrias estimadas pelos métodos de momentos para *Acrocomia aculeata*.

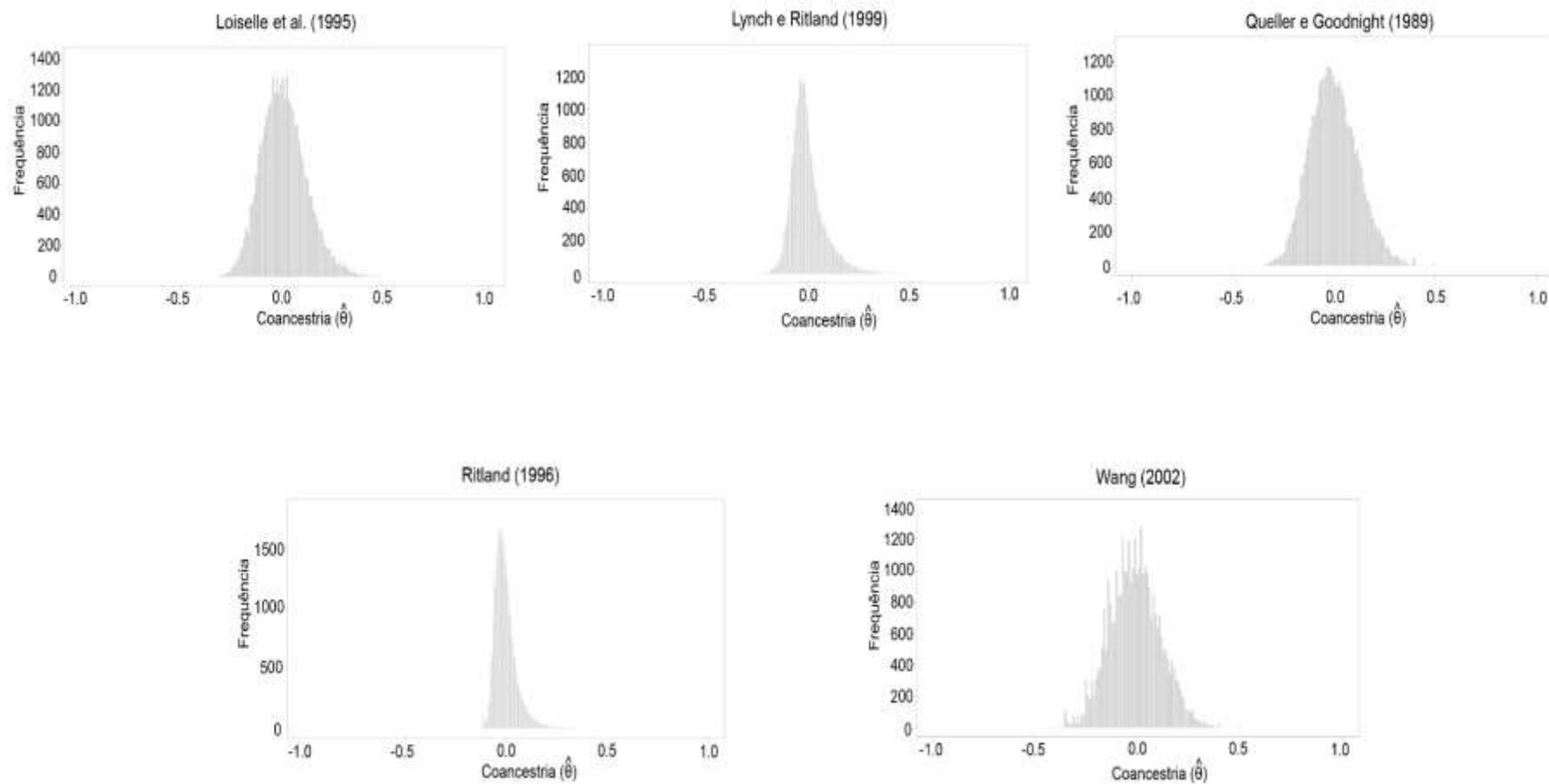


Figura 3 - Distribuição das coancestrias estimadas pelos métodos de momentos para *Dipterix alata*.

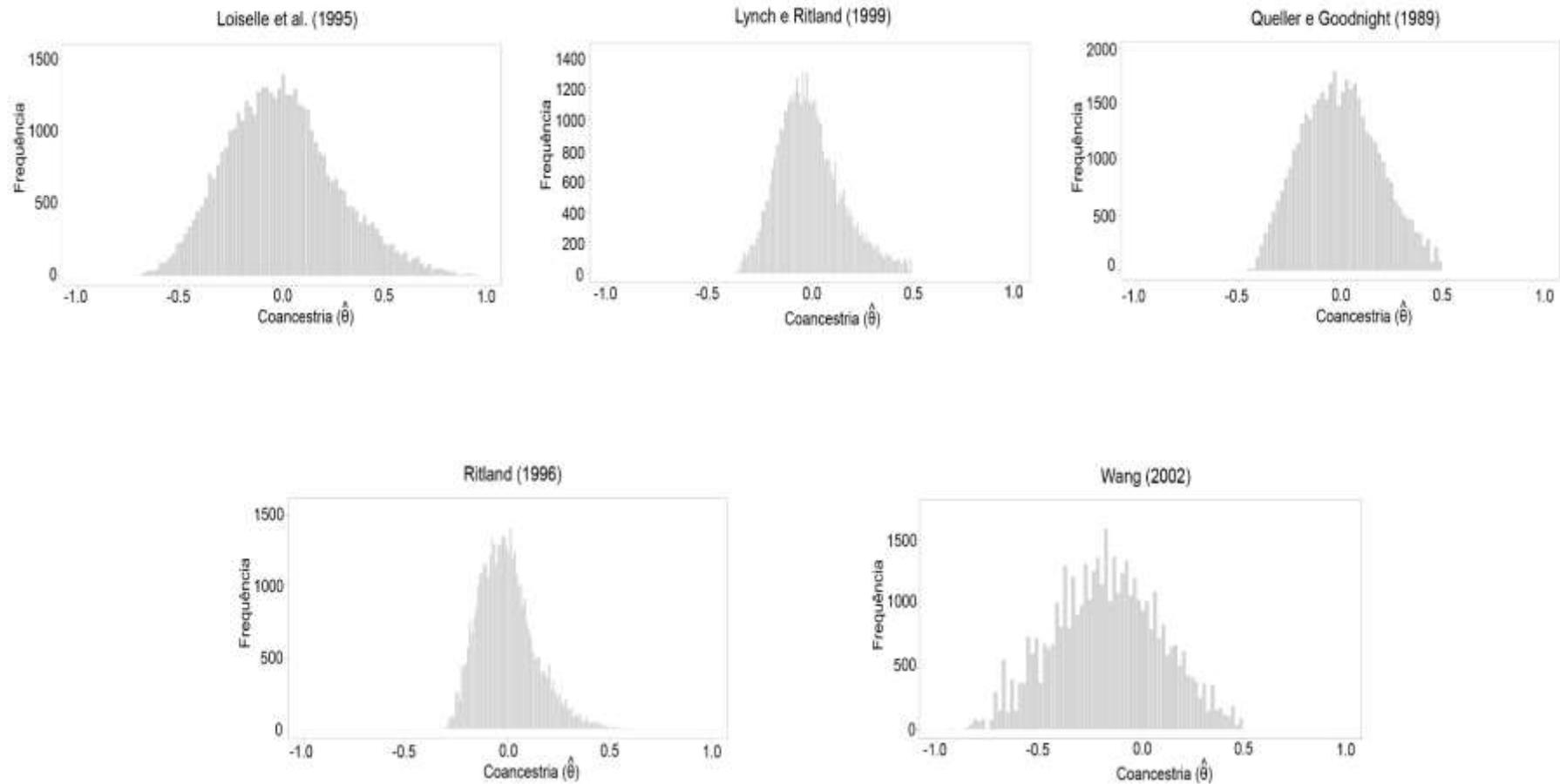


Figura 4 - Distribuição das coancestrias estimadas pelos métodos de momentos para *Hymenaea stagnocarpa*.

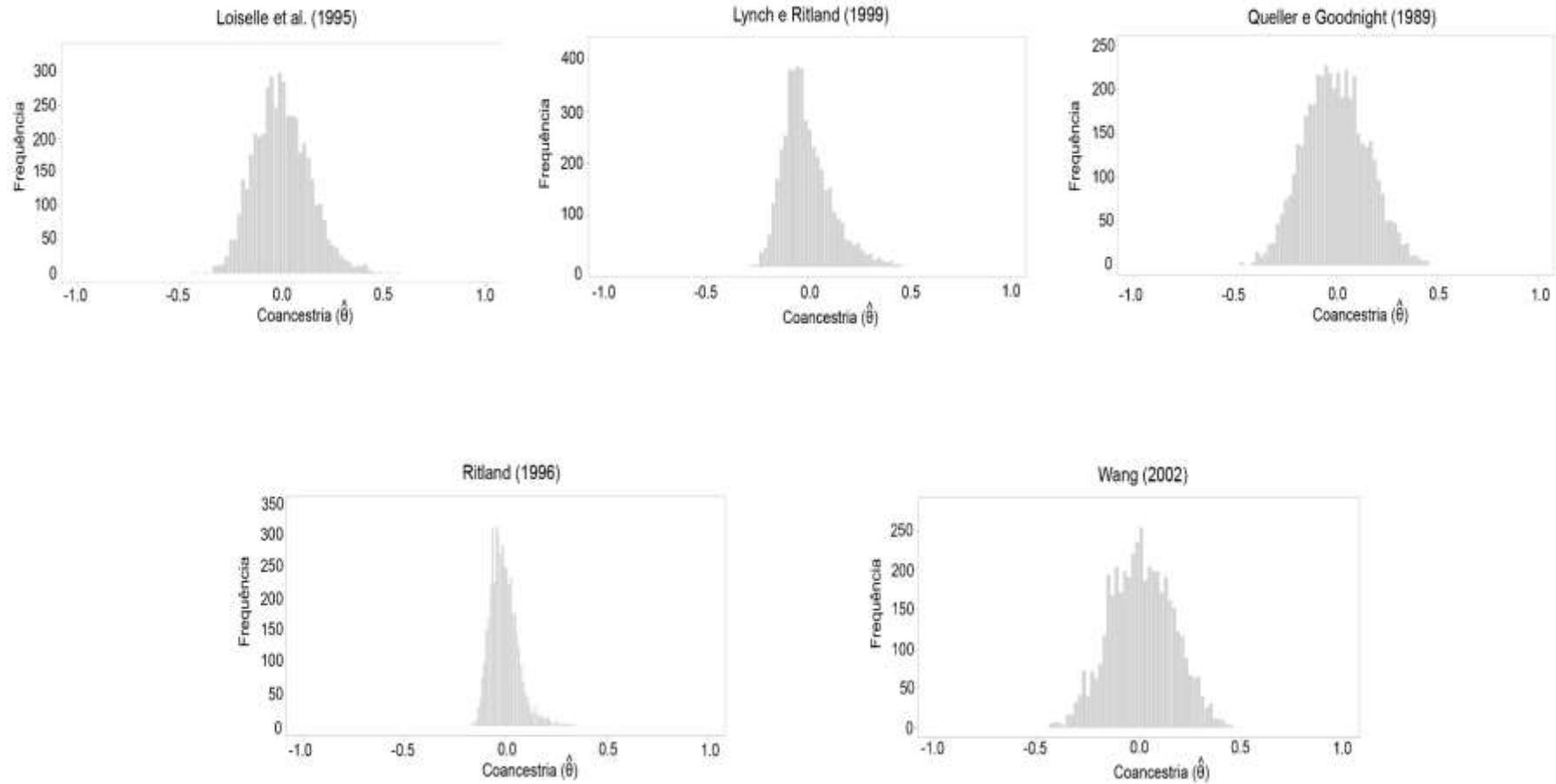
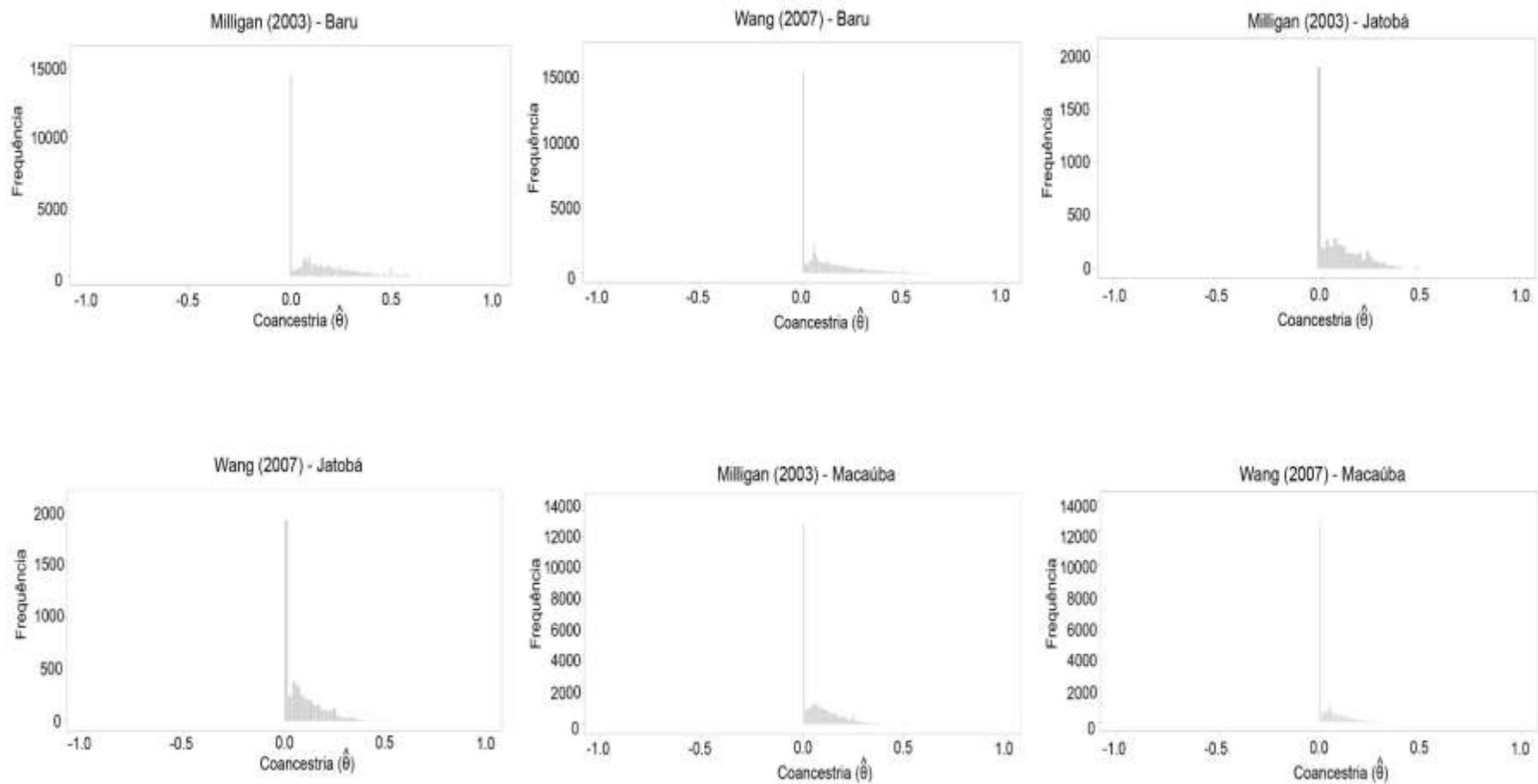


Figura 5 - Distribuição das coancestrias estimadas pelo método da verossimilhança para *Dipteryx alata*, *Hymenaea stigonocarpa* e *Acrocomia aculeata* respectivamente.



As coancestrias ($\hat{\theta}_{xy}$) para as espécies estudadas, estão fora da faixa biologicamente significativa (0 a 0,5) (MILLIGAN, 2003). Em sua maioria, valores negativos são reflexo dos erros de amostragem inerentes aos métodos analisados (Tabela 5), como baixo grau de polimorfismos dos marcadores, erro de genotipagem, pressuposição não atendidas e erro de amostragem (CSILLÉRY et al., 2006; RITLAND, 1996; LYNCH e RITLAND, 1999; WANG, 2002).

Tabela 5 - Média e variância dos estimadores de coancestria para as espécies *Acrocomia aculeata*, *Hymenaea stigonocarpa* e *Dipteryx alata*.

<i>Acrocomia aculeata</i>							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
Média	0,0004	-0,0020	-0,0015	-0,0017	-0,0152	0,1615	0,1285
Variância	0,0132	0,0071	0,0142	0,0038	0,0156	0,0374	0,0269
<i>Hymenaea stigonocarpa</i>							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
Média	-0,0003	-0,0050	-0,0083	-0,0048	0,0123	0,1927	0,1604
Variância	0,0187	0,0145	0,0980	0,0082	0,0242	0,0516	0,0391
<i>Dipteryx alata</i>							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
Média	0,0023	-0,0032	-0,0030	-0,0001	-0,3285	0,2976	0,2524
Variância	0,0748	0,1007	0,1466	0,0266	0,2741	0,1146	0,1016

L: Loiselle et al. (1995); L e R: Lynch e Ritland (1999); Q e G: Queller e Goodnight (1989); R: Ritland (1996); W: Wang (2003); DiaML: Milligan (2003); TrioML: Wang (2007).

As médias das $\hat{\theta}_{xy}$ inferidas pelos estimadores de momentos foram menores ou próximas a zero para todas as espécies, enquanto que para os estimadores de verossimilhança, a média de $\hat{\theta}_{xy}$ foi maior que zero (Tabela 4). Este resultado pode ser explicado pelo fato dos estimadores de verossimilhança serem “truncados”, de modo que os valores negativos gerados são zerados, produzindo uma média maior que os estimadores de momentos que levam em consideração os valores negativos (MILLIGAN, 2003), gerando desta forma, um viés positivo que oferece poucos benefícios as estimativas (Figura 4) (THOMAS, 2005).

Ao se “truncar” os resultados da estimativa (restringindo r_{xy} ao intervalo de 0 a 1 e θ_{xy} ao intervalo de 0 a 0,5), admite-se que nenhum indivíduo diploide é endogâmico, ou seja, os estimadores de verossimilhança não consideram a presença de indivíduos endogâmicos na amostra (BLOUIN, 2003; KONOVALOV e HEG, 2008). Isso tem um significado biológico, pois admite-se que dois indivíduos que apresentam $\hat{\theta}_{xy}$ negativa, compartilham menos alelos do que o esperado com base nas frequências alélicas de Hardy-Weinberg, ou seja, possuem menos alelos idênticos por estado (QUELLER e GOODNIGHT, 1989; KONOVALOV e HEG, 2008).

Bessegua et al. (2012), estimaram coancestria utilizando Ritland (1996), Queller e Goodnight (1989) e Loiselle et al. (1995) em progênies de polinização aberta de *Prosopis alba* Griseb. e os valores médios obtidos foram de 0,05, 0,168 e 0,165 respectivamente. Levando em consideração que a espécie é alógama, apenas o estimador de Ritland (1996) apresentou uma média abaixo do esperado ($\hat{\theta}_{xy}=0,125$). Os valores de coancestria influenciaram no tamanho efetivo da população para coleta de sementes, indicando uma distância mínima entre as matrizes para a coleta.

Caso os autores, utilizassem apenas o estimador de Ritland (1996), sem uma comparação com outros estimadores, a distância mínima das matrizes para a coleta diminuiria em virtude do grau de parentesco estimado, causando cruzamentos com indivíduos aparentados e posteriormente depressão endogâmica.

Cloutier et al. (2006) também analisando a $\hat{\theta}_{xy}$ de *Carapa guianensis* Aubl., uma espécie alógama, por meio do estimador de Loiselle et al. (1995) em favor da distância física das árvores, verificaram que o parentesco se manteve próximo a 0,125 quando as árvores estavam a uma distância inferior a 20 m e com o aumento da distância houve a diminuição do parentesco entre os indivíduos avaliados. Estes resultados demonstram um bom ajuste do estimador de Loiselle et al. (1995), além do efeito da distância no parentesco da população, podendo ser utilizado como parâmetro de seleção de indivíduos num plano de manejo.

Da Silva et al. (2015) trabalhando com uma espécie alógama, estimaram o parentesco (r_{xy}) em *Araucaria angustifolia* (Bertoloni) Otto Kuntze, por meio de quatro estimadores de parentesco separando os indivíduos por classe de distância. Os

autores observaram uma grande variação nas estimativas, indicando a inadequação dos dados aos estimadores utilizados.

Pometti et al. (2013) obteve valores de $\hat{\theta}_{xy}$ média entre progênies de polinização aberta dentro de famílias de *Acacia visco* Lorentz ex Griseb, considerada uma espécie alógama, obteve valores variando de 0,082 a 0,105 para o estimador de Loiselle et al. (1995) e 0,125 a 0,136 para o estimador de Ritland (1996), tendo valores mais próximos do esperado apenas para Ritland (1996), implicando numa variação no número de matrizes para a coleta de sementes dependendo do estimador utilizado.

Bressan et al. (2013) avaliando uma população de *Jatropha curcas* L., espécie considerada mista, amostrando indivíduos de polinização aberta, observaram que as $\hat{\theta}_{xy}$ médias foram de 0,292 a 0,384, superando o valor esperado ($\hat{\theta}_{xy}= 0,125$), sendo estes valores justificados pela ocorrência de cruzamentos entre indivíduos relacionados. Estes resultados demonstram a sensibilidade do estimador ao grau de parentesco dos indivíduos presentes na amostra, sendo que para espécies mistas, o estimador pode gerar um viés positivo, quando há uma maior frequência de indivíduos aparentados e gerar um viés negativo, quando há uma maior frequência de indivíduos sem parentesco.

Estas pesquisas demonstram que a performance dos estimadores é influenciada pelo grau de parentesco dos indivíduos que compõe a amostra de uma população, sendo que quanto maior a heterogeneidade da amostra em relação ao parentesco maior será o viés.

Os resultados das estatísticas de erro aplicadas aos modelos, variaram entre as espécies analisadas nesta pesquisa (Tabela 6). Os estimadores de verossimilhança tiveram os menores valores de *MBE* e *RMSE* quando comparado aos estimadores de momento.

Comparando os estimadores de momento, tem-se que quanto ao *MBE*, nos estimadores para *A. aculeata*, os valores foram negativos, o que implica que as $\hat{\theta}_{xy}$ estimadas foram subestimadas, ou seja, nenhuma das progênies analisadas poderiam ser consideradas meias-irmãs (0,125). Os estimadores Loiselle (1995), Queller e

Goodnight (1989) e Lynch e Ritland (1995), apresentaram os menores valores de *MBE*. Para *RMSE*, os estimadores Ritland (1996), Lynch e Ritland (1995) e Queller e Goodnight (1989), demonstraram menores valores, e melhores performances em relação aos demais.

Tabela 6 - Erros dos quadrados médios (*MBE* em %), raiz dos erros quadrados médios (*RMSE* em %) para coancestrias dos diferentes estimadores para *Acrocomia aculeata*, *Hymenaea stigonocarpa* e *Dipteryx alata*.

<i>Acrocomia aculeata</i>							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-12,46	-12,70	-12,65	-12,67	-14,02	-3,47	-4,76
RMSE	16,95	15,25	17,39	14,09	18,78	9,42	9,03
<i>Hymenaea stigonocarpa</i>							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-12,53	-12,75	-13,07	-12,71	-11,05	-2,86	-4,48
RMSE	18,56	17,55	33,69	15,67	19,03	11,72	10,85
<i>Dipteryx alata</i>							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-12,38	-12,82	-12,65	4,23	-28,93	2,38	0,12
RMSE	18,45	20,30	22,94	20,56	39,01	17,09	15,94

L: Loiselle et al. (1995); L e R: Lynch e Ritland (1999); Q e G: Queller e Goodnight (1989); R: Ritland (1996); W: Wang (2002); DiaML: Milligan (2003); TrioML: Wang (2007).

Para *H. stigonocarpa*, o estimador Queller e Goodnight (1989) apresentou o menor valor de *MBE*, seguido de Lynch e Ritland (1999) e Ritland (1996), e para *RMSE* os estimadores Ritland (1996), Lynch e Ritland (1999) e Loiselle et al. (1995) obtiveram os menores valores (Tabela 6). Em *D. alata*, o estimador Wang (2002) apresentou o menor valor de *MBE*, seguido de Lynch e Ritland (1999) e Queller e Goodnight (1989), e para *RMSE* os estimadores Loiselle et al. (1995), Lynch e Ritland (1999) e Ritland (1996) obtiveram os menores valores (Tabela 6). Embora que para *D. alata* os estimadores não subestimaram as coancestrias o erro foi superior aos resultados para *A. aculeata*, tanto considerando *MBE* como *RMSE*, exceto o *MBE* de Wang (2002).

Estes erros podem ser explicados pelo sistema de reprodução das espécies, pois o sistema afeta diretamente a estrutura genética da população por influenciar a forma como a população será composta (HOLSINGER, 2000), no qual para espécies mistas, a população é uma mistura de indivíduos com diferentes graus de endogamia e parentesco (VENCOVSKY et al., 2001; TAMBARUSSI et al., 2017b).

A diferença nos resultados das correlações entre os estimadores para cada espécie pode ser explicada pela baixa amostragem de indivíduos e os métodos estatísticos utilizados para cada estimador (Figuras 6, 7, 8). Para estimar o parentesco, cada estimador possui fatores específicos na sua composição, como a atribuição de pesos, escolha de um indivíduo de referência, padronização de alelos, entre outros, gerando a variação nos parentescos estimados (THOMAS, 2005; VAN DE CASTEELE et al., 2001).

Figura 6 - Correlações entre os estimadores de coancestria ($\hat{\theta}$) para *Acrocomia aculeata*.

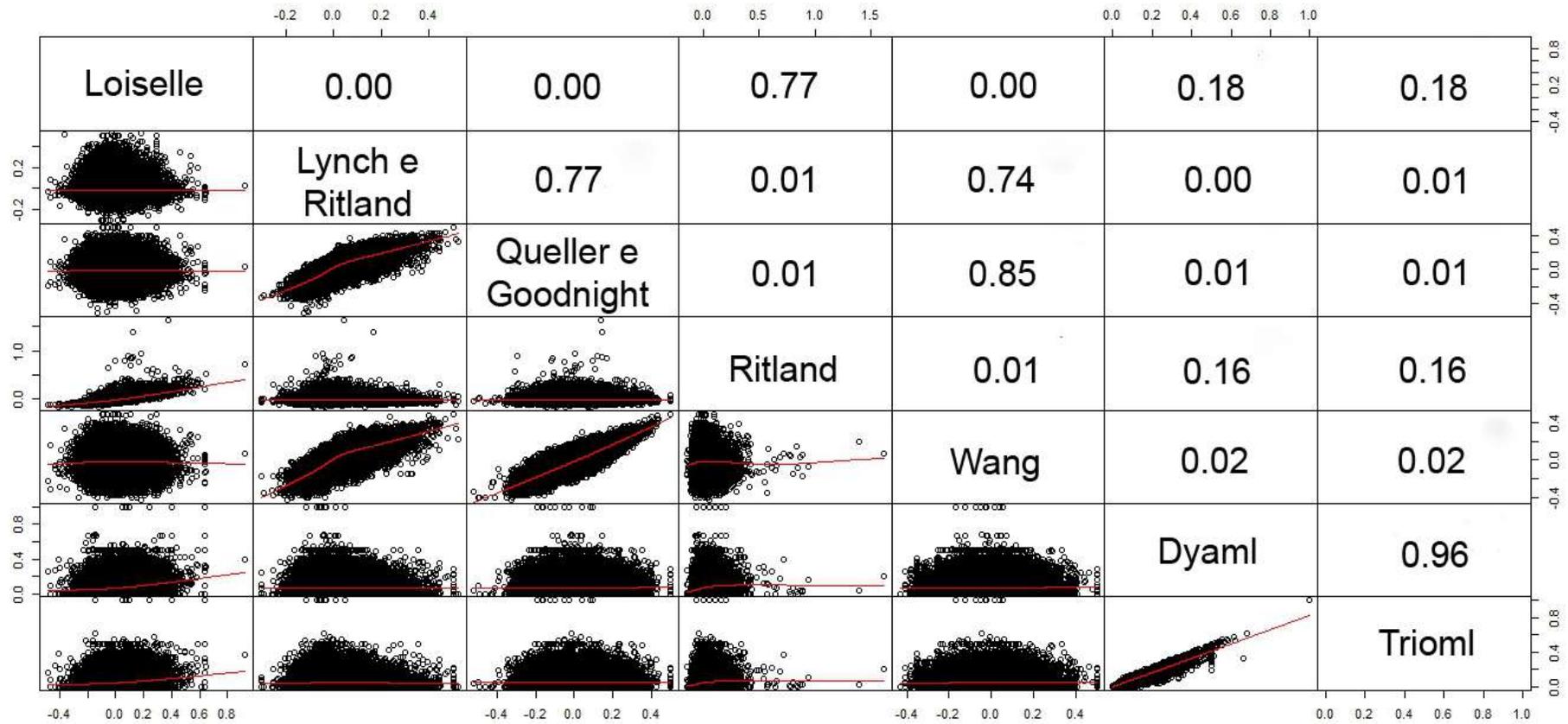


Figura 7 - Correlações entre os estimadores de coancestria ($\hat{\theta}$) em *Hymenaea stigonocarpa*.

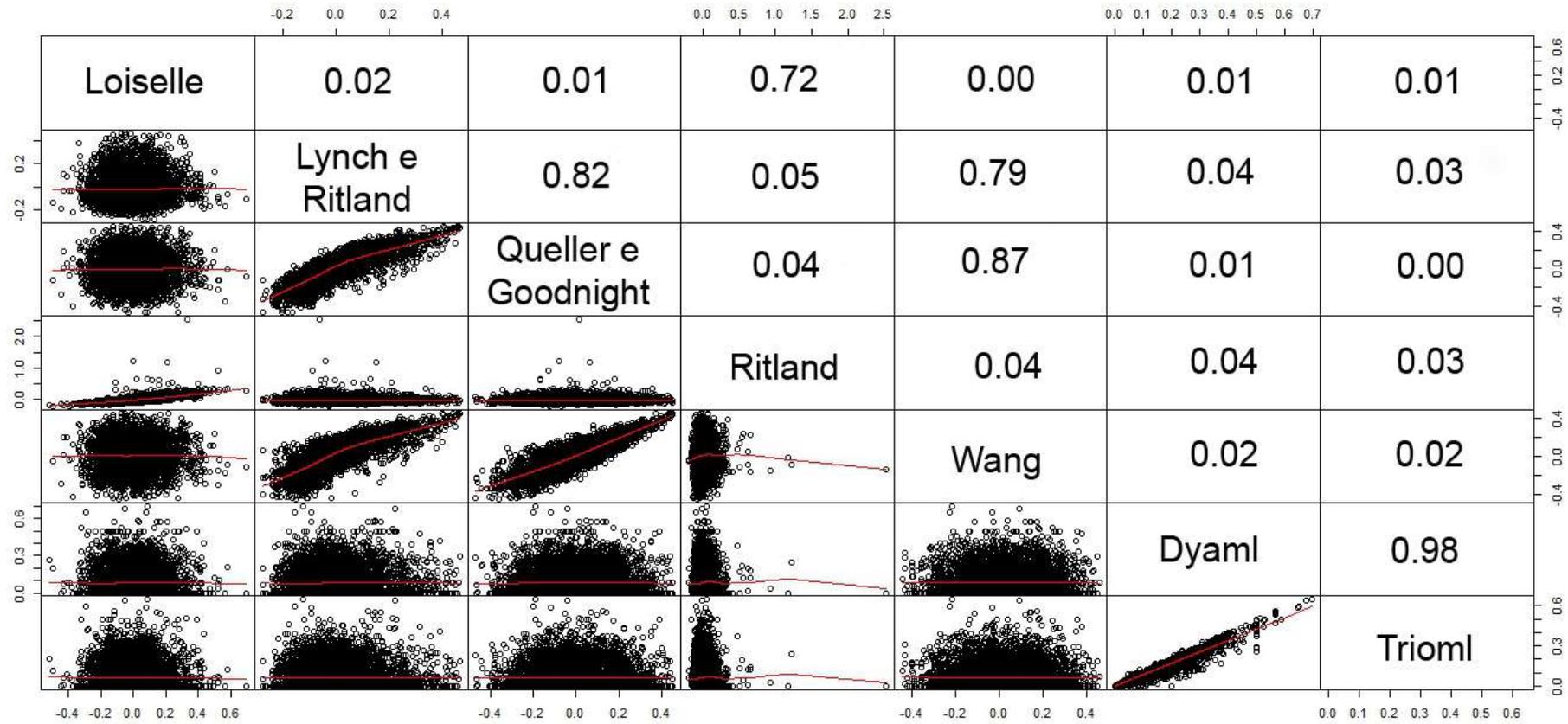
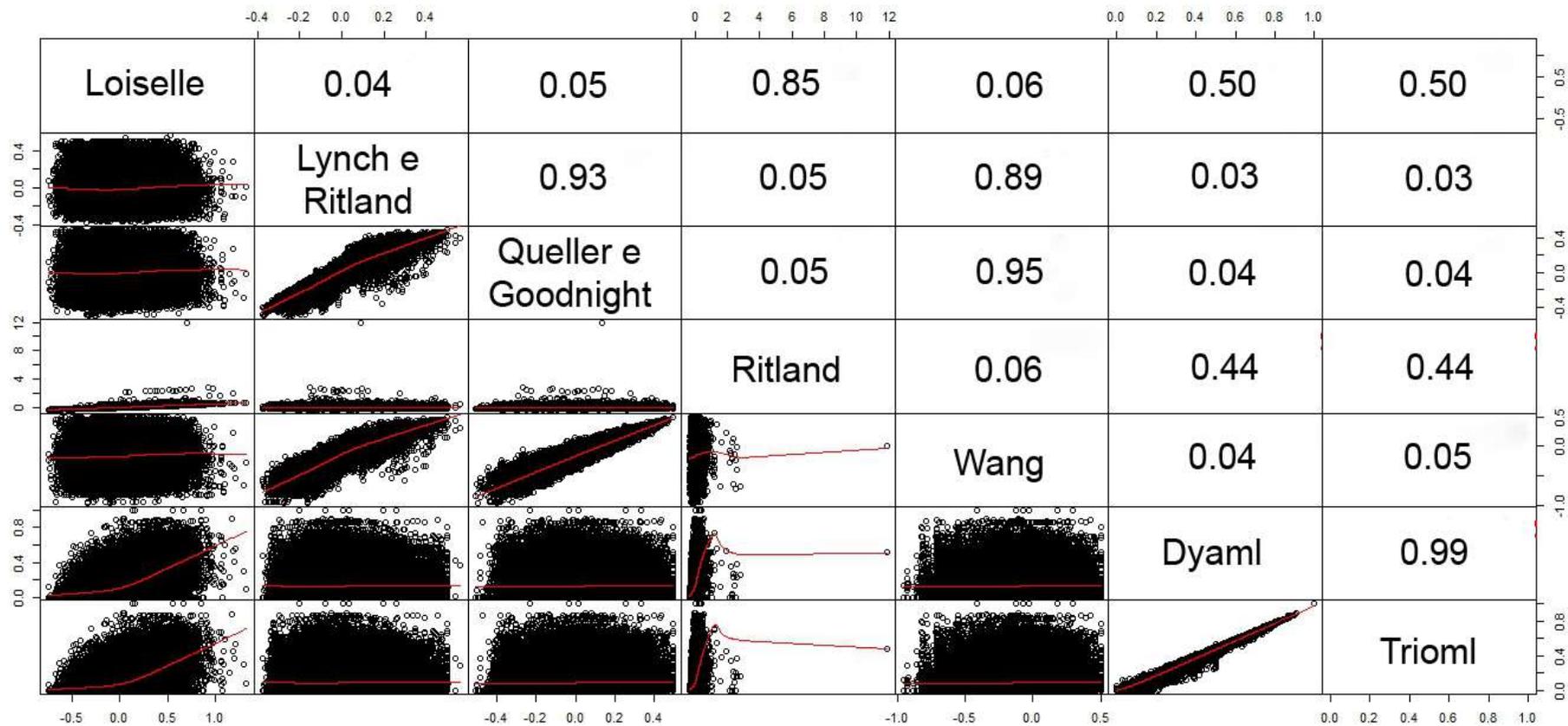


Figura 8 - Correlações entre os estimadores de coancestria ($\hat{\theta}$) em *Dipteryx alata*.



Os valores das correlações para as coancestrias dos diferentes estimadores variaram dentro das espécies, com valores de $r= 0,0$ a $0,96$, $0,0$ a $0,98$ e $0,0$ a $0,99$, para *A. aculeata*, *H. stigonocarpa* e *D. alata*, respectivamente. Desta forma, verifica-se que há heterogeneidade conforme os métodos de estimação. A correlação entre os estimadores Ritland (1996) vs Loiselle (1995) foram de $0,72$ a $0,85$ entre as três espécies (Figuras 6, 7, 8), o que pode ser explicada pelo fato de ambos estimarem parentesco por correlação, cuja diferença está apenas em como locos e alelos são ponderados por cada estimador (HUANG et.al., 2014).

O estimador de Lynch e Ritland (1999) vs Queller e Goodnight (1989) apresentaram valores de correlação de $0,72$ a $0,93$ para espécies estudadas, uma vez que Lynch e Ritland (1999) e Queller e Goodnight (1989) estimam parentesco de forma semelhante (regressão) diferindo apenas o peso dado aos componentes de cada estimador. O estimador de Wang (2002) apresentou correlação tanto com Lynch e Ritland (1999) como com Queller e Goodnight (1989) e isso se dá pelo fato de alguns componentes serem estimados de maneira semelhante (Figura 6,7,8) (BINK et al., 2008).

A correlação entre o estimador de Milligan (2003) vs Wang (2007) foi alta para todas espécies pelo fato que ambos estimam o parentesco por verossimilhança diferindo apenas na quantidade de parâmetros de cada estimador. As correlações foram nulas ou baixas para estimadores de métodos diferentes, como para modelos de correlação (Ritland e Loiselle) em comparação com modelos de regressão (Lynch e Ritland, Queller e Goodnight) uma vez que são assimétricos, fazendo com que cada estimador resulte em coancestrias diferentes (WANG, 2014).

Nenhum dos estimadores representaram de forma satisfatória as espécies estudadas, considerando a quantidade de marcadores utilizados e o grau de parentesco estabelecido (meios-irmãos). Por esta razão, o desenvolvimento de metodologias para verificar a eficácia destes estimadores de relacionamentos são necessárias para aumentar a confiabilidade de decisões levando em consideração a sua aplicação.

4. 2 Análise dos estimadores nas populações simuladas

Para as simulações, os valores de *MBE* e *RMSE* foram menores que os valores observados nas populações reais (Tabela 7).

Tabela 7 - Erros dos quadrados médios (*MBE* em %), raiz dos erros quadrados médios (*RMSE* em %) para as estimativas de coancestrias dos diferentes modelos para as populações simuladas.

1º população simulada (8 locos e 2 alelos)							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-1,6	0,0	-0,5	-1,8	-0,3	-0,5	0,3
RMSE	18,9	18,7	17,2	20,2	17,0	12,6	13,4
2º população simulada (8 locos e 10 alelos)							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-1,4	-1,6	-0,7	-2,7	-0,6	-8,3	-8,4
RMSE	8,8	6,6	6,9	9,3	6,9	9,9	10,0
3º população simulada (20 locos e 2 alelos)							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-2,1	-1,1	-1,1	-2,3	-1,1	-4,8	-5,1
RMSE	13,8	10,8	10,8	13,2	11,4	10,6	10,5
4º população simulada (20 locos e 10 alelos)							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-2,1	-1,8	-2,0	-3,0	-0,9	-10,5	10,6
RMSE	6,4	4,6	8,9	6,6	4,5	11,0	11,1
5º população simulada (15 locos e 2 alelos)							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-3,12	-1,79	-1,77	-3,30	-0,83	–	–
RMSE	14,97	12,48	12,58	15,01	13,05	–	–
6º população simulada (20 locos e 20 alelos)							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-10,32	-2,52	-6,52	-6,85	-3,50	–	–
RMSE	14,48	8,40	9,98	12,87	7,68	–	–
7º população simulada (25 locos e 2 alelos)							
Estatística	L	L e R	Q e G	R	W	DiaML	TrioML
MBE	-1,83	-1,27	-1,26	-2,00	-1,95	–	–
RMSE	12,03	9,79	9,87	12,06	10,46	–	–

L: Loiselle et al. (1995); L e R: Lynch e Ritland (1999); Q e G: Queller e Goodnight (1989); R: Ritland (1996); W: Wang (2003); DiaML: Milligan (2003); TrioML: Wang (2007).

As populações simuladas com menores valores de erro (*MBE* e *RMSE*) foram a segunda (8 locos,10 alelos) a quarta população (20 locos,10 alelos) e a sexta

população (20 locos, 20 alelos), ou seja, aquelas cujo volume de informações para os indivíduos foi maior (Tabela 7). Os resultados indicam uma resposta positiva dos estimadores de momentos em virtude do aumento do número de locos e de alelos, apresentando melhores valores para os estimadores de Lynch e Ritland (1999), Queller e Goodnight (1989) e Wang (2002), enquanto que os estimadores de verossimilhança não apresentam a mesma resposta para essas variáveis e geram um erro similar entre estes estimadores independente da população avaliada. Desta forma, pode-se dizer que para estimar com precisão o parentesco em populações naturais, é necessário o uso de uma maior quantidade de marcadores moleculares e que esses marcadores devem possuir um maior polimorfismo.

Esta resposta positiva dos estimadores de momentos pode ser observada ao se comparar as populações 1 e 2 simuladas, em que o aumento do número de alelos provocou uma redução de pelo menos 50% no *RMSE* de cada estimador (Tabela 7). Todavia, ao se comparar as populações 1 e 3, verificou-se que com o aumento no número de locos, a redução do erro ficou na faixa de 27% (Tabela 7). Este fato demonstra que os estimadores que utilizam métodos de momentos tem uma melhor resposta ao aumento do número de alelos.

Todavia comparando-se a população 4 e 6 que diferem apenas no número de alelos, nota-se que os estimadores não responderam ao aumento no número de alelos, onde a população 4 teve os menores valores de erro na simulação. Isso demonstra a variabilidade no desempenho dos estimadores quanto ao polimorfismo dos marcadores, tornando necessário, uma análise prévia dos estimadores a serem aplicados num estudo.

Van De Casteele et al. (2001) em seu estudo com duas espécies de aves e uma espécie de mamífero utilizaram 10 locos, no qual o desempenho dos estimadores foi afetado pelo uso de diferentes conjuntos de locos para cada espécie estudada e com o aumento no número de locos o desempenho dos estimadores avaliados pode ser melhorado. Pequenos aumentos no número de marcadores produzem grandes reduções no viés e no *RMSE* (ANDERSON e WEIR, 2007). No entanto, nem sempre as populações avaliadas apresentam essa quantidade de informação disponível.

O resultado das simulações mostram que o nível de polimorfismo dos marcadores também afeta os estimadores. Wang (2014), notou que marcadores com

polimorfismo mais elevado tendem a apresentar valores de relacionamento maiores, especialmente para os estimadores Ritland (1996) e Lynch e Ritland (1999). Adjacente a isso, Csilléry et al. (2006) obtiveram respostas positivas ao aumento do número de alelos nas estimativas de parentesco em seu estudo com vertebrados. Os parentescos observados no presente trabalho apresentaram valores similares aos destes trabalhos, com a ocorrência de maiores erros quando se considerou um menor número de marcadores e alelos.

4.3 Implicações do uso de estimadores de coancestria na conservação e melhoramento de espécies

As estimativas de parentesco, utilizando marcadores moleculares, são comumente utilizadas para inferir sobre o grau de parentesco de indivíduos amostrados de uma população e diagnosticar a estrutura genética da população pela incorporação de mais parâmetros (BONATELLI, 2015).

Todavia, os resultados desta pesquisa demonstram que os estimadores produzem viés em relação ao sistema de reprodução da espécie, quantidade de locos e polimorfismo dos marcadores, podendo gerar erros na tomada de decisão em programas de conservação ou melhoramento. Por exemplo, quando o parentesco é estimado para determinar o número de matrizes de uma espécie a serem coletadas sementes para a restauração de uma área ou para implantação de um pomar de semente, com o objetivo de controlar a endogamia, evitando cruzamentos com indivíduos aparentados. Se o modelo (estimador de parentesco) não se ajustar a espécie, o número de indivíduos a serem escolhidos pode ser menor ou maior do que o necessário, podendo gerar endogamia.

Outra situação seria para um plano de manejo florestal, no qual a extração dos indivíduos com potencial madeireiro ocorresse com base na preservação da variabilidade genética dos indivíduos, assim pode se utilizar um estimador de parentesco para determinar se os indivíduos escolhidos para extração possuem parentesco com as árvores que permanecerão na área. Se o estimador não se ajustar a espécie, as árvores que permanecerão podem ser aparentadas (meio-irmãos, irmãos completos), levando a cruzamentos biparentais e possível depressão por endogamia.

Outro cenário com a escolha de um modelo estatístico para estimação dos parâmetros e o ganho genético em um teste de progênie de polinização aberta, onde um estimador de parentesco pode estimar o grau de relação dos indivíduos corroborando para maior precisão do real ganho dos indivíduos analisados.

Uma medida que pode dar maior segurança na utilização destes estimadores é a aplicação de mais de um estimador aliado a informações de taxa de cruzamento, sistema de reprodução, estrutura genética da população, de modo que proporcione maior robustez para inferência de determinado grau de parentesco para os indivíduos de uma população.

5 CONCLUSÃO

O sistema de reprodução afeta diretamente o desempenho dos estimadores, uma vez que este altera a composição e o tipo de relações de parentesco das populações, sendo mais expressivo em sistemas reprodutivos mistos pela diversidade do grau de parentesco que pode ocorrer dentro de uma população;

Os estimadores respondem de maneira diferente ao aumento da informação genética, uma vez que a população com melhores resultados não foi a população com maior nº de locos e nº de alelos/loco, evidenciando uma análise prévia para a escolha de um estimador num estudo com espécies;

O desempenho dos estimadores nas populações naturais das espécies estudadas não foi o ideal, prejudicando a escolha de um desses estimadores para trabalhos posteriores com as espécies avaliadas. Apenas a aplicação dos estimadores para inferir o parentesco esperado não gera estimativas robustas o suficiente para categorizar o nível de relação dos indivíduos analisados, sendo necessário o uso integrado com demais informações genéticas e ecológicas das espécies estudadas.

REFERÊNCIAS

- ALLARD, R. W. Principles of plant breeding. John Wiley & Sons, 1999. 254p.
- ALMEIDA, S. D. Cerrado: aproveitamento alimentar. **Planaltina: Embrapa-CPAC**, 1998.
- ANDERSON, A. D.; WEIR, B. S. A maximum-likelihood method for the estimation of pairwise relatedness in structured populations. **Genetics**, v. 176, n. 1, p. 421-440, 2007.
- ANDRADE, M. H.; VIEIRA, A. S.; AGUIAR, H. F.; CHAVES, J. F.; NEVES, R. M. P. S.; MIRANDA, T. L.; SALUM, A. Óleo do fruto da palmeira macaúba parte I: uma aplicação potencial para indústrias de alimentos, fármacos e cosméticos. **Anais do II ENBTEQ-Encontro Brasileiro sobre Tecnologia na Indústria Química**, p. 17-19, 2006.
- ARISTA, M., BERJANO, R., VIRUEL, J.; ORTIZ, M. Á.; TALAVERA, M.; ORTIZ, P. L. Uncertain pollination environment promotes the evolution of a stable mixed reproductive system in the self-incompatible *Hypochoeris salzmanniana* (Asteraceae). **Annals of Botany**, v. 120, n. 3, p. 447-456, 2017.
- ASORO, F. G.; NEWELL, M. A.; BEAVIS, W. D.; SCOTT, M. P.; JANNINK, J. L. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. **The Plant Genome**, v. 4, n. 2, p. 132-144, 2011.
- BERGNER, L. M.; JAMIESON, I. G.; ROBERTSON, B. C. Combining genetic data to identify relatedness among founders in a genetically depauperate parrot, the Kakapo (*Strigops habroptilus*). **Conservation genetics**, v. 15, n. 5, p. 1013-1020, 2014.
- BERTON, L. H. C.; DE AZEVEDO FILHO, J. A.; SIQUEIRA, W. J.; COLOMBO, C. A. Seed germination and estimates of genetic parameters of promising macaw palm (*Acrocomia aculeata*) progenies for biofuel production. **Industrial Crops and Products**, v. 51, p. 258-266, 2013.
- BESPALHOK, J. C. F.; GUERRA, E. P.; OLIVEIRA, R. Sistemas reprodutivos de plantas cultivadas. In: BESPALHOK F., J.C; Guerra, E.P; Oliveira, R. Melhoramento de Plantas. v. 204, p. 11-18, 2014. Disponível em: <http://www.bespa.agrarias.ufpr.br/paginas/livro/capitulo%204.pdf>. Acesso em: 10 de agosto de 2018.
- BESSEGA, C.; POMETTI, C. L.; EWENS, M.; SAIDMAN, B. O.; VILARDI, J. C. Strategies for conservation for disturbed *Prosopis alba* (Leguminosae, Mimosoideae) forests based on mating system and pollen dispersal parameters. **Tree genetics & Genomes**, v. 8, n. 2, p. 277-288, 2012.
- BINK, M. C.; ANDERSON, A. D.; VAN DE WEG, W. E.; THOMPSON, E. A. Comparison of marker-based pairwise relatedness estimators on a pedigreed plant population. **Theoretical and Applied Genetics**, v. 117, n. 6, p. 843, 2008.

BLOUIN, M. S. DNA-based methods for pedigree reconstruction and kinship analysis in natural populations. **Trends in Ecology & Evolution**, v. 18, n. 10, p. 503-511, 2003.

BRESSAN, E. A.; SEBBENN, A. M.; FERREIRA, R. R.; LEE, T. S. G.; FIGUEIRA, A. *Jatropha curcas* L. (Euphorbiaceae) exhibits a mixed mating system, high correlated mating and apomixis. **Tree genetics & genomes**, v. 9, n. 4, p. 1089-1097, 2013.

BROWNE, L.; KARUBIAN, J. Rare genotype advantage promotes survival and genetic diversity of a tropical palm. **New Phytologist**, v. 218, n. 4, p. 1658-1667 2018.

BONATELLI, I. A. S.; CARSTENS, B. C.; MORAES, E. M. Using next generation RAD sequencing to isolate multispecies microsatellites for *Pilosocereus* (Cactaceae). **PloS one**, v. 10, n. 11, p. 1-14, 2015.

CHARPENTIER, M. J.; FONTAINE, M.C.; CHEREL E.; RENOULT J. P.; JENKINS, T.; BENOIT, L.; BARTHÈS, N.; ALBERTS, S. C.; TUNG, J. Genetic structure in a dynamic baboon hybrid zone corroborates behavioural observations in a hybrid population. **Molecular Ecology**, v. 21, n. 3, p. 715-731, 2012.

CLOUTIER, D.; KANASHIRO, M.; CIAMPI, A. Y.; SCHOEN, D. J. Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl. **Molecular Ecology**, v. 16, n. 4, p. 797-809, 2007.

COIMBRA, M. C.; JORGE, N. Fatty acids and bioactive compounds of the pulps and kernels of Brazilian palm species, guariroba (*Syagrus oleraces*), jervá (*Syagrus romanzoffiana*) and macaúba (*Acrocomia aculeata*). **Journal of the Science of Food and Agriculture**, v. 92, n. 3, p. 679-684, 2012.

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

CSILLÉRY, K., JOHNSON, T., BERARDI, D., CLUTTON-BROCK, T., COLTMAN, D., HANSSON, B., PEMBERTON, J. M. Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. **Genetics**, v. 173, n. 4, p. 2091-2101, 2006.

DA SILVA, E. C. B.; KUBOTA, T. Y. K.; DE MORAES, M. L. T.; SEBBENN, A. M. Coeficientes de herdabilidade e de parentesco em um fragmento florestal de *Araucaria angustifolia* (Bertol.) Kuntze utilizando marcadores genéticos. **Scientia Forestalis**, v.43, n. 105, p.147-153, 2015.

DE MORAES, M. L. T.; KAGEYAMA, P. Y.; SEBBENN, A. M. Mating system in small fragmented populations and isolated trees of *Hymenaea stigonocarpa*. **Scientia Forestalis**, n. 74, p. 75-86, 2007.

DESTA, Z. A.; ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. **Trends in plant science**, v. 19, n. 9, p. 592-601, 2014.

- DEVAUX, C.; LANDE, R.; PORCHER, E. Pollination ecology and inbreeding depression control individual flowering phenologies and mixed mating. **Evolution**, v. 68, n. 11, p. 3051-3065, 2014.
- DIAS, J. D. S.; dos SANTOS, D. T.; BRAGA, M.; ONOYAMA, M. M.; MIRANDA, C. H.; BARBOSA, P. F.; ROCHA, J. D. Produção de briquetes e péletes a partir de resíduos agrícolas, agroindustriais e florestais. **Embrapa Agroenergia-Documentos (INFOTECA-E)**, 2012.
- ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature Reviews Genetics**, v. 5, n. 6, p. 435, 2004.
- FALCONER, D.S.; MACKAY, T. **Introduction to quantitative genetics**. 4.ed. London: Longman Group Ltda, 1996. 464p.
- FALEIRO, F. G. **Marcadores genético-moleculares aplicados a programas de conservação e uso de recursos genéticos**. Planaltina-DF, Embrapa Cerrados, 2007,102p.
- FETZER, D. L.; CRUZ, P. N.; HAMERSKI, F.; CORAZZA, M. L. Extraction of baru (*Dipteryx alata* vogel) seed oil using compressed solvents technology. **The Journal of Supercritical Fluids**, v. 137, p. 23-33, 2018.
- Flora do Brasil 2020 em construção**. Jardim Botânico do Rio de Janeiro. Disponível em: <http://www.herbariovirtualreflora.jbrj.gov.br/reflora/floradobrasil/FB15663>. Acesso em: 02 Mai. 2018.
- FLOWER, C. E.; FANT, J. B.; HOBAN, S.; KNIGHT, K. S.; STEGER, L.; AUBIHL, E.; ROYO, A. A. Optimizing Conservation Strategies for a Threatened Tree Species: In Situ Conservation of White Ash (*Fraxinus americana* L.) Genetic Diversity through Insecticide Treatment. **Forests**, v. 9, n. 4, p. 202, 2018.
- FRIEDMAN, J.; BARRETT, S. C. H. Wind of change: new insights on the ecology and evolution of pollination and mating in wind-pollinated plants. **Annals of Botany**, v. 103, n. 9, p. 1515-1527, 2009.
- GARANT, D.; KRUIK, L. E. B. How to use molecular marker data to measure evolutionary parameters in wild populations. **Molecular Ecology**, v. 14, n. 7, p. 1843-1859, 2005.
- GIUSTINA, L. D.; BALDONI, A. B.; TONINI, H.; AZEVEDO, V. C. R.; NEVES, L. G.; TARDIN, F. D.; SEBBENN, A. M. Hierarchical outcrossing among and within fruits in *Bertholletia excels* Bonpl.(Lecythidaceae) open-pollinated seeds. **Genetics and Molecular Research**, v. 17, n. 1, 2018.
- GONÇALVES DA SILVA, A.; LALONDE, D. R.; QUSE, V.; SHOEMAKER, A.; RUSSELLO, M. A. Genetic approaches refine ex situ lowland tapir (*Tapirus terrestris*) conservation. **Journal of Heredity**, v. 101, n. 5, p. 581-590, 2010.

GOODWILLIE, C.; KALISZ, S.; ECKERT, C. G. The evolutionary enigma of mixed mating systems in plants: occurrence, theoretical explanations, and empirical evidence. **Annual Review of Ecology, Evolution, and Systematics**, v. 36, p. 47-79, 2005.

GRUEBER, C. E.; LAWS, R. J.; NAKAGAWA, S.; JAMIESON, I. G. Inbreeding depression accumulation across life-history stages of the endangered takahe. **Conservation Biology**, v. 24, n. 6, p. 1617-1625, 2010.

GUICHOUX, E.; LAGACHE, L.; WAGNER, S.; CHAUMEIL, P.; LÉGER, P.; LEPAIS, O.; PETIT, R. J. Current trends in microsatellite genotyping. **Molecular Ecology Resources**, v. 11, n. 4, p. 591-611, 2011.

HADLEY WICKHAM (2017). stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.2.0. Disponível em: <https://CRAN.R-project.org/package=stringr> .

HANSEN, O. K.; CHANGTRAGOON, S.; PONOY, B.; LOPEZ, J.; RICHARD, J.; KJAER, E. D. Worldwide translocation of teak—origin of landraces and present genetic base. **Tree Genetics & Genomes**, v. 13, n. 4, p. 87, 2017.

HARDY, O. J.; VEKEMANS, X. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. **Molecular Ecology Resources**, v. 2, n. 4, p. 618-620, 2002.

HOLSINGER, K. E. Reproductive systems and evolution in vascular plants. **Proceedings of the National Academy of Sciences**, v. 97, n. 13, p. 7037-7042, 2000.

HUANG, K.; RITLAND, K.; GUO, S.; SHATTUCK, M.; LI, B. A pairwise relatedness estimator for polyploids. **Molecular Ecology Resources**, v. 14, n. 4, p. 734-744, 2014.

JONES, A. G.; SMALL, C. M.; PACZOLT, K. A.; RATTERMAN, N. L. A practical guide to methods of parentage analysis. **Molecular Ecology Resources**, v. 10, n. 1, p. 6-30, 2010.

JONES, K. L.; GLENN, T. C.; LACY, R. C.; PIERCE, J. R.; UNRUH, N.; MIRANDE, C. M.; CHAVEZ-RAMIREZ, F. Refining the Whooping Crane Studbook by Incorporating Microsatellite DNA and Leg-Banding Analyses. **Conservation Biology**, v. 16, n. 3, p. 789-799, 2002.

KALISZ, S.; VOGLER, D. W. Benefits of autonomous selfing under unpredictable pollinator environments. **Ecology**, v. 84, n. 11, p. 2928-2942, 2003.

KONOVALOV, D. A.; HEG, D. I. K. TECHNICAL ADVANCES: A maximum-likelihood relatedness estimator allowing for negative relatedness values. **Molecular Ecology Resources**, v. 8, n. 2, p. 256-263, 2008.

KOZFKAY, C. C.; CAMPBELL, M. R.; HEINDEL, J. A.; BAKER, D. J.; KLINE, P.; POWELL, M. S.; FLAGG, T. A genetic evaluation of relatedness for broodstock

management of captive endangered Snake River sockeye salmon, *Oncorhynchus nerka*. **Conservation Genetics**, v. 9, n. 6, p. 1421, 2008.

LEMOS, M. R. B.; ALMEIDA SIQUEIRA, E. M.; ARRUDA, S. F.; ZAMBIAZI, R. C. The effect of roasting on the phenolic compounds and antioxidant potential of baru nuts [*Dipteryx alata* Vog.]. **Food Research International**, v. 48, n. 2, p. 592-597, 2012.

LI, C. C.; WEEKS, D. E.; CHAKRAVARTI, A. Similarity of DNA fingerprints due to chance and relatedness. **Human Heredity**, v. 43, n. 1, p. 45-52, 1993.

LIU, Z.; HUANG, C.; ZHOU, Q.; LI, Y.; WANG, Y.; LI, M.; TAKENAKA, A. Genetic analysis of group composition and relatedness in white-headed langurs. **Integrative Zoology**, v. 8, n. 4, p. 410-416, 2013.

LITT, M.; LUTY, J. A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American journal of human genetics**, v. 44, n. 3, p. 397, 1989.

LOISELLE, B. A.; SORK, V. L.; NASON, J.; GRAHAM, C. Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). **American Journal of Botany**, p. 1420-1425, 1995.

LORENZI, H.; SOUZA, H. D.; COSTA, J. D. M.; CERQUEIRA, L. D.; FERREIRA, E. **Palmeiras brasileiras e exóticas cultivadas**. Instituto plantarum de estudos da flora Ltda., 2004.

LUO, Z.W.; ZHANG, Z.; ZHANG, R. M.; PANDEY, M.; GAILING, O.; HATTEMER, H. H.; FINKELDEY, R. Modeling population genetic data in autotetraploid species. **Genetics**, v. 172, n. 1, p. 639-646, 2006.

LYNCH, M.; RITLAND, K. Estimation of pairwise relatedness with molecular markers. **Genetics**, v. 152, p.1753–1766, 1999.

LYNCH, M.; WALSH, B. **Genetics and analysis of Quantitative Traits**. Sunderland, MA: Sinauer, 1998.

MALLECOT, G. **The Mathematic of Heredity**. W. H. Freeman and Co. São Francisco, 1969. 103p.

MARTINS, F. S.; BORGES, L. L.; PAULA, J. R.; CONCEIÇÃO, E. C. Impact of different extraction methods on the quality of *Dipteryx alata* extracts. **Revista Brasileira de Farmacognosia**, v. 23, n. 3, p. 521-526, 2013.

MATTILA A.L.; DUPLOUY, A.; KIRJOKANGAS, M.; LEHTONEN, R.; RASTAS, P.; HANSKI, I. High genetic load in an old isolated butterfly population. **Proceedings of the National Academy of Sciences**, v. 109, n. 37, p. E2496-E2505, 2012.

MIAH, G.; RAFII, M. Y.; ISMAIL, M. R.; PUTEH, A. B.; RAHIM, H. A.; ISLAM, K. N.; LATIF, M. A. A review of microsatellite markers and their applications in rice breeding programs to improve blast disease resistance. **International Journal of Molecular sciences**, v. 14, n. 11, p. 22499-22528, 2013.

MILLIGAN, B. G. Maximum-likelihood estimation of relatedness. **Genetics**, v. 163, n. 3, p. 1153-1167, 2003.

MIRANDA FILHO, J. B. Endogamia ou consanguinidade. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. **Recursos Genéticos e Melhoramento de Plantas**. Rondonópolis: Fundação MT, p.201-230, 2001.

MITTAL, N.; DUBEY, A. K. Microsatellite markers-A new practice of DNA based markers in molecular genetics. **Pharmacognosy Reviews**, v. 3, n. 6, p. 235, 2009.

MORAES, M. L. T.; KAGEYAMA, P. Y.; SEBBENN, A. M. Sistema de reprodução em pequenas populações fragmentadas e em árvores isoladas de *Hymenaea stigonocarpa*. **Scientia Forestalis**, n. 74, p. 75-86, 2007.

MORAES, M. L. T.; SEBBENN, A. M. Pollen dispersal between isolated trees in the Brazilian savannah: a case study of the neotropical tree *Hymenaea stigonocarpa*. **Biotropica**, v. 43, n. 2, p. 192-199, 2011.

NICHOLAS, F. W.; SMITH, C. Increased rates of genetic change in dairy cattle by embryo transfer and splitting. **Animal Science**, v. 36, n. 3, p. 341-353, 1983.

NIELSEN, J. F.; ENGLISH, S.; GOODALL-COPESTAKE, W. P.; WANG, J.; WALLING, C. A.; BATEMAN, A. W.; KRUIK, L. E. Inbreeding and inbreeding depression of early life traits in a cooperative mammal. **Molecular Ecology**, v. 21, n. 11, p. 2788-2804, 2012.

NUCCI, S.M.; AZEVEDO-FILHO, J.A.; COLOMBO, C.A.; PRIOLLI, R.H.G.; COELHO, R.M.; MATA, T.L.; ZUCCHI, M.I. Development and characterization of microsatellites markers from the macaw. **Molecular Ecology Resources**, v. 8, p. 224-226, 2008.

OLIVEIRA, D. A.; MELO JÚNIOR, A. F.; BRANDÃO, M. M.; RODRIGUES, L. A.; MENEZES, E. V.; FERREIRA, P. R. B. Genetic diversity in populations of *Acrocomia aculeata* (Arecaceae) in the northern region of Minas Gerais, Brazil. **Genetics and molecular research**, v. 11, n. 1, p. 531-538, 2012.

OLIVEIRA, M. I. B.; SIGRIST, M. R. Fenologia reprodutiva, polinização e reprodução de *Dipteryx alata* Vogel (Leguminosae-Papilionoideae) em Mato Grosso do Sul, Brasil. **Revista Brasileira de Botânica**, v.31, n.2, p.195-207, 2008.

ORSI, P. R.; BONAMIN, F.; SEVERI, J. A.; SANTOS, R. C.; VILEGAS, W.; HIRUMALIMA, C. A.; DI STASI, L. C. *Hymenaea stigonocarpa* Mart. ex Hayne: A Brazilian medicinal plant with gastric and duodenal anti-ulcer and antidiarrheal effects in experimental rodent models. **Journal of Ethnopharmacology**, v. 143, n. 1, p. 81-90, 2012.

ORSI, P. R.; SEITO, L. N.; DI STASI, L. C. *Hymenaea stigonocarpa* Mart. ex Hayne: A tropical medicinal plant with intestinal anti-inflammatory activity in TNBS model of intestinal inflammation in rats. **Journal of Ethnopharmacology**, v. 151, n. 1, p. 380-385, 2014.

PEMBERTON, J. M. Wild pedigrees: the way forward. **Proceedings of the Royal Society of London B: Biological Sciences**, v. 275, n. 1635, p. 613-621, 2008.

POMETTI, C. L.; BESSEGA, C. F.; VILARDI, J. C.; SAIDMAN, B. O. Comparison of mating system parameters and genetic structure in three natural scenarios of *Acacia visco* (Leguminosae, Mimosoideae). **Plant Systematics and Evolution**, v. 299, n. 4, p. 761-771, 2013.

QUELLER, D. C.; GOODNIGHT, K. F. Estimating relatedness using genetic markers. **Evolution**, v. 43, n. 2, p. 258-275, 1989.

R Core Team (2017). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <https://www.r-project.org/>.

RAMALHO, M. A. P.; ABREU, A. F. B.; SANTOS, J. B. Melhoramento de espécies autógamas. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, p.201-230, 2001.

RITLAND K. Estimators for pairwise relatedness and individual inbreeding coefficients. **Genetical Research**, v.67, p.175–185, 1996a.

RITLAND, K. A marker-based method for inferences about quantitative inheritance in natural populations. **Evolution**, v. 50, n. 3, p. 1062-1073, 1996b.

RITLAND, K. Extensions of models for the estimation of mating systems using independent loci. **Heredity**, London, v.88, p.221-228, 2002.

SANTOS, D. N.; FERREIRA, J. L.; SETOTAW, T. A.; CANCADO, G.; PASQUAL, M.; LONDE, L. C.; VENDRAME, W. A. Genetic structure from the oldest *Jatropha* germplasm bank of Brazil and contribution for the genetic improvement. **Anais da Academia Brasileira de Ciências**, v. 88, n. 4, p. 2363-2374, 2016.

SEBBENN, A. M.; SEOANE, C. E. S. Estimative of inbreeding effective size by genetic markers. **Revista Árvore**, v. 29, n. 1, p. 1-8, 2005.

SERROTE, C. M. L.; REINIGER, L. R. S.; STEFENON, V. M.; CURTI, A. R.; COSTA, L. S.; PAIM, A. F. Simulating the selfing and migration of *Luehea divaricata* populations in the Pampa biome to investigate the conservation potential of their genetic resources. **Genetics and Molecular Research**, v. 15, n. 3, 2016.

SILVA, C. R. S.; ALBUQUERQUE, P. S. B.; ERVEDOSA, F. R.; MOTA, J. W. S.; FIGUEIRA, A.; SEBBENN, A. M. Understanding the genetic diversity, spatial genetic structure and mating system at the hierarchical levels of fruits and individuals of a continuous *Theobroma cacao* population from the Brazilian Amazon. **Heredity**, v. 106, n. 6, p. 973, 2011.

SILVA, S. M.M., MARTINS, K., MESQUITA, A. G.G.; WADT, L. H. O. Parâmetros genéticos para a conservação de *Hymenaea courbaril* L. na amazônia sul-ocidental. **Ciência Florestal**, V. 24, N. 1, 2014.

SMITH, J.M. **Evolutionary genetics**. Oxford: Oxford University Press, 1999. 328 p.

SOBIERAJSKI, G. R.; KAGEYAMA, P. Y.; SEBBENN, A. Sistema de reprodução em nove populações de *Mimosa scabrella* Bentham (*Leguminosaceae*). **Scientia Forestalis**, Piracicaba, n.71, p.37-49, 2006.

SOKAL R. R.; ROHIF F I. **Biometry: the principles and practice of statistics in biological research**. San Francisco: W.H. Freeman, 1969. 776 p.

SOUZA JUNIOR, C. L. Melhoramento de espécies alógamas. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, p.159-200, 2001.

SOUZA, L. M.; KAGEYAMA, P. Y.; SEBBENN, A. M. Mating system in *Chorisia speciosa* A. St.-Hil. (Bombacaceae) natural population. **Brazilian Journal of Botany**, v. 26, n. 1, p. 113-121, 2003.

TAMBARUSSI, E. V.; BOSHIER, D. ; VENCOVSKY, R ; FREITAS, MLM ; SEBBENN, A. M. Inbreeding depression from selfing and mating between relatives in the Neotropical tree *Cariniana legalis* Mart. Kuntze. **Conservation Genetics**, v. 18, p. 225-234, 2017a.

TAMBARUSSI, E., SEBBENN, A., ALVES-PEREIRA, A., VENCOVSKY, R., CAMBUIM, J., DA SILVA, A., MORAES, M., DE MORAES, M. *Dipteryx alata* Vogel (Fabaceae) a neotropical tree with high level of selfing: implication for conservation and breeding programs. **Annals of Forest Research**, v. 60, n.2, p. 243-261, 2017b.

TAYLOR, H. R. The use and abuse of genetic marker-based estimates of relatedness and inbreeding. **Ecology and Evolution**, v. 5, n. 15, p. 3140-3150, 2015.

THOMAS, S. C. The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. **Philosophical Transactions of the Royal Society of London B: Biological Sciences**, v. 360, n. 1459, p. 1457-1467, 2005.

THOMPSON, E. A. Identity by descent: variation in meiosis, across genomes, and in populations. **Genetics**, v. 194, n. 2, p. 301-326, 2013.

THOMPSON, E. A. The estimation of pairwise relationships. **Annals of Human genetics**, v. 39, n. 2, p. 173-188, 1975.

TURCHETTO-ZOLET, A. C.; TURCHETTO, C.; ZANELLA C. M.; PASSAIA G. **Marcadores Moleculares na Era genômica: Metodologias e Aplicações**. 1ª edição, Ribeirão Preto: Sociedade Brasileira de Genética, 2017. 181 p. ISBN 978-85-89265-26-3. Disponível em: <https://www.sbg.org.br/sites/default/files/e_book_marcadores_moleculares_sbg_2017_final.pdf> Acesso: 30 de abril de 2018.

- VAN DE CASTEELE, T.; GALBUSERA, P.; MATTHYSEN, E. A comparison of microsatellite-based pairwise relatedness estimators. **Molecular Ecology**, v. 10, n. 6, p. 1539-1549, 2001.
- VENCOVSKY, R.; PEREIRA, M.B.; CRISÓSTOMO, J.R.; FERREIRA, M.A.J. Genética e melhoramento de populações mistas. In: NASS, L.L.; VALOIS, A.C.C.; MELO, I.S.; VALADARES-INGLIS, M.C. **Recursos genéticos e melhoramento de plantas**. Rondonópolis: Fundação MT, p.231-281, 2001.
- VENCOVSKY, R.; CROSSA, J. Measurements of representativeness used in genetic resources conservation and plant breeding. **Crop science**, v. 43, n. 6, p. 1912-1921, 2003.
- VINSON, C. C.; KANASHIRO, M.; HARRIS, S. A.; BOSHIER, D. H. Impacts of selective logging on inbreeding and gene flow in two Amazonian timber species with contrasting ecological and reproductive characteristics. **Molecular ecology**, v. 24, n. 1, p. 38-53, 2015.
- WANG, J. An estimator for pairwise relatedness using molecular markers. **Genetics**, v. 160, p. 1203–1215, 2002.
- WANG, J. COANCESTRY: a program for simulating, estimating and analysing relatedness and inbreeding coefficients. **Molecular Ecology Resources**, v. 11, n. 1, p. 141-145, 2011.
- WANG, J. Estimating pairwise relatedness in a small sample of individuals. **Heredity**, v. 119, n. 5, p. 302, 2017.
- WANG, J. Marker-based estimates of relatedness and inbreeding coefficients: an assessment of current methods. **Journal of Evolutionary Biology**, v. 27, n. 3, p. 518-530, 2014.
- WANG, J. Triadic IBD coefficients and applications to estimating pairwise relatedness. **Genetics Research**, v. 89, n. 3, p. 135-153, 2007.
- WANG, J. Unbiased relatedness estimation in structured populations. **Genetics**, v. 187, n. 3, p. 887-901, 2011a.
- WEIR, B. S.; ANDERSON, A. D.; HEPLER, A. B. Genetic relatedness analysis: modern data and new challenges. **Nature Reviews Genetics**, v. 7, n. 10, p. 771-780, 2006.
- WHITEHEAD, M. R.; LANFEAR, R.; MITCHELL, R. J.; KARRON, J. D. Plant Mating Systems Often Vary Widely Among Populations. **Frontiers in Ecology and Evolution**, v. 6, p. 38, 2018.
- WRIGHT, S. Coefficients of inbreeding and relationship. **The American Naturalist**, v. 56, n. 645, p. 330-338, 1922.