UNIVERSIDADE FEDERAL DE SANTA MARIA
CENTRO DE CIÊNCIAS RURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA FLORESTAL

Caetano Miguel Lemos Serrote

SIMULAÇÕES DE PARÂMETROS GENÉTICOS A PARTIR DE MICROSSATÉLITES: CONTRIBUIÇÕES PARA A CONSERVAÇÃO DE RECURSOS FLORESTAIS DE BIOMAS BRASILEIROS

Santa Maria, RS
2022

Caetano Miguel Lemos Serrote

SIMULAÇÕES DE PARÂMETROS GENÉTICOS A PARTIR DE MICROSSATÉLITES: CONTRIBUIÇÕES PARA A CONSERVAÇÃO DE RECURSOS FLORESTAIS DE BIOMAS BRASILEIROS

Tese apresentada ao Curso de Doutorado do Programa de Pós-Graduação em Engenharia Florestal, Área de Concentração em Silvicultura, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Doutor em Engenharia Florestal**

Orientadora: Profª. Drª. Lia Rejane Silveira Reiniger

Santa Maria, RS
2022

This study was financied in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) – Finance Code 001

Serrote, Caetano Miguel Lemos
Simulações de parâmetros genéticos a partir de microssatélites: contribuições para a conservação de recursos florestais de biomas brasileiros / Caetano Miguel Lemos Serrote.- 2022.
121 p.; 30 cm

Orientadora: Lia Rejane Silveira Reiniger
Coorientadora: Berta Maria Heinzmann
Tese (doutorado) - Universidade Federal de Santa Maria, Centro de Ciências Rurais, Programa de Pós Graduação em Engenharia Florestal, RS, 2022

1. Conservação genética 2. Easypop 3. Ecossistemas florestais 4. Marcadores moleculares I. Reiniger, Lia Rejane Silveira II. Heinzmann, Berta Maria III. Título.

Sistema de geração automática de ficha catalográfica da UFSM. Dados fornecidos pelo autor(a). Sob supervisão da Direção da Divisão de Processos Técnicos da Biblioteca Central. Bibliotecária responsável Paula Schoenfeldt Patta CRB 10/1728.

Declaro, CAETANO MIGUEL LEMOS SERROTE, para os devidos fins e sob as penas da lei, que a pesquisa constante neste trabalho de conclusão de curso (Tese) foi por mim elaborada e que as informações necessárias objeto de consulta em literatura e outras fontes estão devidamente referenciadas. Declaro, ainda, que este trabalho ou parte dele não foi apresentado anteriormente para obtenção de qualquer outro grau acadêmico, estando ciente de que a inveracidade da presente declaração poderá resultar na anulação da titulação pela Universidade, entre outras consequências legais.

Caetano Miguel Lemos Serrote

SIMULAÇÕES DE PARÂMETROS GENÉTICOS A PARTIR DE MICROSSATÉLITES: CONTRIBUIÇÕES PARA A CONSERVAÇÃO DE RECURSOS FLORESTAIS DE BIOMAS BRASILEIROS

Tese apresentada ao Curso de Doutorado do Programa de Pós-Graduação em Engenharia Florestal, Área de Concentração em Silvicultura, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Doutor em Engenharia Florestal.**

Aprovado em 28 de setembro de 2022

Lia Rejane Silveira Reiniger

Lia Rejane Silveira Reiniger, Dra. (UFSM)
(Presidente/Orientadora)

Candida Elisa Manfio

Candida Elisa Manfio, Profa. Dra. (EPAGRI)

Charlene Moro Stefanel

Charlane Moro Stefanel, Dra. (PMSPS)

Diego Pascoal Golle, Prof. Dr. (UNICRUZ)

Marlove Fátima Brião Muniz, Profa. Dra. (UFSM)

Santa Maria, RS
2022

“Foi o tempo que dedicaste à tua rosa que a fez tão importante”.

Antoine de Saint-Exupéry

Dedico esse trabalho à minha esposa
Angelina e ao meu filho Daniel,
companheiros nos momentos bons e maus.

AGRADECIMENTOS

Á Deus, pela vida e pela força necessária para vencer os desafios do dia-a-dia.

Ao Programa de Pós-Graduação em Engenharia Florestal da Universidade Federal de Santa Maria e ao CNPq pelo apoio material e financeiro para minha formação.

A Universidade Lúrio pela liberação e apoio financeiro para a realização desse sonho.

À minha família, pela base de sustentação, apoio moral e financeiro, durante toda a minha carreira estudantil e profissional.

A minha orientadora, professora Dra. Lia Rejane Silveira Reiniger, por ter apostado em mim e me concedido a oportunidade de ascender em minha carreira acadêmica, pelos valiosos conhecimentos transmitidos, pelo excelente profissionalismo e rigor, combinados com muita simplicidade, atenção e amizade, ingredientes determinantes para o sucesso deste trabalho.

À minha co-orientadora, Profa. Dra. Berta Maria Heinzmann, pela prontidão em colaborar na efetivação deste projeto.

Às minhas colegas do Núcleo de Biotecnologia e Melhoramento, Charlene Stefanel, Karol Buuron da Silva, Silvia Machado e Ana Cristina Ziegler pelo auxílio, troca de conhecimentos, parceria e amizade. Passei por momentos incríveis e experiências inesquecíveis que levare para toda a vida.

Aos amigos que fiz durante a formação, muito obrigado pelo acolhimento, companheirismo e carinho.

Muito obrigado a todos!

RESUMO

SIMULAÇÕES DE PARÂMETROS GENÉTICOS A PARTIR DE MICROSSATÉLITES: CONTRIBUIÇÕES PARA A CONSERVAÇÃO DE RECURSOS FLORESTAIS DE BIOMAS BRASILEIROS

AUTOR: Caetano Miguel Lemos Serrote
ORIENTADORA: Lia Rejane Silveira Reiniger

A crescente destruição de ecossistemas florestais tem preocupado a sociedade em nível global e determinado que venham sendo elaboradas estratégias cada vez mais efetivas para a conservação de seus recursos genéticos. Na presente pesquisa foram empregados dados de marcadores microssatélites em simulações computacionais com a finalidade de contribuir nos esforços relacionados à conservação de recursos florestais de biomas brasileiros. Com essa finalidade, foi utilizado o programa EasyPop para efetuar simulações em populações florestais da maioria dos biomas brasileiros objetivando gerar parâmetros para o planejamento de intervenções visando à conservação desses recursos genéticos. Os resultados obtidos indicam uma tendência de redução de diversidade genética nas populações estudadas como consequência de atividades antrópicas. Nesse contexto, nosso trabalho reforça a necessidade de adoção de medidas com vistas a minimizar a perda de diversidade genética, entre as quais a promoção do fluxo gênico entre populações/fragmentos florestais.

Palavras-chave: Conservação genética; Easypop; Ecossistemas florestais; Marcadores moleculares.

ABSTRACT

SIMULATIONG GENETIC PARAMETERS FROM MICROSATELLITES: CONTRIBUTIONS TO FOREST RESOURCES CONSERVATION IN BRAZILIAN BIOMES

AUTHOR: Caetano Miguel Lemos Serrote
ADVISER: Lia Rejane Silveira Reiniger

The increasing destruction of forest ecosystems has been of concern to society worldwide and demands for more effective conservation strategies for these genetic resources. In this research, microsatellite markers data were used in computer simulations aiming the conservation of forest resources in the Brazilian biomes. The EasyPop program was used to perform simulations in forest populations representing the different Brazilian biomes in order to generate genetic, ecological and reproductive parameters useful for planning interventions for conservation of these resources. The results indicate a trend for low genetic variability in the studied populations as a consequence of anthropic activities. In this context, our work reinforces the need for adopting measures aiming to minimize the loss of genetic diversity, such as promoting gene flow between forest populations/fragments.

Keywords: Genetic conservation; Easypop; Forest ecosystems; Molecular markers.

SUMÁRIO

1 INTRODUÇÃO GERAL

Nas últimas décadas os recursos naturais, incluindo os florestais, vêm sendo destruídos em ritmo acelerado pela alegada necessidade de atender às demandas das populações humana que crescem exponencialmente. Além de estarem sendo destruídos, ecossistemas que albergam populações de espécies florestais vêm sendo fragmentados em subpopulações de menor tamanho e com isolamento espacial entre indivíduos, podendo desencadear processos ecológicos e genéticos com consequências potencialmente desastrosas, representando, assim, a maior ameaça para a manutenção e viabilidade das populações naturais de plantas (LOWE et al., 2005; QUESADA et al., 2013).

Populações de espécies com tamanho reduzido são mais suscetíveis aos efeitos da deriva genética e da endogamia, conduzindo a uma diminuição da variabilidade genética, que é a componente chave para a sustentabilidade das espécies, pois fornece matéria-prima para os processos de adaptação, evolução e sobrevivência, especialmente em um cenário de mudanças ambientais. Em curto prazo, as populações com baixa variabilidade genética têm aumentada sua susceptibilidade a doenças e pragas, perda de alelos de incompatibilidade e fixação de alelos deletérios. Em longo prazo, a perda de variabilidade genética pode conduzir à redução na capacidade das populações em responder a alterações nas pressões de seleção. Assim, a abordagem genética da conservação concentra sua atenção na minimização dos efeitos do isolamento reprodutivo entre indivíduos pertencentes a fragmentos diferentes, através da promoção do fluxo gênico entre eles (NEIGEL, 1997; LOWE et al., 2005).

O tamanho de uma população abaixo do qual os efeitos genéticos são mais severos, tornando o risco de extinção extremamente elevado, é denominado População Mínima Viável (PMV), parâmetro que pode ser estimado tomando-se em consideração aspectos reprodutivos, demográficos e genéticos de uma determinada população. Os dados utilizados na estimação podem ser obtidos mediante o emprego de marcadores de DNA, como, por exemplo, microssatélites (CONSON et al., 2013).

O emprego de marcadores de DNA em estudos sobre conservação tornou-se mais efetivo com o desenvolvimento de programas de computador que permitem

realizar simulações. Tratam-se de programas que usam dados da estrutura genética obtidos com o uso de marcadores de DNA para gerar parâmetros genéticos, ecológicos e reprodutivos associados a esses dados, permitindo fazer projeções do comportamento da variabilidade genética com o decorrer do tempo e, desse modo, planejar com maior efetividade eventuais intervenções visando a conservação (FLATHER, 2011; PENG et al., 2013).

Entre diversos programas de simulação desenvolvidos, no presente trabalho será usado o EASYPOP (BALLOUX, 2001) para estudar os padrões genéticos, ecológicos e reprodutivos que melhor explicam a estrutura genética de diferentes espécies arbóreas florestais que ocorrem em biomas brasileiros, os quais foram obtidos com base em marcadores de DNA do tipo microssatélites. Os resultados obtidos a partir dessas simulações poderão fornecer subsídios para o planejamento visando a conservação, na medida em que permitirão analisar a tendência de erosão da variabilidade genética e, quando necessário, serão sugeridas alternativas para conter essa erosão.

Diante do exposto, o presente trabalho tem como objetivo geral contribuir para a conservação dos recursos genéticos florestais de biomas brasileiros por meio de simulações de parâmetros genéticos obtidos pelo emprego de marcadores microssatélites. Foram realizadas sete pesquisas, organizadas em capítulos, conforme a descrição a seguir. No capítulo I, foi realizada uma revisão de literatura sobre a determinação do Conteúdo de Informação Polimórfica (PIC), que mede a capacidade de um marcador molecular em revelar polimorfismo. No capítulo II, foi realizado um levantamento dos estudos genéticos realizados em biomas brasileiros com o uso de marcadores de DNA do tipo microssatélites. No capítulo III, objetivou-se estimar, por meio de simulações, o padrão do fluxo gênico em fragmentos de *Cabralea canjerana* em desenvolvimento no bioma da Mata Atlântica. No capítulo IV, recorreram-se a simulações para o estudo da estrutura genética de populações *Hancornia speciosa,* do bioma Caatinga. No capítulo V, foi simulado o sistema reprodutivo de populações de H*ymenaea courbaril* em desenvolvimento na Amazônia Sul-Ocidental. No capítulo VI, foi simulado o padrão de fluxo gênico de fragmentos de *Copaifera langsdorffii*, do Cerrado. Finalmente, no capítulo VII foi estudado, por meio

de simulações, o fluxo gênico em remanescentes de *Prosopis rubriflora,* localizados no bioma Pantanal.

1. 2 OBJETIVOS

1.2.1 Objetivo geral

A presente tese tem como objetivo geral contribuir para a conservação dos recursos genéticos florestais de biomas brasileiros por meio de simulações de parâmetros genéticos obtidos a partir de dados de marcadores microssatélites.

1.2.2 Objetivos específicos

Analisar o estado da arte da utilização de marcadores microssatélites em estudos com populações naturais de espécies arbóreas florestais nativas de biomas brasileiros publicados no período compreendido entre os anos de 2017 e 2021.

Identificar e selecionar, por meio de simulações computacionais, estimativas de parâmetros que melhor justificam os parâmetros de estrutura genética obtidos a partir de análises com marcadores microssatélites.

Estimar parâmetros genéticos populacionais de utilidade para a conservação de espécies florestais de biomas brasileiros.

2 REVISÃO DE LITERATURA

2.1 FRAGMENTAÇÃO FLORESTAL E VARIABILIDADE GENÉTICA

Nas últimas décadas tem se observado uma crescente destruição e fragmentação de ecossistemas florestais, por ações antrópicas que incluem construção de estradas, ferrovias, cidades e abertura de campos agrícolas. Habitats ocupando extensas áreas são, com frequência, divididos em porções menores, uma condição que potencializa a ocorrência de processos ecológicos e genéticos com consequências potencialmente desastrosas (ALTER et al., 2007).

Além de isolar reprodutivamente indivíduos que contêm apenas uma pequena amostra do conjunto gênico da população original, a fragmentação florestal pode causar contínua perda de alelos devido à deriva genética (tópico desenvolvido na sequência), caso a população remanescente permaneça isolada por várias gerações, reduzindo, assim, a diversidade de combinações alélicas e a consequente flexibilidade adaptativa da população em face de alterações climáticas, aumentando a probabilidade de seu declínio. Enquanto uma área grande de habitat pode ter sustentado uma única população grande, é possível que nenhum de seus fragmentos possa sustentar o suficiente uma subpopulação para que ela sobreviva por um longo período (ALLARD, 1971; RIDLEY, 2006; VINSON et al., 2015).

Desse modo, pesquisas sobre a conservação dos recursos naturais têm focado no estudo da variabilidade genética existente entre e dentro de populações, considerando-se a população como a unidade básica para a conservação da biodiversidade no nível de espécies individuais. A conservação da variabilidade genética é determinante para a futura evolução e adaptabilidade das populações locais e da espécie no geral, pois a evolução atua sobre uma variabilidade existente na população, permitindo à seleção natural preservar os indivíduos com maior aptidão às condições ambientais correntes (SILVA et al., 2014).

2.2 SISTEMAS REPRODUTIVOS EM PLANTAS

O sistema de reprodução refere-se à forma como as populações de uma espécie recombinam seus genes a cada geração para formar a população

descendente, determinando, desse modo, como a variabilidade genética de uma espécie se organiza no espaço e no tempo. Conforme o modo de reprodução sexual predominante, populações naturais de espécies florestais podem ser classificadas em autógamas, alógamas ou mistas. Existem diversos critérios para classificar uma espécie quanto ao modo reprodutivo, sendo que, de acordo com a classificação comumente utilizada, quando, em suas populações, predomina a autofecundação (acima de 95% das fecundações que ocorrem), a espécie é classificada como autógama; quando predomina a fecundação cruzada (acima de 95%) ela é alógama; espécies intermediárias ou mistas possuem uma taxa de autofecundação entre 5 e 95% e são, normalmente, tratadas no melhoramento como alógamas (DESTRO; MONTALVÁN, 1999).

A grande maioria das espécies de angiospermas é hermafrodita, com flores bissexuais que contêm órgãos reprodutivos tanto androicos como ginoicos (estames e gineceu respectivamente), e somente 6% delas são dioicas (BAWA et al., 1985; MATSUNAGA et al., 2003), o que, neste caso, implica, obrigatoriamente, em alogamia. Apesar de não haver separação espacial dos órgãos sexuais na maioria das angiospermas, as espécies arbóreas tropicais geralmente apresentam sistema misto de reprodução com predomínio de cruzamentos (taxa média estimada em 0,88), embora parte dos cruzamentos ocorra de forma não aleatória, mas, sim, entre árvores aparentadas (cruzamentos endogâmicos) e de forma sistemática entre as mesmas árvores (cruzamentos biparentais), gerando progênies compostas por misturas de irmãos de autofecundação, irmãos germanos ou completos, além de meios irmãos. Esse fato se deve a adaptações evolutivas desenvolvidas por essas espécies para evitar a autofecundação e conservar a variação genética (SEBBENN, 2002; SOBIERAJSKI et al., 2006).

Dentre os vários mecanismos que favorecem a fecundação cruzada, a autoincompatibilidade é considerada o mais frequente (FRANKEL; GALUN, 1977). A autoincompatibilidade é um mecanismo genético-fisiológico que consiste na impossibilidade de produção de zigoto pela autofecundação ou pela fecundação de plantas que apresentam os mesmos alelos de incompatibilidade (SOBIERAJSKI et al., 2006). Sua ocorrência foi registrada em espécies lenhosas do bioma Caatinga, no nordeste brasileiro (LEITE & MACHADO, 2010), na espécie arbustiva *Camellia*

oleifera, nativa da China (LIAO et al., 2014), em *Capparis retusa*, no nordeste da Argentina (BIANCHINI & GIBBS, 2000), em *Dipterocarpus tempehes*, na Malásia (KENTA et al., 2002), em *Eucalyptus globulus*, na Austrália (MCGOWEN et al., 2010), entre diversos estudos realizados no mundo todo. No estudo de LEITE & MACHADO (2010) foram analisada a biologia reprodutiva de 15 espécies do bioma Caatinga, tendo sido observado que autoincompatibilidade em 73,3% delas.

2.3 MARCADORES MICROSSATÉLITES

Diversas técnicas moleculares permitem determinar a variação presente em uma espécie e em suas populações. Entre essas técnicas, incluem-se aquelas destinadas à detecção de variação de sequências de DNA, conhecidas como marcadores de DNA. Elas permitem identificar variabilidade nas sequências de DNA dos indivíduos analisados a um nível de resolução superior ao polimorfismo passível de detecção ao nível morfológico, permitindo estimar parâmetros genéticos úteis para um amplo espectro de estudos, desde identidade genética e construção de mapas genéticos até análises filogenéticas e evolucionárias (KALIA et al, 2011).

Entre diversos marcadores moleculares, os microssatélites (também denominados “Simple Sequence Repeats – SSR”) consistem de sequências curtas de nucleotídeos repetidas, lado a lado (em *tandem*), cujo nível de polimorfismo produzido é devido à variação no número de unidades de repetição em um determinado loco. Os genomas eucariotos são densamente povoados por sequências simples repetidas, consistindo de dois a seis nucleotídeos repetidos em *tandem*. As sequências de DNA que flanqueiam os microssatélites são geralmente conservadas entre os indivíduos de uma mesma espécie, permitindo a seleção de *primers* ou iniciadores específicos que amplificam, via Reação em Cadeia da Polimerase – conhecida pelo acrônimo em língua inglesa “PCR”, fragmentos contendo o DNA repetitivo em todos os genótipos. Resultantes de suas altas taxas mutacionais, alelos polimórficos ocorrem sempre que em um genoma esteja faltando a sequência repetida ou naqueles que têm uma deleção ou uma inserção que modifique a distância entre as repetições, permitindo a comparação do material genético de indivíduos diferentes (KALIA et al, 2011; RODRIGUES et al., 2010).

A hipervariabilidade dos microssatélites é mantida, principalmente, pelo pareamento errôneo dos filamentos polinucleotídicos durante a replicação do DNA, com o deslize (*slipage*) da enzima DNA polimerase. O número de repetições pode aumentar ou diminuir, caso ocorra adição ou perda de uma ou mais unidades de repetição pela DNA polimerase e se houver falha na atividade exonucleásica dessa enzima. Por possuírem elevada reprodutibilidade e serem de herança codominante, esses marcadores vêm sendo bastante utilizados para a estimação de parâmetros genéticos de populações vegetais e para a compreensão de padrões de fluxo gênico e parentesco. Porém, sua utilização em espécies florestais ainda hoje é limitada pela dificuldade de desenvolver iniciadores para essas espécies devido à exigência de um conhecimento prévio, pelo menos de parte, do genoma a ser estudado, ao elevado custo e tempo necessário para seu desenvolvimento (SENAN et al., 2014; BARRANDEGUY; GARCIA, 2016).

2.4 ESTRUTURA GENÉTICA POPULACIONAL

A distribuição heterogênea (não aleatória) dos alelos e genótipos no espaço e no tempo, resultante de forças evolutivas, é denominada *estrutura genética* e pode influenciar significativamente os processos genéticos e ecológicos das espécies, resultando, assim, em agregados de genótipos específicos. O nível de alteração genética entre populações naturais é de importância crítica para o entendimento de muitos processos biológicos, como a diferenciação genética e a manutenção da variabilidade genética em condições de fragmentação de populações (HAMRICK, 1982; ZUCCHI, 2002; STEFENON, 2010).

Diversos parâmetros são utilizados para caracterizar as populações, entre os quais o número de alelos, conteúdo de informação polimórfica, heterozigosidade observada e esperada, coeficiente de endogamia ou índice de fixação. O número total de alelos existente na população constitui-se em um parâmetro populacional importante por estar relacionado à capacidade de as populações se adaptarem às alterações ambientais. Quanto maior o número de alelos na população, maior a diversidade de combinações possíveis que originam diferentes genótipos capazes de se adaptar a determinadas condições do ambiente. Uma medida relativa da quantidade de alelos é a riqueza alélica, a qual corresponde ao número médio de

alelos segregando em um loco dentro de uma população, mas pode ser também abordada em termos de número médio de alelos por loco. Esse parâmetro permite comparar a variabilidade genética de populações com diferente tamanho amostral (baseado no menor tamanho amostral entre as populações) (SANCHEZ-MAZAS et al., 2012; VINSON et al., 2015).

2.5 FLUXO GÊNICO

A variação genética existente entre populações de uma espécie resulta de um balanço entre forças evolutivas que tendem a produzir diferenciação genética local – como a deriva e a seleção natural – e a força que tende a promover uma homogeneização genética entre as populações – o fluxo gênico. O fluxo gênico corresponde à entrada ou saída de genes, sendo capaz de alterar a composição genética original em uma determinada população. O fluxo gênico expressa o número de migrantes entre populações por geração, importante para conectar reprodutivamente populações isoladas (HELLBERG et al., 2002; ZUCCHI, 2002).

Assim, a maior parte dos programas de conservação tem seu enfoque na redução das barreiras para distribuição de genes nas populações de modo a minimizar a perda de genes. Tais barreiras podem ser fisiológicas e/ou de distância, muitas vezes criadas pela fragmentação de populações, resultando na restrição do fluxo de genes entre as subpopulações, propiciando, assim, a endogamia e deriva genética. Desse modo, o fluxo gênico em populações naturais de plantas é fundamental na homogeneização das frequências alélicas entre populações pequenas, de modo a que se comportem como uma grande população com cruzamentos aleatórios, mesmo que separadas geograficamente. Com frequências alélicas que, antes do fluxo eram diferentes, depois do fluxo se tornam homogêneas entre si, minimizando, desse modo, os efeitos da deriva genética (SLATKIN, 1985; ZUCCHI, 2002).

O fluxo gênico entre populações pode ocorrer sob diferentes modelos: (a) modelo de alpondras (*stepping stone*) uni-dimensional, no qual as populações são ordenadas ao longo de uma linha, cada uma recebendo migrantes apenas de populações vizinhas; (b) modelo de alpondras (*stepping stone*) bi-dimensional, uma extensão do modelo anterior para duas dimensões, consistindo de um arranjo

retangular onde cada população (exceto aquelas na borda) está conectada a quatro outras populações; (c) modelo de ilhas, em que o fluxo corre aleatoriamente entre todas as populações do grupo; (d) modelo de ilha hierárquico, uma extensão do modelo de ilhas clássico para dois níveis hierárquicos; (e) modelos espaciais, que Incluem modelos de isolamento por distância, no qual o fluxo ocorre localmente entre vizinhos, em uma população de distribuição uniforme; e (f) modelo continente-ilha, baseado no modelo de ilhas e que assume migração unidirecional de uma fonte grande continental com uma frequência alélica fixada para uma população menor isolada (BALLOUX, 2001; KIMURA; WEISSS, 1964).

2.6 POPULAÇÃO MÍNIMA VIÁVEL E TAMANHO EFETIVO POPULACIONAL

O estabelecimento de prioridades para a conservação de espécies e de objetivos quantitativos de manejo passam pela estimação dos riscos aleatórios de extinção enfrentados pelas populações isoladas. Os primeiros trabalhos sobre a estimativa de risco de extinção tiveram seu foco na Análise de Viabilidade Populacional (AVP) e outros métodos relacionados que visavam estimar o tamanho limiar, abaixo do qual os riscos de extinção seriam considerados extremamente altos. Foi, então, proposto o conceito de *População Mínima Viável* (PMV), que visa compreender a viabilidade (sobrevivência) em longo prazo de populações, relacionada com a probabilidade de extinção, sob a premissa de que populações maiores são mais aptas que aquelas menores para conter os efeitos estocásticos que ameaçam as populações em extinção. Assim, População Mínima Viável corresponde ao menor tamanho necessário para que uma população tenha predeterminada probabilidade de persistência para um dado período de tempo (KAGEYAMA; LEPSCH-CUNHA, 2001; FLATHER, 2011).

A população mínima viável é abordada em termos de T*amanho Efetivo Populacional,* que incorpora variáveis relacionadas à estrutura genética da população. O Tamanho Efetivo Populacional permite quantificar o efeito da deriva (tamanho efetivo de variância) e da endogamia (tamanho efetivo de endogamia), relacionados com a representatividade genética de amostras de indivíduos de uma população finita, sendo, então, um parâmetro importante na amostragem para coleta de sementes e na

delimitação de uma área mínima viável para conservação *in situ* de uma espécie (WRIGHT, 1931). A população mínima viável pode ser também, abordada em termos de Área Mínima Viável (AMV), parâmetro muito utilizado como a expressão do tamanho de área mínima de um fragmento florestal capaz de manter populações viáveis (SILVA et al., 2014).

2.7 SIMULAÇÃO DE DADOS GENÉTICOS

O comportamento futuro de uma dada espécie ou população pode ser previsto através da sua história evolutiva do passado ao presente. Muitos fatores que contribuem podem ser acessados através da biologia reprodutiva e de marcadores moleculares, haja vista que a história evolutiva é difícil de acessar diretamente (PENG et al., 2013).

Estudos sobre o sistema reprodutivo podem ser conduzidos para determinar, entre outros parâmetros, taxas de cruzamento, de autofecundação, a proporção de cruzamentos endogâmicos ou biparentais, a distância de dispersão de pólen, a taxa de fluxo gênico entre populações. Tais estudos podem ser eficientemente realizados mediante o uso de marcadores de DNA, em especial àqueles de herança codominante e altamente polimórficos, como os microssatélites, desde que se adote uma amostragem delineada adequadamente para atingir os objetivos do estudo (SEBBENN, 2006).

O procedimento envolve a utilização de modelos genéticos para descrever padrões de reprodução, como o modelo de reprodução e o modelo de cruzamentos correlacionados. Devido ao custo e disponibilidade de amostras genéticas, falta de conhecimento das variantes causais que contribuem para os fenótipos observados e intratabilidade matemática de modelos evolutivos complexos, simulações de computador têm sido amplamente utilizadas para, entre várias aplicações, prever resultados sob cenários genéticos reais, comparar e verificar os métodos ou ferramentas e estimar parâmetros de modelos evolutivos (PENG et al., 2013).

Com o avanço cada vez mais crescente na área da informática, novos métodos e programas de simulação vêm sendo desenvolvidos na área da biologia evolutiva para inferir com maior acurácia eventos passados a partir de dados atuais. Pela

natureza inerentemente desconhecida dos eventos passados, esses programas permitem desvendar as relações evolutivas entre caracteres moleculares simples, genes, genomas e espécies.

Inúmeros programas já foram desenvolvidos para realizar simulação de dados genéticos. Entre estes, o EASYPOP (BALLOUX, 2001) é um programa destinado a simular conjuntos de dados sob uma ampla gama de condições, permitindo explorar problemas muito complexos em genética de populações. Permite simular dados taxas e modelos de migração, taxas e modelos de mutação, sistema reprodutivo, entre outros parâmetros, com base em análises de marcadores moleculares. As entradas do programa estão limitadas aos parâmetros escolhidos pelo utilizador para qualquer simulação, incluindo parâmetros ecológicos e reprodutivos, ao passo que as saídas consistem de estatísticas de estruturação genética das populações tais como heterozigosidade observada e esperada e as estatísticas F. Isso permite, a partir de dados reais de estruturação genética obtidos mediante marcadores moleculares, realizar simulações visando encontrar o modelo ecológico e reprodutivo que melhor explica tais resultados. Esses arquivos podem ser analisados diretamente em diversos programas como o Fstat (GOUDET, 1995) e o Arlequin (EXCOFFIER et al., 2005).

O programa Fstat permite estimar diversidades gênicas e estatísticas de diferenciação a partir de marcadores genéticos codominantes ou de dados gerados em programas de simulação. Os seguintes parâmetros podem ser obtidos pelo programa: frequência alélica, número observado e esperado de genótipos, variabilidade genética por amostra e loco, número de alelos, riqueza alélica, estimadores de diversidade e genética e testes de equilíbrio de Hardy-Weinberg (GOUDET, 1995). Por seu turno, Arlequim fornece um amplo conjunto de testes estatísticos a fim de estudar características genéticas e demográficas de populações. O programa estima os índices de diversidade genética, número de alelos e frequências de haplótipos, testes de equilíbrio de ligação, neutralidade seletiva e equilíbrio demográfico, parâmetros de expansão populacional e análises de subdivisão populacional, por meio da Análise de Variância Molecular (AMOVA) (EXCOFFIER et al., 2005).

2.8 REFERÊNCIAS

BALLOUX, F. EASYPOP (version 1.7). A computer program for the simulation of population genetics. **Journal of Heredity**, v.92, n.3, p**.**301-302, 2001. Disponível em: <https://academic.oup.com/jhered/article/92/3/301/2187330>. Acessado: Mar. 18, 2019. doi: 10.1093/jhered/92.3.301.

BARRANDEGUY, M.E.; GARCIA, M.V. **Microsatellites as a tool for the study of microevolutionary process in native forest trees**. In: ABDURAKHMONOV, I. Y. Microsatellite markers. In TechOpen, 2016. 27p. DOI: 10.5772/65042. Disponível em: <https://www.intechopen.com/books/microsatellite-markers/microsatellites-as-a-tool-for-the-study-of-microevolutionary-process-in-native-forest-trees >. Acesso em: 14 mar. 2021.

BIANCHINI, M.B.; GIBBS, P.E. Late-acting self-incompatibility in *Capparis retusa* (Capparaceae), a species of Chaco woodland in NE Argentina. **Revista Brasileira de Botânica**, v.23, n.4, p.395-400, 2000.

CONSON, A.R. et al. Genetic structure of the Atlantic Rainforest tree species *Luehea divaricata* (Malvaceae). **Genetica**, v.141, n.4-6, 205-215, 2013. Disponível em: <https://doi.org/10.1007/s10709-013-9719-4>. Acessado: Nov. 20, 2019. doi: 10.1007/s10709-013-9719-4.

FLATHER, C.H. et al. Minimum viable populations: Is there a ‘magic number’ for conservation practitioners? **Trends in Ecology and Evolution,** v.26, n.6, p.307–316, 2011. Disponível em: < https://pubmed.ncbi.nlm.nih.gov/21458878>. Acessado: Nov. 21, 2019. doi: 10.1016/j.tree.2011.03.001.

KAGEYAMA, P.; LEPSCH-CUNHA, N.M. Singularidade da biodiversidade nos trópicos. In: GARAY, I.; DIAS, B. (eds*.)* **Conservação da Biodiversidade em ecossistemas tropicais: Avanços conceituais e revisão de novas metodologias de avaliação e monitoramento***.* São Paulo: Vozes, 2001, p.199-214.

KALIA, R. K.; RAI, M. K.; KALIA, S.; SINGH, R.; DHAWAN, A. K. Microsatellite markers: an overview of the recent progress in plants. **Euphytica**, v.177, n.3, p.309–334, 2010. doi:10.1007/s10681-010-0286-9.

KENTA, T.; SHIMIZU, K.K.; NAKAGAWA, M.; OKADA, K.; HAMID, A.A.; NAKASHIZUKA, T. Multiple factors contribute to outcrossing in a tropical emergent *Dipterocarpus tempehes*, including a new pollen-tube guidance mechanism for self-incompatibility. **American Journal of Botany***,* v.89, n.1, p.60–66, 2002.

KIMURA, M.; WEISSS, G.H. The stepping stone model of population structure and the decrease of genetic correlation with distance. **Genetics**, v.49, n.4, p.561-576, 1964.

LEITE, A.V.; MACHADO, I.C. Reproductive biology of woody species in Caatinga, a dry forest of northeastern Brazil. **Journal of Arid Environments**, v.74, n.11, p.1374e1380, 2010.

LIAO, T.; YUAN, D.-Y.; ZOU, F.; GAO, C.; YANG, Y.; ZHANG, L.; TAN, X.-F. Self-sterility in *Camellia oleifera* may be due to the prezygotic late-acting self-incompatibility. ***PLoS One****,* v.9, n.6, p. e99639, 2014.

LOWE, A. J. et al. Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. **Heredity**, v.95, n.4, p.255-273, 2005. Disponível em: <https://www.nature.com/articles/6800725>. Acessado: Out. 18, 2018. doi: 10.1038/sj.hdy.6800725.

MCGOWEN, M.H.; VAILLANCOURT, R.E.; PILBEAM, D.J.; POTTS, B.M. Sources of variation in self-incompatibility in the Australian forest tree, *Eucalyptus globulus*. **Annals of Botany**, v.105, n.5, p.737–745, 2010.

NEIGEL, J. E. A comparison of alternative strategies for estimating gene flow from genetic markers. **Annual Review of Ecology and Systematics**, v.28, n.1, p.105-128, 1997. Disponível em: <https://www.annualreviews.org/doi/full/10.1146/annurev.ecolsys.28.1.105>. Accessado: Mai. 4, 2019. doi: 10.1146/annurev.ecolsys.28.1.105.

PENG, B. et al. Genetic Simulation Resources: a website for the registration and discovery of genetic data simulators. **Bioinformatics,** v.29, n.8, p.1101-1102, 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624809/>. Accessado: Jun. 11, 2018. doi: 10.1093/bioinformatics/btt094.

QUESADA, M. et al. Long-term effects of habitat fragmentation on mating patterns and gene flow of a tropical dry forest tree, *Ceiba aesculifolia* (Malvaceae: *Bombacoideae*).

American Journal of Botany, v.100, n.6, p.1095-1101, 2013. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/23720432/>. Acessado: Nov. 12, 2018. doi: 10.3732/ajb.1200542.

SENAN, S.; KIZHAKAYIL, D.; SASIKUMAR, B.; SHEEJA, T.E. Methods for development of microsatellite markers: an overview. **Notulae Scientia Biologicae**, v.6, n.1, p.1-13, 2014.

3 CAPÍTULO I - DETERMINING THE POLYMORPHISM INFORMATION CONTENT OF A MOLECULAR MARKER

A Review Article

Serrote, C.M.L.S.; Reiniger, L.R.S.; Silva, K.B.; Rabaiolli, S.M.S.; Stefanel, C.M. Determining the Polymorphism Information Content of a molecular marker. **Gene,** v.726, p.144-175, 2020. https://doi.org/10.1016/j.gene.2019.144175.

Abstract

This review was carried out with the purpose of contributing to the discussion on the equations used in calculating the Polymorphism Information Content (PIC) of molecular markers. PIC measures the ability of a marker to detect polymorphisms, and therefore has enormous importance in selecting markers for genetic studies. We perform a summary analysis of PIC and its difference in relation to heterozygosity, another parameter used to evaluate the quality of a marker, presenting and discussing the several equations registered in the literature for both dominant and codominant markers. Finally, we present a brief direction on estimating PIC for dominant markers.
Keywords: Genetic studies; Marker quality; PIC calculation; Polymorphism.

Introduction

Molecular markers correspond to a class of genetic markers used to evaluate genetic differences between two or more individuals. They are capable of revealing the existing polymorphism in a set of genetically related individuals. The technology of molecular markers enables the genetic characterization of large numbers of genotypes through relatively simple and inexpensive procedures, and are fundamental in genetic improvement, phylogeography and analysis of similarity or genetic distance, identification of plant accesses, animals, microorganism isolates, and support in taxonomic studies.

The ability to reveal polymorphisms is fundamental in selecting molecular markers. If a marker is not able to detect the genetic differences existing in a set of individuals, its usefulness is not effective. In qualitative terms, a marker is considered polymorphic if it has at least two alleles and the most frequent allele has a frequency of up to 99% (Shete et al., 1999). Quantitatively, the polymorphism degree can be

measured by heterozygosity (Nei and Roychoudhury, 1974) and the Polymorphism Information Content (PIC) (Botstein et al., 1980). Other parameters include the Marker Index (product of the total number of loci per initiator times the fraction of polymorphic loci) and Resolution Power (parameter indicating discriminatory marker power). For example, efficiency of RAPD, ISSR, AFLP and ISTR markers for the detection of polymorphisms and genetic relationships in *Dioscorea* spp. was evaluated by Velasco-Ramírez et al. (2014). The PIC values were quite similar for all markers (≈0.48), heterozygosity was lower for ISTR (0.36), and higher for other markers (RAPD = 0.43; ISSR = 0.45 and AFLP = 0.47).

Inferences on the attributes of molecular markers contribute to reducing costs and improving the results, since the discrimination of the individuals can be carried out by a smaller number of primers, using only the most informative ones. They can also serve as screening for genetic studies of populations with the objectives to investigate evolutionary and ecological processes.

In the present review we focus our attention on Polymorphism Information Content, a widely used parameter in the literature to evaluate the discriminatory power of molecular markers and to study the genetic diversity in populations of several taxa. We discriminate the use of PIC for dominant and co-dominant markers. The literature presents several equations for calculating PIC, some of which are suitable for the class of marker used, whether dominant or codominant.

Thus, the present review aims to present a brief discussion about these equations, beginning with a distinction of this parameter with heterozygosity and ending with a brief direction on estimating PIC for dominant markers.

Heterozygosity and polymorphism information content

The heterozygosity of a marker is the likelihood of an individual being heterozygous at the marker site and depends on the number of alleles and their frequency in the population. The heterozygosity value ranges from zero (without heterozygosity) to 1.0 (high number of alleles with equal frequency). The following equations enable the heterozygosity of a marker to be calculated:

$$H = 1 - \sum_{i=1}^{k} p_i^2$$ (Nei, 1978) (1)

$$H = \prod_{i=1}^{k} 2p_i$$ (2)

k = number of alleles; pi = frequency of the i^{th} allele.

Equation 1 originates from that used to estimate the genetic diversity in a population taking into account the frequency of homozygous individuals (Nei, 1978), instead of considering the frequencies of the heterozygous individuals (Equation 2), which was elaborated by the authors of the present work with the objective to expand it for multiple alleles. Equation 1 was derived from Equation 1 by using the capital pi (Π) notation of productory. The heterozygosity determined by using Equation 1 and Equation 2 is called expected heterozygosity (He) or genetic diversity, because it is less sensitive to the sample size in relation to the observed heterozygosity (Ho), which is obtained by the ratio between the number of heterozygous individuals and the total number of individuals in the population (Chesnokov; Artemyeva, 2015).

Polymorphism Information Content (PIC) of a marker corresponds to its ability to detect the polymorphism among individuals of a population, and the higher that capacity, the greater its value. It is one of the indicators of marker quality in genetic studies. PIC values for co-dominant markers range from 0 (monomorphic) to 1 (very highly informative, with several alleles of equal frequency). According to Botstein et al. (1980), markers with PIC values greater than 0.5 are considered to be very informative, values between 0.25 and 0.50 are somewhat informative, and values lower than 0.25 are not very informative. For example, a marker that reveals six alleles but one allele is found to be in very high frequency, has a lower discriminatory capacity than another with six alleles but with similar frequencies (Smith et al., 1997). Therefore, markers with PIC values above 0.5 are more recommended to genetic studies while those below 0.25 are not recommended. There is no available classification for dominant markers, which PIC values range from 0 to 0.5. Low PIC values are associated with low or high loci frequencies. According to Grativol *et al.* (2010), loci with high frequency prevail among individuals which exhibit a tendency to monomorphism, resulting in low

PIC value. However, similarly less frequent loci also tend to monomorphism, since they present the prevalence of absent bands, also resulting in lower PIC values.

PIC calculation for co-dominant markers

Botstein (1980) developed the following equation for calculating PIC with the use of codominant markers (RFLP):

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} 2p_i^2 p_j^2 \quad \text{(Botstein, 1980)} \tag{3}$$

In which: n = number of alleles; p_i and p_j = allele frequency in population i and j, respectively.

Considering a single population, the PIC is determined by:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 - \left(\sum_{i=1}^{n} p_i^2\right)^2 + \sum_{i=1}^{n} p_i^4 \quad \text{(Botstein, 1980)} \tag{4}$$

Among the softwares developed for genetic analysis that help to determine the PIC, Genetix (Belkhir et al., 1999), Genes (Cruz, 2013) and Cervus (Kalinowski et al., 2007) may be highlighted in their various versions. In addition, this parameter can also be calculated online at the following site: https://www.liverpool.ac.uk/~kempsj/pic.html, which is suitable for allele numbers equal to or less than 10, being required to enter the frequency values of each band (allele). In its simplest form, the PIC for codominant markers is calculated in the same way as heterozygosity:

$$PIC = 1 - \sum_{i=1}^{n} p_i^2 \quad \text{(Botstein, 1980)} \tag{5}$$

This simplified equation is only applied when assuming inbred lineages (homozygotes), since the third and fourth terms cancel out when there is only a single allele. Despite some authors used equation 3 in their studies, such as Mei et al. (2012) to select highly informative SSR markers for screening barley genotypes, several authors have used this short equation. Landjeva et al. (2006) obtained PIC values

ranging from 0.10 to 0.80 for 21 microsatellites used to evaluate the genetic diversity among Bulgarian winter wheat (*Triticum aestivum* L.) varieties. PIC values ranging from 0.13 to 0.78 were obtained for a set of 30 polymorphic simple sequence repeats (SSR) loci by Muis et al. (2015) when assessing the genetic diversity of S3 maize genotypes resistant to downy mildew in Indonesia. Smith et al. (1997) used SSRs and restriction fragment length polymorphism (RFLP) loci to evaluate their utility as molecular markers in maize pedigree. PIC values ranged from 0.06 to 0.91 for SSRs and from 0.10 to 0.84 for RFLPs. Mean values for PIC for SSRs and RFLPs were similar, approximately 0.62. Likewise, Senior et al. (1998) used 70 SSRs for determining genetic similarities and relationships in maize in the United States. PIC values for these primers ranged from 0.17 to 0.92. The effectiveness of SSR loci were also tested in local species of *Citrullus colocynthis*, in Iran, in which seven were polymorphic and, therefore, were used to access the genetic diversity among the samples (Avval, 2017).

Expressed sequence tags (ESTs) represent a useful tool for gene and marker discovery, which are attractive for gene mapping, functional studies, genome annotation and comparative genomics. Thirty out of 46 developed EST-SSRs for olive trees were classified as informative markers (PIC > 0.5) and nine as suitable markers for gene mapping (PIC > 0.7). PIC values were in a range from 0.040 and 0.833, with a mean value 0.539 (Arbeiter et al., 2017).

Another class of co-dominant markers, largely used before the discovery of DNA markers, are the izoenzymes. Despite they have high reproduciblity, the level of detectable variation due to changes in DNA sequence is reduced as they reveal polymorphism at the protein level. Three proteins used in soybean (*Glycine max* L.) yielded PIC values from 0 to 0.223 (Jain et al., 2017). In another study, Nagy; Marton (2006) obtained PIC values ranging from 0.04 to 0.55 (mean 0.27) in maize lines.

In all these studies a diversity of PIC values was obtained from values close to zero to values close to one, which enables the selection of higher polymorphic markers for further genetic studies. This equation is more frequently used because of its simplicity in relation to the full equation (4). However, equation (5) determines heterozygosity, another measure of polymorphism, and overestimates PIC values. In addition, the simplification of the equation suggests, as already mentioned, to be inbred

lines, which is not the case in most cases. Therefore, it is more advisable to use the full equation (4).

PIC calculation for dominant markers

For dominant markers, the PIC value indicates the probability of finding that marker in two different states (present/absent) in two randomly selected individuals in a population. Its value ranges from zero for monomorphic markers to 0.5 for markers present in 50% of individuals and absent in the remaining 50%. Because these markers have only two possible alleles, the PIC definition is similar to heterozygosity. We propose the following classification on the informativeness for dominant markers, based on PIC values: low (0 to 0.10), medium (0.10 to 0.25), high (0.30 to 0.40) and very high (0.40 to 0.50). The following equations are commonly used for determining PIC for dominant markers:

$PIC = 2f\ (1 - f)$ (ROLDAN-RUIZ *et al.*, 2000)

(6)

In which: f = the frequency of present bands in the developing gel; and 1-f = frequency of absent bands.

$PIC = 1 - [f^2 + (1 - f)^2]$ (De Riek et al., 2001)

(7)

In which: *f* is the frequency of the marker in the data set.

In equation 7, *f* can be the frequency of amplified loci or the frequency of unamplified loci. Several authors have used these equations in their studies with dominant markers. AFLP combinations used in molecular characterization of common been landraces in Mexico by Langarica et al. (2014) demonstrated an average PIC value of 0.25. Guo et al. (2014) used moderately informative (P = 0.358) markers to distinguish whipgrass (*Hemarthria compressa* L.) clones assess the genetic diversity and population structure among those genotypes. ISSR markers used to characterize the genetic diversity in a population of *Schizolobium amazonicum*, in the State of

Espírito Santo – Brazil, by Silva Júnior et al. (2017), were also moderately informative (PIC = 0.37). Aiming to characterize trees, fruits and the genetic diversity of natural populations of mangaba in Sergipe - Brazil, Silva et al. (2017) used ISSR markers that which PIC values averaged 0.26. Panyanitikoon et al. (2018) assessed the genetic variation in cucumber (*Cucumis sativus* L.) germplasm using RAPD markers with an average PIC value of 0.27.

Single nucleotide polymorphisms (SNPs), have been used in genetic studies. In spite of being co-dominant, SNPs are biallelic markers (Vignal et al., 2017) and, therefore, PIC for these markers is calculated as for dominant markers and the maximum value is 0.5 (Singh et al., 2013). SNP markers used in Estimation of Genetic Diversity and Population Structure of Indian Rice Varieties by Singh et al. (2013) ranged from 0.03 to 0.37 with an average PIC value 0.23. García et al. (2018) found informative SNPs (PIC = 0.30) to investigate genetic variation in a juniper species (*Juniperus phoenicea* ssp. *turbinata*). Chen et al. (2011) found PIC values from 0.01 to 0.375 with an average of 0.285 in a collection of rice varieties (*Oryza sativa* L.) by using SNPs.

There is also an online software to determine the PIC for dominant markers, which was developed by Abuzayed et al. (2017), and is available at: http://plantmolgen.iyte.edu.tr/GDdom/.

A simplified equation to estimate the PIC for dominant markers

In dominant markers we elaborate a binary matrix of presence and absence of bands as a result of the polymorphism revealed by the marker. If we consider the frequency of bands present as *p*, and the frequency of absents as *q*, we can establish the following general equation to estimate the Polymorphism Information Content (PIC):

$$PIC = 1 - (p^2 + q^2)$$

(8)

This equation considers polymorphism based on the frequencies of homozygous individuals for the Hardy-Weinberg equilibrium, as opposed to equation (6) which takes into account the frequency of heterozygous individuals. It is a simplified equation for quick calculation of PIC, derived from the ones cited in this review.

The following (Table 2) is an example of using the equation from a binary matrix of presence and absence of bands in electrophoresis gel (where 1 corresponds to the presence of the band, and 0 to the absence of the band) for 10 analyzed genotypes (G1 to G10).

Table 2: A sample of present (1) and absent (0) bands matrix and the PIC calculation for a dominant marker.

Band No.	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	p	q	$1 – (p^2 + q^2)$
1	1	0	0	1	1	1	0	1	1	1	0.7	0.3	0.42
2	1	1	1	1	0	0	1	1	0	0	0.6	0.4	0.48
3	0	0	1	1	1	1	1	0	1	1	0.7	0.3	0.42
4	0	0	0	0	0	1	0	0	0	1	0.2	0.8	0.32
5	1	1	1	1	1	1	0	1	1	0	0.8	0.2	0.32
6	0	1	1	1	1	1	1	1	1	0	0.8	0.2	0.32
7	1	1	1	1	1	1	0	0	1	1	0.8	0.2	0.32
8	1	0	0	0	0	1	1	1	0	1	0.5	0.5	0.5
9	1	0	1	1	1	1	0	1	1	1	0.8	0.2	0.32
10	1	1	1	0	1	1	0	0	0	1	0.6	0.4	0.48
												PIC	0.39

The obtained result indicates a high polymorphism revealed by the marker, according to the PIC scale for dominant markers proposed in the present review. A simulation of PIC values as a function of the *p* and *q* values between zero and one are presented below in Table 3 and Figure 1. Note that the maximum PIC value is 0.5, which is reached when the frequency of present bands is equal to the frequency of absent bands. At the other extreme, when no band is produced or if all genotypes present all bands, the marker is absolutely unable to reveal polymorphisms in the set of evaluated genotypes.

Table 3: Simulation of PIC values as a function of the frequencies of present and absent bands in a developing gel generated by Polymerase Chain Reaction (PCR).

p	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
q	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
PIC	0	0.18	0.32	0.42	0.48	0.5	0.48	0.42	0.32	0.18	0

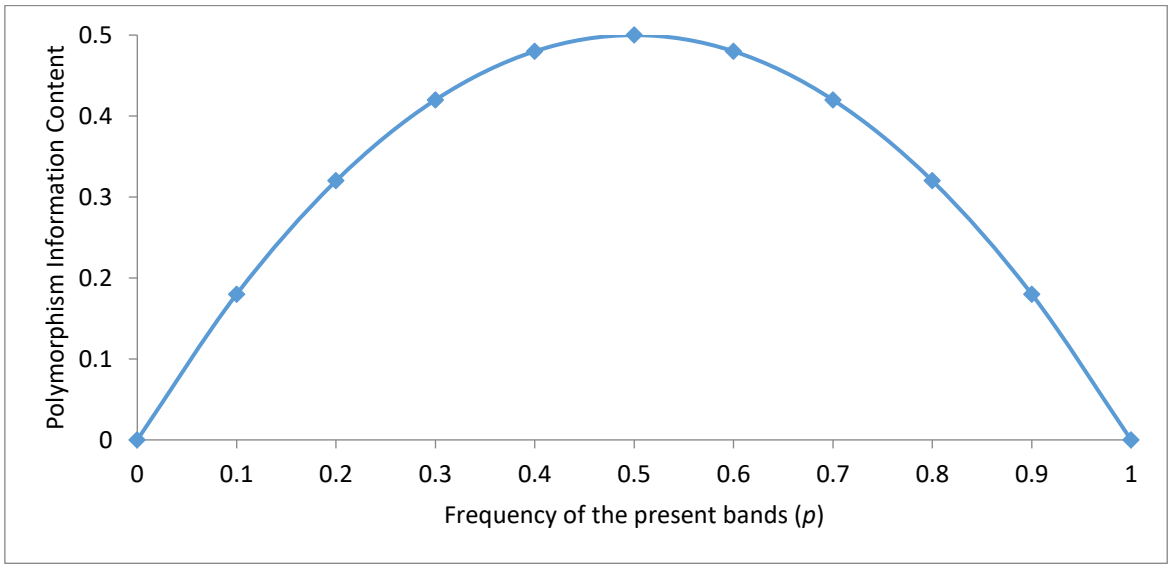


Figure 1: Graphical representation of the simulation of PIC values as a function of the frequencies of present and absent bands in a PCR-generated fragment gel.

Taking into account the previously referenced definition of polymorphism (Shete et al., 1999), the high PIC values are obtained when the frequency of bands (present or absent) is around 0.5. As we move left or right of the graphic (when the difference between the frequencies of band presence and absence increases), PIC values reduces.

Final remarks

There are several equations to estimate Polymorphism Information Content (PIC) and which are similar to heterozygosity, another measure used to assess the quality of a marker. The equation choice depends on the type of marker used (dominant or codominant). A simplified equation to estimate the PIC for dominant markers was presented herein, which can be used in genetic analyzes with this class of markers.

References

Abuzayed, M., N. El-Dabba, A. Frary, and S. Doganlar. 2017. GDdom: An online tool for calculation of dominant marker gene diversity. *Biochemical Genetics* 55:155-157. doi:10.1007/s10528-016-9779-0.

Anderson J.A., Churchill G.A., Autrique J.E., Tanksley S.D.and Sorrells M.E. 1993. Optimising parental selection forgenetic linkage maps. *Genome* 36: 181–186. doi.org/10.1139/g93-024.

Arbeiter, A.B.; Hladnik, M. Jakše, J.; Bandelj, D. 2017. Identification and validation of novel EST-SSR markers in olives. *Scientia Agricola* (Piracicaba, Braz.) 74(3): 215-225. doi.10.1590/1678-992x-2016-0111.

Avval. S.E. 2017. Assessing Polymorphism Information Content (PIC) using SSR molecular markers on local species of *Citrullus colocynthis*. Case Study: Iran, Sistan-Balouchestan Province. *Journal of Molecular Biology Research*; 7(1):42-47. doi:10.5539/jmbr.v7n1p42.

Belkhir, K. Genetix: Logiciel sous WindowsTM pour la génétique des populations. Laboratoire Génome, Populations, Interactions. CNRS UPR 9060, 1999.

Botstein, D., White, R. L., Skalnick, M. H., and Davies, R. W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphism. *American Journal of Human Genetics* 32: 314-331.

Chen H, He H, Zou Y, Chen W, Yu R, et al. (2011) Development and application of a set of breeder friendly SNP markers for genetic analyses and molecular breeding of rice (*Oryza sativa* L.). *Theoretical Applied Genetics* 123: 869-879, doi: 10.1007/s00122-011-1633-5.

Chesnokov, Y.V.; Artemyeva, A. 2015. Evaluation of the measure of polymorphism information of genetic diversity. *Agricultural Biology* 50(5):571–578. doi.org/10.15389/agrobiology.2015.5.571eng

Cruz, C.D. 2013. GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum* 35:271-276. http://dx.doi.10.4025/actasciagron.v35i3.21251

De Riek, J.; Calsyn, E.; Everaert, I.; Van, Bockstaele, E.; De Loose, M. AFLP. 2001. AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theoretical and Applied Genetics 103: 1254-1265. doi: 10.1007/s001220100710.

García, C.; Guichoux, E.; Hampe, A. 2018. A comparative analysis between SNPs and SSRs to investigate genetic variation in a juniper species (*Juniperus phoenicea*

ssp. *turbinata*). *Tree Genetics & Genomes*, 14: 87 https://doi.org/10.1007/s11295-018-1301-x

Grativol, C.; Lira-Medeiros, C.F.; Hemerly, A.S.; Ferreira, P.C.G. 2010. High efficiency and reliability of inter-simple sequence repeats (ISSR) markers for evaluation of genetic diversity in Brazilian cultivated *Jatropha curcas* L. accessions. *Molecular Biology Reports* 10:547-557. doi: 10.1007/s11033-010-0547-7.

Guo, Z-H.; Fu, K-X.; Zhang, X-Q.; Bai, S-Q.; Fan, Y.; Peng, Y.; Huang, L-K.; Yan, Y-H.; Liu, W.; Ma, X. 2014. Molecular Insights into the Genetic Diversity of *Hemarthria compressa* Germplasm Collections Native to Southwest China. *Molecules*, 19, 21541-21559. doi:10.3390/molecules191221541

Jain, R.K.; Joshi, A.; Singh, M.; Sen, Y.; Kaur, G. 2017. Isoenzyme diversity and phylogenetic relationship among soybean (*Glycine max* L.) genotypes. *Journal of Pharmacognosy and Phytochemistry,* 6 (5): 1260-1265.

Kalinowski S.T, Taper M.L, Marshal T.C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099–1106. doi: 10.1111/j.1365-294X.2007.03089.x

Landjeva, S.; Korzun, V.; Ganeva, G. Evaluation of Genetic Diversity Among Bulgarian Winter Wheat (*Triticum aestivum* L.) Varieties During the Period 1925–2003. 2006. Using microsatellites. genetic resources and crop evolution (2006) 53:1605–1614. Doi: 10.1007/s10722-005-8718-4.

Langarica, H.R.G., Serna, R.R.; Delgado, S.H.; Perez, N.M. 2014. Morphological and molecular characterization of common bean landraces cultivated in the semi-arid Mexican high plateau. *Boletin de la Sociedad Argentina de Botanica* 49: 525-540.

Mei, L.; Ping, J.; Wang, D.; Zhang, Z.; Luo, S., Yang, M.; Qian, G. 2012. Malt genotypic screening of Polymorphism Information Content (PIC) of SSR markers based on physiological traits in barley. Molecular Biology 1:101. doi:10.4172/2168-9547.1000101.

Muis, A.; Nonci, N.; Pabendon, M.B. 2015. Genetic diversity of S3 maize genotypes resistant to downy mildew based on SSR markers. *Indonesian Journal of Agricultural Science* 16: 79-86. http://dx.doi.org/10.21082/ijas.v16n2.2015.pp.79-86

Nagy, E., Marton, L. 2006. The use of morphological traits and genetic markers to estimate genetic relationships in maize. *Cereal Research Communications,* 34 (2-3): 887-894. doi.10.1556/CRC.34.2006.2-3.216.

Nei, M., and Roychoudhury, A. K. 1974. Sampling variances of heterozygosity and genetic distance. *Genetics* 76: 379-390.

Nei, M. 1978. Molecular Evolutionary Genetics. Columbia University Press, New York, 512 pp.

Ott, J. 1992. Strategies for characterizing highly polymorphic markers in human gene mapping. *American Journal of Human Genetics* 51: 283-290.

Panyanitikoon, H.; Khanobdee, C.; Jantasuriyarata, C.; Samipaka, S. 2018. Genetic variation in cucumber (*Cucumis sativus* L.) germplasm assessed using random amplified polymorphic DNA markers. *Agriculture and Natural Resources* 52(5): 497-502.

Prevost, A.; Wilkinson, M.J. 1999. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theoretical and Applied Genetics* 98: 107-112. doi:10.1007/s001220051046.

Roldan-Ruiz, I., J. Dendauw, E. Van Bockstaele, A. Depicker, and M. De Loose, 2000. AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). *Molecular Breeding* 6: 125-134. doi: 10.1023/A:1009680614564.

Senior, M.L., Murphy, J.P., Goodman, M.M. and Stuber, C.W. 1998. Utility of SSRs for determining genetic similarities and relationships in maize using an agarose gel system. *Crop Science* 38: 1088-1098. doi: 10.2135/cropsci1998.0011183X003800040034x.

Shete, S.; Tiwari, H.; Elston, R.C. 2000. On estimating the heterozygosity and polymorphism information content value. Theoretical Population Biology 57: 265-271. doi:10.1006/tpbi.2000.1452.

Silva, A.V.C.; Amorim, J.A.E.; Vitória, M.F.; Ledo, A.S.; Rabbani, A.R.C. 2007. Characterization of trees, fruits and genetic diversity in natural populations of mangaba. Ciência e Agrotecnologia, 41 (3):255-262. doi:10.1590/1413-70542017413048416.

Silva Júnior, A.L.; Souza, L.C.; Pereira, A.G.; Caldeira, M.V.W.; Miranda, F.D. 2017. Genetic diversity of *Schizolobium parahyba* var. *amazonicum* (Huber ex. Ducke) Barneby, in a forest area in Brazil. *Genetics and Molecular Research* 16:1-10, doi:10.4238/gmr16039774.

Singh, N.; Choudhury, D.R.; Singh, A.K.; Kumar, S.; Srinivasan, K.; Tyagi, R.K.; Singh, N.K.; Singh, R. 2013. Comparison of SSR and SNP markers in estimation of genetic diversity and population structure of Indian rice varieties. *PLoS One* 8(12): e84136, doi:10.1371/journal.pone.0084136.

Smith, J.S.C.; Chin, E.C.L. Shu, H.; Smith, O.S.; Wall, S.J.; Senior, M.L.; Mitchell, S.E.; Kresovich, S.; Ziegle, J. 1997. An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.) - Comparisons with data from RFLPs and pedigree. *Theoretical and Applied Genetics* 95: 163-173. doi:10.1007/s001220050544

Varshney, R. K.; Chabane, K.; Hendre, P. S.; Aggarwal, R. K.; Graner, A. 2007. Comparative assessment of EST-SSR, EST-SNP and AFLP markers for evaluation of genetic diversity and conservation of genetic resources using wild, cultivated and elite barleys. *Plant Science* 173: 638-649. https://doi.org/10.1016/j.plantsci.2007.08.010.

Velasco-Ramírez, A.P.; Torres-Morán, M.I.; Molina-Moret, S.; Sánchez-González, J.J.; Santacruz-Ruvalcaba, F. 2014. Efficiency of RAPD, ISSR, AFLP and ISTR markers for the detection of polymorphisms and genetic relationships in camote de cerro (*Dioscorea* spp.). *Electronic Journal of Biotechnology* 17: 65-71. doi:10.1016/j.ejbt.2014.01.002.

Vignal, A.; Milan, D.; Sancristobal, M.; Eggen, A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution,* 34: 275–305. doi: 10.1051/gse:2002009.

4 CAPÍTULO II - MICROSATELLITES ARE IMPORTANT FOR FOREST GENETIC RESOURCES CONSERVATION IN BRAZILIAN BIOMES

Review

(Artigo aceito na revista *Acta Botanica Brasilica*)

ABSTRACT

Microsatellites are short sequences repeats that make up the genomes of eukaryotes and prokaryotes. They are of great importance as DNA markers for studies in several fields of genetics. In the present review, we searched for studies published in the five years period of 2017 to 2021 regarding the use of microsatellites in studies with forest tree species from the Brazilian biomes, in order to examine the importance of these markers for forest resources conservation. We searched scientific papers in journals indexed on the Scopus and Web of Science databases. There were found 38 peer reviewed articles that used microsatellites in the Brazilian biomes. The Atlantic Forest was the biome with more studies (35.9%) and most of the studies were published in 2018 (34.2%). In addition, most of the studied species belonged to the Fabaceae family (34.2%). The conclusions and recommendations made in these studies ratify the great contribution of microsatellite markers in the conservation of native forest species in Brazilian biomes.

Keywords: Brazilian ecosystems, genetic diversity, genetic structure, forest conservation, SSR markers

Introduction

Habitat fragmentation is one of the most issues of concern in conservation biology. Once large and continuous populations are split into smaller fragments, primarily by human disturbances such as land clearing and conversion, the genetic diversity is negatively affected

(Franklin *et al.* 2002). Genetic diversity is a key component for the sustainability of species as it enables communities to adapt to changing environments. For this reason, efforts in forest conservation include genetic tools to analyze the genetic diversity among individuals and populations (Jump *et al.* 2009).

The genetic diversity and its distribution among groups can be quantified through the use of genetic markers, such as Simple Sequence Repeats – SSR. This class of markers contributes to increase efficiency in genetic studies as they are neutral, can be accessed regardless of the plant's development stage and the environment, and do not compromise the viability of the specimens under study since small amounts of tissue are necessary, allowing additional analyzes to be performed (Garcia *et al.* 2004).

Since its development in the 1980s by Litt & Luty (1989), microsatellites or Simple Sequence Repeats (SSRs) have been used in genetic studies. These markers are abundant and broadly distributed in eukaryotic and prokaryotic genomes. Due to high rates of DNA replication error within microsatellites, the length of a microsatellite shows intra- and interspecific variation. For this reason, microsatellites are used for designing PCR-based markers for population genetic characterizations, genome mapping, tagging trait-associated genes during marker-assisted selection, among other applications (Wang *et al.* 2018). The use of SSRs is qualified by the information obtained, mainly because their co-dominant inheritance, allowing access to complete genetic information (Garrido-Cardenas *et al.* 2018).

The aim of this study was to analyze the state of the art of using microsatellite markers in scientific articles regarding genetic studies of natural forest populations in Brazilian biomes published from 2017 to 2021.

Material and methods

This article is a bibliographic review on the studies regarding the use of microsatellite markers in genetic analyzes of Brazilian forest ecosystems. It brings together manuscripts published in the last five years (2017 - 2021). The research was restricted to scientific articles in journals indexed on the Scopus and Web of Science databases. The access to these databases was performed through the Coordination for the Improvement of Higher Education Personnel (CAPES, Brazil) platform. The following keywords were used: "SSR", "microsatellites", “forest” and “Brazil", in association with the Boolean operator AND. Only publications with some application in forest tree species conservation were selected.

The obtained data were analyzed in order to get the dimension of the relevance of this class of DNA markers in studies for the forest resources conservation. The data were submitted to descriptive analysis, followed by a brief critical analysis of the results.

Results

There were found 38 peer reviewed scientific articles that used microsatellite markers in forest tree species from different Brazilian biomes from 2017 to 2021. The Atlantic Forest was the biome with more studies (n = 14; 35.9%), followed by Cerrado (n = 11; 28.2%) and Amazon (n = 10; 25.6), while Pantanal (n = 2; 5.1%) and Caatinga (n = 2; 5.1%) were the less studied biomes. No study was recorded in the Pampa biome (Table 1).

Table 1. List of studies published in the Brazilian biomes using microsatellite markers from 2017 to 2021.

Study number	Biome	Target species	Family	Authors and publication year
1	Atlantic Forest	*Cariniana estrellensis* and *Cariniana legalis*	Lecythidaceae	Souza et al. (2018b)
2	Atlantic Forest	*Anadenanthera colubrina* and *Anadenanthera peregrina*	Fabaceae	Feres et al. (2021)
3	Atlantic Forest	*Centrolobium tomentosum*	Fabaceae	Sujii et al. (2017)

4	Atlantic Forest	*Casearia sylvestris*	Salicaceae	Araujo et al. (2017)
5	Atlantic Forest	*Myroxylon peruiferum*	Fabaceae	Schwarcz et al. (2018)
6	Atlantic Forest	*Myroxylon peruiferum*	Fabaceae	Silvestre et al. (2018)
7	Atlantic Forest	*Rhizophora mangle*	Rhizophoraceae	Francisco et al. (2018)
8	Atlantic Forest	*Eschweilera ovata*	Lecythidaceae	Santos et al. (2019)
9	Atlantic Forest	*Cedrela fissilis*	Meliaceae	Gandara et al. (2019)
10	Atlantic Forest	*Eugenia involucrata*	Myrtaceae	Stefanel et al. (2021)
11	Atlantic Forest	*Luehea divericata*	Malvaceae	Silva et al. (2021)
12	Atlantic Forest	*Campomanesia xanthocarpa*	Myrtaceae	Petry et al. (2021)
13	Atlantic Forest	*Anadenanthera peregrina*	Fabaceae	Cortelete et al. (2021)
14	Atlantic Forest	*Schinus terebinthifolia*	Anacardiaceae	Velasques et al. (2021)
15	Cerrado	*Dipteryx alata*	Fabaceae	Berti et al. (2017)
16	Cerrado	*Dipteryx alata*	Fabaceae	Guimarães et al. (2019)
17	Cerrado	*Casearia grandiflora*	Salicaceae	Costa et al. (2017)
18	Cerrado	*GenipaAmericana*	Rubiaceae	Manoel et al. (2017)
19	Cerrado	*Eugenia dysenterica*	Myrtaceae	Boaventura-Novaes et al. (2018)
20	Cerrado	*Hymenaea stigonocarpa*	Fabaceae	Moraes et al. (2018)
21	Cerrado	*Qualea grandiflora*	Vochysiaceae	Potascheff et al. (2019)
22	Cerrado	*Dimorphandra wilsonii*	Fabaceae	Muniz et al. (2020)
23	Cerrado	*Hancornia speciosa*	Apocynaceae	Chaves et al. (2020)
24	Cerrado	*Astronium fraxinifolium*	Anacardiaceae	Manoel et al. (2021)
25	Amazon	*Bertholletia excelsa*	Lecythidaceae	Cabral et al. (2017)
26	Amazon	*Bertholletia excelsa*	Lecythidaceae	Giustina et al. (2017)
27	Amazon	*Bertholletia excelsa*	Lecythidaceae	Giustina et al. (2018)
28	Amazon	*Bertholletia excelsa*	Lecythidaceae	Martins et al. (2018)
29	Amazon	*Bertholletia excelsa*	Lecythidaceae	Vieira et al. (2019)
30	Amazon	*Bertholletia excelsa*	Lecythidaceae	Baldoni et al. (2020)
31	Amazon	*Genipa Americana*	Rubiaceae	Ruzza et al. (2018)
32	Amazon	*Theobroma speciosum*	Malvaceae	Dardengo et al. (2018)
33	Amazon	*Hymenaea courbaril*	Fabaceae	Rocha et al. (2019)
34	Amazon	*Caryocar villosum*	Caryocaraceae	Francisconi et al. (2021)
35	Caatinga	*Prosopis palida* and *Prosopis juliflo*	Fabaceae	Freitas et al. (2019)
36	Caatinga	*Spondias tuberosa*	Anacardiaceae	Santos et al. (2021b)
37	Pantanal	*Prosopis rubriflora* and *Prosopis ruscifolia*	Fabaceae	Alves et al. (2018a)

38	Pantanal	*Prosopis rubriflora*	Fabaceae	Alves et al. (2018b)

Twelve botanic families were represented in the studies. Most of the studied species belonged to the Fabaceae family (n = 13; 34.2%), followed by the Lecythidaceae family (n = 8; 21.1%). On the other hand, Caryocaraceae, Meliaceae, Apocynaceae, Rhizophoraceae and Vochysiaceae were represented by only one species (2.6%). The other studied families were Myrtaceae (n = 3; 5.3%), Anacardiaceae (n = 3; 5.3%), Salicaceae (n = 2; 2.6%), Rubiaceae (n = 2; 26%) and Malvaceae (n = 2; 2.6%). In addition, most of the studies were published in 2018 (n = 13; 34.2%), followed by 2021 (n = 8; 21.1%), 2017 (n = 7; 18.4%), 2019 (n = 7; 18.4%), while in 2020 least number of studies (n=3; 7.9%) were published .

Atlantic Forest biome

Most of the published studies with microsatellite markers were performed in the Atlantic Forest. There was a record of 14 studies (36.8%) in this biome. In a study that involved two species of the *Cariniana* genus, microsatellite markers were used to access the gene flow pattern in fragmented populations of *Cariniana estrellensis* and *C. legalis*. For both species, there were high levels of seed (38.5–61.5%) and pollen (80.1–100%) immigration. No self–fertilization was detected, but there was evidence of mating among related trees (8.9 - 12.5%). The effective size in most of the populations varied from 10 to 33 and these values are lower than suggested for short-term conservation (Ne < 70) (Souza *et al.* 2018b).

In another study that involved two species, microsatellite markers were used to study the mating system and gene flow for *Anadenanthera colubrina* and *A. peregrina*. The analyses revealed that *A. colubrina* is a mixed mating species (multilocus outcrossing rate = 0.619) while *A. peregrina* is a predominantly outcrossing species (multilocus outcrossing rate = 0.905). For

both species, high indices of biparental inbreeding were observed (0.159 and 0.216 respectively), resulting in low effective pollination neighborhood sizes (Feres *et al.* 2021).

Sujii *et al.* (2017) used nuclear and plastid microsatellite markers to assess genetic parameters of juvenile and adult individuals in two *Centrolobium tomentosum* restoration areas, one corresponding to a disturbed fragment and the other, a large and well-preserved protection area. The authors reported that the restoration program was successful as they observed high genetic diversity and low inbreeding in the restoration areas, which values were similar to the natural remnants, suggesting gene flow between those areas.

In characterizing two *Casearia sylvestris* populations, microsatellite markers revealed high allelic variation in both populations (number of alleles = 101 and 117; allelic richness = 12.5 and 14.4), despite what the authors considered high inbreeding (F_{IS} = 0.640 and 0.363). Due to low gene flow, the authors found significant genetic divergence between populations (F_{ST} = 0.103)(Araujo *et al*. 2017).

In genotyping *Myroxylon peruiferum* populations from reforested and remnant natural areas, Schwarcz *et al.* (2018) evaluated the potential of forest restoration for the production of high genetic diversity tree populations in previously deforested areas. Due to the intense gene flow, no significant differences were found between areas in terms of inbreeding (F_{IS} = 0.20) or genetic diversity (H_E = 0.31 - 0.43; allelic richness = 2.41 - 2.94).

Eight microsatellite loci used to study the reproductive system and genetic diversity in *Myroxylon peruiferum* revealed a mixed reproductive system in this species with evidence of biparental inbreeding at the rate of 0.118. Genetic diversity was low (allelic richness = 1.40 - 4.82; H_E = 0.29 - 0.52) and the effective sizes for seedlings were much lower (Ne = 27.54 - 34.86) to those recommended for short-term conservation ($Ne \geq 100$) (Silvestre *et al.* 2018).

In studying the genetic diversity and the reproductive system of adult individuals and seeds from a *Rhizophora mangle* population, four microsatellite loci yielded a fixation index of -0.222

and 0.030 for adults and seeds respectively, and a multilocus outcrossing rate (tm) of 0.921. The coancestry coefficient was 0.180, similar to the expected for half-sib progenies (0.125). Based on these results, the authors estimated that 62 adult trees are needed for seed collection for short-term conservation (Francisco *et al.* 2018).

The genetic variability and gene pool sharing analysis of *Eschweilera ovata* revealed that there was moderate genetic diversity, particularly in conservation units with full protection, and that there was gene pool sharing between the subpopulations, which might reflect the historical gene flow that occurred before forest fragmentation (Santos *et al*. 2019).

Through the use of nine microsatellite loci, Gandara *et al.* (2019) investigated the genetic structure and diversity of undisturbed and disturbed *Cedrela fissilis* fragments. Genetic diversity was higher within than among fragments, with observed and expected heterozygosities ranging from 0.48 to 0.63 and from 0.55 to 0.70, respectively. The fragments showed moderate genetic structure (F_{ST} = 0.10). Therefore, authors suggested protecting all fragments instead of single isolated fragments.

A set of microsatellites used to analyze the variability and genetic structure in *Eugenia involucrata* fragments revaled high levels of genetic variability (3.67 alleles per locus; H_O = 0.815; H_E = 0.625), most of which (93%) was distributed within the fragments (Stefanel *et al.* 2021).

Silva *et al*. (2021) studied the genetic variability of three *Luehea divaricata* natural fragments and observed high genetic variability, most of which (77%) distributed within fragments, high gene flow (Nm= 3.853) and low genetic differentiation (F_{ST} = 0.072).

The evaluation of the patterns of genetic diversity, fine-scale spatial genetic structure and historical gene flow in *Campomanesia xanthocarpa* fragments revealed that the fragments presented moderate to high levels of genetic diversity and there was observed the isolation by

adaptation pattern, which implied the need for maintenance of the current remnants to assure the conservation of the private alleles (Petry *et al.* 2021).

The diversity and genetic structure of *Anadenanthera peregrina* were used as strategies for *ex situ* conservation. From a planted population, 42 alleles were detected and negative values for F_{IS} were observed, indicating escape of inbreeding in the population. According to the authors, the findings revealed the importance of *ex situ* conservation of the evaluated genotypes, allowing future use of the population as a seed orchard (Cortelete *et al.* 2021).

Microsatellite markers were used in studies of genetic structure among *Schinus terebinthifolia* populations from different ecological groups. Genetic structure revealed differences among populations (37.72%) and significant fixation rates based on F_{ST} ($P < 0.001$). The patterns of distribution for the species did not follow the isolation by distance or similarity by environmental conditions. The most divergent genotype group was found at the ombrophilous forest, which evidences that conservation efforts should be undertaken to prevent losses of biodiversity in that area (Velasques *et al.* 2021).

Cerrado biome

Eleven studies (28.9%) used microsatellites to analyze the genetic diversity in forest tree species from Cerrado, one of which was performed also in the Pantanal biome. Two of these studies involved *Dipteryx alata.* The analysis of the genetic diversity of three natural populations of *Dipteryx alata* revealed that these populations presented moderate genetic diversity ($H_O = 0{,}618$; $H_E = 0{,}715$) which is fundamental for their survival along the generations (Berti *et al.* 2017). In another study, the genetic diversity of a *Dipteryx alata* progeny from a germplasm collection, revealed that the number of alleles was 50 and, due to the high effective population size (Ne = 96), the germplasm collection had sufficient representativeness for use as a base population for breeding programmes (Guimarães *et al.* 2019).

In studying the genetic structure of *Casearia grandiflora* in conserved and disturbed populations, there was observed moderate divergence between populations (F_{ST} = 0.14) and higher proportion of genetic diversity (85%) was distributed within populations, which were not structured. In addition, less urbanized populations had greater genetic diversity, confirming the effectiveness of protected areas in genetic diversity conservation (Costa *et al*. 2017).

Microsatellites used to investigate the impact of spatial isolation on pollen and seed flow in a *Genipa americana* population detected a minimum immigration of pollen (6%) and seeds at 4% and mating among relatives (20-40%), indicating genetic connectivity with other populations (Manoel *et al.* 2017).

The assessment of the pattern of phenotypic and molecular genetic divergence among natural subpopulations of *Eugenia dysenterica* suggested that the species has a spatial genetic structure which must be taken into account for managing its genetic resources for both conservation and breeding purposes (Boaventura-Novaes *et al.* 2018).

Microsatellite loci were used to investigate the pollen and seed dispersal and mating patterns in *Hymenaea stigonocarpa*. The species presented a mixed mating system, with variations in the outcrossing rate (0.53 - 1.0). Pollen and seed dispersal occurred over long distances (>8 km) and the dispersal patterns were isolated by distance. Selfing resulted in a higher inbreeding depression than mating among relatives (Moraes *et al.* 2018).

The analysis of reproductive success, pollen dispersal and mating system of *Qualea grandiflora* trees revealed that the mean pollen dispersal distance (524.7 m) and the effective number of pollen donors per mother-tree (Nep = 12.7) were higher than for roadside trees (60.9 m, Nep = 4.6). The results indicated that the spatial isolation of roadside trees decreased pollinator movements (Potascheff *et al*. 2019).

The genetic diversity evaluation of ten *Dimorphandra wilsonii* populations resulted in 4 to 13 alleles per locus, and heterozygosity values per locus ranged from 0.113 to 0.940 for H_O and from 0.219 to 0.796 for H_E (Muniz *et al.* 2020).

A study aiming to compare quantitative and molecular variation within and among botanical varieties and subpopulations of *Hancornia speciosa* revealed a low degree of divergence among the botanical varieties and significant structuring among the subpopulations within varieties. According to the authors, divergent selection shaped the genetic structure among the botanical varieties for some traits, while genetic drift and uniform selection influenced the variation among the subpopulations (Chaves *et al.* 2020).

Pollen and seed flow for *Astronium fraxinifolium*, investigated through parentage analysis and microsatellite loci, revealed that a large proportion of pollen (76.5%) and seeds (57%) immigrated from trees outside the sampled populations and the dispersion followed a pattern of isolation by distance (Manoel *et al*. 2021).

Amazon Forest biome

In this biome were recorded ten studies (26.3%) using microsatellite markers in forest tree species. *Bertholletia excelsa* was the most studied species with six studies. Cabral *et al.* (2017) assessed the genetic diversity of a *B. excelsa* population and observed high genetic diversity (H_O = 0.512; H_E = 0.491) and no inbreeding. The analysis of half-sib progenies from different *B. excelsa* trees, by Giustina *et al.* (2017), revealed greater genetic diversity between families than among progenies from the same family. In studying the mating system in a *B. excelsa* population, Giustina *et al.* (2018) observed that outcrossing rates varied between trees (0.49–1.0) and fruits (0.53–1.0), but seeds were predominantly produced by outcrossing (0.92). Martins *et al.* (2018) observed moderate genetic diversity and high seed-dispersal distances in *B. excelsa* populations. Studying two *B. excelsa* populations, Vieira *et al.* 2019 found 70 alleles,

H_O was 0.43 and H_E was 0.82. The analysis of the genetic diversity of *B. excelsa* revealed greater genetic diversity between populations than within populations, allelic variation ranged from four to nine alleles, and heterozygosity ranged from 0.32 to 0.80 (Baldoni *et al*. 2020).

The analysis of the genetic diversity and population structure of three *Genipa americana* populations revealed 17 alleles, the expected heterozygosity ranged from 0.35 to 0.67 and remained higher than the observed heterozygosity. The populations presented high inbreeding (F_{IS} = 0.40), probably because of fragmentation (Ruzza *et al.* 2018).

In studying the effects of fragmentation on the genetic structure of *Theobroma speciosum* populations, Dardengo *et al.* (2018) found that most of the genetic diversity was distributed within groups (83%), which means that no significant effect of fragmentation was observed. However, given the small number of reproductive individuals in the population, the authors warned that the process of continuous fragmentation might increase inbreeding and favor genetic drift, leading populations to inbreeding depression and diversity loss.

In another study, SSR markers used to analyzed the genetic diversity of five *Hymenaea courbaril* populations detected 10.29 alleles per locus, H_E and H_O averaged 0.85 and 0.29 per locus, respectively (Rocha *et al*. 2019).

The genetic diversity and structure analysis of *Caryocar villosum* revealed low inbreeding (F_{IS} = 0.127; F_{IT} = 0.173) and low differentiation between regions (F_{ST} = 0.06). Most of the variation (89%) was found to occur within regions (Francisconi *et al*. 2021).

Caatinga biome

In Caatinga, only two studies (5.3%) were found in our search. Freitas *et al.* (2019) found low levels of genetic diversity (two alleles per locus and H_E = 0.181) and inbreeding (F_{IS} = -0.007) for *Prosopis palida* and *Prosopis juliflora*, suggesting the presence of genetic bottleneck and probable events of founders.

The diversity and genetic structure analysis of *Spondias tuberosa* accessions resulted in similarity coefficients from 0.30 to 0.84, indicating the existence of divergence among the accessions which can be used to increase the germplasm bank genetic diversity of this species (Santos *et al*. 2021b).

Pantanal biome

Only two studies (5.3%) with microsatellites were recorded in Pantanal, one of which was performed also in Cerrado. Microsatellites employed by Alves *et al.* (2018a) in collections of *Prosopis rubriflora* and *Prosopis ruscifolia* from Cerrado and Pantanal resulted in similar levels of genetic diversity for both species (H_E = 0.59 and H_E = 0.60 respectively) and there was evidence of genetic bottleneck in 64% of *P. rubriflora* sampled area and in 36% of *P. rusciflia* sampled areas.

In a study of the reproductive system of *Prosopis rubriflora*, Alves *et al*. (2018b) found that the species is preferably allogamous and the obtained progeny was composed, predominantly by half-sibs (79%). Coancestry coefficients ranged from 0.158 to 0.162 and there were high levels of crossings due to several mechanisms that prevent selfing.

Discussion

The Fabaceae family was the most studied in the Brazilian biomes. This finding is supported by the fact that Fabaceae is among the richest families in most Brazilian ecosystems. According to Lima et al. (2015), there are 222 native genera and 2807 species. In the Caatinga, for example, this family constitutes about a third of the richness of the biome with 86 genera and 320 species (Córdula et al. 2014).

Regarding the evolution of the number of studies published within the analyzed period, there was a trend of increase until 2018. However, the studies decreased in 2019 and 2020,

coincidently, during the COVID-19 pandemic outbreak. This period was characterized by lower allocation of research resources and increased socio-political tension in Brazil. The syndemic theory by Merrill Singer could explain the negative effect of COVID-19 on researchers productivity (Singer 1996). In 2021, probably, due to the vaccination process and the return of many activities, the publications returned its growing.

Considering the importance of microsatellites and the growing need for genetic studies in Brazilian biomes, there were expected more studies for a five years period. However, forest species are low studied in Brazil because they are neglected. On the other hand, the need for prior knowledge of the species genome limits the use of this class of molecular markers in genetic analysis. Thus, the abundance of alternative molecular markers, which are more accessible, may have contributed to the less use of microsatellites in genetic analyzes in Brazilian biomes.

Despite these limitations, researches that use microsatellites are important due to the quality of the information accessed, mainly in terms of genomic coverage, codominance characteristics and heritability.

Most of the studies were performed in the Atlantic Forest biome, while Pampa and Pantanal are the biomes with less studies. In general, the genetic diversity decreased in all biomes, primarily due to anthropic activities, according to the authors. In addition, the use of microsatellites was helpful to propose proper alternatives for conservation. However, there was observed a lack of standardization of the genetic diversity statistics. For example, Berti *et al.* (2017) considered $H_O = 0.618$ and $H_E = 0.715$ as moderate genetic diversity, while Cabral *et al.* (2017) considered $H_O = 0.512$ and $H_E = 0.491$ high genetic diversity. The probable reasons for this discrepancy may be the life history traits of each species and the heterogeneity between ecosystems. For example, Pantanal has richer ecosystems than Caatinga and Pampa.

Most of the genetic diversity in all biomes is distributed within groups. This pattern is consistent with the predominance of allogamy in the plant kingdom (Bawa *et al.* 1985) and the gene exchange allows recombination, increasing the genetic diversity within groups.

In spite of fragmentation, the studies ratified the role of gene flow in connecting isolated populations, thus preventing the loss of genetic diversity in the Brazilian biomes. In fact, allowing gene flow among populations of a species is one successful alternative to reduce the negative effects associated to small populations such us inbreeding and genetic drift. It is supported by studies carried out to compare disturbed and undisturbed areas, in which disturbance did not affect genetic diversity when high levels of gene flow were observed (Sujii et al. 2017; Costa et al. 2017; Gandara et al. 2019). According to Hellberg *et al.* (2002), gene flow is essential in connecting reproductively isolated populations, thereby reducing genetic differentiation among them. The studies ratified the predominance of crossing in the tropical forest tree species (Sobierajski *et al*. 2006). Although most of the flowering plants in nature are hermaphrodite, many species developed mechanisms, such as self-incompatibility, to prevent selfing in order to allow gene exchange and avoid gene erosion (Bawa *et al* 1985; Sobierajski *et al*. 2006).

Final remarks

Our research highlighted the prominent contribution of microsatellites in genetic studies in the Brazilian forest biomes as well as for the genetic conservation. In general, the studies reinforce that human activities are reducing genetic diversity in the Brazilian biomes. In addition, the studies ratified the role of gene flow in connecting isolated populations, thus reducing the probability of species extinction. Thus, in order to slow down the loss of genetic

diversity, it is recommended to maintain a large number of individuals and allow connectivity among isolated fragments.

References

Aguiar BI, Gandara FB, Freitas MLM, Moraes MLT, Kageyama PY, Sebbenn AM. 2019. Validation of microsatellite loci for *Balfourodendron riedelianum* through analysis of Mendelian inheritance, genetic linkage, and genotypic linkage disequilibrium. Silvae Genetica 68: 81-84.

Alves FM et al. 2018a. Genetic structure of two *Prosopis species* in Chaco areas: A lack of allelic diversity diagnosis and insights into the allelic conservation of the affected species. Ecology and Evolution 8(13): 1-17.

Alves FM, Sartori ALB, Zucchi MI, Azevedo-Tozzi AMG, Tambarussi EV, Souza AP. 2018b. A high level of outcrossing in the vulnerable species *Prosopis rubriflora* in a Chaco remnant. Australian Journal of Botany 66(4): 360-368.

Araujo F et al. 2017. Genetic diversity of *Casearia sylvestris* populations in remnants of the Atlantic Forest. Genetics and Molecular Research 16: gmr16019105.

Balbino E, Martins G, Morais S, Almeida C. 2019. Genome survey and development of 18 microsatellite markers to assess genetic diversity in *Spondias tuberosa* Arruda Câmara (Anacardiaceae) and cross-amplification in congeneric species. Molecular Biology Reports 46: 3511-3517. DOI: 10.1007/s11033-019-04768-w

Baldoni AB et al. 2020. Genetic diversity of Brazil nut tree (*Bertholletia excelsa* Bonpl.) in southern Brazilian Amazon. Forest Ecology and Management 458: 117795. DOI: 10.1016/j.foreco.2019.117795

Bawa KS. 1974. Breeding systems of tree species of a lowland tropical community. Evolution 28(1): 95-92.

Bawa KS, Perry DR, Beach JH. 1985. Reproductive biology of tropical lowland rain forest trees. I. Sexual systems and incompatibility mechanisms. American Journal of Botany 72(32): 331-345.

Berti CLF et al. 2017. Diversidade genética de populações naturais de *Dipteryx alata* localizadas nos municípios de Brasilândia/MS, Indiara/GO e Itarumã/GO estimada por marcadores microssatélites. Cultura Agronômica 26(2): 203-216.

Boaventura-Novaes CRD, Novaes E, Mota EES, Telles MPC, Coelho ASG, Chaves LJ. 2018. Genetic drift and uniform selection shape evolution of most traits in *Eugenia dysenterica* DC. (Myrtaceae). Tree Genetics & Genomes 14:76. DOI: 10.1007/s11295-018-1289-2

Braun M, Esposito T, Huettel B, Harand AP. 2019. Development and characterization of eleven microsatellite loci for the tropical understory tree *Paypayrola blanchetiana* Tul. (Violaceae). Molecular Biology Reports 46: 2529-2532. DOI: 10.1007/s11033-019-04622-z

Cabral JC et al. 2017. Diversity and genetic structure of the native Brazil nut tree (*Bertholletia excelsa* Bonpl.) population. Genetics and Molecular Research 16 (3): gmr16039702. DOI: 10.4238/gmr16039702

Carvalho T, Garcia G, Carvalho CS, Palma-Silva C, Culot L. 2017. Development of microsatellite loci for *Cryptocarya mandioccana* Meisner (Lauraceae) and their genotyping success in different tissues. Hoehnea 44(4): 580-587. DOI: 10.1590/2236-8906-22/2017

Chakraborty R. 1993. Analysis of genetic structure of populations: meaning, methods, and implications. In: Majumder PP (ed.). Human population genetics. New York, Plenum Press. p. 189–206.

Chaves LJ, Ganga RMD, Guimarães RA, Caldeira AJR. 2020. Quantitative and molecular genetic variation among botanical varieties and subpopulations of *Hancornia speciosa* Gomes (Apocynaceae). Tree Genetics & Genomes 16(4). DOI: 10.1007/s11295-020-01444-0

Córdula E, Morim MP, Alves M. 2014. Morfologia de frutos e sementes de Fabaceae ocorrentes em uma área prioritária para a conservação da Caatinga em Pernambuco, Brasil. Rodriguésia 65 (2). DOI: 10.1590/S2175-78602014000200012.

Cornacini MR et al. 2021. Detection and application of novel SSR markers from transcriptome data for *Astronium fraxinifolium* Schott, a threatened Brazilian tree species. Molecular Biology Reports 48: 3165–3172. DOI: 10.1007/s11033-021-06338-5

Cortelete MA, Silva Júnior AL, Pereira MLS, Miranda FD, Caldeira MVW. 2021. Molecular characterization as strategy for *ex situ* conservation of *Anadenanthera peregrina* (L.) Speg. Scientia Forestalis 49(129): e3443. DOI: 10.18671/scifor.v49n129.23

Costa MF et al. 2017. Chloroplast diversity of *Casearia grandiflora* in the Cerrado of Piauí State. Genetics and Molecular Research 16(1): 1-9.

Dantas LG, Alencar L, Huettel B, Pedrosa-Harand A. 2019. Development of ten microsatellite markers for *Alibertia edulis* (Rubiaceae), a Brazilian savanna tree species. Molecular Biology Reports 46:4593–4597. DOI: 10.1007/s11033-019-04819-2

Dardengo JFE, Rossi AAB, Varella TL. 2018. The effects of fragmentation on the genetic structure of *Theobroma speciosum* (Malvaceae) populations in Mato Grosso, Brazil. Revista de Biología Tropical 66(1): 218-226.

Duarte JF, Carvalho SVA, Nazareno AG, Dos Santos RM, de Carvalho D. 2021. Development of microsatellite markers for *Calyptranthes clusiifolia* (Myrtaceae). Molecular Biology Reports 48:3027–3030. DOI: 10.1007/s11033-021-06280-6

Feres JM, Nazareno AG, Borges LM, Guidugli MC, Bonifacio-Anacleto F, Alzate-Marin AL. 2021. Depicting the mating system and patterns of contemporary pollen flow in trees of the genus *Anadenanthera* (Fabaceae). PeerJ 9:e10579. DOI: 10.7717/peerj.10579

Francisco, P. M.; Tambarussi, E. V.; Alves, F. M. Bajay, S.; Ciampi-Guillardi, M.; Souza, A. P. 2018. Genetic diversity and mating system of *Rhizophora mangle* L. (Rhizophoraceae) in Northern Brazil revealed by microsatellite analysis. Cerne 24(4): 295-302.

Francisconi AF et al. 2021. Genetic structure and diversity identify incipient domestication of Piquiá [*Caryocar villosum* (Aubl.) pers.] along the lower Tapajós River, Brazilian Amazonia. Genetic Resources and Crop Evolution 68(4): 1487-1501. DOI: 10.1007/s10722-020-01078-0

Franklin AB, Noon BR, George TL. 2002. What is habitat fragmentation? Studies in Avian Biology 25:20-29.

Freitas LS, Melo CAF, Gaiotto FA, Ronan X, Corrêa RX. 2019. SSR based genetic diversity analysis in diploid algaroba (*Prosopis* spp.) population. Journal of Agricultural Science 11(1): 179-190.

Gandara FB et al. 2019. The effects of habitat loss on genetic diversity and population structure of *Cedrela fissilis* Vell. Tropical Plant Biology 12: 282–292. DOI: 10.1007/s12042-019-09234-2.

Garcia AAF, Benchimol LL, Barbosa AMM, Geraldi IO, Souza Jr CL, Souza AP. 2004. Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. Genetics and Molecular Biology 27 (4): 579-588.

Garrido-Cardenas JA, Mesa-Valle C, Manzano-Agugliaro F. 2018. Trends in plant research using molecular markers. Planta 247: 543–557.DOI: 10.1007/s00425-017-2829-y

Giustina LD et al. 2017. Genetic diversity between and within half-sib families of Brazil nut tree (*Bertholletia excelsa* Bonpl.) originating from native forest of the Brazilian Amazon.

Genetics and Molecular Research 16 (4): gmr16039839. DOI: doi.org/10.4238/gmr16039839

Giustina LD et al. 2018. Hierarchical outcrossing among and within fruits in *Bertholletia excelsa* Bonpl. (Lecythidaceae) open-pollinated seeds. Genetics and Molecular Research 17 (1): gmr16039872. DOI: 10.4238/gmr16039872.

Góes BD, Ruas EA, Benício LM, Cassiano D, de Souza FP, Ruas PM. 2019. Development and characterization of microsatellite loci for *Campomanesia xanthocarpa* (Myrtaceae) and cross amplification in related species. Acta Scientiarum. Biological Sciences 41: e43454. DOI: DOI: 10.4025/actascibiolsci.v42i1.43454

Gonçalves AR, Chaves LJ, Telles MPC. 2019. Genetic variability and effective population size in *Hymenaea stigonocarpa* (Fabaceae) germplasm collection: tools for breeding programs and genetic conservation. Genetica 147: 359-368. DOI: 10.1007/s10709-019-00076-0 2

Guimarães RA, Miranda KMC, Mota EES, Chaves LJ, Telles MPC, Soares TN. 2019. Assessing genetic diversity and population structure in a *Dipteryx alata* germplasm collection utilizing microsatellite markers. Crop Breeding and Applied Biotechnology 19(3): 329-336.

Hellberg ME, Burton RS, Neigel J, Palumbi SR. 2002. Genetic assessment of connectivity among marine populations. Bulletin of Marine Science 70(1): 273-290.

Jump, A.S.; Marchant, R.A.; Penuelas, J. 2009. Environmental change and the option value of genetic diversity. Trends in Plant Science 14(1):51-58.

Lemos RPM et al. 2018. Characterization of plastidial and nuclear SSR markers for understanding invasion histories and genetic diversity of *Schinus molle* L. Biology 7 (43): 1-10. DOI: 10.3390/biology7030043

Lima HC et al. 2015. Fabaceae. In: Lista de Espécies da Flora do Brasil. Jardim Botânico do Rio de Janeiro. http://floradobrasil.jbrj.gov.br/jabot/floradobrasil/FB115. 20 Jun. 2022.

Litt M, Luty JA. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. The American Journal of Human Genetics 44: 397–401.

Manoel RO et al. 2021. Landscape barriers to pollen and seed flow in the dioecious tropical tree *Astronium fraxinifolium* in Brazilian savannah. PLoS ONE 16(8): e0255275. DOI: 10.1371/journal. pone.0255275

Manoel RO et al. 2017. Low levels of pollen and seed flow in a riparian forest fragment of the dioecious tropical tree *Genipa americana* L. Forestry Research and Engineering: International Journal 1(1): 00003. DOI: 10.15406/freij.2017.01.00003

Martins K, Santos RSO, Campos T, Wadt LHO. 2018. Pollen and seed dispersal of Brazil nut trees in the southwestern Brazilian Amazon. Acta Amazonica 48: 217-223.

Messias PA, Alves FM, Pinheiro F, Souza AP, Koch I. 2020. Development and transferability of microsatellite markers for a complex of *Aspidosperma* Mart. & Zucc. (Apocynaceae) species from South American seasonally dry tropical forests. Brazilian Journal of Botany 43:139-145. DOI: 10.1007/s40415-020-00586-8

Moraes MA et al. 2018. Long-distance pollen and seed dispersal and inbreeding depression in *Hymenaea stigonocarpa* (Fabaceae: Caesalpinioideae) in the Brazilian savannah. Ecology and Evolution 8:7800-7816. DOI: 10.1002/ece3.4253

Muniz AC et al. The protected tree *Dimorphandra wilsonii* (Fabaceae) is a population of inter-specific hybrids: recommendations for conservation in the Brazilian Cerrado/Atlantic Forest ecotone. Annals of Botany 126: 191-203, 2020. DOI: 10.1093/aob/mcaa066

Petry VS, Stefenon VM, Machado LO, da Costa NCF, Klabunde GHF, Nodari RO. 2021. Patterns of genetic diversity, spatial genetic structure and gene flow in *Campomanesia xanthocarpa*: insights from SSR markers of different genomic origins. Anais da Academia Brasileira de Ciências 93(Suppl. 3): e20210134. DOI 10.1590/0001-3765202120210134

Potascheff CM et al. 2019. Stepping stones or stone dead? Fecundity, pollen dispersal and mating patterns of roadside *Qualea grandiflora* Mart. trees. Conservation Genetics 20: 1355-1367. DOI: 10.1007/s10592-019-01217-w

Ramanatha Rao V, Hodgkin, T. 2002. Genetic diversity and conservation and utilization of plant genetic resources. Plant Cell Tissue Organ Culture 68: 1-19.

Rocha LF, Paula NR, Nazareno AG, Carvalho D. 2020. Development and characterization of nuclear microsatellite markers for *Eremanthus erythropappus* and their transferability across related species. Biological Research 53(30):1-5.

Rocha VD et al. 2019. Genetic diversity of *Hymenaea courbaril* L. in the Mato Grosso Amazon: implications for conservation. Floresta 49(4): 745-754.

Ruzza DAC et al. 2018. The genetic diversity and population structure of *Genipa Americana* (Rubiaceae) in Northern Mato Grosso, Brazil. Genetics and Molecular Research 17 (4): gmr18017. DOI: 10.4238/gmr18017

Saiki FA et al. 2017. Development and validation of the first SSR markers for *Mimosa scabrella* Benth. Genetics and Molecular Research 16 (1): gmr16019571.

Santos AS et al. 2019. Gene pool sharing and genetic bottleneck effects in subpopulations of *Eschweilera ovata* (Cambess.) Mart. ex Miers (Lecythidaceae) in the Atlantic Forest of southern Bahia, Brazil. Genetics and Molecular Biology 42 (3): 655-665. DOI: 10.1590/1678-4685-GMB-2018-0140

Santos V, Santos CAF, de Oliveira VR, Costa AES, da Silva FFS. 2021b. Diversity and genetic structure of *Spondias tuberosa* (Anacardiaceae) accessions based on microsatellite loci. Revista de Biología Tropical 69(2): 640-648.

Santos VN, Gama RNCS, Santos CAF. 2021a. Development and transferability of microsatellite loci for *Spondias tuberosa* (Anacardiaceae: Sapindales), a species endemic

to the Brazilian semi-arid region. Genetics and Molecular Research 20 (2): gmr18778. DOI: 10.4238/gmr18778

Sarzi DS, Justolin B, Silva C, Lemos RPM, Stefenon VM. 2019. Discovery and characterization of SSR markers in *Eugenia uniflora* L. (Myrtaceae) using low coverage genome sequencing. Anais da Academia Brasileira de Ciências 91: e20180420. DOI: 10.1590/0001-3765201920180420

Schwarcz KD et al. 2018. Shelter from the storm: Restored populations of the neotropical tree *Myroxylon peruiferum* are as genetically diverse as those from conserved remnants. Forest Ecology and Management 410: 95-103.

Sebbenn, A.M. 2002. Número de árvores matrizes e conceitos genéticos na coleta de sementes para reflorestamentos com espécies nativas. Revista do Instituto Florestal 14(2): 115-132.

Selkoe KA, Toonen RJ. 2006. Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters 9(5): 615-629.

Silva et al. 2021. Variabilidade genética de fragmentos naturais de *Luehea divaricata* Mart. & Zucc. no bioma Mata Atlântica. Biodiversidade Brasileira 11(4): 4-11.

Silvestre EA et al. 2018. Mating system and effective population size of the overexploited Neotropical tree (*Myroxylon peruiferum* L.f.) and their impact on seedling production. Journal of Heredity 109(3):264-271.

Singer M. 1996. A dose of drugs, a touch of violence, a case of AIDS: conceptualizing the SAVA syndemic. Free Inquiry in Creative Sociology 24(2):99-110.

Sobierajski GR, Kageyama PY, Sebbenn AM. 2006. Sistema de reprodução em nove populações de *Mimosa scabrella* Bentham (Leguminosaceae). Scientia Forestalis 71: 37-49.

Souza DCL, Rossini BC, de Souza FB, · Sebbenn AM, Marino CL, de Moraes MKT. 2018a. Development of microsatellite markers for *Myracrodruon urundeuva* (F.F. & M.F.

Allemão), a highly endangered species from tropical forest based on next-generation sequencing. Molecular Biology Reports 45(1): 71-75. DOI:10.1007/s11033-017-4142-z

Souza et al. 2018b. Historic pollen and seed dispersal in fragmented populations of the two largest trees of the Atlantic Forest. Forestry Research and Engineering: International Journal 2(2): 100-109. DOI: 10.15406/freij.2018.02.00033

Stefanel CM, Reiniger LRS, Serrote CML, Stefenon VM, Lemos RPM. Variability and genetic structure in fragments of *Eugenia involucrata* De Candolle established through microsatellite markers. Ciência Rural 51(5): e20200008, 2021. DOI: 10.1590/0103-8478cr20200008

Sujii PS et al. 2017. Recovery of genetic diversity levels of a Neotropical tree in Atlantic Forest restoration plantations. Biological Conservation 211: 110-116.

Velasques J et al. 2021. Genetic and chemodiversity in native populations of *Schinus terebinthifolia* Raddi along the Brazilian Atlantic forest. Scientific Reports 11(1), 1-11. DOI: 10.1038/s41598-021-00015-0

Vidal Jr. JD, Cortez MB, Alves FM, Koehler S, de Souza AP, Koch I. 2018. Development and cross-validation of microsatellite markers for *Rauvolfia weddeliana* Müll.Arg. (Apocynaceae) species complex. Brazilian Journal of Botany 41(3): 681–686. DOI: 10.1007/s40415-018-0465-9

Vieira FS et al. 2019. Genetic diversity of Brazil-nut populations naturally occurring in the municipality of Alta Floresta, MT, Brazil. Genetics and Molecular Research 18 (2): gmr18174. DOI: 10.4238/gmr18174

Wang X et al. 2018. Comparative genome-wide characterization leading to simple sequence repeat marker development for Nicotiana. BMC Genomics 19(500): 1-12.

5 CAPÍTULO III - SIMULATION OF PATTERN OF GENE FLOW IN CANJERANA FRAGMENTS IN THE BRAZILIAN ATLANTIC RAINFOREST FOR EVALUATING GENETIC CONSERVATION STRATEGIES

Original article

Serrote, C.M.L.S.; Reiniger, L.R.S.; Costa, L.S.; Stefanel, C.M.; Silva, K.B.; Rabaiolli, S.M.S. Simulation of pattern of gene flow in Canjerana fragments in the Brazilian Atlantic Rainforest for evaluating genetic conservation strategies. **Ciência Rural,** v.49, n.8, e20180764, 2019. https://doi.org/10.1590/0103-8478cr20180764.

ABSTRACT

Gene flow is important for the conservation of genetic resources to allow connectivity of geographically isolated populations and which genetic variability is reduced. Gene movement is a function of flow rate and model. Understanding how gene flow occurs can contribute to the conservation and selection of priority populations that could benefit from an eventual intervention. Simulation softwares allow making inferences about past events based on current datasets or predict future phenomena under real genetic scenarios. Adverse phenomena can be predicted and actions can be taken to avoid them. The aim of this study was to identify a model and the gene flow rates that could explain genetic structure of eight forest fragments of *Cabralea canjerana* in development in the Brazilian Atlantic Rainforest. To do this, simulations were performed with the EASYPOP software using a microsatellite marker dataset obtained for the species by Melo and collaborators, in 2012, 2014 and 2016. We tested five models and nine migration rates and we selected the model that produced values closer to those previously obtained for them. Criteria used for selection were the observed and expected heterozygosity and the Wright's F Statistics obtained in the simulations. The gene flow model selected was the isolation by distance model that used a rate of 0.1. We observed high levels of genetic differentiation among

the fragments as result of their reproductive isolation. To allow homogenization of the allelic frequencies through gene flow, the solution would be to create ecological corridors with the aim of connecting distant fragments.

Keywords: *Cabralea canjerana*, reproductive isolation, genetic variability, EASYPOP, microsatellites.

RESUMO

O fluxo gênico, cuja efetividade é função do modelo e da taxa, assume especial importância na conservação de recursos genéticos por permitir a conectividade de populações isoladas geograficamente, sujeitas à redução da variabilidade genética. O entendimento de como o fluxo gênico ocorre pode contribuir no planejamento de ações para a conservação e na seleção de populações prioritárias para uma eventual intervenção. Programas de simulação permitem inferir sobre eventos passados, a partir de dados atuais ou prever fenômenos futuros sob cenários genéticos reais. Fenômenos adversos podem ser previstos e medidas podem ser tomadas para contorná-los. O objetivo deste estudo foi identificar o modelo e a taxa de fluxo gênico que melhor explicam a estrutura genética de oito fragmentos da espécie arbórea florestal *Cabralea canjerana*, em desenvolvimento na região brasileira do bioma Mata Atlântica. Foram realizadas simulações com o programa EASYPOP usando dados de marcadores microssatélites obtidos por Melo e colaboradores, em 2012, 2014 e 2016, sendo testados cinco modelos e nove taxas de migração, selecionando-se o modelo que apresentou os valores mais próximos daqueles que foram publicados. O modelo de fluxo gênico entre os fragmentos foi o de isolamento por distância a uma taxa de 0.1. Foram observados elevados índices de diferenciação genética entre os fragmentos em decorrência do seu isolamento reprodutivo. Desse modo, sugere-se a

construção de corredores ecológicos com vistas a conectar fragmentos distantes e, desta forma, permitir a homogeneização das frequências alélicas por meio do fluxo gênico.

Palavras-chave: *Cabralea canjerana*, isolamento reprodutivo, variabilidade genética, EASYPOP, microssatélites.

INTRODUCTION

In general, most conservation programs are focused on reducing gene flow barriers between isolated populations that carry a small fraction of the total variability in relation to continuous populations. Populations of a species can be reproductively isolated by a physical barrier (e.g. mountains) or by fragmentation, which is often caused by several types of human activities including construction of roads, railways, cities, and expansion of farmland. In small fragments, events like genetic drift and inbreeding may occur, which can reduce genetic variability (PRIMACK & RODRIGUES, 2001; BROQUET et al., 2010).

Over time, populations with low genetic variability have low evolutionary potential and limited ability to respond to environmental changes in long term. While a large area habitat can hold a single large population, it is possible that none of its fragments can maintain a subpopulation for an extended period (PRIMACK & RODRIGUES, 2001).

Population connectivity and redistribution of genetic variability among the individuals allows the homogenization of allelic frequencies, thus providing the viability of populations with sizes below that required to control for the adverse effects of drift and inbreeding. Thus, small isolated populations start to behave as a large panmitic population with greater genetic variability and this reduces the risk of its extinction.

Furthermore, the restoration of gene flow homogenizes the allelic frequencies, thereby maintaining genetic variability and polymorphism (SLATKIN, 1985; ZUCCHI et al., 2003; PINHO & HEY, 2010).

Improving our understanding of gene flow (model and rate) between geographically isolated populations is important for planning conservation strategies, including the selection of priority populations that would benefit from an eventual intervention. Wright's island model is a conventional model used in population genetics to evaluate gene flow between populations and allows the estimation of the actual number of migrants (Nm) by means of the F_{ST} statistic for an *n* set of populations (WRIGHT, 1965). However, this model does not have realistic assumptions for all populations, such as the equilibrium between migration and genetic drift, if there is equal gene exchange between populations, and if all populations have equivalent sources of migrants. Also, the estimates obtained with this model do not reflect the contemporary variation in gene exchange between populations or current changes in the dispersal process.

Due to these limitations, there is need for more accurate methods for unraveling the models and gene flow rates for specific populations in order to ensure eventual success of genetic conservation programs. With advances in computer technology, simulation software was developed to make inferences about past events from actual data, as well as those that perform simulations forward in time, allowing predictions of future phenomena using real genetic scenarios (YUAN et al., 2012; DALQUEN et al., 2012). In this simulation research we focalized *Cabralea canjerana* ssp. *canjerana* (Meliaceae), a dioecious forest species that is considered model to Brazilian Atlantic Rainforest conservation studies (KLEIN, 1979). The Atlantic Rainforest is the Brazilian biome most affected by environmental fragmentation. Today's the biome presents itself

as a mosaic composed of few relatively large areas. Thus, the forest fragments of diverse sizes and shapes, assume fundamental importance for the perenniality of the Atlantic Rainforest biome (ZAÚ, 1998).

The objective of the present study was to identify the model and the gene flow rate that best explain the genetic structure of eight fragments of *C. canjerana* using data from microsatellite markers.

MATERIAL AND METHODS

Simulations

The EASYPOP Version 2.0.1 software (BALLOUX, 2001) was used to simulate different models and flow rates, which were selected based on how well they could explain the genetic structure of eight *C. canjerana* fragments obtained by six microsatellite markers from MELO et al. (2014). In these population fragments studied were sampled: (#1) 27, (#2) 33, (#3) 31, (#4) 33, (#5) 17, (#6) 11, (#7) 39, and (#8) 30 individuals. All of the populations were located in the Atlantic Rainforest biome in the Environmental Protection Area Fernão Dias, which is in the south of the state of Minas Gerais, in Brazil (Figure 1 and Table 1). This area has suffered in the last 60 years a great deforestation for extractivism and potato planting. The Environmental Protection Area Fernão Dias has 180,073 hectares and was created in 1997 as an environmental compensation plan for the duplication of BR 381 (MELO, 2012).

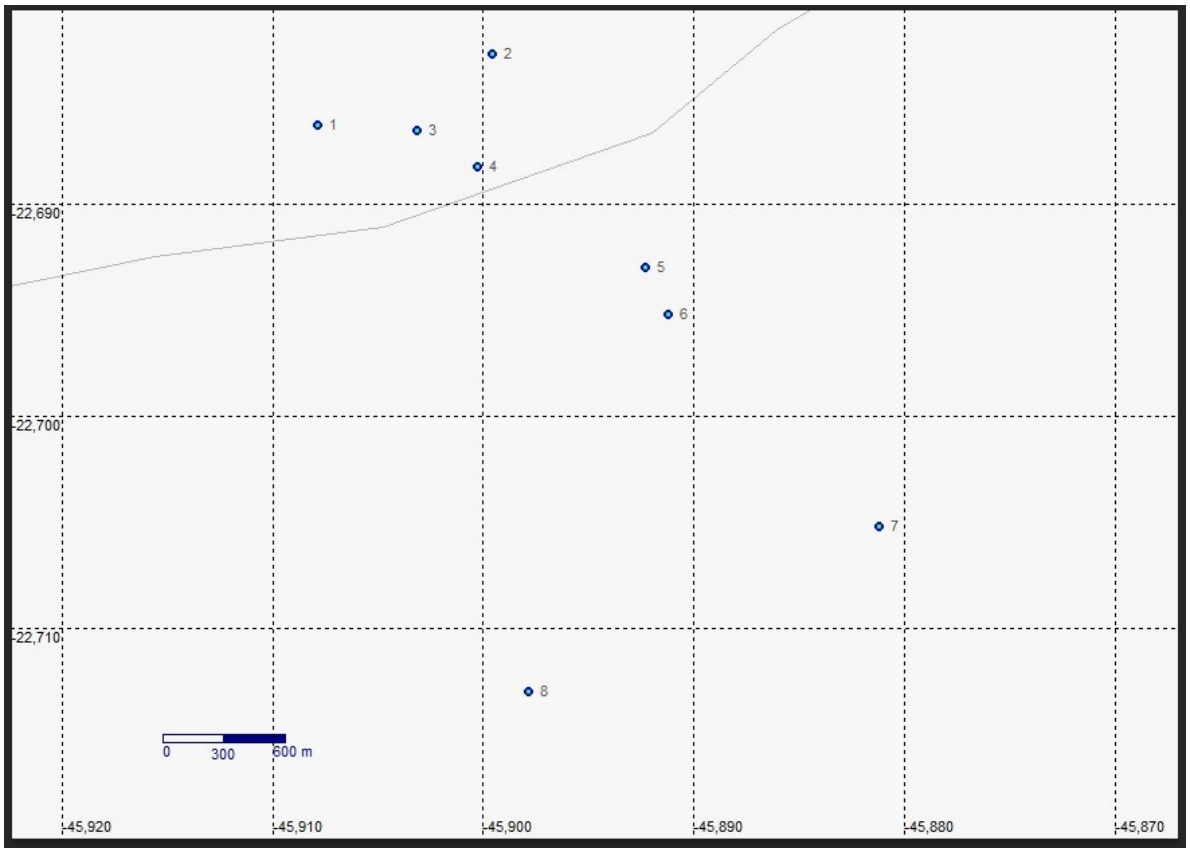


Figure 1 - Geographical location of the eight *C. canjerana* fragments used for the simulations and molecular characterization in the present study. Source: MELO et al. (2014).

Table 1. Geographical traits of the studied *C. canjerana* fragments

Fragments	Latitude S	Longitude W	Elevation (m)
#1	22º 41’ 11”	45º 54’ 28”	1681
#2	22º 40’ 59”	45º 53’ 58”	1667
#3	22º 41’ 12”	45º 54’ 11”	1630
#4	22º 41’ 18”	45º 54’ 06”	1603
#5	22º 41’ 35”	45º 53’ 32”	1598

#6	22º 41’ 43”	45º 53’ 28”	1600
#7	22º 42’ 19”	45º 52’ 52”	1810
#8	22º 42’ 47”	45º 53’ 52”	1928

Source: Adapted from MELO et al. (2014)

Simulation parameters

In this study, we tested the following migration models: one-dimension stepping stone, island, hierarchical stepping stone ('contact zone'), hierarchical island, and spatial (isolation by distance). For the hierarchical stepping stone ('contact zone') model, two groups were considered with four fragments in each group. The first group consisted of 1, 2, 3, and 4; and the second group was composed of 5, 6, 7 and 8 fragments. The contact zone between the groups was composed of fragments 4 and 5. In the hierarchical island model, two archipelagos were composed of four islands (fragments) each. The archipelago for each fragment was the same composition as the previous model groups and the distribution of the fragments used was the same as that in the hierarchical stepping stone model.

In the spatial model simulation, two dimensions based on the real geographic position of the fragments were used. The average distance of pollen dispersion was 127 m (MELO & FRANCESCHINELLI, 2016). For the seeds, the same dispersal value was used because, according to SLATKIN (1985), seed dispersal distance estimates are similar to those of pollen.

The following parameters were considered in our simulations: a dioic species, migration rates between 0.1 and 0.9, six loci, mixed mutation Single-step Model (SSM) with a proportion of K allele Model (KAM) mutation events (with a ratio of 0.1) and 25 possible allelic states, maximum variability of the initial population, and a mutation rate

of 0.00005, which is in accordance with the mean observed in nature (PIRES et al., 2011). The generation number for each simulation was 100. For each model and migration rate combination 100 independent replicates were conducted. We selected the model that showed values of observed and expected heterozygosity and Wright's F Statistics closest to those previously obtained with microsatellite markers by MELO et al. (2014).

RESULTS AND DISCUSSION

Model selection

Based on observed and expected heterozygosity average values of 0.69 and 0.72, respectively, obtained by MELO et al. (2014) we pre-selected the following models for gene flow: the one-dimensional stepping stone model with migration rates of 0.5 and 0.6; island model, which has a migration rate of 0.1; the hierarchical stepping stone model with a migration rate 0.2; and the spatial model, which has a migration rate of 0.1. Using these models, we obtained statistics for genetic structure similar to those obtained with molecular markers (Table 2).

Table 2. Observed heterozygosity estimates (H_O), expected heterozygosity (H_S), fixation index between fragments (F_{ST}) and total inbreeding (F_{IT}) generated by simulations (software EASYPOP) in function of different models and migration rates in eight fragments of *C. canjerana*.

		Migration rates								
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
One-dimension	H_O	0.577	0.643	0.643	0.666	0.697	0.696	0.716	0.729	0.728
Stepping Stone	H_S	0.581	0.643	0.645	0.667	0.698	0.699	0.716	0.727	0.726
Model	F_{ST}	0.266	0.185	0.159	0.140	0.103	0.100	0.072	0.066	0.062

	F_{IT}	0.271	0.184	0.163	0.141	0.104	0.104	0.072	0.064	0.060
Island Model	H_O	0.709	0.733	0.747	0.755	0.759	0.768	0.763	0.765	0.673
	H_S	0.714	0.730	0.741	0.752	0.756	0.761	0.759	0.761	0.668
	F_{ST}	0.101	0.051	0.032	0.019	0.014	0.011	0.008	0.005	0.002
	F_{IT}	0.107	0.047	0.025	0.016	0.010	0.002	0.002	0.001	-0.006
Hierarchical Stepping Stone	H_O	0.640	0.665	0.694	0.733	0.728	0.747	0.735	0.744	0.730
	H_S	0.642	0.673	0.700	0.734	0.733	0.746	0.738	0.744	0.729
	F_{ST}	0.181	0.126	0.098	0.068	0.059	0.045	0.045	0.047	0.051
	F_{IT}	0.184	0.137	0.106	0.069	0.064	0.043	0.049	0.046	0.050
Hierarchical Island Model	H_O	0.772	0.769	0.776	0.765	0.768	0.762	0.765	0.756	0.760
	H_S	0.767	0.765	0.776	0.764	0.765	0.759	0.759	0.754	0.760
	F_{ST}	0.004	0.004	0.000	0.000	0.000	0.000	0.002	0.003	0.004
	F_{IT}	-0.002	0.000	0.000	0.000	-0.003	-0.005	-0.007	0.001	0.003
Spatial Model (isolation by distance)	H_O	0.709	0.720	0.758	0.742	0.754	0.771	0.755	0.763	0.777
	H_S	0.704	0.723	0.754	0.741	0.753	0.768	0.751	0.763	0.770
	F_{ST}	0.107	0.047	0.027	0.021	0.014	0.010	0.008	0.006	0.005
	F_{IT}	0.101	0.050	0.023	0.019	0.012	0.007	0.003	0.006	-0.005

The exclusion criteria used in the model selection was pollen dispersal distance; the average distance obtained from the microsatellite markers was 127.33 m, and a range of 42 - 1,040 m. Given that physiological barriers may prevent gene exchange between fragments, the statistical pairwise difference (R_{ST}) was assessed. The data obtained by MELO & FRANCESCHINELLI (2016) showed a strong correlation with the Mantel test ($r= 0.836$, $p< 0.0001$) (Table 3). It was; therefore, considered that fragments

with a distance of more 1,040 m could not exchange genes with each other. The fulfillment of the first law of geography (TOBLER, 1970) was also observed, with less genetic differentiation occurring among the closest fragments and vice versa.

Table 3. R_{ST} pairwise (upper diagonal) and geographic distance (km, lower diagonal) based on the Euclidean distance for the eight fragments of *C. canjerana*. All R_{ST} pairwise values were significant at 95%.

	Frag. 01	Frag. 02	Frag. 03	Frag. 04	Frag. 05	Frag. 06	Frag. 07	Frag. 08
Frag. 01	-	0.0110	0.0093	-0.0091	0.0951	0.0798	0.1626	0.3357
Frag. 02	0.9400	-	-0.0014	-0.0075	0.0843	0.0798	0.1856	0.3519
Frag. 03	0.4200	0.6600	-	0.0034	0.0698	0.0684	0.1404	0.3084
Frag. 04	0.7400	0.6200	0.3600	-	0.0986	0.0975	0.1991	0.3628
Frag. 05	1.7800	1.3900	1.4100	1.0700	-	0.0168	0.0984	0.2520
Frag. 06	2.0300	1.6500	1.6600	1.3100	0.3100	-	0.0832	0.2323
Frag. 07	3.5300	3.1100	3.1300	2.7800	1.7400	1.4900	-	0.0868
Frag. 08	3.2600	3.3500	2.9700	2.7500	2.1400	2.0100	1.7600	-

Source: MELO (2012).

Based on the pollen dispersal distance and geographic arrangement of the population fragments, the fragment groups1, 2, 3, and 4 were exchange genes with each other. This exchange was also observed between fragments 5 and 6, but not in the groups that were over 1,040 m apart (MELO & FRANCESCHINELLI, 2016). Thus, only the spatial model of isolation by distance with migration rate of 0.1 could explain the pattern of gene flow between the eight *C. canjerana* fragments studied (Table 2). The difference of statistics in migration rate 0.1 in relation to others is significant, by *t*

test at 5% error probability. Spatial model and migration rate are the most suitable for the genetic characterization of these population fragments.

Characterization of the Fragments

The gene flow rate between *C. canjerana* fragments was 0.1, as indicated in the isolation-by-distance model. Only proximal fragments could exchange genes between them, if they were limited to the maximum dispersion distance of 1,040 m; however, a migration rate of 0.1 underestimates the actual movement of pollen and seeds, since it considers only the successful migrants.

According to ENNOS (1994), the estimation of gene flow in plant populations is complex because it is done in two ways - through both pollen and seeds. Gene flow via pollen involves the dispersal of pollen of one population and the successful fertilization the flowers of another population; while gene flow via seeds involves the dispersal of seeds of one population and their successful establishment in a new population (RIDLEY & MALLORY-SMITH, 2015). A complete description of gene flow between plant populations should ideally include the evaluation of the relative importance of both pollen and seeds as gene flow agents. This analysis is very complex and highlights the value of simulation models, which can be based on the genetic structure that is actually observed in the population.

Due to species population fragmentation, both the observed and expected heterozygosity decreased over 100 generations (Figure 2A). This reduction in heterozygosity occurred due to the loss of alleles for genetic drift. Simulations performed by STEFENON & COSTA (2012) with EASYPOP software indicated an increase in the inbreeding coefficient over generations in small populations compared to larger populations, resulting in lower genetic variability. However, random mating

contributed to an observed heterozygosity that was similar to that expected from the Hardy-Weinberg equilibrium in all generations. This result confirms the importance of intra-population gene flow for the conservation of genetic variability within these fragments. A study of *Luehea divaricata* fragments in the Pampa biome showed that the largest fraction of genetic variability was distributed within the fragments due to high crossing rates (NAGEL et al., 2015).

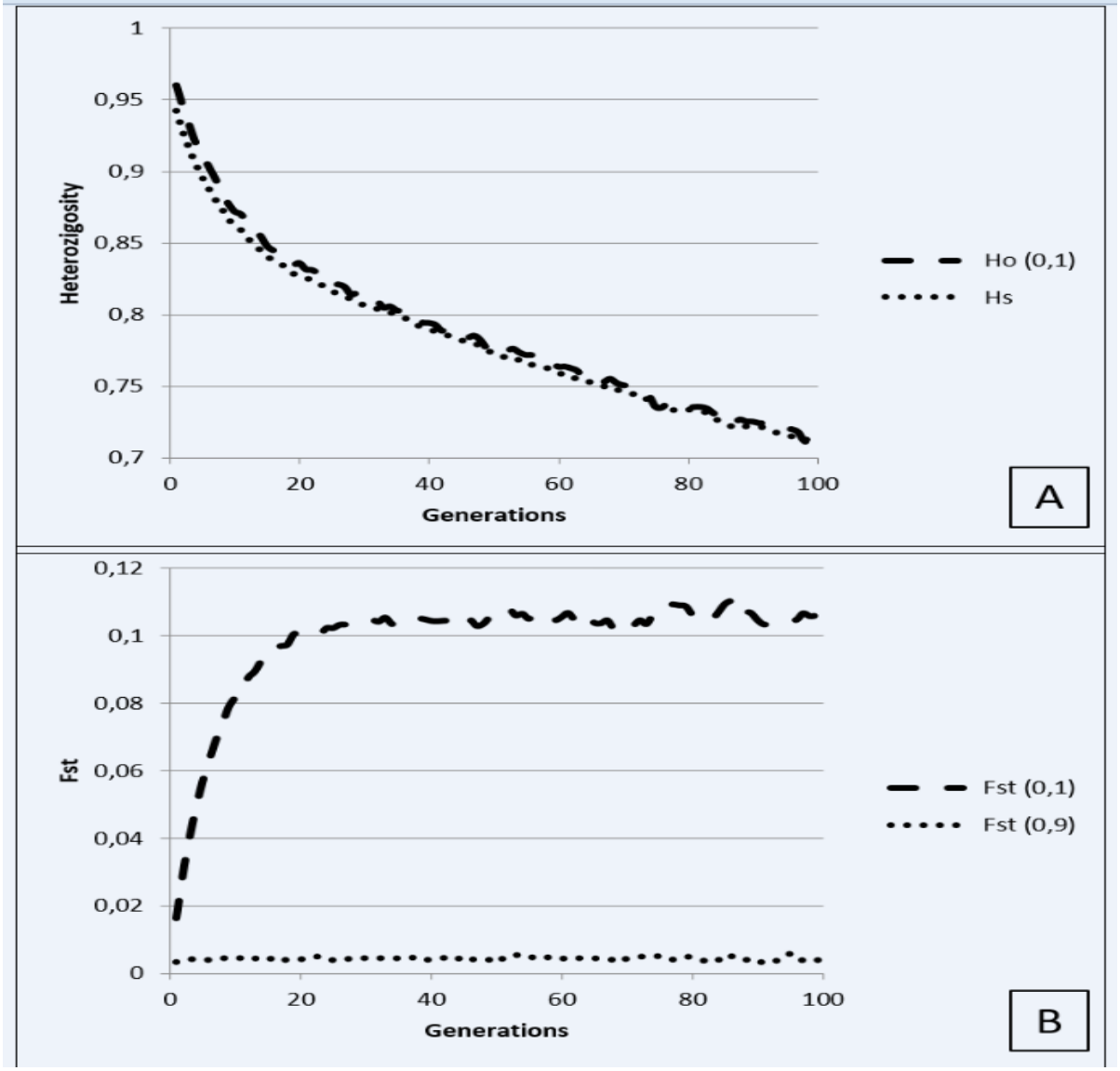


Figure – 2 Geographical location of the eight *C. canjerana* fragments used for the simulations and molecular characterization in the present study. Source: MELO et al. (2014).

Another effect of gene flow is the reduction of genetic divergence between populations. To study this effect in the *C. canjerana* fragments, we plotted F_{ST} values with migration rates of 0.1 and 0.9 (Figure 2B). There was a differentiating coefficient near zero when the migration rate was 0.9, resulting in homogeneous fragments. This indicates that fragmentation of the population was not strongly affected by inbreeding, which is attributed almost exclusively to two-parent crossings within the fragments. A migration rate of 0.1 caused high levels of inbreeding due to reproductive isolation that limited the gene exchange between the fragments. Genetic differentiation was more pronounced in the first generations, as observed in the stabilization after the thirtieth generation. This delayed stabilization suggested that the loss of genetic variability in all fragments did not allow genetic differentiation to continue growing.

The decrease in heterozygosity suggested the same behavior in the future, requiring the adoption of interventional measures that reduce the risk of population extinction. Similarly, the simulation using a higher migration rate allowed us to predict how much genetic variability was lost due to the low migration rate, further confirming the need for the development of strategies for the conservation of these genetic fragments.

Assuming that all fragments of the first group (1, 2, 3, 4) exchange genes between each other, as a function on their geographical distance and the correlation between the geographical distance and the value of R_{ST}, we proposed the creation of ecological corridors that would connect the fragments 7 and 8, 6 and 7, 5 and 8, and 6 and 8 as a potential intervention plan. This plan could also include a contact area between fragments 4 and 5, which would follow the hierarchical island model for gene flow, and form a second island between the fragments 5, 6, 7 and 8. A more economical but less effective alternative could include ecological corridors only among

the 4 and 5, 6 and 7, and 7 and 8 fragments, in which case the gene flow would be governed under the mixed model of islands and stepping stones.

Implications for conservation

The studied area has a deforestation history of more than 60 years, which led to its fragmentation. In addition, *C. canjerana* does not have long-distance pollen dispersers. The main pollen dispersers are the moths of the Lepidoptera order, which have short flight behavior, pollinating nearby trees (PIZO & OLIVEIRA, 1998). The dispersion of fruits and seeds is zoocorical, mainly by birds, attracted by the orange color of the aril (PIZO, 1997). As a result, high levels of genetic differentiation were observed among fragments due to their reproductive isolation characterized by low migration rates. This condition propitiates the occurrence of inbreeding and genetic drift that lead to progressive loss of genetic variability and, ultimately, population extinction.

Therefore, the findings in this study allowed a better management of the area toward its conservation by enabling gene flow among fragments, thus increasing the genetic variability.

CONCLUSION

The eight fragments of *C. canjerana* studied have high levels of genetic differentiation due to reproductive isolation (the gene flow model is isolation by distance), and because a gene flow rate equivalent to 0.1 is not sufficient for the homogenization of allelic frequencies. In this condition these fragments demand special attention to reduce their vulnerability to adverse events. We suggested the

creation of ecological corridors in order to increase the flow rate, connect the distant fragments, and provide homogenization of the fragments via gene flow.

REFERENCES

BALLOUX, F. EASYPOP (version 1.7). A computer program for the simulation of population genetics. **Journal of Heredity,** v.92, p.301-302, 2001. Available from: <https://doi.org/10.1093/jhered/92.3.301>. Accessed: Sept. 18, 2018. doi: 10.1093/jhered/92.3.301.

BROQUET, T. et al. Genetic bottlenecks driven by population disconnection. **Conservation Biology**, v.24, p.1596-1605, 2010. Available from: <https://doi.org/10.1111/j.1523-1739.2010.01556.x >. Accessed: Sept. 10, 2018. doi: 10.1111/j.1523-1739.2010.01556.x.

DALQUEN, D.A. et al. ALF—a simulation framework for genome evolution. **Molecular Biology and Evolution,** v.29, p.1115-1123, 2012. Available from: <https://doi.org/10.1093/molbev/msr268>. Accessed: Sept. 10, 2018. doi: 10.1093/molbev/msr268.

ENNOS, R.A. Estimating the relative rates of pollen and seed migration among plant populations. **Heredity**, v.72, p.250-259, 1994. Available from: <https://doi.org/10.1038/hdy.1994.35>. Accessed: Sept. 18, 2018. doi: 10.1038/hdy.1994.35.

KLEIN,R.M. Ecologia da flora e vegetação do Vale do Itajaí. **Sellowia**, v.31, p.1-164, 1979.

MELO, A.T.O. **Fluxo gênico e estrutura genética espacial de *Cabralea canjerana* (Vell.) Mart. (MELIACEAE) em fragmentos florestais de Mata Atlântica.** 2012. 88f. Master's thesis (Master in Genetics and Plant Breeding) – Postgraduate course in Genetics and Plant Breeding, Federal University of Goiás.

MELO, A.T.O. et al. High genetic diversity and strong spatial structure in ***Cabralea canjerana*** (Vell.) Mart. (Meliaceae): implications to Brazilian Atlantic Forest tree conservation. **Brazilian Journal of Nature Conservation,** v.12, p.129-133, 2014. Available from: <https://doi.org/10.1016/j.ncon.2014.08.001>. Accessed: Sept. 5, 2018. doi: 10.1016/j.ncon.2014.08.001.

MELO, A.T.O.; FRANCESCHINELLI, E.V. Gene flow and fine-scale spatial genetic structure in ***Cabralea canjerana*** (Meliaceae), a common tree species from the Brazilian Atlantic forest. **Journal of Tropical Ecology**, v.32, p.135-145, 2016. Available from: <https://doi.org/10.1017/S0266467416000067>. Accessed: Aug. 22, 2018. doi: 10.1017/S0266467416000067.

NAGEL, J.C.et al. Historical gene flow within and among populations of ***Luehea divaricata*** in the Brazilian Pampa. **Genetica**, v.143, p.317-29, 2015. Available from: <https://link.springer.com/article/10.1007/s10709-015-9830-9>. Accessed: Aug. 22, 2018. doi: 10.1007/s10709-015-9830-9.

PINHO, C.; HEY, J. Divergence with Gene Flow: Models and Data. **Annual Review of Ecology, Evolution, and Systematics***,* v.41, p.215-230, 2010. Available from: <https://doi.org/10.1146/annurev-ecolsys-102209-144644>. Accessed: Aug. 21, 2018. doi: 10.1146/annurev-ecolsys-102209-144644.

PIRES, I.E. et al. **Genética florestal**. Viçosa: Arka, 2011. 1st ed.

PRIMACK, R.B.; RODRIGUES, E. **Biologia da conservação.** Londrina: Planta. 2001. 1st ed.

PIZO, M. A. Seed dispersal and predation in two populations of *Cabralea canjerana* (Meliaceae) in the Atlantic Forest of southeastern Brazil. **Journal of Tropical Ecology** v.13, p.559- 578, 1997. Available from: <https://www.jstor.org/stable/2560180>. Accessed: Feb. 27, 2019. doi: 10.1017/S0266467400010713

PIZO, M. A.; OLIVEIRA, P. S. Interaction between ants and seeds of a nonmyrmecochorous neotropical tree, *Cabralea canjerana* (Meliaceae), in the Atlantic forest of southeast Brazil. **American Journal of Botany**, v.85, p.669–674, 1998. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/21684948>. Accessed: Feb. 27, 2019.

RIDLEY, C.E.; MALLORY-SMITH, C. Gene flow and Invasiveness in Bioenergy Systems. In: QUINN, L.D. et al. **Bioenergy and Biological Invasions: Ecological, Agronomic, and Policy Perspectives on Minimizing Risk**. Croydon: CPI Group, 2015. p: 52-66.

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

STEFENON V.M.; COSTA, L.S. A simulation study on the behavior of allelic richness and inbreeding coefficient over generations in fragmented populations of tree species. **Annals of Forest Research**, v.55, p.3-10, 2012. Available from: <http://afrjournal.org/index.php/afr/article/view/70>. Accessed: Sept. 18, 2018. doi: 10.15287 / afr.2012.70.

TOBLER, W.R. A computer movie simulating urban growth in the Detroit region. **Economic Geography**, v.46, p.234-240, 1970. Available from: <https://www.jstor.org/stable/143141>. Accessed: Sept. 15, 2018. doi: 10.2307 / 143141.

WRIGHT, S. The interpretation of population structure by F-statistics with special regard to systems of mating. **Evolution**, v.19, p.395-420, 1965. Available from: <https://doi.org/10.1111/j.1558-5646.1965.tb01731.x>. Accessed: Sept, 15, 2018. doi: 10.1111/j.1558-5646.1965.tb01731.x.

YUAN, X. An overview of population genetic data simulation. **Journal of Computational Biology**, v.19, p.42-54, 2012. Available from: <https://doi.org/10.1089/cmb.2010.0188>. Accessed: Sept. 12, 2018. doi: 10.1089/cmb.2010.0188.

ZAÚ, A.S. Fragmentação da Mata Atlântica: Aspectos teóricos. **Floresta e Ambiente**, v.5, p.160-170, 1998.

ZUCCHI, M.I. et al. Genetic structure and gene flow in ***Eugenia dysenterica*** DC in the Brazilian Cerrado utilizing SSR markers. **Genetics and Molecular Biology**, v.26, p.449-457, 2003. Available from: <http://www.scielo.br/scielo.php?pid=S1415-47572003000400008&script=sci_arttext>. Accessed: Sept. 12, 2018. doi: 10.1590/S1415-47572003000400008.

6 CAPÍTULO IV - GENETIC STRUCTURE SIMULATION FOR *HANCORNIA SPECIOSA* POPULATIONS IN NORTHEAST BRAZIL

Original article

Serrote, C.M.L.S.; Reiniger, L.R.S.; Costa, L.S.; Rabaiolli, S.M.S.; Stefanel, C.M.; Silva, K.B.; Ziegler, A.C.F. Genetic structure simulation for *Hancornia speciosa* populations in Northeast Brazil. **Ciência Rural,** v.53:1, e20210176, 2023. https://doi.org/10.1590/0103-8478cr20180764.

ABSTRACT

Mangaba tree is a fruit tree species whose natural populations are fragmented by anthropic actions. For this reason, studies assessing the impact of fragmentation on the diversity and genetic structure of these populations are required in order to establish suitable conservation strategies. In our study, we used data from analyzes through microsatellite markers in computer simulations to estimate the rates of migration and selfing of six mangaba populations. The studied populations are located in the northeastern states of Ceará, Pernambuco and Sergipe. We tested different selfing and migration rates and selected the combination that showed values of observed and expected heterozygosity closest to those previously obtained with microsatellite markers. According to our simulations, selfing and migration were moderate. This may have led to an increase in inbreeding and genetic drift, resulting in low genetic diversity. We recommend expanding the area and reducing disturbance to promote the occurrence of pollinators, which play an important role in increasing genetic diversity.

Keywords: Easypop, genetic diversity, habitat fragmentation, mangaba tree, microsatellites.

RESUMO

A mangabeira é uma espécie frutífera cujas populações naturais se encontram fragmentadas por ações antrópicas. Desse modo, são necessários estudos sobre a avaliação do impacto da fragmentação sobre a diversidade e estrutura genética dessas populações para o estabelecimento de estratégias de conservação adequadas. No presente estudo, foram utilizados dados de análises com marcadores microssatélites em simulações computacionais para estimar as taxas de migração e autofecundação de seis populações de mangaba. As populações estudadas estão localizadas nos estados nordestinos do Ceará, Pernambuco e Sergipe. Foram testadas diferentes taxas de autofecundação e migração, e selecionada a combinação que apresentou valores de heterozigosidade observada e esperada mais próximos dos obtidos com marcadores microssatélites. Com base nas simulações, a autofecundação foi de 0,3 e a taxa de migração variou de 0,5 a 0,6, valores que podem ter conduzido ao aumento da endogamia e deriva genética, resultando em baixa diversidade genética. Recomenda-se a expansão da área e a redução de perturbações para promover a ocorrência de polinizadores, que desempenham um papel importante no aumento da diversidade genética.

Palavras-chave: Easypop, diversidade genética, fragmentação de habitat, mangabeira, microssatélites.

INTRODUCTION

Habitat fragmentation is the process by which a large and continuous habitat is split into smaller and more isolated remnants, primarily by human disturbances such as land clearing and conversion of vegetation from one type to another. This phenomenon affects many ecosystems and species such as mangaba tree (*Hancornia speciosa* Gomes), a fruitful species native to Brazil (FRANKLIN et al., 2002; MOURA et al., 2005; SCHLAEPFER et al., 2018). It is a perennial species native to several regions and ecosystems in Brazil, extending along the coast,

occurring naturally in marginal soils, subject to long periods of drought such as savannas and semi-arid areas in the Northeast. It is also reported throughout the Cerrado region of Central Brazil to the Pantanal biome (LEDERMAN et al., 2000). Among many consequences, habitat fragmentation in forest ecosystems affects the phenology, pollination patterns and reproductive success of species. The reduction in population size may reduce the density of reproductive trees, limit pollen availability and propitiate the occurrence of inbreeding and genetic drift effects. Besides the size reduction, the lack of connectivity can affect pollen and seed dispersal and limit gene flow among isolated fragments. These factors can lead to progressive loss of genetic diversity and, ultimately, to extinction (DUMINIL et al., 2016). For this reason, it is important to assess the impact of fragmentation on the genetic diversity and structure of forest tree populations in order to establish appropriate strategies for conservation.

Some conservation programs focus on promoting gene flow among isolated fragments. Gene flow is an evolutionary force that enables the allele exchange, thus increasing the genetic diversity. Moreover, the reproductive system plays an important role in the genetic diversity since selfing limits the pollen and seed dispersion and the potential for recombination between alleles from different individuals, in contrast to crossing. As a result, it is expected that crossing species preserve higher genetic diversity than autogamous species (JULLIEN et al., 2019). In our study, we used data from microsatellite markers in computer simulations to estimate the selfing and migration rate of six *H. speciosa* populations from the Caatinga biome, covering the northeastern Brazilian states of Ceará, Pernambuco and Sergipe. The software used for the simulations was the Easypop 2.0.1 (BALLOUX, 2001).

MATERIALS AND METHODS

The Easypop 2.0.1 software was used to simulate different rates of selfing (0.1, 0.3, or 0.5) and migration (from 0.1 to 0.9, with steps of 0.1), which were selected based on how well they could explain the genetic structure of six *H. speciosa* populations obtained by microsatellite markers from AMORIM et al. (2015). In the referred study, the average observed heterozygosity was 0.47, and the average expected heterozygosity was 0.60. The studied populations were: Reserva do Cajú (SE), Abaís (SE), Barrados Coqueiros (SE), Jacarecoara (CE), Tapera (CE) and Tamandaré (PE), all located in northeast Brazil, in the Caatinga biome (Table 1). A total of 94 indivíduals were collected, 59 from the state of Sergipe (SE), 20 from Ceará (CE), and 15 from Pernambuco (PE). According to the authors (AMORIM et al., 2015), in some cases, the total number of individuals was exactly the population size because these are remnant populations.

Table 1 – Sample sizes and geographic coordinates of the studied *Hancornia speciosa* populations used in the simulations.

Population	Sample size	Geographic coordinates
Reserva do Caju (SE)	19	11°11'6"S and 37°11'18"W
Abaís (SE)	20	11°18'18"S and 37°17'18"W
Barrados Coqueiros (SE)	20	10°49'10"S and 36°56'52"W
Jacarecoara (CE)	14	4°07'10"S and 38°10'34"W
Tapera (CE)	6	3°56'20"S and 38°20'18"W
Tamandaré (PE)	15	8°43'50"S and 35°6'10"W
Total	94	

Source: AMORIM et al. (2015).

For the simulations, we considered diploid hermaphrodite species, with non-random mating system and non-clonal reproduction. According to the study by DARRAULT; SCHLINDWEIN (2006), it is self-incompatible, demanding different genotypes of the species and specific pollinators for crossing and fruit production to occur. The flowers have a complex

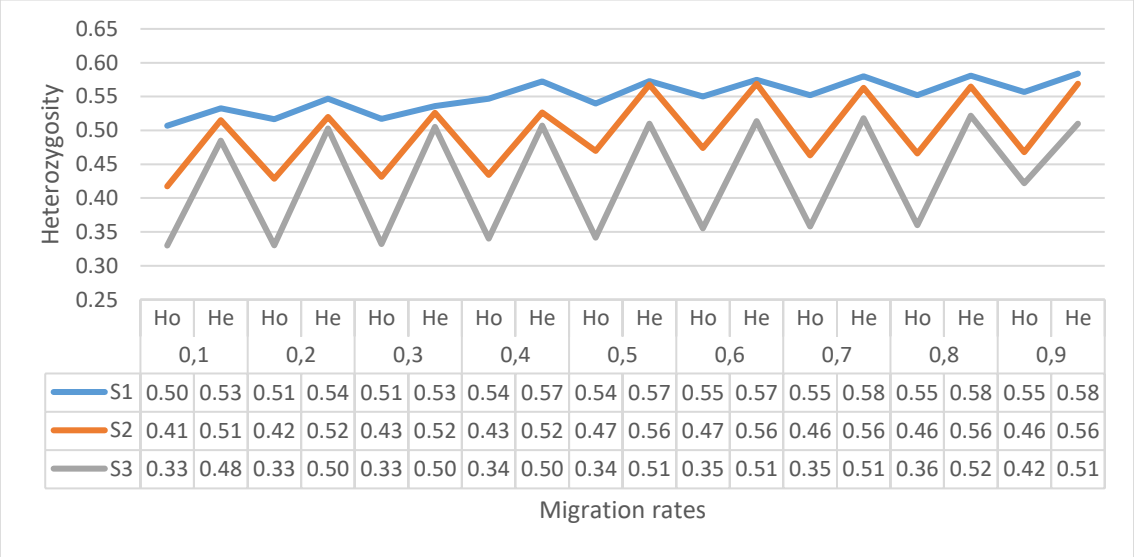
pollination mechanism co-adapted to pollination by moths and butterflies. The fruits are dispersed by large and medium-sized mammals. A spatial migration model was considered, with coordinates based on the geographic location of the populations which were obtained from the geographical coordinates (latitude and longitude) according to the map. Regarding the mutation settings, we assumed nine loci evolving according to the single-step mutation model (SSM), which is assumed to be typical for microsatellites (VALDES et al., 1993), and considered a proportion of 0.1 K-allele model (KAM) events, under 42 possible allelic states, according to the data obtained by Amorim et al. (2015). A mutation rate of 0.0001 mutations per locus and per generation was assumed. The genetic variability of the initial population was considered the maximum, and authors simulated 100 generations. Each combination of selfing and migration rates was replicated 100 times. The observed and expected heterozygosities from AMORIM et al. (2015) were used to select the model settings that presented values closest to the field observations. The two independent sample *t*-tests at a 5% level of probability were used to compare the observed and expected heterozygosities of the selected combination of selfing and migration rates with other combinations.

RESULTS AND DISCUSSION

According to our simulations, there is a clear trend of increasing heterozygosity (both H_O and H_E) with increasing migration rate and decreasing selfing (Figure 1). Heterozygosity is one of the parameters of genetic diversity. While gene flow increases the genetic diversity by enabling the allele exchange among populations (SMITH et al., 2020), selfing reduces genetic diversity by limiting the pollen and seed dispersion and; therefore, the potential for recombination (JULLIEN et al., 2019).

The closest values of H_O and H_E to those obtained through the use of microsatellites

(0.47 and 0.60, respectively) were generated under the selfing of 0.3 and the migration rate of 0.5 to 0.6 and these values were significantly different from the other rates, based on two independent sample t-tests at a 5% level of probability of error. Migration rates are higher than selfing rates, which suggested that spatial patterns may be more important than the degree of inbreeding for *H. speciosa.*



	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He	Ho	He
	0,1		0,2		0,3		0,4		0,5		0,6		0,7		0,8		0,9	
S1	0.50	0.53	0.51	0.54	0.51	0.53	0.54	0.57	0.54	0.57	0.55	0.57	0.55	0.58	0.55	0.58	0.55	0.58
S2	0.41	0.51	0.42	0.52	0.43	0.52	0.43	0.52	0.47	0.56	0.47	0.56	0.46	0.56	0.46	0.56	0.46	0.56
S3	0.33	0.48	0.33	0.50	0.33	0.50	0.34	0.50	0.34	0.51	0.35	0.51	0.35	0.51	0.36	0.52	0.42	0.51

Figure 1 - Estimates of observed (H_O) and expected (He) heterozygosities generated by simulations, for different rates of selfing and migration.

The rate of selfing of 0.3 generated from the simulations implies an outcrossing rate of 0.7 which enables us to classify the mode of reproduction of the population as mixed, with a predominance of outcrossing, according to the classification of DESTRO & MONTALVÁN (1999). This author classifies vegetal populations as autogamous (outcross from 0 to 0.05), mixed (outcross from 0.05 to 0.95) and allogamous (outcross from 0.95 to 1.0). Outcrossing in hermaphrodite plants, like *H. speciosa* is usually enforced by self-incompatibility. According to TAKAYAMA & ISOGAI, 2005), self-incompatibility is an important mechanism in

flowering plants that prevents selfing and thereby generate and maintain genetic diversity within a species. The estimated rate of self-incompatibility among tropical tree species is 0.88 (SOBIERAJSKI et al., 2006). In studying the reproductive system of *Shorea congestiflora* and *S. trapezifolia* from Sri Lankan tropical rain forests, MURAWSKI et al. (1994) observed differences in the outcrossing rates between the two species due to differences in the degree of self-incompatibility. This finding ratifies the importance of self-incompatibility in promoting outcrossing; therefore, increasing genetic diversity.

Regarding the migration settings, the simulations derived a rate of 0.5 to 0.6. Migration in these populations occurs, probably, through seed dispersal, mediated by long-distance dispersers, considering the long distances among them. Gene flow is the exchange of alleles among populations, capable of altering the original gene composition. It can be translated as the number of migrants among populations each generation (SLATKIN, 1985). High levels of gene flow increase the genetic diversity in populations (BURCZYK et al., 2004), as observed by plotting the heterozygosity (observed and expected) and the migration rate (Figure 1).

Gene flow in plant populations is an evolutionary force that promotes the homogenization of gene frequencies between populations. Through gene flow, the allele frequencies between populations become more similar. It is opposite to genetic drift that promotes genetic differentiation between populations (ELLSTRAND, 1992). Some conservation programs focus on reducing barriers to gene flow in fragmented populations (KWAK et al., 1998; AULER et al., 2002; KAMM et al., 2010). Such barriers can be physical or distance barriers, often created by populations fragmented into small groups which are more susceptible to genetic drift (NEIGEL, 1997). In another study, BUSCHBOM et al. (2011) observed effective long-distance gene flow that contributed considerably to the genetic

diversity of an oak stand at St Sibay east of the Ural Mountains. This result reinforces the importance of gene flow in the long-term persistence of this relict stand.

In our simulation, there was a gradual decrease in the genetic diversity parameters number of alleles, observed heterozygosity and expected heterozygosity over generations, probably due to anthropic activities (Figure 2). The number of alleles determines the probability for a population to survive over generations under adverse events such as forest fires, insect and disease outbreaks, drought, which result in the loss of part of the original genetic diversity. High genetic diversity increases the chance of existing pre-adapted genotypes to future perturbations. When populations are fragmented, there is a decrease in the number of alleles in each fragment (GAMFELDT & KÄLLSTRÖM, 2007). In analysing the genetic diversity of *Dalbergia odorifera*, LIU et al. (2019) obtained a higher number of alleles in larger populations.

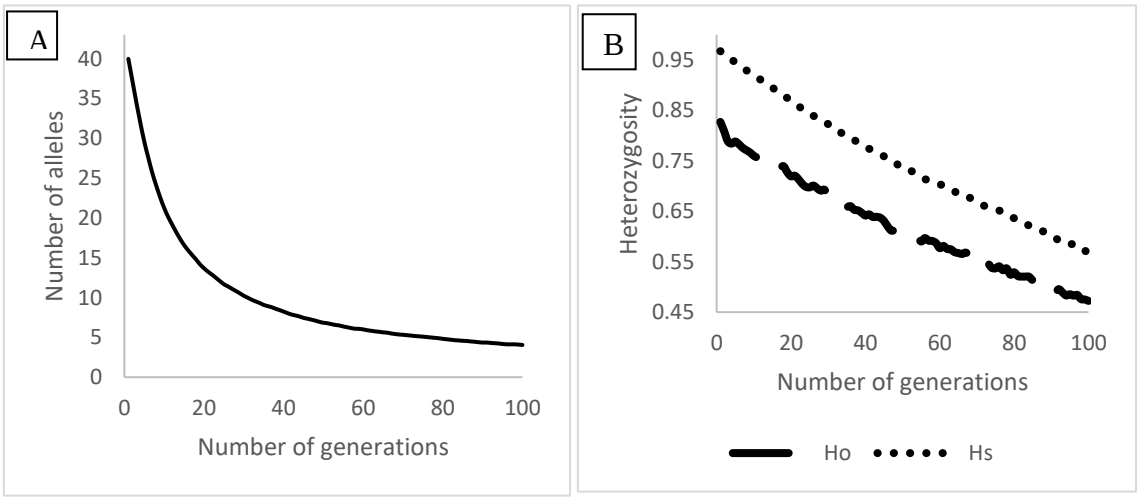


Figure 2 - Evolution of the number of alleles (A) and heterozygosity (B) over time. H_O: observed heterozygosity; He: expected heterozygosity.

The heterozygosity of a population is defined as the probability that two randomly sampled gene copies are different, being a good measure of genetic diversity (NEI, 1978). The possibilities of combining different alleles rely on the number of alleles in the population.

Therefore, the decrease in the number of alleles contributes to reducing the heterozygosity. It can be confirmed by the similar behavior of graphs for these parameters (Figure 2). In addition, when the lost alleles include self-incompatibility alleles, it leads to an increase in selfing, increasing the proportion of homozygosity (VINSON et al., 2015).

Our results reinforced the importance of conserving high genetic diversity in forest populations in order to ensure the sustainability of this important natural resource. The studied populations of *H. speciosa* result from a long fragmentation process by anthropic activities. Furthermore, the genetic erosion of this species can also be explained by human selection throughout domestication due to the real estate expansion and the intensification of agriculture in areas of natural occurrence (SOARES et al., 2019). According to the simulations, the genetic diversity of this species will diminish in the long-running. The long distances that separate these populations today limit the connectivity through gene flow. Therefore, local mutations are the unique source for new alleles. which are recombined through crossing, contributing to a level of genetic diversity in these populations. Thus, local strategies must be adopted to contain genetic erosion. We recommend enriching the area with high genetic diversity germplasm and minimizing disturbance to promote the occurrence of pollinators.

CONCLUSION

The *H. speciosa* populations have a selfing rate of 0.3 and a migration rate of 0.5 to 0.6. The observed selfing and migration rates could be the result of genetic drift and isolation, respectively, reducing the genetic diversity. Strategies to contain the loss of genetic diversity must be adopted locally. We suggest expanding the area and reducing the disturbance to promote the occurrence of pollinators, which are responsible for increasing the genetic diversity by promoting crossing.

REFERENCES

AMORIM, J.A.E. et al. Diversity and genetic structure of mangaba remnants in states of northeastern Brazil. **Genetics and Molecular Research,** v.14, n.1, p.823-833, 2015. Available from: <https://pubmed.ncbi.nlm.nih.gov/25730022/>. Accessed: Mar. 14, 2020. doi: 10.4238/2015.February.2.7.

AULER, N.M.F. et al. The genetics and conservation of *Araucaria angustifolia*: I. Genetic structure and diversity of natural populations by means of nonadaptive variation in the State of Santa Catarina, Brazil. **Genetics and Molecular Biology,** v.25, p.329–338, 2002. Available from: <https://www.scielo.br/j/gmb/a/wSv5Rs7wXpnVQ4NYmh35LfG/?lang=en>. Accessed: Jul. 02, 2021. doi: 10.1590/S1415-47572002000300014.

BALLOUX, F. "EASYPOP (Version 1.7): A Computer Program for the Simulation of Population Genetics. **Journal of Heredity**, v.92, n.3, p.301-302, 2001. Available from: <https://academic.oup.com/jhered/article/92/3/301/2187330>. Accessed: Mar. 14, 2019. doi: 10.1093/jhered/92.3.301.

BUSCHBOM, J. et al. Efficient long-distance gene flow into an isolated relict oak stand. **Journal of Heredity,** v.102, n.4, p.464-472, 2011. Available from: <https://academic.oup.com/jhered/article/102/4/464/899137>. Accessed: Jul. 03, 2021. doi: 10.1093/jhered/esr023.

BURCZYK, J. et al. Gene flow in forest trees: how far do genes really travel. **Forest Genetic**, v.11, p.1-14, 2004. Available from: <http://www.as.wvu.edu/~sdifazio/peer8.pdf>. Accessed: Out. 03, 2020.

DESTRO D.; MONTALVÁN, R. **Melhoramento Genético de Plantas**. Londrina: Ed. UEL, 1999. 820p.

DARRAULT, R. O.; SCHLINDWEIN, C. Polinização**.** In: SILVA JUNIOR, J. F.; LEDO, A. S. (Org.). **A cultura da mangabeira**. Aracaju: Embrapa Tabuleiros Costeiros, 2006. p. 43-56.

DUMINIL et al. Relationships between population density, fine-scale genetic structure, mating system and pollen dispersal in a timber tree from African rainforests. **Heredity,** v.116, p.295–303, 2016. Available from: <https://www.nature.com/articles/hdy2015101#citeas>. Accessed: Jul. 04, 2021. doi: 10.1038/hdy.2015.101.

ELLSTRAND, N.C. Gene flow among seed plant populations. In: ADAMS, W.T. et al. **Population genetics of forest trees**. Dordrecht: Forestry Sciences, Springer, v. 42, 1992, Available from: < https://link.springer.com/chapter/10.1007%2F978-94-011-2815-5_13>. Accessed: Jul. 03, 2021. doi: 10.1007/978-94-011-2815-5_13.

GAMFELDT, L.; KÄLLSTRÖM, B. Increasing intraspecific diversity increases predictability in population survival in the face of perturbations. **Oikos,** v.116, n.4, p.700-705, 2007. Available from: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0030-1299.2007.15382.x>. Accessed: Out. 07, 2020. doi: 10.1111/j.0030-1299.2007.15382x.

JULLIEN, M. et al. Structure of multilocus genetic diversity in predominantly selfing populations. **Heredity,** v.123, p.176–191, 2019. Available from: <https://www.nature.com/articles/s41437-019-0182-6#citeas>. Accessed: Jul 04, 2021. doi: 10.1038/s41437-019-0182-6.

KAMM, U. et al. Open areas in a landscape enhance pollen-mediated gene flow of a tree species: evidence from northern Switzerland. **Landscape Ecology**, v.25, p.903-911, 2010. Available from: <https://link.springer.com/article/10.1007/s10980-010-9468-z>. Accessed: Jul. 03, 2021. doi: 10.1007/s10980-010-9468-z.

KWAK, M. M. et al. Pollen and gene flow in fragmented habitats. **Applied Vegetation Science**, v.1, n.1, p.37–54, 1998. Available from: < https://www.jstor.org/stable/1479084>. Accessed: Jul. 03, 2021. doi:10.2307/1479084.

LEDERMAN, I. E. et al. Mangaba (*Hancornia speciosa* Gomez). In: LENDERMAN, I. E. **Série Frutas Nativas.** Jaboticabal: Funep, 2000. 35p.

LIU, F. et al. Genetic diversity of the endangered *Dalbergia odorifera* revealed by SSR markers. Available from: <https://www.mdpi.com/1999-4907/10/3/225>. Accessed: Jan. 15, 2021. **Forests,** v.10, n.225, p.1-17, 2019. doi:10.3390/f10030225.

MOURA, N.F. et al. Seleção de marcadores RAPD para o estudo da estrutura genética de populações de *Hancornia speciosa* Gomez. **Bioscience Journal,** v.21, p.119-125, 2005. Available from: <http://www.seer.ufu.br/index.php/biosciencejournal/article/view/6615>. Accessed: Nov. 25, 2020.

MURAWSKI, D.A. et al. Outcrossing rates of two endemic Shorea species from Sri Lankan tropical rain forests. **Biotropica**, v.26, n.1, p.23-29, 1994. Available from: <https://www.cabi.org/isc/abstract/19936793200>. Accessed: Dec. 20, 2020. doi: 10.2307/2389107.

NEI, M. **Molecular evolutionary genetics**. New York: Columbia University Press, 1978. 512p.

NEIGEL, J. E. A comparison of alternative strategies for estimating gene flow from genetic markers. **Annual Review of Ecology and Systematics,** v.28, n.1, p:105-28, 1997. Available from: <https://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.28.1.105>. Accessed: Jan. 29, 2020. doi: 10.1146/annurev.ecolsys.28.1.105.

SCHLAEPFER, D.R. et al. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. **Ecosphere**, v.9, n.10, e02488, 2018. Available

from: <https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecs2.2488>. Accessed: Jan. 23, 2021. doi: 10.1002/ecs2.2488.

SLATKIN, M. Gene flow in natural populations. **Annual Review of Ecology and Systematics**, v.16, p.393–430, 1985. Available from: <https://www.annualreviews.org/doi/abs/10.1146/annurev.es.16.110185.002141>. Accessed: Jul 04, 2021. doi: 10.1146/annurev.es.16.110185.002141.

SMITH, A.L et al. Global gene flow releases invasive plants from environmental constraints on genetic diversity. **PNAS,** v.117, n.8, p.4218–4227, 2020. Available from: <https://www.pnas.org/content/pnas/117/8/4218.full.pdf>. Accessed: Jul 04, 2021. doi: 10.1073/pnas.1915848117.

SOARES, L. A. S. S. et al. Anthropogenic disturbances eroding the genetic diversity of a threatened Palm tree: A multiscale approach. **Frontiers in Genetics**, v.10, p.1-12, 2019. Available from: <https://www.frontiersin.org/articles/10.3389/fgene.2019.01090/full>. Accessed: Jun 05, 2021. doi: 10.3389/fgene.2019.01090.

SOBIERAJSKI, G.R. et al. Sistema de reprodução em nove populações de *Mimosa scabrella* Bentham (Leguminosaceae). **Scientia Forestalis**, Piracicaba, n.71, p.37-49, 2006. Available from: <https://www.ipef.br/publicacoes/scientia/nr71/cap04.pdf>. Accessed: Jan. 23, 2021.

TAKAYAMA, S.; ISOGAI, A. Self-incompatibility in plants. **The Annual Review of Plant Biology**, v.56, p.467–89, 2005. doi: 10.1146/annurev.arplant.56.032604.144249. Available from: <https://pdfs.semanticscholar.org/725d/beaeff1bb6b5441887253ca8c4de71f3b5d3.pdf>. Accessed: Feb. 14, 2021. doi: 10.1146/ annurev.arplant.56.032604.144249.

VALDES, A.M. et al. Allele frequencies at microsatellite *loci*: the stepwise mutation model revisited. **Genetics**, v.133, n.3, p.737-749, 1993. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1205356/>. Accessed: May. 10, 2021.

VINSON, C.C. et al. 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. Available from: <https://pubmed.ncbi.nlm.nih.gov/25402015/>. Accessed: Feb. 10, 2021. doi.10.1111/mec.13002.

7 CAPÍTULO V- SIMULATING THE REPRODUCTIVE SYSTEM OF *HYMENAEA COURBARIL* L. POPULATIONS IN THE SOUTHWESTERN AMAZON FOREST

(Manuscrito submetido à revista *Pesquisa Florestal Brasileira* em 19/11/2021)

Abstract: In this study we used genetic structure data from three *Hymenaea courbaril* populations growing in the southwestern Amazon forest, obtained through microsatellite markers, to simulate their reproductive system. We tested different selfing, migration and clonal reproduction rates using the Easypop software and selected the combination that showed values of observed and expected heterozygosity closest to those previously obtained with microsatellite markers. According to our simulations, selfing was 0.1, migration rate was 0.1 and the clonal reproduction rate was 0 to 0.2. These results suggest the presence of both sexual and clonal reproduction in these populations. The populations showed a gradual decrease in genetic diversity due to low gene flow. To minimize the loss of genetic diversity we suggest expanding the area and replacing the explored trees using high genetic diversity germplasm. Our study stressed the importance of using computer simulations in genetic conservation

Keywords: computational simulations, conservation biology, Easypop, genetic diversity, microsatellite markers, reproductive mode.

INTRODUCTION

The Amazon rainforest is the largest continuous rainforest ecosystem in the world. It is characterized by high heterogeneity of habitats, soil and micro-climates, resulting in high biodiversity. This biodiversity has attracted worldwide attention regarding the sustainable exploitation and conservation of this ecosystem. However, sustainable exploitation requires research on the impact of forest management activities on the distribution, structure, genetic

diversity and ecology (pollination and dispersion) of potential timber species as deforestation of the Amazon advances at an alarming pace (Werth & Avissar, 2002, Silva et al. 2014).

Among several consequences, deforestation causes populations fragmentation and size reduction. Small-sized populations are more susceptible to genetic drift and inbreeding effects, leading to genetic diversity reduction. Genetic drift is a random process in which allele frequencies within a population fluctuate by chance, being capable of leading to extinction. Inbreeding is the mating of individuals that are closely related through common ancestry, leading to homozygosity (Matesanz et al. 2017).

Genetic variability is the key component for the sustainability of species, as it provides raw material for adaptation, evolution and survival under environmental changes. In the short term, populations with low genetic diversity have high susceptibility to diseases and pests, loss of incompatibility alleles and fixation of harmful alleles. In the long term, the loss of genetic diversity can lead to low ability of populations to respond to changes in selection pressures (Lowe et al. 2005).

The genetic diversity of populations is affected by the reproductive system. According to the prevailing sexual reproduction mode, natural populations of forest species can be classified as autogamous (when selfing prevails), allogamous (when outcrossing prevails) or mixed. Outcrossing favors the increase of genetic diversity through the recombination of genes while selfing and clonal reproduction reduce the genetic diversity.

In this sense, aspects related to the reproductive system should be considered in genetic diversity conservation programs, in addition to the gene flow, which is important to reduce inbreeding. In this study, we used Easypop (Balloux 2001) program to perform simulations of genetic data obtained through the use of microsatellite molecular markers by Silva et al. (2014),

in order to estimate the rates of selfing, migration and clonal reproduction of three *Hymenaea courbaril* populations in southwestern Amazon forest.

MATERIAL AND METHODS

We used the computer program Easypop version 2.0.1 (Balloux, 2001) to simulate different rates of selfing (0.1, 0.3 or 0.5), migration (0.1, 0.3, 0.5, 0.7 or 0.9) and clonal reproduction (0.0, 0.2, 0.4 or 0.6) in order to select the combination that would result in parameters similar to Silva et al. (2014) when studying the genetic structure of three *Hymenaea courbaril* L. populations through the use of microsatellite markers. The studied populations were sampled from Lábrea (AM), Capixaba (AC) and Porto Acre (AC), in the southwestern Amazon forest. In the referred study, the average observed heterozygosity was 0,6197, and the average expected heterozygosity was 0,6963.

For the simulations, we considered diploid hermaphrodite specie, with non-random mating and no clonal reproduction. A spatial migration model was considered, which coordinates were based on the geographic location of the populations, obtained from the real geographical coordinates (latitude and longitude) according to the map (Silva, unpublished data). With respect to the mutation settings, we assumed 10 loci evolving according to the single-step mutation model (SSM), with a proportion of 0.1 K-allele model (KAM) events, under 107 possible allelic states, according to the data obtained in the original study. A mutation rate of 0.0001 mutations per locus per generation was assumed. The genetic variability of the initial population was considered the maximum, and we simulated 100 generations. For each combination of selfing and migration rates, we replicated 100 replicates.

We selected the combination that showed values of observed and expected heterozygosity closest to those previously obtained with microsatellite markers.

RESULTS

Values of observed (Ho) and expected (He) heterozygosities generated from Easypop simulations for different selfing, migration and clonal reproduction rates are summarized in Table I. There is clear trend of increasing heterozygosity (both H_O and H_E) with increasing migration rate and reducing selfing and clonal reproduction. The closest values of Ho and He to those obtained through the use of microsatellite markers (0,6197 and 0,6963 respectively) were generated under the selfing of 0.1, the migration rate of 0.1 and the clonal reproduction rate of 0 to 0.2 (Table 1). These values were significantly different from the other rates, based on two independent sample t-tests at a 5% level of probability of error.

Table 1. Estimates of observed (H_O) and expected (H_E) heterozygosities generated by simulations, for different rates of selfing, migration and clonal reproduction.

Selfing rates	Migration rates	Clonal reproduction rates							
		0		0,2		0,4		0,6	
		H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E
0,1	0,1	0,654	0,693	0,653	0,693	0,656	0,691	0,650	0,681
	0,3	0,683	0,711	0,674	0,705	0,684	0,725	0,686	0,713
	0,5	0,689	0,721	0,685	0,724	0,685	0,721	0,689	0,722
	0,7	0,697	0,731	0,678	0,712	0,689	0,718	0,702	0,734
0,3	0,1	0,568	0,684	0,567	0,689	0,559	0,672	0,565	0,676
	0,3	0,630	0,711	0,563	0,686	0,585	0,695	0,550	0,669
	0,5	0,566	0,679	0,557	0,686	0,583	0,704	0,577	0,690

0,5	0,7	0,572	0,697	0,581	0,700	0,580	0,692	0,588	0,707
	0,1	0,418	0,628	0,417	0,621	0,445	0,657	0,423	0,616
	0,3	0,435	0,657	0,431	0,641	0,442	0,646	0,445	0,649
	0,5	0,436	0,656	0,442	0,655	0,443	0,658	0,452	0,668
	0,7	0,459	0,677	0,436	0,648	0,444	0,668	0,443	0,657

Gene flow enables the gene exchange among populations, thus increasing the genetic diversity (Madsen et al., 1995). On the other hand, selfing and clonal reproduction limit the gene flow and the potential for recombination, resulting in low genetic diversity in populations with these reproductive systems (Loveless & Hamrick 1987).

The simulation carried out over 100 generations for the selected model indicated a gradual reduction in the number of alleles (Figure 1) and heterozygosity (Figure 2). The number of alleles and heterozigosity are parameters of the genetic diversity usually used in genetic analyzes. These parameters enable assessing the conservation state and evaluating the effects of conservation programs.

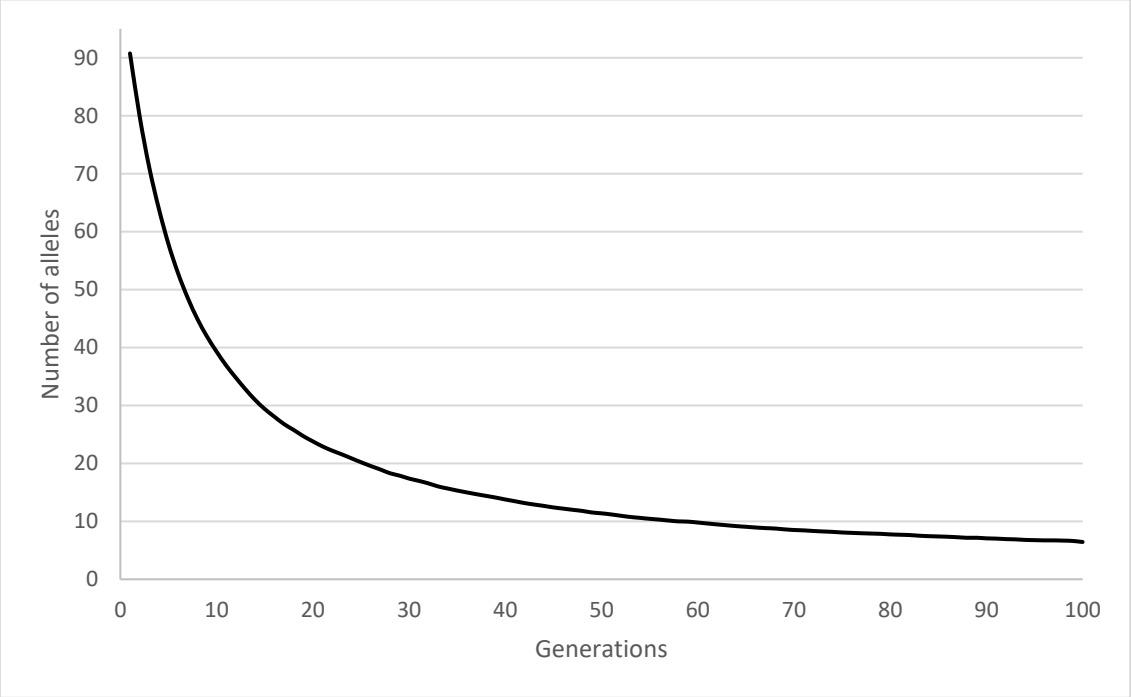


Figure 1. Simulation of number of alleles heterozygosities across 100 generations in three *Hymenaea courbaril* populations

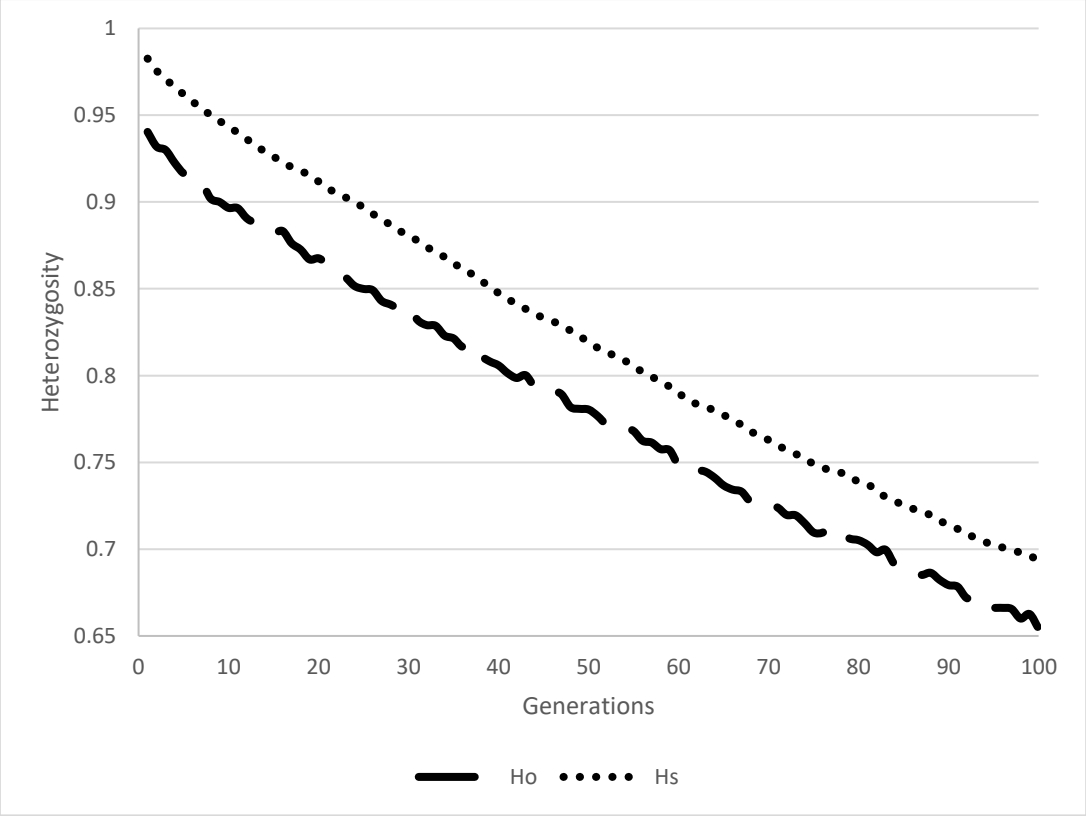


Figure 2. Simulation of observed (H_O) and expected (H_S) heterozygosities across 100 generations in three *Hymenaea courbaril* populations

DISCUSSION

Our results suggest evidence of natural clonal reproduction in *Hymenaea courbaril*, although it has not yet been officially registered. Clonal reproduction is common in natural forests (Silvertown 2008). It was registered in the congener species *Hymenaea stiginicarpa* (Moreno, unpublished data). This reproductive system refers to the reproduction by which new plants are produced from roots, rhizomes, stems, tubers, leaves and inflorescences.

Regarding sexual reproduction, the rate of selfing of 0.1 generated from simulations implies an outcrossing rate of 0.9. This finding enables us to classify the populations as mixed, with a predominance of outcrossing, according to the classification of Destro & Montalván (1999). This author classifies populations as autogamous (outcross from 0 to 0.05), mixed (outcross from 0.05 to 0.95) and allogamous (outcross from 0.95 to 1). Outcrossing in *H. courbaril* was described by Bawa (1974) and confirmed by Crestana et al. (1985). According to these authors, bats are the main pollinators of this species, in addition to some daytime visitors such as hymenopterans, dipterans and hummingbirds.

Mixed reproduction system in hermaphrodite vegetal species, like *Hymenaea courbaril*, suggests the presence of self-incompatibility mechanism in these species that reduce selfing. Self-incompatibility is a genetic mechanism frequently used by angiosperms to prevent inbreeding and the subsequent loss of genetic variability (Bawa 1974). In studying *Ceiba aesculifolia* populations, Quesada et al. 2013 observed outcrossing rates close to 1 due to a self-incompatibility mechanism present in this species.

There was low gene flow among populations due to the distances among them. *Hymenaea courbaril* has animal-mediated fruits and seeds dispersal, however, gene flow is limited by the distance among the populations (Silva et al. 2014).

In small-sized populations, there is increased loss of alleles by genetic drift and decreased heterozygosity through the increase in inbreeding level (Ellstrand & Elam 1993). Simulations carried out by Stefenon & Costa (2012) using the Easypop program indicated that smaller populations had greater increase in the inbreeding coefficient over generations, resulting in less genetic variability, compared to larger populations.

Another effect of fragmentation is the deviation of allele frequencies from the expected in the Hardy-Weinberg equilibrium. Under Hardy-Weinberg equilibrium, observed heterozygosity equals expected heterozygosity. However, inbreeding reduces observed heterozygosity from the expected heterozygosity (Ellstrand & Elam 1993) as shown in Figure 2.

For this reason, gene flow, a opposite force to genetic drift and inbreeding, has been explored in programs that prioritize genetic diversity such as restoration and conservation. In our previous work (Serrote et al. 2019), we simulated the effect of two extreme migration rates (0.1 and 0.9) on the genetic diversity of *Cabralea canjerana* fragments. The high migration rate homogenized the fragments, decreased the genetic differentiation coefficient among the fragments and increased the genetic diversity, in contrast to the low migration rate that increased inbreeding, increased genetic differentiation and reduced genetic diversity.

The populations of *Hymenaea courbaril* result from a long fragmentation process by anthropic activities, which resulted in a gradual loss of genetic diversity. The distances that currently separate these populations limit the interpopulational gene flow due to the distances covered by pollinators and fruit and seed dispersers. For this reason, strategies to contain the loss of genetic diversity should be local, for each population. One strategy could be expanding the area and replacing the explored trees using high genetic diversity germplasm. The results

of our study contribute to the sustainable management of *Hymenaea courbaril* through promoting the increase in genetic diversity.

CONCLUSION

Based in our simulation, the three *Hymenaea courbaril* populations reproduce both sexually and clonally. The main mode of sexual reproduction is through outcrossing (90%). The populations showed a gradual decrease in genetic diversity due to low gene flow (migration rate equal to 0.1). One strategy to minimize the loss of genetic diversity could be expanding the area and replacing the explored trees using high genetic diversity germplasm. This study stressed the importance of using computer simulations in genetic conservation.

REFERENCES

BALLOUX, F. 2001. Easypop (version 1.7). A computer program for the simulation of population genetics. J Hered 92: 301-302.

BAWA K. S. 1974. Breeding systems of tree species of a lowland tropical community. Evol 28: 85-92.

CRESTANA CSM, DIAS IS, MARIANO G. 1985. Ecologia de polinização de *Hymenaea stilbocarpa* Hayne, o Jatobá. Silvicultura em São Paulo 17: 31-37.

Destro D & Montalván R. 1999. Melhoramento Genético de Plantas. Londrina: Ed. UEL, 818p.

ELLSTRAND NC & ELAM DR. 1993. Population genetic consequences of small population size: Implications for plant conservation. Ann Rev Ecol Syst 24: 217-242.

LOVELESS MD & HAMRICK JL. 1987. Distribuición de la variación en species de árboles tropicales. Rev Biol Trop 35: 165-75.

LOWE AJ, BOSHIER D, WARD M, BACLES CFE & NAVARRO C. 2005. Genetic resource impacts of habitat loss and degradation; reconciling empirical evidence and predicted theory for neotropical trees. Heredity 95: 255-273.

MADSEN T, STILLE B & SHINE R. 1995. Inbreeding depression in an isolated population of adders *Vipera berus*. Biol Conserv 75: 113-118.

MATESANZ S, TESO MLR, GARCÍA-FERNANDEZ A & ESCUDERO A. 2017. Habitat fragmentation differentially affects genetic variation, phenotypic plasticity and survival in populations of a gypsum endemic. Frontiers in Plant Science 8:843.

QUESADA M, HERRERÍAS-DIEGO Y, LOBO JA, SÁNCHEZ-MONTOYA G, ROSAS F & AGUILAR R. 2013. Long-term effects of habitat fragmentation on mating patterns and gene flow of a tropical dry forest tree, *Ceiba aesculifolia* (Malvaceae: Bombacoideae). American Journal of Botany 100: 1095-101.

SERROTE CML, REINIGER LRS, COSTA LS, STEFANEL CM, SILVA KB & RABAIOLLI SMS. 2019. Simulation of pattern of gene flow in Canjerana fragments in the Brazilian Atlantic Rainforest for evaluating genetic conservation strategies. Cienc Rural 49: e20180764.

SILVA SMM, MARTINS K, MESQUITA AGG & WADT LHO. 2014. Parâmetros genéticos para a conservação de *Hymenaea courbaril* L. na Amazônia Sul-Ocidental. Cienc. Florest 24: 87-95.

SILVERTOWN J. 2008. The evolutionary maintenance of sexual reproduction: evidence from the ecological distribution of asexual reproduction in clonal plants. Int J Plant Sci 169: 157–168.

STEFENON VM & COSTA LS. 2012. A simulation study on the behavior of allelic richness and inbreeding coefficient over generations in fragmented populations of tree species. Annals of Forest Research 55: 3-10.

WERTH D & AVISSAR R. 2002. The local and global effects of Amazon deforestation. J Geophys Res 107: 8034.

8 CAPÍTULO VI - GENE FLOW AFFECTS THE GENETIC DIVERSITY OF *COPAIFERA LANGSDORFFII* DESF. IN CERRADO

(Manuscrito aceito na revista *Acta Botanica Brasilica*)

ABSTRACT

Gene flow allows connectivity of geographically isolated populations and which genetic diversity is reduced. Understanding how gene flow occurs is essential to conservation and selection of priority populations for intervention. In this study, the rates of migration and selfing of *Copaifera langsdorffii* Desf. sampled in Cerrado fragments in the State of São Paulo, Brazil, were estimated through computer simulations. Different selfing and migration rates were tested and the combination that showed values of observed and expected heterozygosity closest to those previously obtained with microsatellite markers was selected. According to the simulations, selfing and migration were low. Due to the high geographical distances among the fragments, gene flow was limited and may have led to low genetic diversity in the fragments. It is recommend enriching the area with high genetic diversity germplasm and reducing disturbance to promote the occurrence of pollinators, in order to increase the genetic diversity.

Keywords: fragmentation, reproductive isolation, genetic diversity, EASYPOP, microsatellites.

INTRODUCTION

Ecosystems vulnerability is an issue of great concern worldwide. Human disturbances, such as land clearing, negatively affect the environmental sustainability of forest ecosystems that lead to habitat fragmentation (Schlaepfer *et al*., 2018).

Among the consequences of habitat fragmentation, it propitiates the occurrence of inbreeding and genetic drift, especially when there is a lack of connectivity among fragments through gene flow, thus, reducing the genetic diversity, which is important for environmental sustainability (Duminil *et al*., 2016). For this reason, many conservation programs are focused on reducing gene flow barriers between isolated populations (Birnbaum *et al.*, 2013). Understanding the gene flow pattern between geographically isolated populations enables conservationists to plan conservation strategies, including the selection of priority populations for intervention (Ellstrand, 1992).

Moreover, genetic diversity is affected by the reproductive system of a species. Selfing limits pollen and seed dispersion and the potential for gene recombination, in contrast to crossing. For this reason, allogamous species are expected to preserve higher genetic diversity than autogamous species (Dias *et al.*, 2004).

In this study, the rates of gene flow and selfing of *Copaifera langsdorffii* trees located in Cerrado fragments were estimated through computer simulations.

Easypop version 2.0.1 (Balloux, 2001) was used to simulate different rates of selfing (0.1, 0.3 or 0.5) and migration (from 0.1 to 0.9, with steps of 0.1) of four *Copaifera langsdorffii* fragments. All fragments (Assis, Itirapina, Pedregulho and Brotas) are located in the Cerrado biome in the State of São Paulo, Brazil. Criteria used for selection were the values of observed and expected heterozygosity ($H_O = 0.68$; $H_E = 0.87$) reported by Antiqueira *et al*. (2014).

For the simulations, we considered a diploid hermaphrodite species, with non-random mating and no clonal reproduction. A spatial migration model was considered, with coordinates based on the geographic location (latitude and longitude) of the genotyped populations. Regarding the mutation settings, there were assumed eight loci evolving according to the single-step mutation model (SSM), with a proportion of 0.1 *K*-allele model (KAM) events, under 103

possible allelic states (Antiqueira *et al*., 2014). We considered a mutation rate of 0.0001 mutations per locus per generation and assumed 100 generations. Each combination of selfing and migration rates was replicated 100 times. The observed and expected heterozygosities from Antiqueira *et al*. (2014) were used to select the model settings that presented values closest to the field observations. The two independent sample *t*-tests at a 5% level of probability were used to compare the observed and expected heterozygosities of the selected combination of selfing and migration rates with other combinations.

According to the simulations, there is an increase in heterozygosity (observed and expected) as selfing reduces and migration rate increases. The closest values of H_O and H_E to those obtained using microsatellite markers (H_O = 0.68; H_E = 0.87) were obtained under the selfing of 0.3 and the migration rate of 0.1 and these values were significantly different from the other rates, based on two independent sample t-tests at a 5% level of probability of error (Figure 1).

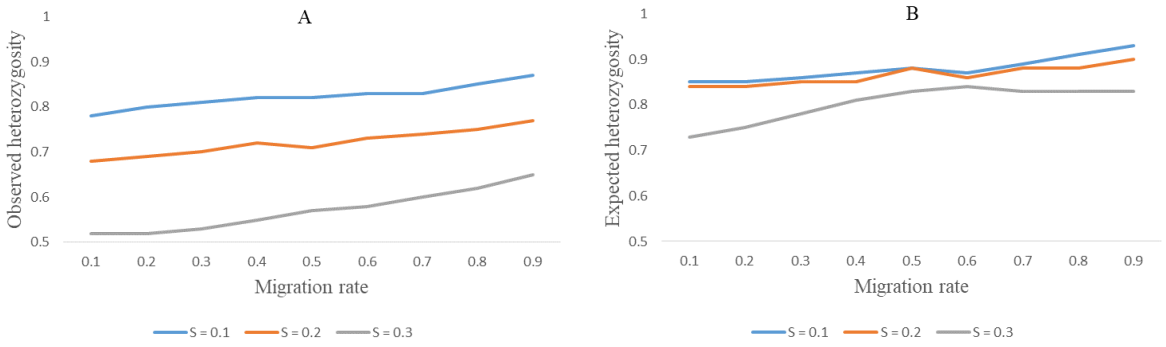


Figure 1. Estimates of observed (A) and expected (B) heterozygosities generated by simulations, for different rates of selfing (S) and migration.

The low selfing rate can be justified by the reproductive mode of this species. According to Carvalho (2003), *C. langsdorffii* is predominantly allogamous. Outcrossing in hermaphrodite plants, like *C. langsdorffii* is usually enforced by self-incompatibility, an important mechanism

in flowering plants that prevents selfing and consequently generating and mantaining the genetic diversity (Takayama & Isogai, 2005). Therefore, although most flowering plants are hermaphrodites, outcross prevails in the plant kingdom. In a study on the reproductive system of *Copaifera langsdorffii* from a preserved area in Minas Gerais, Brazil, Oliveira *et al*. (2002) found high rates of multilocus and unilocus outcrossing (0.917 and 0.877, respectively).

In regard to migration, there is low gene flow among the studied fragments. Gene flow via pollen in *Copaifera langsdorffii* is performed by bees and seed dispersal is zoochoric. The main animal seed dispersers are monkeys and birds, which both swallow arils and eventually regurgitate the seeds close to the mother plant, limiting gene flow (Martins *et al.*, 2008; Sebbenn *et al.*, 2011). Assis and Pedregulho are the most distant fragments (460 km away), followed by Assis – Itirapina (350 km), Pedregulho – Itirapina (280 km), Pedregulho – Brotas (280 km), Assis – Brotas (275 km) and Itirapina – Brotas (24 km) (Antiqueira *et al*., 2014). Thus, the long distances between the fragments and the dispersal syndrome in this species can justify the low gene flow obtained in our simulations.

According to Ellstrand (1992), gene flow in plant populations is an evolutionary force that promotes the homogenization of gene frequencies between populations in contrast to genetic drift that promotes genetic differentiation between populations. For this reason, various conservation programs focus on reducing barriers (physical or distance barriers) to gene flow in fragmented populations (Birnbaum *et al.*, 2013). In a study by Sebbenn *et al*. (2011), there was restricted gene flow in a fragmented population of *Copaifera langsdorffii* from the Brazilian Atlantic forest.

This study reinforces the importance of tree species genetic diversity in order to ensure its conservation. The studied populations of *C. langsdorffii* result from a long fragmentation process by anthropic activities. On the other hand, the long distances that separate these

populations limit the connectivity through gene flow. Therefore, local strategies must be adopted to contain genetic erosion, such as enriching the area with high genetic diversity germplasm and reducing disturbance to promote the occurrence of pollinators in these fragments.

REFERENCES

Antiqueira LMOR, Souza RGVC, Bajay MM, Kageyama PY. 2014. Genetic structure and diversity of *Copaifera langsdorffii* Desf. in Cerrado fragments of the São Paulo state, Brazil. Revista Árvore 38(4): 667-675.

Balloux F. EASYPOP (version 1.7). 2001. A computer program for the simulation of population genetics. Journal of Heredity 92(3): 301-302.

Birnbaum K, DeSalle R, Peters CM, Benfey PN. 2003. Integrating gene flow, crop biology, and farm management in on-farm conservation of avocado (*Persea americana*, Lauraceae). American Journal of Botany 90(11): 1619-1627.

Buschbom J, Yanbaev Y, Degen B. 2011. Efficient long-distance gene flow into an isolated relict oak stand. Journal of Heredity 102(4): 464-472.

Carvalho PER. 2003. Espécies Arbóreas Brasileiras. Coleção Espécies Arbóreas Brasileiras. Vol. I. Brasília, DF: Embrapa Informações Tecnológica; Colombo, PR: Embrapa Florestas. 1039 p.

Dias PMB, Dall'Agnol M, Schifino-Wittmann MT. 2004. Genetic diversity in the Brazilian species of Adesmia DC (Leguminosae) as assessed by RAPD. Plant Genetic Resources: Characterization and Utilization 2: 43-50.

Duminil J, Abessolo DTM, Bourobou DN, Doucet J-L, Loo J, Hardy OJ. 2016. High selfing rate, limited pollen dispersal and inbreeding depression in the emblematic African rain forest tree *Baillonella toxisperma* – Management implications. Forest Ecology and Management 379: 20-29.

Ellstrand NC. 1992. Gene flow by pollen: implications for plant conservation genetics. Oikos 63: 77-86.

Martins K, Dos Santos JD, Gaiotto FA, Moreno MA, Kageyama PY. 2008. Estrutura genética populacional de *Copaifera langsdorffii* Desf. (Leguminosae – Caesalpinioideae) em fragmentos florestais no Pontal do Paranapanema, SP, Brasil. Revista Brasileira de Botânica 31(1): 61-69.

Oliveira AF, Carvalho D, Rosado SCS. 2002. Taxa de cruzamento e sistema reprodutivo de uma população natural de *Copaifera langsdorffii* Desf. na região de Lavras (MG) por meio de isoenzimas. Brazilian Journal of Botany 25(3): 331-338.

Schlaepfer DR, Braschler B, Rusterholz H-P, Baur B. 2018. Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9(10): e02488.

Sebbenn AM, Carvalho ACM, Freitas MLM, Moraes SMB, Gaino APSC, da Silva JM, Jolivet C, Moraes MLT. 2011. Low levels of realized seed and pollen gene flow and strong spatial genetic structure in a small, isolated and fragmented population of the tropical tree *Copaifera langsdorffii* Desf. Heredity 106: 134-145.

Takayama S, Isogai A. 2005. Self-incompatibility in plants. The Annual Review of Plant Biology 56: 467–89.

9 CAPÍTULO VII - STUDY OF GENE FLOW IN *PROSOPIS RUBRIFLORA* REMNANTS FROM THE PANTANAL BIOME THROUGH COMPUTER SIMULATIONS

(Manuscrito submetido à revista *Biodiversidade Brasileira* em 17/08/2022)

ABSTRACT – Understanding the pattern of gene flow between isolated populations is crucial in forest conservation. In this study was estimated the pattern of gene flow in eight *Prosopis rubriflora* fragments from the Pantanal biome, in Brazil, through computer simulations. Different models and rates of migration were tested and selected the combination that showed values of observed and expected heterozygosity closest to those previously obtained with microsatellite markers. According to the simulations, gene flow is performed through the isolation by distance model in a rate of 0.3. The high geographical distances between the fragments may have limited the gene flow, resulting in a decrease of genetic diversity over generations. It is recommend adopting local strategies to contain genetic erosion, such as enriching the area with high genetic diversity germplasm and reducing disturbance to promote the occurrence of pollinators in these fragments.
Keywords**:** Forest fragmentation; genetic diversity; EASYPOP; microsatellites.

Estudo do fluxo gênico em remanescentes de *Prosopis rubriflora* do bioma Pantanal por meio de simulações computacionais

RESUMO – A compreensão do padrão de fluxo gênico entre populações isoladas é crucial para a conservação florestal. Neste estudo foi estimado o padrão de fluxo gênico em oito fragmentos de *Prosopis rubriflora* do bioma Pantanal, no Brasil, por meio de simulações computacionais. Foram testados diferentes modelos e taxas de migração e foi selecionada a combinação que apresentou valores de heterozigosidade observada e esperada mais próximos aos obtidos anteriormente com marcadores microssatélites. De acordo com a simulação, o fluxo gênico é efetivado segundo o modelo de isolamento por distância sob a taxa de 0,3. As grandes distâncias geográficas entre os fragmentos podem ter limitado o fluxo gênico, resultando na redução da diversidade genética ao longo das gerações. Recomenda-se a adoção de estratégias locais para

conter a erosão genética, como o enriquecimento da área com germoplasma de alta diversidade genética e reduzir perturbações para promover a ocorrência de polinizadores nesses fragmentos.
Palavras-chave: Fragmentação florestal; diversidade genética, EASYPOP, microssatélites.

Estudio del flujo de genes en remanentes de *Prosopis rubriflora* del bioma Pantanal mediante simulaciones por computadora

RESUMEN: Comprender el patrón del flujo de genes entre poblaciones aisladas es crucial para la conservación de los bosques. En este estudio, se estudió el patrón de flujo de genes en ocho fragmentos de *Prosopis rubriflora* del bioma Pantanal, en Brasil, mediante simulaciones por computadora. Se probaron diferentes modelos y tasas de migración y se seleccionó la combinación que presentaba valores de heterocigosidad observada y esperada más similar a los obtenidos previamente con marcadores microsatélites. Según la simulación, el flujo génico se realiza según el modelo de aislamiento por distancia a tasa de 0,3. Las grandes distancias geográficas entre los fragmentos pueden tener limitado el flujo de genes, lo que da como resultado una diversidad genética reducida a lo largo de las generaciones. Se recomienda la adopción de estrategias locales para contener la erosión genética, como enriquecer el área con germoplasma de alta diversidad genética y reducir las perturbaciones para promover la ocurrencia de polinizadores en estos fragmentos.

Palabras llave: Fragmentación de bosques; diversidad genética, EASYPOP, microsatélites.

Introduction

Human activities in forest ecosystems such as logging, conversion to agriculture, and road construction result in forest fragmentation. Due to their small size, forest fragments are more susceptible to the genetic drift and inbreeding effects than continuous forests. These factors are recognized as determinants to the biodiversity erosion. For this reason, special attention is required for fragmented forests (Slatkin, 1985, Pinho & Hey, 2010).

Connecting fragments through gene flow is an important issue in conservation genetics as it allows the sharing of genetic diversity, reducing the fragments vulnerability (Slatkin,

1985). Most conservation programs are focused on reducing gene flow barriers between isolated fragments. Improving the understanding on the gene flow pattern among geographically isolated populations allows conservationists to plan conservation strategies, including the selection of priority populations for intervention (Broquet *et al*., 2010).

Wright's island model is a conventional model used in population genetics to evaluate gene flow between populations. This model allows the estimation of the actual number of migrants (Nm) by means of the F_{ST} statistic for a set of populations (Wright, 1965). However, this model does not have realistic assumptions for all populations, such as the equilibrium between migration and genetic drift, if there is equal gene exchange between populations, and if all populations have equivalent sources of migrants. Also, the estimates obtained through this model do not reflect the contemporary variation in gene exchange between populations or current changes in the dispersal process.

For these reasons, more accurate methods to investigate the gene flow pattern for specific populations are needed to ensure the success of conservation programs. In this study, the EasyPop program (Balloux, 2001) was used to simulate models and rates of gene flow in eight fragments of the forest tree species *Prosopis rubriflora* growing in the Pantanal biome, in Brazil, using data obtained from microsatellite markers analyzes.

Material and Methods

Molecular data of eight *Prosopis rubriflora* fragments obtained through microsatellite markers by Alves *et al*. (2018) were used in simulation of migration model and rate using the Easypop version 2.0.1 program (Balloux, 2001). All fragments are located in the Pantanal biome in the Brazilian State of Mato Grosso do Sul (Figure 1; Table 1). Criteria used for selection were the values of observed and expected heterozygosity ($H_O = 0.56$; $H_E = 0.58$) reported by Alves *et al*. (2018).

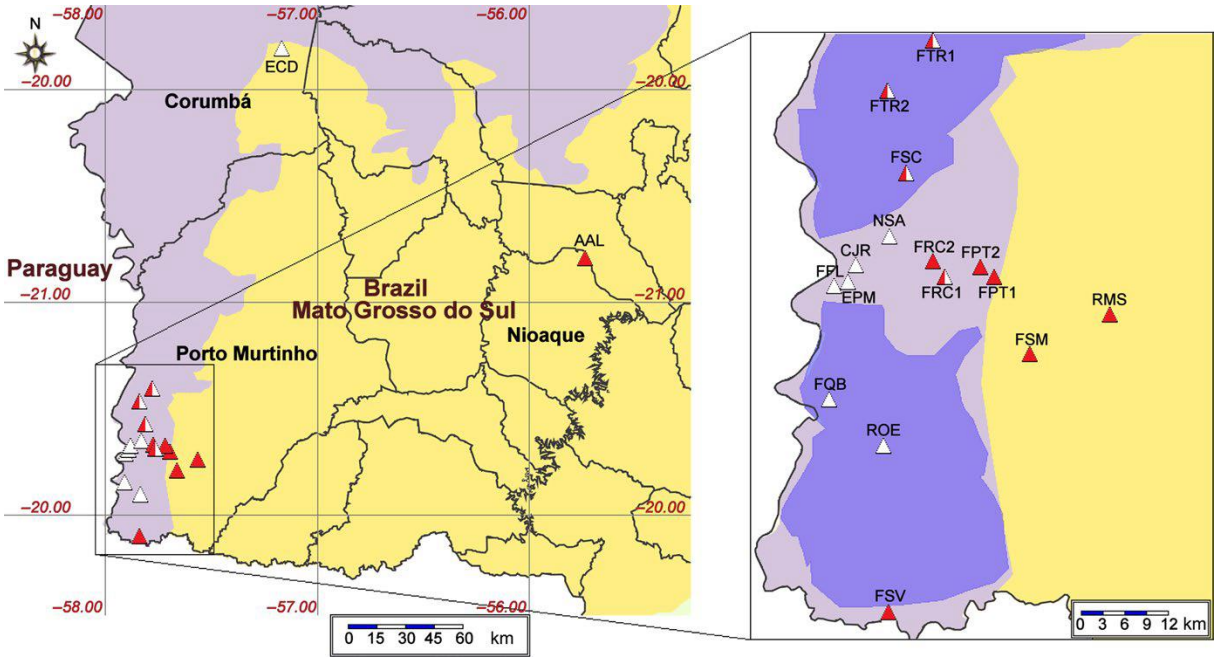


Figure 1 – Locations at which samples were collected. The sampled areas for *Prosopis rubriflora* are represented by red triangles. Only the purple areas, which represent the Pantanal domain, were used in the simulations. FPT1: Patolá Farm Area 1; FPT2: Patolá Farm Area 2; FSC: Santa Cristina Farm; FTR1: Tereré Farm Area 1; FTR2: Tereré Farm Area 2; FRC1: Retiro Conceição Farm Area 1; FRC2: Retiro Conceição Farm Area 2; FSV: Santa Vergínia Farm. Source: Alves et al. (2018).

Table 1 – Fragments at which samples of *Prosopis rubriflora* used in the simulations were collected.

Sampled area (initials)	Geographic coordinates	H_O	H_E
Patolá Farm Area 1 (FPT1)	57°42′11″W–21°42′09″S	0.40	0.49
Patolá Farm Area 2 (FPT2)	57°43′16″W–21°41′21″S	0.59	0.61
Santa Cristina Farm (FSC)	57°48′36″W–21°34′35″S	0.64	0.61
Tereré Farm Area 1 (FTR1)	57°46′43″W–21°24′42″S	0.61	0.60
Tereré Farm Area 2 (FTR2)	57°49′59″W–21°28′40″S	0.57	0.60
Retiro Conceição Farm Area 1 (FRC1)	57°45′49″W–21°42′08″S	0.53	0.57
Retiro Conceição Farm Area 2 (FRC2)	57°46′43″W–21°41′05″S	0.52	0.55
Santa Vergínia Farm (FSV)	57°50′01″W–21°06′42″S	0.58	0.61
Average		0.56	0.58

Source: Adapted from Alves et al. (2018).

There were tested different rates of migration (from 0.1 to 0.9 in steps of 0.2) and the following migration models: one-dimension stepping stone, island and spatial. For the spatial model simulation, two dimensions based on the real geographic position of the fragments were used. The following parameters were considered in all simulations: a dioic species, twenty loci, 400 generations, mixed Single Step Mutation Model (SSM) with a proportion of 0.1 *K*-allele model (KAM) events, under 100 possible allelic states (Alves *et al*., 2018). Each combination of selfing and migration rates was replicated 100 times. We selected the model that showed values of observed and expected heterozygosity closest to those previously obtained with microsatellite markers by Alves *et al*. (2018).

Results

Based on the observed and expected heterozygosity average values of 0.56 and 0.58, respectively, from Alves *et al*. (2018), gene flow among the studied fragments is performed through the spatial model in a rate of 0.3 (Table 2). The difference of statistics in migration rate 0.1 in relation to others is significant, by t test at 5% error probability.

Table 2 – Estimates of observed heterozygosity (H_O) and expected heterozygosity (H_E) generated by simulations for different models and migration rates in eight *P. rubriflora* fragments.

		Migration rates				
		0.1	0.3	0.5	0.7	0.9
	H_O	0.52	0.54	0.58	0.00	0.64
One-dimension stepping stone model	H_E	0.60	0.65	0.68	0.69	0.75
	H_O	0.43	0.48	0.51	0.57	0.60
Island model	H_E	0.60	0.60	0.63	0.67	0.70
	H_O	0.52	0.57	0.61	0.63	0.64
Spatial model	H_E	0.60	0.60	0.63	0.67	0.70

By simulating the behavior of the number of alleles over generations it is observed a clear trend of gene erosion due to fragmentation (Figure 2). Through fragmentation, large populations are split into smaller patches, propitiating the occurrence of inbreeding and genetic drift, resulting in the loss of the genetic diversity (Cuénin *et al*., 2019).

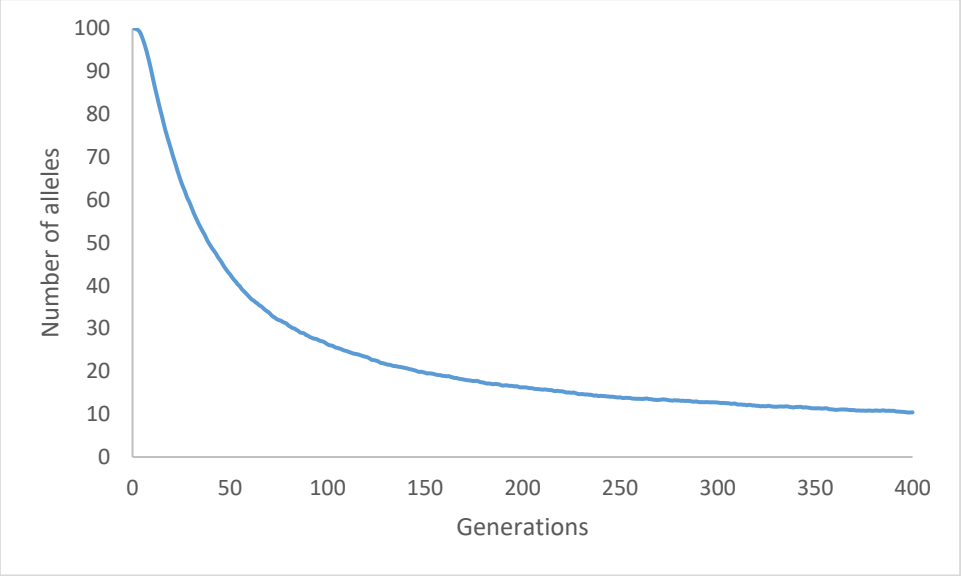


Figure 2 – Behavior of the number of alleles over generations for eight *Prosopis rubriflora* fragments from the Pantanal biome, Brazil.

Discussion

Isolationn by distance is a term used to describe patterns of population genetic variation that derive from limited gene flow. It is defined as a decrease in the genetic similarity among populations as the geographic distance between them increases. In other words, due to limited dispersal, individuals geographically closer tend to be genetically more similar than individuals that are far apart. As a consequence, populations can be structured in a hierarchical way and experience high genetic differentiation (Jensen *et al*., 2005; Meirmans, 2012).

Due to the isolation by distance model, the studied fragments present low rate of gene flow. Once gene flow is limited to closer fragments, there is a minimum possibility of gene exchange between Santa Vergínia Farm and the rest of the fragments. On the other hand, most of the gene flow is performed between Patolá Farm Area 1, Patolá Farm Area 2, Retiro Conceição Farm Area 1 and Retiro Conceição Farm Area 2, according to the distances between these fragments (Figure 1). According to Ellstrand & Ellam (1993), in a fragmented habitat, as distances

between patches increase, gene flow decreases, further increasing differentiation through higher inbreeding. This pattern was also observed in a study in *F. mauritiana* patches in a fragmented forest from the Mascarene Islands (Cuénin *et al*., 2019).

The estimation of gene flow in plant populations is complex because it is performed both through pollen and seeds (Ennos, 1994). Gene flow via pollen involves pollen dispersal from one population and the successful fertilization into flowers of another population; on the other hand, gene flow via seeds involves seeds dispersal from one population and their successful establishment in a new population (Ridley & Mallory-Smith, 2015). A complete description of gene flow between plant populations should ideally include the evaluation of the relative importance of both pollen and seeds. However, it is a very complex analysis, which makes computer simulation an important tool for studies on the gene flow pattern in forest populations.

Gene flow in is an evolutionary force that promotes homogenization of allelic frequencies between populations in contrast to genetic drift that promotes genetic differentiation. For this reason, it is fundamental reducing barriers to gene flow in fragmented populations for forest conservation (Ellstrand, 1992).

Simulations performed with the EASYPOP software showed an increase in inbreeding over generations in small populations compared to larger populations. However, random mating contributed to increase the heterozygosity in all generations. This result ratifies the importance of gene flow for genetic diversity conservation (Stefenon & Costa, 2012).

This study reinforces the importance of genetic diversity for forest conservation. The long distances that separate the studied fragments of *P. rubriflora* limit the connectivity through gene flow. Therefore, we propose local strategies to contain genetic erosion, such as enriching the area with high genetic diversity germplasm and reducing disturbance to promote the occurrence of pollinators in these fragments.

Conclusion

The *P. rubriflora* fragments showed continuous decrease in genetic diversity over generations due to the pattern of gene flow. These fragments exchange genes through the

isolation by distance in a rate of 0.3. In this condition, local strategies must be adopted to contain the loss of genetic diversity. We suggest expanding the areas using high genetic diversity germplasm and reducing disturbance to promote the occurrence of pollinators and seed dispersers.

References

Alves FM, et al. Genetic structure of two Prosopis species in Chaco areas: A lack of allelic diversity diagnosis and insights into the allelic conservation of the affected species. Ecology and Evolution, 8(13): 6558-6574, 2018.

Balloux F. EASYPOP (version 1.7). A computer program for the simulation of population genetics. Journal of Heredity, 92: 301-302, 2001.

Broquet T et al. Genetic bottlenecks driven by population disconnection. Conservation Biology, 24(6): 1596-1605, 2010.

Cuénin N, Flores O, Rivière E, Lebreton G, Reynaud B, Martos F. Great genetic diversity but high selfing rates and short-distance gene flow characterize populations of a tree (Foetidia; Lecythidaceae) in the fragmented tropical dry forest of the Mascarene Islands. Journal of Heredity, 110(3): 287–299, 2019.

Ellstrand NC. Gene flow by pollen: implications for plant conservation genetics. Oikos, 63: 77-86, 1992.

Ellstrand NC, Ellam DR. Population genetic consequences of small population size: implication for plant conservation. Annual Review of Ecology, Evolution, and Systematic, 24:217-242, 1993.

Ennos RA. Estimating the relative rates of pollen and seed migration among plant populations. Heredity, 72: 250-259, 1994.

Jensen JF, Bohonak AJ, Kelley STN. Isolation by distance, web service. BMC Genetics, 6:13, 2005.

Meirmans PG. The trouble with isolation by distance. Molecular Ecology, 21(12): 2839-2846, 2012.

Pinho C, Hey J. Divergence with gene flow: models and data. Annual Review of Ecology, Evolution, and Systematics 41: 215-230, 2010.

Ridley CE, Mallory-Smith C. 2015. Gene flow and Invasiveness in Bioenergy Systems. p:52-66. In: Quinn LD, Matlaga DP, Barney JN (orgs.). Bioenergy and biological invasions: ecological, agronomic, and policy perspectives on minimizing risk. CPI Group, Croydon. 170p.

Slatkin, M. Gene flow in natural populations. Annual Review of Ecology, Evolution, and Systematic 16: 393-430, 1985.

Stefenon VM, Costa LS. A simulation study on the behavior of allelic richness and inbreeding coefficient over generations in fragmented populations of tree species. Annals of Forest Research 55: 3-10, 2012.

Wright S. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19: 395-420, 1965.

10 DISCUSSÃO GERAL

Em face das crescentes ameaças à diversidade genética, o uso de tecnologias cada vez mais apuradas revela-se de extrema importância para contribuir na conservação dos recursos genéticos. A descoberta da técnica de marcadores moleculares revolucionou o campo de pesquisa em diversas áreas da genética, incluindo os estudos da diversidade genética. Nesse estudo buscou-se aprofundar a aplicabilidade de uma das classes de marcadores moleculares, os microssatélites (SSRs), em estudos visando a conservação de espécies florestais de diversos biomas brasileiros. Esses marcadores apresentam as vantagens de serem amplamente distribuídos nos genomas de eucariotos e de procariotos, possuírem herança co-dominante e elevadas taxas de mutação (GARRIDO-CARDENAS et al*.* 2018).

Assim, a revisão realizada na presente tese ratificou a ampla utilização de microssatélites em estudos envolvendo espécies florestais, porém, abaixo do esperado, devido a necessidade de conhecimento prévio do genoma da espécie a ser estudada (FERREIRA & GRATTAPAGLIA, 1998). Essa limitação restringe o uso de microssatélites a espécies com genoma pelo menos parcialmente sequenciado ou aquelas em que foram realizados testes de transferibilidade de outras espécies. Por outro lado, o baixo número de estudos com microssatélites em biomas brasileiros pode estar relacionado ao fato das espécies florestais serem pouco estudadas no Brasil, a redução de pesquisas no período da pandemia da COVID-19 e a existência de outras classes de marcadores. Ainda assim, os microssatélites se apresentam como sendo de extrema importância para a conservação de espécies florestais no Brasil.

Os estudos com marcadores moleculares permitem o monitoramento do comportamento da diversidade genética e permitem a recomendação de estratégias mais adequadas para a sua conservação. Entretanto, análises mais apuradas demandam a escolha de bons marcadores com base na sua capacidade de revelar polimorfismo entre indivíduos analisados, sendo a heterozigosidade (NEI; ROYCHOUDHURY, 1974) e o Conteúdo de Informação Polimórfica (PIC) (BOTSTEIN et al., 1980) os parâmetros mais usados. Assim, a revisão de literatura realizada na presente tese, relativa à determinação do PIC para marcadores moleculares, revelou-

se de extrema importância na medida em que irá auxiliar na qualidade das análises genéticas, através da escolha correta dos marcadores a usar, além de ter contribuído com uma equação simples para a determinação do PIC para marcadores dominantes, contribuindo assim na conservação dos recursos genéticos florestais.

Por outro lado, o desenvolvimento de programas de simulação de dados genéticos tornou mais efetivos os estudos sobre diversidade genética na medida em que permitem também predizer os efeitos dos fatores atuando sobre populações de espécies florestais (PENG et al., 2013). Nesse sentido, os estudos que serviram de base para as simulações na presente tese forneceram informações em relação a situação atual de espécies florestais em diversos biomas brasileiros, ao passo que as simulações forneceram estatísticas adicionais importantes para a tomadas de decisões visando a conservação.

Por exemplo, por meio das simulações, foi observado que o fluxo gênico entre fragmentos de *Cabralea canjerana* localizados na Mata Atlântica, entre remanescentes de *Prosopis rubriflora,* do bioma Pantanal, e entre fragmentos de *Copaifera langsdorffii*, do Cerrado, ocorre sob o modelo de isolamento por distância a uma taxa reduzida e que, em função disso, os níveis de diferenciação genética foram altos. Por seu turno, populações de *Hancornia speciosa* do bioma Caatinga foram classificadas, quanto ao modo reprodutivo, como mistas, com uma taxa de autofecundação de 0,3 e que trocavam genes a uma taxa de 0,5 a 0,6, fato que resultou em baixa diversidade genética. As simulações sugeriram também a presença de reprodução clonal em populações de *Hymenaea courbaril* localizadas na Amazônia Sul-Ocidental, fato previamente observado na espécie congênere *Hymenaea stiginicarpa* (MORENO, 2009).

Em função desses resultados foram realizadas recomendações tais como a criação de corredores ecológicos, que é a principal medida adotada em diversos estudos para prevenir ou restaurar populações subdivididas, pois conectam habitats isolados que passam a trocar genes como se uma única população de tamanho maior (BRUDVIG et al., 2009; CHRISTIE et al., 2015; LEMOS et al., 2015). Outras recomendações incluíram a manutenção ou expansão das áreas isoladas, o enriquecimento das áreas com germoplasma de alta diversidade genética e por forma a aumentar o tamanho e reduzir os efeitos da deriva genética e da endogamia,

substituição de árvores exploradas com germoplasma de elevada diversidade genética, redução da perturbação para promover a ocorrência de polinizadores, que promovem o aumento da diversidade genética através da promoção de cruzamentos.

Dessa forma, percebe-se a contribuição das simulações de dados de marcadores microssatélites na geração de parâmetros genéticos importantes para recomendar alternativas adequadas para a conservação de recursos florestais de biomas brasileiros. Espera-se que esta seja uma tendência mundial numa altura em que os níveis de desmatamento e consequente erosão genética seguem em ritmo acelerado.

11 CONSIDERAÇÕES FINAIS

A presente tese de doutorado objetivou, em complementação à dissertação de mestrado, aplicar a ferramenta de simulação de dados genéticos de populações de espécies florestais abrangendo os seis biomas brasileiros. Assim, em continuidade da simulação de um trabalho do bioma Pampa realizado no mestrado, na presente tese foram simulados dados dos biomas Amazônia, Mata Atlântica, Cerrado, Pantanal e Caatinga.

Os trabalhos usados na revisão e nas simulações da presente tese, os quais recorreram ao uso de marcadores microssatélites para as análises, confirmaram a crescente redução da diversidade genética nos biomas brasileiros como resultado do desmatamento acelerado. Esse fato ratifica a elevada potencialidade da contribuição dessa classe de marcadores de DNA para estudos genéticos com espécies florestais nesses biomas, como ferramenta para auxiliar em programas da conservação da biodiversidade por meio da estimação da variabilidade genética e sua distribuição entre e dentro de grupos ou populações. Ainda assim, o levantamento de estudos que fizeram o uso de microssatélites em biomas brasileiros revelou uma baixa quantidade de estudos no período considerado (2017 - 2021), fato que pode ser atribuido a redução de pesquisas no período da pandemia da COVID-19, a existência de outras classes de marcadores e ao fato da pesquisa com espécies florestais brasileiras ser negligenciada.

Por seu turno, os programas de simulação usados na tese (EasyPop e Fstat) mostraram-se eficientes na estimação de parâmetros como taxas de migração, de autofecundação, de reprodução clonal e modelos de migração, importantes para direcionar a adoção de estratégias mais efetivas de conservação dos recursos genéticos florestais. Assim, recomenda-se a inclusão de simulações nos programas de conservação florestal nos biomas brasileiros em complementação às análises com marcadores moleculares.

12 REFERÊNCIAS

BOTSTEIN, D.; WHITE, R. L.; SKALNICK, M. H.; DAVIES, R. W. Construction of a genetic linkage map in man using restriction fragment length polymorphism. **American Journal of Human Genetics**, v. 32, n. 3, p. 314-331. 1980. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1686077/>. Acesso em: 23 jun. 2020.

BRUDVIG, L.A.; DAMSCHEN, E.I.; TEWKSBURY, J.J.; HADDAD, N.M.; LEVEY, D.J. Landscape connectivity promotes plant biodiversity spillover into non-target habitats. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 23, p. 9328–9332, 2009. https://www.pnas.org/doi/10.1073/pnas.0809658106. Acesso em: 02 set. 2022.

CHRISTIE MR, KNOWLES LL. Habitat corridors facilitate genetic resilience irrespective of species dispersal abilities or population sizes. **Evolutionary Applications,** v. 8, n. 5, p. 454-463, 2015. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/eva.12255>. Acesso em: 02 set. 2022.

FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética.** 3. ed. Brasília: Embrapa-Cenargen, 1998, 220 p.

GARRIDO-CARDENAS JA, MESA-VALLE C, MANZANO-AGUGLIARO F. 2018. Trends in plant research using molecular markers. **Planta,** v. 247, p. 543–557. Disponível em: <https://link.springer.com/article/10.1007/s00425-017-2829-y>. Acesso em: 01 set. 2022.

LEMOS, R.P.M.; D'OLIVEIRA, C.B.; STEFENON, V.M. Genetic structure and internal gene flow in populations of *Schinus molle* (Anacardiaceae) in the Brazilian Pampa. **Tree Genetics & Genomes** v. 11, n. 75, 2015. Disponível em: < https://link.springer.com/article/10.1007/s11295-015-0885-7>. Acesso em: 02 set. 2022.

MORENO, M. A. **Estrutura genética e diversidade clonal de Jatobá-do-Cerrado (Hymenaea stiginicarpa Mart. ex Hayne) em duas populações no Cerrado do Estado de São Paulo.** 2009. 115 p. Dissertação (Mestrado em Recursos Florestais) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, SP, 2009.

NEI, M.; ROYCHOUDHURY, A. K. Sampling variances of heterozygosity and genetic distance. **Genetics** v. 76, n. 2, p. 379-390, 1974. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/4822472/>. Acesso em: 12 mar. 2021.

PENG, B. et al. Genetic Simulation Resources: a website for the registration and discovery of genetic data simulators. **Bioinformatics,** v.29, n.8, p.1101-1102, 2013. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624809/>. Accesso em: Jun 11, 2018.