UNIVERSIDADE ESTADUAL DE SANTA CRUZ

PROGRAMA DE PÓS-GRADUAÇÃO EM ECOLOGIA E CONSERVAÇÃO DA BIODIVERSIDADE

ALESANDRO SOUZA SANTOS

GENÉTICA DA CONSERVAÇÃO DE ÁRVORES: PERSPECTIVAS E COMPREENSÃO DA BASE GENÉTICA EM FLORESTAS TROPICAIS

ILHÉUS, BAHIA
2019

ALESANDRO SOUZA SANTOS

GENÉTICA DA CONSERVAÇÃO DE ÁRVORES: PERSPECTIVAS E COMPREENSÃO DA BASE GENÉTICA EM FLORESTAS TROPICAIS

Tese apresentada ao Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade da Universidade Estadual de Santa Cruz, como parte dos requisitos para obtenção do título de doutor.

Orientadora: Dra. Fernanda Amato Gaiotto

Ilhéus, Bahia,

2019

S237 Santos, Alesandro Souza.
Genética da conservação de árvores: perspectivas e compreensão da base genética em florestas tropicais / Alesandro Souza Santos. – Ilhéus, BA: UESC, 2019.
122f. : il.

Orientadora: Fernanda Amato Gaiotto
Tese (Doutorado) – Universidade Estadual de Santa Cruz. Programa de Pós-Graduação em Ecologia e Conservação da Biodiversidade.
Inclui referências e apêndice.

1. Florestas tropicais. 2. Mata atlântica. 3. Árvores. 4. Genômica. 5. Paisagem (Análise). 6. Marcadores genéticos. 7. Sequenciamento genético. 8. Sondas DNA. 9. Triagem genética. I. Título.

CDD 577.34

ALESANDRO SOUZA SANTOS

Bacharel em Ciências Biológicas

Mestre em Ecologia e Conservação da Biodiversidade

GENÉTICA DA CONSERVAÇÃO DE ÁRVORES: PERSPECTIVAS E COMPREENSÃO DA BASE GENÉTICA EM FLORESTAS TROPICAIS

Ilhéus, 31/07/2019

Banca avaliadora:

Dra. Fernanda Amato Gaiotto
Universidade Estadual de Santa Cruz – UESC
(Orientadora)

Dr. Leandro Lopes Loguercio
Universidade Estadual de Santa Cruz – UESC

Dr. Jose Carlos Morante-Filho
Universidade Estadual de Feira de Santana – UEFS

Dra. Flora Bittencourt Lima
Universidade Federal do Pará – UFPA

Dra. Jannaína Velasques da Costa Pinto
Universidade Federal do Sul da Bahia – UFSB

Dedico essa tese ao meu pai Manoel Teles e minha mãe Marilene Nascimento,
por tudo que vocês representam em minha vida.

AGRADECIMENTOS

Agradeço a Deus por ter abençoado meus passos, dando forças e determinação em todos os momentos da minha vida. Agradeço aos meus pais, Marilene e Manoel pelo amor, carinho, incentivos e por tantas outras coisas, que não caberiam nas 122 páginas dessa tese de doutorado. A conquista desse título de doutor em Ecologia teve contribuições de inúmeras pessoas. Mas, gostaria de expressar minha gratidão aos meus queridos irmãos: Arlene, Roque, Arlete, Horácio, Orlando e Alane. Vocês sempre estiveram comigo em todos os momentos, são irmãos sensacionais, que apesar da distância física, sempre estiveram e sempre estarão em meu coração. Agradeço também, aos meus tios Antônio e Marinalva, pelo carinho, amor e zelo que sempre tiveram comigo. Agradeço a toda minha família pelo apoio de sempre, em especial á minha querida prima Iraildes. Agradeço à minha companheira Thâmara, por tudo que você representa em minha vida, essa tese também é sua! Obrigado por fazer parte da minha vida, sem você, eu não teria conseguido chegar até o final do doutorado. Obrigado pela paciência em ouvir minhas lamentações quando as coisas não estavam dando certo (quase os 4 anos! rssrs), pelos incentivos, pelas intermináveis discussões científicas (até dormindo, rsrsrs) e, principalmente por seu amor. Agradeço às pessoas que direta ou indiretamente me ajudaram nessa caminhada, em especial aos amigos Ícaro Menezes e Eudes Oliveira pela amizade e apoio em todos os momentos.

Agradeço aos atuais membros e aos egressos do grupo de genética da conservação da UESC, pela amizade, discussões científicas, confraternizações e muitas resenhas compartilhadas nos últimos nove anos! Até parece que foi ontem que entrei pela primeira vez no laboratório de marcadores moleculares, onde aprendi todas as técnicas laboratoriais que sei e, aprendi desde um simples gel de agarose até as análises genéticas dos trabalhos da graduação, mestrado e doutorado. Todos vocês foram essenciais para minha formação e estiveram sempre presentes na minha caminhada, da época da graduação em biologia até o presente momento! Não irei citar nomes, porque foram muitos e todos fazem parte dessa conquista.

Gostaria de agradecer ao programa de pós-graduação em Ecologia e Conservação da Biodiversidade da UESC. Em especial ao corpo docente pelo ensinamento durante o mestrado e doutorado, compartilhando um pouco de seus

conhecimentos. Agradeço também, às secretárias Iky e Amábille, pela amizade, compreensão, pela agilidade e eficiência! Vocês foram essenciais neste processo.

Agradeço à Universidade Estadual de Santa Cruz (nº: 00220.1100.1681) e CNPq (nº: 426828/2016-0) pelo financiamento do projeto e CAPES pela concessão da bolsa de doutorado. À equipe da rede de pesquisa SISBIOTA, pela concepção teórica do desenho amostral do projeto, principalmente, à coordenadora Dra. Deborah Faria. Deixo também meus sinceros agradecimentos à Dra. Eliana Cazetta, por estar sempre aberta a discussões científicas e pela coordenação do financiamento do projeto. Por fim, agradeço à amizade e ensinamentos da minha querida orientadora Dra. Fernanda Gaiotto, que me recebeu de braços abertos no grupo de genética da conservação da UESC. Obrigado Fê, não tem palavras para expressar minha gratidão por tudo que você representa nessa caminhada, e obrigado por não ter deixado a "peteca cair" quando até eu mesmo duvidava, rssrsr.

Peço desculpas, caso não tenha realizado os merecidos agradecimentos a todas as pessoas que direta ou indiretamente estiveram presentes nessa caminhada! Mas, certamente, sou uma pessoa abençoada por Deus, por ter convivido e aprendido com tantas pessoas maravilhosas!

RESUMO

As florestas tropicais são reconhecidas globalmente como *hotspots* de biodiversidade. Porém, as atividades humanas relacionadas à ocupação e uso do solo têm ameaçado essa diversidade biológica. Dentro das regiões tropicais, o cenário de perda e fragmentação florestal é particularmente preocupante para a Mata Atlântica Brasileira, que possui elevada riqueza e endemismo de espécies e foi reduzida á cerca de 11-16% de sua extensão original. Em geral, a redução de floresta representa uma das principais ameaças, impactando negativamente a riqueza de espécies, processos ecológicos, grupos funcionais, fluxo gênico e a diversidade genética. No entanto, ainda existe uma grande lacuna de conhecimento para uma melhor compreensão dos efeitos antrópicos na biodiversidade tropical, se considerarmos o grande número de espécies presentes nesse ambiente, e que cada espécie pode responder de maneira distinta às perturbações. Nesse contexto, estudos que possibilitem acessar a variabilidade genética das populações naturais de espécies tropicais são extremamente importantes do ponto de vista ecológico e conservacionista, pois são capazes de detectar a fragilidade das populações diante de futuros distúrbios antrópicos. Assim, as análises de diversidade genética podem auxiliar a prever futuras respostas das espécies em ambientes modificados por ações humanas, contribuindo em estratégias mais eficientes de conservação. No presente trabalho, avaliamos a diversidade genética de duas espécies de plantas em remanescentes florestais de Mata Atlântica no Brasil. Além disso, descrevemos o estado da arte em estudos de genômica da paisagem de plantas silvestres, com diretrizes para subsidiar a melhor relação custo benefício para estudos futuros em florestas tropicais. No geral, evidenciamos com nossa revisão de literatura, uma grande lacuna acerca do conhecimento com a abordagem de genômica da paisagem para espécies de árvores tropicais. Além disso, a fim de otimizar a relação custo benefício dos estudos de genômica da paisagem, sugerimos utilizar a abordagem Pool-Seq para futuros estudos. No estudo empírico com ***Euterpe edulis*** **Mart (Arecaceae)**, as 17 populações na Bahia e as seis em São Paulo, apresentaram elevada diversidade gênica. Sendo, a maior diversidade reportada para os grupos de São Paulo se comparados aos grupos da Bahia. Acredita-se que essa diferença seja ocasionada pelas características demográficas dos diferentes morfotipos da espécie. Além disso, observamos que a porcentagem de floresta na paisagem nem a riqueza de aves dispersoras de sementes atuam como um *drive* importante para a diversidade gênica de *E. edulis,* e acreditamos que essa

diversidade seja moldada pela ampla distribuição da espécie na Mata Atlântica, o que lhe confere alta plasticidade fenotípica e adaptação local. No estudo empírico com ***Eschweilera ovata*** **(Cambess.) Mart. ex Miers (Lecythidaceae)**, os resultados obtidos com as análises utilizando cpSSR e nSSR indicam moderada diversidade genética e efeito de gargalo genético nas subpopulações avaliadas, indicando uma possível susceptibilidade ao processo de perda e fragmentação florestal. Por outro lado, as análises de estrutura genética revelaram o compartilhamento de pool gênico entre as subpopulações, o que pode ser um reflexo do fluxo gênico histórico que ocorreu antes da fragmentação florestal na região de estudo.

Palavras-Chave: Mata Atlântica, Árvores, Genômica da paisagem, cpSSR, SSR.

ABSTRACT

Tropical forests are globally recognized as a hotspot of biodiversity. However, human activities of land use and occupation have threat this biological diversity. In tropical regions, forest loss and fragmentation involve a particular concern for the Brazilian Atlantic Forest, which has a high species richness and endemism and has been reduced to about 11-16% of its original extension. In general, it has been observed in the literature that forest reduction represents one of the main threats, impacting negatively on species richness, ecological processes, functional groups, gene flow and genetic diversity. However, there is a knowledge gap for a better understanding the anthropic effects on tropical biodiversity, considering the large number of species present in Atlantic Forest, and that each species can respond differently to each disturbance. In this context, studies allowing the access to genetic variability of natural populations of tropical species are extremely important to ecological and conservationist purposes. Additionally, genetic parameters are useful to predict the fragility of populations facing threats in a scenario of anthropogenic disturbances, which are rapidly changing the environment. Thus, genetic studies can help to predict species responses in modified environments by human actions, contributing to design efficient conservation strategies. In the present research, we evaluated the genetic diversity of two plant species in forest remnants of Brazilian Atlantic Forest. In addition, we describe the state of the art in genomic studies of the landscape of wild plants, with guidelines to subsidize the best cost-benefit relation for future studies in tropical forests. Overall, our literature review evidenced a large gap in the knowledge with the approach of landscape genomics in tropical tree species. In addition, to optimize the cost-effectiveness of landscape genomics studies, we conclude suggesting the using of Pool-Seq approach for future studies. In our empirical study with ***Euterpe edulis*** **Mart (Arecaceae)**, 17 populations in Bahia and the six in São Paulo showed high genetic diversity. The greater genetic diversity was found for the groups from São Paulo when compared to the groups from Bahia. We believe that this difference is caused by some demographic characteristics and different morphotypes of the species. Furthermore, we observed that neither the percentage of forest in the landscape nor the richness of seed-dispersing birds act as an important drive for *E. edulis* gene diversity and we believe that this diversity is shaped by the wide distribution of the species in the Atlantic Forest, which gives it high phenotypic plasticity and local adaptation. Finally, in the empirical study with

***Eschweilera ovata* (Cambess.) Mart. ex Miers (Lecythidaceae)**, our results obtained by cpSSR and nSSR markers indicated moderate genetic diversity and bottleneck effect in the evaluated subpopulations, indicating a possible susceptibility to the process of forest loss and fragmentation. On the other hand, analyzes of genetic structure revealed the sharing of gene pool among subpopulations, which may be reflecting the historical gene flow, which probably occurred before the forest fragmentation in the study region.

Key words: Atlantic forest, Trees, Landscape genomics, cpSSR, SSR.

LISTA DE TABELAS

Capítulo 1

Table 1: Biological models and descriptive statistics calculated in 35 studies that assessed local adaptation in wild plant populations in situ using markers SNPs…………………………………………………………………………………49

Capítulo 2

Tabela 1: Porcentagem de cobertura florestal estimada nas 23 áreas amostradas no estudo**,** número de indivíduos coletados, estimativas de diversidade gênica (H_E) para os SNPs neutros e a riqueza de aves dispersoras de sementes de *E. edulis*...73

Capítulo 3

Table 1: Wilcoxon test for *E. ovata* subpopulations with nSSR markers……………………………………………………………………………..105

Table 2: *E. ovata* subpopulation genetic parameters based on the cpSSR markers……………………………………………………………………………105

Table 3: Genetic structure analysis (F_{ST}) with cpSSR markers between areas in pairs in the lower diagonal and geographic distance in upper diagonal (Km)……………………………………………………………………………......106

LISTA DE FIGURAS

Capítulo 1

Figure 1: Geographical, ecological and species distribution found in the 35 research articles, included in this study, which evaluated local adaptation in wild plant populations in their natural occurrence area using SNPs markers, are shown; (**A**) Percentage of studies published per continent; (**B**) Percentage of studies distributed per plant family; (**C**) The points distributed on the map indicating the sampling of the studies in the different terrestrial ecosystems……………………………………………51

Figure 2: Relation between sampling design and the detection of SNPs with environmental association **(A)** number of SNPs with environmental association as a function of the number of SNPs used in the studies; **(B)** number of SNPs with environmental association as a function of the total number of individuals in each study; **(C)** number of SNPs used as a function of the total number of individuals in each study. The colors of the dots represent the different methods used in the studies (Random used random regions of DNA and individualized sequencing; Random Pool-Seq used random regions of DNA and Pool-Seq technique; Targeted used only gene regions or expressed sequence tags and individualized sequencing; Targeted Pool-Seq used only gene regions or expressed sequence tags and sequencing in Pool-Seq)………..................................52

Capítulo 2

Figura 1: Distribuição das populações amostradas de E. edulis dentro dos estados da Bahia e São Paulo. **A)** Região sul da Bahia, porção Norte; **B)** Região sul da Bahia, porção Sul; **C)** Floresta Ombrófila (SP1_IN, SP1_CB) e Floresta Semidicidua (SP2_CA, SP2_CS, SP2_DI, SP2_SG) situadas no estado de São Paulo; **D)** Exemplos de paisagens com diferentes porcentagens de floresta e suas respectivas classes de uso da terra na escala de 2km...74

Figura 2: Distribuição das 23 populações de *E. edulis* com base nas frequências alélicas dos SNPs neutros, avaliadas por meio da Análise de Componentes Principais (PCA)..75

Figura 3: Boxplot da diversidade gênica com teste de média *ScottKnoot* para os quatros grupos formados pelas 23 populações de *E. edulis*..76

Capítulo 3

Figure 1: Map of Brazil. The state of Bahia is highlighted, indicating the region of the study, with the current forest cover. Blue and brown lines show network linkages identified by EDENetworks (Kivelä *et al.*, 2015) between nodes (sampling sites), using cpSSR markers. Line thickness is proportional to linkage strength and node size is proportional to the number of haplotypes for each sampling sites de *E. ovata*. Map data: Atlas of the Atlantic Forest Forest Remnants of the year 2016, obtained from SOS Mata Atlântica……………………………………………………………………...107

Figure 2: Allelic richness with standard deviation values for *E. ovata* subpopulations with nSSR markers……………………………………………………………………108

Figure 3: Observed heterozygosity (Ho) and expected heterozygosity (He) with standard deviation values for *E. ovata* subpopulations with nSSR markers……………………………………………………………………………………..109

Figure 4: Fixation index (Fis) estimated values with nSSR markers and 95 % confidence intervals based on 10.000 bootstrap values for *E. ovata* subpopulations …………………………………………………………………………110

Figure 5: F_{ST} estimated values with nSSR markers and 95 % confidence intervals based on 10.000 bootstrap values of *E. ovata* sampled in four subpopulations……………………………………………………………………..110

Figure 6: Genetic differentiation among the four *E. ovata* subpopulations based on nSSR markers. A) The graph represents the individuals as dots and the subpopulations

as ellipses according to their original groups, which model 95% of the corresponding variability plotted. The subpopulations (ellipses) are plotted within the orthogonal space defined by the first two eigenvalues of the PCA (inserts). B) Diagram representing the gene pool of the individuals associated with each of the subpopulations. Each vertical bar represents an individual with the height of the column segments showing the probability of being assigned to one of the four subpopulations. Dark lines correspond to the collection location……………………………………………………………111

Figure 7: Genetic diversity partitioned from the *E. ovata* subpopulations with cpSSR markers. Note: H_S = Genetic diversity within the population; D_{ST} = Diversity due to genetic differentiation; H_T = Total genetic diversity…………………………………112

Figure 8: Network analysis among the individuals of *E. ovata* with cpSSR markers. Grey and green lines show network linkages identified by EDENetworks (Kivelä *et al.*, 2015) between nodes (individuals). The colors of the nodes represent the sampling sites, the line thickness is proportional to linkage strength and node size is proportional to the number of linkages for each node……………………………………………………..113

SUMÁRIO

INTRODUÇÃO GERAL 17

REFERÊNCIAS 22

Capítulo 1 28

Landscape Genomics of Wild Plants: Insights on Sampling Design Strategies for Improving Cost-Benefit Ratio of Studies 29

Abstract 30

1. Introduction 31

2. Methodology 34

3. Results 37

4. Discussion 38

5. Guidance for study of local adaptation in wild plant species 42

References 44

LIST OF TABLE AND FIGURE 49

Capítulo 2 53

Diversidade genômica de uma palmeira tropical em paisagens antrópicas da Mata Atlântica 54

Resumo 55

1. Introdução 56

2. Metodologia 57

2.1. Área de estudo 57

2.2. Desenho amostral 58

2.3. Amostragem de *E. edulis* nas paisagens 59

2.4. Extração de DNA genômico e utilização de sondas para captura de sequências do genoma de *E. edulis* 60

2.5. Identificação de Polimorfismo de Nucleotídeo Único (SNP) 60

2.6. Detecção de marcadores SNPs *outliers* 61

2.7. Estimativa da diversidade genética e análises estatísticas 63

3. Resultados 63

4. Discussão 64

Referências 68

LISTA DE TABELA E FIGURAS 73

APÊNDICE 77

Capítulo 3 79

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 80

Abstract 81

1. Introduction 82

2. Material and Methods 84

2.1 Study area 84

2.2 DNA extraction, development of specific microsatellite primers and genotyping with cpSSR and nSSR 85

2.3 Analysis of nuclear microsatellite markers (nSSR) and population genetics ... 87

2.4 Population genetic analysis with chloroplastid microsatellite markers (cpSSR) 88

3. Results 89

3.1 Development of microsatellite markers nuclear (nSSR) for *E. ovata* and population analysis 89

3.2 Population analysis for *E. ovata* using microsatellite markers plastid (cpSSR) 90

4. Discussion 91

References 97

LIST OF TABLES AND FIGURES 105

APÊNDICE 114

CONCLUSÃO GERAL DA TESE 122

INTRODUÇÃO GERAL

A humanidade tem modificado drasticamente o meio ambiente, promovendo as principais mudanças que ocorrem atualmente no planeta e, por esta razão, defende-se a mudança de época geológica do holoceno para o antropoceno (CRUTZEN, 2006; STEFFEN et al. 2011; MCGILL et al. 2015). Essas alterações tem impactado de diversas maneiras o meio ambiente, seja pela exploração direta e predatória de seus recursos, seja pelas transformações nos padrões de biodiversidade e biogeoquímica global (LIU et al. 2013; ALBERTI 2015; DORRESTEIJN et al. 2015).

Dentre as ameaças provenientes das atividades antrópicas, pode-se citar a introdução de espécies exóticas, a poluição do ar, a defaunação e as mudanças climáticas globais que ameaçam a biodiversidade (BELLARD et al. 2012; BASHALKHANOV et al. 2013; GALETTI et al. 2013; PARMESAN & HANLEY, 2015; CARVALHO et al. 2016). Além disso, diversas atividades humanas relacionadas a ocupação e uso do solo, tem ocasionado uma rápida redução e fragmentação das florestas nos ecossistemas terrestres, provocando declínio da biodiversidade (HANSEN et al. 2013; HADDAD et al. 2015). Por esse motivo, a conversão desses sistemas naturais em antrópicos são as principais preocupações e alvos de estudos dos conservacionistas deste século, que buscam compreender como essas alterações impactam a biodiversidade (FAHRIG, 2003; MCGILL et al. 2015; PARMESAN & HANLEY, 2015; WILSON et al. 2016; CARVALHO et al. 2017).

Os impactos antrópicos associados a perda e fragmentação de florestas, tem atingido principalmente as regiões tropicais, que são reconhecidas globalmente pela sua grande biodiversidade (MANNION et al. 2014; SLIK et al. 2015). Acrescenta-se ainda, que esses ecossistemas tropicais estão em constante perigo, incluindo uma perda anual de 2.101 km^2 de cobertura florestal (HANSEN et al. 2013). Assim, a biodiversidade associada com esses ambientes corre sérios riscos de ser perdida (HANSEN et al. 2013; HADDAD et al. 2015; SLIK et al. 2015; BARLOW et al. 2018). Dentro das regiões tropicais, o cenário de destruição de florestas é particularmente preocupante para Mata Atlântica brasileira, que possui elevada riqueza e endemismo de espécies (THOMAS et al. 1998; MARTINI et al. 2007). Entretanto, esse domínio florestal encontra-se reduzido à aproximadamente 11-16% de sua extensão original, distribuídos em sua maioria em fragmentos pequenos e isolados, com grande empobrecimento de espécies (TABARELLI et al. 2004; RIBEIRO et al. 2009).

Nesse cenário atual de ameaça da Mata Atlântica, torna-se urgente os estudos que visam compreender os impactos gerados por essas ações antrópicas na biodiversidade. De modo geral, tem sido observado que a redução de floresta representa uma das principais ameaças, impactando negativamente a riqueza de espécies (RIGUEIRA et al. 2013; LIMA & MARIANO-NETO, 2014), os processos ecológicos (MORAN & CATTERALL, 2014), os grupos funcionais (MORANTE-FILHO et al. 2015), a diversidade filogenética (ANDRADE et al. 2015), o fluxo gênico e a diversidade genética (ZHANG et al. 2012; CARVALHO et al. 2015; SANTOS et al. 2016). Contudo, ainda existe uma grande lacuna de conhecimento, para uma melhor compreensão dos efeitos antrópicos na biodiversidade tropical, se considerar o grande número de espécies presente nesse ambiente, e que cada espécie, pode responder de maneira distinta a essa perturbação (MORANTE-FILHO et al. 2015).

Além disso, como as ações antrópicas estão modificando rapidamente o meio ambiente, existem muitas lacunas de conhecimento sobre como as espécies irão responder evolutivamente a tais mudanças e especialmente se serão capazes de ser adaptarem a esses ambientes modificados (REZNICK et al. 2001; KAWECKI & EBERT, 2004; DAVIS et al. 2005; ORR, 2005; LEIMU & FISCHER, 2008; MANEL et al. 2010; HOFFMANN et al. 2015; BENESTAN et al. 2016). Assim, os estudos que possibilitam acessar a variabilidade genética das populações naturais das espécies, são extremamente importantes do ponto de vista ecológico, evolutivo e conservacionista (KELLER et al. 2012; PARISOD & HOLDEREGGER, 2012; ĆALIĆ et al. 2016; CHRISTMAS et al. 2016; RELLSTAB et al. 20117; SHIH et al. 2018). Estes estudos são capazes de detectar a fragilidade das populações frente às ameaças em um cenário de distúrbios antrópicos (SPIELMAN et al. 2004; HANSEN et al. 2005; FITZPATRICK et al 2015; JARAMILLO-CORREA et al. 2015). Dessa maneira, a abordagem genética conservacionista, pode auxiliar na predição das respostas adaptativas a estes cenários, contribuindo em estratégias eficientes de conservação (STEANE et al. 2014; TORRES-FLOREZ et al. 2018).

De uma madeira geral, o acesso a variabilidade genética para espécies de plantas, tem sido feito através dos marcadores moleculares (FRANKHAM et al. 2002; AVISE, 2004). Sendo revolucionado através do grande avanço na coleta de dados genético no final da década de 80, com o desenvolvimento da técnica de PCR (Reação em cadeia da DNA polimerase) que possibilitou a amplificação do DNA com maior facilidade e precisão (MULLIS & FALOONA, 1987; SAIKI et al. 1988). Com o

surgimento dessa técnica, houve um aumento no uso dos marcadores moleculares, em especial do tipo microssatélites, possibilitando assim os estudos de genética da conservação (TORRES-FLOREZ et al. 2018). Essa popularidade deveu-se, principalmente, por se tratar de um marcador codominante e por possuir alto grau de polimorfismo, o que permite caracterizar melhor geneticamente as populações naturais (SCHLOTTERER, 2000; ELLEGREN, 2004). Além disso, os microssatélites são excelentes ferramentas para acessar a variabilidade genética proveniente de herança uni ou biparental (MARTINS et al. 2011; LI et al. 2012). Os marcadores baseados em herança biparental geralmente possuem maiores taxas de mutação se comparado com os marcadores uniparental (PROVAN et al. 1999). Com essas características de herança uni ou biparental é possível avaliar como eventos mais antigos (uniparental, ex: microssatélites com base no DNA cloroplastidial) e mais recentes (biparental, microssatélites com base no DNA nuclear) influenciaram e/ou influenciam a diversidade genética (MARTINS et al. 2011; ZHANG et al. 2012; SANTOS et al. 2019 in press).

Apesar dos marcadores microssatélites serem eficientes para estudos que visam caracterizar geneticamente as populações de uma dada espécie, eles são, em sua maioria, neutros, ou seja, não são diretamente influenciados pela ação da seleção natural (SCHLOTTERER, 2000; ELLEGREN, 2004). Por isso, esses marcadores são utilizados apenas para compreender e caracterizar os processos relacionados à deriva e dispersão nas populações de interesse. Nesse sentido, estudos que visam avaliar processos relacionados à seleção natural, devem utilizar outros marcadores genéticos, como por exemplo, os marcadores SNPs (Single Nucleotide Polymorphism) (STORFER et al. 2010; STEANE et al. 2014; HOFFMANN et al. 2015).

Os marcadores SNPs são amplamente distribuídos e com elevada frequência no genoma, podendo estar localizados em regiões neutras ou funcionais do genoma (INTERNATIONAL SNP MAP WORKING GROUP et al. 2001; SYVÄNEN et al. 2001). Por isso, são extremamente indicados para estudos de genética da conservação. Com esse tipo de marcador é possível fazer inferências relacionadas aos processos demográficos e de deriva genética (a partir de regiões neutras do genoma), assim como, de processos relacionados com a seleção natural e adaptações locais (regiões funcionais do genoma) (MANEL et al. 2010; PARISOD et al. 2012; MANEL et al. 2013; ĆALIĆ et al. 2016; FORESTER et al 2016; PLUESS et al. 2016). Inicialmente, esse tipo de ferramenta molecular era muito custosa e disponível apenas para espécies tidas como

modelos (MUIR et al. 2016). Contudo, devido ao surgimento e desenvolvimento de novas tecnologias de sequenciamento (por exemplo: NGS, *Next Generation Sequencing*), houve uma diminuição considerável dos custos relacionados a obtenção dos SNPs (MUIR et al. 2016), possibilitando a utilização de um grande número de SNPs, distribuídos em regiões neutras ou gênicas de espécies com pouco ou nenhum conhecimento acerca do genoma (CHRISTMAS et al. 2016). Desta forma, tornou-se possível mensurar com maior grau de precisão, como mudanças ocasionadas pelo homem estão impactando as espécies ao nível genômico (HANSEN et al. 2012; JARAMILLO-CORREA et al. 2015). Porém, estudos com abordagem genômica, ainda carecem de avanços metodológicos, principalmente no que tange às estratégias de amostragem e métodos de sequenciamento do genoma, para maximizar a relação custo benefício desses estudos em espécies não modelos, em especial para o contexto socioeconômico das regiões tropicais (LOTTERHOS & WHITLOCK, 2015; BENESTAN et al. 2016; MANEL et al. 2016; AHRENS et al. 2018; BARLOW et al. 2018; SANTOS & GAIOTTO in press).

No presente trabalho, foi realizada uma abordagem empírica de genética e genômica populacional e, uma revisão de genômica da paisagem, utilizando espécies de plantas como modelos biológicos. Como objetivo geral, buscou-se (i) avaliar a diversidade genética de duas espécies de plantas em remanescentes florestais de Mata Atlântica Brasleira; e (ii) descrever o estado da arte em estudos de genômica da paisagem de plantas silvestres, com diretrizes para subsidiar a melhor relação custo benefício para estudos futuros, na realidade socioeconômica das florestas tropicais. A presente tese foi desenvolvida em três capítulos, inseridos especificamente nas seguintes abordagens:

Capítulo 1: Realizou-se uma revisão para avaliar a importância do desenho amostral, em termos do (i) número de populações amostradas, (ii) número de indivíduos amostrados por população, (iii) número total de indivíduos amostrados e (iv) número de SNPs usados em diferentes estudos, no discernimento dos mecanismos moleculares subjacentes à adaptação local de espécies de plantas selvagens.

Capítulo 2: Avaliou-se como a redução de floresta em escala de paisagem e a riqueza de aves dispersoras de sementes influenciam na diversidade genômica de *Euterpe edulis* Mart (Arecaceae), usando SNPs inseridos em regiões gênicas.

Capítulo 3: Avaliou-se a diversidade genética de cinco subpopulações de *Eschweilera ovata* (Cambess.) Mart. ex Miers (Lecythidaceae), em remanescentes florestais de Mata Atlântica no sul da Bahia, por meio de marcadores microssatélites nucleares e cloroplastidiais.

REFERÊNCIAS

AHRENS, Collin W. et al. The search for loci under selection: trends, biases and progress. **Molecular ecology**, v. 27, n. 6, p. 1342-1356, 2018.

ALBERTI, Marina. Eco-evolutionary dynamics in an urbanizing planet. **Trends in Ecology & Evolution**, v. 30, n. 2, p. 114-126, 2015.

ANDRADE, Edila R. et al. Effects of habitat loss on taxonomic and phylogenetic diversity of understory Rubiaceae in Atlantic forest landscapes. **Forest Ecology and Management**, v. 349, p. 73-84, 2015.

AVISE, John C. **Molecular markers, natural history and evolution**, Second Edition edn. Sinauer Associates, Inc., Sunderland, MA 01375 USA, 2004.

BARLOW, Jos. et al. The future of hyperdiverse tropical ecosystems. **Nature**, v. 559, n. 7715, p. 517, 2018.

BASHALKHANOV, Stanislav; ECKERT, Andrew J.; RAJORA, Om P. Genetic signatures of natural selection in response to air pollution in red spruce (P icea rubens, Pinaceae). **Molecular ecology**, v. 22, n. 23, p. 5877-5889, 2013.

BELLARD, Céline. et al. Impacts of climate change on the future of biodiversity. **Ecology letters**, v. 15, n. 4, p. 365-377, 2012.

BENESTAN, Laura M. et al. Conservation genomics of natural and managed populations: building a conceptual and practical framework. **Molecular ecology**, v. 25, n. 13, p. 2967-2977, 2016.

ĆALIĆ, Irina. et al. Recent landscape genomics studies in forest trees—what can we believe?. **Tree genetics & genomes**, v. 12, n. 1, p. 3, 2016.

CARVALHO, Carolina S. et al. Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. **Heredity**, v. 115, n. 3, p. 216, 2015.

CARVALHO, Carolina S. et al. Defaunation leads to microevolutionary changes in a tropical palm. **Scientific Reports**, v. 6, p. 31957, 2016.

CARVALHO, Carolina S. et al. Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. **Conservation Genetics**, v. 18, n. 2, p. 467-478, 2017.

CHRISTMAS, Matthew J. et al. Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. **Molecular ecology**, v. 25, n. 17, p. 4216-4233, 2016.

CRUTZEN, Paul J. In: The “anthropocene”. In: **Earth system science in the anthropocene**. Springer, Berlin, Heidelberg, 2006. p. 13-18.

DAVIS, Margaret B.; SHAW, Ruth G.; ETTERSON, Julie R. Evolutionary responses to changing climate. **Ecology**, v. 86, n. 7, p. 1704-1714, 2005.

DORRESTEIJN, Ine et al. Incorporating anthropogenic effects into trophic ecology: predator–prey interactions in a human-dominated landscape. **Proc. R. Soc. B**, v. 282, n. 1814, p. 20151602, 2015.

ELLEGREN, Hans. Microsatellites: simple sequences with complex evolution. **Nature reviews genetics**, v. 5, n. 6, p. 435, 2004.

FAHRIG, Lenore. Effects of habitat fragmentation on biodiversity. **Annual review of ecology, evolution, and systematics**, v. 34, n. 1, p. 487-515, 2003.

FITZPATRICK, Matthew C.; KELLER, Stephen R. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. **Ecology Letters**, v. 18, n. 1, p. 1-16, 2015.

FORESTER, Brenna R. et al. Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. **Molecular ecology**, v. 25, n. 1, p. 104-120, 2016.

FRANKHAM, Richard; BRISCOE, David A.; BALLOU, Jonathan D. **Introduction to conservation genetics**. Cambridge university press, 2002.

GALETTI, Mauro. et al. Functional extinction of birds drives rapid evolutionary changes in seed size. **Science**, v. 340, n. 6136, p. 1086-1090, 2013.

HADDAD, Nick M. et al. Habitat fragmentation and its lasting impact on Earth's ecosystems. **Science Advances**, v. 1, n. 2, p. e1500052, 2015.

HANSEN, Matthew C. et al. High-resolution global maps of 21st-century forest cover change. **science**, v. 342, n. 6160, p. 850-853, 2013.

HANSEN, Michael M. et al. Monitoring adaptive genetic responses to environmental change. **Molecular ecology**, v. 21, n. 6, p. 1311-1329, 2012.

HOFFMANN, Ary et al. A framework for incorporating evolutionary genomics into biodiversity conservation and management. **Climate Change Responses**, v. 2, n. 1, p. 1, 2015.

INTERNATIONAL SNP MAP WORKING GROUP et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. **Nature**, v. 409, n. 6822, p. 928, 2001.

JARAMILLO-CORREA, Juan-Pablo. et al. Molecular proxies for climate maladaptation in a long-lived tree (Pinus pinaster Aiton, Pinaceae). **Genetics**, v. 199, n. 3, p. 793-807, 2015.

KAWECKI, Tadeusz J.; EBERT, Dieter. Conceptual issues in local adaptation. **Ecology letters**, v. 7, n. 12, p. 1225-1241, 2004.

KELLER, Stephen R. et al. Local adaptation in the flowering-time gene network of balsam poplar, Populus balsamifera L. **Molecular biology and evolution**, v. 29, n. 10, p. 3143-3152, 2012.

LEIMU, Roosa; FISCHER, Markus. A meta-analysis of local adaptation in plants. **PloS one**, v. 3, n. 12, p. e4010, 2008.

LI, Junmin; JIN, Zexin; TAN, Tan. Genetic diversity and differentiation of Sinocalycanthus chinensis populations revealed by chloroplast microsatellite (cpSSRs) markers. **Biochemical systematics and ecology**, v. 41, p. 48-54, 2012.

LIMA, Marília M; MARIANO-NETO, Eduardo. Extinction thresholds for Sapotaceae due to forest cover in Atlantic Forest landscapes. **Forest ecology and management**, v. 312, p. 260-270, 2014.

LIU, Bingliang. et al. Anthropogenic halo disturbances alter landscape and plant richness: A ripple effect. **PloS one**, v. 8, n. 2, p. e56109, 2013.

LOTTERHOS, Katie E.; WHITLOCK, Michael C. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. **Molecular ecology**, v. 24, n. 5, p. 1031-1046, 2015.

MANEL, Stéphanie. et al. Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. **Molecular Ecology**, v. 19, n. 17, p. 3760-3772, 2010.

MANEL, Stéphanie.; HOLDEREGGER, Rolf. Ten years of landscape genetics. **Trends in ecology & evolution**, v. 28, n. 10, p. 614-621, 2013.

MANEL, Stéphanie. et al. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. **Molecular ecology**, v. 25, n. 1, p. 170-184, 2016.

MANNION, Philip D. et al. The latitudinal biodiversity gradient through deep time. **Trends in Ecology & Evolution**, v. 29, n. 1, p. 42-50, 2014.

MARTINI, Adriana M Z. et al. A hot-point within a hot-spot: a high diversity site in Brazil's Atlantic Forest. **Biodiversity and Conservation**, v. 16, n. 11, p. 3111-3128, 2007.

MARTINS, K. et al. Genetic structure based on nuclear and chloroplast microsatellite loci of Solanum lycocarpum A. St. Hil.(Solanaceae) in Central Brazil. **Genet Mol Res**, v. 10, p. 665-677, 2011.

MCGILL, Brian J. et al. Fifteen forms of biodiversity trend in the Anthropocene. **Trends in ecology & evolution**, v. 30, n. 2, p. 104-113, 2015.

MORAN, Cath.; CATTERALL, Carla P. Responses of seed-dispersing birds to amount of rainforest in the landscape around fragments. **Conservation biology**, v. 28, n. 2, p. 551-560, 2014.

MORANTE-FILHO, José Carlos. et al. Birds in anthropogenic landscapes: the responses of ecological groups to forest loss in the Brazilian Atlantic Forest. **PLoS One**, v. 10, n. 6, p. e0128923, 2015.

MUIR, Paul. et al. The real cost of sequencing: scaling computation to keep pace with data generation. **Genome biology**, v. 17, n. 1, p. 53, 2016.

MULLIS, Kary B.; FALOONA, Fred A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. **Methods in enzymology**. v.155, p. 335–350,1987.

ORR, H. Allen. The genetic theory of adaptation: a brief history. **Nature Reviews Genetics**, v. 6, n. 2, p. 119, 2005.

PARISOD, Christian.; HOLDEREGGER, Rolf. Adaptive landscape genetics: pitfalls and benefits. **Molecular ecology**, v. 21, n. 15, p. 3644-3646, 2012.

PARMESAN, Camille.; HANLEY, Mick E. Plants and climate change: complexities and surprises. **Annals of botany**, v. 116, n. 6, p. 849-864, 2015.

PLUESS, Andrea R. et al. Genome–environment association study suggests local adaptation to climate at the regional scale in Fagus sylvatica. **New Phytologist**, v. 210, n. 2, p. 589-601, 2016.

PROVAN, Jim. et al. A low mutation rate for chloroplast microsatellites. **Genetics**, v. 153, n. 2, p. 943-947, 1999.

RELLSTAB, Christian. et al. Local adaptation (mostly) remains local: reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. **Heredity**, v. 118, n. 2, p. 193, 2017.

REZNICK, David N.; GHALAMBOR, Cameron K. The population ecology of contemporary adaptations: what empirical studies reveal about the conditions that promote adaptive evolution. In: **Microevolution Rate, Pattern, Process**. Springer, Dordrecht, 2001. p. 183-198.

RIGUEIRA, Dary M G.; DA ROCHA, Pedro L B.; MARIANO-NETO, Eduardo. Forest cover, extinction thresholds and time lags in woody plants (Myrtaceae) in the Brazilian Atlantic Forest: resources for conservation. **Biodiversity and conservation**, v. 22, n. 13-14, p. 3141-3163, 2013.

RIBEIRO Milton C.; Metzger JP.; Martensen AC.; Ponzoni FJ.; Hirota MM. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. **Biol Conserv** 142:1141–1153. doi: 10.1016/j.biocon.2009.02.021, 2009

SAIKI, Randall K. et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. **Science**, v. 239, n. 4839, p. 487-491, 1988.

SANTOS, Alesandro S. et al. Landscape-scale deforestation decreases gene flow distance of a keystone tropical palm, Euterpe edulis Mart (Arecaceae). **Ecology and evolution**, v. 6, n. 18, p. 6586-6598, 2016.

SCHLÖTTERER, Christian. Evolutionary dynamics of microsatellite DNA. **Chromosoma**, v. 109, n. 6, p. 365-371, 2000.

SHIH, Kai-Ming. et al. Adaptive genetic divergence despite significant isolation-by-distance in populations of taiwan cow-tail fir (Keteleeria davidiana var. formosana). **Frontiers in plant science**, v. 9, p. 92, 2018.

SLIK, JW Ferry. et al. An estimate of the number of tropical tree species. **Proceedings of the National Academy of Sciences**, v. 112, n. 24, p. 7472-7477, 2015.

SPIELMAN, Derek; BROOK, Barry W.; FRANKHAM, Richard. Most species are not driven to extinction before genetic factors impact them. **Proceedings of the National Academy of Sciences**, v. 101, n. 42, p. 15261-15264, 2004.

STEANE, Dorothy A. et al. Genome-wide scans detect adaptation to aridity in a widespread forest tree species. **Molecular Ecology**, v. 23, n. 10, p. 2500-2513, 2014.

STEFFEN, Will. et al. The Anthropocene: conceptual and historical perspectives. **Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences**, v. 369, n. 1938, p. 842-867, 2011.

STORFER, Andrew et al. Landscape genetics: where are we now?. **Molecular ecology**, v. 19, n. 17, p. 3496-3514, 2010.

SYVÄNEN, Ann-Christine. Accessing genetic variation: genotyping single nucleotide polymorphisms. **Nature Reviews Genetics**, v. 2, n. 12, p. 930, 2001.

TABARELLI, Marcelo.; DA SILVA, José Maria C.; GASCON, Claude. Forest fragmentation, synergisms and the impoverishment of neotropical forests. **Biodiversity & Conservation**, v. 13, n. 7, p. 1419-1425, 2004.

THOMAS, Wm Wayt. et al. Plant endemism in two forests in southern Bahia, Brazil. **Biodiversity & Conservation**, v. 7, n. 3, p. 311-322, 1998.

TORRES-FLOREZ, Juan P. et al. The coming of age of conservation genetics in Latin America: what has been achieved and what needs to be done. **Conservation Genetics**, v. 19, n. 1, p. 1-15, 2018.

WILSON, Maxwell C. et. al. Habitat fragmentation and biodiversity conservation: key findings and future challenges. **Landscape Ecology**. v. 31, n. 2. p. 219–227, 2016.

ZHANG, Xin. et al. Habitat loss other than fragmentation per se decreased nuclear and chloroplast genetic diversity in a monoecious tree. **PLoS One**, v. 7, n. 6, p. e39146, 2012.

Capítulo 1

Landscape Genomics of Wild Plants: Insights on Sampling Design Strategies for Improving Cost-Benefit Ratio of Studies

Citação

Santos, AS and Gaiotto FA. **Landscape Genomics of Wild Plants: Insights on Sampling Design Strategies for Improving Cost-Benefit Ratio of Studies.** Submitted for publication in Scientific Reports.

Landscape Genomics of Wild Plants: Insights on Sampling Design Strategies for Improving Cost-Benefit Ratio of Studies

Alesandro Souza Santos* and Fernanda Amato Gaiotto

Alesandro Souza Santos* and Fernanda Amato Gaiotto

Applied Ecology & Conservation Lab, Santa Cruz State University - Rodovia Ilhéus-Itabuna, km 16 Ilheus, BA, Brazil - ZIP Code: 45662-901.

*****Corresponding author:

Alesandro Souza Santos

alesandrouesc@gmail.com

Abstract

To avoid local extinction due to the changes in their natural ecosystems, introduced by anthropogenic activities, species undergo local adaptation. Landscape genomics through genome–environment association studies, has helped evaluate the local adaptation in natural populations. However this approach is still a developing discipline, requiring refinement of guidelines in sampling design, especially for studies conducted in the backdrop of stark socioeconomic realities of the rainforest ecologies. In this study we aimed to devise strategies to improve the cost-benefit ratio of landscape genomics studies by surveying sampling designs and genome sequencing strategies used in existing studies. We conducted meta-analyses to evaluate the importance of sampling designs, in terms of (i) number of populations sampled, (ii) number of individuals sampled per population, (iii) total number of individuals sampled, and (iv) number of SNPs used in different studies, in discerning the molecular mechanisms underlying local adaptation of wild plant species. Using the linear mixed effects model, we demonstrated that the total number of individuals sampled and the number of SNPs used, significantly influenced the detection of loci underlying the local adaptation. Thus, based on our findings, in order to optimize the cost-benefit ratio of landscape genomics studies, we suggest focusing on increasing the total number of individuals sampled and using a targeted (e.g. sequencing capture) Pool-Seq approach and/or a random (e.g. RAD-Seq) Pool-Seq approach to detect SNPs and identify SNPs under selection for a given environmental cline. We also found that the existing molecular evidences are inadequate in predicting the local adaptations to climate change in tropical forest ecosystems.

Keywords: Rainforest, climate changes, local adaptation, SNPs, conservation genomics.

1. Introduction

Anthropogenic activities are transforming natural systems, drastically changing the environmental conditions in a way which poses major threat to global biodiversity (Liu et al. 2013; Wilson et al. 2016). Anthropogenic modifications can lead to reduction and fragmentation of natural environments, or to changes in climatic conditions (Fahrig 2003; Bellard et al. 2012; Parmesan and Hanley 2015). Species respond to such changes in the natural habitat, by (i) phenotypic plasticity, (ii) migrating from their natural habitats, in search of conditions, fit for survival and having ample resources, or (iii) adapting to the new environment to avoid local extinction (local adaptation) (Davis et al. 2005; Nicotra et al. 2010; Bellard et al. 2012; Alberto et al. 2013).

Faced with local disturbances, a species may exhibit local adaptation, when individuals on an average have a superior fitness in their home environment, compared to a transplanted individual (Williams 1966; Reznick and Ghalambor 2001; Kawecki and Ebert 2004; Alberto et al. 2013; Pluess et al. 2016; Rellstab et al. 2017). Therefore, local adaptation is driven by the action of natural selection, which acts on the individual's phenotypes, determining the characteristics that will be favored under certain environmental conditions (Reznick and Ghalambor 2001; Kawecki and Ebert 2004; Rellstab et al. 2016; Roschanski et al. 2016). Historically, local adaptation has been studied either through translocation experiments, especially between environments, or trials in the greenhouse under controlled environmental conditions (Alberto et al. 2013; Steane et al. 2014; Rellstab et al. 2015). The drawbacks of these

two approaches include the requirement of ample financial resources and time, which is generally scarce for studies involving long-lived species such as trees (Steane et al. 2014; Rellstab et al. 2015). Landscape genomics, a study that identifies genetic variations that confer local adaptation, is used to remedy these limitations (Manel et al. 2010). In this approach, significant differences in allele frequency between populations of the target species, indicates that individuals in the population are experiencing selection pressure; possibly in response to change in some environmental factor, such as changes in soil type, radiation, water stress, and temperature (Turner et al. 2010; Di Pierro et al. 2016; Mosca et al. 2016; Sork et al. 2016). Therefore, landscape genomics studies analyze frequency distribution changes in molecular markers, such as single nucleotide polymorphism (SNPs), (Steane et al. 2014; Ćalić et al. 2016) in relation to given environmental factors. The SNPs are commonly used in wild local adaptation studies because their location and functional annotations are known, and they are widely distributed throughout the genome (Gaut 2012; Fischer et al. 2013). Additionally, the advent of high-throughput sequencing technology (HTS) has made the sequencing of millions of SNPs from the genome, possible, at moderate cost and it is not time intensive (Andrews and Luikart 2014; Christmas et al. 2016). Thus, through the use of landscape genomics, it is now possible to find correlations between genomic regions and the variable environmental characteristics (Rellstab et al. 2015; Ćalić et al. 2016). Therefore, this approach is used to pinpoint the environmental change in nature, that affects ecology (e.g.: climatic changes) of a species and can influence its adaptive genetic potential (Parisod and Holderegger 2012; Zhou et al. 2014; Rajora et al. 2016).

In landscape genomics studies, outlier loci method has been used in tandem with genome–environment association (GEA) method, to evaluate local adaptation in natural populations (Ćalić et al. 2016). In the outlier loci method, changes in among-population

frequency distribution of given allele(s)/loci, significantly different from that seen in the absence of natural selection, are identified, and these alleles/loci are considered to be under natural selection pressure (Rellstab et al. 2015; Ćalić et al. 2016). While, in the GEA analysis, occurrence of high correlation between the allele frequencies with one or more environmental variables is considered an indicator of local adaptation (Gugger et al. 2016; Pluess et al. 2016). However, as the outlier method does not specify the environmental forces at work on the locus under selection, studies using landscape genomics approach use the GEA results (Bashalkhanov et al. 2013; Shafer et al. 2015) to fix the causative environmental factor.

Considering that landscape genomics is a relatively new discipline, it requires refinement within the scope of sampling design, for GEA studies (Schoville et al. 2012; Robasky et al. 2014; Benestan et al. 2016; Manel et al. 2016; Ahrens et al. 2018). Studies, based on simulations, have predicted that the number of SNP markers used and the number of populations sampled influence the inferences from landscape genomics analysis (De Mita et al. 2013; Lotterhos and Whitlock 2015). However, a recent review, evaluated the impacts of sampling design on inferences from empirical studies on landscape genomics and it does not take into account the particularities of tropical regions (Ahrens et al. 2018). In a way, our approach complements that of Ahrens et al. (2018), as they focused on the limitations of the different techniques, the lack of standardization and non-availability of information in conducting these studies. However, unlike others, our study, by reporting on empirically observed patterns in landscape genomics, aims to shape sampling design strategies for improving the cost-benefit ratio of future works, conducted in laboratories using population genomics for the first time. Guidelines put forth in this study will be helpful, in particular, in subsidizing the costs of studies being carried out in the stark socioeconomic realities of

tropical regions, which are global biodiversity hotspots under grave anthropogenic threat.

2. Methodology

To perform this study, all research articles, published to date (at the time of writing, September 2018), which evaluated local adaptations in populations of wild plants were pooled. The papers were analyzed for (i) number of populations sampled, (ii) number of individuals sampled per population, (iii) total number of individuals sampled, and (iv) number of SNPs used. Care was taken to ensure that all data were obtained in empirical studies assessing local adaptation in wild plant populations. The Scopus (https://www.scopus.com) and Google Scholar (http://scholar.google.com.br/) databases were searched for title, abstract, and articles using the following keywords: environmental variables and SNPs, landscape genetics and SNPs, spatial analysis and SNPs, landscape genomics and SNPs, population genomics and SNPs, adaptive genetic variation, and local adaptation and SNPs. Search results were filtered to remove papers and reviews from clinical, biomedical, veterinary, and immunology areas. Then, a second filter was applied to retain only plants ecology studies. Finally, papers using simulations, or involving exotic and crop species in plantations, or using greenhouse experiments were removed and the remaining 35 empirical studies on in situ wild plants were selected for the present work. Using this data set and ArcGIS (10.2), a map was created depicting the distribution of the localities where researches were performed, in the different terrestrial biomes (Fig. 1). A double check was done at this stage to ensure if all the 35 papers chosen for this study, had evaluated in situ local adaptation in wild plant populations, using SNPs markers. Then, we compiled the following information from the papers: 1) authors names; 2) publication year, 3) country of study and

geographical coordinates; 4) botanical family studied; 5) species; 6) total number of individuals; 7) number of populations; 8) number of SNPs used; 9) number of SNPs under selection; and 10) method used to generate the SNPs data.

For this study, we considered as SNPs under selection, the SNPs from the pooled data that had a significant association with a geoclimatic variable, such as temperature, precipitation, latitude, longitude, elevation, evapotranspiration, and drought. They were used as a proxy to detect the potential for natural selection in wild plant populations. The number of SNPs under selection could be influenced by the methods employed to analyze the correlation between allelic frequency and environmental variables (De Mita et al. 2013), therefore most articles in the literature make inferences from such studies, using conservative criteria (Robasky et al. 2014; Benestan et al. 2016; Forester et al. 2016). It is known that each method has different types of limitations and caveats (e.g., vary in the rates of false positives or of false negatives), which can cause errors in the estimates and consequently in the estimate of the number of SNPs under selection (De Mita et al. 2013). For this reason, we chose to be conservative, using only the results obtained either simultaneously by at least two methods of analysis or strictly controlled for rates of false positives or false negatives (e.g., studies that controlled the population structure in the analyses). Thus, although limitations of the different methodologies used were not fully circumvented, we believe that the results reported in our study, portrays the findings of the literature, in a way seeking to minimize, as much as possible, the biases inherent to the different analysis.

The methods used to generate the SNPs data were subdivided into Random, Random Pool-Seq, Targeted, and Targeted-Pool-Seq categories. The Random category included studies that used random regions of DNA and individualized sequencing for the preparation of libraries. The Random-Pool-Seq category is distinct from Random, in

that the data is sequenced with the Pool-Seq technique (Pool-Seq, an equimolar amount of DNA is taken from each individual from a population and pooled for sequencing). In the Targeted category, we included studies that used only specific gene regions or expressed sequence tags with individualized sequencing. Finally, Targeted Pool-Seq category differed from Targeted, such that the data is sequenced by Pool-Seq technique mentioned above.

We also assessed the influence of sample size on number of SNPs under selection by determining the average of the individual number of samples from each population which was arrived at by dividing the total number of individuals and the number of populations in each study. Additionally, the percentage of SNPs under selection was calculated to estimate the number of genes under natural selection in the genome.

To evaluate the influence of each variable on the number of SNPs under selection, we used linear mixed effects model of the lme4 package of R program (Bates et al. 2015). The variables selected in this study, such as the number of populations, number of individuals, average number of individuals sampled per population, and number of SNPs, were used to determine their influence on the number of SNPs under selection and, consequently, the detection of natural selection in plant species. As this type of model requires the fulfillment of the assumptions of normality and homoscedasticity of the residues, we performed the scaling of each variable separately. We used the rank function in R program to reduce the amplitude of our data, meet the assumptions, and obtain the goodness of fit of the statistical models. For each variable, an independent model was created by considering each variable as the fixed effect and the others as the random effect, such that:

nº of SNPs under selection ~ nº of SNPs used + (1| nº of individuals) + (1| nº of pop) + (1| nº of ind_pop).

where, the number of SNPs under selection is the response variable; number of SNPs used is the fixed-effect variable; the terms in parentheses are the random effect variables and the number 1 indicates that the intercept is random between the observations of each variable.

The linear mixed effects model in the lme4 package does not provide the results as the coefficient of determination (r^2), due to theoretical problems or difficulty of implementation. For this reason, we used the marginal r^2 (Nakagawa and Schielzeth 2013) to calculate the amount of variance explained by the fixed-effect variables in each model. Finally, we used simple linear regression analysis of the number of SNPs used in the function of the total number of individuals. To make the graphs with the significant results of all the analyses, we used the ggplot2 package in R (Wickham 2016).

3. Results

Based on the 35 selected empirical studies, we analyzed 44 observations involving 36 different species for the number of SNPs under selection (Table 1). These studies were conducted mainly in Europe and North America (Fig. 1A), in ten plant families, dominance by Pinaceae (53% of the species) (Fig. 1B), distributed mainly in ecosystems of temperate broadleaf and mixed forest, temperate grasslands and Mediterranean forests (Fig. 1C). The total number of individuals ranged from 22 to 2,574, and the number of SNPs used varied between 33 and 2,091,957. However, the number of SNPs under selection (SNPs with significant association with some

geoclimatic variable) varied from 2 to 2,522 with the percentage of SNPs under selection ranging from 0.02 to 78.14 (Table 1).

In the analysis using the linear mixed-effects models, we verified the effect of each variable (related to sample size and SNP number) individually on the number of SNPs under selection, whereas the others were used as random effect, as mentioned in the methodology section. On entering, the number of populations, as fixed effect variable, the model showed no significant relation with the number of SNPs under selection (marginal r^2 = 0.011, p=0.38). A similar result was obtained when the mean number of individuals was used as the fixed effect variable (Marginal r^2=0.041, p=0.16). However, when the number of SNPs was used as the fixed effect variable, a positive and significant relationship was observed with the number of SNP under selection (Fig. 2A). When the total number of individuals was used as a fixed effect variable in the model, a negative and significant relation was observed in relation to the number of SNPs under selection (Fig. 2B). Subsequently, it was verified that the total number of individuals had a negative and significant relation with the number of SNPs used (Fig. 2C).

4. Discussion

The data obtained in the systematic review of studies in wild plant populations showed great heterogeneity in the number of individuals, populations, and quantity of SNPs. Our results revealed that the total number of individuals and number of SNPs used are of fundamental importance in detecting signs of natural selection in wild plant populations. We also found an enormous knowledge gap in the neotropics, underscoring the fact that predicting the response of the tropical forests to geoclimatic changes based on available data is not possible. Most of the studies had been conducted in the

ecosystems of temperate and mixed forests of the European and North American habitats.

We found few empirical studies evaluating SNPs under natural selection, with only 36 plants species under analysis. From those studies, we registered ten families, where 53% of the species belonged to the Pinaceae family. This result evidences the poor understanding of local adaptation processes for different species. Considering that tropical forests have approximately 11,371 species of trees (Slik et al. 2015), we have a limited vision to predict how species may respond to anthropogenic disturbances like climate change, fragmentation, and forest loss (Hansen et al. 2012; Cullingham et al. 2014). In addition, ~ 81% of studies were conducted in the European continent or in North America. Although much of the plant diversity is located in tropical regions (Mannion et al. 2014; Slik et al. 2015), the ability of these species to adapt in situ in response to environmental changes has not been studied (Storfer et al. 2010). We believe that this lack of knowledge for the tropics is a reflection of the socioeconomic conditions of this region as well as the lack of specialists in the area since population genomics is relatively recent and the first studies in the tropics are just emerging (see Table 1). Considering that tropical regions could be severely affected by climate change (Bellard et al. 2012), our results have shown the great need for studies that seek to understand how geoclimatic variables influence local adaptation. Through these studies, it will be possible to predict how tropical species will respond to climate change and assist in their conservation and management (Hansen et al. 2012; Fitzpatrick and Keller 2015; Shafer et al. 2015; Benestan et al. 2016).

Our study also evidenced that the number of populations and mean number of individuals in a population do not influence the ability to detect natural selection. This emphasizes the importance of sampling, at the best, the whole extent of the area that is

being influenced by the environmental variable of interest, rather than focusing on the number of populations sampled or on the mean number of individuals per population. If many populations, from similar environments, are sampled, the probability of detecting local adaptation does not increase (Lotterhos and Whitlock 2015). Conversely, even if few populations are sampled in contrasting environmental conditions, it would increase the power to detect local adaptation (Lotterhos and Whitlock 2015). Thus, it will be more effective to cover the environmental heterogeneity, avoiding efforts both in the expansion of the number of populations and in the average number of individuals per population.

Our results showed that the number of SNPs used explained 59% of the variation in the number of SNPs under selection and therefore can positively influence the detection of natural selection signals. This could occur because mostly a low percentage of genes in a genome are under natural selection (Table 1); therefore, increasing the genome sampling, also increases the chance of identifying SNPs under selection. Interestingly, a general pattern observed was that studies using Random, Targeted Pool-Seq, and Random Pool-Seq methodologies used more SNPs and had a greater number of SNPs under selection. Therefore, it would be beneficial if future studies used one of these methodologies, especially Random Pool-Seq and Targeted Pool-Seq, to optimize the cost-benefit ratio. In this context, studies that aim to evaluate the indicators of natural selection in populations of wild plants could use the Targeted Pool-Seq approach, in cases where genomic information of the species is available (e.g., genome size and gene annotation). By using targeted sequencing in DNA pools, cost benefit ratio of such studies would be increased. In contrast, for the species with little or no genomic information available and for laboratories that are working on landscape genomics for the first time, Random Pool-Seq is the best strategy to increase the

chances of detecting SNPs under selection. An interesting alternative to make landscape genomics financially feasible would be the use of grouped sequencing of individuals by techniques like RADseq that do not require any prior knowledge of the genome (Andrews et al. 2016). While it is still advantageous to use the Pool-Seq technique which can increase accuracy in allele frequency estimates, compared to individualized sequencing, making it an excellent tool to be used in landscape genomic studies (Futschik and Schlötterer 2010; Bansal et al. 2011; Schlötterer et al. 2014). Moreover, it also allows to reconcile the sampling of a large number of individuals needed in population genomics, with thousands or millions of SNPs, to evaluate adaptation in natural populations (Fischer et al. 2013; Gautier et al. 2013; Rellstab et al. 2016; Frachon et al. 2018). It is important to note that the Pool-Seq technique has some limitations, for example, not providing information for each individual separately (see Schlötterer et al. 2014). Therefore, the use of this technique is only suitable for studies aiming for population inferences without the need of information about individuals.

By fixing the variable in the linear mixed effects model as total number of individuals, we found that the total sample size was inversely related to the number of SNPs under selection and to the number of SNPs used. Although significant, these correlations explain only 10% of the variation in the number of SNPs under selection and 14% in the number of SNPs used. However, much of the variation in the number of SNPs under selection (59%) could be explained by fixing the variable as the number of SNPs used. Therefore we believe that the significantly negative correlation between the total number of individuals and the number of SNPs under selection is a reflection of the inverse relationship observed between the total number of individuals and the number of SNPs used. Thus, the negative relation seen between the total number of individuals and the number of SNPs under selection is an artifact generated by the lower

capacity detection of natural selection when the number of SNPs used decreases. The negative relationship observed between the number of individuals sampled and the number of SNPs used could be due to incomplete genomic information because of the costs associated with sequencing of whole genomes (Gautier et al. 2013). Among species for which some genomic information is available, size consideration of the genome becomes important; because the larger genomes may need more number of SNPs to be sampled to give a full perspective, thus impacting the number of individuals that can be analyzed given the budgetary constraints. Also, investing a huge portion of the budget for sampling in the field to increase the sample size, would affect the laboratory stages of the study by limiting resource allocation. Thus, the negative relation found in this study, can be related to the costs of individual sequencing of each sample. In population genomics studies, the conflicting demands between the total number of individuals and the number of SNPs used (Gautier et al. 2013), is circumvented by using the Pool-Seq technique (Table 1). This technique has also been used in plant species with incomplete or total absence of genomic information (Schlötterer et al. 2014). This indicates that the same strategy can be used in the landscape genomics studies of species in tropical regions, which are major biodiversity hotspots of the world (Mannion et al. 2014; Slik et al. 2015), with little or no genomic information available.

5. Guidance for study of local adaptation in wild plant species

Our study involving systematic review of studies about wild plant populations, allows us to put forth the recommendations that in future landscape genomics studies, the experimental design should focus on increasing the total number of individuals sampled along the environmental heterogeneity under analysis (see Lotterhos and

Whitlock 2015). In addition, we also suggest that researchers seek to evaluate the adaptation in wild plant populations using the Pool-Seq technique to increase the number of SNPs and accuracy in allele frequency estimates (see Schlötterer et al. 2014). This technique has been used in population genomics as a powerful tool under different study scenarios, including model species (Frachon et al. 2018), allopolyploid species (Hirao et al. 2017) as well as for species with little or no genomic information available (Shih et al. 2018). Following these guidelines, it will be possible to extrapolate findings from studies that are performed almost exclusively on species of temperate climate and expanding them to tropical species. In this way, landscape genomics can be conducted in tropical regions in relation to anthropogenic disturbances, in spite of existing budgetary constraints, and the data so generated can be used to develop strategies for management and conservation of biodiversity.

Acknowledgments

The authors thank the Brazilian National Council of Scientific and Technological Development (CNPq) for the FAG research productivity fellowship; and Coordination for the Improvement of Higher Education Personnel (CAPES) for granting a scholarship to ASS. The authors also thank Marina Corrêa Côrtes, Luciana Aparecida Carlini Garcia, and Ricardo Dobrovolski for the corrections in the first version of the manuscript.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

Ahrens CW, Rymer PD, Stow A, et al (2018) The search for loci under selection: trends, biases and progress. Molecular ecology. 27, 1342-1356.

Alam Z, Roncal J, Peña-Castillo L (2018) Genetic variation associated with healthy traits and environmental conditions in Vaccinium vitis-idaea. BMC genomics. 19, 4. doi.org/10.1186/s12864-017-4396-9.

Alberto FJ, Aitken SN, Alía R, et al (2013) Potential for evolutionary responses to climate change–evidence from tree populations. Global Change Biology. 19, 1645-1661. doi.org/10.1111/gcb.12181.

Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. Molecular Ecology. 23, 1661-1667.doi.org/10.1111/mec.12686.

Andrews KR, Good JM, Miller MR, et al (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics. 17, 81-92. 10.1038/nrg.2015.28.

Bansal V, Tewhey R, LeProust EM, et al (2011) Efficient and cost effective population resequencing by pooling and in-solution hybridization. PLoS One. 6, e18353.

Bashalkhanov S, Eckert AJ, Rajora OP (2013) Genetic signatures of natural selection in response to air pollution in red spruce (Picea rubens, Pinaceae). Molecular ecology. 22, 5877-5889.doi: 10.1111/mec.12546.

Bates D, Mächler M, Bolker B, et al (2015) Fitting linear mixedeffects models using lme4. J Stat Softw. 67:1–48. doi:10.18637/ jss.v067.i01.

Bellard C, Bertelsmeier C, Leadley P, et al (2012) Impacts of climate change on the future of biodiversity. Ecology letters. 15, 365-377. doi: 10.1111/j.1461-0248.2011.01736.x.Impacts.

Benestan LM, Ferchaud AL, Hohenlohe PA, et al (2016) Conservation genomics of natural and managed populations: building a conceptual and practical framework. Molecular ecology. 25, 2967-2977. doi: 10.1111/mec.13647.

Ćalić I, Bussotti F, Martínez-García PJ, et al (2016) Recent landscape genomics studies in forest trees—what can we believe? Tree genetics & genomes. 12, 3.doi: 10.1007/s11295-015-0960-0.

Christmas MJ, Biffin E, Breed MF, et al (2016) Finding needles in a genomic haystack: targeted capture identifies clear signatures of selection in a nonmodel plant species. Molecular ecology. 25,4216-4233.doi: 10.1111/mec.13750.

Cullingham CI, Cooke JE, Coltman DW (2014) Cross-species outlier detection reveals different evolutionary pressures between sister species. New Phytologist, 204, 215-229. doi: 10.1111/nph.12896.

Davis MB, Shaw RG, Etterson JR (2005) Evolutionary responses to changing climate. Ecology. *86*, 1704-1714. doi: 10.1890/03-0788.

De Kort H, Vandepitte K, Mergeay J, et al (2015) The population genomic signature of environmental selection in the widespread insect-pollinated tree species Frangula alnus at different geographical scales. Heredity.115, 415-425. doi: 10.1038/hdy.2015.41.

De Mita S, Thuillet AC, Gay L, et al (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. Molecular

ecology. 22, 1383-1399. doi: 10.1111/mec.12182.

Di Pierro EA, Mosca E, Rocchini D, et al (2016) Climate-related adaptive genetic variation and population structure in natural stands of Norway spruce in the South-Eastern Alps. Tree genetics & genomes. 12,16.doi: 10.1007/s11295-016-0972-4.

Di Pierro EA, Mosca E, González-Martínez SC, et al (2017) Adaptive variation in natural Alpine populations of Norway spruce (Picea abies [L.] Karst) at regional scale: Landscape features and altitudinal gradient effects. Forest Ecology and Management. 405, 350-359. doi.org/10.1016/j.foreco.2017.09.045.

Eckert AJ, Bower AD, González-Martínez SC, et al (2010) Back to nature: ecological genomics of loblolly pine (Pinus taeda, Pinaceae). Molecular ecology. 19, 3789-3805.

Eckert AJ, Maloney PE, Vogler DR, et al (2015) Local adaptation at fine spatial scales: an example from sugar pine (Pinus lambertiana, Pinaceae). Tree Genetics & Genomes. 11, 1614-2942. doi.org/10.1007/s11295-015-0863-0.

Fahrenkrog AM, Neves LG, Resende Jr MF, et al (2017) Population genomics of the eastern cottonwood (Populus deltoides). Ecology and evolution. 7, 9426-9440.

Fahrig L (2003) Effects of habitat fragmentation on biodiversity. Annual review of ecology, evolution, and systematics. 34, 487-515. doi: 10.1146/annurev.ecolsys.34.011802.132419.

Fischer MC, Rellstab C, Tedder A, et al (2013) Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. Molecular ecology. 22, 5594-5607. doi: 10.1111/mec.12521.

Fitzpatrick MC, Keller SR (2015) Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future environmental adaptation. Ecology Letters. 18, 1-16. doi: 10.1111/ele.12376.

Forester BR, Jones MR, Joost S, et al (2016) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular ecology. 25, 104-120. doi: 10.1111/mec.13476.

Frachon L, Bartoli C, Carrère S, et al (2018) A genomic map of climate adaptation in Arabidopsis thaliana at a micro-geographic scale. Frontiers in plant science. 9, 967. doi:10.3389/fpls.2018.00967.

Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics. 186, 207-218.

Gaut B (2012) Arabidopsis thaliana as a model for the genetics of local adaptation. Nature Genetics. 44, 115.doi: 10.1038/ng0612-732a.

Gautier M, Foucaud J, Gharbi K, et al (2013) Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. Molecular Ecology. 22, 3766-3779.

Geraldes A, Farzaneh N, Grassa C J, et al (2014) Landscape genomics of Populus trichocarpa: the role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. Evolution. 68, 3260-3280.

Gugger PF, Cokus SJ, Sork VL (2016) Association of transcriptome-wide sequence variation with climate gradients in valley oak (Quercus lobata). Tree genetics & genomes. 12, 15.doi:

10.1007/s11295-016-0975-1.

Hamlin JA, Arnold ML (2015) Neutral and selective processes drive population differentiation for Iris hexagona. Journal of Heredity. 106, 628-636.

Hansen MM, Olivieri I, Waller DM, et al (2012) Monitoring adaptive genetic responses to environmental change. Molecular ecology. 21, 1311-1329.doi: 10.1111/j.1365-294X.2011.05463.x.

Hirao AS, Onda Y, Shimizu-Inatsugi R, et al (2017) Cost-Effective Discovery of Nucleotide Polymorphisms in Populations of an Allopolyploid Species Using Pool-Seq. American Journal of Molecular Biology. 7, 153-168. doi:10.4236/ajmb.2017.74012.

Jaramillo-Correa JP, Rodríguez-Quilón I, Grivet D, et al (2015) Molecular proxies for climate maladaptation in a long-lived tree (Pinus pinaster Aiton, Pinaceae). Genetics. 199, 793-807.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecology letters. 7, 1225-1241.

Keller SR, Levsen N, Olson MS, et al (2012) Local adaptation in the flowering-time gene network of balsam poplar, Populus balsamifera L. Molecular biology and evolution. 29, 3143-3152.

Lanes ÉC, Pope NS, Alves R, et al (2018) Landscape genomic conservation assessment of a narrow-endemic and a widespread morning glory from Amazonian savannas. Frontiers in plant science. 9, 532. doi:10.3389/fpls.2018.00532.

Lind BM, Friedline CJ, Wegrzyn JL, et al (2017) Water availability drives signatures of local adaptation in whitebark pine (Pinus albicaulis Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. Molecular ecology. 26, 3168-3185.

Liu B, Su J, Chen J, et al (2013) Anthropogenic halo disturbances alter landscape and plant richness: A ripple effect. PloS one. 8, e56109. doi: 0.1371/journal.pone.0056109.

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular ecology. 24, 1031-1046.

Manel S, Joost S, Epperson BK, et al (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. Molecular Ecology. 19, 3760-3772. doi:10.1111/j.1365-294X.2010.04717.x.

Manel S, Perrier C, Pratlong M, et al (2016) Genomic resources and their influence on the detection of the signal of positive selection in genome scans. Molecular ecology. 25, 170-184.doi: 10.1111/mec.13468.

Mannion PD, Upchurch P, Benson RB, et al (2014) The latitudinal biodiversity gradient through deep time. Trends in ecology & evolution. 29, 42-50.doi: 10.1016/j.tree.2013.09.012.

Martins K, Gugger PF, Llanderal-Mendoza J, et al (2018) Landscape genomics provides evidence of climate associated genetic variation in Mexican populations of Quercus rugosa. Evolutionary Applications. 1-17. doi.org/10.1111/eva.12684.

Mosca E, Eckert AJ, Di Pierro EA, et al (2012) The geographical and environmental determinants of genetic diversity for four alpine conifers of the E uropean A lps. Molecular ecology. 21, 5530-5545.

Mosca E, González-Martínez SC, Neale DB (2014) Environmental versus geographical determinants of genetic structure in two subalpine conifers. New Phytologist. 201, 180-192.

Mosca E, Gugerli F, Eckert AJ, et al (2016) Signatures of natural selection on Pinus cembra and P. mugo along elevational gradients in the Alps. Tree genetics & genomes. 12, 9.doi: 10.1007/s11295-015-

0964-9.

Modesto IS, Miguel C, Pina-Martins F, et al (2014) Identifying signatures of natural selection in cork oak (Quercus suber L.) genes through SNP analysis. Tree genetics & genomes. 10, 1645-1660.

Nakagawa S, Schielzeth H (2013) A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution. 4, 133-142.

Nicotra AB, Atkin OK, Bonser SP, et al (2010) Plant phenotypic plasticity in a changing climate. Trends in plant science. 15, 684-692.

Parisod C, Holderegger R (2012) Adaptive landscape genetics: pitfalls and benefits. Molecular ecology. 21, 3644-3646.doi: 10.1111/j.1365-294X.2012.05675.x

Parmesan C, Hanley ME (2015) Plants and climate change: complexities and surprises. Annals of botany. 116, 849-864.doi: 10.1093/aob/mcv169.

Pluess AR, Frank A, Heiri C, et al (2016) Genome–environment association study suggests local adaptation to climate at the regional scale in Fagus sylvatica. New Phytologist. 210, 589-601.doi: 10.1111/nph.13809.

Prunier J, Gerardi S, Laroche J, et al (2012) Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. Molecular Ecology. 21, 4270-4286.

Rajora OP, Eckert AJ, Zinck JW (2016) Single-locus versus multilocus patterns of local adaptation to climate in eastern white pine (Pinus strobus, Pinaceae). PloS one. 11, e0158691.doi: 10.1371/journal.pone.0158691.

Rellstab C, Gugerli F, Eckert AJ, et al (2015) A practical guide to environmental association analysis in landscape genomics. Molecular Ecology. 24, 4348-4370. doi: 10.1111/mec.13322.

Rellstab C, Zoller S, Walthert L, et al (2016) Signatures of local adaptation in candidate genes of oaks (Quercus spp.) with respect to present and future climatic conditions. Molecular ecology. 25, 5907-5924.doi:10.1111/mec.13889.

Rellstab C, Fischer MC, Zoller S, et al (2017) Local adaptation (mostly) remains local: reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. Heredity. 118, 193.doi: 10.1038/hdy.2016.82.

Reznick DN, Ghalambor CK (2001) The population ecology of contemporary adaptation: what empirical studies reveal about the conditions that promote adaptive evolution. Genetica. 112–113, 183–198.

Robasky K, Lewis NE, Church GM (2014) The role of replicates for error mitigation in next-generation sequencing. Nature Reviews Genetics. 15, 56.doi: 10.1038/nrg3655.

Roschanski AM, Csilléry K, Liepelt S, et al (2016) Evidence of divergent selection for drought and cold tolerance at landscape and local scales in Abies alba Mill. in the French Mediterranean Alps. Molecular ecology. 25, 776-794.doi: 10.1111/mec.13516

Ruiz Daniels R, Taylor RS, Serra-Varela MJ, et al (2018) Inferring selection in instances of long-range colonization: The Aleppo pine (Pinus halepensis) in the Mediterranean Basin. Molecular ecology. 27, 3331-3345.

Scalfi M, Mosca E, Di Pierro EA, et al (2014) Micro-and macro-geographic scale effect on the molecular imprint of selection and adaptation in Norway spruce. PloS one. 9, e115499.

Schlötterer C, Tobler R, Kofler R, et al (2014) Sequencing pools of individuals—mining genome-wide

polymorphism data without big funding. Nature Reviews Genetics. 15, 749.doi: 10.1038/nrg3803.

Schoville SD, Bonin A, François O, et al (2012) Adaptive genetic variation on the landscape: methods and cases. Annual Review of Ecology, Evolution, and Systematics. 43, 23-43.doi: 10.1146/annurev-ecolsys-110411-160248.

Shafer AB, Wolf JB, Alves PC, et al (2015) Genomics and the challenging translation into conservation practice. Trends in Ecology & Evolution. 30, 78-87.doi: 10.1016/j.tree.2014.11.009.

Shih KM, Chang CT, Chung JD, et al (2018) Adaptive genetic divergence despite significant isolation-by-distance in populations of taiwan cow-tail fir (Keteleeria davidiana var. formosana). Frontiers in plant science. 9, 92. doi=10.3389/fpls.2018.00092.

Slik JF, Arroyo-Rodríguez V, Aiba SI, et al (2015) An estimate of the number of tropical tree species. Proceedings of the National Academy of Sciences. 112, 7472-7477.

Sork VL, Squire K, Gugger PF, et al (2016) Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, Quercus lobata. American Journal of Botany. 103, 33-46.doi: 10.3732/ajb.1500162.

Steane DA, Potts BM, McLean E, et al (2014) Genome-wide scans detect adaptation to aridity in a widespread forest tree species. Molecular Ecology. 23, 2500-2513.doi: 10.1111/mec.12751.

Storfer A, Murphy MA, Spear SF, et al (2010) Landscape genetics: where are we now? Molecular ecology. 19, 3496-3514.doi:10.1111/j.1365-294X.2010.04691.x.

Tsumura Y, Uchiyama K, Moriguchi Y, et al (2014) Genetic differentiation and evolutionary adaptation in Cryptomeria japonica. G3: Genes, Genomes, Genetics. g3-114.

Turner TL, Bourne EC, Von Wettberg EJ, et al (2010) Population resequencing reveals local adaptation of Arabidopsis lyrata to serpentine soils. Nature genetics. 42, 260.doi: 10.1038/ng.515.

Wickham H (2016) ggplot2: elegant graphics for data analysis. Springer.

Williams GC (1966) Adaptation and Natural Selection. Princeton University Press, Princeton, New Jersey.

Wilson MC, Chen XY, Corlett RT, et al (2016) Habitat fragmentation and biodiversity conservation: key findings and future challenges. Landscape Ecology. 31, 2. doi: 10.1007/s10980-015-0312-3.

Zhou Y, Zhang L, Liu J, et al (2014) Climatic adaptation and ecological divergence between two closely related pine species in Southeast China. Molecular ecology. 23, 3504-3522.doi: 10.1111/mec.12830.

LIST OF TABLE AND FIGURE

Table 1 Biological models and descriptive statistics calculated in 35 studies that assessed local adaptation in wild plant populations in situ using markers SNPs.

Author	Species	Pop	Ind	Ind_Pop	SNPs	SNPs_GEA	SNPs_GEA*	Method
Eckert et al., 2010	*Pinus taeda*	54	682	12.63	1730	118	6.82	Targeted
Mosca et al., 2012	*Abies alba Mill*	37	1183	31.97	249	3	1.20	Targeted
Mosca et al., 2012	*Larix decidua Mill*	24	920	38.33	267	5	1.87	Targeted
Keller et al., 2012	*Populus balsamifera L*	31	443	14.29	335	11	3.28	Targeted
Prunier et al., 2012	*Picea mariana*	41	593	14.46	47	9	19.15	Targeted
Mosca et al., 2012	*Pinus cembra L*	24	860	35.83	459	11	2.40	Targeted
Mosca et al., 2012	*Pinus mugo*	27	935	34.63	693	14	2.02	Targeted
Fischer et al., 2013	*Arabidopsis halleri*	5	100	20.00	2091957	1037	0.05	Random-Pool-Seq
Bashalkhanov et al., 2013	*Picea rubens*	5	900	180.00	33	6	18.18	Targeted
Mosca et al., 2014	*Abies alba Mill*	36	1108	30.80	231	5	2.16	Targeted
Tsumura et al., 2014	*Cryptomeria japonica*	14	186	13.29	3930	25	0.64	Targeted
Mosca et al., 2014	*Larix decidua Mill*	22	824	37.50	233	7	3.00	Targeted
Modesto et al., 2014	*Quercus suber L*	16	96	6.00	44	5	11.36	Targeted
Scalfi et al., 2014	*Picea abies [L.] Karst)*	12	300	25.00	227	2	0.88	Targeted
Cullingham et al., 2014	*Pinus contorta var. latifolia*	13	368	28.31	399	22	5.51	Targeted
Cullingham et al., 2014	*Pinus banksiana*	4	100	25.00	399	8	2.01	Targeted
Geraldes et al., 2014	*Populus trichocarpa*	25	424	16.96	28135	58	0.21	Targeted
De Kort et al., 2015	*Frangula alnus subsp. alnus*	25	619	24.76	183	143	78.14	Targeted
Hamlin et al., 2015	*Iris hexagona*	8	92	11.50	750	70	9.33	Random
Eckert et al., 2015	*Pinus lambertiana*	10	241	24.10	475	14	2.95	Targeted
Jaramillo-Correa et al., 2015	*Pinus pinaster Aiton*	36	772	21.44	266	18	6.77	Targeted
Roschanski et al., 2016	*Abies alba Mill*	4	376	94.00	267	8	3.00	Targeted
Christmas et al., 2016	*Dodonaea viscosa ssp. Angustissima*	17	89	5.24	8462	93	1.10	Targeted
Pluess et al., 2016	*Fagus sylvatica*	79	234	2.96	144	16	11.11	Targeted
Gugger et al., 2016	*Quercus lobata*	12	22	1.83	220427	79	0.04	Targeted
Sork et al., 2016	*Quercus lobata*	13	45	3.46	195	8	4.10	Targeted
Rellstab et al., 2016	*Quercus robur*	24	465	19.38	3576	181	5.06	Targeted-Pool-Seq
Rellstab et al., 2016	*Quercus petraea*	18	350	19.44	3576	224	6.26	Targeted-Pool-Seq
Rellstab et al., 2016	*Quercus pubescens*	17	326	19.18	3576	304	8.50	Targeted-Pool-Seq
Di Pierro et al., 2016	*Norway spruce*	24	826	34.40	214	10	4.67	Targeted
Di Pierro et al., 2016	*Picea abies [L.] Karst*	23	826	35.91	214	7	3.27	Targeted
Mosca et al., 2016	*Pinus cembra L*	18	678	37.67	455	74	16.26	Targeted
Mosca et al., 2016	*Pinus mugo*	20	673	33.65	663	60	9.05	Targeted
Rajora et al., 2016	*Pinus strobus*	29	638	22.00	44	2	4.55	Targeted
Di Pierro et al., 2017	*Picea abies [L.] Karst*	18	687	38.17	175	19	10.86	Targeted
Lind et al., 2017	*Pinus albicaulis Engelm*	8	244	30.50	116231	1780	1.53	Random

Fahrenkrog et al., 2017	*Populus deltoides*	50	168	3.36	79969	2.522	3.15	Targeted-Pool-Seq
Frachon et al., 2018	*Arabidopsis thaliana*	168	2.574	15.32	1638649	300	0.02	Random-Pool-Seq
Lanes et al., 2018	*Ipomoea cavalcantei*	1	122	122.00	34102	2239	6.57	Random
Lanes et al., 2018	*Ipomoea maurandioides*	4	254	63.50	23181	1814	7.83	Random
Shih et al., 2018	*Keteleeria davidiana var. formosana*	5	62	12.40	13914	15	0.11	Random-Pool-Seq
Martins et al., 2018	*Quercus rugosa Née*	17	103	6.06	5354	97	1.81	Random-Pool-Seq
Ruiz Daniels et al., 2018	*Pinus halepensis Mill*	46	1.326	28.83	294	7	2.38	Targeted
Alam et al., 2018	*Vaccinium vitis-idaea subsp. Minus*	56	56	1.00	1586	132	8.32	Random
Descriptive statistics	**Minimum**	**1**	**22**	**1.00**	**33**	**2**	**0.02**	
	Maximum	**168**	**2574**	**180**	**2091957**	**2522**	**78.14**	
	Mean	**25.91**	**520.2**	**29.48**	**97420**	**263.2**	**7**	
	Standard deviation	**27.28**	**471.82**	**32.31**	**395181.4**	**614**	**11.99**	

(Ind, Number of individuals; Ind_Pop, Average number of individuals per population; SNPs, Number of used SNPs; Pop, Number of populations; SNPs_GEA, Number of SNPs under selection; SNPs_GEA*, Percentage of SNPs under selection; Method, Method used to generate the SNP data)

Fig 1 Geographical, ecological and species distribution found in the 35 research articles, included in this study, which evaluated local adaptation in wild plant populations in their natural occurrence area using SNPs markers, are shown; (**A**) Percentage of studies published per continent; (**B**) Percentage of studies distributed per plant family; (**C**) The points distributed on the map indicating the sampling of the studies in the different terrestrial ecosystems.

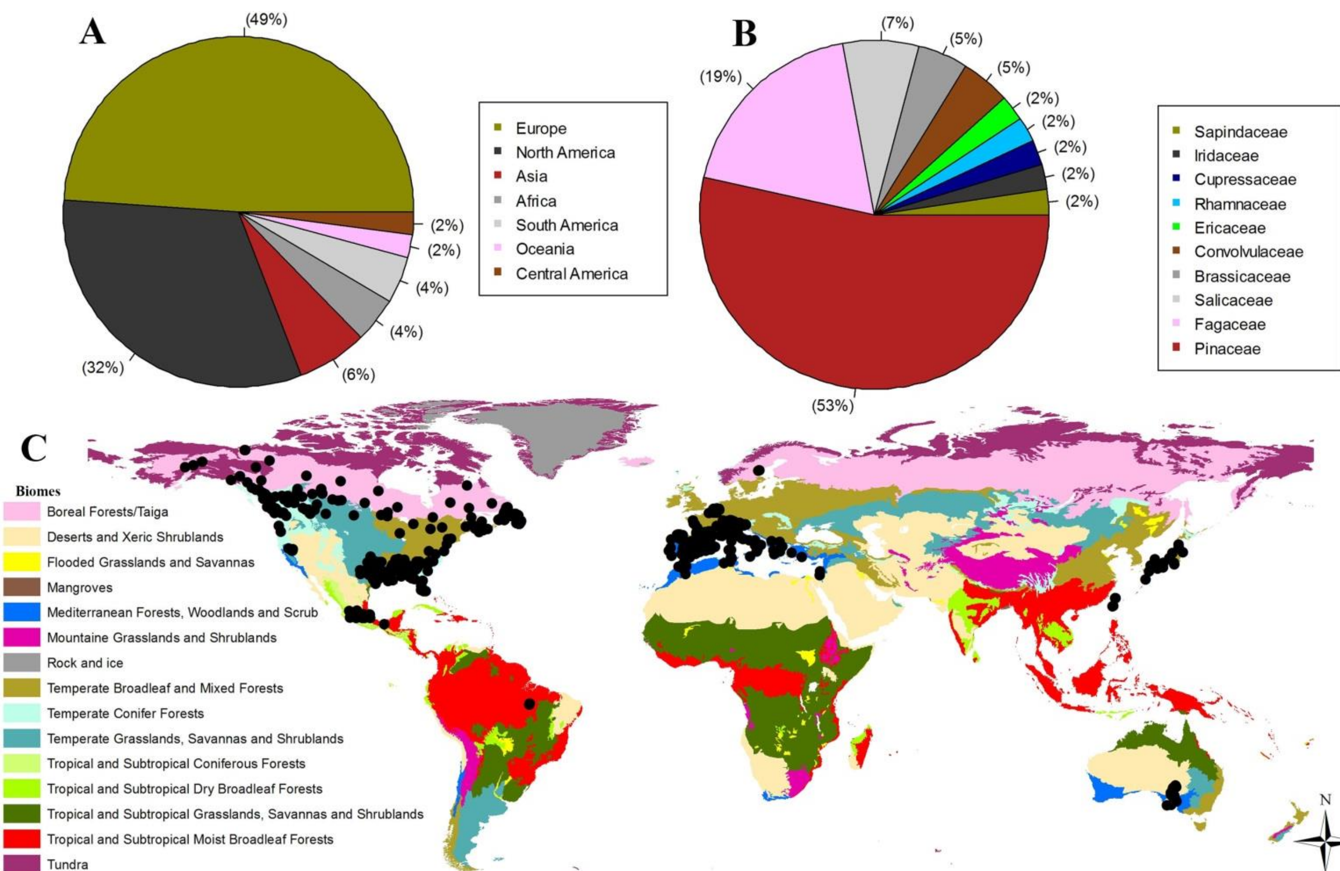
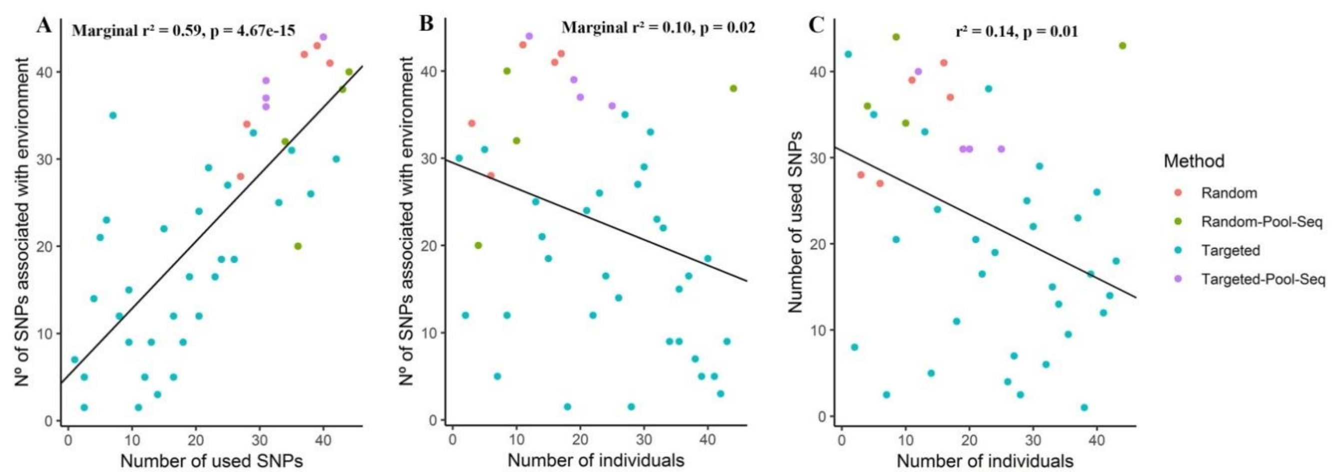


Fig 2 Relation between sampling design and the detection of SNPs with environmental association **(A)** number of SNPs with environmental association as a function of the number of SNPs used in the studies; **(B)** number of SNPs with environmental association as a function of the total number of individuals in each study; **(C)** number of SNPs used as a function of the total number of individuals in each study. The colors of the dots represent the different methods used in the studies (Random used random regions of DNA and individualized sequencing; Random Pool-Seq used random regions of DNA and Pool-Seq technique; Targeted used only gene regions or expressed sequence tags and individualized sequencing; Targeted Pool-Seq used only gene regions or expressed sequence tags and sequencing in Pool-Seq).



Capítulo 2

Diversidade genômica de uma palmeira tropical em paisagens antrópicas da Mata Atlântica

Citação

Alesandro Souza Santos, Eliana Cazetta, Deborah Faria, Thâmara Moura Lima, Maria Teresa Gomes Lopes, Carolina da Silva Carvalho, Alessandro Alves-Pereira, Fernanda Amato Gaiotto. **Diversidade genômica de uma palmeira tropical em paisagens antrópicas da Mata Atlântica**. Manuscrito será submetido para a Molecular Ecology.

Diversidade genômica de uma palmeira tropical em paisagens antrópicas da Mata Atlântica

Alesandro Souza Santos[1], Eliana Cazetta[1], Deborah Faria[1], Thâmara Moura Lima[2], Maria Teresa Gomes Lopes[3], Carolina da Silva Carvalho[4], Alessandro Alves-Pereira[5], Fernanda Amato Gaiotto[1]

[1]Laboratório de Ecologia Aplicada à Conservação, Pós-Graduação em Ecologia e Conservação da Biodiversidade, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna, km 16, Ilhéus, BA, Brasil, CEP 45662-900

[2]Pós-Graduação em Genética e Biologia Molecular, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna, km 16, Ilhéus, BA, Brasil, CEP 45662-900

[3]Universidade Federal do Amazonas, 69.077-000, Manaus, AM, Brasil

[4]Universidade Estadual Paulista (UNESP), Instituto de Biociências, Departamento de Ecologia, Campus Rio Claro, São Paulo, 13506-900, Brasil.

[5]Departamento de Biologia de Planta, Universidade Estadual de Campinas, Av. Cândido Rondon, 400, Cidade Universitária Zeferino Vaz, Campinas, SP 13083-875, Brasil.

*Autor correspondente: alesandrouesc@gmail.com

Resumo

O ser humano tem reduzido e fragmentado florestas nos diferentes ecossistemas terrestres do planeta. Essa ação é especialmente preocupante para Mata Atlântica Brasileira que embora abrigue elevada riqueza e endemismo de espécies, foi reduzida a apenas 11-16% da sua extensão original. Nossa pesquisa foi conduzida em 17 fragmentos de Mata Atlântica do estado da Bahia e seis de São Paulo, utilizando *Euterpe edulis* Martius, uma palmeira tropical ameaçada de extinção. Investigamos a influência da redução de floresta em escala de paisagem e da riqueza de aves sob a diversidade gênica de *E. edulis,* utilizando 7.490 SNPs (*single nucleotide polymorphism*) intragênicos. Além disso, como os fragmentos florestais na Bahia apresentam baixa densidade populacional de *E. edulis* e reduzida riqueza de aves, se comparadas com as florestas em São Paulo, esperávamos encontrar uma maior diversidade gênica nessas últimas. As 23 populações formaram quatro grupos, sendo que a diversidade gênica decresce das localidades do estado de São Paulo, para as regiões do estado da Bahia. Acreditamos que essa diversidade seja moldada por diversos processos ecológicos e evolutivos simultaneamente, que estão diferindo entre as localidades de estudo e entre diferentes morfotipos da espécie que ocorrem na Bahia e em São Paulo. Por outro lado, a porcentagem de floresta na paisagem nem a riqueza de aves dispersoras de sementes são um *drive* para a diversidade gênica de *E. edulis,* e acreditamos que essa diversidade seja moldada pela ampla distribuição da espécie na Mata Atlântica, o que lhe confere alta plasticidade fenotípica e adaptação local.

Palavras-Chave: Floresta tropical, Genômica da paisagem, Pool-seq, Sondas de captura, HTS.

1. Introdução

Atividades antrópicas para suprir a demanda por recursos da população humana tem alterado drasticamente o meio ambiente, convertendo paisagens naturais em ambientes completamente modificados, ocasionando assim uma rápida redução e fragmentação dos ecossistemas florestais (Liu et al. 2013; Mcgill et al. 2015). Em especial, as alterações antrópicas têm atingindo as floretas tropicais, colocando em risco sua altíssima biodiversidade e endemismo de espécies (Morante-Filho et al. 2015; Slik et al. 2015; Rocha-santos et al. 2016). Assim, as diversas atividades humanas têm provocado declínios nos diferentes níveis da biodiversidade, como na riqueza de espécies de animais e plantas, em processos ecológicos, na diversidade filogenética e genética (Andrade et al. 2015; Carvalho et al. 2015; Morante-Filho et al. 2016; Benchimol et al. 2017). Dessa maneira, as floretas tropicais e toda sua biodiversidade associada estão em constante perigo.

Nesse contexto, estudos que possibilitem acessar a variabilidade genética das populações naturais de espécies tropicais são extremamente importantes do ponto de vista ecológico, evolutivo e conservacionista (Santos et al. 2015; Carvalho et al. 2017; Brancalion et al. 2018; Lanes et al. 2018). Estudos genéticos são capazes de detectar a fragilidade das populações diante das constantes perturbações causadas por ações humanas (Carvalho et al. 2016). Além disso, uma investigação genética mais detalhada ao nível de genômica da paisagem possibilita compreender como tais distúrbios ambientais alteram a diversidade genética de populações naturais (Orr 2005; Schoville et al. 2012; Fischer et al. 2013; Rellstab et al. 2015; Ćalić et al. 2016). Assim, a genômica da paisagem pode auxiliar a prever futuras respostas das espécies em ambientes modificados por ações humanas, contribuindo com a elaboração de estratégias eficientes de conservação (Fitzpatrick and Keller 2015; Benestan et al. 2016; Sork 2017). Apesar disso, raros estudos têm utilizado a genômica da paisagem para investigar os impactos das ações antrópicas em florestas tropicais (Lanes et al. 2018).

Nesse sentido, para compreender a influência de paisagens tropicais antrópicas sobre a variação genômica em populações naturais, selecionamos como modelo biológico, a espécie *Euterpe edulis* Martius (Arecaceae) e como modelo de ambiente, a Mata Atlântica Brasileira. Esse domínio florestal possui elevada riqueza e endemismo de espécies, e encontra-se reduzido à aproximadamente 11-16% de sua extensão original (Ribeiro et al. 2009). A espécie é nativa do Brasil, e possui grande importância

ecológica, com seus frutos consumidos por cerca de 58 espécies de aves e 20 espécies de mamíferos, sendo considerada espécie chave para manutenção da fauna em período de escassez de recursos (Galetti et al. 2013). Em termos econômicos, *E. edulis* é um dos produtos não madereiro mais explorados da Mata Atlântica Brasileira, através do extrativismo ilegal do meristema apical para consumo humano (Brancalion et al. 2012; Elena et al. 2014). Esse extrativismo provoca a morte dos indivíduos adultos e por consequência, a redução populacional, que em conjunto com a perda e fragmentação da Mata Atlântica, conduziu a espécie para a lista da flora Brasileira ameaçada de extinção (CNCFlora 2012).

Em termos gerais, estudos realizados na Mata Atlântica com *E. edulis* utilizando marcadores microssatélites, demonstraram que o desmatamento e a defaunação tem influenciado negativamente a espécie, reduzindo sua diversidade genética e distância de fluxo gênico (Carvalho et al. 2015, 2017; Santos et al. 2016). Além disso, estudos demonstram diferenças genéticas entre os morfotipos da espécie e adaptação para distintos ambientes florestais (Coelho 2010; Brancalion et al. 2018). Contudo, nenhum estudo até o presente momento utilizou uma abordagem de genômica da paisagem para avaliar os impactos antrópicos na espécie. Dessa maneira, a presente pesquisa foi conduzida em fragmentos de Mata Atlântica do estado da Bahia e de São Paulo, utilizando marcadores SNPs (*single nucleotide polymorphism*) inseridos em regiões gênicas, como objetivo de investigar: Como a redução de floresta em escala de paisagem e a riqueza de aves dispersoras de sementes influenciam na diversidade gênica de *E. edulis*? Além disso, como é conhecido que as florestas na Bahia apresentam baixa densidade populacional de *E. edulis* e reduzida riqueza de potenciais dispersores, se comparadas com as florestas em São Paulo, esperamos encontrar uma maior diversidade gênica nessas últimas.

2. Metodologia

2.1. Área de estudo

O estudo foi realizado em 17 áreas na região sul do estado da Bahia e, em seis áreas localizadas no estado de São Paulo, Brasil (Figura 1). As áreas amostradas na

Bahia estão inseridas em uma região composta pelos maiores remanescentes de Mata Atlântica do nordeste do Brasil (Ribeiro et al. 2009) e a vegetação é do tipo ombrófila densa (Thomas et al. 1998). Essa região é considerada prioritária para conservação devido à elevada riqueza e centro de endemismo de espécies arbóreas (Martini et al. 2007). Contudo, nos últimos 30 anos, o sul da Bahia tem passado por pressões antrópicas de conversão de floresta nativa em diferentes tipos de uso da terra (Morante-Filho et al. 2016), com destaque para as plantações de cacau (*Theobroma cacao*), seringueira (*Hevea brasiliensis*) e *Eucalyptus* sp, além de pastagem para gado (Figura 1D). Essa conversão de floresta nos diferentes usos da terra tem exibido diferentes trajetórias, com a Região Sul da Bahia, porção Sul (RS_BA) sendo mais desflorestada se comparada com a Região Sul da Bahia, porção Norte (RN_BA), onde estão inseridos grandes fragmentos de floresta, como a reserva biológica de Una. Assim, apesar da similaridade florística, topográfica e tipo do solo, devido as diferentes dinâmicas de desflorestamento, o fator localidade foi considerado nas análises (Figura 1A-B).

As áreas de estudo no estado de São Paulo estão inseridas em duas diferentes fitofisionomias da Mata Atlântica. Sendo, quatro delas inseridas em vegetação de Floresta Semidecidua (FS_SP) e duas em Floresta Ombrófila densa (FO_SP) (Carvalho et al. 2016; Figura 1C). Dessa maneira, como existe diferença na regularidade climática entre as fitofisionomias que pode influenciar de maneira distinta na adaptação de *E. edulis* (Brancalion et al. 2018), esse fator foi considerado nas análises.

2.2. Desenho amostral

Este estudo utilizou o desenho amostral da rede de pesquisa SISBIOTA (CNPq Nº 563216/2010-7), através do uso do mapeamento da região sul baiana, com imagens do satélite QuickBird e WorldView, obtidas a partir de 2011 e de imagens do RapidEye obtidas entre os anos de 2009 e 2010. Estas imagens aliadas ao processo de checagem das áreas em campo ajudaram a construir um mapa de uso da terra de 3.470 km^2 (Figura 1A-B). Este mapa foi utilizado para identificar e quantificar as diferentes classes do uso do solo na escala 1:10. 000. Para estimar a porcentagem de cobertura vegetal na paisagem, foi considerado apenas o montante de floresta nativa. A paisagem foi definida como uma área circular em torno de um ponto central localizado dentro de um fragmento de floresta (Ver Morante-Filho et al. 2015). Assim, foi calculada a porcentagem de cobertura florestal para 40 paisagens dentro da área mapeada, das quais,

selecionamos 17 que possuíam distância mínima de 1 km entre elas e perfaziam um gradiente de cobertura florestal. Para as áreas amostradas em São Paulo, utilizamos imagens de alta resolução do Google Earth versão 7.3.2 através do OpenLayers plugin in Quantum GIS 2.14.3 (QGIS Development Team 2016). Em seguida, classificamos visualmente as imagens em floresta e não floresta e quantificamos a porcentagem de floresta em cada uma das seis áreas em estudo, seguindo os mesmos procedimentos realizados nas localidades da Bahia.

Posteriormente, como é conhecida a importância da escala espacial em análise de paisagem (Rocha-santos et al. 2016), calculamos a porcentagem de floresta com intervalos a cada 250m, com os raios variando entre 500m - 2 km, totalizando seis escalas de cobertura florestal. Para selecionar a escala mais adequada para as análises de diversidade gênica, utilizamos a função multifit no programa R com base no Critério de Informação de Akaike (AIC) (Huais 2018). Por fim, foi selecionado o modelo com melhor ajuste, correspondente a cobertura florestal na escala de 2 km.

2.3. Amostragem de *E. edulis* nas paisagens

Para este estudo, utilizou-se uma abordagem de fragmento-paisagem (Fahrig 2013), instalando três parcelas de 50x10m em um fragmento localizado no centro de cada paisagem entre dezembro de 2015 e fevereiro de 2016. A escolha deste tamanho amostral se justifica porque um estudo demografico de *E. edulis* no sul da Bahia demostrou que a utilização deste tamanho é suficiente para a coleta de 50 indivíduos da classe jovem II, que são os indivíduos entre 0.15 – 1.00 m de altura de inserção das folhas, os mais afetados pela redução de floresta na paisagem (Oliveira 2014). O número de indivíduos coletados por área foi estabelecido visando maximizar o poder de acurácia na detecção de SNPs na estratégia de sequenciamento em pool seq utilizada no estudo (Schlötterer et al. 2014), mas, em algumas áreas foram coletados menos que 50 indivíduos (Tabela 1). A alocação da primeira parcela foi aleatória dentro do fragmento com a distância mínima de 30m da borda e entre cada uma das três parcelas. Em seguida, coletamos material foliar de todos os indivíduos da classe jovem II de *E. edulis*. Nas áreas com mais de 50 indivíduos coletados na soma das três parcelas, realizamos sorteios para selecionar 50 indivíduos aleatórios para as análises genéticas em cada uma das 23 paisagens (17 da Bahia e 6 de São Paulo).

2.4. Extração de DNA genômico e utilização de sondas para captura de sequências do genoma de *E. edulis*

O DNA genômico foi extraído dos tecidos foliares dos 1.113 indivíduos coletados nas 23 áreas de estudo (Tabela 1), usando o protocolo Doyle and Doyle (1990). Em seguida, todas as amostradas foram quantificadas em gel de agarose 0,8% e, quantidades equimolares dos DNAs dos indivíduos advindos de uma mesma área de coleta foram usados para formar 23 pools, um para cada população, separadamente. Optamos pela estratégia de pool-seq, visto sua eficiência e custos reduzidos para obtenção de dados de SNPs em estudos de genômica populacional, que geralmente requer um grande número de amostras. Esta metodologia tem sido relatada como a melhor para reduzir os custos relacionados com o sequenciamento de última geração em análise de espécies não modelo (Schlötterer et al. 2014), como é o caso de *E. edulis*.

Em seguida, foram preparadas bibliotecas com barcodes para cada uma das 23 amostras em pool e, utilizadas 20 mil sondas específicas que foram previamente desenvolvidas para capturar regiões transcritas do genoma de *E. edulis* (Lopes et al. 2015 – dados não publicados). O DNA capturado por essas sondas foi sequenciado usando o Illumina HiSeq 2000 (Illumina, San Diego, CA, EUA). Após o sequenciamento, a extremidade 3' foi cortada para remover bases de baixa qualidade (índice de qualidade de leitura <20), e as *reads* foram filtradas para remover aquelas com mais de 10% do índice de qualidade. Posteriormente, as *reads* selecionadas após as etapas de "*trimagem*" e filtragem foram alinhadas usando MOSAIK 2.2. (Lee et al. 2014).

2.5. Identificação de Polimorfismo de Nucleotídeo Único (SNP)

Para identificar os SNPs foi usado o programa Freebayes 0.9.15 (Garrison and Marth 2012), considerando o número de amostras contidas em cada pool e utilizando a opção de pooled-discrete. O programa VCFtools 0.1.12a (Danecek et al. 2011) foi usado para filtrar os SNPs restantes, usando o filtro de qualidade (minQ 10, max-missing-count 3, min-alleles 2, max-alleles 2, min-meanDP 3, max- meanDP 750, maf 0.01, mac 1). Após esses dois filtros, foram obtidos 323.839 SNPs. Contudo, como

ainda existia uma variação grande na qualidade de mapeamento (QUAL) e profundidade de sequenciamento (DP), principais parâmetros utilizados na abordagem de *PoolSeq*, foi necessário realizar um filtro adicional, retendo apenas os marcadores com QUAL ≥ 100 e DP ≥ 100 (Schlötterer et al. 2014). Em seguida, foram retidos apenas os marcadores com QUAL e DP de sequenciamento entre os quantis de 5% e 95% (i.e, retenção de 90% dos marcadores com base nas distribuições de QUAL e DP). Destes, optou-se por reter apenas um SNP com o maior DP para cada um dos *contigs*, buscando minimizar possível desequilíbrio de ligação, resultando em 8.296 SNPs. Além disso, devido os pacotes computacionais considerar apenas SNPs bialélicos, um total de 7.632 marcadores foram utilizados para as análises genéticas.

2.6. Detecção de marcadores SNPs *outliers*

Os métodos para identificação de sinais de seleção em SNPs possuem limitações inerentes aos seus pressupostos e modelos genéticos que não necessariamente refletem a estrutura genética populacional em questão e outras covariáveis, revelando falso-positivos (Lotterhos and Whitlock 2015). Assim, para minimizar os impactos de falso-positivos, e tirando vantagem de diferentes pressupostos, foram utilizados três abordagem (*SelEstim*, *pcadapt* e *BayPass*) na identificação de marcadores SNPs sob seleção (*outliers*).

O programa *SelEstim* é baseado em estatística Bayesiana e considera a aproximação por difusão da distribuição das frequências alélicas em uma população subdividida em um número de demes que trocam migrantes entre si (Vitalis et al. 2014), como descrito no modelo de ilhas de Wright (1931). Para a identificação de SNPs *outliers* foram realizadas previamente 25 análises piloto com 5.000 iterações da Cadeia de Markov Monte Carlo (MCMC) para a calibração dos parâmetros estimados pelo programa. Posteriormente, foram realizadas 200.000 iterações *burn-in* seguidas de 500.000 iterações MCMC, amostradas a cada 40 iterações. A convergência e autocorelação da análise foram verificadas com o pacote *coda* (Plummer et al. 2006) na plataforma R (R Core Team 2018). Foram considerados SNPs *outliers* os marcadores com estimativa de divergência Kullback-Leibler (KLD) ≥ 2.3186, correspondente a alpha ≤ 0.01.

Em contraste, o método implementado no programa *pacadapt* não assume um modelo genético explícito e não requer grupos de indivíduos pré-definidos (Luu et al.

2017). Ele realiza *scans* genômicos por meio de análises de componentes principais (PCA) e assume que marcadores *outliers* estarão relacionados com a estrutura populacional sugerida pela PCA. A cada marcador, foi atribuído um valor de z-score que quantifica a relação do marcador com os primeiros *K* componentes principais que melhor explicam a estrutura genética observada. Com base nos z-scores, foi calculada uma distância de Mahalanobis para cada marcador em relação aos demais marcadores. Assim, foram considerados SNPs *outliers* aqueles que apresentaram desvios significativos destas distâncias em relação à média dos valores dos demais marcadores. Esta análise foi realizada com o pacote *pcadapt* (Luu et al. 2017) na plataforma R (R Core Team 2018). Para essa análise, foram considerados $K = 5$ componentes principais, que representou o ponto em que a inclusão de mais componentes não aumentou a proporção cumulativa de variação explicada pela PCA. Como ponto de corte, foram considerados SNPs *outliers* os marcadores com q-valores ≤ 0.1, assumindo que 10% dos SNPs *outliers* podem ser falso-positivos.

A estimativa XtX foi implementada no programa *BayPass,* considerando explicitamente a estrutura de covariância entre as frequências alélicas das populações (Ω), resultante de sua história evolutiva compartilhada (Günther and Coop 2013; Gautier 2015). Com base em dados simulados usando as estimativas iniciais de Ω, novas estimativas de XtX foram calculadas para a definição dos valores limiares para a identificação dos locos *outliers*. Assim, as estimativas de covariância das frequências alélicas entre populações foram obtidas a partir de 20 análises piloto com 2.000 iterações de MCMC, seguidas por um período de burn-in de 50.000 e 500.000 iterações de MCMC. Após a obtenção da matriz de covariância de frequências alélicas entre populações, a calibração dos valores de XtX foi realizada com as mesmas configurações considerando 5.000 marcadores simulados com a função *simulate.baypass*. A partir desta análise, foram considerados locos *outliers,* os marcadores com estimativas de XtX acima do percentil de 99% das estimativas de calibração.

Nesse estudo, foram considerados como marcadores hipoteticamente sob seleção (SNPs *outliers*), aqueles identificados por pelo menos dois destes três testes simultaneamente. Posteriormente, o número e a sobreposição de SNPs *outliers* identificados por esses métodos, foram representados através do diagrama de Venn, usando o pacote VennDiagram em programa R (Chen and Boutros 2011).

2.7. Estimativa da diversidade genética e análises estatísticas

Para estimar a diversidade e o agrupamento genético das 23 áreas de estudo, utilizamos apenas o subconjunto contendo os SNPs neutros. Com base nas frequências alélicas desses SNPs, foi estimada a diversidade gênica de Nei (1978), corrigida para tamanhos populacionais finitos, para cada loco e para cada população no programa R (R Core Team 2018), de acordo com a fórmula: $h = \frac{2n}{2n-1}(1 - \sum p_i^2)$. Onde h = diversidade gênica, n = número de indivíduos por população, p = frequência alélica do loco i. Para obter a distribuição das populações ao longo da diversidade genética com base nas frequências alélicas desses SNPs, utilizamos a análise de componentes principais (PCA), no pacote *LEA* (Frichot and François 2015) no programa R (R Core Team 2018). Posteriormente, utilizamos uma análise de ANOVA desbalanceada seguida pelo teste de *ScottKnoot* para avaliar se existia diferença na diversidade gênica entre os grupos indicados pela PCA.

Para avaliar se a disponibilidade de aves dispersoras de sementes de *E. edulis* poderia afetar na diversidade gênica, utilizamos como variável preditora a riqueza de aves potencialmente dispersoras de sementes da espécie na região do estudo (Morante-Filho et al. 2015; Carvalho 2017). Para avaliar a influência da porcentagem de floresta na paisagem e da riqueza de aves na diversidade gênica nas 22 áreas que possuíamos informações acerca dessas variáveis, utilizamos o modelo de efeitos mistos lineares do pacote lme4 do programa R (Bates et al. 2015), considerando como fator aleatório os grupos genéticos indicados na PCA. Assim, foram construídos os modelos combinando as variáveis preditoras em questão. Em seguida, colocamos os modelos para concorrer e selecionamos o melhor modelo através do Critério de Informação de Akaike (AIC) para pequeno tamanho amostral (Ver Anderson and Burnham 2004). Após selecionar o melhor modelo, ele foi comparado com o modelo nulo através do teste de ANOVA no programa R.

3. Resultados

Dos 7.632 SNPs utilizados para as análises genéticas, 91 SNPs foram identificados como *outliers* pelo método *SelEstim*, 1.266 SNPs pelo método *pcadapt* e 93 SNPs pelo método XtX. Quando considerado a sobreposição de pelos menos dois

dos três métodos, foram identificados 142 SNPs *outliers* (Figura 1, apêndice), e 7.490 SNPs neutros.

A diversidade gênica nas 23 populações variou entre 0.16 – 0.22 com média de 0.18 (Tabela 1). Na análise de PCA, com base nas frequências alélicas dos SNPs neutros, foi observado que as populações se agruparam de acordo com a localidade de origem (FS_SP, FO_SP, RS_BA e RN_BA) em quatro grupos, com os dois eixos da PCA explicando 50.58% da variação (Figura 2). Nós observamos através do teste de ScottKnoot diferença significativa na diversidade gênica entre os grupos, com os maiores valores encontrado para FS_SP, FO_SP, RS_BA e RN_BA, respectivamente (Figura 3).

Quando avaliada a escala de paisagem com maior efeito sob a diversidade gênica, o modelo com 2 km teve o maior poder explicativo (Tabela 2, apêndice). Na análise utilizando os modelos lineares de efeito misto, a riqueza de aves teve maior pode explicativo que a porcentagem de floresta na paisagem (Tabela 3, apêndice), mas não influencia significativamente na diversidade gênica de *E. edulis* (Tabela 4, apêndice).

4. Discussão

Nesse estudo, foram utilizadas 20 mil sondas de capturas desenvolvidas para regiões gênicas do genoma de *E. edulis* e identificados 7.490 SNPs neutros, empregados para acessar a diversidade gênica de 23 populações em pool-seq. Essas populações apresentaram alta diversidade gênica, independente da quantidade de floresta remanescente na paisagem ou da riqueza de aves potencialmente dispersoras de sementes. Assim, foi possível realizar, pela primeira vez, um estudo de genômica em escala de paisagem para uma espécie nativa da Mata Atlântica Brasileira, utilizando uma amostragem robusta a nível genômico, de indivíduos e populações. Dessa maneira, esse trabalho inédito representa um importante avanço para conservação de uma espécie ameaçada de extinção e de grande importância ecológica e econômica (CNCFlora 2012).

Em relação às análises para detecção de SNPs *outliers*, elas tiveram uma variação muito grande no número de SNPs identificados, demonstrando a importância da utilização simultânea dos três métodos, minimizando possíveis falsos-positivos ou falsos-negativos (Lotterhos and Whitlock 2015; Ahrens et al. 2018). Assim, garantindo

uma maior confiabilidade na distinção dos SNPs neutros daqueles sob ação da seleção natural, permitindo que fossem prosseguidas as análises genômicas e estatísticas utilizando apenas os SNPs considerados neutros. Além disso, importante ressaltar que esses SNPs *outliers* possibilitarão estudos futuros de associação com variáveis ambientais, representando um avanço na compreensão da adaptação de populações de *E. edulis* em paisagens antrópicas.

As 23 populações de *E. edulis* apresentaram elevada diversidade gênica, como esperado para uma espécie com ampla distribuição geográfica (Lowe et al. 2018). Assim, a alta diversidade gênica reportada pode ser essencial para garantir uma maior resiliência frente ás ações antrópicas que essas populações estão expostas, possibilitando um maior ajuste adaptativo ao longo do tempo (Kawecki and Ebert 2004; Orr 2005; Mable 2018). Além disso, é importante ressaltar que as populações em estudo apresentaram maior diversidade gênica se comparadas com outras populações empregando SNPs para a mesma espécie (Brancalion et al. 2018). No entanto, como utilizamos um maior número de SNPs e de indivíduos por população, a maior diversidade gênica pode ser devido a esses fatores amostrais. Além disso, importante considerar também que utilizamos sondas específicas para capturar regiões gênicas e sequenciamento em pool de amostras, enquanto Brancalion e colaboradores (2018) utilizaram a técnica de DNA associado à restrição de dupla digestão (ddRAD). Assim, a maior diversidade gênica registrada no presente trabalho pode ser devido ao nosso maior esforço amostral a nível genômico e de indivíduos, associada à eficiência do método pool-seq utilizando sequências alvos (Gautier et al. 2013; Schlötterer et al. 2014). Importante salientar ainda que as 23 populações formaram quatro grupos, sendo dois no estado de São Paulo (Floresta Ombrófila- FO_SP; Floresta Semidecídua- FS_SP) e dois no estado da Bahia (Região Sul, porção Norte - RN_BA; Região Sul, porção Sul - RS_BA), com as frequências alélicas explicando 50.58% da variação genética entre as populações.

Quando avaliados os grupos, foi observada maior diversidade gênica para FS_SP e FO_SP se comparados com os grupos RS_BA e RN_BA. Essa diferença pode ser decorrente de diversos fatores ecológicos e evolutivos atuando conjuntamente nessas populações de *E. edulis* (Ellegren and Galtier 2016; Carvalho et al. 2017; Lowe et al. 2018). No entanto, nossa hipótese é que a diferença na diversidade gênica encontrada entre os grupos de SP e BA esteja relacionada principalmente com as características dos morfotipos da espécie (Coelho 2010; Melito et al. 2014). No geral, os dois grupos da

BA são formados principalmente por indivíduos de *E. edulis* que possuem bainhas na color amarelada ou vermelha, conhecido como ecótipo “Bahia” (Bovi et al. 1987; Coelho 2010; Melito et al. 2014). Esse ecótipo possui menor densidade de indivíduos e baixa produção de sementes se comparado com as populações de *E. edulis* com bainhas na cor verde observados em SP (Silva et al. 2009; Melito et al. 2014). Assim, acreditamos que a menor quantidade de indivíduos adultos disponíveis para reprodução, associada com uma menor produção de sementes, possa ser um dos possíveis mecanismos que explique a menor diversidade gênica de *E. edulis* nas populações amostradas na BA. Além disso, um estudo baseado em marcadores microssatélites demonstrou diferença genética entre populações constituídas por *E. edulis* com bainha verde em relação ao ecótipo “Bahia” (Coelho 2010). Dessa maneira, acreditamos que as características demográficas inerentes desse ecótipo seja uma das possíveis explicações para sua menor diversidade gênica, que pode estar relacionada também a adaptação dos seus indivíduos para os solos pobres em nutrientes, comumente observados na região sul da Bahia (Silva et al. 2009; Brancalion et al. 2012; Melito et al. 2014).

Quando considerado os grupos dentro de cada estado (BA ou SP), nossa hipótese é que a extração ilegal da espécie esteja moldando a diversidade gênica por meio da redução de parentesco entre os indivíduos. Estudos demográficos realizados com *E. edulis* tem demonstrado um maior número de indivíduos adultos na porção norte da Bahia (grupo RN_BA) se comparada com a porção sul (grupo RS_BA) e, maior número de indivíduos nas populações do grupo FO_SP em relação ao grupo FS_SP (Oliveira 2014; Carvalho 2017). Além disso, considerando que as populações dos grupos FS_SP e RS_BA estão inseridas em paisagens altamente desmatadas (ver Figura 1), facilitaria o acesso de pessoas para realizar extração ilegal de palmito em indivíduos adultos de *E. edulis* (Gardner et al. 2009; Laurance et al. 2011). Assim, considerando que indivíduos geograficamente próximos são mais aparentados e facilmente explorados, a extração ilegal poderia aumentar a distância física e reduzir o parentesco médio entre os indivíduos adultos remanescentes (Santos et al. 2015). Com isso, um posterior cruzamento entre os indivíduos remanescentes poderia aumentar a diversidade gênica nos indivíduos jovens por meio da recombinação genética (Kramer et al. 2008). Dessa forma, explicaria a maior diversidade gênica encontrada no grupo de plantas em estágio juvenil II de RS_BA em relação ao grupo RN_BA e, a maior diversidade gênica do grupo FS_SP se comparado com o grupo FO_SP.

Embora possa haver um aumento inicial da diversidade gênica com o cruzamento de indivíduos menos aparentados nessas populações dos grupos FS_SP e RS_BA (Kramer et al. 2008). Os jovens podem apresentar um maior grau de parentesco entre si, visto que eles são oriundos de um menor número de indivíduos adultos (Oliveira 2014; Carvalho 2017). Dessa maneira, espera-se que a cada geração aumente mais o grau de parentesco, e uma consequente erosão da diversidade gênica por depressão endogâmica nessas populações que muitas vezes estão inseridas em fragmentos pequenos e isolados (Mable 2018; Ver Figura1). Acrescenta-se ainda, o maior isolamento genético entre as populações dos grupos FS_SP e RS_BA, se comparadas com as populações dos grupos FO_SP e RN_BA (Ver Figura 2). Deste modo, esse padrão observado reforça nossa hipótese que as populações dos grupos FS_SP e RS_BA podem ser vulneráveis à erosão genética em longo prazo, apesar da elevada diversidade gênica. Assim, o fluxo gênico entre as populações inseridas dentro desses grupos pode ser de fundamental importância para manutenção da diversidade genética á longo prazo (Santos et al. 2016).

Por fim, foi observado que a porcentagem de floresta na paisagem nem a riqueza de aves dispersoras de sementes afetaram a diversidade gênica de *E. edulis*, como esperado (Galetti et al. 2013; Carvalho et al. 2015, 2016). Assim, embora seja conhecido que essas variáveis sejam importantes em processos microevolutivo de *E. edulis* (Galetti et al. 2013; Carvalho et al. 2015, 2016), elas não são um *drive* determinante para manutenção da diversidade gênica da espécie. Dessa maneira, acreditamos que a diversidade gênica seja determinada principalmente pela ampla distribuição da espécie na Mata Atlântica, o que lhe confere alta plasticidade fenotípica da espécie.

Em resumo, as populações de *E. edulis* apresentaram alta diversidade gênica que provavelmente está sendo moldada por diversos processos ecológicos e evolutivos simultaneamente, que estão diferindo entre as regiões de estudo e entre diferentes morfotipos da espécie (Coelho 2010; Galetti et al. 2013; Melito et al. 2014; Carvalho et al. 2017; Mable 2018). Embora a diferenciação morfológica possa ser devida á plasticidade fenotípica esperada para espécies com ampla distribuição geográfica, como é o caso de *E. edulis*, a adaptação local também pode ocasionar diferenciação entre populações (Bragg et al. 2015). Portanto, ainda que a plasticidade fenotípica seja importante para sobrevivência da espécie, existem diferenças genéticas entre os morfotipos que possivelmente estão relacionadas à adaptação local (Coelho 2010;

Nicotra et al. 2010; Melito et al. 2014). Dessa maneira, considerando a amplitude e as diferentes características ambientais e ecológicas de ocorrência da espécie, estudos futuros devem investigar á importância da plasticidade fenotípica para adaptação local da espécie (Rellstab et al. 2017).

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo financiamento da pesquisa (Universal 426828/2016-0) e bolsa de produtividade em pesquisa da FAG; Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão de bolsa ao ASS; Pro-Reitoria de Pesquisa e Pós-Graduação da UESC pelo financiamento da pesquisa (PROPP 00220.1100.1320).

Referências

Ahrens CW, Rymer PD, Stow A, et al (2018) The search for loci under selection : trends, biases and progress. Mol Ecol. doi: 10.1111/mec.14549

Anderson DR, Burnham K (2004) Model selection and multi-model inference. Second. NY: Springer-Verlag, 63

Andrade ER, Jardim JG, Santos BA, et al (2015) Forest ecology and management effects of habitat loss on taxonomic and phylogenetic diversity of understory Rubiaceae in Atlantic forest landscapes. For Ecol Manage 349:73–84. doi: 10.1016/j.foreco.2015.03.049

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw. doi: 10.18637/jss.v067.i01

Benchimol M, Mariano-neto E, Faria D, et al (2017) Translating plant community responses to habitat loss into conservation practices_ Forest cover matters. Biol Conserv 209:499–507. doi: 10.1016/j.biocon.2017.03.024

Benestan LM, Ferchaud A-L, HOHENLOHE PA, et al (2016) Conservation genomics of natural and managed populations : building a conceptual and practical framework. Mol Ecol 25:2967–2977

Bovi ML, Godoy Jr G and Sáes LA (1987) Pesquisas com o gênero Euterpe e Bactrisno Instituto Agronômico de Campinas. O agronômico 39(2): 129-174

Bragg JG, Supple MA, Andrew RL, Borevitz JO (2015) Genomic variation across landscapes : insights and applications. New Phytol 207:953–967.

Brancalion PHS, Oliveira GCX, Zucchi MI, et al (2018) Phenotypic plasticity and local adaptation favor range expansion of a Neotropical palm. Ecol Evol 8:7462–7475. doi: 10.1002/ece3.4248

Brancalion PHS, Vidal E, Lavorenti NA, et al (2012) Soil-mediated effects on potential Euterpe edulis (Arecaceae) fruit and palm heart sustainable management in the Brazilian Atlantic Forest. For Ecol Manage 284:78–85

Ćalić I, Bussotti F, Martínez-García PJ, Neale DB (2016) Recent landscape genomics studies in forest trees — what can we believe? Tree Genet Genomes 12:3. doi: 10.1007/s11295-015-0960-0

Carvalho C da S, Ballesteros-Mejia L, Ribeiro MC, et al (2017) Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. Conserv Genet 18:467–478. doi: 10.1007/s10592-016-0921-7

Carvalho CDS (2017) Variação fenotípica , genética e dispersão de sementes. Rio Claro. Universidade Estadual Paulista

Carvalho CS, Galetti M, Colevatti RG, Jordano P (2016) Defaunation leads to microevolutionary changes in a tropical palm. Sci Rep 6:31957. doi: 10.1038/srep31957

Carvalho CS, Ribeiro MC, Côrtes MC, et al (2015) Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm. Heredity (Edinb) 115:216–224. doi: 10.1038/hdy.2015.30

Chen H, Boutros PC (2011) VennDiagram: A package for the generation of highly-customizable Venn and Euler diagrams in R. BMC bioinformatics, 12(1), 35

CNCFlora. *Euterpe edulis* in Lista Vermelha da flora brasileira versão 2012.2 Centro Nacional de Conservação da Flora. Disponível em <http://cncflora.jbrj.gov.br/portal/pt-br/profile/Euterpe edulis>. Acesso em 11 junho 2019

Coelho GM (2010) Estrutura genética populacional em morfotipos de Euterpe edulis. Universidade Estadual de Santa Cruz

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics, *27*(15), 2156-2158

Doyle JJ, JL Doyle (1987) Isolation of plant DNA from fresh tissue. Focus 12:13–15

Elena A, Rother DC, Brancalion PS, et al (2014) Forest ecology and management can overharvesting of a non-timber-forest-product change the regeneration dynamics of a tropical rainforest? The case study of Euterpe edulis. For Ecol Manage 324:117–125. doi: 10.1016/j.foreco.2013.09.001

Ellegren H, Galtier N (2016) Determinants of genetic diversity. Nat Rev Genet 17:422–433. doi: 10.1038/nrg.2016.58

Fahrig L (2013) Rethinking patch size and isolation effects: The habitat amount hypothesis. J Biogeogr 40:1649–1663. doi: 10.1111/jbi.12130

Fischer MC, Rellstab C, Tedder A, et al (2013) Population genomic footprints of selection and associations with climate in natural populations of Arabidopsis halleri from the Alps. Mol Ecol 22:5594–5607. doi: 10.1111/mec.12521

Fitzpatrick MC, Keller S (2015) Ecological genomics meets community-level modelling of biodiversity : mapping the genomic landscape of current and future environmental. Ecol Lett 18:1–16. doi: 10.1111/ele.12376

Frichot E, François E (2015) LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution, 6(8): 925-929

Galetti M, Guevara R, Côrtes MC, et al (2013) Functional extinction of birds drives rapid evolutionary changes in seed size. Science (80-) 340:1086–1091. doi: 10.1126/science.1233774

Gardner TA, Barlow J, Chazdon R, et al (2009) Prospects for tropical forest biodiversity in a human-modified world. Ecol Lett 12:561–582. doi: 10.1111/j.1461-0248.2009.01294.x

Garrison E, Marth G (2012) Haplotype-based variant detection from short-read sequencing.arXiv preprint: arXiv:1207.3907

Günther T and Coop G (2013) Robust identification of local adaptation from allele frequencies. Genetics, 195, 205–220

Gautier M (2015) Genome-wide scan for adaptive differentiation and association analysis with population-specific covariables. Genetics, 201: 1555-1579

Gautier M, Foucaud J, Gharbi K, et al (2013) Estimation of population allele frequencies from next-generation sequencing data: Pool-versus individual-based genotyping. Mol Ecol 22:3766–3779. doi: 10.1111/mec.12360

Huais PY (2018) multifit : An R function for multi-scale analysis in landscape ecology. Landsc Ecol 33:1023–1028. doi: 10.1007/s10980-018-0657-5

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecol Lett 7:1225–1241. doi: 10.1111/j.1461-0248.2004.00684.x

Kramer AT, Ison JL, Ashley M V., Howe HF (2008) The paradox of forest fragmentation genetics. Conserv Biol 22:878–885. doi: 10.1111/j.1523-1739.2008.00944.x

Lanes ÉC, Pope NS, Alves R, et al (2018) Landscape genomic conservation assessment of a narrow-endemic and a widespread morning glory from amazonian savannas. Front Genet 9:1–13. doi: 10.3389/fpls.2018.00532

Laurance WF, Camargo JLC, Luizão RCC, et al (2011) The fate of amazonian forest fragments : A 32-year investigation. Biol Conserv 144:56–67. doi: 10.1016/j.biocon.2010.09.021

Lee W-P, Stromberg MP, Ward A, et al (2014) MOSAIK: A hash-based algorithm for accurate next-generation sequencing short-read mapping.PLoS ONE9: e90581

Liu B, Su J, Chen J, et al (2013) Anthropogenic halo disturbances alter landscape and

plant richness: A ripple effect. PLoS One 8:1–8. doi: 10.1371/journal.pone.0056109

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Mol Ecol 24:1031–1046. doi: 10.1111/mec.13100

Lowe AJ, Breed MF, Caron H, et al (2018) Standardized genetic diversity life history correlates for improved genetic resource management of Neotropical trees. Divers Distrib 24:730–741. doi: 10.1111/ddi.12716

Luu K, Bazin E, Blum MG (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources, 17(1): 67-77

Mable BK (2018) Conservation of adaptive potential and functional diversity: Integrating old and new approaches. Conserv Genet 20:89–100. doi: 10.1007/s10592-018-1129-9

Martini AMZ, Fiaschi P, Amorim AM, Paixão JL da (2007) A hot-point within a hot-spot : A high diversity site in Brazil’s Atlantic Forest. Biodivers Conserv 16:3111–3128. doi: 10.1007/s10531-007-9166-6

Mcgill BJ, Dornelas M, Gotelli NJ, Magurran AE (2015) Fifteen forms of biodiversity trend in the anthropocene. Trends Ecol Evol 30:104–113. doi: 10.1016/j.tree.2014.11.006

Melito MO, Faria JC, Amorim AM, Cazetta E (2014) Demographic structure of a threatened palm (Euterpe edulis Mart.) in a fragmented landscape of Atlantic Forest in northeastern Brazil. Acta Bot Brasilica 28:249–258. doi: 10.1590/S0102-33062014000200011

Morante-Filho JC, Arroyo-Rodríguez V, Faria D (2016) Patterns and predictors of b - diversity in the fragmented Brazilian Atlantic forest : a multiscale analysis of forest specialist and generalist birds. J Anim Ecol 85:240–250. doi: 10.1111/1365-2656.12448

Morante-Filho JC, Faria D, Mariano-Neto E, Rhodes J (2015) Birds in anthropogenic landscapes: The responses of ecological groups to forest loss in the Brazilian Atlantic forest. PLoS One 10:e0128923. doi: 10.1371/journal.pone.0128923

Morante-Filho JC, Arroyo-Rodríguez V, Lohbeck M, et al (2016) Tropical forest loss and its multitrophic effects on insect herbivory. Ecology 97:3315–3325

Nei, M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590

Nicotra AB, Atkin OK, Bonser SP, et al (2010) Plant phenotypic plasticity in a changing climate. Trends Plant Sci 15:684–692

Oliveira LC (2014) Efeitos da perda de habitat na densidade de indivíduos e fenologia reprodutiva de Euterpe edulis na região sul da Bahia Ilhéus. Universidade Estadual de Santa Cruz

Orr HA (2005) The genetic theory of adaptation: a brief history. Nat Rev Genet 6:119–127. doi: 10.1038/nrg1523

Plummer M, Best N, Cowles K, Vines K (2006) CODA: Convergence diagnosis and output analysis for MCMC. R News 6: 7-11

R Core Team (2018) R: A language and environment for statistical computing. R foundation for statistical computing, Vienna. https://www.R-project.org

Rellstab C, Fischer MC, Zoller S, et al (2017) Local adaptation (mostly) remains local : reassessing environmental associations of climate-related candidate SNPs in Arabidopsis halleri. Heredity 118:193. doi: 10.1038/hdy.2016.82

Rellstab C, Gugerli F, Eckert AJ, et al (2015) A practical guide to environmental association analysis in landscape genomics. Mol Ecol 24:4348–4370. doi: 10.1111/mec.13322

Ribeiro MC, Metzger JP, Martensen AC, et al (2009) The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biol Conserv 142:1141–1153. doi: 10.1016/j.biocon.2009.02.021

Rocha-Santos L, Pessoa MS, Cassano CR, et al (2016) The shrinkage of a forest : Landscape-scale deforestation leading to overall changes in local forest structure. Biol Conserv 196:1–9. doi: 10.1016/j.biocon.2016.01.028

Santos AS, Cazetta E, Dodonov P, et al (2016) Landscape-scale deforestation decreases gene flow distance of a keystone tropical palm, Euterpe edulis Mart (Arecaceae). Ecol Evol 6:6586–6598. doi: 10.1002/ece3.2341

Santos AS, Cazetta E, Morante Filho JC, et al (2015) Lessons from a palm: genetic diversity and structure in anthropogenic landscapes from Atlantic Forest, Brazil. Conserv Genet 16:1295–1302. doi: 10.1007/s10592-015-0740-2

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat Rev Genet 15:749. doi: 10.1038/nrg3803

Schoville SD, Bonin A, François O, et al (2012) Adaptive genetic variation on the landscape : methods and cases adaptive genetic variation on the landscape : methods and cases. Annu Rev Ecol Evol Syst 43:23–43. doi: 10.1146/annurev-ecolsys-110411-160248

Silva MDGCPC, Martini AMZ, Araújo QRD (2009) Estrutura populacional de Euterpe edulis Mart. no Sul da Bahia, Brasil. Revista Brasileira de Botânica 32(2): 393-403

Slik JWF, Arroyo-Rodríguezb V, Aiba S-I, et al (2015) An estimate of the number of tropical tree species. Pnas 112:E4628–E4629. doi: 10.1073/pnas.1512611112

Sork VL (2017) Genomic studies of local adaptation in natural plant populations. J Hered 109:3–15. doi: 10.1093/jhered/esx091

Thomas WMW, Carvalho AMV de, Amorim AA, et al (1998) Plant endemism in two forests in southern Bahia , Brazil. Biodivers Conserv 322:311–322

Vitalis R, Gautier M, Dawson KJ, Beaumont MA (2014) Detecting and measuring selection from gene frequency data. Genetics 196: 799-817

Wright S (1931) Evolution in mendelian populations. Genetics 16: 97–159

LISTA DE TABELA E FIGURAS

Tabela 1. Porcentagem de cobertura florestal estimada nas 23 áreas amostradas no estudo**,** número de indivíduos coletados, estimativas de diversidade gênica (H_E) para os SNPs neutros e a riqueza de aves dispersoras de sementes de *E. edulis*.

Localidade	% de floresta	Nº	H_E	Aves
RS_BA	18	50	0.18	9
RS_BA	27	50	0.18	4
RS_BA	41	50	0.18	8
RS_BA	52	50	0.18	2
RS_BA	64	50	0.17	5
RS_BA	64	50	0.17	5
RS_BA	71	50	0.18	5
RN_BA	37	50	0.16	5
RN_BA	42	50	0.16	5
RN_BA	43	48	0.16	6
RN_BA	45	50	0.17	7
RN_BA	50	50	0.17	6
RN_BA	51	50	0.16	7
RN_BA	61	49	0.16	8
RN_BA	79	49	0.16	6
RN_BA	85	50	0.17	4
RN_BA	96	50	0.16	NA
FS_SP	16	30	0.22	9
FS_SP	19	47	0.22	8
FS_SP	27	45	0.20	6
FS_SP	49	49	0.21	13
FO_SP	64	50	0.20	20
FO_SP	100	46	0.20	21
Média	**52.21**	**48.39**	**0.18**	**7.68**

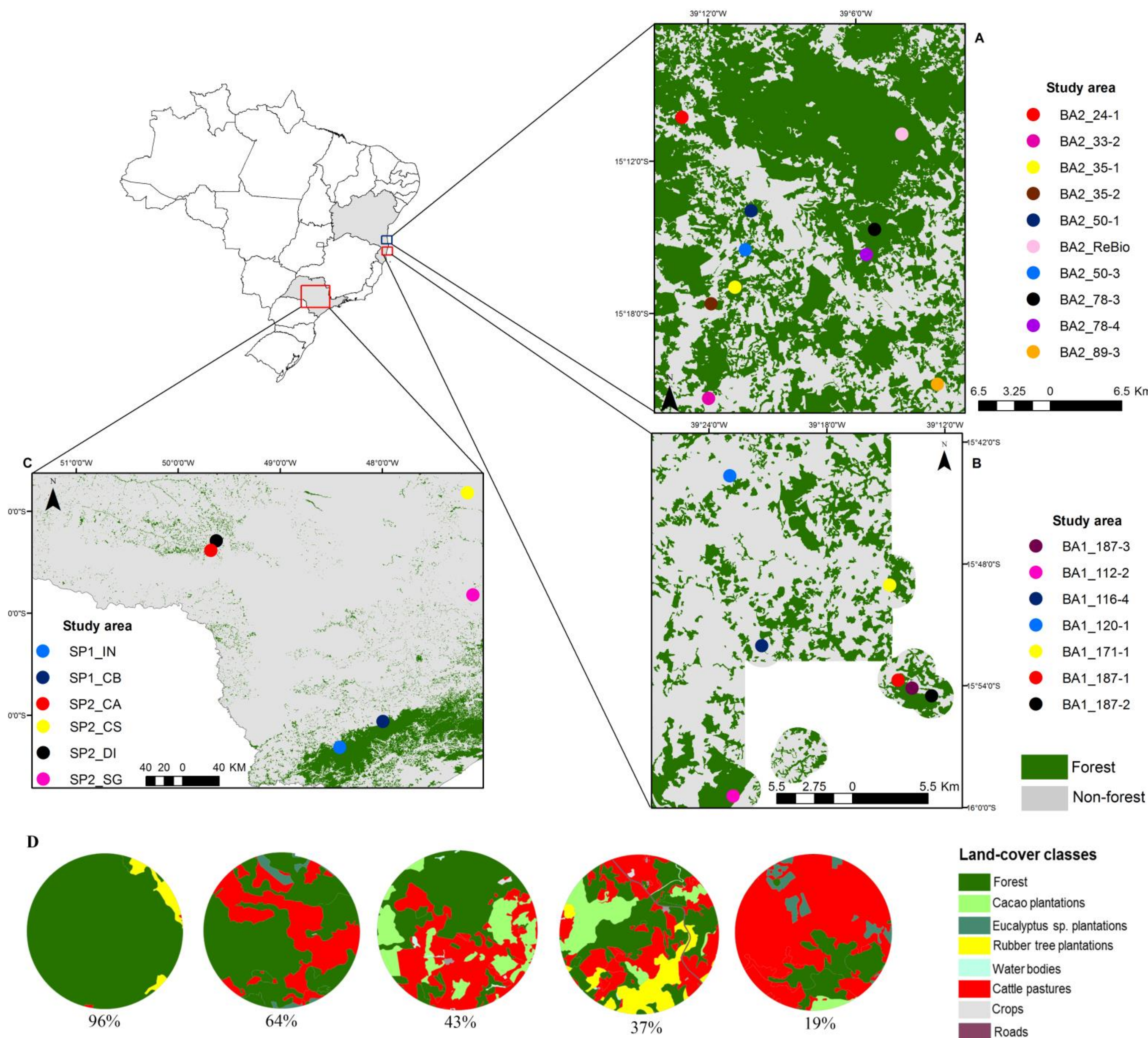


Figura 1. Distribuição das populações amostradas de E. edulis dentro dos estados da Bahia e São Paulo. **A)** Região sul da Bahia, porção Norte; **B)** Região sul da Bahia, porção Sul; **C)** Floresta Ombrófila (SP1_IN, SP1_CB) e Floresta Semidicidua (SP2_CA, SP2_CS, SP2_DI, SP2_SG) situadas no estado de São Paulo; **D)** Exemplos de paisagens com diferentes porcentagens de floresta e suas respectivas classes de uso da terra na escala de 2km.

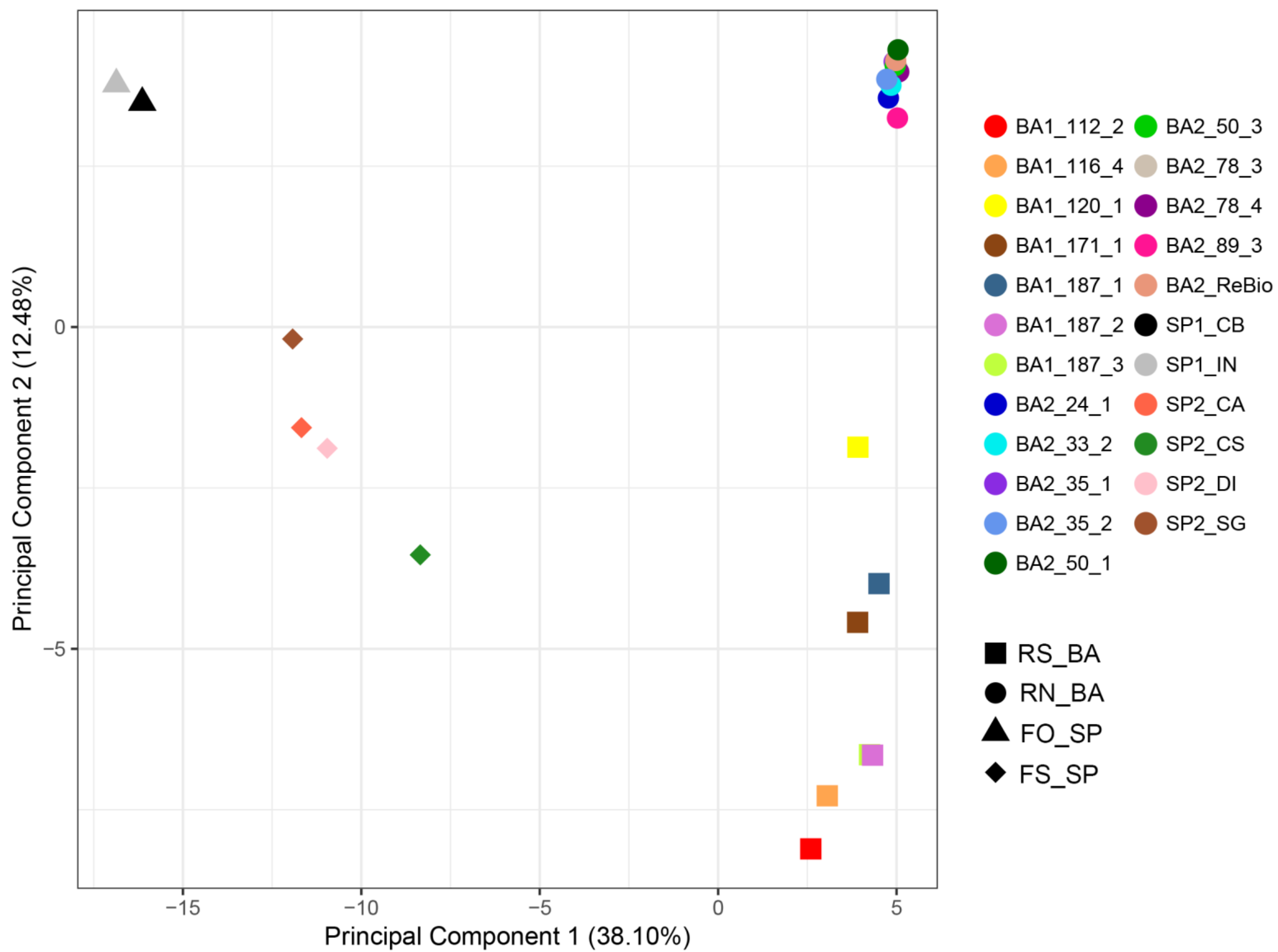


Figura 2. Distribuição das 23 populações de *E. edulis* com base nas frequências alélicas dos SNPs neutros, avaliadas por meio da Análise de Componentes Principais (PCA).

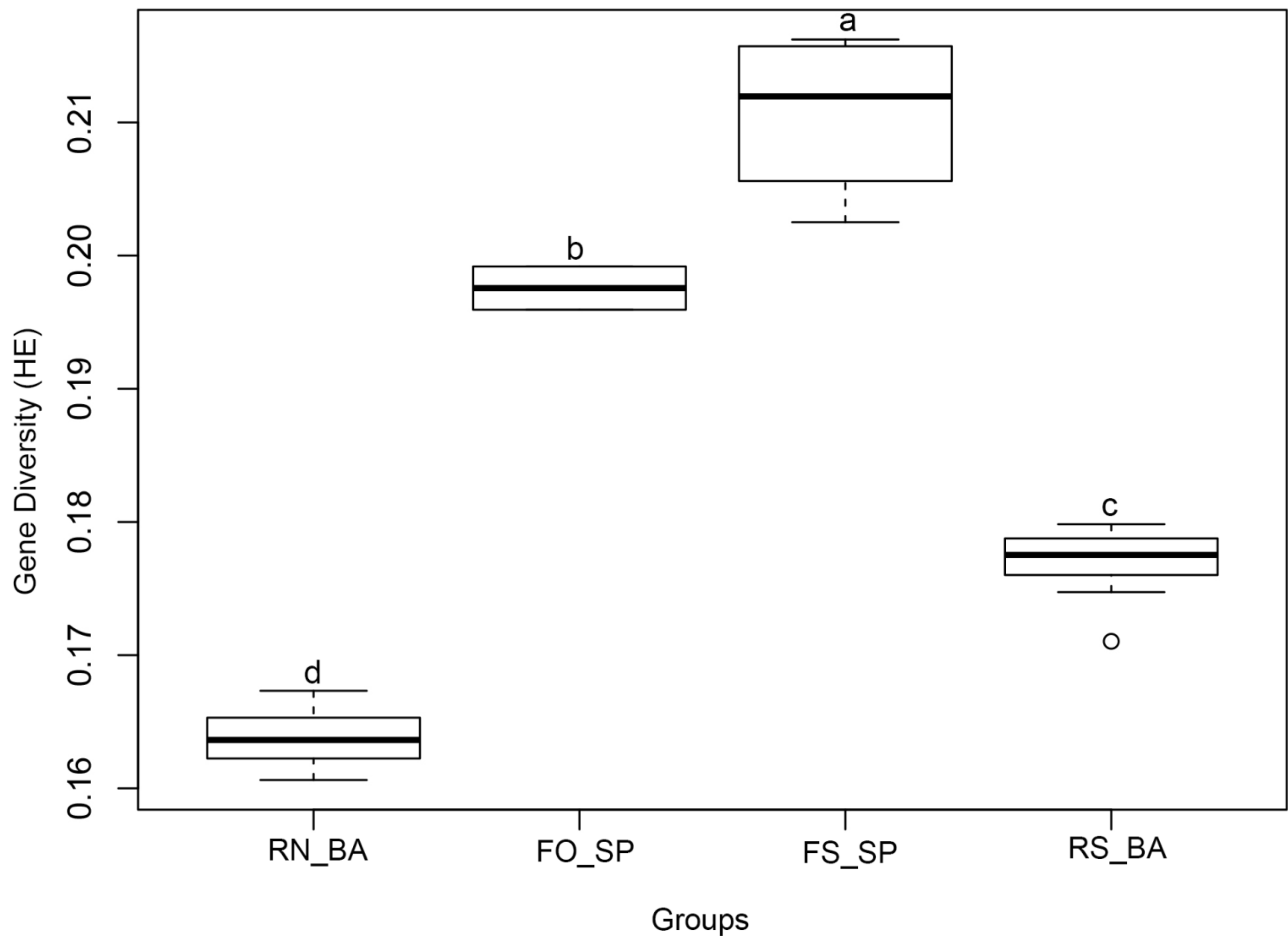


Figura 3. Boxplot da diversidade gênica com teste de média *ScottKnoot* para os quatros grupos formados pelas 23 populações de *E. edulis*.

APÊNDICE

Tabela 1: Resumo da seleção de modelo com base no Critério de Informação de Akaike (AIC) para as sete escalas de porcentagem de floresta na paisagem.

[1]Escala	AIC	Estimates	p.values
500m	-114.4661	8.866121e-05	0.53581339
750m	-114.0554	-2.200953e-05	0.89590611
1000m	-114.0970	-3.824190e-05	0.81580945
1250m	-114.5184	-1.145786e-04	0.51203198
1500m	-114.9915	-1.592652e-04	0.35604102
1750m	-116.1857	-2.329336e-04	0.16621368
2000m	**-117.1129**	**-2.722672e-04**	**0.09763152**

[1]Nota: Porcentagem de floresta para diferentes escalas, sendo a melhor escala destacada em negrito.

Tabela 2: Seleção de modelo para efeito da riqueza de aves e da porcentagem de floresta na paisagem sob a diversidade gênica (H_E) de *E. edulis*.

Modelo	AICc	dAICc	df	weight
HE ~ Riqueza de aves	**-137.8**	**0.0**	**4**	**0.82**
HE ~ % Floresta	-134.7	3.1	4	0.18
HE ~ % Floresta + Riqueza de aves	-117.9	19.8	5	<0.001

Tabela 3: Resultado na análise de ANOVA comparando o modelo nulo com o modelo contendo a riqueza de aves.

```
Models:
Modelo1: HE ~ 1 + (1 | grupo)

Modelo2: HE ~ aves + (1 | grupo)

         Df  AIC     BIC    logLik  deviance  Chisq Chi Df Pr(>Chisq)
Modelo1: 3  -162.78 -159.51  84.391  -168.78
Modelo2: 4  -161.57 -157.20  84.783  -169.57   0.7854    1    0.3755
```

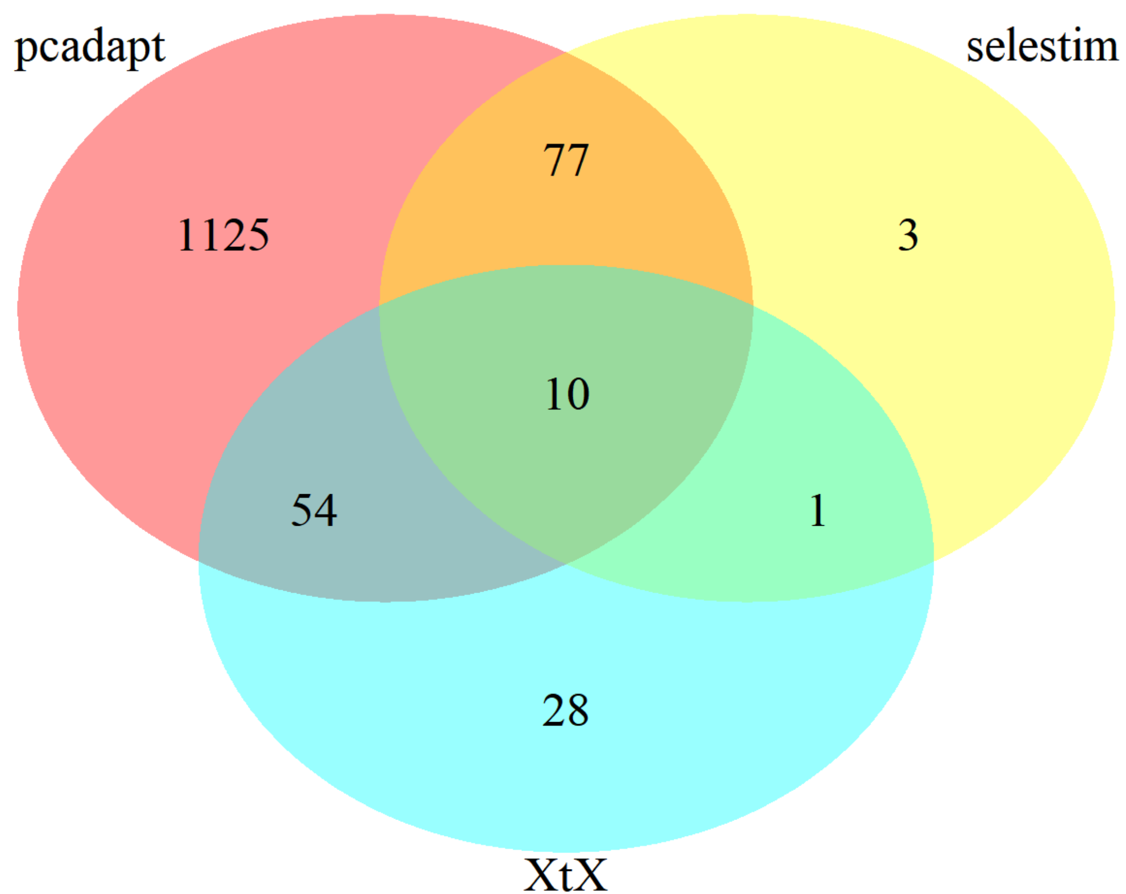


Figura 1. Diagrama de Venn para os SNPs detectados como *outliers* (sob seleção) utilizando o pcadapt, selestim e o XtX nas 23 populações de *E. edulis*.

Capítulo 3

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

Citação

Santos AS, Borges DB, Vivas CV, Berg CVD, Rodrigues PS, Tarazi R, Gaiotto FA. **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.** Accepted for publication in Genetics and Molecular Biology

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

Alesandro S. Santos[1], Daniela B. Borges[2], Caio V. Vivas[2], Cassio V.D. Berg[3], Polliana S. Rodrigues[4], Roberto Tarazi[5] and Fernanda Amato Gaiotto[4*]

[1]Laboratório de Ecologia Aplicada à Conservação, Pós-Graduação em Ecologia e Conservação da Biodiversidade, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna, km 16, Ilhéus, BA, Brazil, CEP 45662-900

[2]Pós-Graduação em Genética e biologia molecular, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna, km 16, Ilhéus, BA, Brazil, CEP 45662-900

[3]Universidade Estadual de Feira de Santana, Departamento de Ciências Biológicas, Av. Transnordestina s/n, 44030-900, Feira de Santana, BA.

[4]Laboratório de Marcadores Moleculares, Centro de Biotecnologia e Genética, Universidade Estadual de Santa Cruz, Rodovia Ilhéus-Itabuna, km 16, Ilhéus, BA, Brazil, CEP 45662-900

[5]Currently a Discovery Breeder at Bayer CropScience in Trindade, GO, Brazil.

*correspondent author: gaiotto@uesc.br

Abstract

Forest loss and fragmentation are the main threats to the maintenance of the Atlantic Forest, an important global biodiversity hotspot. Due to the current critical level of deforestation, it is strategic to establish ecological corridors to facilitate species dispersion and gene flow among fragments. Our study aimed to investigate the *Eschweilera ovata* genetic variability and the gene pool sharing of this tree in five forest remnants in southern Bahia, Brazil, using nuclear (nSSR) and plastid (cpSSR) microsatellite markers. The cpSSR marker analysis revealed the domain of four haplotypes, evidencing that 80% of the present individuals have only four maternal origins, which reflects a founder effect and or genetic bottleneck. The results obtained with the cpSSR and nSSR analyses indicate moderate genetic diversity, with an emphasis on the conservation units with full protection, which had the best parameters of all areas evaluated. Another indication of susceptibility of these populations to the process of forest loss and fragmentation is the strong genetic bottleneck observed. On the other hand, the genetic structure analyses (F_{ST} and discriminant analysis of principal components) revealed gene pool sharing between the subpopulations, which may be a reflection of the historical gene flow that occurred before the forest fragmentation.

Keywords: Functional connectivity, tree species, chloroplast, founder effect, SSR development.

1. Introduction

The Atlantic Forest in the east of Brazil is an important center of endemism associated with the high species richness of several taxa, especially tree species (Thomas *et al.*, 1998; Myers *et al.*, 2000; Martini *et al.*, 2007; Murray-smith *et al.*, 2008). Despite its notorious importance for biodiversity conservation, this phytogeographical domain has been reduced to approximately 11-16% of its original extent (Ribeiro *et al.*, 2009). This anthropogenic forest reduction represents one of the main threats to the permanence of species in this environment, reducing the resources and increasing the risks of population extinctions (Pessoa *et al.*, 2016; Benchimol *et al.*, 2017). Additionally, the negative effects of forest reduction on biodiversity can be potentialized through interactions, causing structural isolation and decreasing species dispersal and gene flow among forest remnants (Tabarelli *et al.*, 2004; Pardini *et al.*, 2010; Lima and Mariano-Neto, 2014).

Considering the conservation scope and the current state of the reduction and fragmentation of the Atlantic Forest, it is strategic to establish ecological corridors between the forest remnants (MMA, 2008), connecting them by the dispersion of species (Damschen *et al.*, 2006). Enabling gene flow between the fragments, corridors decrease the deleterious effects caused by isolation and the reduction of forest area that decrease populations and their genetic diversity, thus increasing the risks of local extinctions over time (Carvalho *et al.*, 2015).

In the last few decades, genetic studies with a conservationist approach have mainly used molecular markers such as microsatellites to access the variability and distribution of the genetic diversity of tree species (Gaiotto *et al.*, 2003; Santos *et al.*, 2016; Carvalho *et al.*, 2017; Torres-Florez *et al.*, 2017). Some of these studies used microsatellite markers with biparental and uniparental inheritance patterns to uncover

historical colonization patterns and the functional connection among tree species populations in the past (Martins *et al*., 2011; Zhang *et al*., 2012). Several studies have shown founder effects in plant populations by using chloroplastidial microsatellites (cpSSRs), which have maternal inheritance in most species of angiosperms and act as functionally haploid and non-recombinant markers (Corriveau and Coleman, 1988; Weising and Gardner, 1999; Parducci *et al*., 2001; Li *et al*., 2012; Tong *et al*., 2013). Additionally, studies combining nuclear and chloroplast microsatellite markers have shown convergence in the results, suggesting that they are complementary approaches (Zhang *et al*., 2012; Tóth *et al*., 2017).

In this way, due to the current critical state of the reduction and fragmentation of the Atlantic Forest, we evaluated the genetic variability and the gene pool sharing of a tree species, *Eschweilera ovata* (Cambess.) Mart. ex Miers (Lecythidaceae), in five forest remnants in southern Bahia, Brazil, using cpSSR and nSSR markers. These remnants are located in a region with high endemism and plant richness where there is an effort to implement ecological corridors that connect the forest fragments and conservation units (Fonseca *et al*., 2004). This tree species has a large occurrence (Atlantic Forest, Amazon Forest, Cerrado and Caatinga), with a zoochoric syndrome of seed dispersal and pollen. Besides that, it has been indicated for recovery of degraded areas. It is also economically exploited to make the berimbau, a musical instrument used in the capoeira orchestra and a cultural symbol of the state of Bahia, Brazil (Lorenzi, 1998). In this study, we tested the following hypotheses: (1) The subpopulations of *E. ovata* have high genetic diversity for both molecular markers, considering the wide occurrence and density of the species; (2) The subpopulations still share a gene pool, reflecting the functional connectivity prior to the process of forest loss and fragmentation.

2. Material and Methods

2.1 Study area

The study was conducted in five areas in southern Bahia, Brazil, a portion of the Atlantic Forest considered a priority for conservation actions due to its biological importance (Thomas *et al.*, 1998; Martini *et al.*, 2007). Using ArcGIS (10.2), a map of the study region, including all areas and forest remnants, was created (Figure 1). Four of the five locations are inside a protected area (PA): Reserva Biológica de Una (ReBio), which has 11,400 hectares in Una; Parque Municipal Boa Esperança (PMBE), which has 437 hectares, in the urban area of Ilhéus; Reserva Particular do Patrimônio Natural (RPPN) Mãe da Mata which has 13 hectares in Ilhéus; and RPPN Capitão, which has 660 hectares in Itacaré. The fifth area corresponds to an arboreal restinga (Restinga) located between the PMBE and ReBio in Ilhéus, which is not in a PA. In each study area, we collected the leaf samples of 15 *E. ovata* adult individuals for the genetic analysis and georeferenced the trees using GPS (GPS Garmin Map 62 s, USA). Sampling was performed to cover as much of each area as possible, avoiding the sampling of individuals geographically close to each other to reduce the chances of kinship between them and to sample the genetic pool of each collection area (Distance varying between 10 and 2.000m between the individuals sampled). As is known, the estimates the of populations genetics parameters are influenced by number and frequency of informative alleles (Frequency $\geq$ 0.05) (Hale *et al.*, 2012). For this reason, to evaluate the accuracy of estimated allelic frequencies with the 15 individuals sampled in each of the four subpopulations using nSSR markers, we calculated the allele frequencies and the mean allelic pattern (number of different alleles, number different alleles with a Frequency >= 5% and Shannon's information index) with standard

deviation using the GenAlex 6.5 program (Peakall and Smouse, 2012). Thus, if the number of individuals is influencing the estimates mentioned above, there will be great heterogeneity of these measures among the populations as a reflection of the sampling error.

2.2 DNA extraction, development of specific microsatellite primers and genotyping with cpSSR and nSSR

DNA was obtained according to the CTAB protocol (Doyle and Doyle, 1987), and the quantitation was performed in agarose gel (0.8%).

An enriched library was obtained by hybridization with $(CT)_8$ and $(GT)_8$ biotinylated probes, and the positive fragments for the microsatellites were amplified and sequenced on an ABI 3100 genetic analyzer (Applied Biosystems, Foster City, CA). We used the Primer3 software (Untergasser *et al.*, 2007) to design primer pairs.

To characterize the new nuclear microsatellite marker set (nSSR) and perform genetic analysis of populations, we used individual DNA sampled at four subpopulations (ReBio, MM, Cap and AR), totaling 60 trees. The PCR components were prepared with a mix containing 7.5 ng of genomic DNA, 10 µl of Top Taq Master Mix kit (Quiagen), 0.8 µM of the primer reverse, 0.4 µM of the forward primer, and 0.16 µM of M13 tail (CAC GACGTTGTAAAACGA) labeled with fluorochrome (6-FAM, VIC, PET, or NED, Applied Biosystems, Foster City, CA). The amplification reaction was placed on the Veriti 96-Well Thermal Cycler (Applied Biosystems, Foster City, CA) and consisted of an initial denaturation step at 94 °C for 1 min followed by 30 cycles of an denaturation step at 94 °C for 1 min, an annealing step at the specific annealing temperature for 45 s, and a extension step at 72 °C for 1 min; 8 cycles for the M13 tail amplification at 94 °C for 1 min, 53 °C for 45 s, and 72 °C for 1 min; and

followed by the final extension step at 72 °C for 10 min. The amplified fragments were injected into capillary electrophoresis in a multiload system by ABI 3500 genetic analysis (Applied Biosystems, Foster City, CA). The peaks were analyzed using the GeneMarker software 2.6.2 (SoftGenetics, State College, PA, USA).

Ten loci were amplified in all sampled individuals (75 trees) from the five subpopulations (ReBio, MM, Cap and AR) using chloroplastidial microsatellites (cpSSR) (Weising and Gardner, 1999) with a M13 tail and respective tail fluorophores (6FAM, VIC, NED and PET, Applied Biosystems, Foster City, CA). Amplifications were performed by PCR in a thermal cycler Life Pro, 96, gradient (TC-96 / G / H (B) A) Bioer, China). The final concentration of the PCR reagents in 13 µL was: 0.75 ng. µL-1 of genomic DNA; 0.325 mM of each deoxyribonucleotide (dATP, dCTP, dGTP, dTTP) (Invitrogen); 1.3X PCR buffer (Invitrogen), 4.0 mM MgCl2, 0.325 mg (Invitrogen); 0.2 mM forward primer with the M13 tail; 0.4 mM reverse primer; 0.06 mM M13 complementary primer with the fluorophore; 1 U Taq DNA polymerase (Invitrogen) and completed with MilliQ water to the final volume. The PCR was performed under the following conditions: a denaturation step at 94 °C for 2 min 30 cycles of a denaturation step at 94 °C for 1 min, an annealing step with a temperature gradient from 48 °C to 62 °C for each primer pair for 1 min and one extension step at 72 °C for 1 min. The second PCR step was programmed for 8 cycles of a denaturation step at 94 °C for 1 min, one annealing step at 53 °C, (specified for the tail primers) for 1 min, one extension step at 72 °C for 1 min, and one final extension step at 72 °C for 7 min. The PCR product genotyping was performed on a multiload system (ABI 3130xl, Applied Biosystems, Foster City, CA), and the GSLIZ500 was used as the size fragment standard (Applied Biosystems, Foster City, CA). To analyze the electropherograms and obtain the genotypes, the GeneMarker program version 2.2.0 was used.

2.3 Analysis of nuclear microsatellite markers (nSSR) and population genetics

To calculate the alleles per locus (A), probability of exclusion of paternity (Q) and identity index (I) we used the CERVUS 3.0.6 program (Kalinowski *et al.*, 2007). We used the divBasic function of the diveRsity package in the R software (Keenan *et al.*, 2013) to calculate allelic richness, expected heterozygosity (H_E), observed heterozygosity (H_O) and inbreeding coefficient (*f*).The confidence interval of H_E, H_O and *f* at 95% was calculated using divBasic function with 10000 bootstraps and to plot the results of these analysis, the ggplot2 package was used in the R software (Wickham, 2010). To verify the existence of genetic bottlenecks in the subpopulations, we employed the Wilcoxon test in software BOTTLENECK version 1.2.02 02 (Cornuet and Luikartt, 1996). As recommended for studies with few individuals per population and few loci (<20 loci), the Wilcoxon test with the infinite allele model (IAM), stepwise mutation model (SMM) and two phase model (TPM), which allows multiple-step mutations were applied (Piry *et al.*, 1999). For the TPM model, the proportion of SMM in TPM = 0.000, and variance of the geometric distribution for TPM = 0.36, which correspond to the most sensitive values for most microsatellites (Piry *et al.*, 1999).

The analysis of molecular variance (AMOVA) was calculated in GenAlex 6.5 (Peakall and Smouse, 2012). To evaluate the genetic structure of the subpopulations analyzed, we calculated F_{ST} using the divBasic function of the diveRsity package (Keenan *et al.*, 2013) and discriminant analysis of principal components (DAPC) (Jombart *et al.*, 2010) using the Adegenet 2.0.0 package in the R software (Jombart *et al.*, 2008). To calculate the 95% confidence interval of the F_{ST} with 10,000 bootstraps, we used the divBasic function of the diveRsity package in the R software. Posteriorly, to detect if the geographic distance (km) influenced the genetic distance between the

populations (F_{ST}), a simple linear regression analysis was performed in the R software (http://www.r-project.org/). To identify the number of clusters, we used the find.clusters function, retaining all main components, and the best fit cluster number was inferred through the Bayesian information criterion (BIC), selecting the lowest BIC as the ideal cluster. To describe the relationship between the identified clusters the generic function dapc was used, retaining the number of principal components that incorporated 80% of the cumulative variance (PCA eigenvalues) and the number of discriminant functions (DA eigenvalues) to maximize the variation between groups. In the graphical representation of the results obtained in the dapc analysis, we used the scatter function to assign individuals to clusters according to their gene pool and the compoplot function to produce a diagram representing the genetic pool of the individuals associated with the clusters of the subpopulations.

2.4 Population genetic analysis with chloroplastid microsatellite markers (cpSSR)

To analyse the cpSSR loci we used the Haplotype Analysis software version 1.04 (Eliades and Eliades, 2009), calculating the following genetic parameters: the number of haplotypes (N_H), number of unique haplotypes (P_H), number of effective haplotypes (H_{Ne}), haplotype richness (H), haplotypic diversity (H_E), total genetic differentiation (F_{ST}) between areas, and average genetic distance between individuals (D^2sh). The linear regression analysis was performed in the R software (http://www.r-project.org/) to evaluate if the pairwise genetic (F_{ST}) and geographical distances (km) are associated.

To visualize and analyze the genetic relationships without assuming a priori a cluster of individuals or populations, we used the EDENetworks program (Kivelä *et al.*, 2015). The networks analysis consist of nodes (individuals or populations) linked

according to their genetic relationships. In the construction of networks of subpopulations and individuals, automatic thresholding was used. The automatic thresholding was detected automatically by EDENetwork using the percolation threshold. The automatic thresholding used for network analysis was slightly below the percolation threshold, so that the network remained connected. In the network at the subpopulation level, Fst was used as distance with automatic threshold (0.42). In the network at the individuals level, allele sharing was used as distance with automatic thresholding (0.14).

3. Results

3.1 Development of microsatellite markers nuclear (nSSR) for *E. ovata* and population analysis

We developed 13 new nuclear primer pairs for *E. ovata* which had (i) an average of 5.6 alleles, (ii) H_O ranging between 0.214 - 1.000, (iii) H_E between 0.198 - 0.878, and (iv) a fixation index (F) in the range of -0.333 to 0.4763 (see Table S1 and Table S2, Supporting information). The combined probability of exclusion (Q) of these 13 loci was $0.9x10^{-9}$, and the identity index (I) was $3x28^{-10}$. The sequence, allele amplitude, and annealing temperature for each primer are in Table S1.

In the population analysis, it was verified that the 13 loci present similar distribution of the allelic frequencies, average number of alleles, frequency of informative alleles and estimated genetic diversity (Figure S1 and Figure S2, Supporting information). The four subpopulations showed low numbers of effective alleles (AR= 2.610; MM=2.883; Cap=2.685; ReBio= 2.841), low allelic richness (Figure 2) and presence of private alleles were observed in all subpopulations (AR =8, MM =5, Cap =4 and ReBio =2). The observed heterozygosity (Ho) values were

moderate to high (Figure 3). However, the Ho values were always higher than the expected heterozygosity (Figure 3), which was reflected in the negative values of the inbreeding coefficient (Figure 4) and significant genetic bottleneck in the four subpopulations (Table 1).

On the other hand, in the AMOVA, there were observed variances of 18% within and 80% among individuals, while there was only 2% variance among subpopulations. The F_{ST} analysis revealed that subpopulations had low and not significant genetic differentiation (Figure 5), despite of relatively large distance between some subpopulations (up to 89 km).

It was verified that geographic distance did not explain the pattern of genetic distance between subpopulations, as they were not significantly correlated ($R^2 = 0.207$, $p = 0.36$). The discriminant analysis of principal components revealed that the four collection areas form three genetic clusters (K = 3, BIC = 66.06), being a group formed by MM and ReBio and the other two groups formed separately by areas AR and CAP (Figure 6A), with sharing of the gene pool between the groups (Figure 6).

3.2 Population analysis for *E. ovata* using microsatellite markers plastid (cpSSR)

Seven of the 10 cpSSR primers tested in *E. ovata* were successfully amplified. Three were monomorphic (cpSSR1, cpSSR3 and cpSSR7), and four were polymorphic (cpSSR2, cpSSR4, cpSSR5 and cpSSR6), producing an average of three alleles per locus and 14 haplotypes. Of these haplotypes, four were found exclusively in PMBE (H1, H8, H9 and H14), two were exclusively found inside ReBio (H3 e H4), one in AR (H5), and another was exclusively found in Cap (H11). On the other hand, six of the 14 haplotypes (H2, H6, H7, H10, H12 and H13) were shared by two or more study areas. The dominance of four haplotypes (H10, H6, H3 and H13) was observed in these areas,

evidencing that 80% of the analyzed individuals only presented four maternal origins (see Table S3, Supporting information). We registered, for the intra-population analysis, an average of five haplotypes. The PMBE subpopulation had the highest number (7) of haplotypes, in contrast to the RPPN Cap subpopulation, which had the lowest (3) (Table 2). When we analyzed the number of effective haplotypes, the five areas had an average of 2.32, representing about half of the average number of haplotypes found. These subpopulations showed, on average, low richness (4) and haplotype diversity (0.55), with low genetic distances between individuals (0.70). Overall, the highest levels of genetic diversity were found in PMBE and ReBio (Table 2).

When we assessed the partitioned genetic diversity, we observed low levels of total genetic diversity (HT) and, consequently, low levels within subpopulations (HS), with higher values in PMBE and ReBio than in the other locations (Figure 7). In addition, the diversity due to genetic differentiation (D_{ST}) was low in all subpopulations, with the lowest value in PMBE (Figure 7).

The F_{ST} analysis revealed that subpopulations had low genetic differentiation (Table 3), despite of relatively large distance between some subpopulations (up to 89 km). It was verified that geographic distance did not explain the pattern of genetic distance between subpopulations, as they were not significantly correlated ($R^2 = 0.08$, $p = 0.43$). In the network analysis at the population level, sharing of the gene pool was observed, reflecting the connection between subpopulations (Figure 1), and a substructure between individuals, forming two groups (gray and green) that share alleles between individuals from different subpopulations (Figure 8).

4. Discussion

We demonstrated a gene pool sharing among forest remnants by a insect-pollinated, animal-dispersed widely distributed tree in the Atlantic Forest. This knowledge can be the key in the recovery process of degraded areas, mainly in the protected areas (PAs). This indicates the relevance of integrating such important remnants in Atlantic Forest management plans, as suggested by ecological corridors project in that region. Considering the difficulties of managing protected areas in Brazil due to the lack of knowledge about biodiversity in the PAs (Oliveira *et al*., 2017), conservation genetics can help public managers make decisions to maximize the potential of PAs in maintaining diversity and in genetic connection between these areas and with the surrounding forest remnants (Torres-Florez *et al*., 2017).

The new set of nSSR markers developed for *E. ovata* showed sufficient polymorphism for individual identification and kinship analysis, providing an excellent molecular tool for population genetic studies. In addition, we demonstrated that the 15 individuals sampled and genotyped with 13 loci were sufficient to represent the number of alleles and frequency of the informative alleles and genetic diversity of each of the four subpopulations. This small number of individuals may have been sufficient for population genetic analysis, due to the strategy of sampling distant individuals geographically within each population. Another important factor is that loci with high heterozygosity and with similar distribution of the allelic frequency may enable genetic studies with reduced sample size (Hale *et al*., 2012). In this regard, this set of tools will be useful in population studies, allowing access to information about the genetic variability of *E. ovata*; in addition, they can be used to develop alternative strategies for the management and conservation of tree species in the Atlantic Rainforest.

When genetic diversity was accessed with the cpSSR marker, it revealed that the five subpopulations sampled have few haplotypes, with large proportion of individuals

(80%) with only four haplotypes. This finding suggests that these subpopulations of *E. ovata* in the ecological corridor of the Atlantic Forest in southern Bahia were mainly founded by four maternal lineages, reflecting a low variety of seed sources and a strong founding effect or a recent genetic bottleneck (Parducci *et al*., 2001; Logossa *et al*., 2011; Li *et al*., 2012; Wang, 2013). On the other hand, the presence of unique haplotypes in the different subpopulations reflects the importance of the areas for the maintenance of genetic variability, especially the integral protection area (PMBE and ReBio) that retains six of the eight unique haplotypes.

The intrapopulational analysis showed that *E. ovata* subpopulations have a low average haplotypic diversity compared to those of other plants species (Petit *et al*., 2005). In addition, these populations have a low number of effective haplotypes and a low haplotypic richness, reflecting the low genetic distance found among individuals (Wang *et al*., 2011; Wójkiewicz and Wachowiak, 2016). These results may be a consequence of the narrow genetic base of the individuals that founded the subpopulations or resulting from a genetic bottleneck, reinforcing the importance of maintaining functional connectivity through gene flow between these subpopulations, since they may be more sensitive to external disturbances and more prone to genetic drift (Parducci et al., 2001; Martins et al., 2011; Tong et al., 2013). In this context, taking into account the critical state of the fragmentation and reduction of the Atlantic Forest (Ribeiro et al., 2009), one way to maintain contemporary gene flow is with the management of regional land use to produce a heterogeneous mosaic (e. g: Through of ecological corridors or step-stones between forest remnants) (Fonseca *et al*., 2004; Damschen *et al*., 2006; Brudvig *et al*., 2009).

The four subpopulations evaluated with the nSSR loci had genetic diversity (H_E and H_O) estimates similar to those found in rare and overexploited palm and tree species

from the Atlantic Forest in southern Bahia, such as *Euterpe edulis* Mart. (H_E = 0.64 and H_O = 0.58) and *Melanoxylon brauna* Schott (H_E = 0.57 and H_O = 0.53) (Borges *et al*., 2014; Santos *et al*., 2015). However, these diversity indexes are low compared to those of other, more abundant species in the same environment, such as *Licania hypoleuca* Benth. H_E = 0.84 and H_O = 0.71(França *et al*., 2015). In addition, it is also important to emphasize that the H_O of *E. ovata* was larger than the H_E, indicating a possible imbalance between the evolutionary forces, such as inbreeding or genetic bottleneck. Notably, the values of inbreeding (*f*) were negative in all subpopulations, which could be a strong indicator of preferential crosses between unrelated individuals. On the other hand, we used the Wilcoxon test to demonstrate that this excess in H_O was also a consequence of the genetic bottleneck in *E. ovata* subpopulations. Once that, the genetic bottleneck may develop an excess of transient heterozygosity, after a recent change in effective population size, when heterozygotes have a selective advantage, resulting in a higher H_O than H_E (Cornuet and Luikartt, 1996). The four subpopulations evaluated showed excess heterozygosity as a consequence of genetic bottleneck, as revealed by the infinite allele model (IAM). As demonstrated by Cornuet and Luikartt, (1996), the IAM is a more sensitive model for detecting the genetic bottleneck than the SMM and TPM models, indicating that the nSSR loci of *E. ovata* satisfied the requirements of the infinite allele model.

The results obtained with the two sets of markers (nSSR and cpSSR) demonstrated convergence, with the results of AMOVA (nSSRs) and haplotype diversity analysis (cpSSR) revealing that the diversity within the subpopulations contributed more to the overall composition than the genetic variation among the subpopulations. This result is probably the consequence of allele sharing and the ancestral haplotype lineages that founded these populations (Wang, 2013). In addition,

we highlight that areas of integral protection may play an important role in maintaining genetic variability. Since human interference is prohibited in these areas, it may have favored the maintenance of unique haplotypes and alleles (Bruner *et al.*, 2001). Another interesting point to emphasize is that despite the little difference, it was possible to observe that the areas of integral protection (ReBio and PMBE), possess the largest genetic indices in comparison with the other areas. The lowest genetic indexes observed in the two sustainable use PAs may reflect the selective logging of *E. ovata* in RPPN Cap and the forest reduction in RPPN MM, with both anthropogenic impacts occurring before the implementation of the PAs (personal communication from the land owners). In this sense, considering that genetic variability plays an important role in natural selection, maintaining the ability of populations to adapt to their habitats over time (Leimu and Fischer, 2008). No matter how small the loss of genetic variability, it requires attention, especially in the subpopulation of Restinga (AR) that is outside conservation unit and, consequently, more susceptible to human impacts and loss of alleles and haplotypes.

The genetic structure analysis (F_{ST}) revealed that the studied subpopulations had very low genetic structure when analyzed with both markers, and this genetic pattern is not influenced by geographic distance. However, the discriminant analysis of principal components (DAPC) revealed that the four collection areas form three genetic clusters. But, it is worth noting that many individuals from an a priori subpopulation are inserted into eclipses of other subpopulations and there are overlaps of eclipses between subpopulations, which corroborates the observed low Fst values. We also added that there is a great mixture of the gene pool in the individuals of the different clusters, reinforcing the results obtained in the F_{ST} analysis and showing that the genetic diversity is mainly contained among the individuals, as indicated by AMOVA. Another

important point is that the network analysis showed substructure between individuals, forming two groups (green and gray), formed by individuals from different subpopulations. In addition, it is important to highlight the connections between the individuals of the two groups, showing that although they belong to a certain group (green or gray), the individuals also have genetic similarities with individuals from the other group. Thus, although two groups are observed, they are genetically connected, which may be a reflection of the allele sharing among individuals in these groups. Overall, there is the same pattern in results obtained with the DAPC (nSSR), network analysis (cpSSR) and Fst analyses for both markers converge at the same biological explanation: that, these subpopulations are functionally connected by the dispersion of pollen and seeds. However, it is worth noting that, even with gene pool sharing, it is possible that these subpopulations are having impaired contemporary gene flow because of the fragmentation of the forest in this specific landscape (details on the map). However, additional studies evaluating contemporary gene flow are needed to evaluate this issue. In addition, considering that large infrastructure projects of the Brazilian government, such as the construction of the southern port and the west - east railroad, may negatively impact some of these areas of study (mainly PMBE and Cap), causing suppression and forest fragmentation, and with great potential for disorderly urban expansion. It would be important in the context of conservation that the public power take into account in the regional planning that these remnants of Atlantic Forest have functional connectivity (genetic). In this way, for implement efficient conservation strategies, we recommend that the five studied subpopulations should be considered as a single ecological entity during the regional planning of different land uses.

Acknowledgments

The authors thank Mr. Ronaldo Santana; Instituto de Estudos Socioambientais do Sul da Bahia – IESB; Instituto Chico Mendes de Conservação da Biodiversidade – ICMBio; and the Secretaria de Meio Ambiente de Ilhéus for the sampling licenses. They thank the Fundação de Amparo à Pesquisa do Estado da Bahia (FAPESB) for granting a scholarship to ASS; and Fapesb – Pronex (#PNX0014/2009), CAPES-PNPD (#02736/09-8) for financing the project and granting a scholarship to RT. The authors also thank the Brazilian National Council of Scientific and Technological Development (CNPq) for the FAG research productivity fellowship.

Conflict of Interest

The authors declare no conflicts of interest.

References

Benchimol M, Mariano-neto E, Faria D, Rocha-santos L, Pessoa MS, Sanches F, Custodio D and Cazetta E (2017) Translating plant community responses to habitat loss into conservation practices Forest cover matters. Biol Conserv209:499–507.doi: 10.1016/j.biocon.201703024.

Borges DB, Mariano-Neto E and Gaiotto FA (2014) Development of microsatellite primers for Melanoxylon brauna (Fabaceae): an endangered and endemic tree from the Brazilian Atlantic Forest. Conserv Genet Resour. doi: 10.1007/s12686-014-0288-3.

Brudvig LA, Damschen EI, Tewksbury JJ, Haddad NM and Levey DJ (2009) Landscape connectivity promotes plant biodiversity spillover into non-target

habitats. Proc Natl Acad Sci 106:9328–9332.

Bruner AG, Gullison RE, Rice RE and Fonseca GAB (2001) Effectiveness of Parks in Protecting Tropical Biodiversity. 291:125–128.

Carvalho CS, Ribeiro MC, Côrtes MC, Galetti M and Collevatti RG (2015) Contemporary and historic factors influence differently genetic differentiation and diversity in a tropical palm Contemporary and historic factors in fl uence differently genetic differentiation and diversity in a tropical palm. Heredity (Edinb) 115:216–224. doi: 10.1038/hdy.2015.30.

Carvalho CS, Ballesteros-Mejia L, Ribeiro MC, Côrtes MC, Santos AS and Collevatti RG (2017) Climatic stability and contemporary human impacts affect the genetic diversity and conservation status of a tropical palm in the Atlantic Forest of Brazil. Conservation Genetics, 18(2), 467-478.

Cornuet JM and Luikartt G (1996) Description and Power Analysis of Two Tests for Detecting Recent Population Bottlenecks From Allele Frequency Data. Genetics 144:2001–2014.

Corriveau JL and Annette WC (1988) Rapid screening method to detect potential biparental inheritance of plastid DNA and results for over 200 angiosperm species. American Journal of Botany 75: 1443-1458.

Damschen EI, Haddad NM, Orrock JL, Tewksbury JJ and Levey DJ (2006) Corridors Increase Plant Species Richness at Large Scales. Science (80-) 313:1284–1286. doi: 10.1126/science.1130098.

Doyle JJ, Doyle JL (1987) Isolation of plant DNA from fresh tissue. Focus 12:13–15.

Eliades N-G and Eliades DG (2009) User ' s Manual H APLOTYPE A NALYSIS Software for Analysis of Haplotype Data.

Fonseca GD, Alger K, Pinto LP, Araújo M and Cavalcanti R (2004) Corredores de biodiversidade: o corredor central da Mata Atlântica. Corredores ecológicos:

uma visão integradora de ecossistemas. Ibama, Ministério do Meio Ambiente, 47-65.

França DDS, Neres J and Gaiotto FA (2015) Development and characterization of 26 microsatellite loci in Licania hypoleuca (Chrysobalanaceae) and transferability to Licania hoehnei. Biochem Syst Ecol 59:204–208. doi: 10.1016/j.bse.2015.01.024.

Gaiotto FA, Grattapaglia D and Vencovsky R (2003) Genetic Structure, Mating System, and Long-Distance Gene Flow in Heart of Palm (Euterpe edulis Mart.). J Hered 94:399–406. doi: 10.1093/jhered/esg087.

Hale ML, Burg TM and Steeves TE (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PloS one, *7*(9), e45170.

Kalinowski ST, ML Taper, TC Marshall (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol. Ecol. 16:1099–1106.

Keenan K, McGinnity P, Cross TF, Crozier WW, and Prodohl PA (2013) diveRsity: An R package for the estimation and exploration of population genetics parameters and their associated errors. Methods in Ecology and Evolution, 4: 782-788. doi:10.1111/2041-210X.12067.

Jombart T, Devillard S and Balloux F (2010) Discriminant analysis of principal components : a new method for the analysis of genetically structured populations. BMC Genet 11:94.

Jombart T, Lyon D and Biome L De (2008) adegenet : a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. doi: 10.1093/bioinformatics/btn129.

Kivelä M, Arnaud-Haond S and Saramäki J (2015) EDENetworks: A User-Friendly Software to Build and Analyze Networks in Biogeography, Ecology and Population Genetics. Mol Ecol Resour 15:117–122. doi: 10.1111/1755-0998.12290.

Leimu R and Fischer M (2008) A meta-analysis of local adaptation in plants. PLoS One 3:1–8. doi: 10.1371/journal.pone.0004010.

Li J, Jin Z and Tan T (2012) Genetic diversity and differentiation of Sinocalycanthus chinensis populations revealed by chloroplast microsatellite (cpSSRs) markers. Biochem Syst Ecol 41:48–54. doi: 10.1016/j.bse.2011.12.019.

Lima MM and Mariano-Neto E (2014) Extinction thresholds for Sapotaceae due to forest cover in Atlantic Forest landscapes. For Ecol Manage 312:260–270. doi: 10.1016/j.foreco.2013.09.003.

Logossa ZA, Camus-Kulandaivelu, A. LÉTIZI, Allal, F, Vaillant A, Sanou H, Kokou K, Bouvet J and Jean-Marc (2011) Molecular data reveal isolation by distance and past population expansion for the shea tree (Vitellaria. Mol Ecol 20:4009–4027. doi: 10.1111/j.1365-294X.2011.05249.x.

Lorenzi H (1998) Árvores brasileiras: manual de identificação e cultivo de plantas arbóreas nativas do Brasil. Nova Odessa: Plantarum, v.2.

Martini AMZ, Fiaschi P, Amorim AM and Paixão JL da (2007) A hot-point within a hot-spot : a high diversity site in Brazil's Atlantic Forest. Biodivers Conserv 16:3111–3128. doi: 10.1007/s10531-007-9166-6.

Martins K, Chaves LJ, Vencovsky R and Kageyama PY (2011) Genetic structure based on nuclear and chloroplast microsatellite loci of Solanum lycocarpum A . St . Hil . (Solanaceae) in Central Brazil. Genet Mol Res 10:665–677. doi: 10.4238/vol10-2gmr1046.

Ministério do Meio Ambiente - MMA (2008) Experiências em implementação de corredores ecológicos /Ministério do Meio Ambiente. Secretaria de Biodiversidade e Florestas. Departamento de Áreas Protegidas. Programa Piloto para a Proteção das Florestas Tropicais do Brasil, Brasília: MMA, 2008.80 p.

Murray-smith C, Brummitt NA, Oliveira-filho ARYT, Bachman S, Moat J, Lughadha EMNIC and Lucas EVEJ (2008) Plant Diversity Hotspots in the Atlantic Coastal Forests of Brazil. Conserv Biol 23:151–163. doi: 10.1111/j.1523-1739.2008.01075.x.

Myers N, Mittermeier RA, Mittermeier CG, Fonseca GAB and Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403.6772 (2000): 853.

Oliveira U, Soares-Filho BS, Paglia AP, Brescovit AD, Carvalho CJ, Silva DP and Stehmann JR (2017) Biodiversity conservation gaps in the Brazilian protected areas. Sci Rep 7:1–9. doi: 10.1038/s41598-017-08707-2.

Pardini R, de Bueno AA, Gardner TA, Prado PI and Metzger JP (2010) Beyond the fragmentation threshold hypothesis: Regime shifts in biodiversity across fragmented landscapes. PLoS One. doi: 10.1371/journal.pone.0013666.

Parducci L, Szmidt AE, Madaghiele A and Vendramin GG (2001) Genetic variation at chloroplast microsatellites (cpSSRs) in Abies nebrodensis (Lojac .) Mattei and three neighboring Abies species. Theor Appl Genet 102:733–740.

Peakall R, Smouse PF (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—An update. Bioinformatics 28:2537–2539. doi:10.1111/j.1471-8286.2005.01155.x.

Pessoa MS, Rocha-Santos L, Talora DC, Faria D and Mariano-Neto E (2016) Fruit biomass availability along a forest cover gradient. Biotropica 49:45–55. doi: 10.1111/btp.12359.

Petit RJ, Duminil J, Fineschi S, Hampe A, Salvini D and Vendramin GG (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol 14:689–701. doi: 10.1111/j.1365-294X.2004.02410.x.

Piry S, Luikart G and Cornuet J-M (1999) Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. J Hered 90:502–503. doi: 10.1093/jhered/90.4.502.

Ribeiro MC, Metzger JP, Martensen AC, Ponzoni FJ and Hirota MM (2009) The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. Biol Conserv 142:1141–1153. doi: 10.1016/j.biocon.2009.02.021.

Rocha-santos L, Pessoa MS, Cassano CR, Talora DC, Orihuela RLL, Mariano-neto E, Morante-filho JC, Faria D and Cazetta E (2016) The shrinkage of a forest : Landscape-scale deforestation leading to overall changes in local forest structure. Biol Conserv J 196:1–9. doi: 10.1016/j.biocon.2016.01.028.

Santos AS, Cazetta E, Dodonov P, Faria D and Gaiotto FA (2016) Landscape-scale deforestation decreases gene flow distance of a keystone tropical palm, Euterpe edulis Mart (Arecaceae). Ecol Evol 6:6586–6598. doi: 10.1002/ece3.2341.

Santos AS, Cazetta E, Morante Filho JC, Baumgarten J, Faria D and Gaiotto FA (2015) Lessons from a palm: genetic diversity and structure in anthropogenic landscapes from Atlantic Forest, Brazil. Conserv Genet 16:1295–1302. doi: 10.1007/s10592-015-0740-2.

Tabarelli M, Cardoso Da Silva JM and Gascon C (2004) Forest fragmentation, synergisms and the impoverishment of neotropical forests. Biodivers Conserv 13:1419–1425. doi: 10.1023/B:BIOC.0000019398.36045.1b.

Thomas WW, Garrison J and Arbela AL (1998) Plant endemism in two forests in southern Bahia , Brazil. Biodivers Conserv 7:311–322.

Tong Y, Wu C and Gao L (2013) Characterization of chloroplast microsatellite loci from whole chloroplast genome of Camellia taliensis and their utilization for evaluating genetic diversity of Camellia reticulata (Theaceae). Biochem Syst Ecol 50:207–211. doi: 10.1016/j.bse.2013.04.003.

Torres-Florez JP, Johnson WE, Nery MF, Eizirik E, Oliveira-Miranda MA and Jr PMG (2017) The coming of age of conservation genetics in Latin America : what has been achieved and what needs to be done. Conserv Genet 1–15. doi: 10.1007/s10592-017-1006-y.

Tóth EG, Vendramin GG, Bagnoli F, Cseke K and Höhn M (2017) High genetic diversity and distinct origin of recently fragmented Scots pine (Pinus sylvestris L .) populations along the Carpathians and the Pannonian Basin. Tree Genet Genomes. doi: 10.1007/s11295-017-1137-9.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R and Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic acids research, 35(suppl_2), W71-W74.

Wang Y (2013) Chloroplast microsatellite diversity of Opisthopappus Shih (Asteraceae) endemic to China. Plant Syst Evol 299:1849–1858. doi: 10.1007/s00606-013-0840-8.

Wang Y, Guo J and Zhao G (2011) Chloroplast microsatellite diversity of Clintonia udensis (Liliaceae) populations in East Asia. Biochem Syst Ecol 39:22–30. doi: 10.1016/j.bse.2010.12.015.

Weising K and Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of

dicotyledonous angiosperms. Genome 42:9–19.

Wickham H (2010) ggplot2: elegant graphics for data analysis. *J Stat Softw*, *35*(1), 65-88.

Wójkiewicz B and Wachowiak W (2016) Substructuring of Scots pine in Europe based on polymorphism at chloroplast microsatellite loci. Flora 220:142–149.

Zhang X, Shi M, Shen D and Chen X (2012) Habitat Loss other than Fragmentation per se Decreased Nuclear and Chloroplast Genetic Diversity in a Monoecious Tree. PLoS One 7:e39146. doi: 10.1371/journal.pone.0039146.

LIST OF TABLES AND FIGURES

Table 1: Wilcoxon test for *E. ovata* subpopulations with nSSR markers.

Supopulation	IAM	TPM	SMM
AR	0.04016*	0.55371	0.77258
Rebio	0.00012*	0.00201*	0.05493*
MM	0.00153*	0.07324	0.34241
Cap	0.00085*	0.07324	0.24866

Note: AR= Restinga; ReBio= Reserva Biológica de Una; MM= RPPN Mãe da Mata; PMBE= Parque Municipal Boa Esperança; Cap= RPPN capitão; Infinite Allele Model (IAM), Stepwise Mutation Model (SMM), and Two Phase Model (TPM). *represents significance at alpha ≤ 0.05.

Table 2: *E. ovata* subpopulation genetic parameters based on the cpSSR markers.

Population	PA	AH	N	N_H	P_H	H_{Ne}	H_R	H_E	D^2sh
AR	Not	-	15	4	1	2.273	3	0.600	0.163
MM	Sustainable use	13	15	5	0	1.800	4	0.476	0.147
CAP	Sustainable use	660	15	3	1	1.316	2	0.257	0.041
ReBio	Integral protection	11,400	15	6	2	2.528	5	0.648	0.463
PMBE	Integral protection	437	15	7	4	3.689	6	0.781	2.707
Mean	-	-	15	5	1.6	2.321	4	0.552	0.704

Note: PA= Protected area; AH = Amount of hectares; N= Number of individuals per population; N_H= Number of haplotypes; P_H= Number of unique haplotypes; H_{Ne}= Number of effective haplotypes; H_R= Haplotypes richness; H_E= Haplotype diversity;

D^2sh= Average genetic distance between individuals; AR= Restinga; ReBio= Reserva Biológica de Una; MM= RPPN Mãe da Mata; PMBE= Parque Municipal Boa Esperança; Cap= RPPN capitão.

Table 3: Genetic structure analysis (F_{ST}) with cpSSR markers between areas in pairs in the lower diagonal and geographic distance in upper diagonal (Km).

	AR	ReBio	MM	Cap	PMBE
AR	-	24	16	67	21
ReBio	0.046	-	37	89	43
MM	0.060	0.052	-	52	6
Cap	0.013	0.079	0.092	-	46
PMBE	0.005	0.035	0.036	0.015	-

Note: AR= Restinga; ReBio= Reserva Biológica de Una; MM= RPPN Mãe da Mata; PMBE= Parque Municipal Boa Esperança; Cap= RPPN capitão.

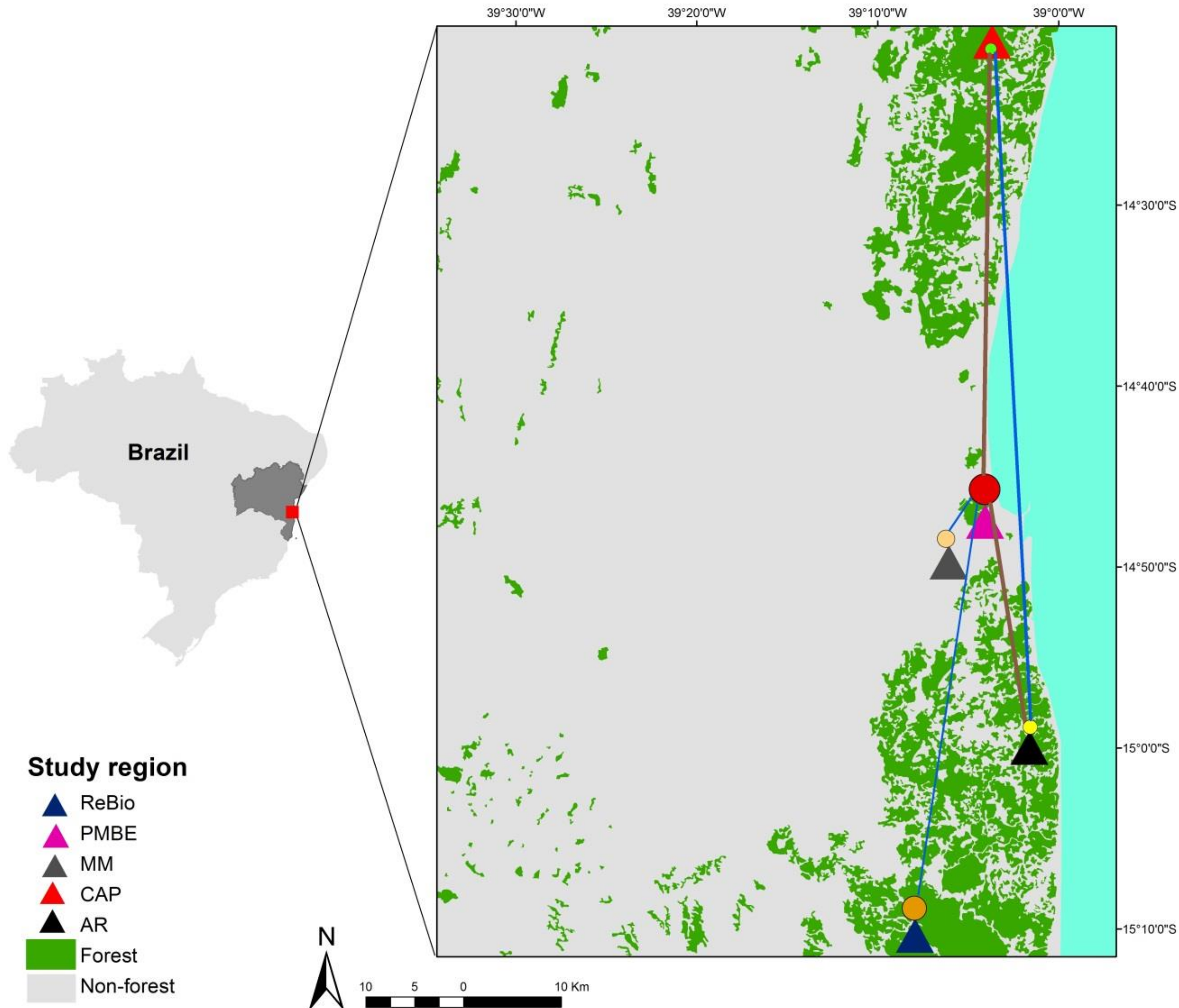


Figure 1 Map of Brazil. The state of Bahia is highlighted, indicating the region of the study, with the current forest cover. Blue and brown lines show network linkages identified by EDENetworks (Kivelä *et al*., 2015) between nodes (sampling sites), using cpSSR markers. Line thickness is proportional to linkage strength and node size is proportional to the number of haplotypes for each sampling sites de *E. ovata*. Map data: Atlas of the Atlantic Forest Forest Remnants of the year 2016, obtained from SOS Mata Atlântica.

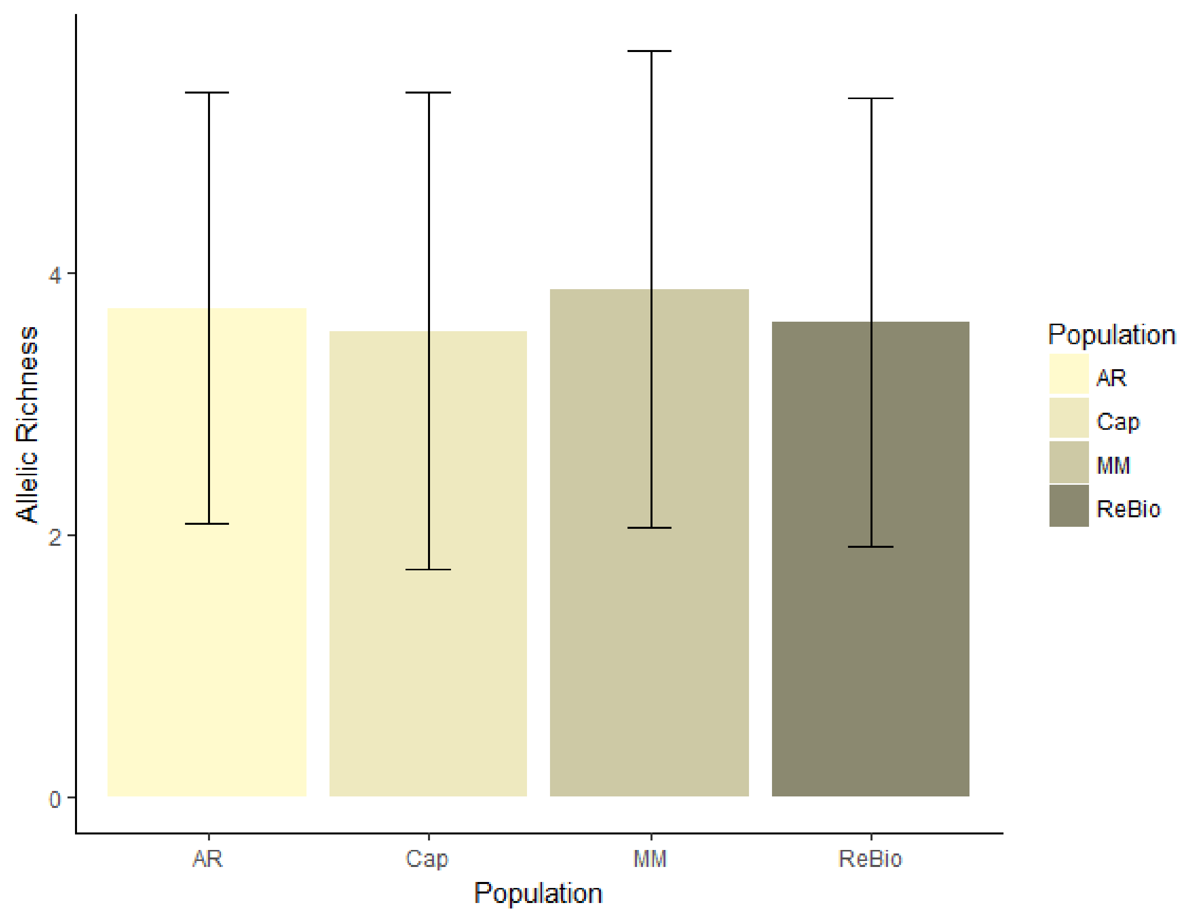


Figure 2 Allelic richness with standard deviation values for *E. ovata* subpopulations with nSSR markers.

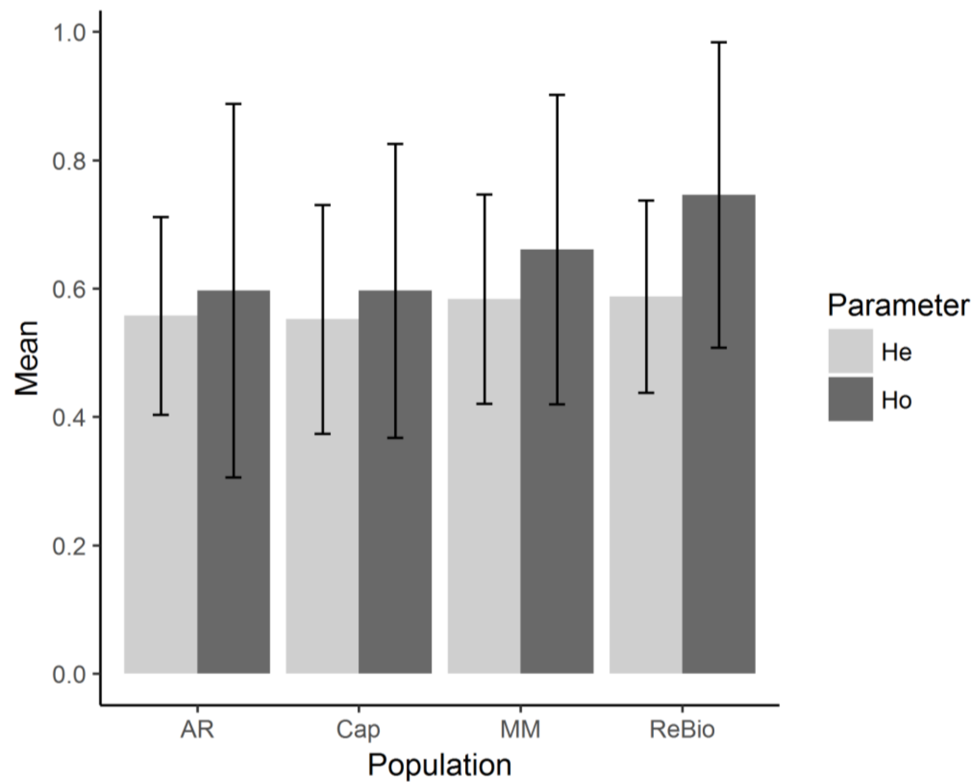


Figure 3 Observed heterozygosity (Ho) and expected heterozygosity (He) with standard deviation values for *E. ovata* subpopulations with nSSR markers.

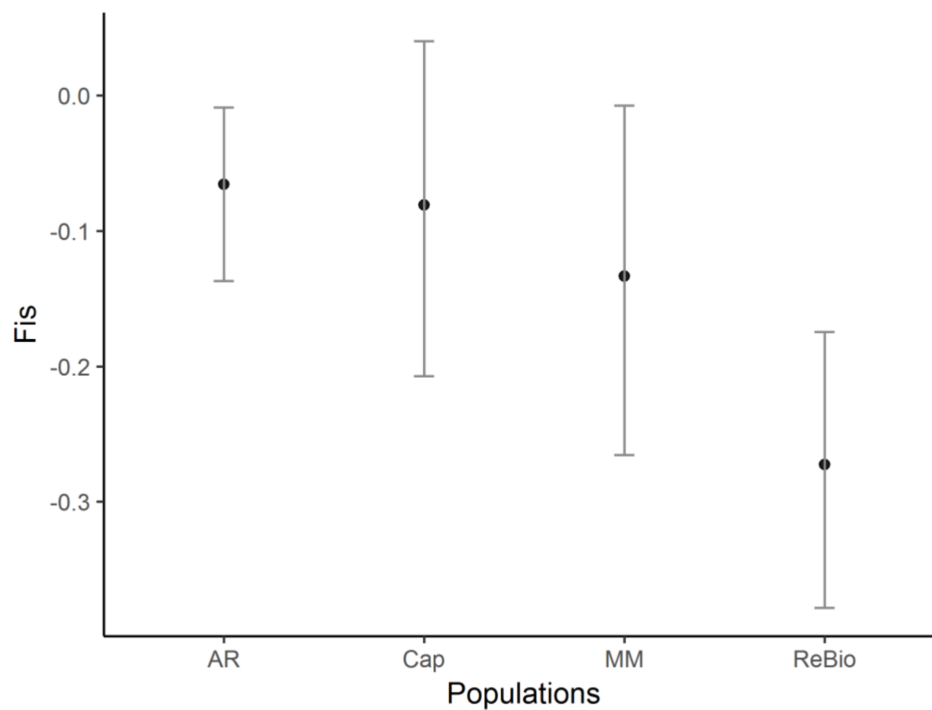


Figure 4 Fixation index (Fis) estimated values with nSSR markers and 95 % confidence intervals based on 10.000 bootstrap values for *E. ovata* subpopulations.

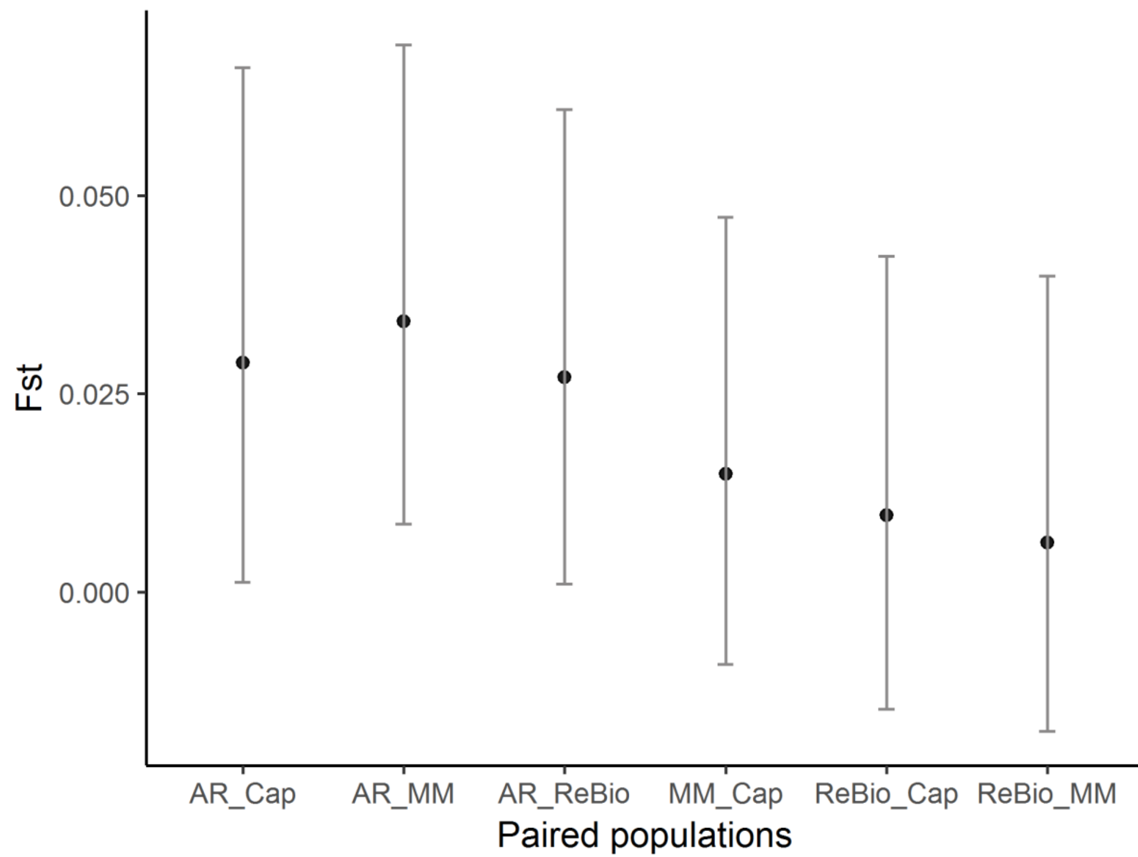


Figure 5 F_{ST} estimated values with nSSR markers and 95 % confidence intervals based on 10.000 bootstrap values of *E. ovata* sampled in four subpopulations.

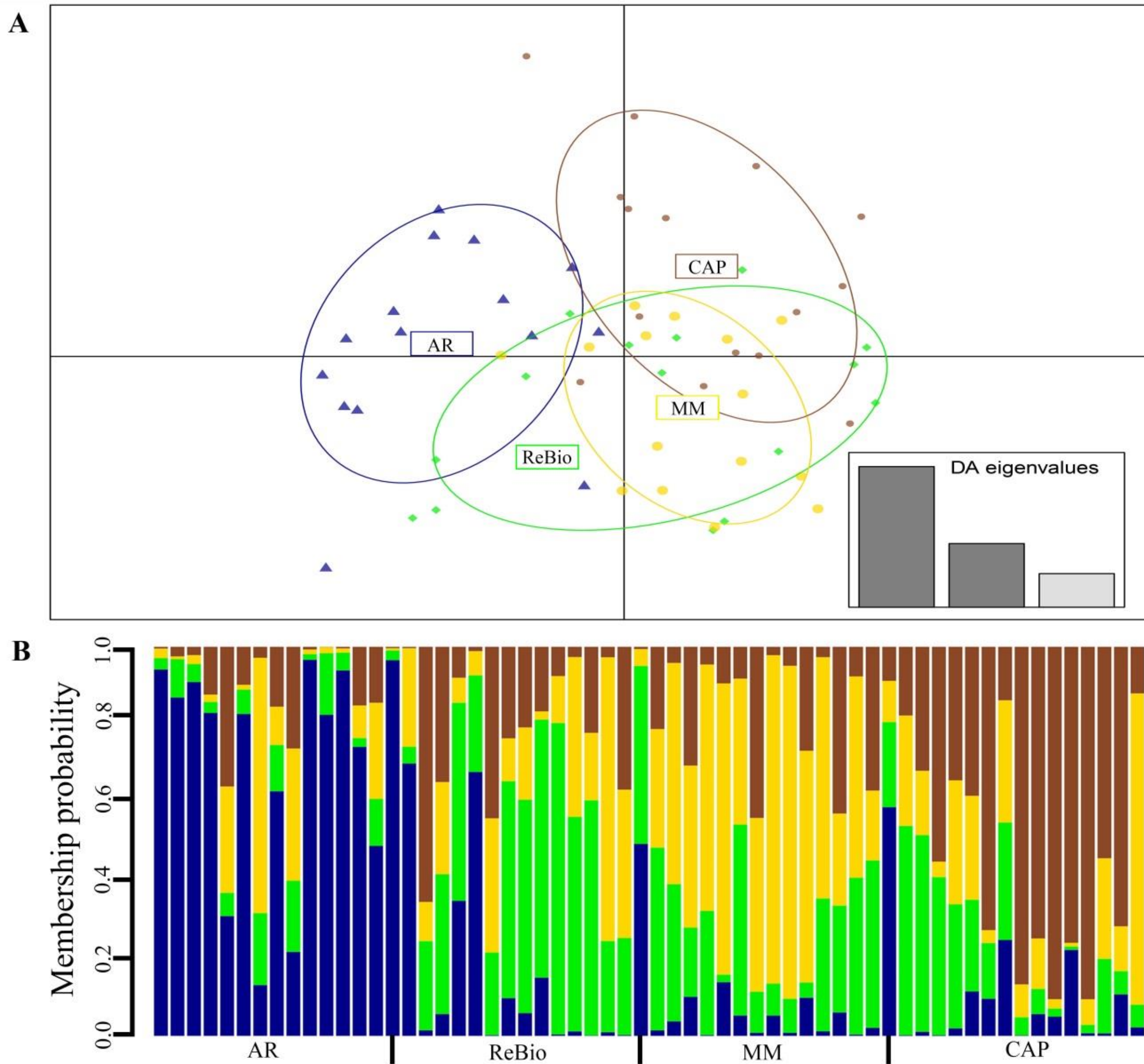


Figure 6 Genetic differentiation among the four *E. ovata* subpopulations based on nSSR markers. A) The graph represents the individuals as dots and the subpopulations as ellipses according to their original groups, which model 95% of the corresponding variability plotted. The subpopulations (ellipses) are plotted within the orthogonal space defined by the first two eigenvalues of the PCA (inserts). B) Diagram representing the gene pool of the individuals associated with each of the subpopulations. Each vertical bar represents an individual with the height of the column segments showing the

probability of being assigned to one of the four subpopulations. Dark lines correspond to the collection location.

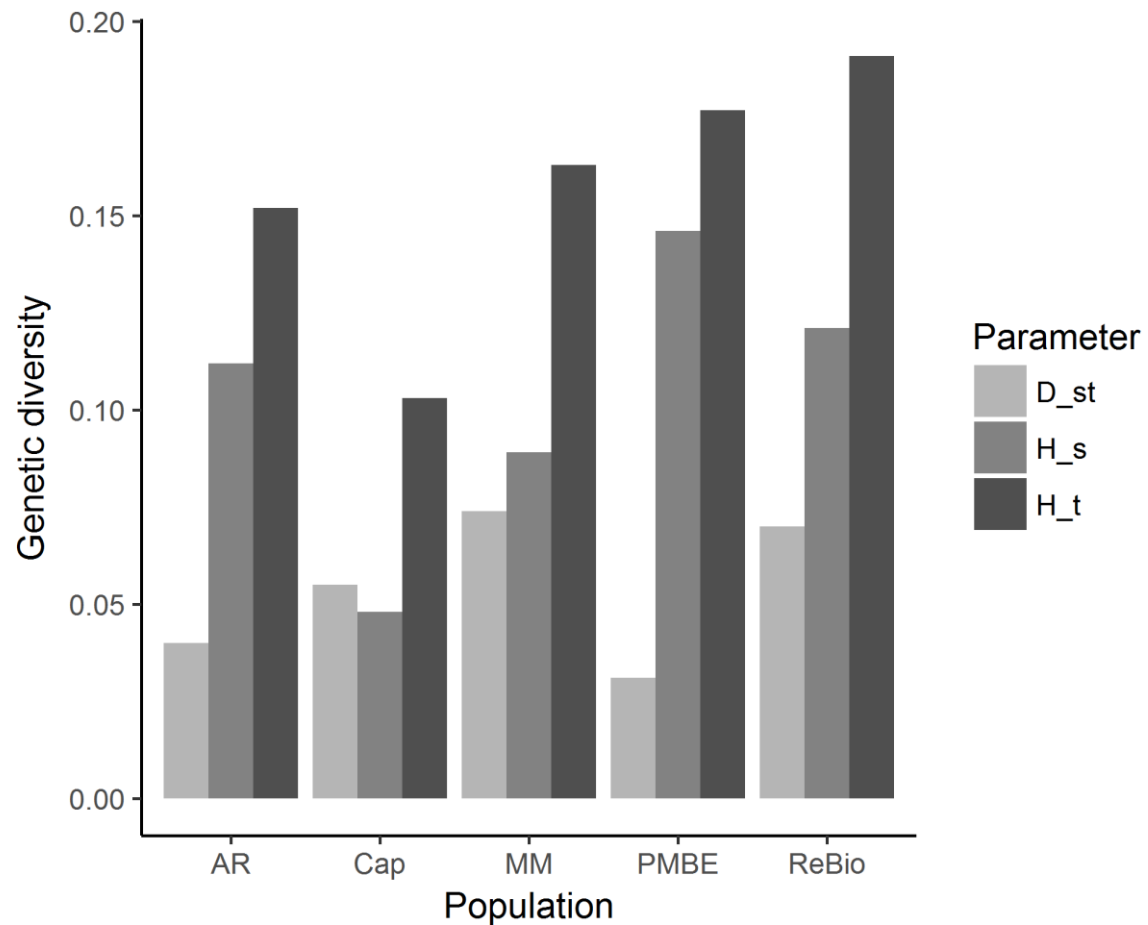


Figure 7 Genetic diversity partitioned from the *E. ovata* subpopulations with cpSSR markers. Note: H_S = Genetic diversity within the population; D_{ST} = Diversity due to genetic differentiation; H_T = Total genetic diversity.

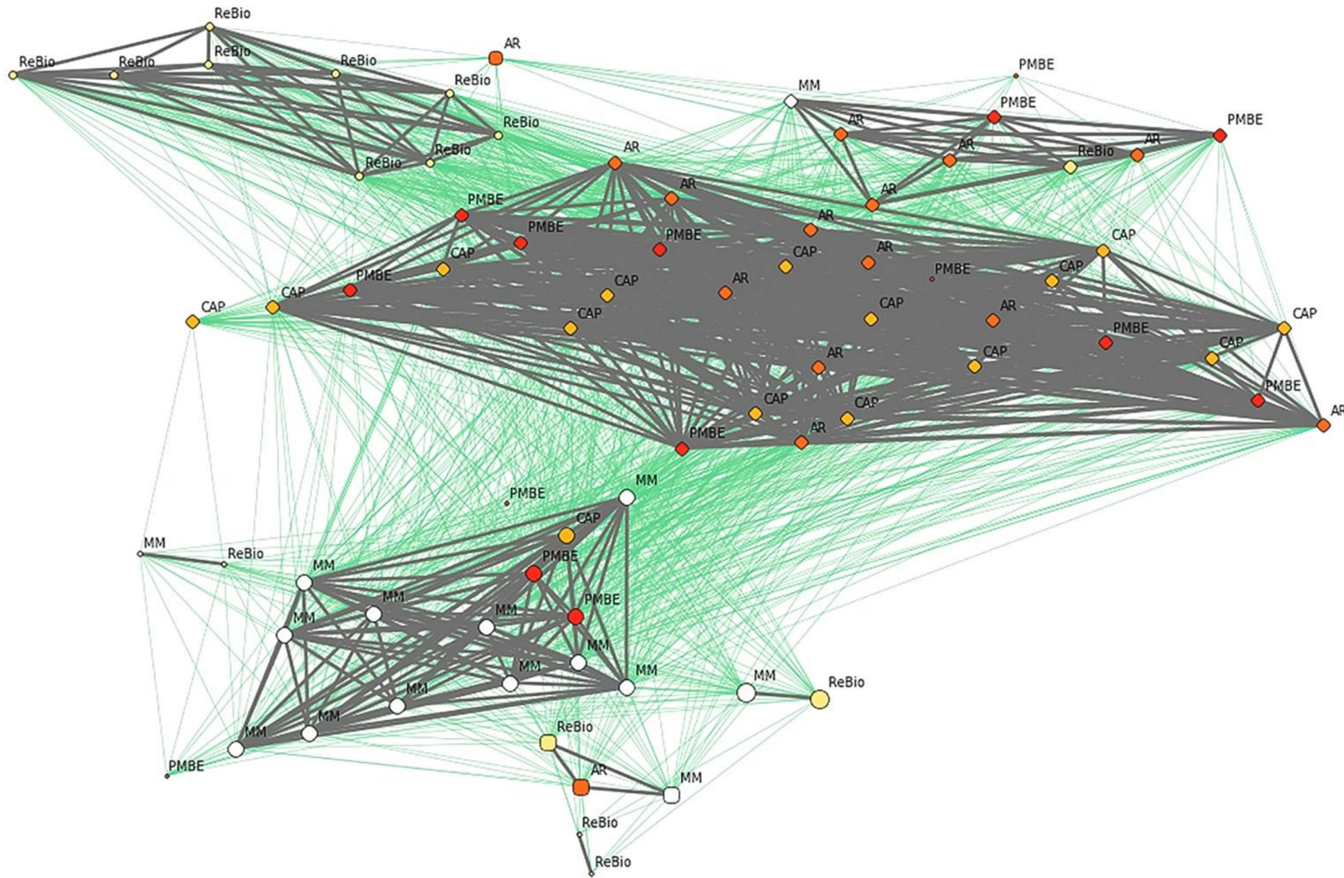


Figure 8 Network analysis among the individuals of *E. ovata* with cpSSR markers. Grey and green lines show network linkages identified by EDENetworks (Kivelä *et al*., 2015) between nodes (individuals). The colors of the nodes represent the sampling sites, the line thickness is proportional to linkage strength and node size is proportional to the number of linkages for each node.

APÊNDICE

Table S1: Characterization of the new 13 microsatellite loci specific to *E. ovata*.

Loci	Primer sequence (5'-3')	T (ºC)	Allele size range	Na	EHW
EO04	F: TTAgAgTTggTgAgTgCATATT	58º	120-190	12	ns
	R:CAgCAACCTCTAgCATACTgT				
EO07	F: TgTgCTTACgCgTggACTAC	48º	130-190	5	ns
	R: AgACCgAgAgCTACAggATCA				
EO11	F: AgTTgTCTgAAACCATCACT	48º	300-400	2	ns
	R: CAACACAgCCCTACAAA				
EO16	F: CgCTgCTATCAACgAgACTg	58º	170-230	2	ns
	R: TgCTATCTgAggCTgACACg				
EO24	F: TAgACTCCggCATgATAC	58º	160-230	3	ns
	R: AACTCAgTgAATAACAATACAgT				
EO25	F: gAACACTgCTggAgTgATTATT	58º	110-160	4	ns
	R: gTgAgTgCATTCTTCTgCAA				
EO26	F: CTTggAACCTgACCAgCAgT	58º	260-320	8	ns
	R: TCAACAgAACCgATTCACACA				
EO29	F: TTCTgACTCTggTgTTgATT	48º	240-310	8	ns
	R: CACAATTAgAggCAgAATCC				
EO31	F: ATggATAATTCCTCCgATgg	58º	170-270	4	ns
	R: ACgCATgCATgATgAgATAC				
EO39	F: CAgCgAggAATggACgATA	58º	220-270	4	ns
	R: TAgAgCAgAgCCTgATCgTg				
EO40	F: AgCATgATACgTTCTgTTggAg	58º	200-260	5	ns

R: gCACgAgAACCgAgTCTgTA				
EO47 :TTCTgTTAAgCTCTACATgTCTgATT	60º	250-330	3	ns
R: CATCTgTCTgTCggCTgAg				
EO63 F: TTCCTgAAATCTggTTTCgTTT	58º	270-340	13	ns
R: CgCCgTATCAgAgCCATAAT				

Note: T (ºC) = Temperature annealing; Na = Number of alleles; EHW= Hardy–Weinberg equilibrium; ns=Not significant.

Table S2: Characterization of 13 microsatellite loci in four subpopulations of *E. ovata*.

	AR					
Loci	A	H_O	H_E	Q	I	F
EO04	7	0.286	0.841	0.808	0.062	0.4763
EO07	5	0.733	0.674	0.619	0.161	-0.0861
EO11	2	0.267	0.239	0.174	0.618	-0.0672
EO16	2	1	0.517	0.281	0.375	-0.333
EO24	5	0.467	0.503	0.482	0.287	-0.0187
EO25	4	0.6	0.469	0.362	0.355	-0.1633
EO26	5	0.533	0.584	0.45	0.264	0.0504
EO29	4	0.133	0.605	0.447	0.257	0.6343

EO31	2	0.467	0.434	0.255	0.425	-0.0526
EO39	3	1	0.646	0.492	0.212	-0.2682
EO40	4	1	0.683	0.545	0.182	-0.2211
EO47	3	0.8	0.522	0.334	0.343	-0.2383
EO63	10	0.467	0.811	0.812	0.066	0.2722

	ReBio					
Loci	A	H_O	H_E	Q	I	F
EO04	7	0.714	0.878	0.857	0.043	0.0846
EO07	5	0.933	0.68	0.636	0.152	-0.2352
EO11	2	0.533	0.405	0.245	0.447	-0.1534
EO16	2	1	0.517	0.281	0.375	-0.333
EO24	4	0.867	0.618	0.511	0.217	-0.2281
EO25	2	0.333	0.37	0.232	0.476	0.0354
EO26	4	0.667	0.637	0.563	0.191	-0.0614
EO29	5	0.733	0.749	0.674	0.121	-0.0321
EO31	3	0.667	0.48	0.32	0.368	-0.1914
EO39	2	0.933	0.515	0.281	0.376	-0.3042
EO40	4	1	0.756	0.655	0.122	-0.1605

EO47	2	1	0.517	0.281	0.375	-0.333
EO63	8	0.333	0.775	0.734	0.097	0.404

	MM					
Loci	A	H_O	H_E	Q	I	F
EO04	9	0.6	0.874	0.878	0.037	0.1767
EO07	5	0.8	0.536	0.561	0.204	-0.2046
EO11	2	0.267	0.198	0.174	0.618	-0.0672
EO16	2	1	0.519	0.281	0.375	-0.333
EO24	4	0.667	0.628	0.552	0.197	-0.0264
EO25	3	0.467	0.362	0.385	0.302	0.0674
EO26	6	0.5	0.742	0.673	0.126	0.1564
EO29	5	0.733	0.655	0.536	0.2	-0.0831
EO31	4	0.5	0.423	0.396	0.34	-0.0385
EO39	2	1	0.537	0.281	0.375	-0.333
EO40	5	0.933	0.712	0.629	0.144	-0.1609
EO47	2	0.8	0.476	0.275	0.386	-0.2499
EO63	7	0.333	0.785	0.804	0.066	0.4069

	Cap					

Loci	A	H_O	H_E	Q	I	F
EO04	10	0.533	0.874	0.868	0.041	0.2281
EO07	3	0.600	0.536	0.403	0.299	-0.0967
EO11	2	0.214	0.198	0.151	0.672	-0.0507
EO16	2	1.000	0.519	0.281	0.375	-0.3330
EO24	3	0.667	0.628	0.482	0.223	-0.0449
EO25	3	0.286	0.362	0.275	0.464	0.0922
EO26	5	0.692	0.742	0.663	0.127	0.0053
EO29	6	0.533	0.655	0.582	0.183	0.0959
EO31	2	0.571	0.423	0.251	0.434	-0.1664
EO39	3	0.857	0.537	0.341	0.334	-0.2586
EO40	4	0.786	0.712	0.598	0.156	-0.0720
EO47	2	0.714	0.476	0.268	0.398	-0.2173
EO63	6	0.308	0.785	0.731	0.095	0.4158

Note: AR= Restinga; ReBio= Reserva Biológica de Una.; MM= RPPN Mãe da Mata; Cap= RPPN Capitão; A= Number of alleles, H_O= observed heterozygosity, H_E= expected heterozygosity, Q= probability of paternity exclusion, i= identity index; F= fixation index.

Table S3: Frequency of *E. ovata* haplotypes in the five study areas.

Haplotypes	**[2]Combination**	**NI**	**AR**	**ReBio**	**MM**	**PMBE**	**Cap**
H1	222 159 130 124	1				0.067	
H2	227 159 125 124	2		0.067	0.067		
H3	227 159 126 124	9		0.600			
H4	227 160 125 124	2		0.133			
H5	227 160 126 124	1	0.067				
H6	228 159 125 124	14			0.733	0.133	0.067
H7	228 159 125 125	2		0.067	0.067		
H8	228 159 125 127	1				0.067	
H9	228 159 126 123	1				0.067	
H10	228 159 126 124	29	0.600			0.467	0.867
H11	228 159 126 125	1					0.067
H12	228 160 125 124	3	0.067	0.067	0.067		
H13	228 160 126 124	8	0.267	0.067	0.067	0.133	
H14	228 160 126 125	1				0.067	
Total		75	1.000	1.000	1.000	1.000	1.000

Note: AR= Restinga; ReBio= Reserva Biológica de Una; MM= RPPN Mãe da Mata; PMBE= Parque Municipal Boa Esperança; Cap= RPPN capitão; [2]Combination = combination of alleles in the polymorphic cpSSR; NI= Number of individuals with the haplotype.

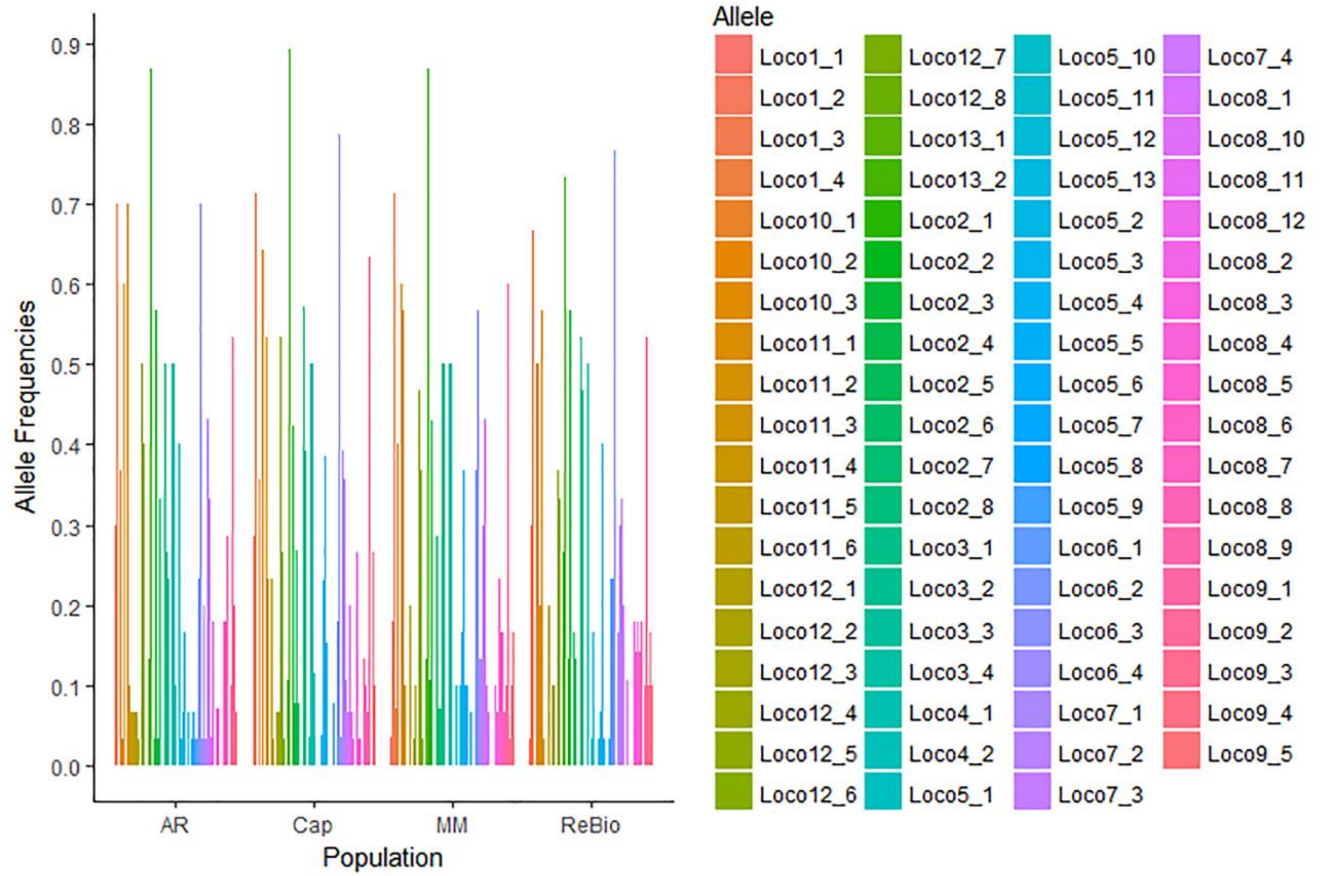


Figure S1 Allele frequencies of the 13 nSSR loci in four subpopulations of *E. ovata*.

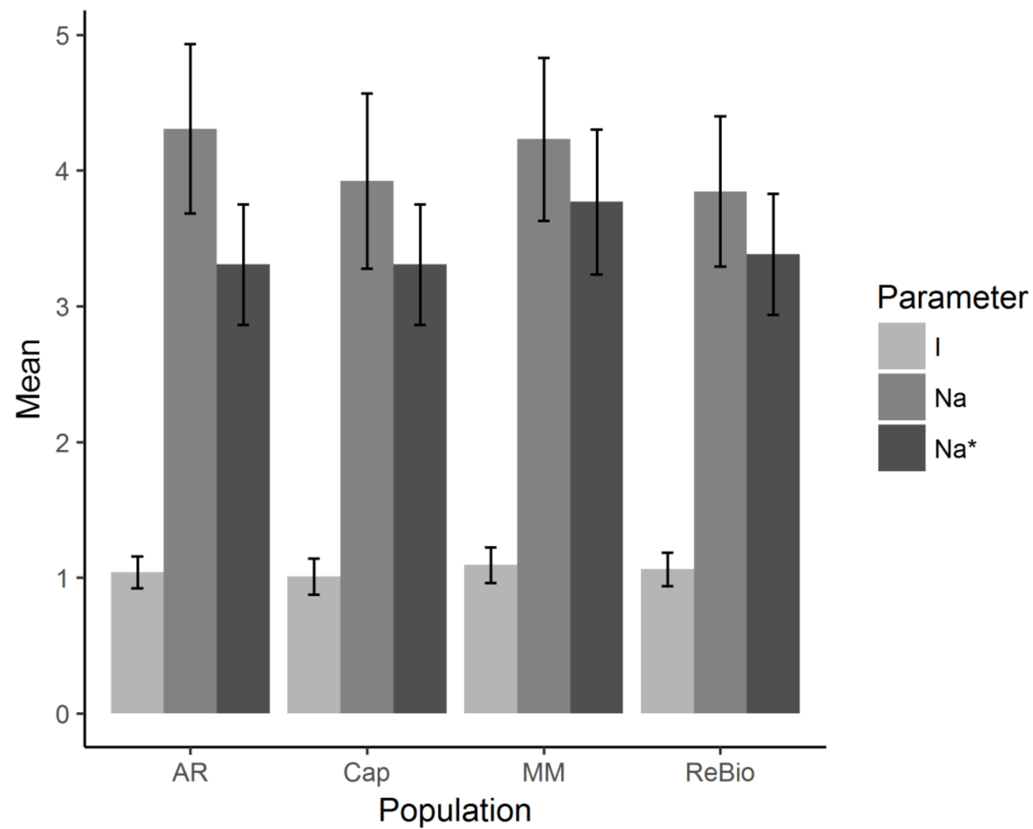


Figure S2 Mean allelic pattern with standard deviation for the 13 nSSR loci in four populations of *E. ovata*. Note: I = Shannon's information index = -1* Sum (pi * Ln (pi)); Na = Nº of different alleles; Na* = Nº of different alleles with a Frequency >= 5%.

CONCLUSÃO GERAL DA TESE

A revisão de literatura de genômica da paisagem com plantas silvestres evidenciou que o número de indivíduos e de SNPs utilizados são de fundamental importância na detecção de sinais de seleção natural nas populações. Além disso, cerca de 81% dos estudos foram realizados nos biomas de florestas temperadas e mistas dos habitats europeus e norte-americanos. Assim, sugerimos que para popularizar o uso dessa abordagem em estudos futuros em florestas tropicais, a melhor relação custo benefício seria o sequenciamento em pool de amostras.

Nessa tese, realizamos pela primeira vez um estudo de genômica em escala de paisagem para uma espécie nativa da Mata Atlântica Brasileira, avaliando 23 populações de ***Euterpe edulis*** **Mart (Arecaceae)** com sequenciamento em pool de amostras, utilizando 7.490 SNPs localizados em regiões transcritas. Essas populações apresentaram alta diversidade gênica, independente da quantidade de floresta remanescente na paisagem ou da riqueza de aves potencialmente dispersoras de sementes. Acreditamos que essa diversidade gênica seja moldada por diversos processos ecológicos e evolutivos simultaneamente, como por exemplo, as divergências bióticas e abióticas entre as regiões de estudo e por conterem diferentes morfotipos da espécie. Assim, esse trabalho inédito representa um importante avanço para conservação de uma espécie ameaçada de extinção e de grande importância ecológica e econômica.

Em um segundo estudo empírico, reportamos pela primeira vez um conjunto de marcadores microssatélites nucleares para ***Eschweilera ovata*** **(Cambess.) Mart. ex Miers (Lecythidaceae)**, uma espécie com importância ecológica e econômica. Nesse trabalho, demonstramos o compartilhamento de pool genético da espécie entre cinco remanescentes florestais por meio dos marcadores microssatélites nucleares e cloroplastidiais. Esse resultado é um indicativo da relevância de integrar esses importantes remanescentes florestais nos planos de manejo da Mata Atlântica, como sugerido pelo projeto de corredores ecológicos na região do estudo.

De maneira geral, os três capítulos apresentados nessa tese proporcionam importantes avanços na área de genética da conservação de plantas tropicais, apresentando novas perspectivas para estudos futuros. Assim, possibilitando uma melhor compreensão dos efeitos das perturbações antrópicas na diversidade genética dessas espécies e propor estratégias eficientes de conservação.